

Alma Mater Studiorum - Università degli Studi di Bologna
Facoltà di Economia
Dipartimento di Scienze Economiche

Dottorato di ricerca in Economia XIX ciclo
Settore Scientifico Disciplinare: SECS-P/05 ECONOMETRIA

Topics in Econometrics of Financial Markets

Relatore: Chiar.mo Prof. Sergio Pastorello Tesi di Dottorato di Ricerca di:
Coordinatore: Chiar.mo Prof. Andrea Ichino Laura Coroneo

Contents

- Introduction** **2**

- 1 A Quantile Regression Approach to Intraday Seasonality** **5**
 - 1.1 Introduction and Motivation 6
 - 1.2 Market and Data 11
 - 1.3 Quantile Regression as Density Regression 12
 - 1.4 Estimation Results 16
 - 1.5 Intraday Value at Risk 18
 - 1.6 Conclusions 24
 - 1.7 Appendix 26

- 2 Forecasting the yield curve using large macroeconomic information** **40**
 - 2.1 Introduction 41
 - 2.2 Model 46
 - 2.2.1 Alternative models 49
 - 2.3 Data 51
 - 2.4 Estimation procedure 53
 - 2.4.1 Model selection 53
 - 2.5 Estimation Results 56
 - 2.6 Out-of-sample forecast 58
 - 2.6.1 Forecast performances 59

2.7	Conclusions	62
2.8	Appendix	63
3	How Arbitrage-Free is the Nelson-Siegel Model?	81
3.1	Introduction	82
3.2	Modeling framework	85
3.2.1	The Nelson-Siegel model	86
3.2.2	Gaussian arbitrage-free models	88
3.2.3	Motivation	90
3.3	Data	92
3.4	Estimation Procedure	93
3.4.1	Resampling procedure	94
3.5	Results	96
3.5.1	Testing results	96
3.5.2	In-sample comparison	99
3.5.3	Out-of-sample comparison	100
3.6	Conclusion	103
	References	124

List of Figures

1.1	Kernel estimates at different hours of the day.	32
1.2	15 minutes returns	33
1.3	Location and scale shifts in the pdf through the quantile function	34
1.4	Estimated parameters	35
1.5	Seasonal component	36
1.6	Seasonality in the quantiles	37
1.7	Seasonality and the tails	38
1.8	VaR	39
2.1	Yield data	74
2.2	Macro-yields model in sample fit: yields	75
2.3	Macro-yields model in sample fit: key macro variables/1	76
2.4	Macro-yields model in sample fit: key macro variables/2	77
2.5	Estimated macro-yields factors	78
2.6	Smoothed square forecast errors	79
2.7	Smoothed forecast errors	80
3.1	Nelson-Siegel factor loadings	116
3.2	No-Arbitrage Latent factors and Nelson and Siegel factors	117
3.3	Zero-coupon yields data	118
3.4	No-Arbitrage loadings of the Nelson and Siegel factors	119
3.5	Distribution of the estimated loadings for a^{NA}	120
3.6	Distribution of the estimated loadings for $b^{NA}(1)$	121

3.7 Distribution of the estimated loadings for $b^{NA}(2)$ 122
3.8 Distribution of the estimated loadings for $b^{NA}(3)$ 123

List of Tables

1.1	Descriptive statistics at different hours of the day	28
1.2	GARCH(1,1) estimates with Student-t distribution	29
1.3	Kupiec test on the VaR forecasts	29
1.4	Christoffersen's likelihood ratio test on the VaR forecasts	30
1.5	GARCH(1,1) estimates of the standardized returns	31
1.6	Kupiec test on the VaR forecasts	31
2.1	Summary statistics of the US zero-coupon data	68
2.2	Macroeconomic series	69
2.3	Model selection	71
2.4	Summary statistics of the estimated factors	71
2.5	Goodness of fit	72
2.6	Out-of-sample performance	73
3.1	Summary statistics of the US zero-coupon data	104
3.2	Autocorrelations	105
3.3	Parameter estimates	106
3.4	Estimation results for a^{NA}	108
3.5	Estimation results for $b^{NA}(1)$	109
3.6	Estimation results for $b^{NA}(2)$	110
3.7	Estimation results for $b^{NA}(3)$	111
3.8	Summary statistics for the resampled parameters	112
3.9	Measures of fit	113

3.10 Out-of-sample performance 114
3.11 Diebold-Mariano test statistics 115

Introduction

In past years, improvements in the computational speed of computers and the availability of large datasets have further fostered the research in forecasting analysis for financial time series. This thesis contributes to the empirical literature on financial forecasting by addressing issues related to the problem of improving models performance by exploiting larger information sets.

The first paper investigates the distribution of high frequency financial returns, with special emphasis on the seasonality. With the availability of detailed information on trades and quotes, due to the implementation of electronic trading systems, intraday data has become a major pole of interest for researchers and financial agents that practice intraday trading. Within the day there are significant variations in asset prices, which imply different evaluations of the return's distribution through the day, these variations are partly deterministic and due to the intraday seasonality. Intraday value at risk evaluations therefore depend on the time of the day. If an intraday trader does not take this seasonality into account in her risk estimations, she will underestimate the expected loss at the opening and closing and overestimate it at noon. I propose a quantile regression approach (Koenker and Basset, 1978) to model the distribution of high frequency financial returns and to forecast intraday value at risk. This choice is motivated by several reasons. First of all, not only the volatility of high frequency financial returns presents seasonal movements, but also the skewness and kurtosis. Moreover, quantile regression does not assume the existence of any moment, is distribution free and robust to the presence of outliers or jumps. Using 15 minutes quote midpoints of three stocks traded

at the Spanish stock exchange from January 2001 to December 2003, I show that indeed the conditional probability distribution depends on the time of the day. Results, in terms of quantiles, permit straightforward intraday risk evaluations, such as value at risk. I show how the intraday value at risk at 2.5%, 1% and 0.5% confidence levels depend on the time of the day and I perform out-of-sample value at risk forecasts. The tests performed on the out-of-sample value at risk forecasts confirm that the model is able to provide good risk assessments and to outperform standard approaches (Gaussian and Student-t GARCH).

In the second paper, I focus on the problem of forecasting yields including large datasets of macroeconomic information. The interaction between financial markets and macroeconomic conditions has raised the necessity of developing new financial models which are able to efficiently summarize the macroeconomic information. I propose an innovative way to exploit the linkages between macro variables and yields. Rather than including in the yield curve model macroeconomic variables as factors, I directly extract the latent factors from a data set composed of both yields (seventeen series) and macro variables (one hundred-eighteen) which includes real variables (sectoral industrial production, employment and hours worked), nominal variables (consumer and producer price indices, wages, and money aggregates) and asset prices (stock prices and exchange rates). To identify the yield curve factors, I follow the approach based on the Nelson and Siegel (1987) curve imposing restrictions only on the loadings relative to the yields, leaving the loadings relative to macro variables free. This allows to use latent yield curve factors which are enriched with information from macro variables, thereby keeping parsimony. I estimate the model by maximum likelihood, combining EM algorithm and Kalman filter, using monthly observations from January 1970 to December 2000. Results show that out-of-sample forecast performances improves at mid and long horizons (i.e. 6 and 12-months ahead) compared with the forecasts generated by a model estimated using only the yields, a model augmented with three key macro variables, a model augmented with the first three principal components extracted from the

same dataset of macroeconomic variables and the random walk (which is a standard benchmark for yield curve forecasting).

In the third paper, I test whether the Nelson and Siegel (1987) yield curve model is arbitrage-free in a statistical sense. Fixed-income wealth managers in public organizations, investment banks and central banks rely heavily on Nelson and Siegel (1987) type of models to fit and forecast yield curves. Despite its empirical merits and wide-spread use in the finance community, two theoretical concerns can be raised against the Nelson-Siegel model. It is not theoretically arbitrage-free and it falls outside the class of affine yield curve models. I estimate the Nelson and Siegel factors and use them as exogenous factors in an essentially-affine term structure model to estimate the implied arbitrage-free factor loadings. For the no-arbitrage model with time-varying term premia, I use the two-step approach of Ang, Piazzesi and Wei (2006). Using a non-parametric resampling technique and zero-coupon yield curve data from the US market covering the period from January 1970 to December 2000 and spanning 18 maturities from 1 month to 10 years, I find that estimated parameters from the no-arbitrage model are not statistically different from those obtained from the Nelson-Siegel model, at a 95 percent confidence level. To corroborate this result, I show that the Nelson-Siegel model performs as well as its no-arbitrage counterpart in an out-of-sample forecasting experiment. I therefore conclude that the Nelson and Siegel yield curve model is compatible with arbitrage-freeness.

Chapter 1

A Quantile Regression Approach to Intraday Seasonality

ABSTRACT: This paper investigates the distribution of high frequency financial returns, with special emphasis on the seasonality. Using quantile regression, we show the expansions and shrinks of the probability law through the day for three years of 15 minutes sampled stock returns. Returns are more dispersed and less concentrated around the median at the hours near the opening and closing. We provide intraday value at risk assessments and we show how it adapts to changes of dispersion over the day. The tests performed on the out-of-sample forecasts of the VaR show that the model is able to give good risk assessments and it outperforms Gaussian and Student's t GARCH models.

Keywords: High frequency returns, Quantile Regression, Seasonality, Intraday VaR.

JEL classification: C14, C22, C53, G10.

This chapter is adapted from the paper "Intraday seasonality of returns distribution. A quantile regression approach and intraday VaR estimation" written with David Veredas (Universite Libre de Bruxelles), CORE DP 2006/77.

1.1 Introduction and Motivation

The interest of intraday seasonal patterns of the probability law of high frequency financial returns rests on two facts. First, intraday data has become a major pole of interest for researchers and financial agents that practice and analyze high frequency trading. This practice and this analysis is used in an array of instruments such as derivative pricing, efficient estimation of security's beta, liquidity analysis, responses to news arrivals, and any operation that involves risk measures. For instance, high frequency hedge fund managers often open and close positions within the day. For these managers intraday risk evaluation is an important tool to follow the market and to build optimal intraday trading strategies.

Second, the analysis of risk is intimately related with the analysis of probabilities and, therefore, the analysis of the conditional probability distribution. Asset returns are realizations of a random variable and their behavior is fully described by their conditional probability law. Any function, such as the density function, describing this law conveys information about the likelihood that the next realization will take a certain value. But within the day these odds depend partly on a deterministic seasonal component that makes the probability density function to expand or shrink as a function of the time of the day. This effect is illustrated in the kernel densities for returns at different hours of the day shown in Figure 1.1. Data are 15 minutes sampled returns for three stocks (large, medium and small caps) traded at the Spanish stock exchange.¹ The kernel density estimates for returns at different times of the day vary significantly. Around lunch the density is more peaked and the tails are thinner while it is more dispersed at the hours near the opening and closing.

One of the most common risk-related intraday measures, that make use of the probability law of returns, is value at risk.² Value at risk evaluations depend very

¹All over the analysis we use standardized returns for comparison purposes. Otherwise the scale of the plots depends on the price, perturbing the interpretation.

²There are many other risk-measures that may be constructed from the intraday return's distribution, such as volatility or left extreme tail analysis. Alternatively we may, exploit the

much on the time of the day. If an intraday trader does not take this seasonality into account in her risk estimations, she will underestimate the expected loss at the opening and closing and overestimate it at noon.

[FIGURE 1.1 AROUND HERE]

Moreover, not only volatility presents seasonal movements, but also skewness and kurtosis. Table 1.1 shows descriptive statistics for the data grouped according to the hour of the day. We report the sample mean, the sample standard deviation, the skewness and the kurtosis indices for four different hours.³ These estimates are proxies of the intraday behavior of the probability law. There is no evidence of an intraday seasonal pattern in the sample mean of returns. However, there is a very clear U-shaped pattern in the sample standard deviation, as found in many former studies. In addition to this, the last two columns suggest the presence of an intraday seasonality in the skewness and kurtosis indices. While the movements in the skewness are small in magnitude, for the kurtosis index there are large variations during the day. For all the stocks, there is a significant increase in the thickness of the tails just after 15:00, right before the opening of the NYSE.

[TABLE 1.1 AROUND HERE]

The standard approach to analyze the conditional distribution function of intraday asset returns is to fit a model for the second moment as a function of two components. One for the dynamics and another for the seasonality. If returns are Gaussian, the second moment provides information enough to describe the conditional probability law, as all the odd moments are zero and the even moments

intraday distribution to construct daily measures that can be used to compute daily volatilities and daily value at risk. Yet, the way dynamic models for the density aggregate (e.g. aggregation of quantile regression models as the one we present here) is still an open research question.

³The table displays the bias adjusted skewness index computed as $\frac{\sqrt{n(n-1)}}{n-2}m_3/s^3$ where n is the number of days (there is one observation for each day), m_3 is the third sample central moment and s_2 is the sample standard deviation. And the bias adjusted kurtosis computed as $\frac{n-1}{(n-1)(n-3)}((n+1)\frac{m_4}{s^4} - 3(n-1))$ where m_4 is the fourth sample central moment.

are functions of the second moment. This property of the Gaussian distribution is very appealing but, at the same time, this distribution is not able to reproduce the tail behavior present in the data. This is one of the reasons for which it is now commonly accepted that asset returns are not normally distributed. More flexible distributions, such as the Student's t distribution, are needed. However, the drawback of these laws is that moments beyond the second are either zero -e.g. the third moment- or functions of an invariant tail index -e.g. the fourth moment. For instance, a GARCH model with a Student's t distribution has constant kurtosis given by a function of the estimated degrees of freedom, which is not consistent with the data features. A possible solution to overcome this problem would be to fit models for different moments, similarly to Hansen (1994) or Harvey and Siddique (1999) among others, but it is not clear the functional forms that these models should take and/or which regressors to use.

Since our interest is the analysis of the seasonality of the conditional distribution, a natural alternative is to model directly the conditional probability. Among all the functions that characterize the conditional probability (density, cumulative, characteristic, Laplace, hazard, etc), the conditional quantiles are the better suited due to the existence of quantile regression, introduced by the seminal work of Koenker and Basset (1978). Indeed quantile regression has a number of useful features. First, quantile regression is one of the possible ways to characterize the conditional probability law and, since there is a one to one relation with all the other possible characterizations, it allows, indirectly, to analyze the effect of the time of the day on the density function of asset returns. Second, quantile regression does not assume the existence of any moment. In fact, it does not assume anything about the moments. Often it happens that the tails of returns are so thick that some important moments do not exist. For instance, Table 1.2 shows the estimated parameters of a GARCH(1,1) with Student's t distribution. The estimated degrees of freedom for the three stocks are very low. So low that, according to the model, kurtosis does not exist for any of them and variance does not even exist for one of

them.⁴

[TABLE 1.2 AROUND HERE]

Third, quantile regression is robust in the sense that the estimated coefficients are not sensitive to outliers on the dependent variable. This is particularly useful in the analysis of high frequency financial returns since often we do find outliers or, at least, observations that are remarkably different to the rest of the process. For instance, Figure 3.3 shows the actual returns for the three stocks we analyze. For all there is at least one observation that is unusually high. Fourth, quantile regression is a distribution free model. This is a very compelling feature. It does not rely on any distribution specification but, ironically, it is an estimate of the conditional probability distribution. As noted earlier, and shown in Table 1.2, assuming a parametric distribution for intraday asset returns entails a series of problems that sometimes, e.g. infinite variance, are difficult to overcome.

[FIGURE 3.3 AROUND HERE]

The use of quantile regression in asset returns is not new. One of the first to use it are Engle and Manganelli (2004) who introduce the CAViaR (Conditional Autoregressive Value at Risk). CAViaR extends the traditional linear quantile regression to a nonlinear framework and develop a new test of model adequacy, the Dynamic Quantile (DQ) test, using the criterion that each period the probability of exceeding the VaR must be independent of all the past information. Gouriéroux and Jasiak (2005) introduce a new dynamic quantile model univariate series and panel data as well as the Quantile Factor Model. Less related, Bouye and Salmon (2003) use quantile regression in a copula context, that is they deduce the form

⁴We are here abusing a bit of the estimation results. According to the table, the variance does exist for the stock ANA, as the estimated degrees of freedom is higher than two, though very close to it. However, in the software we use -the GARCH toolbox of Matlab- the parameter is constrained to be greater than two (as it often happens due to the way in which the Student's t distribution is written). Therefore, in some sense, it can be expected that the unbounded estimator for the degrees of freedom to be below two.

of the non linear conditional quantile regression implied by the copula. As for intraday VaR, Giot (2005) quantify intraday VaR (15 and 30 minute returns) using normal GARCH, Student GARCH, RiskMetrics and Log-ACD models. He shows that Student GARCH model performs best. Last, Dionne et al. (2005) investigate the use of tick-by-tick data for market risk measurement and propose an intraday Value at Risk at different horizons based on irregularly time-spaced high-frequency data by using an intraday Monte Carlo simulation.

Using quote midpoints of three stocks traded at the Spanish stock exchange from January 2001 to December 2003, we show that indeed the conditional probability distribution depends on the time of the day. At the opening and closing the density flattens and the tails become thicker, while in the middle of the day returns concentrate around the median and the tails are thinner. Results are intuitive, in the sense that they confirm the general perception that in the opening and closing the probabilities of finding large price fluctuations are higher than at lunch. Results, in terms of quantiles, permit straightforward intraday risk evaluations, such as value at risk. We show the intraday variation of the maximum expected loss at 2.5%, 1% and 0.5% confidence levels. The maximum expected loss is maximal at the opening and closing and minimal at lunch time. Failure rates tests, based on Kupiec (1995) and Christoffersen (1998) confirm that the model is able to provide good forecasts of the maximum expected loss. Comparison with standard approaches (Gaussian and Student-t GARCH) show that the latter miss the correct probabilities and that quantile regression outperforms them.

The structure of the paper is as follows. Section 1.2 introduces the data and the market structure. Section 1.3 briefs quantile regression, the model that is used for estimation and how to interpret results in term of density functions. Section 1.4 shows the estimation results, while Section 1.5 contains intraday value at risk forecast and evaluation with the quantile regression model as well as its GARCH competitors. Section 1.6 concludes.

1.2 Market and Data

Data come from the Spanish Stock Exchange (SSE), the 9th world largest stock exchange in terms of capitalization (the 3th among continental European markets), and the 7th in terms of total value of share trading (the 3th in continental Europe) according to the World Federation of Exchanges. The Spanish stock exchange interconnection system is the electronic platform that connects, since 1995, the four exchanges that compose the SSE (Barcelona, Bilbao, Madrid, and Valencia). This system holds all the Spanish stocks that achieve pre-determined minimum levels of trading frequency and liquidity. Every order submitted to the system is electronically routed to a centralized limit order book (LOB) to proceed with its immediate execution or storage. The matching of orders is, therefore, computerized. The LOB on the brokers' screens is updated each time there is a cancelation, execution, modification or new submission. The SSE is organized as an order-driven market with a daily continuous trading session from 9:00 a.m. to 5:30 p.m. and two call auctions that determine the opening and closing prices.

During the continuous trading session, a trade takes place if and only if an order hits the quotes. Pre-arranged trades are not allowed during the continuous session, and price-improvements are impossible. There are no market makers and there is no floor trading. The market is governed by a strict price-time priority rule, but an order may lose priority if modified. Stocks are quoted in euros. The minimum price variation (tick) equals 0.01 for prices below 50 and 0.05 for prices above 50. The minimum trade size is one share. There are three basic types of orders: market, limit, and market-to-limit. Market orders are executed against the best prices on the opposite side of the book. Any excess that cannot be executed at the best bid or ask quote is executed at less favorable prices by walking down (up) the book until the order is fulfilled. Market-to-limit orders do not specify a limit price but are limited to the best opposite-side price on the book at the time of entry. Any excess that cannot be executed is converted into a limit order at that price. Finally, limit orders are to be executed at the limit price or better. Any unexecuted part of

the order is stored in front of the book at the limit price. By default, orders expire at the end of the session.

The official market index of the SSE is the IBEX-35, which includes the 35 most liquid and active stocks of the exchange, weighted by market capitalization. Its composition is regularly revised every semester. Our initial sample is formed by the 35 index constituents from January 2001 to December 2003. The data used in this study consists of 15 minutes sampled quote midpoints during 3 years, from January 2001 to December 2003, of the 35 companies listed in the IBEX-35. For each stock there are 34 intraday observations for a total of 25.430 observations. For simplicity, we will report the analysis only on 3 of the 35 companies but the results are valid for all of them and they are available upon request. Among the 35 companies of the IBEX-35, we report the results for Telefonica (TEF), Endesa (ELE) and Aciona (ANA) that are, respectively, a big, medium and small company, weighting, approximately, 20%, 6% and 0.8% in the index.

1.3 Quantile Regression as Density Regression

The probability law of a random variable r_t can be characterized by means of different functions. Some, like the density or the cumulative functions, are common. Others, like the quantile function, the hazard function or the characteristic function are less used. Yet, any can be written as a function of the others and hence the knowledge of one implies the knowledge of the others. The quantile function is particularly compelling in the context of conditional distributions. This is due to the existence of a solid theory on quantile regression (see Koenker, 2005, for a survey). Let $Q_{r_t}(\tau)$ be the τ -th quantile of r_t . It is well known that

$$\begin{aligned} f(r_t) &= \frac{\partial}{\partial r_t} F(r_t) \quad \text{and} \\ Q_{r_t}(\tau) &= F^{-1}(\tau) = \inf\{r_t : F(r_t) \geq \tau\}, \quad \tau \in (0, 1), \end{aligned}$$

where $f(r_t)$ is the probability density function, pdf hereafter, and $F(r_t)$ is the cumulative distribution function, cdf hereafter. Top row of Figure 1.3 shows this idea. The density is symmetric around the mean, which implies that the quantiles are also symmetric around the median (that equals the mean). The density is centered at zero, and hence the quantile function at the median, $Q_{r_t}(0.5)$, is zero.

One may question what happens with the pdf and the quantile function if there is a location-scale shift. Second to fourth rows of Figure 1.3 illustrate these cases. The second row shows a positive location shift in the density, which produces a parallel upward shift of the quantile function. Or, inversely, if the quantile function shifts, the density shifts the location. It is worth noticing that after the shift the quantile function at the median, $Q_{r_t}(0.5)$, is not zero anymore as the mean in the pdf is not zero anymore either.

[FIGURE 1.3 AROUND HERE]

The third row shows the effect of a positive scale shift in the density. This shift produces an expansion of the quantile function, or, inversely, an expansion of the quantile function implies a positive scale shift in the density.⁵ The expansion in the quantile function implies an increase in the dispersion of the quantiles. This happens when we compare the probability law at, for instance, lunch and the closing, as already noted in Figure 1.1. By contrast, the dispersion of the observations decreases if we compare the probability law the opening and at lunch which means a contraction of the quantiles. Finally, last row illustrates a positive location shift and a scale shift in the density, implying an asymmetric shift -a mix of shift and expansion effects- in the quantile function. More complex shifts are possible. For instance a one-sided expansion in the quantile implies an increase of the dispersion in only one side of the density, creating skewness. Fat tails can be also created in the density if the quantile are stretched only at the highest and lowest values, say 1% and/or 99% quantiles. In sum, either a location shift or a scale shift, or

⁵These type of shifts are of particular relevance for this article.

both, in the pdf has a clear representation in terms of quantiles, as both functions contain the same information about the random variable of interest.

The understanding of the effect of these shifts and how the quantile and the density function are affected by them is important in a conditional context. In fact, the movements in the densities of Figure 1.1 are produced by the intraday seasonality. It is therefore meaningful to model how the probability distribution evolves conditional to the time of the day. Quantile regression (QR henceforth), introduced by Koenker and Bassett (1978), is the appropriate tool. The problem of finding the τ -th unconditional quantile can be expressed as the solution of a simple linear optimization problem. Generalizing these results to the case in which the quantiles are linear functions of some explanatory variables leads to the QR method. The fundamental difference of QR with respect to mean regression is that the latter considers the effect of the regressor on the mean of the regressand while QR considers the effect of the regressor on the specific τ -th quantile of the regressand. Hence, for a sufficiently narrow grid of τ , the QR method can fully describe the quantile function. The basic QR model is

$$Q_{r_t}(\tau|x_t) = \omega(\tau) + \sum_{j=1}^J \beta_j(\tau)x_{jt}, \quad \tau \in (0, 1), \quad (1.3.1)$$

where the intercept $\omega(\tau)$ and the slope parameters $\beta_j(\tau)$ are functions of τ . While in the mean regression model there is a unique parameter β_j that describes the effect that x_{jt} has on the conditional mean of r_t , in QR for each $\tau \in (0, 1)$ there is a parameter $\beta_j(\tau)$ that describes the effect of x_{jt} on the τ -th conditional quantile of r_t . In other words, QR measures the effect of the regressors on each quantile of the conditional distribution of the dependent variable. In this way it allows to analyze how a shock in the regressors affects the different quantiles and hence the pdf of returns.

The set of regressors x_{jt} is divided in two parts. One accounts for the intraday seasonality, the main object of interest, and the second controls for the dynamics.

As for the seasonality, we model it using a Fourier series of order 3:

$$seas_d(\tau) = \sum_{j=1}^3 \alpha_j(\tau) \cos\left(2\pi j \frac{d}{34}\right) + \gamma_j(\tau) \sin\left(2\pi j \frac{d}{34}\right), \quad (1.3.2)$$

where 34 is the number of intraday time intervals (for the 15 minutes sampled returns) and d denotes the time of the day in ordinal sense (i.e. the sequence 1, 2, ..., 34).⁶ Fourier series are convenient expressions for seasonality as the combination of cosines and sines is flexible enough to capture virtually any seasonal pattern. The cosine component of the first Fourier series reaches the maximum at the opening and at the closing, the hours of the day in which the dispersion is higher, and has the minimum at lunch time, the time of the day in which the dispersion is minimal. We therefore expect this cosine term to capture most of the seasonal pattern.

To control for the dynamics, we follow Koenker (2005) choosing one lag of the absolute value of returns: $\beta(\tau)|r_{t-1}|$. More lags or other functions of r_t to capture the dynamics as, for instance, square returns are possible. However, in a robust setting, the choice of absolute values is more sensible.⁷ Putting all the elements together the model we estimate is

$$Q_{r_t}(\tau|d, |r_{t-1}|) = \omega(\tau) + \beta(\tau)|r_{t-1}| + \sum_{j=1}^3 \alpha_j(\tau) \cos\left(2\pi j \frac{d}{D}\right) + \gamma_j(\tau) \sin\left(2\pi j \frac{d}{D}\right). \quad (1.3.3)$$

Estimation has been implemented in GAUSS using a modified version of the library Qreg.⁸ Parameters are estimated using the interior point method, as described by Portnoy and Koenker (1997). The chosen grid of quantiles is (0.05,

⁶We tried higher orders of the Fourier series but results do not change substantially.

⁷We have also tried with more lags of absolute returns and results, available upon request, don't change qualitatively.

⁸Qreg, a GAUSS library for computing quantile regression, D. Jacomy, J. Messines and T. Roncalli (2000), Groupe de Recherche Operationelle, Credit Lyonnais, France: http://gro.creditlyonnais.fr/content/rd/home_gauss.htm. All the codes have also been translated into Matlab, using the function *lp_fnm* of Daniel Morillo and Roger Koenker, translated from Ox to Matlab by Paul Eilers 1999, modified by Roger Koenker 2000 and by Paul Eilers 2004.

0.10,..., 0.95) and the limiting covariance matrix has been computed in GAUSS using the procedure described in Appendix.

1.4 Estimation Results

Parameters in equation (1.3.3) depend on the quantile considered, τ . There are as many parameters as quantiles times the number of explanatory variables plus those in the intercept $\omega(\tau)$. Because this number may become large, in our case is 168 per stock, we follow the literature, see for instance Koenker (2005), and we present all the results graphically. This presentation nicely dovetails with Figure 1.3 as the interpretation of density movements in terms of quantiles applies. Figure 1.4 shows the estimated parameters of model (1.3.3) for TEF, ELE and ANA respectively. Every point is an estimated parameter for a different quantile. We also plot the 5% point-wise confidence intervals.

Top left plots of each panel show the intercept parameters, $\hat{\omega}(\tau)$ while the coefficients for past absolute returns, $\hat{\beta}(\tau)$, are in the top right plots. For all the stocks, the magnitude of the lagged value of the return is an important source of variation. But it affects differently the different quantiles of the distribution. The median is unaffected by a shock in $|r_{t-1}|$. Following the logic of Figure 1.3, there is no location shift and hence the median remains unchanged for any value of $|r_{t-1}|$. It changes however any quantile beyond and below 50%. For a given past absolute return, the effect on the extreme quantiles is larger than for the quantiles near the median. Exemplifying, if return at $t - 1$ was zero, the density, conditional to the time of the day, remains unchanged. If, by contrast, at $t - 1$ there is a large movement in returns, the density becomes more sparse around the median, that remains unchanged, increasing the probabilities of finding a large price variation the next period. If return at $t - 1$ is small, the density shrinks, decreasing the probabilities of finding large price variations.

The remaining six plots show the estimated values of the parameters for the Fourier series. The second line refers to the estimated parameters of the first

Fourier series, the third one to the estimated parameters of the second Fourier series and the last one to the third Fourier series. The coefficients for the cosine terms, the alphas, are larger, for all stocks, than the sinus ones, the gammas. This is due to the fact that the cosine series peak at the opening and the closing, the times of the day at which trading activity is more intense and return dispersion is bigger. None of the coefficients is different from zero for $\tau = 0.5$, meaning that the median is not affected by past observation nor the time of the day. In other words, no profit strategies based on the time of the day are found. Consequently, since also the estimated coefficient of $|r_{t-1}|$ for $\tau = 0.5$ is zero, the conditional median is equal to the unconditional one, that is zero. Figure 1.5 shows the estimated seasonal component, $se\hat{a}_d(\tau)$, computed as in (1.3.2). The plots read as follows. Each line is the seasonal component for a specific hour of the day for different quantiles. The estimated seasonal components displays different shapes within the day and some conclusions can be drawn. First, the shape and the magnitude of the seasonal component is fairly similar for all the stocks. In particular, there is no seasonal behavior at the median. But there is beyond it and becomes more remarkable as we approach the extreme quantiles. Second, the seasonal component is clearly different at the opening and the closing, with values that are negative for taus smaller than 0.5 and positive for taus bigger than 0.5. Third, the seasonal component at 13:00 and 14:00 displays exactly the opposite behavior with respect to the one at the opening and closing, but with a smaller magnitude.

To better see how the conditional distribution of returns moves though the day, Figure 1.6 plots the conditional quantiles of the 15 minutes returns for different hours of the day. Rewriting equation (1.3.3) conditional to a particular value of past absolute return and on different hours of the day, we have

$$Q_{r_t}(\tau|d, \overline{|r_{t-1}|}) = \omega(\tau) + \beta(\tau)\overline{|r_{t-1}|} + seas_d(\tau).$$

The choice of the conditioning value of $|r_{t-1}|$ has a quantitative but not qualitative effect. For a given τ , $\beta(\tau)\overline{|r_{t-1}|}$ is constant, while the term $seas_d(\tau)$ changes

according to the hour of the day (as shown in Figure 1.5). The only effect that the chosen value of $|r_{t-1}|$ has is to shift all the conditional quantiles at the same τ by the same amount. The figure reads as follows: the closer the line is to the horizontal zero line, the more concentrated is the density around the median. And the further it is, the more dispersed it is. The time of the day at which there is the largest seasonal effect is at 17:00, the closure of the market, followed by the effect at 9:30, the opening. At these hours the conditional density becomes more dispersed. In the opposite direction, for all the companies, are the seasonal effects at 13:00 and 14:00. They decrease (in absolute value) the conditional quantiles, decreasing the dispersion. This effect can be associated to a reduced trading activity during the lunch break.

1.5 Intraday Value at Risk

As shown in Section 1.3, there is a one to one relation among the quantile and density functions. This is particularly appealing in the construction of risk measures, which are intimately related with the analysis of the tails of the density function. Using the results of Figure 1.6 and equation (1.7.2) in the Appendix, we can compute the conditional density at different quantiles. Figure 1.7 shows the tails of these densities.⁹ Each point of the conditional density is derived from its relative conditional quantile. As expected, the density mass at the extremes is way larger around the opening and closing than around lunch. This seasonal tail behavior has to be taken into account in the computation of intraday risk measures, such as VaR.

Value at Risk was developed to provide a single number that could summarize the information about the risk in a portfolio. Over the last ten years, this technique has been increasingly used by banks and regulators all over the world as a way to

⁹A full picture of the density is possible but not relevant as the financial interest lies on the tails and not around the median. And, moreover, it has been shown earlier that nothing interesting happens around the median.

estimate possible losses related to the trading of financial assets, i.e. as a tool designed to quantify and forecast market risk. In particular, the goal of VaR is to assess the possible loss that can be incurred by a trader or bank, for a given portfolio of assets, over a given time period and for a certain confidence level. The time period and the confidence level are the two major parameters that should be chosen in a way appropriate to the overall goal of risk measurement. When the primary goal is to satisfy external regulatory requirements, such as bank capital requirements of the Basel II Capital Accord, the confidence level is typically small, 1%, and the time horizon is long (usually a 10 day period). However for an internal risk management model, used by a company to control the risk exposure, the typical confidence level is even smaller and the time horizon shorter. In particular, for active market participants such as high frequency traders, floor traders or market makers, the time horizon of their returns is shorter and the corresponding trading risk must be assessed on such short time intervals. Therefore a VaR model that characterizes the market risk on an intraday basis is useful for market participants (such as intraday traders and market makers) involved in frequent intraday trades.

The VaR at a confidence level of τ for a given portfolio is the loss at the τ percent probability level, which can simply be defined as the τ empirical quantile of the conditional distribution of returns:

$$Pr[r_t < VaR_t(\tau|\mathfrak{S}_{t-1})] = \tau \Leftrightarrow VaR_t(\tau|\mathfrak{S}_{t-1}) = Q_{r_t}(\tau|\mathfrak{S}_{t-1}).$$

From an empirical point of view, the computation of the $VaR_t(\tau|\mathfrak{S}_{t-1})$ of a portfolio of assets requires the computation of the empirical quantile at level τ of the distribution of the future returns of the portfolio given the information set available at time $t - 1$, \mathfrak{S}_{t-1} . Engle and Manganelli (1999) introduced nonlinear QR as a method for computing VaR. The originality of our model relies on two points: the use of high frequency data to forecast VaR at intraday time horizon and the use of the Fourier series to model the intraday seasonality of returns in a quantile regression framework. Our model defines the information set available up to time

$t - 1$, \mathfrak{S}_{t-1} , as including the lagged absolute value of returns, $|r_{t-1}|$, and the three deterministic Fourier series, that are indexed by the time of the day d

$$VaR_t(\tau|d, |r_{t-1}|) = Q_{r_t}(\tau|d, |r_{t-1}|).$$

The one step ahead out-of-sample VaR forecast is conducted using a rolling window scheme, a method popular among practitioners since Fama and MacBeth (1973) and Gonedes (1973). The use of rolling windows is justified by parameter instability, which can distort the out-of-sample forecast. The window size is adapted to the liquidity of the stock. For the most liquid stock, TEF, we use a rolling window of 2000 observations, for ELE a window of 2500 observations and for ANA the less liquid stock a bigger window of 3000 observations.¹⁰ This choice is motivated by the fact that in the same time span, there is a different number of transactions. While for the most liquid stocks (like TEF) in 2000 observations of 15 minutes returns there is enough information due to the high number of transactions, for the less liquid stocks (like ANA) this time interval is too short because it includes a fewer number of transactions.¹¹ This lead to 23.430, 22.930 and 22.430 one-step ahead forecasts for TEF, ELE and ANA respectively.

Figure 1.8 displays the last 500 observations of the 15 minutes sampled returns for TEF, ELE and ANA with the relative VaR forecasts at the confidence levels of 2.5%, 1% and 0.5%.¹² The estimated VaRs show clearly the effects of the two components that we used to model the conditional quantiles. The seasonal component is responsible of the deterministic daily oscillations, while the dynamic one is amplifying or reducing the oscillations to take into account the dispersion clustering. Moreover, as the confidence level of the VaR decreases, the dynamic component

¹⁰2500 observation of 15 minutes return correspond to 58 days (three months), while 2500 observation of 15 minutes observation cover a time span of 73 days (four months) and, finally, 3000 observations are equivalent to 88 days (five months).

¹¹The choice of the optimal window, although relevant in this literature, is out of the scope of this paper.

¹²These are reasonable confidence levels for intraday market risk evaluations as Basel threshold is 1%.

becomes more relevant.

At first sight, it looks that the estimated VaR for the three confidence levels and for all the stocks are close enough to the data, i.e. we are not overestimating the risk, and that the number of times that the realized returns are below the estimated VaR is not too big. As a check, we computed the failure rates. That is the percentage of times that the observations are below the VaR. If the VaR is well specified, then the empirical failure rates, denoted by \hat{f} , should be close enough to the confidence level. Table 1.3 reports the empirical failure rates for the three stocks and for the confidence levels of 2.5%, 1% and 0.5%. The values that are in parenthesis refer to the confidence intervals computed according to the Kupiec (1995) test. The null hypothesis of the test is that the empirical failure rate, \hat{f} , is equal to the confidence level of the VaR, τ . The 5% confidence interval for τ is given by $\hat{f} \pm 1.96\sqrt{\hat{f}(1 - \hat{f})/N}$, where N is the total number of observations that we are evaluating, that is 23.430, 22.930 and 22.430 for TEF, ELE and ANA respectively. For all the stocks and all the confidence levels, the confidence interval contains the theoretical confidence levels of 2.5%, 1% and 0.5% respectively, therefore we do not reject the null hypothesis that the empirical failure rates are equal to the theoretical ones for all the confidence levels of the VaR and for all the stocks.

A test that is equivalent to the Kupiec's is the likelihood ratio test of unconditional coverage developed by Christoffersen (1998). This test is based on a hit variable, that takes value 1 if there is a success, that is if the realized return is bigger than the expected VaR, and 0 otherwise, and therefore distributed according to a binomial distribution. The test is

$$LR_{uc} = -2 \log \left(\frac{(1 - \tau)^{n_0} \tau^{n_1}}{(1 - \hat{f})^{n_0} \hat{f}^{n_1}} \right) \sim \chi_1^2,$$

where n_0 is the number of failures and n_1 the number of successes. The first panel of Table 1.4 reports the values of the test with the relative p-values. The conclusions are similar to Kupiec's test. The model is able to predict well the VaR for all the stocks and for all the confidence levels considered.

However, a drawback of the Kupiec and the likelihood ratio test of Christoffersen is that they just count the number of successes and of failures, testing only the equality between the VaR violations and the confidence level. In a risk management framework, it is also important that the VaR violations are not correlated in time. The likelihood ratio test of independence, Christoffersen (1998), examines serial independence of VaR estimates. As the previous likelihood ratio test, this test is built starting from a hit variable that takes values according to

$$I_t = \begin{cases} 1, & \text{if } r_t > VaR_t(\tau|d, |r_{t-1}|); \\ 0, & \text{otherwise.} \end{cases}$$

The likelihood ratio test of independence tests the null of independence against the alternative of a first order Markov process of the violations. Denoting with n_{ij} the number of observation of I with value i followed by j , the likelihood ratio test of independence can be expressed as

$$LR_{ind} = -2 \log \left(\frac{(1 - \hat{f})^{n_{00}+n_{10}} \hat{f}^{n_{01}+n_{11}}}{(1 - \hat{f}_{01})^{n_{00}} \hat{f}_{01}^{n_{01}} (1 - \hat{f}_{11})^{n_{10}} \hat{f}_{11}^{n_{11}}} \right) \sim \chi_1^2,$$

where \hat{f}_{01} is the percentage of successes after a failure and \hat{f}_{11} is the percentage of successes after a success. The null of the test is that $\hat{f}_{01} = \hat{f}_{11} = \hat{f}$. Third panel of Table 1.4 reports the value of the test with the relative p-values. The null of independence of the violations is accepted for all the stocks and all the confidence levels of the VaR. Finally, as a more powerful tool, we performed the joint likelihood ratio test of independence and coverage. The Christoffersen's likelihood ratio test of conditional coverage

$$LR_{cc} = LR_{uc} + LR_{ind} \sim \chi_2^2,$$

in which the null of the unconditional coverage is tested against the alternative of the independence test. Bottom panel of Table 1.4 reports the results. For all the

confidence levels, we do not reject the null of conditional coverage confirming that the model is well specified. The tests results show the ability of the model to provide good out-of-sample forecasts of the intraday VaR confirming the importance of well specifying the intraday seasonality. This component, as shown in Figure 1.8, seems to have a crucial role in the determination of the intraday VaR.

We compare the performance of the VaR using quantile regression with the benchmark in risk modelling: GARCH type of models. To account for the intraday seasonality in the variance, we follow a common used approach, see for example Giot (2005), which is to seasonally adjust the return series:

$$\tilde{r}_t = \frac{r_t}{\sqrt{\phi_d}},$$

where ϕ_d is the deterministic intraday seasonal component. The latter is defined as the expected volatility conditioned on the time of the day, where the expectation is computed by averaging the squared raw returns for each time of the day. If \tilde{r}_t has no mean effects, a GARCH(1,1) can be written as

$$\begin{aligned}\tilde{r}_t &= \varepsilon_t h_t \\ h_t^2 &= \omega + \alpha \tilde{r}_{t-1}^2 + \beta h_{t-1}^2\end{aligned}$$

where $\omega > 0$, $\alpha \geq 0$ and $\beta \geq 0$ and ε_t is an *i.i.d.* sequence of random variables following either a Gaussian $N(0, 1)$ or a Student-t $St(0, 1, \nu)$. Once that we have estimated the parameters, we compute the forecast of the variance of the deseasonalized returns and the intraday VaR for r_t at a confidence level τ as

$$\begin{aligned}VaR_t(\tau|d, \underline{\tilde{r}_{t-1}})_G &= z_\tau^G \sqrt{\hat{h}_t \phi_d} \\ VaR_t(\tau|d, \underline{\tilde{r}_{t-1}})_{St} &= z_\tau^{St} \sqrt{\hat{h}_t \phi_d}\end{aligned}\tag{1.5.1}$$

where z_τ^G and z_τ^{St} denote the τ -th quantiles of a standard Gaussian and Student-t distributions respectively, $\underline{\tilde{r}_{t-1}}$ is the set of past adjusted returns and \hat{h}_t is the one-step ahead forecasted variance.

Table 1.5 reports the estimation results for the two distributions. Notice that these results refer to the standardized returns series, while in Table 1.2 we reported the estimation results on raw data. These results confirm that there is a strong seasonal component in the intraday return and that forgetting about that can be misleading. Indeed, for all the deseasonalized returns, Table 1.5, the estimated degrees of freedom of the Student t model are larger than the ones obtained on the raw returns even if the increase for ANA seems to be marginal. Yet, both the Gaussian and the Student-t models are close to be integrated. Table 1.6 reports Kupiec test and the empirical failure rates the VaR forecast for the three stocks, computed like in (1.5.1) and using the same rolling windows as for quantile regression. Both models fail to forecast correctly -for all the stocks and all the confidence levels.¹³ With a Gaussian distribution failure rates are systematically bigger than the theoretical ones while the GARCH(1,1) with a Student t distribution does the opposite. This means that the Gaussian model underestimates the risk -assigns too little mass to the tails of the distribution- and the Student t overestimates it -assigns too much mass to the tails. This makes evident the advantage of using a semiparametric method such as quantile regression that does not require any assumption on the underlying distribution.

1.6 Conclusions

We investigate intraday seasonal patterns on the probability law of high frequency financial returns. Within the day there are significant variations in asset prices, which imply different evaluations of the tails of the return's distribution through the day. And these variations are partly deterministic and due to the intraday seasonality. As returns are realizations of a random variable and as such their behavior is fully described by their conditional probability law. To analyze the intraday behavior of the probability law, we use quantile regression, where the

¹³We do not show results for the *LR* tests as the model already failed to pass the simple Kupiec test. They are available under request.

regressors are Fourier series that capture the time of the day and past absolute returns.

Using quote midpoints of three stocks traded at the Spanish stock exchange from January 2001 to December 2003, we show that indeed the conditional probability distribution depends on the time of the day. At the opening and closing the density flattens and the tails become thicker, while in the middle of the day returns concentrate around the median and the tails are thinner. Results are intuitive, in the sense that they confirm the general perception that in the opening and closing the probabilities of finding large price fluctuations are higher than at lunch. Results, in terms of quantiles, permit straightforward intraday risk evaluations, such as value at risk. We show the intraday variation of the maximum expected loss at 2.5%, 1% and 0.5% confidence levels. The maxima expected losses are, as expected, maximal at the opening and closing and minimal at lunch time. Moreover the test performed on the out-of-sample forecasts of the value at risk show that the model is able to provide good risk assessments contrary to the standard GARCH(1,1) models.

1.7 Appendix

In this appendix, we describe the estimation procedure that we followed for the estimation of the asymptotic covariance matrix of the QR estimates. We follow Koenker (2005). Consider the basic model presented in equation (1.3.1). The asymptotic distribution of the QR estimator in a non-iid setting

$$\sqrt{T}(\hat{\beta}(\tau) - \beta(\tau)) \rightarrow N(0, \tau(1 - \tau)H_T^{-1}J_T H_T^{-1})$$

where

$$J_T(\tau) = T^{-1} \sum_{t=1}^T x_t x_t'$$

and

$$H_T(\tau) = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^n x_t x_t' f_t(\xi_t(\tau)) \quad (1.7.1)$$

and $f_t(\xi_t(\tau))$ denotes the conditional density of the r_t evaluated at the τ -th percent conditional quantile. The asymptotic covariance among estimates at different quantiles has blocks

$$Cov(\sqrt{T}(\hat{\beta}(\tau_t) - \beta(\tau_t)), \sqrt{T}(\hat{\beta}(\tau_s) - \beta(\tau_s))) = [\tau_t \wedge \tau_s - \tau_t \tau_s] H_T(\tau_t)^{-1} J_T H_T(\tau_s)^{-1}$$

The conditional density $f_t(\xi_t(\tau))$ in (1.7.1) is estimated using the Hendricks and Koenker (1992) sandwich form. This estimation procedure requires at first to compute the optimal bandwidth for each τ , h_T . To do it, we used the optimal bandwidth suggested by Bofinger(1975)

$$h_T = T^{1/5} \left(\frac{4.5\phi^4(\Phi^{-1}(\tau))}{(2\Phi^{-1}(\tau)^2 + 1)^2} \right)^{1/5}$$

where T is the sample size, ϕ is the normal pdf and Φ^{-1} is the normal quantile function, i.e. the inverse of the normal cdf. Last, we re-perform the QR estimation for the grids $\tau - h_n$ and $\tau + h_n$.

As showed in Figure 1.3, the cdf can be obtained inverting the quantile function and, once that we have the cdf, we can recover the density function differentiating. Following this intuition, Hendricks and Bofinger suggest to estimate the conditional density function as

$$\hat{f}_t = \max\{0, 2h_T/(x'_t\hat{\beta}(\tau + h_T) - x'_t\hat{\beta}(\tau - h_T) - \varepsilon)\} \quad (1.7.2)$$

where $\hat{\beta}(\tau + h_n)$ and $\hat{\beta}(\tau - h_n)$ are the estimated parameters at $\tau - h_n$ and $\tau + h_n$ and ε is a small tolerance parameter that we fixed to 0.01 to avoid dividing by zero.

Table 1.1: Descriptive statistics at different hours of the day

TEF				
	Mean	S. Dev	Skew	Kurt
09:30	0.000	0.065	-0.283	7.053
12:00	0.002	0.035	-0.196	5.945
15:15	0.000	0.035	-0.274	7.296
17:15	-0.002	0.048	-0.559	6.062

ELE				
	Mean	S. Dev	Skew	Kurt
09:30	-0.004	0.063	-0.076	6.512
12:00	0.001	0.039	0.353	9.017
15:15	-0.001	0.033	-1.075	12.684
17:15	0.002	0.051	0.183	6.066

ANA				
	Mean	S. Dev	Skew	Kurt
09:30	-0.008	0.136	-0.604	8.077
12:00	0.005	0.080	-0.249	8.368
15:15	-0.003	0.084	-2.338	34.049
17:15	-0.002	0.111	0.253	6.290

The first column reports the time of the day to which the statistics refer. The second displays the sample mean of all the observation at the selected time of the day. The third the sample standard deviation. The fourth the bias corrected skewness and the last one shows the bias adjusted kurtosis.

Table 1.2: GARCH(1,1) estimates with Student-t distribution

	ω	α	β	ν
TEF	0.048	0.270	0.730	3.48
ELE	0.097	0.354	0.646	3.35
ANA	0.164	0.424	0.576	2.49

Student's t GARCH(1,1) $h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}$ estimates. ν stands for degrees of freedom.

Table 1.3: Kupiec test on the VaR forecasts

	VaR(2.5%)		VaR(1%)		VaR(0.5%)	
TEF	2.37	(2.17 2.56)	1.00	(0.88 1.13)	0.54	(0.44 0.63)
ELE	2.33	(2.13 2.52)	1.04	(0.91 1.17)	0.58	(0.48 0.67)
ANA	2.54	(2.34 2.75)	1.06	(0.93 1.20)	0.58	(0.48 0.68)

Empirical failure rates of the VaR forecasts at the confidence levels of 2.5% 1% and 0.5%. Confidence intervals in parenthesis and values are in percentage.

Table 1.4: Christoffersen's likelihood ratio test on the VaR forecasts

LR_{uc}						
	VaR(2.5%)		VaR(1%)		VaR(0.5%)	
TEF	1.68	(0.19)	0.00	(0.96)	0.66	(0.42)
ELE	2.81	(0.10)	0.41	(0.52)	2.52	(0.11)
ANA	0.16	(0.69)	0.83	(0.36)	2.72	(0.10)

LR_{ind}						
	VaR(2.5%)		VaR(1%)		VaR(0.5%)	
TEF	1.06	(0.30)	0.15	(0.69)	1.38	(0.24)
ELE	0.96	(0.33)	1.96	(0.16)	1.38	(0.24)
ANA	0.50	(0.48)	1.26	(0.26)	0.06	(0.80)

LR_{cc}						
	VaR(2.5%)		VaR(1%)		VaR(0.5%)	
TEF	2.74	(0.25)	0.16	(0.92)	2.04	(0.36)
ELE	3.77	(0.15)	2.37	(0.31)	3.89	(0.14)
ANA	0.66	(0.72)	2.09	(0.35)	2.78	(0.25)

Christoffersen's likelihood ratio test for the the VaR forecasts at the confidence levels of 2.5% 1% and 0.5%. The first panel presents results for the Christoffersen's likelihood ratio test of unconditional coverage, LR_{uc} with p-values in parenthesis. The second panel presents results for Christoffersen's likelihood ratio test of independence, LR_{ind} with p-values in parenthesis. Last panel presents results for Christoffersen's joint likelihood ratio test of coverage and independence, LR_{cc} with p-values in parenthesis.

Table 1.5: GARCH(1,1) estimates of the standardized returns

Gaussian				
	ω	α	β	
TEF	0.001	0.032	0.967	
ELE	0.002	0.029	0.970	
ANA	0.022	0.060	0.920	

Student's t				
	ω	α	β	ν
TEF	0.001	0.035	0.965	6.018
ELE	0.008	0.062	0.934	4.681
ANA	0.125	0.266	0.734	2.501

GARCH(1,1) $h_t = \omega + \alpha \tilde{r}_{t-1}^2 + \beta h_{t-1}$ estimates. ν stands for degrees of freedom.

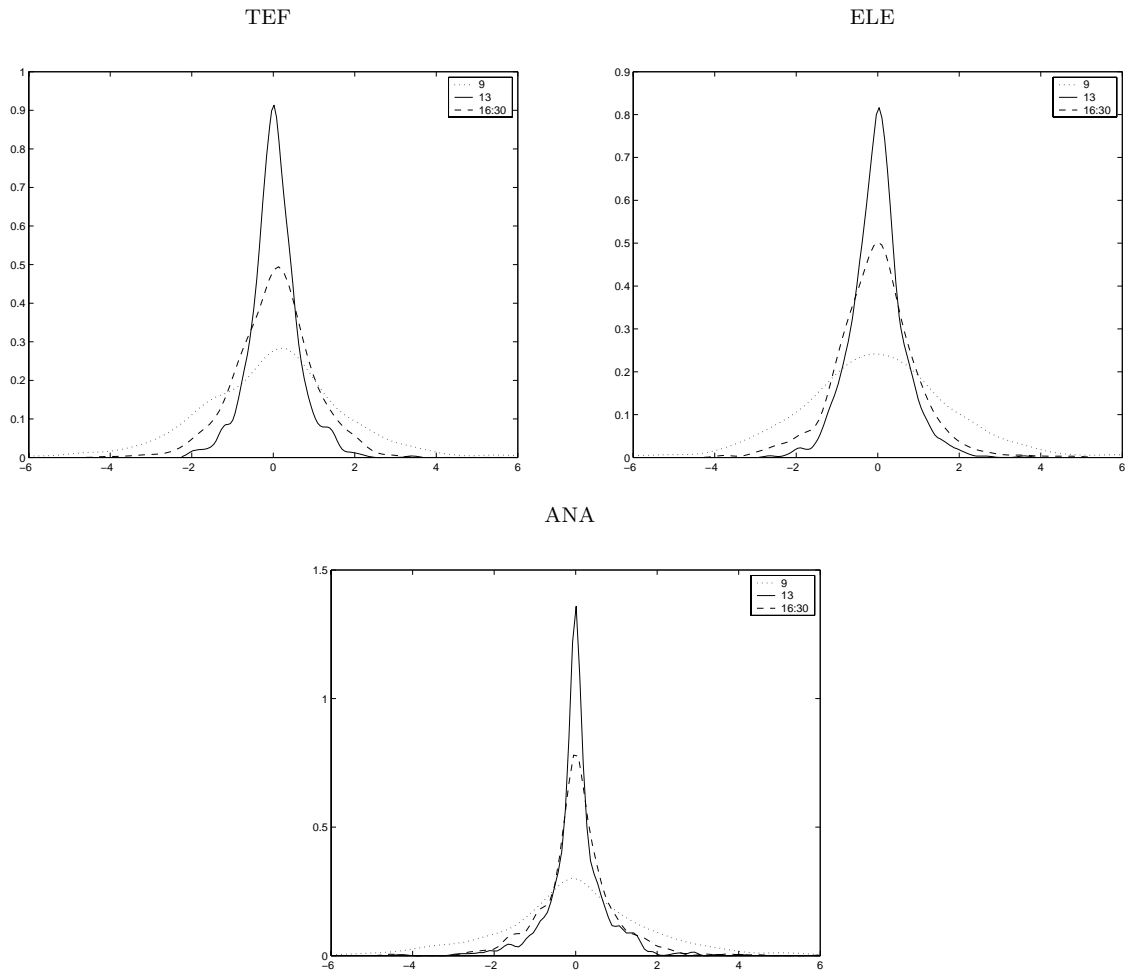
Table 1.6: Kupiec test on the VaR forecasts

Gaussian									
	VaR(2.5%)			VaR(1%)			VaR(0.5%)		
TEF	2.78	(2.57	2.99)	1.56	(1.40	1.72)	1.07	(0.94	1.20)
ELE	2.86	(2.64	3.07)	1.62	(1.46	1.79)	1.08	(0.95	1.22)
ANA	3.14	(2.91	3.37)	2.07	(1.88	2.25)	1.56	(1.39	1.72)

Student's t									
	VaR(2.5%)			VaR(1%)			VaR(0.5%)		
TEF	1.62	(1.46	1.78)	0.63	(0.53	0.73)	0.37	(0.29	0.44)
ELE	1.20	(1.06	1.34)	0.42	(0.34	0.50)	0.18	(0.13	0.24)
ANA	0.38	(0.30	0.46)	0.10	(0.06	0.14)	0.03	(0.01	0.05)

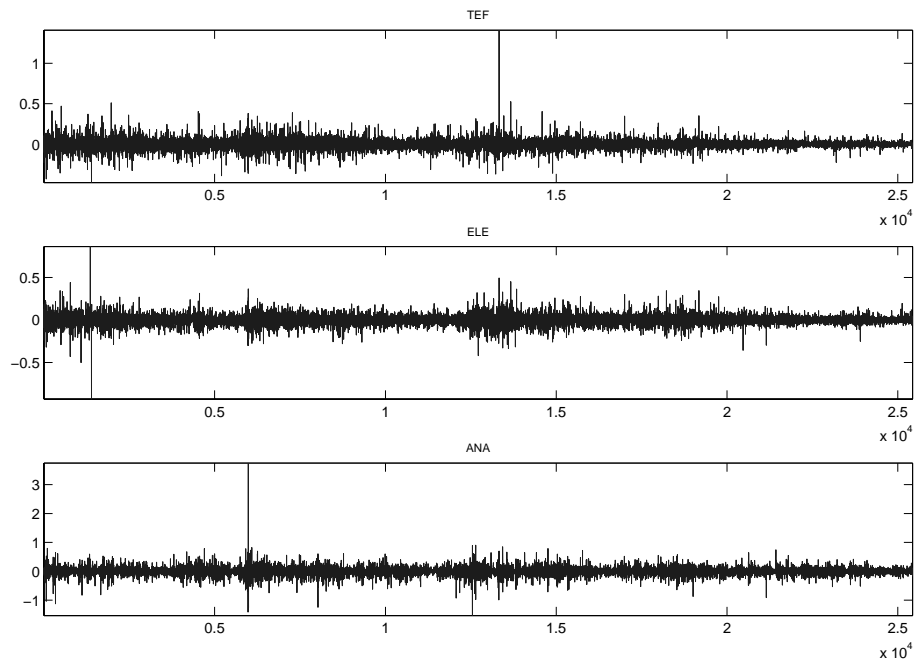
Empirical failure rates of the VaR forecasts using a GARCH(1,1) at the confidence levels of 2.5% 1% and 0.5%. Confidence intervals in parenthesis and values are in percentage.

Figure 1.1: Kernel estimates at different hours of the day.



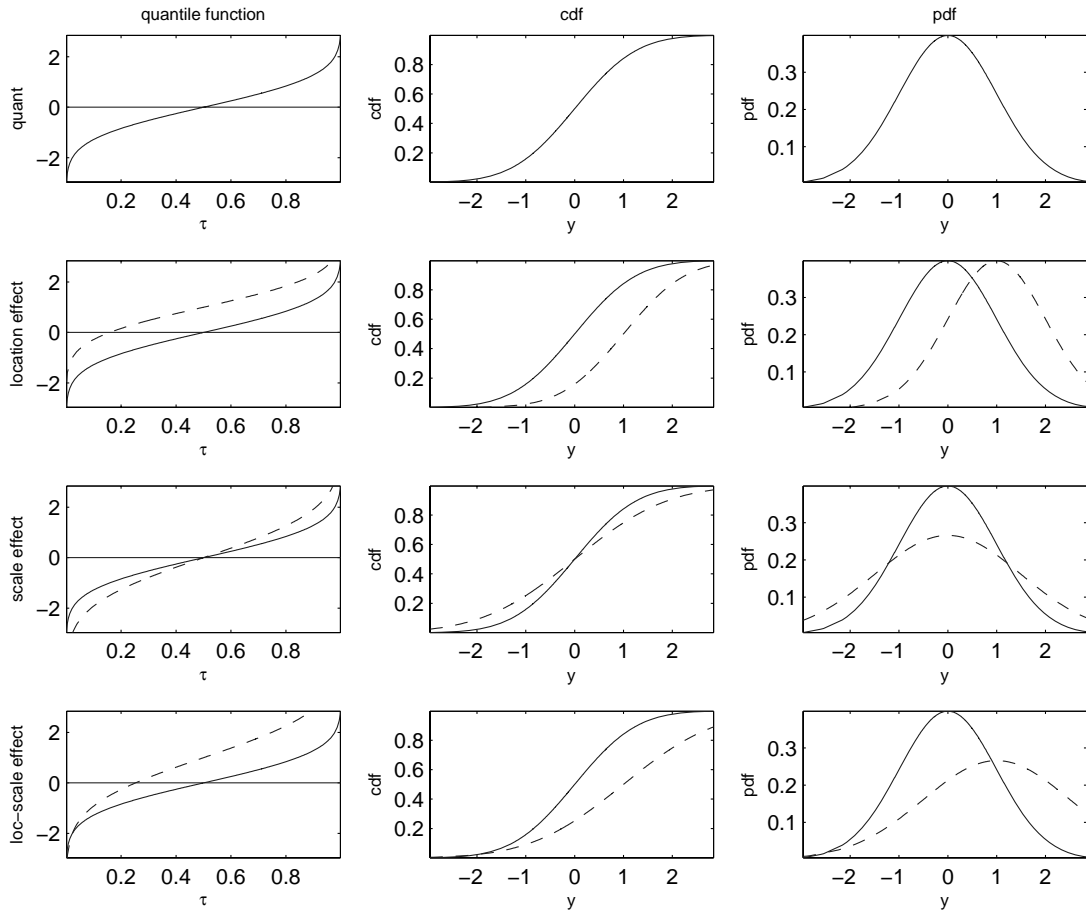
Nonparametric density estimate of the 15 minutes returns at different hours of the day. For each day, we included the observation at the selected hour, therefore each sample contains a number of observation equal to the number of days. The estimate is based on a Gaussian kernel with optimal bandwidth.

Figure 1.2: 15 minutes returns



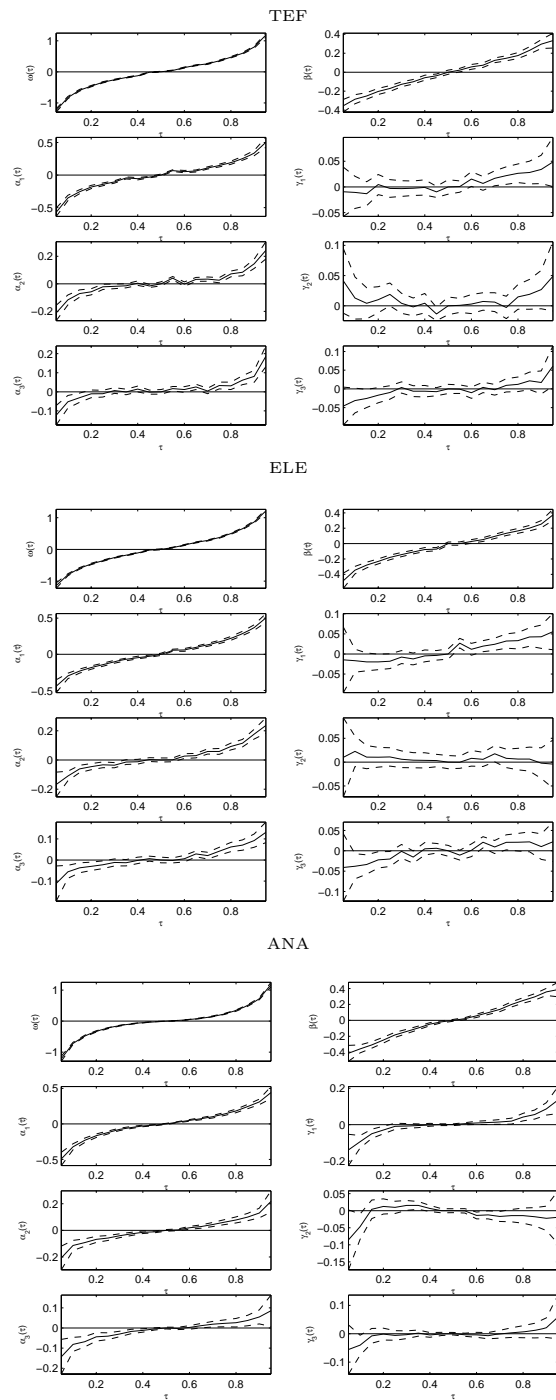
Standardized 15 minutes returns. The sample period runs from January 2001 to December 2003. For each stock there are 34 intraday observations for a total of 25 400.

Figure 1.3: Location and scale shifts in the pdf through the quantile function



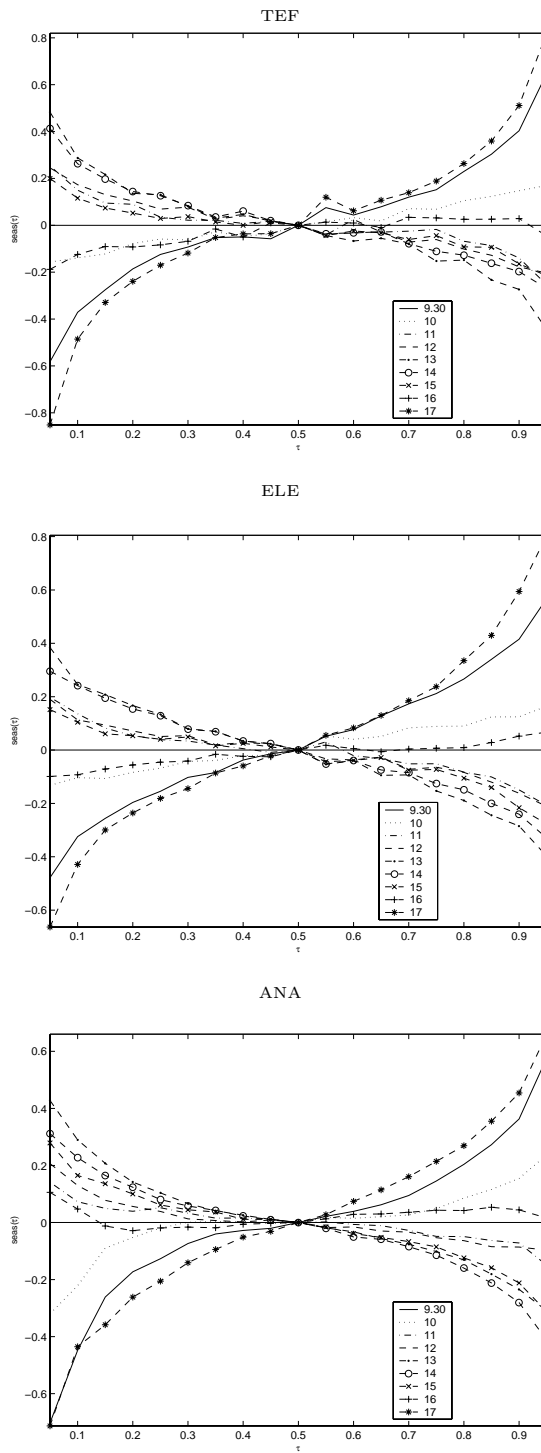
Top row shows the pdf, cdf and quantile function of a standardized normal. For the other three rows, the continuous line indicates the pdf, cdf and quantile function of the standardized normal. The dashed line in the second row refers to the pdf, cdf and quantile function of a normal with mean 1 and variance 1. In the third row the dashed line indicates a normal with mean 0 and variance 1.5 and in the last row a normal with mean 1 and variance 1.5.

Figure 1.4: Estimated parameters



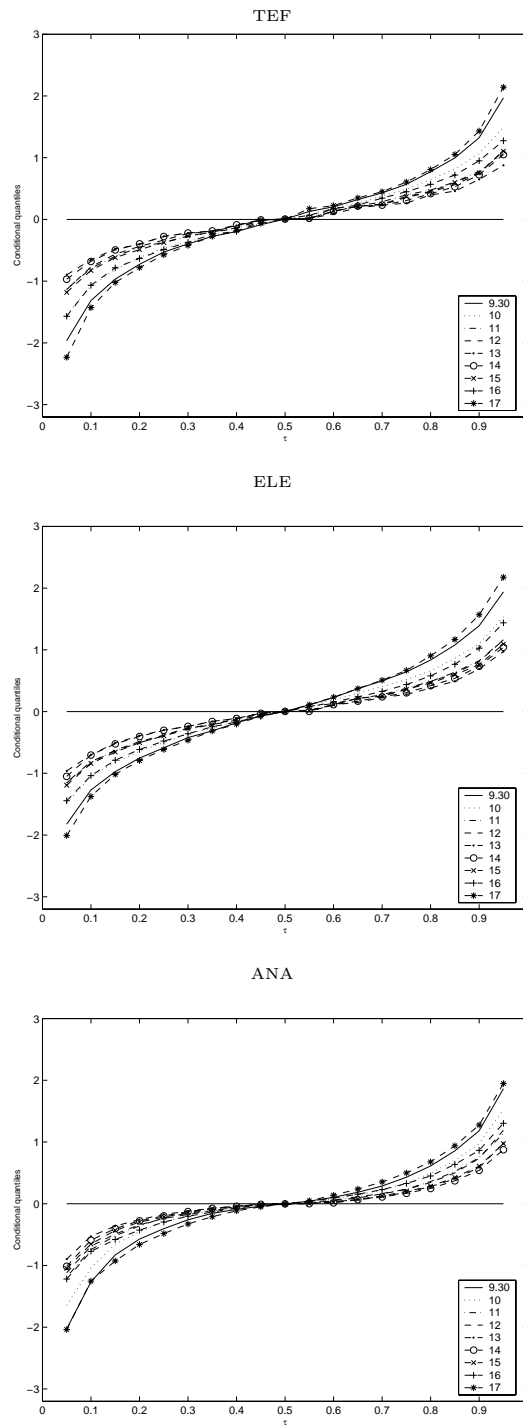
The figure displays the estimated parameters of equation (1.3.3). The continuous line indicates the estimated parameters for each τ quantile. The dashed one refers to the 5% point-wise confidence intervals.

Figure 1.5: Seasonal component



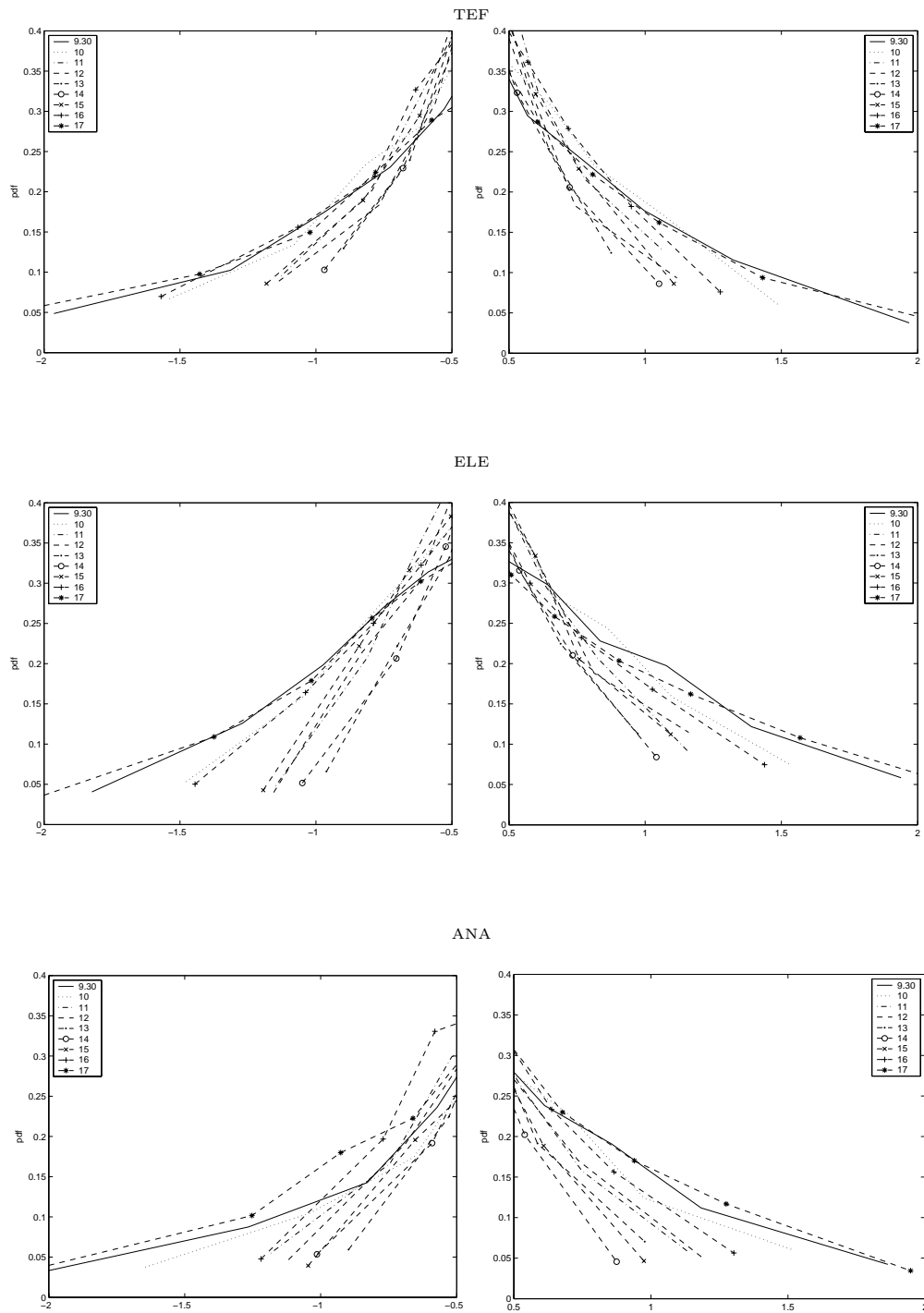
Estimated seasonal component $séas_t(\tau)$, as presented in equation (1.3.2), for different times of the day.

Figure 1.6: Seasonality in the quantiles



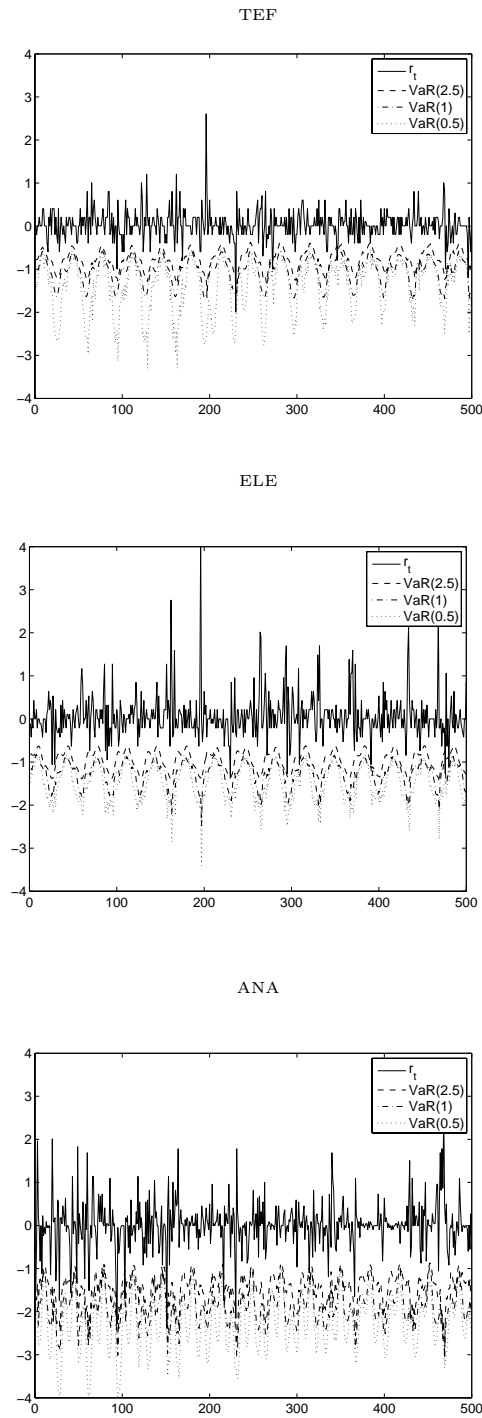
Conditional quantiles of r_t given $|r_{t-1}|$ equal to its 50 percent empirical quantile, $Q_{r_t}(\tau|t, |r_{t-1}| = Q_{|r_{t-1}|}(0.50))$, and for different times of the day

Figure 1.7: Seasonality and the tails



Left and right tail of the conditional densities of r_t given $|r_{t-1}|$ equal to its 50 percent empirical quantile for different times of the day. The conditional density is computed using equation (1.7.2) in the Appendix.

Figure 1.8: VaR



Last 500 standardized 15 minutes returns and the relative out of sample Value at Risk forecast at confidence levels 2.5%, 1% and 0.5%. 39

Chapter 2

Forecasting the yield curve using large macroeconomic information

ABSTRACT: This paper investigates whether macroeconomic indicators are helpful in forecasting the yield curve. We incorporate a large number of macroeconomic predictors within the Nelson and Siegel (1987) model for the yield curve which can be cast in a common factor model representation. Estimation is performed by EM algorithm and Kalman filter using a data set composed by 17 yields and 118 macro variables. Results show that incorporating large macroeconomic information improves the accuracy of out-of-sample yield forecasts at medium and long horizons.

Keywords: Yield Curve, Factor Models, Forecasting, Large Cross-Sections, Quasi Maximum Likelihood.

JEL classification: C33, C53, E43, E44.

This chapter is adapted from the working paper "Forecasting the term structure of interest rates using a large panel of macroeconomic data" with Domenico Giannone (ECB and Universite Libre de Bruxelles) and Michele Modugno (ECB and Universite Libre de Bruxelles).

2.1 Introduction

The interaction between the yield curve and macro variables is a clear phenomenon manifested in the behavior of market agents and policy makers. On one hand, market participants closely monitor macro data releases and try to assess their impact in the yields, see for example Fleming and Remolona (1999) and Furfine (2001). On the other one, central banks, in the standard view of the monetary transmission mechanisms, react to current macroeconomic situation, stimulating aggregate demand and controlling inflation by fixing the short term interest rates. Following the expectation theory, long term interest rates depend on present and future expected short term interest rates. This suggests that macro variables can incorporate important information in order to forecast the behavior of market and central bank practitioners, and thereby the evolution of the yield curve.

The term structure of interest rates is characterized by a high degree of correlation among yields with different maturities. This collinearity can be explained by few sources of co-movement. As a consequence, a parsimonious representation of the yield curve can be obtained by modeling fewer factors than observed maturities. Accordingly, the two main approaches for yield curve modeling can be cast in a factor model representation, which differ from each other for the restrictions imposed on the model parameters. The first approach, the Nelson and Siegel (1987) model, is a parsimonious model based on the relation between the yields and their corresponding maturities. This model is able to reproduce the historical stylized facts concerning the average shape of the yield curve, the variety of shapes assumed at different times and the strong persistence of yields. Moreover, Diebold and Li (2006) reinterpret the Nelson and Siegel model as a dynamic three latent factors model with restricted loadings and show that it is able to provide good forecasts of the yield curve. The second approach, the no-arbitrage term structure models, is characterized by restrictions on the factor loadings that rule out arbitrage opportunities. These models impose a structure on the factor loadings such that the resulting yield curves, in the maturity dimension, are compatible with the time-

series dynamics of the yield curve factors. This consistency between the dynamic evolution of the yield curve factors, and hence the yields at different maturities, is what ensures the absence of arbitrage opportunities and makes these models particularly useful for derivative pricing.¹ However the imposition of the no-arbitrage restrictions on the term structure models imply that the resulting model is clearly not parsimonious and therefore not suitable for forecasting purposes.² This is confirmed by Duffee (2002) who finds that this type of models forecasts the yield curve poorly. Accordingly, given that the main focus of this paper is on forecasting of the term structure of interest rates, we adopt the more parsimonious Nelson and Siegel approach.

In the literature there has been a lot of interest in the relationship between macroeconomic variables and the yield curve. In their seminal paper, Ang and Piazzesi (2003) study the interactions between yields and macroeconomic variables augmenting the standard no-arbitrage affine term structure model with two observable macroeconomic factors, measuring inflation and real activity. The idea behind this model is to use macroeconomic variables to capture the variability of the yields not explained by the latent factors, improving the forecasting performance of the model. Their main conclusion is that macroeconomic factors help in forecasting the yield curve. Following this finding, several papers investigate the links between the yield curve and macroeconomic variables, incorporating macro determinants as factors, into multi factor no-arbitrage affine models. Among others Dai and Philippon (2005), Dewachter and Lyrio (2006), Kozicki and Tinsley (2001), Wu (2006). Mönch (2005) uses, as additional factors, principal components extracted from a large macroeconomic data- set, instead of single macro variables.³ Rudebusch and Wu (2004) and Hördhal, Tristani and Vestin (2006) develop a theoretical framework which allows to identify the sources of co-movements by structural macroeconomic

¹For more details about the no-arbitrage term structure models see Duffie and Kan (1996), Litterman and Scheinkman (1991) and Dai and Singleton (2000).

²For a more detailed comparison between the Nelson and Siegel model and the no-arbitrage term structure models see Chapter 3.

³A similar approach is used to forecast the excess bond returns in Ludvigson and Ng (2005).

relations. On the other side, following the Nelson and Siegel approach, Diebold, Rudebush and Aruoba (2006) introduce a yield curve model where, in addition to the Nelson and Siegel latent factors, they include some observable macroeconomic factors. They show that observable macroeconomic factors have strong effects on the future yield curve and that there is evidence of reverse influence. Mönch (2006) proposes to use principal components extracted from a large data set of macro economic variables to augment the Nelson and Siegel latent factors. Favero, Niu and Sala (2007) and De Pooter, Ravazzolo and Van Dijk (2007) investigate the impact of macro variables on the forecast of yields. They provide an exhaustive comparison of the existing yield curve models, with and without macro factors, and they find that additional factors extracted from large macro dataset are important for yield curve forecasting.

To summarize, the general idea behind the previous literature is to use macro variables as extra factors to capture the co-movement among yields not explained by the yield curve latent factors. One can raise three criticisms against this approach. First, the idea behind factor models is parsimony. Augmenting the number of factors goes against this notion, specially if the three latent factors already explain most of the variation of the yields. Moreover, these latent factors are frequently identified as proxies of macro variables, therefore adding macro variables can be redundant. Second, adding macro variables as factors has not been proven successful at improving the out-of-sample performance of these models. This can be due to the fact that the gain of exploiting a larger information set does not counterbalance the loss in terms of lack of parsimony. Third, this approach allows to exploit only a small set of macro variables. One could use principal components to summarize the information content of a larger macro variables set, but in this way it would not be possible to understand which macroeconomic variables are important in fitting and forecasting the yield curve. Moreover, in the presence of a high correlation among the principal components and the latent yields curve factors, there would be a problem of parsimony.

In this paper, we propose a model for forecasting the yield curve using parsimoniously a large amount of macroeconomic information. We suggest an innovative way to exploit the linkages between macroeconomic variables and yields. Rather than including macroeconomic variables as factors in the yield curve model, we directly extract the latent factors from a data set composed of both yields (seventeen series) and macro variables (one hundred-eighteen). The macroeconomic variables considered in the analysis include real variables (sectorial industrial production, employment and hours worked), nominal variables (consumer and producer price indices, wages, and money aggregates) and asset prices (stock prices and exchange rates). To identify the yield curve factors, we impose the Nelson and Siegel restrictions on the loadings relative to the yields, leaving the loadings relative to macro variables free. This allows to enrich the yield curve latent factors with the information contained in the macroeconomic variables. This approach allows to preserve parsimony and include a large amount of information at the same time, since it is not necessary to augment the standard three latent factor models in order to include the additional information coming from the macroeconomic variables. Indeed, as shown by Diebold and Li (2006), the Nelson and Siegel factors are highly correlated with some macro variables, in particular with measures of inflation and industrial production. This means that the sources of co-movement for the yields co-move with the rest of the economy. Accordingly, in the aim of the factor model literature, extracting the latent factors from a panel of yields enriched with a large amount of macroeconomic variables, allows us to better identify them. Moreover, by looking at the loadings it is possible to discriminate which macro variables have significant information content for each factor. Related to this work is Law (2006), who extracts the latent factors, using a no arbitrage model, from twenty-four macro series plus the yields. We differ from him in several aspects. First, we exploit a broader quantity of information. Second, we perform an out of sample forecast exercise. Third, we use the Nelson and Siegel approach.

We estimate the model by maximum likelihood combining EM algorithm and

Kalman filter. Doz, Giannone and Reichlin (2006) show that this procedure makes maximum likelihood estimation of factor models feasible for large cross sections, in the sense that it delivers consistent estimates for large cross sections and large sample sizes (for any relative size of the time span and cross sectional dimension). Consistency is guaranteed even if the hypothesis of orthogonality and of absence of serial correlation are violated for the idiosyncratic part. Moreover, this methodology allows us to impose the crucial restrictions on the loadings to identify the factors.

Results show that the out-of-sample forecasting performance improves at middle and long horizons (i.e. 6 and 12-months ahead) compared with the forecasts generated by a model estimated using only the yields, the ones generated by a model *à la* Diebold, Rudebusch and Aruoba (2006) where the Nelson and Siegel factors are augmented with three macroeconomic variables (the manufacturing capacity utilization, the federal funds rate and the annual price inflation), a model *à la* Mönch (2006) where the Nelson and Siegel factors are augmented with the first three principal components extracted from the same macro dataset and the ones generated by a random walk.

The paper is organized as follows. Section 2.2 introduces the *macro-yields* model and the four alternative models considered in the analysis. Section 2.3 presents the data describing how the yields are constructed and providing a description of the macroeconomic dataset. Section 2.4 describes the estimation technique used, with a more detailed description in Appendix, and derives the modified Bai and Ng (2002) information criterion that we use for model selection. Section 2.5 shows the estimation results of the *macro-yields* model and the in sample performances of the proposed models. The importance of using parsimoniously macro variables becomes clear in Section 2.6 where we compare the forecasting performances of the proposed model. Section 2.7 concludes.

2.2 Model

The Nelson and Siegel (1987) model is a parsimonious model to fit yields of different maturities at a specific point in time. Diebold and Li (2006) reinterpret the Nelson and Siegel model as a latent factor model, where the evolution in time of the yields depends on three latent factors identified as level, slope and curvature through the restrictions on their relative factor loadings. Denoting with $y_t(\tau_i)$ the yield of maturity τ_i at time t , the Nelson and Siegel model, as reinterpreted by Diebold and Li (2006) can be expressed as

$$y_t(\tau_i) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} \right) + C_t \left(\frac{1 - e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i} \right) + v_t(\tau_i) \quad (2.2.1)$$

where the level, slope and curvature are denoted by L_t , S_t and C_t , and $v_t(\tau_i)$ is the residual, or pricing error. The predetermined loadings $(1, \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}, \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i})$ allow to identify the three factors as level, slope and curvature of the yield curve because of the effects that they have on its shape. The loadings relative to the first factor, equal to one for all maturities, imply that an increase in L_t increases all yields equally, shifting the level of the yield curve. The loadings of the second factor are high for short maturities decaying to zero for the long ones. Accordingly, as increase in S_t increases the slope of the yield curve. The loadings relative to C_t are zero for the shortest and the longest maturities, reaching the maximum for medium maturities. Therefore, an increase of C_t augments the curvature of the yield curve. The parameter λ governs the exponential decay rate, a small value of λ can better fit the yield curve at long maturities, while large values can better fit it at short maturities. Diebold and Li (2006) keep this parameter constant over time. Rewriting equation (2.2.1) in vector notation

$$y_t = \Gamma_{yf}^* f_t + v_{y,t} \quad (2.2.2)$$

where y_t collects the yields of different maturities available at time t , Γ_{yf}^* is the matrix of restricted factor loadings with row i equal to $(1, \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i}, \frac{1-e^{-\lambda\tau_i}}{\lambda\tau_i} - e^{-\lambda\tau_i})$ and f_t is the vector of factors $(L_t, S_t, C_t)'$.

The aim of this paper is to introduce a parsimonious model that exploits all the information about the state of the economy in order to fit and forecast the yield curve. To summarize all the information in the macroeconomic variables, we do not add any specific macroeconomic variable as a factor, neither we add principal components extracted from a macroeconomic data set. We rather extract the level, slope and curvature from a large panel composed by yields and macroeconomic variables. Generalizing the Nelson and Siegel factor model of equation (2.2.2), we have

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \Gamma_{yf}^* \\ \Gamma_{xf} \end{pmatrix} f_t + \begin{pmatrix} v_{y,t} \\ v_{x,t} \end{pmatrix} \quad (2.2.3)$$

where y_t is the vector of yields, x_t is a large set of macroeconomic variables and f_t collects the yield curve latent factors. To identify the three unobservable factors as level, slope and curvature, we restrict the matrix Γ_{yf}^* of factor loadings relative to the yields *à la* Nelson and Siegel as in equation (2.2.2). While the matrix Γ_{xf} , that collects the loadings relative to the macro variables, is left unrestricted.

Rather than including macro variables as additional factors, we use them to extract the Nelson and Siegel factors. Yields co-move with the whole economy, therefore the few sources that generate the evolution of the yields have to be related to the whole economy. This implies that extracting the Nelson and Siegel factors not only from yields but also from macro variables allows to use the extra information of the macro variables to better identify the factors. This feature is in line with the previous macro-finance literature, which links the level with different measures of inflation and the slope with capacity utilization or industrial production, see among others Diebold and Li (2006), Diebold, Rudebusch and Aruoba (2006), Rudebusch and Wu (2004) and Hördhal et al (2002). Moreover, using a

large set of macro-variables allows to discriminate, through the loadings, which variables are related to the factors and useful to forecast out-of-sample the yields.

The model presented in equation (2.2.3) can be easily extended to allow the presence of additional unobservable and unidentified factors in the following way

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \Gamma_{yf}^* & 0 \\ \Gamma_{xf} & \Gamma_{xg} \end{pmatrix} \begin{pmatrix} f_t \\ g_t \end{pmatrix} + \begin{pmatrix} v_{y,t} \\ v_{x,t} \end{pmatrix} \quad (2.2.4)$$

In this model, both yields and macro-variables participate in determining the yield curve factors (level, slope and curvature) collected in the vector f_t , while the factors collected in g_t are determined only by the macro variables and are unidentified since we do not impose any restriction on the matrix of factor loadings Γ_{xg} . In this model, even if we added additional and unidentified factors to explain the variation in the large data set of macro variables, the yields still load only on the three yield curve factors. This is consistent with the standard view that three factors are able to exploit all the information in the yield curve.⁴

The model presented in equation (2.2.4) can be easily put in a state-space representation. The *macro-yields* model that we propose is

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \Gamma_{yf}^* & 0 \\ \Gamma_{xf} & \Gamma_{xg} \end{pmatrix} \begin{pmatrix} f_t \\ g_t \end{pmatrix} + \begin{pmatrix} v_{y,t} \\ v_{x,t} \end{pmatrix} \quad (2.2.5)$$

$$\begin{pmatrix} f_t \\ g_t \end{pmatrix} = A \begin{pmatrix} f_{t-1} \\ g_{t-1} \end{pmatrix} + \begin{pmatrix} u_{f,t} \\ u_{g,t} \end{pmatrix} \quad (2.2.6)$$

where $v = (v_{y,t}, v_{x,t}) \sim iid N(0, R)$ and $u = (u_{f,t}, u_{g,t}) \sim iid N(0, Q)$, with a diagonal variance matrix of the idiosyncratic disturbances R and a non-diagonal variance matrix of the shocks driving the common factors Q .

⁴We also estimated the model allowing Γ_{yg} to be different from zero and, as expected, the estimated loadings were small in magnitude and not significant. All the results presented in the paper do not change allowing Γ_{yg} to be different than zero and are available upon request.

2.2.1 Alternative models

We compare the *macro-yields* model, presented in equations (2.2.5) - (2.2.6), with three alternative models: the *only yields* model, the *basic macro-yields* model and the *large macro-yields* model.

The *only yields* model uses only the information contained in the yields series to extract the yield curve factors. This model is a generalization of the Diebold and Li (2006) one, which has been showed to outperform several models in forecasting the U.S. yields. It can be represented as

$$y_t = \Gamma_{yf}^* f_t + v_{y,t}, \quad v_{y,t} \sim iid N(0, R) \quad (2.2.7)$$

$$f_t = A f_{t-1} + u_{f,t}, \quad u_{f,t} \sim iid N(0, Q) \quad (2.2.8)$$

where the matrix of factor loadings of the yields Γ_{yf}^* is restricted *à la* Nelson and Siegel. This model can be obtained from the *macro-yields* model, presented in equations (2.2.5) - (2.2.6), imposing the following restrictions: the macro variables do not participate in the determination of the yield curve factors, i.e. $\Gamma_{xf} = 0$, and there are only the three yield curve factors, i.e. $\Gamma_{xg} = 0$.

The *basic macro-yields* model augments the yields curve factors with a minimal set of fundamental macro variables as extra factors to capture basic macroeconomic dynamics. In particular, following Diebold, Rudebush and Aruoba (2006), we consider as additional factors the manufacturing capacity utilization (CU), the federal funds rate (FFR) and the annual price inflation (INFL). Therefore imposing $g_t = (CU_t, FFR_t, INFL_t)$, the *basic macro-yields* model can be written as

$$y_t = \Gamma_{yf}^* f_t + \Gamma_{yg} g_t + v_{y,t}, \quad v_{y,t} \sim iid N(0, R) \quad (2.2.9)$$

$$\begin{pmatrix} f_t \\ g_t \end{pmatrix} = A \begin{pmatrix} f_{t-1} \\ g_{t-1} \end{pmatrix} + u_t, \quad u_t \sim iid N(0, Q) \quad (2.2.10)$$

with Γ_{yf}^* restricted *à la* Nelson and Siegel to identify the three yield curve factors. This model can be obtained from the *macro-yields* model, presented in equations

(2.2.5) - (2.2.6), imposing that the yield curve factors are extracted only from the yields, i.e. $\Gamma_{xf} = 0$, that the yields load both on the yield curve factors and the macro factors, i.e. $\Gamma_{yg} \neq 0$, that the additional factors are equal to the manufacturing capacity utilization (CU), the federal funds rate (FFR) and the annual price inflation (INFL), i.e. $g_t = (CU_t, FFR_t, INFL_t)$, and coincide with the additional macro variables, i.e. $\Gamma_{xg} = I$ and $R_x = 0$.

The *large macro-yields* model exploits a larger information set with respect to the *basic macro-yields* models. This model augments the three yield curve factors, extracted only from the yields, with the first three principal components extracted from the large data-set of macroeconomic variables. Denoting with PC_t the vector of three principal components at time t , the *large macro-yields* model can be represented as

$$y_t = \Gamma_{yf}^* f_t + \Gamma_{yg} PC_t + v_{y,t}, \quad v_{y,t} \sim iid N(0, R) \quad (2.2.11)$$

$$\begin{pmatrix} f_t \\ PC_t \end{pmatrix} = A \begin{pmatrix} f_{t-1} \\ PC_{t-1} \end{pmatrix} + u_t, \quad u_t \sim iid N(0, Q) \quad (2.2.12)$$

with Γ_{yf}^* restricted *à la* Nelson and Siegel to identify the three yields curve factors. Also this model can be considered as a restricted version of the *macro-yields* model, presented in equations (2.2.5) - (2.2.6), with the following restrictions: the yield curve factors are extracted only from the yields, i.e. $\Gamma_{xf} = 0$, the yields load both on the yield curve factors and the macro factors, i.e. $\Gamma_{yg} \neq 0$, the additional factors are equal to the principal components extracted from a large dataset of macro variables $g_t = PC_t$ and coincide with the additional macro variables, i.e. $\Gamma_{xg} = I$ and $R_x = 0$. The *large macro-yields* model is closely related to the model proposed in Mönch (2006) and, through the comparison with the *macro-yields* model, we want to emphasize the importance of extracting the factors from both the yields and the macro series.

The information set used for the analysis expands passing from the *only yields* to the *basic macro-yields* and to the *large macro-yields* models. While the *large*

macro-yields and the *macro-yields* model use the same information set. However the *macro-yields* model, presented in equations (2.2.5) - (2.2.6), is the only model that includes a large amount of macroeconomic information and has only three factors in the observation equation of the yields.

2.3 Data

The data-set used for the empirical analysis contains monthly observations of zero-coupon yields and a large set of macro variables from January 1970 to December 2000.

The zero-coupon yields have maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. This data, available on Diebold's home page, are constructed from end-of-month price quotes (bid/ask average) for U.S. Treasuries taken from the CRSP government bonds files. CRSP filters the data eliminating bonds with option features (callable and flower bonds), and bonds with special liquidity problems (notes and bonds with less than one year to maturity, and bills with less than one month to maturity), and then converts the filtered bond prices to unsmoothed Fama-Bliss (1987) forward rates. Then these unsmoothed forward rates are converted into unsmoothed Fama-Bliss zero-coupon yields, using fixed maturities. To pool the data in the fixed maturities listed above a month is defined as 30,4375 days, given that not every month has the same maturities, the data are obtained by linearly interpolating nearby maturities. For example in each month there are many bonds with either 30, 31, 32, 33 or 34 days to maturities, and by interpolating them it is possible to get the yields with maturity of one month (of 30.4375 days).

Summary statistics for the zero-coupon yields used in this paper are presented in Table 2.1. 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 and autocorrelations are very high and increasing with maturity.

[TABLE 2.1 AROUND HERE]

Figure 2.1 shows a plot of the zero-coupon yields for the period considered and highlights how the yields at different maturities tend to move together through time. Correlations between yields of different maturities are high, specially for yields with maturities that are close to each other.

[FIGURE 2.1 AROUND HERE]

The macro dataset is the same as used in Giannone, Reichlin and Sala (2004) and consists of 118 monthly US series. We exclude all interest and spread series, except for the federal funds rate, from the original panel dataset of 132 series. The federal funds rate closely follows the federal fund target rate, which is the key monetary policy instrument for the US Federal Reserve, and should therefore be important for capturing the movements of the short end of the term structure. The variables contained in the macro dataset include real variables (sectorial industrial production, employment and hours worked), nominal variables (consumer and producer price indices, wages, and money aggregates), asset prices (stock prices and exchange rate). Table 2.2 lists the series included in the macro dataset.

[TABLE 2.2 AROUND HERE]

We transform the monthly recorded macro series, whenever appropriate, to ensure stationarity by using levels, log levels, monthly differences, monthly log differences or annual log differences. The last column in Table 2.2 lists the applied transformation. In general, for real variables such as employment and industrial production we use the monthly growth rates. We use first differences for series already expressed in rates: unemployment rate and capacity of utilization. We do not transform in first differences the federal funds rate to be able to extract level, slope and curvature from the data.

2.4 Estimation procedure

The *macro-yields* model, presented in equation (2.2.5) - (2.2.6), allows to identify the Nelson and Siegel yield curve factors through the restrictions imposed in the relative factor loadings, Γ_{yf}^* . Thus the *macro-yields* model is a restricted dynamic factor model and it cannot be estimated by standard principal components, since this estimator does not allow to impose the necessary restrictions on the factor loadings. For this reason, using the results in Doz, Giannone and Reichlin (2006), we estimate the *macro-yields* model using quasi maximum likelihood.

The procedure proposed by Doz, Giannone and Reichlin (2006) combines EM algorithm and Kalman filter. This method makes feasible maximum likelihood estimation of factor models for large cross sections providing consistent estimates for any relative size of the time span and of the cross sectional dimension. Moreover, this procedure guarantees consistency even when hypothesis of orthogonality and absence of serial correlation of the idiosyncratic component are violated.

The estimation procedure alternates Kalman filter extraction of the factors to the maximization of the likelihood. In particular, for given parameters of the model, we use the Kalman filter to extract the factors. Then given the extracted factors, we maximize the Gaussian likelihood function implied by the Kalman filter using the EM algorithm. The estimation procedure is described in details in the Appendix.

As shown in section 2.2.1, the alternative models considered in this paper can be considered as restricted versions of the *macro-yields* model. For this reason, we use the same estimation procedure for all the models included in the analysis.

2.4.1 Model selection

The *macro-yields* model, presented in equation (2.2.5) - (2.2.6), is a general framework that allows for the presence of both the three Nelson and Siegel yield curve factors f_t and any number of unidentified and unobservable factors g_t . Therefore,

in order to estimate the model, we need a statistical criteria to select the number of factors to include.

The most used procedure to determine the number of factors in approximate factor models is the information criterion proposed by Bai and Ng (2002). The idea is to choose the number of factors that maximizes the general fit of the model using a penalty function to account for the loss in parsimony. The general form of the information criterion IC_3 introduced by Bai and Ng (2002) is

$$IC(r) = \log(V(r, \hat{F}^r)) + rg(N, T), \quad g(N, T) = \frac{\log C_{NT}^2}{C_{NT}^2} \quad (2.4.1)$$

where r denotes the number of factors, \hat{F}^r are the estimated factors and $V(r, \hat{F}^r)$ is the sum of squared residuals (divided by NT) when r factors are estimated. Moreover, the penalty function $g(N, T)$ is a function of both N and T and depends on C_{NT}^2 , the convergence rate of the principal component estimator, $C_{NT}^2 = \min\{T, N\}$.

The *macro-yields* model is estimated by quasi maximum likelihood and not by principal components, for this reason the IC information criterion, as presented in equation (2.4.1), cannot be used. However, in Corollary 2 of Bai and Ng (2002) it is shown that the IC information criterion can be applied to any consistent estimator of the factors provided that the penalty function is derived from the correct convergence rate. Thus, in order to apply this criterion to the *macro-yields* model, it is necessary to substitute the convergence rate of the quasi maximum likelihood estimator in equation (2.4.1).

Doz, Giannone and Reichlin (2006) in Proposition 1 show that the quasi maximum likelihood estimator of the common factors converges to the true value at a rate equal to $C_{NT}^{*2} = \min\{\sqrt{T}, \frac{N}{\log N}\}$. Therefor substituting C_{NT}^{*2} in equation (2.4.1), we obtain the modified Bai and Ng information criterion IC^* which can be

used when the common factors are estimated by maximum likelihood:

$$IC^*(r) = \log(V(r, \hat{F}^r)) + r \frac{\log \min\{\sqrt{T}, \frac{N}{\log N}\}}{\min\{\sqrt{T}, \frac{N}{\log N}\}}. \quad (2.4.2)$$

The estimated modified Bai and Ng information criterion $IC^*(r)$ for different specification of the *macro-yields* model is reported in Table 2.3. The model has been estimated on the full sample, from January 1970 to December 2000, varying the number of factors included. In particular, the model with three factors, i.e. $r = 3$, includes only the three Nelson and Siegel yield curve factors extracted from both yields and macro variables. While the models with $r = 4, 5$ and 6 include also one, two or three unidentified additional factors. The table also reports the values of the sum of the variance of the idiosyncratic components, denoted by $RR(r, \hat{F}^r)$, for each specification of the model.

[TABLE 2.3 AROUND HERE]

The modified Bai and Ng information criterion indicates that the best model is the model with the three Nelson and Siegel yield curve factors plus one unidentified factor, i.e. $r = 4$. This is also confirmed by the fact that the strongest reduction in the sum of the variances of the idiosyncratic components is obtained passing from the three to the four factors specification. Intuitively, this result can be explained by the large dimension of the data-set (17 yields plus 118 macroeconomic variables) that cannot be explained only through the three Nelson and Siegel yield curve factors. Figures 2.2-2.4 show the in sample fit of the *macro-yields* model with only the three Nelson and Siegel yield curve factors, i.e. $r = 3$, and adding the unidentified factor, i.e. $r = 4$.

[FIGURES 2.2-2.4 AROUND HERE]

The in sample fit of the yields does not improve passing from the three to the four factors specification, as it can be seen in Figure 2.2. However, the picture is completely different for the macroeconomic variables. Figures 2.3-2.4 highlight how the

fourth factor is important to capture the dynamics of most of the macroeconomic variables. Indeed, the three Nelson and Siegel yield curve factors do a poor job in fitting most of the macroeconomic variables, except price indexes. Figure 2.4 shows that the yield curve factors are by themselves able to fit really well the producer price index (PPI), the consumer price index (CPI) and the personal consumption expenditure implicit price deflator (PCE). This is due to the fact that the first Nelson and Siegel factor is highly correlated with inflation, as also confirmed by Diebold, Rudebush and Aruoba (2005).

Following this findings from now on we will refer to the *macro-yields* model as the model with the three Nelson and Siegel yield curve factors plus one unidentified factor.

2.5 Estimation Results

Estimation of the *macro-yields* model, requires a joint procedure to extract the latent factors, to identify the first three as the Nelson and Siegel yields curve factors and to estimate the loadings of all the 118 macroeconomic variables on the extracted factors. As explained in section 2.4, we address this issue estimating the model by quasi maximum likelihood.

Figure 2.5 shows the estimated factors of the *macro-yields* model and, for comparison purposes, also the relative Nelson and Siegel factors.

[FIGURE 2.5 AROUND HERE]

The first three factors of the *macro-yields* model, which we identified as the Nelson and Siegel factors, indeed are really close to the original Nelson and Siegel factors. The difference between the original Nelson and Siegel factors and the *macro-yields* Nelson and Siegel factors comes from the fact that the *macro-yields* factors include not only the information contained in the yields but also the one contained in the macroeconomic series. The factor that is more affected by the macroeconomic information is the curvature, since the *macro-yields* curvature factor is way more

persistent than the original Nelson and Siegel curvature. This is also confirmed in Table 2.4 where we report summary statistics of the estimated *macro-yields* factors and of the Nelson and Siegel factors.

[TABLE 2.4 AROUND HERE]

Table 2.4 highlights also a certain difference in the persistence of the *macro-yields* slope factors and the Nelson and Siegel one. In general, the *macro-yields* factors tend to be more persistent than the Nelson and Siegel ones. The last panel of Figure 2.5 reports the fourth factor, the unidentified one. The plot highlights how this factor accounts for the macroeconomic situation. Indeed, during all the recessions in the sample the unidentified factor decreases drastically. Summary statistics for the unidentified factor in Table 2.4 show that it has zero mean and almost unit variance, but a high degree of persistency.

Table 2.5 displays the goodness of fit of the *macro-yields* model compared with the alternative models presented in section 2.2.1. In particular, the table reports the mean square error of the *only yields* model (OY), the *basic macro-yields* (BMY), the *large macro-yields* (LMY) and the *macro-yields* (MY) models for selected maturities.

[TABLE 2.5 AROUND HERE]

The four models display almost the same performances in fitting the term structure of interest rates except for the shortest yield, where adding a large set of macro-variables clearly worsens the fit. Moreover, the *basic macro-yields* and the *large macro-yields models*, even if they are the only models with six factors, they do not display a significant improvement with respect to the other two models, namely the *only yields* model and the *macro-yields* model. Macro variables do not improve the fit of the yield curve, but this does not imply that they do not have leading information for the yields. The aim of the following chapter is to show that a large set of macro variables helps in forecasting the yields provided they are used parsimoniously.

2.6 Out-of-sample forecast

The out-of-sample forecasts of the *only yields*, *basic macro-yields*, *large macro-yields* and *macro-yields* models are obtained iteratively. As mentioned in Section 2.2.1, the *only yields*, *large macro-yields* the *basic macro-yields* models are nested in the *macro-yields* model.⁵ Therefore rewriting the *macro-yields* model (2.2.5)-(2.2.6) in compact notation we obtain a general representation of all the models presented

$$z_t = \Gamma F_t + v_t, \quad v_t \sim iidN(0, R) \quad (2.6.1)$$

$$F_t = AF_{t-1} + u_t, \quad u_t \sim iidN(0, Q) \quad (2.6.2)$$

where $F_t = (f_t, g_t)$, $v_t = (v_{y,t}, v_{x,t})$ and $u_t = (u_{f,t}, u_{g,t})$. We generate iterative forecasts for all the models at first projecting forward the factors

$$\hat{F}_{t+h|t} = \hat{A}^h \hat{F}_t$$

and then computing the out-of-sample forecast given the projected factors

$$\hat{z}_{t+h|t} = \hat{\Gamma} \hat{F}_{t+h|t}$$

To evaluate the prediction accuracy at a given forecast horizon h , we use the mean square forecast error (MSFE), the average square error between time t_0 and t_1 for the h -months ahead forecast of the yield with maturity τ , using a particular model m

$$MSFE_{t_0}^{t_1}(\tau, h, m) = \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} (\hat{y}(\tau)_{t+h|t}^m - y(\tau)_{t+h})^2 \quad (2.6.3)$$

where $y(\tau)_{t+h}$ is the realized yield with maturity τ at time $t+h$ and $\hat{y}(\tau)_{t+h|t}^m$ is the

⁵The only yield model is obtained setting $\Gamma_{xf} = \Gamma_{xg} = \Gamma_{yg} = 0$ in equations (2.2.5)-(2.2.6). The *basic macro-yields* model is obtained setting $x_t = (CU_t FFR_t INFL_t)$, $\Gamma_{xf} = 0$, $\Gamma_{xg} = I$, $\Gamma_{yg} \neq 0$ and $R_x = 0$ in the same equations. And the *large macro-yields* model can be obtained setting $g_t = PC_t$, $\Gamma_{xf} = 0$, $\Gamma_{xg} = I$, $\Gamma_{yg} \neq 0$ and $R_x = 0$ in equations (2.2.5)-(2.2.6).

h-steps ahead forecast of the yield with maturity τ made at time t with a particular model m .

Forecast results for yields are usually expressed as ratios of the MSFEs of the considered model and the MSFE of a random walk, which is a naïve model very difficult to outperform given the high persistency of the yields. The random walk h-steps ahead prediction at time t of the yield with maturity τ is

$$\hat{y}_{t+h|t}(\tau) = y_t(\tau)$$

where the optimal predictor does not change regardless of the maturity of the yield and the forecast horizon.

Ang and Piazzesi (2003), Mönch (2006), Favero et al. (2007) and De Pooter et al. (2007) found that macroeconomic variables help in forecasting the yield curve. For this reason, it can be expected that the *macro-yields* model will outperform the *only yields* one. However, from the comparison of the *macro-yields* model with the *basic macro-yields* and the *large macro-yields*, it will be possible to show that, not only it is important to use large information to forecast the yields, but it is also crucial to extract the yield curve factors from both the yields and the macro variables in order to be able to capture the co-movement between the yields and the whole economy.

2.6.1 Forecast performances

We forecast the yields estimating each model recursively using data from January 1970 until the time that the forecast is made, beginning in January 1985 to December 2000. We use the random walk as benchmark, therefore we construct ratios of each model's MSFEs over the random walk MSFEs. Table 2.6 reports these ratios for the *only yields* (OY), *basic macro-yields* (BMY), *large macro-yields* (LMY) and *macro-yields* (MY) models, for selected maturities.

[TABLE 2.6 AROUND HERE]

The out-of-sample performances of the *only yields*, the *basic macro-yields* and the *large macro-yields* models are similar. This, rather than being interpreted as evidence that the macroeconomic variables are not useful in forecasting the yields, should be considered as a consequence of the lack of parsimony of the *basic macro yields* model and the *large macro-yields* models, which both include six factors.

The *only yields* model can be seen as a restricted *basic macro-yields*, or *large macro-yields*, model where the loadings of the yields on the observable macroeconomic factors are zero. Therefore, if the macro variables would not be useful in forecasting the yields, it should be expected that the *only yields* model would outperform the *basic macro yields* and *large macro-yields* models. This is not the case, meaning that the macro variables are helpful in forecasting the yields but in the *basic macro-yields* and the *large macro-yields* model they are used in a non parsimonious way.

The *macro-yields* model is suited to solve this problem and to exploit a large amount of macro information in a parsimonious way. Indeed the relevance of the macroeconomic variable in forecasting the yields, specially on medium and long horizons, becomes evident looking at the out-of-sample forecasting performance of the *macro-yields* model. For 6 and 12 months ahead, the *macro-yields* model not only outperforms the *only yields*, the *basic macro yields* and the *large macro-yields* models but also the random walk. However, at the shortest horizon, i.e. one month ahead, the random walk in most cases provides the best forecast.

To investigate the out-of-sample performances of all the models over time, Figures 2.6-2.7 plot the smoothed square forecast errors and the smoothed forecast errors of all the models considered for some selected maturities. The smoothed square forecast errors are computed as a 30 months moving average of the squared forecast errors, while the smoothed forecast errors are computed as a 30 months moving average of the forecast errors.

[FIGURES 2.6-2.7 AROUND HERE]

The square forecast errors of all the models have a similar pattern both for 6-months and 12-months ahead forecasts. In general, they start to increase just before the recession of July 1981 - November 1982, they peak after the recession and then they decline at the end of the period. However, it is possible to distinguish different behaviors across the models. The *macro-yields* model at both 6-months and 12-months ahead outperforms all the other models, and often also the random walk, for almost all the sample except at the two last years. The opposite happens for the basic *macro-yields* model, which at both 6-months and 12-months ahead is the worst model for almost all the sample, but just at the last year it slightly outperforms the other models. This can indicate that large macroeconomic information is particularly useful just before and after the recessions, provided it is used parsimoniously. While in the other periods, few macro indicators are enough to convey information about the state of the economy. Figure 2.6 highlights also a bad performance of the benchmark, the random walk, at the beginning and the end of the sample.

The forecast errors plotted in Figure 2.7 are also particularly small at the beginning and at the end of the sample, while they increase just before the recession with all the models, sometimes except the *macro-yields* one, overestimating the yields. However the conclusions, slightly change looking at the forecast errors. In this case, the random walk is always one of the best models, providing small forecast errors. The *macro-yields* model outperforms all the competitive models in almost the whole sample, and often also the random walk. The forecast errors also indicate that the *only yields* model, which is the only model that does not use any macro information, is the only model to systematically overestimate the yields.

In conclusion, the *macro-yields* model at 6 and 12-months ahead outperforms on average all the competing models and also the random walk. This result is driven from the fact that the model is particularly able to outperform the others during and just after the recessions. The *macro-yields* model is the only model able to provide forecast errors that are almost constant in time, while all the other

models exhibit high variability of the forecast errors with huge peaks just after the recession.

2.7 Conclusions

We propose a new framework to fit and forecast the yield curve using parsimoniously a large amount of macroeconomic information. Our approach is based on a factor model, where the factors are extracted directly from a panel of 17 yields and 118 macro variables. The loadings of the yields on the first three factors are characterized by restrictions *à la* Nelson and Siegel that allow us to identify these first three factors as the level, the slope and the curvature of the yield curve. This is an innovative way to use macro variables to forecast the yields, given that the most recent literature was using a small set of macro variables as extra regressors. We show that our approach outperforms the existing methods for all the maturities at mid and long horizons (i.e. 6 and 12-month ahead).

2.8 Appendix

The more general version of our model can be written as

$$\begin{aligned} z_t &= \Gamma F_t + v_t \quad v_t \sim N(0, R) \\ F_t &= AF_{t-1} + u_t \quad u_t \sim N(0, Q) \end{aligned}$$

where R is diagonal and $F_t = [f_t \quad g_t]'$ and we can partition the matrix Γ as

$$\Gamma = \begin{bmatrix} \Gamma_{yf}^* & 0 \\ \Gamma_{xf} & \Gamma_{xg} \end{bmatrix}$$

Where the identification of the yield curve factors is achieved through the restrictions on their loadings coming from the Nelson and Siegel representation, with

$$\Gamma_{yf}^* = \begin{bmatrix} 1 & \frac{1-e^{\lambda\tau_1}}{\lambda\tau_1} & \frac{1-e^{\lambda\tau_1}}{\lambda\tau_1} - e^{\lambda\tau_1} \\ 1 & \frac{1-e^{\lambda\tau_2}}{\lambda\tau_2} & \frac{1-e^{\lambda\tau_2}}{\lambda\tau_2} - e^{\lambda\tau_2} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{\lambda\tau_N}}{\lambda\tau_N} & \frac{1-e^{\lambda\tau_N}}{\lambda\tau_N} - e^{\lambda\tau_N} \end{bmatrix}$$

Following Diebold and Li (2006), we fix $\lambda = 0.0609$, the value that maximizes the loading on the curvature factor for the yields with maturity to 30 months.⁶

The parameters are estimated by maximum likelihood combining EM algorithm and Kalman filter. As shown in Doz, Giannone and Reichlin (2006), maximum likelihood estimation of a dynamic approximate factor model, when the panel of time series is large, is feasible, in the sense that it guarantees consistency, and represents a valid alternative to principal components. Moreover this methodology is particularly suitable in our case since it allows to impose restriction on the model.

Assuming $F_1 \sim N(\pi_1, V_1)$ and labeling the time series $(z_1, z_2, \dots, z_T) = \{\mathbf{z}\}$ and $(F_1, F_2, \dots, F_T) = \{F\}$, and the parameters $\{\Gamma, R, A, Q, \pi_1, V_1\} = \theta$, the log-

⁶Using the ECM algorithm is also possible to estimate λ , but despite the increase in the computation burden, the results remain substantially unchanged.

likelihood is:

$$\begin{aligned}
L(\{z\}\{F\}; \theta) &= - \sum_{t=1}^T \left(\frac{1}{2} \left[z_t - \begin{bmatrix} \Gamma_{yf} & 0 \\ \Gamma_{xf} & \Gamma_{xg} \end{bmatrix} F_t \right]' R^{-1} \left[z_t - \begin{bmatrix} \Gamma_{yf} & 0 \\ \Gamma_{xf} & \Gamma_{xg} \end{bmatrix} F_t \right] \right) + \\
&\quad - \frac{T}{2} \log |R| - \sum_{t=2}^T \left(\frac{1}{2} [F_t - AF_{t-1}]' Q^{-1} [F_t - AF_{t-1}] \right) + \\
&\quad - \frac{T-1}{2} \log |Q| \frac{1}{2} [F_1 - \pi_1]' V^{-1} [F_1 - \pi_1] + \\
&\quad - \frac{1}{2} \log |V_1| - \frac{T(p+k)}{2} \log 2\pi
\end{aligned}$$

The EM algorithm alternates Kalman filter extraction of the factors to the maximization of the likelihood. In particular, for given parameters of the model we use the Kalman filter to extract the factors (E step). Then given the extracted factors, we maximize the Gaussian likelihood function implied by the Kalman filter (M step).

Therefore in the E step, we compute the expected log-likelihood

$$\mathcal{Q} = E[L(\{z\}\{F\}; \theta) | \{z\}]$$

which depends on three expectations

$$\hat{F}_t \equiv E[F_t | \{z\}]$$

$$P_t \equiv E[F_t F_t' | \{z\}]$$

$$P_{t,t-1} \equiv E[F_t F_{t-1}' | \{z\}]$$

And in the M step, we re-estimate the parameters $\theta = \{\Gamma_{yg}, \Gamma_{xf}, \Gamma_{xg}, R, A, Q, \pi_1, V_1\}$ taking the corresponding partial derivative of the expected log likelihood, setting to zero, and solving.

- Output matrix: since we have the restriction on the upper blocks, we derive the first order conditions by blocks. We denote by f_t the yield curve factors

and by g_t the macro factors, such that $F_t = \begin{bmatrix} f_t & g_t \end{bmatrix}$, and by y_t the yields and x_t the macro variables, such that $z = \begin{bmatrix} y_t & x_t \end{bmatrix}'$.

- submatrix $\begin{bmatrix} \Gamma_{xf} & \Gamma_{xg} \end{bmatrix}$

$$\frac{\partial \mathcal{Q}}{\partial \begin{bmatrix} \Gamma_{xf} & \Gamma_{xg} \end{bmatrix}} = E \left[- \sum_{t=1}^T R_{xx}^{-1} (x_t - \begin{bmatrix} \Gamma_{xf} & \Gamma_{xg} \end{bmatrix} F_t) F_t' | \{z\} \right] = 0$$

$$\begin{bmatrix} \Gamma_{xf} & \Gamma_{xg} \end{bmatrix} = \sum_{t=1}^T E [x_t F_t' | \{z\}] \left(\sum_{t=1}^T E [F_t F_t' | \{z\}] \right)^{-1}$$

$$\Rightarrow \begin{bmatrix} \Gamma_{xf}^{new} & \Gamma_{xg}^{new} \end{bmatrix} = \sum_{t=1}^T (x_t \hat{F}_t') \left(\sum_{t=1}^T P_t \right)^{-1}$$

- Output noise covariance:

$$\frac{\partial \mathcal{Q}}{\partial R^{-1}} = E \left[- \sum_{t=1}^T \frac{1}{2} (z_t - \Gamma F_t) (z_t - \Gamma F_t)' + \frac{T}{2} R | \{z\} \right] = 0$$

$$R = \frac{1}{T} \sum_{t=1}^T E [z_t z_t' - z_t F_t' \Gamma' - \Gamma F_t z_t' + \Gamma F_t F_t' \Gamma' | \{z\}]$$

$$\Rightarrow R^{new} = \frac{1}{T} \sum_{t=1}^T (z_t z_t' - z_t \hat{F}_t' \Gamma'^{new} - \Gamma^{new} \hat{F}_t z_t' + \Gamma^{new} P_t \Gamma'^{new})$$

- State dynamics matrix:

$$\frac{\partial \mathcal{Q}}{\partial A} = E \left[- \frac{1}{2} \sum_{t=2}^T Q^{-1} (F_t - A F_{t-1}) F_{t-1}' | \{z\} \right] = 0$$

$$A = \sum_{t=2}^T E [F_t F_{t-1}' | \{z\}] \left(\sum_{t=2}^T E [F_{t-1} F_{t-1}' | \{z\}] \right)^{-1}$$

$$\Rightarrow A^{new} = \left(\sum_{t=2}^T P_{t,t-1} \right) \left(\sum_{t=2}^T P_{t-1} \right)^{-1}$$

- State noise covariance:

$$\frac{\partial \mathcal{Q}}{\partial Q^{-1}} = E \left[\frac{T-1}{2} Q - \frac{1}{2} \sum_{t=2}^T (F_t - AF_{t-1})(F_t - AF_{t-1})' | \{z\} \right] = 0$$

$$Q = \frac{1}{T-1} \sum_{t=2}^T E [F_t F_t' - F_t F_{t-1}' A' - AF_{t-1} F_t' + AF_{t-1} F_{t-1}' A' | \{z\}]$$

$$Q = \frac{1}{T-1} \sum_{t=2}^T E [F_t F_t' - AF_{t-1} F_t' | \{z\}]$$

$$\Rightarrow Q^{new} = \frac{1}{T-1} \left(\sum_{t=2}^T P_t - A^{new} \sum_{t=2}^T P_{t-1,t} \right)$$

- Initial state mean:

$$\frac{\partial \mathcal{Q}}{\partial \pi_1^{new}} = E [(F_1 - \pi_1) V_1^{-1} | \{z\}] = 0$$

$$\frac{\partial \mathcal{Q}}{\partial \pi_1^{new}} = (\hat{F}_1 - \pi_1) V_1^{-1} = 0$$

$$\Rightarrow \pi_1^{new} = \hat{F}_1$$

- Initial state covariance:

$$\frac{\partial \mathcal{Q}}{\partial V_1^{-1}} = E \left[\frac{1}{2} V_1 - \frac{1}{2} (F_1 - \pi_1)(F_1 - \pi_1)' | \{z\} \right] = 0$$

$$V_1 = E [F_1 F_1' - \pi_1' F_1 - F_1' \pi_1 + \pi_1 \pi_1' | \{z\}]$$

$$\Rightarrow V_1^{new} = P_1 - \hat{F}_1 \hat{F}_1'$$

Now we go back to the E step and update the expectations. Using $F_{t|\tau}$ to denote $E(F_t | \{z\}_{t=1}^\tau)$ and $V_{t|\tau}$ to denote the $Var(F_t | \{z\}_{t=1}^\tau)$, we obtain the following

Kalman filter forward recursions

$$\begin{aligned}
F_{t+1|t} &= AF_{t|t} \\
V_{t+1|t} &= AV_{t|t}A' + Q \\
K_t &= V_{t|t-1}\Gamma(\Gamma'V_{t|t-1}\Gamma + R)^{-1} \\
F_{t|t} &= F_{t|t-1} + K_t(z_t - \Gamma F_{t|t-1}) \\
V_{t|t} &= V_{t|t-1} - K_t\Gamma'V_{t|t-1}
\end{aligned}$$

where $F_{1|0} = \pi_1$ and $V_{1|0} = V_1$. To compute $\hat{F}_t \equiv F_{t|T}$ and $P_t \equiv V_{t|T} + F_{t|T}F_{t|T}'$ one performs a set of backward recursion using

$$\begin{aligned}
J_t &= V_{t|t}A'V_{t+1|t}^{-1} \\
F_{t|T} &= F_{t+1|T} + J_t(F_{t+1|T} - AF_{t|T}) \\
V_{t|T} &= V_{t+1|T} + J_t(V_{t+1|T} - V_{t+1|t})J_t'
\end{aligned}$$

Moreover $P_{t,t-1} \equiv V_{t,t-1|T} + F_{t|T}F_{t-1|T}'$ can be obtained through the backward recursion

$$V_{t,t-1|T} = V_{t|t}J_{t-1}' + J_t(V_{t+1,t|T} - AV_{t|t})J_{t-1}'$$

which is initialized $V_{T,T-1|T} = (I - K_T\Gamma)AV_{T-1|T-1}$.

The estimation procedure is initialized using the factors extracted by the two steps OLS procedure introduced by Diebold and Li (2006). These factors are centered around their means and are standardized with the average of the yield standard deviations. The means of these factors multiplied by Γ_{yf} are used to center the yields, that is equivalent to center the yields with the mean of the means of the yields, and are standardized by the mean of the standard deviations. The macroeconomic data are centered around their own means and standardized by their standard deviations.

Table 2.1: Summary statistics of the US zero-coupon data

τ	mean	std dev	min	max	$\rho(1)$	$\rho(2)$	$\rho(3)$	$\rho(12)$
3	6.75	2.66	2.73	16.02	0.97*	0.94*	0.91*	0.71*
6	6.98	2.66	2.89	16.48	0.97*	0.94*	0.91*	0.73*
9	7.10	2.64	2.98	16.39	0.97*	0.94*	0.91*	0.73*
12	7.20	2.57	3.11	15.82	0.97*	0.94*	0.91*	0.74*
15	7.31	2.52	3.29	16.04	0.97*	0.94*	0.91*	0.75*
18	7.38	2.50	3.48	16.23	0.98*	0.94*	0.92*	0.75*
21	7.44	2.49	3.64	16.18	0.98*	0.95*	0.92*	0.76*
24	7.46	2.44	3.78	15.65	0.98*	0.94*	0.92*	0.75*
30	7.55	2.36	4.04	15.40	0.98*	0.95*	0.92*	0.76*
36	7.63	2.34	4.20	15.77	0.98*	0.95*	0.93*	0.77*
48	7.77	2.28	4.31	15.82	0.98*	0.95*	0.93*	0.78*
60	7.84	2.25	4.35	15.01	0.98*	0.96*	0.94*	0.79*
72	7.96	2.22	4.38	14.98	0.98*	0.96*	0.94*	0.80*
84	7.99	2.18	4.35	14.98	0.98*	0.96*	0.94*	0.78*
96	8.05	2.17	4.43	14.94	0.98*	0.96*	0.95*	0.81*
108	8.08	2.18	4.43	15.02	0.98*	0.96*	0.95*	0.81*
120	8.05	2.14	4.44	14.93	0.98*	0.96*	0.94*	0.78*

Descriptive statistics of monthly yields at different maturities τ for the sample from January 1970 to December 2000. $\rho(p)$ refers to the sample autocorrelation of the series at lag p and * denotes significance at 95 percent confidence level. Confidence intervals are computed according to Box and Jenkins (1976).

Table 2.2: Macroeconomic series

Code	Description	Transf.	
1	a0m052	Personal income (AR, bil. chain 2000 \$)	4
2	A0M051	Personal income less transfer payments (AR, bil. chain 2000 \$)	4
3	A0M224_R	Real Consumption (AC) A0m224/gmdc	4
4	A0M057	Manufacturing and trade sales (mil. Chain 1996 \$)	4
5	A0M059	Sales of retail stores (mil. Chain 2000 \$)	4
6	IPS10	INDUSTRIAL PRODUCTION INDEX - TOTAL INDEX	4
7	IPS11	INDUSTRIAL PRODUCTION INDEX - PRODUCTS, TOTAL	4
8	IPS299	INDUSTRIAL PRODUCTION INDEX - FINAL PRODUCTS	4
9	IPS12	INDUSTRIAL PRODUCTION INDEX - CONSUMER GOODS	4
10	IPS13	INDUSTRIAL PRODUCTION INDEX - DURABLE CONSUMER GOODS	4
11	IPS18	INDUSTRIAL PRODUCTION INDEX - NONDURABLE CONSUMER GOODS	4
12	IPS25	INDUSTRIAL PRODUCTION INDEX - BUSINESS EQUIPMENT	4
13	IPS32	INDUSTRIAL PRODUCTION INDEX - MATERIALS	4
14	IPS34	INDUSTRIAL PRODUCTION INDEX - DURABLE GOODS MATERIALS	4
15	IPS38	INDUSTRIAL PRODUCTION INDEX - NONDURABLE GOODS MATERIALS	4
16	IPS43	INDUSTRIAL PRODUCTION INDEX - MANUFACTURING (SIC)	4
17	IPS307	INDUSTRIAL PRODUCTION INDEX - RESIDENTIAL UTILITIES	4
18	IPS306	INDUSTRIAL PRODUCTION INDEX - FUELS	4
19	PMP	NAPM PRODUCTION INDEX (PERCENT)	1
20	A0m082	Capacity Utilization (Mfg)	2
21	LHEL	INDEX OF HELP-WANTED ADVERTISING IN NEWSPAPERS (1967=100;SA)	2
22	LHELX	EMPLOYMENT: RATIO; HELP-WANTED ADS:NO. UNEMPLOYED CLF	2
23	LHEM	CIVILIAN LABOR FORCE: EMPLOYED, TOTAL (THOUS.,SA)	4
24	LHNAG	CIVILIAN LABOR FORCE: EMPLOYED, NONAGRIC.INDUSTRIES (THOUS.,SA)	4
25	LHUR	UNEMPLOYMENT RATE: ALL WORKERS, 16 YEARS & OVER (%;SA)	2
26	LHU680	UNEMPLOY.BY DURATION: AVERAGE(MEAN)DURATION IN WEEKS (SA)	2
27	LHU5	UNEMPLOY.BY DURATION: PERSONS UNEMPL.LESS THAN 5 WKS (THOUS.,SA)	4
28	LHU14	UNEMPLOY.BY DURATION: PERSONS UNEMPL.5 TO 14 WKS (THOUS.,SA)	4
29	LHU15	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 WKS + (THOUS.,SA)	4
30	LHU26	UNEMPLOY.BY DURATION: PERSONS UNEMPL.15 TO 26 WKS (THOUS.,SA)	4
31	LHU27	UNEMPLOY.BY DURATION: PERSONS UNEMPL.27 WKS + (THOUS,SA)	4
32	A0M005	Average weekly initial claims, unemploy. insurance (thous.)	4
33	CES002	EMPLOYEES ON NONFARM PAYROLLS - TOTAL PRIVATE	4
34	CES003	EMPLOYEES ON NONFARM PAYROLLS - GOODS-PRODUCING	4
35	CES006	EMPLOYEES ON NONFARM PAYROLLS - MINING	4
36	CES011	EMPLOYEES ON NONFARM PAYROLLS - CONSTRUCTION	4
37	CES015	EMPLOYEES ON NONFARM PAYROLLS - MANUFACTURING	4
38	CES017	EMPLOYEES ON NONFARM PAYROLLS - DURABLE GOODS	4
39	CES033	EMPLOYEES ON NONFARM PAYROLLS - NONDURABLE GOODS	4
40	CES046	EMPLOYEES ON NONFARM PAYROLLS - SERVICE-PROVIDING	4
41	CES048	EMPLOYEES ON NONFARM PAYROLLS - TRADE, TRANSPORTATION, AND UTILITIES	4
42	CES049	EMPLOYEES ON NONFARM PAYROLLS - WHOLESALE TRADE	4
43	CES053	EMPLOYEES ON NONFARM PAYROLLS - RETAIL TRADE	4
44	CES088	EMPLOYEES ON NONFARM PAYROLLS - FINANCIAL ACTIVITIES	4
45	CES140	EMPLOYEES ON NONFARM PAYROLLS - GOVERNMENT	4
46	A0M048	Employee hours in nonag. establishments (AR, bil. hours)	4
47	CES151	AV. WEEKLY HRS OF PROD OR NONSUP WORKERS ON PRIV NONFAR - GOODS PROD	1
48	CES155	AV. WEEKLY HRS OF PROD OR NONSUP WORKERS ON PRIV NONFAR - MFG OVERTIME	2
49	aom001	Average weekly hours, mfg. (hours)	1
50	PMEMP	NAPM EMPLOYMENT INDEX (PERCENT)	1
51	HSFR	HOUSING STARTS:NONFARM(1947-58);TOTAL FARM&NONFARM(1959-)(THOUS.,SA	3
52	HSNE	HOUSING STARTS:NORTHEAST (THOUS.U.)S.A.	3
53	HSMW	HOUSING STARTS:MIDWEST(THOUS.U.)S.A.	3
54	HSSOU	HOUSING STARTS:SOUTH (THOUS.U.)S.A.	3
55	HSWST	HOUSING STARTS:WEST (THOUS.U.)S.A.	3
56	HSBR	HOUSING AUTHORIZED: TOTAL NEW PRIV HOUSING UNITS (THOUS.,SAAR)	3
57	HSBNE	HOUSES AUTHORIZED BY BUILD. PERMITS:NORTHEAST(THOU.U.)S.A	3
58	HSBMW	HOUSES AUTHORIZED BY BUILD. PERMITS:MIDWEST(THOU.U.)S.A.	3
59	HSBSOU	HOUSES AUTHORIZED BY BUILD. PERMITS:SOUTH(THOU.U.)S.A.	3
60	HSBWST	HOUSES AUTHORIZED BY BUILD. PERMITS:WEST(THOU.U.)S.A.	3
61	PMI	PURCHASING MANAGERS' INDEX (SA)	1
62	PMNO	NAPM NEW ORDERS INDEX (PERCENT)	1
63	PMDEL	NAPM VENDOR DELIVERIES INDEX (PERCENT)	1
64	PMNV	NAPM INVENTORIES INDEX (PERCENT)	1
65	A0M008	Mfrs' new orders, consumer goods and materials (bil. chain 1982 \$)	4
66	A0M007	Mfrs' new orders, durable goods industries (bil. chain 2000 \$)	4
67	A0M027	Mfrs' new orders, nondefense capital goods (mil. chain 1982 \$)	4
68	A1M092	Mfrs' unfilled orders, durable goods indus. (bil. chain 2000 \$)	4
69	A0M070	Manufacturing and trade inventories (bil. chain 2000 \$)	4
70	A0M077	Ratio, mfg. and trade inventories to sales (based on chain 2000 \$)	2

Code	Description	Transf.	
71	FM1	MONEY STOCK: M1(CURR,TRAV.CKS,DEM DEP,OTHER CK'ABLE DEP)(BIL\$,SA)	5
72	FM2	MONEY STOCK:M2(M1+O'NITE RPS,EUROS,G/P&B/D MMMFS&SAV&SM TIME DEP(BIL\$,	5
73	FM3	MONEY STOCK: M3(M2+LG TIME DEP,TERM RP'S&INST ONLY MMMFS)(BIL\$,SA)	5
74	FM2DQ	MONEY SUPPLY - M2 IN 1996 DOLLARS (BCI)	4
75	FMFBA	MONETARY BASE, ADJ FOR RESERVE REQUIREMENT CHANGES(MIL\$,SA)	5
76	FMRRA	DEPOSITORY INST RESERVES:TOTAL,ADJ FOR RESERVE REQ CHGS(MIL\$,SA)	5
77	FMRNBA	DEPOSITORY INST RESERVES:NONBORROWED,ADJ RES REQ CHGS(MIL\$,SA)	5
78	FCLNQ	COMMERCIAL & INDUSTRIAL LOANS OUSTANDING IN 1996 DOLLARS (BCI)	5
79	FCLBMC	WKLY RP LG COM'L BANKS:NET CHANGE COM'L & INDUS LOANS(BIL\$,SAAR)	1
80	CCINRV	CONSUMER CREDIT OUTSTANDING - NONREVOLVING(G19)	5
81	AOM095	Ratio, consumer installment credit to personal income (pct.)	2
82	FSPCOM	S&P'S COMMON STOCK PRICE INDEX: COMPOSITE (1941-43=10)	4
83	FSPIN	S&P'S COMMON STOCK PRICE INDEX: INDUSTRIALS (1941-43=10)	4
84	FSDXP	S&P'S COMPOSITE COMMON STOCK: DIVIDEND YIELD (% PER ANNUM)	2
85	FSPXE	S&P'S COMPOSITE COMMON STOCK: PRICE-EARNINGS RATIO (%NSA)	4
86	FYFF	INTEREST RATE: FEDERAL FUNDS (EFFECTIVE) (% PER ANNUM,NSA)	1
87	CP90	Commercial Paper Rate (AC)	2
88	FYAAAC	BOND YIELD: MOODY'S AAA CORPORATE (% PER ANNUM)	2
89	FYBAAC	BOND YIELD: MOODY'S BAA CORPORATE (% PER ANNUM)	2
90	EXRUS	UNITED STATES;EFFECTIVE EXCHANGE RATE(MERM)(INDEX NO.)	4
91	EXRSW	FOREIGN EXCHANGE RATE: SWITZERLAND (SWISS FRANC PER U.S.\$)	4
92	EXRJAN	FOREIGN EXCHANGE RATE: JAPAN (YEN PER U.S.\$)	4
93	EXRUK	FOREIGN EXCHANGE RATE: UNITED KINGDOM (CENTS PER POUND)	4
94	EXRCAN	FOREIGN EXCHANGE RATE: CANADA (CANADIAN \$ PER U.S.\$)	4
95	PWFSA	PRODUCER PRICE INDEX: FINISHED GOODS (82=100,SA)	5
96	PWFCSA	PRODUCER PRICE INDEX:FINISHED CONSUMER GOODS (82=100,SA)	5
97	PWIMSA	PRODUCER PRICE INDEX:INTERMED MAT.SUPPLIES & COMPONENTS(82=100,SA)	5
98	PWCMSA	PRODUCER PRICE INDEX:CRUDE MATERIALS (82=100,SA)	5
99	PSM99Q	INDEX OF SENSITIVE MATERIALS PRICES (1990=100)(BCI-99A)	5
100	PMCP	NAPM COMMODITY PRICES INDEX (PERCENT)	1
101	PUNEW	CPI-U: ALL ITEMS (82-84=100,SA)	5
102	PU83	CPI-U: APPAREL & UPKEEP (82-84=100,SA)	5
103	PU84	CPI-U: TRANSPORTATION (82-84=100,SA)	5
104	PU85	CPI-U: MEDICAL CARE (82-84=100,SA)	5
105	PUC	CPI-U: COMMODITIES (82-84=100,SA)	5
106	PUCD	CPI-U: DURABLES (82-84=100,SA)	5
107	PUS	CPI-U: SERVICES (82-84=100,SA)	5
108	PUXF	CPI-U: ALL ITEMS LESS FOOD (82-84=100,SA)	5
109	PUXHS	CPI-U: ALL ITEMS LESS SHELTER (82-84=100,SA)	5
110	PUXM	CPI-U: ALL ITEMS LESS MIDICAL CARE (82-84=100,SA)	5
111	GMDC	PCE,IMPL PR DEFL:PCE (1987=100)	5
112	GMDCD	PCE,IMPL PR DEFL:PCE; DURABLES (1987=100)	5
113	GMDCN	PCE,IMPL PR DEFL:PCE; NONDURABLES (1996=100)	5
114	GMDCS	PCE,IMPL PR DEFL:PCE; SERVICES (1987=100)	5
115	CES275	AV. HOURLY EARNINGS OF PROD OR NONSUP WORKERS ON PRIV NO - GOODS PROD	5
116	CES277	AV. HOURLY EARNINGS OF PROD OR NONSUP WORKERS ON PRIV NO - CONSTRUCTION	5
117	CES278	AV. HOURLY EARNINGS OF PROD OR NONSUP WORKERS ON PRIV NO - MANUFACTURING	5
118	HHSNTN	U. OF MICH. INDEX OF CONSUMER EXPECTATIONS(BCD-83)	2

The table lists the macro series included in the macroeconomic dataset. The first column counts the position in the dataset, the second reports the code of the series, the third shows the name of the variable. The last column reports the transformations applied to original series. These tranformations are coded as: 1:=no transformation (levels are used), 2:= monthly differences, 3:= logarithm of the level, 4:= monthly first differences of the log levels (in percentage), 5:= annual first differences of the log levels (in percentage). The sample period is January 1970 - December 2000 (372 observations).

Table 2.3: Model selection

r	3	4	5	6
$RR(r, \hat{F}^r)$	90.12	69.40	64.69	57.14
$IC^*(r)$	0.06	-0.05	0.03	0.06

Notes: $RR(r, \hat{F}^r)$ is the sum of the variance of the idiosyncratic component and $IC^*(r)$ is the modified Bai and Ng information criteria presented in equation (2.4.1). Both measures are computed for different specifications of the *macro yields* model (i.e. with the 3 Nelson and Siegel factors plus 1, 2 or 3 unidentified factors).

Table 2.4: Summary statistics of the estimated factors

	L_{MY}	L_{NS}	S_{MY}	S_{NS}	C_{MY}	C_{NS}	U_{MY}
Mean	8.26	8.26	-1.58	-1.58	0.19	0.19	0.00
Var	4.29	4.32	3.75	3.67	1.84	3.27	1.06
$\rho(1)$	0.98	0.98	0.96	0.94	0.95	0.79	0.92
$\rho(2)$	0.96	0.97	0.90	0.87	0.90	0.64	0.83
Min	4.20	4.43	-5.85	-5.62	-3.21	-5.25	-4.44
Max	14.28	14.15	5.36	5.32	3.71	7.62	2.37

Notes: summary statistics of the estimated factors of the *macro-yields* model. L_{MY} denotes the level factor of the *macro-yields* model, S_{MY} is the slope of the *macro-yields* model, C_{MY} the curvature and U_{MY} is the unidentified factor. The table also reports the relative summary statistics for the Nelson and Siegel factors L_{NS} , S_{NS} and C_{NS} .

Table 2.5: Goodness of fit

maturities	OY	BMY	LMY	MY
3	0.063	0.051	0.046	0.266
12	0.010	0.010	0.010	0.017
36	0.006	0.006	0.006	0.011
60	0.008	0.008	0.008	0.007
120	0.024	0.024	0.024	0.039

Notes: MSE of the *only yields* model (OY), *basic macro-yields* model (BMY), the *large macro-yields* (LMY) and *macro-yields* model (MY) on the sample 1970:1 to 2000:12.

Table 2.6: Out-of-sample performance

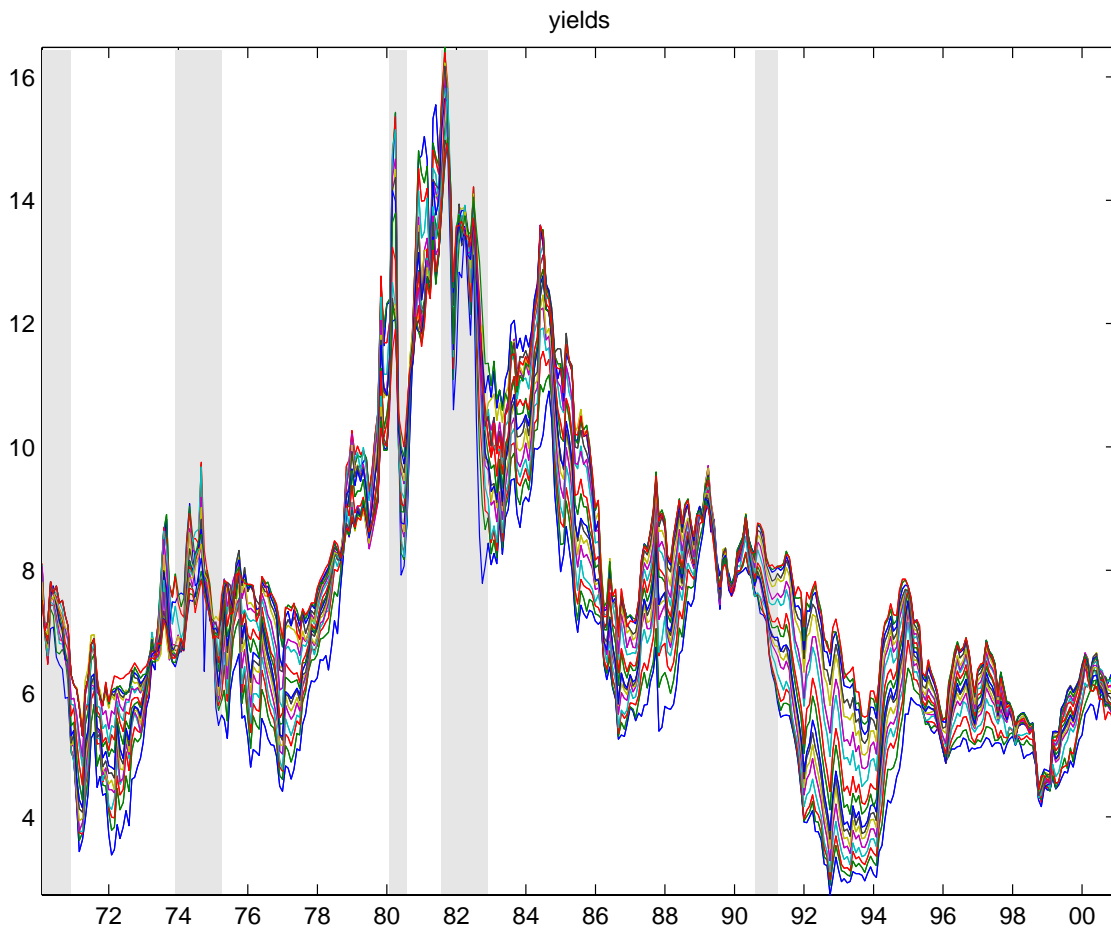
1-month ahead				
maturities	OY	BMY	LMY	MY
3	1.00	1.28	1.15	4.33
12	1.22	1.30	1.17	1.24
36	1.26	1.20	1.11	0.98
60	1.17	1.01	1.05	0.97
120	1.10	1.07	1.03	1.68

6-months ahead				
maturities	OY	BMY	LMY	MY
3	1.05	1.31	1.06	0.96
12	1.22	1.40	1.25	0.99
36	1.10	1.13	1.12	0.87
60	1.01	1.01	1.02	0.82
120	0.93	0.90	0.94	0.86

12-months ahead				
maturities	OY	BMY	LMY	MY
3	1.02	1.01	1.05	0.69
12	1.04	1.04	1.13	0.80
36	0.95	0.91	1.02	0.74
60	0.87	0.82	0.92	0.69
120	0.77	0.73	0.82	0.65

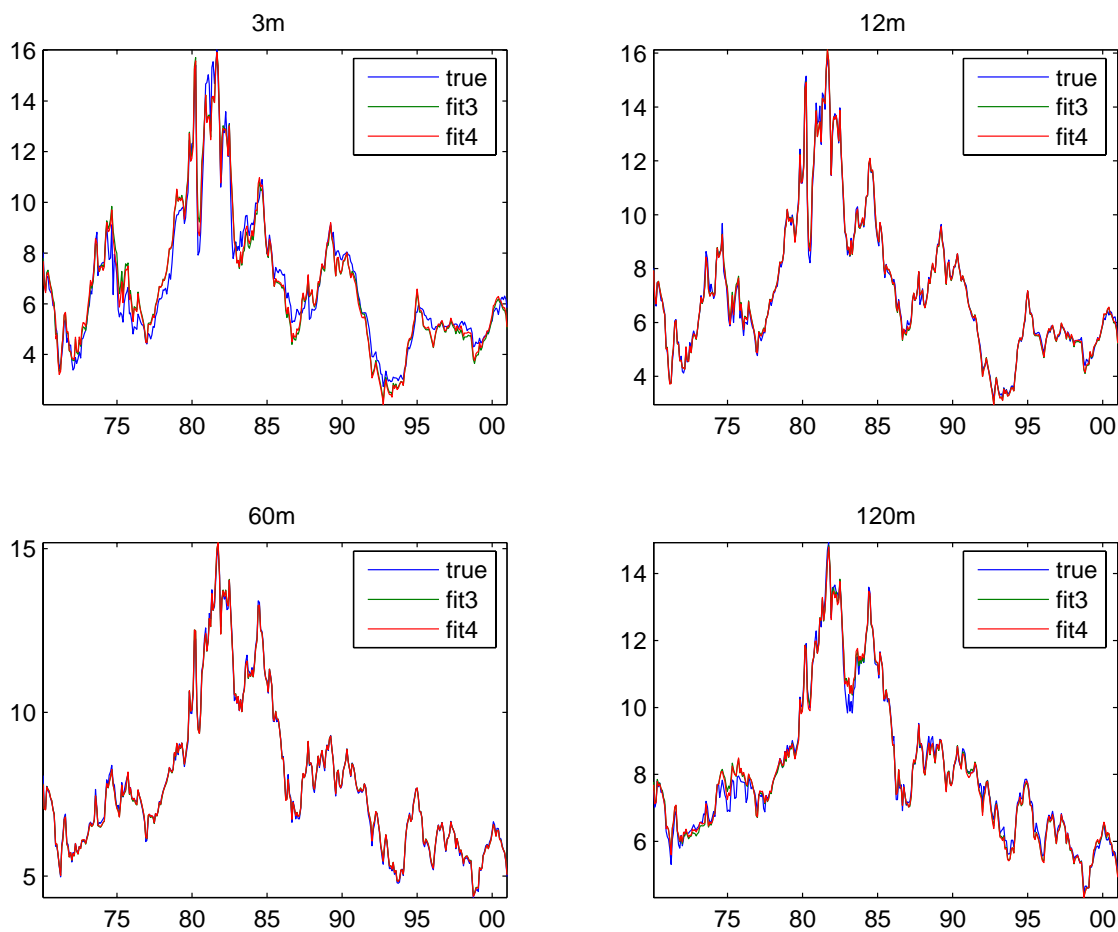
Notes: ratios of the MSFEs of the *only yields* model (OY), *basic macro-yields* model (BMY), *large macro-yields* model (LMY) and *macro-yields* model (MY) on the MSFE of the random walk, evaluated on the sample 1985:1 to 2000:12.

Figure 2.1: Yield data



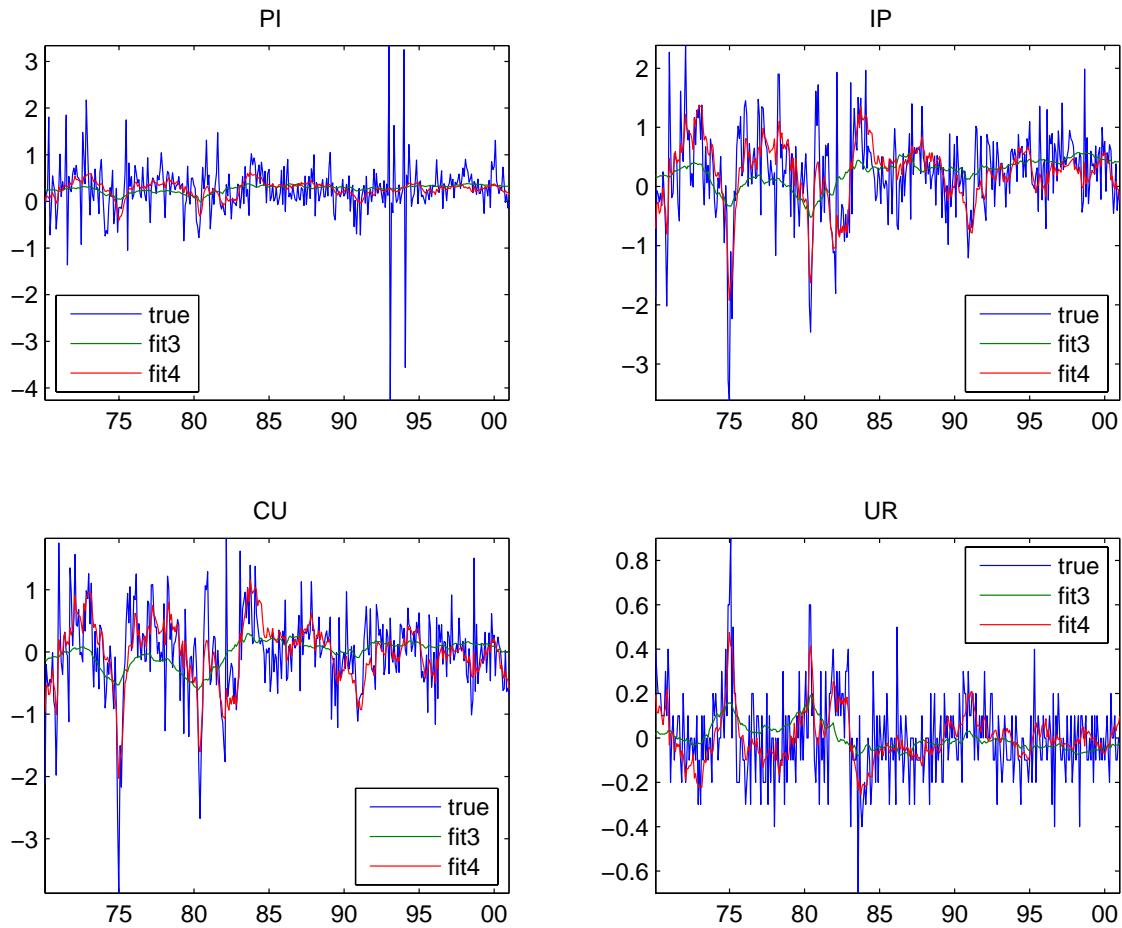
U.S. zero-coupon yield curve data at monthly frequency from 1970:1 to 2000:12 at maturities 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months. The grey-shaded areas indicate the recessions as defined by the NBER.

Figure 2.2: Macro-yields model in sample fit: yields



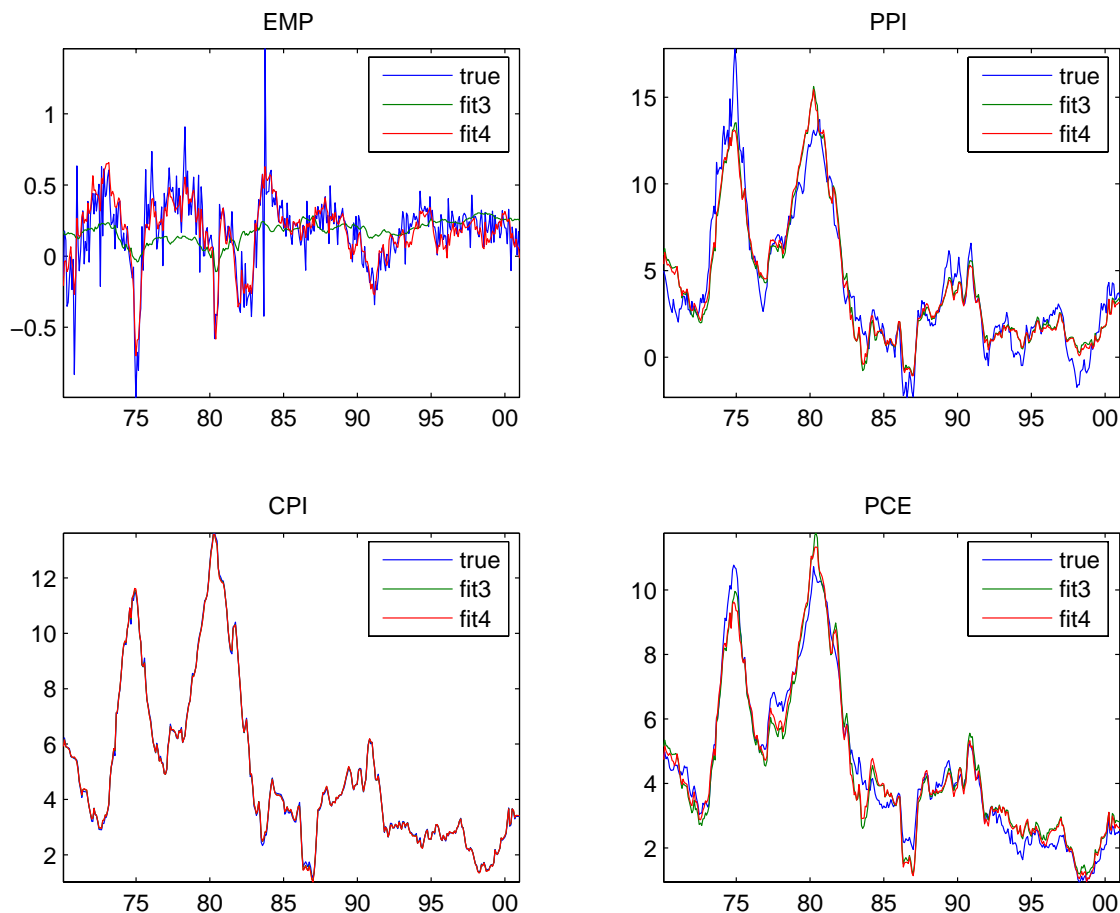
The figure displays the observed yields, for selected maturities, in blue and the relative in sample fit of the *macro-yields* model. The green line refers to the *macro-yields* model with only three factors, identified as the level, slope and curvature. The red line refers to the *macro-yields* model with four factors: the level, slope and curvature plus one unidentified factor. The first plot refers to the yields with maturity 3 months, the second to the yields with maturity 12 months, the third one to the yields with maturity 60 months and the last one to the yields with maturity 120 months

Figure 2.3: Macro-yields model in sample fit: key macro variables/1



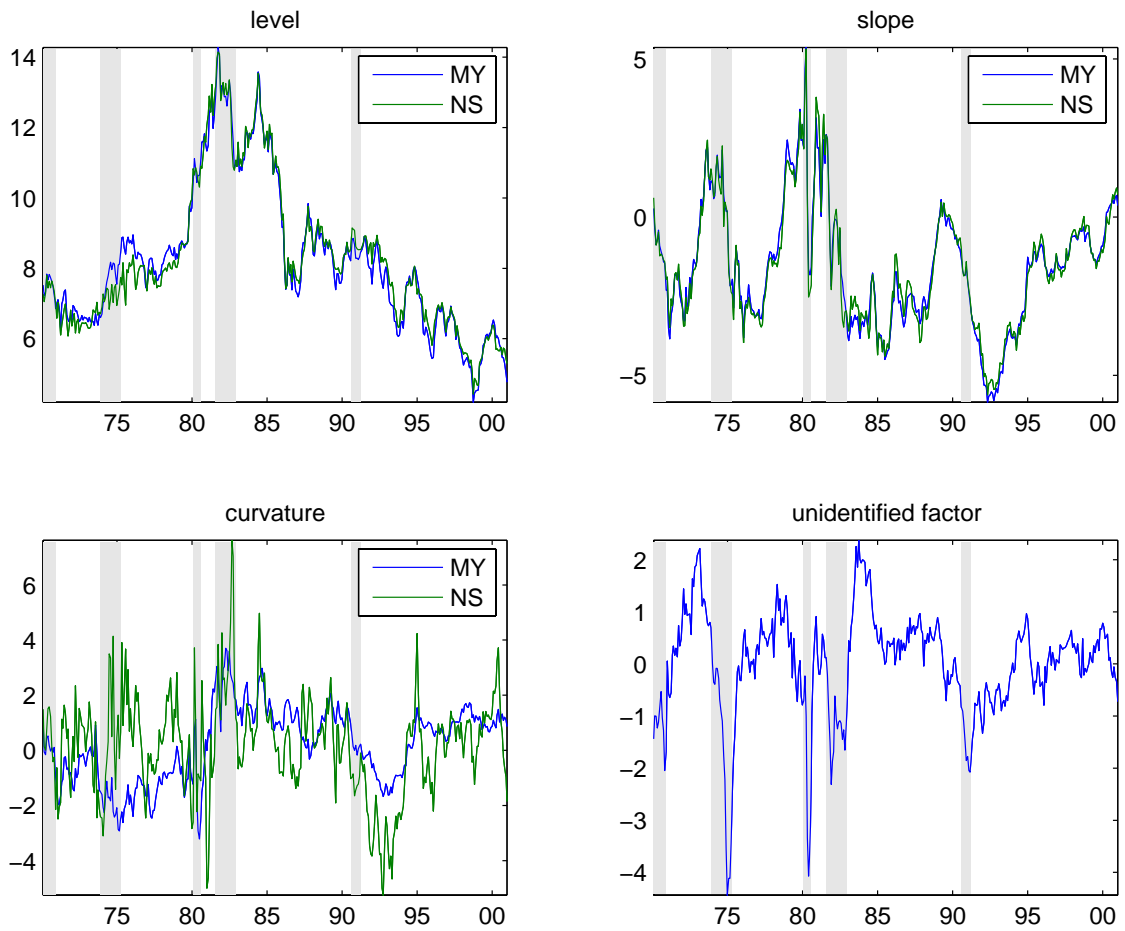
The figure displays some observed key macroeconomic variables in blue and the relative in sample fit of the *macro-yields* model. The green line refers to the *macro-yields* model with only three factors, identified as the level, slope and curvature. The red line refers to the *macro-yields* model with four factors: the level, slope and curvature plus one unidentified factor. The first plot refers to the personal income PI (the first variable in table 2.2), the second to the industrial production IP (variable number 6 in table 2.2), the third to the capacity of utilization CU (variable number 20 in table 2.2) and the last one to the unemployment rate UR (variable number 25 in table 2.2).

Figure 2.4: Macro-yields model in sample fit: key macro variables/2



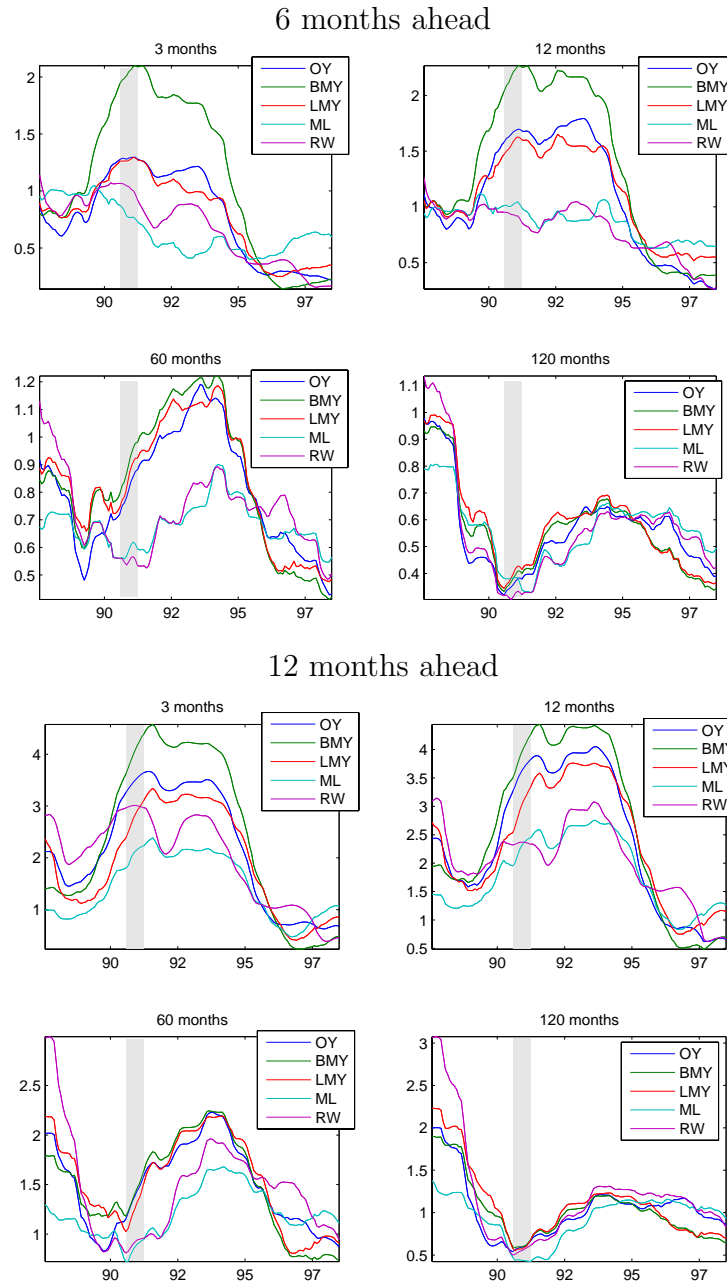
The figure displays some observed key macroeconomic variables in blue and the relative in sample fit of the *macro-yields* model. The green line refers to the *macro-yields* model with only three factors, identified as the level, slope and curvature. The red line refers to the *macro-yields* model with four factors: the level, slope and curvature plus one unidentified factor. The first plot refers to the employment EMP (variable number 33 table 2.2), the second to the producer price index PPI (variable number 95 in table 2.2), the third to the consumer price index CPI (variable number 101 in table 2.2) and the last one the personal consumption expenditure implicit price deflator PCE (variable number 111 in table 2.2).

Figure 2.5: Estimated macro-yields factors



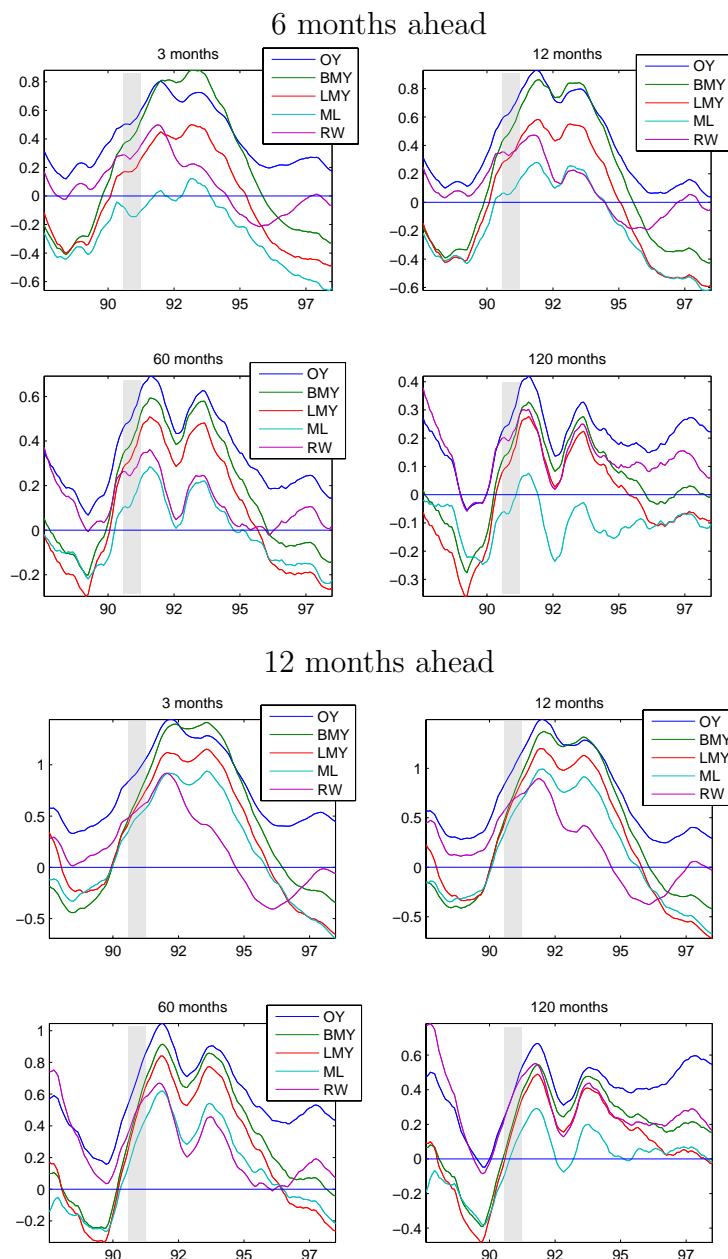
Estimated factors of the *macro-yields* model. The first plot displays the estimated level of the *macro-yields* model (MY) in blue and the relative Nelson and Siegel factor (NS) in green. The second plot refers to the slope, the third to the curvature and the last one to the unidentified factor. The gray-shaded area refers to the recessions.

Figure 2.6: Smoothed square forecast errors



Notes: 30 months moving average square forecast errors for the OY (*only yields*), BMY (*basic macro-yields*), LMY (*large macro-yields*) and MY (*macro-yields*) models for yields with maturity 3, 12, 60 and 120 months. The MSFE is shown for the out-of-sample period 1985:1-2000:12 for a 6-month horizon in top panel and a 12-month horizon in the bottom panel. The shadowed area indicates the recession between July 1990 and March 1991.

Figure 2.7: Smoothed forecast errors



Notes: 30 months moving average forecast errors for the OY (*only yields*), BMY (*basic macro-yields*), LMY (*large macro-yields*) and MY (*macro-yields*) models for yields with maturity 3, 12, 60 and 120 months. The MSFE is shown for the out-of-sample period 1985:1-2000:12 for a 6-month horizon in top panel and a 12-month horizon in the bottom panel. The shadowed area indicates the recession between July 1990 and March 1991.

Chapter 3

How Arbitrage-Free is the Nelson-Siegel Model?

ABSTRACT: This paper tests whether the Nelson and Siegel (1987) yield curve model is arbitrage-free in a statistical sense. Theoretically, the Nelson-Siegel model does not ensure the absence of arbitrage opportunities, as shown by Bjork and Christensen (1999). Still, central banks and public wealth managers rely heavily on it. Using a non-parametric resampling technique and zero-coupon yield curve data from the US market, we find that the no-arbitrage parameters are not statistically different from those obtained from the NS model, at a 95 percent confidence level. We therefore conclude that the Nelson and Siegel yield curve model is compatible with arbitrage-freeness.

Keywords: Nelson-Siegel model; No-arbitrage restrictions; affine term structure models; non-parametric test.

JEL classification: C14, C15, G12.

This chapter is adapted from the paper "How arbitrage free is the Nelson and Siegel model?" written with Ken Nyholm (ECB) and Rositsa Vidova-Koleva (ECB and Universitat Autònoma de Barcelona), ECB Working Paper 2008, No 874.

3.1 Introduction

Fixed-income wealth managers in public organizations, investment banks and central banks rely heavily on Nelson and Siegel (1987) type models to fit and forecast yield curves. According to BIS (2005), the central banks of Belgium, Finland, France, Germany, Italy, Norway, Spain, and Switzerland, use these models for estimating zero-coupon yield curves. The European Central Bank (ECB) publishes daily Eurosystem-wide yield curves on the basis of the Soderlind and Svensson (1997) model, which is an extension of the Nelson-Siegel model.¹ In its foreign reserve management framework the ECB uses a regime-switching extension of the Nelson-Siegel model, see Bernadell, Coche and Nyholm (2005).

There are at least four reasons for the popularity of the Nelson-Siegel model. First, it is easy to estimate. In fact, if the so-called time-decay-parameter is fixed, then Nelson-Siegel curves are obtained by linear regression techniques. If this parameter is not fixed, one has to resort to non-linear regression techniques. In addition, the Nelson-Siegel model can be adapted in a time-series context, as shown by Diebold and Li (2006). In this case the Nelson-Siegel yield-curve model can be seen as the observation equation in a state-space model, and the dynamic evolution of yield curve factors constitutes the transition equation. As a state-space model, estimation can be carried out via the Kalman filter. Second, by construction, the model provides yields for all maturities, i.e. also maturities that are not covered by the data sample. As such it lends itself as an interpolation and extrapolation tool for the analyst who often is interested in yields at maturities that are not directly observable.² Third, estimated yield curve factors obtained from the Nelson and Siegel model have intuitive interpretations, as level, slope (the difference between the long and the short end of the yield curve), and curvature of the yield curve. This interpretation is akin to that obtained by a principal component analysis (see,

¹For Eurosystem-wide yield curves see <http://www.ecb.int/stats/money/yc/html/index.en.html>.

²This is relevant e.g. in a situation where fixed-income returns are calculated to take into account the roll-down/maturity shortening effect.

e.g. Litterman and Scheinkman (1991) and Diebold and Li (2006)). Due to the intuitive appeal of the Nelson-Siegel parameters, estimates and conclusions drawn on the basis of the model are easy to communicate. Fourth, empirically the Nelson-Siegel model fits data well and performs well in out-of-sample forecasting exercises, as shown by e.g. Diebold and Li (2006) and De Pooter, Ravazzolo and van Dijk (2007).

Despite its empirical merits and wide-spread use in the finance community, two theoretical concerns can be raised against the Nelson-Siegel model. First, it is not theoretically arbitrage-free, as shown by Bjork and Christensen (1999). Second, as demonstrated by Diebold, Ji and Li (2004), it falls outside the class of affine yield curve models defined by Duffie and Kan (1996) and Dai and Singleton (2000).

The Nelson-Siegel yield curve model operates at the level of yields, as they are observed, i.e. under the so-called empirical measure. In contrast, affine arbitrage-free yield curve models specify the dynamic evolution of yields under a risk-neutral measure and then map this dynamic evolution back to the physical measure via a functional form for the market price of risk. The advantage of the no-arbitrage approach is that it automatically ensures a certain consistency between the parameters that describe the dynamic evolution of the yield curve factors under the risk-neutral measure, and the translation of yield curve factors into yields under the physical measure. An arbitrage-free setup will, by construction, ensure internal consistency as it cross-sectionally restricts, in an appropriate manner, the estimated parameters of the model. It is this consistency that guarantees arbitrage freeness. Since a similar consistency is not hard-coded into the Nelson-Siegel model, this model is not necessarily arbitrage-free.³

The main contribution of the current paper is to conduct a statistical test for the equality between the factor loadings of Nelson-Siegel model and the implied arbitrage-free loadings. In the context of a Monte Carlo study, the Nelson and Siegel factors are estimated and used as exogenous factors in an essentially-affine

³An illustrative example of this issue for a two-factor Nelson-Siegel model is presented by Diebold, Piazzesi and Rudebusch (2005).

term structure model to estimate the implied arbitrage-free factor loadings. The no-arbitrage model with time-varying term premia is estimated using the two-step approach of Ang, Piazzesi and Wei (2006), while we use the re-parametrization suggested by Diebold and Li (2006) as our specification of the Nelson-Siegel model.

In a recent study Christensen, Diebold and Rudebusch (2007) reconcile the Nelson and Siegel modelling setup with the absence of arbitrage by deriving a class of dynamic Nelson-Siegel models that fulfill the no-arbitrage constraints. They maintain the original Nelson-Siegel factor-loading structure and derive mathematically, a correction term that, when added to the dynamic Nelson-Siegel model, ensures the fulfillments of the no-arbitrage constraints. The correction term is shown to impact mainly very long maturities, in particular maturities above the ten-year segment.

While being different in setup and analysis method, our paper confirms the findings of Christensen et al. (2007). In particular, we find that the Nelson-Siegel model is not significantly different from a three-factor no-arbitrage model when it is applied to US zero-coupon yield-curve data. In addition, we outline a general method for empirically testing for the fulfillment of the no-arbitrage constraints in yield curve models that are not necessarily arbitrage-free. Our results furthermore indicate that non-compliance with the no-arbitrage constraints is most likely to stem from "mis-specification" in the Nelson-Siegel factor loading structure pertaining to the third factor, i.e. the one often referred to as the curvature factor.

Our test is conducted on U.S. Treasury zero-coupon yield data covering the period from January 1970 to December 2000 and spanning 18 maturities from 1 month to 10 years. We rely on a non-parametric resampling procedure to generate multiple realizations of the original data. Our approach to regenerate yield curve samples can be seen as a simplified version of the yield-curve bootstrapping approach suggested by Rebonato, Mahal, Joshi, Bucholz and Nyholm (2005).

In summary, we (1) generate a realization from the original yield curve data using a block-bootstrapping technique; (2) estimate the Nelson-Siegel model on the

regenerated yield curve sample; (3) use the obtained Nelson-Siegel yield curve factors as input for the essentially affine no-arbitrage model; (4) estimate the implied no-arbitrage yield curve factor loadings on the regenerated data sample. Steps (1) to (4) are repeated 1000 times in order to obtain bootstrapped distributions for the no-arbitrage parameters. These distributions are then used to test whether the implied no-arbitrage factor loadings are significantly different from the Nelson-Siegel loadings.

Our results show that the Nelson Siegel factor loadings are not statistically different from the implied no-arbitrage factor loadings at a 95 percent level of confidence. In an out-of-sample forecasting experiment, we show that the performance of the Nelson-Siegel model is as good as the no-arbitrage counterpart. We therefore conclude that the Nelson and Siegel model is compatible with arbitrage-freeness at this level of confidence.

3.2 Modeling framework

Term-structure factor models describe the relationship between observed yields, yield curve factors and loadings as given by

$$y_t = a + bX_t + \epsilon_t, \tag{3.2.1}$$

where y_t denotes a vector of yields observed at time t for N different maturities; y_t is then of dimension $(N \times 1)$. X_t denotes a $(K \times 1)$ vector of yield curve factors, where K counts the number of factors included in the model. The variable a is a $(N \times 1)$ vector of constants, b is of dimension $(N \times K)$ and contains the yield curve factor loadings. ϵ_t is a zero-mean $(N \times 1)$ vector of measurement errors.

The reason for the popularity of factor models in the area of yield curve modeling is the empirical observation that yields at different maturities generally are highly correlated. So, when the yield for one maturity changes, it is very likely that yields at other maturities also change. As a consequence, a parsimonious representation of

the yield curve can be obtained by modeling fewer factors than observed maturities.

This empirical feature of yields was first exploited in the continuous-time one factor models, where, in terms of equation (3.2.1), $X_t = r_t$, r_t being the short rate, see e.g. Merton (1973), Vasicek (1977), Cox, Ingersoll and Ross (1985), Black, Derman and Toy (1990), and Black and Karasinski (1993).⁴ A richer structure for the dynamic evolution of yield curves can be obtained by adding more yield curve factors to the model. Accordingly, X_t becomes a column-vector with a dimension equal to the number of included factors.⁵ The multifactor representation of the yield curve is also supported empirically by principal component analysis, see e.g. Litterman and Scheinkman (1991).

Multifactor yield curve models can be specified in different ways: the yield curve factors can be observable or unobserved, in which case their values have to be estimated alongside the other parameters of the model; the structure of the factor loadings can be specified in a way such that a particular interpretation is given to the unobserved yield curve factors, as e.g. Nelson and Siegel (1987) and Soderlind and Svensson (1997); or the factor loadings can be derived from no-arbitrage constraints, as in, among many others, Duffee (2002), Ang and Piazzesi (2003) and Ang, Bekaert and Wei (2007).

Yield curve models that are linear functions of the underlying factors can be written as special cases of equation (3.2.1).⁶ In this context, the two models used in the current paper are presented below.

3.2.1 The Nelson-Siegel model

The Nelson and Siegel (1987) model, as re-parameterized by Diebold and Li (2006), can be seen as a restricted version of equation (3.2.1) by imposing the following

⁴The merit of these models mainly lies in the area of derivatives pricing.

⁵Yield curve factor models are categorized by Duffie and Kan (1996) and Dai and Singleton (2000).

⁶Excluded from this list are naturally the quadratic term structure models as proposed by Ahn, Dittmar and Gallant (2002).

constraints:

$$a^{NS} = 0 \tag{3.2.2}$$

$$b^{NS} = \left[1 \quad \frac{1 - \exp(-\lambda\tau)}{\lambda\tau} \quad \frac{1 - \exp(-\lambda\tau)}{\lambda\tau} - \exp(-\lambda\tau) \right], \tag{3.2.3}$$

where λ is the exponential decay rate of the loadings for different maturities, and τ is time to maturity. This particular loading structure implies that the first factor is responsible for parallel yield curve shifts, since the effect of this factor is identical for all maturities; the second factor represents minus the yield curve slope, because it has a maximal impact on short maturities and minimal effect on the longer maturity yields; and, the third factor can be interpreted as the curvature of the yield curve, because its loading has a hump in the middle part of the maturity spectrum, and little effect on both short and long maturities. In summary, the three factors have the interpretation of a yield curve level, slope and curvature.

[FIGURE 3.1 AROUND HERE]

A visual representation of the Nelson and Siegel factor loading structure is given in Figure 3.1. By imposing the restrictions (3.2.2) to (3.2.3) on equation (3.2.1) we obtain

$$y_t = b^{NS} X_t^{NS} + \epsilon_t^{NS}, \tag{3.2.4}$$

where $X_t^{NS} = [L_t \quad S_t \quad C_t]$ represents the Nelson-Siegel yield curve factors: Level, Slope and Curvature, at time t .

Empirically the Nelson-Siegel model fits data well, as shown by Nelson and Siegel (1987), and performs relatively well in out-of-sample forecasting exercises (see among others, Diebold and Li (2006) and De Pooter et al. (2007)). However, as mentioned in the introduction, from a theoretical viewpoint the Nelson-Siegel yield curve model is not necessarily arbitrage-free (e.g. see Bjork and Christensen (1999)) and does not belong to the class of affine yield curve models (e.g. see

Diebold et al. (2004)).

3.2.2 Gaussian arbitrage-free models

The Gaussian discrete-time arbitrage-free affine term structure model can also be seen as a particular case of equation (3.2.1), where the factor loadings are cross-sectionally restricted to ensure the absence of arbitrage opportunities. This class of no-arbitrage (NA) models can be represented by

$$y_t = a^{NA} + b^{NA} X_t^{NA} + \epsilon_t^{NA}, \quad (3.2.5)$$

where the underlying factors are assumed to follow a Gaussian VAR(1) process

$$X_t^{NA} = \mu + \Phi X_{t-1}^{NA} + u_t,$$

with $u_t \sim N(0, \Sigma \Sigma')$ being a $(K \times 1)$ vector of errors, μ is a $(K \times 1)$ vector of means, and Φ is a $(K \times K)$ matrix collecting the autocorrelation coefficients. The elements of a^{NA} and b^{NA} in equation (3.2.5) are defined by

$$a_\tau^{NA} = -\frac{A_\tau}{\tau}, \quad b_\tau^{NA} = -\frac{B_\tau}{\tau}, \quad (3.2.6)$$

where, as shown by e.g. Ang and Piazzesi (2003), A_τ and B_τ satisfy the following recursive formulas to preclude arbitrage opportunities

$$A_{\tau+1} = A_\tau + B'_\tau (\mu - \Sigma \lambda_0) + \frac{1}{2} B'_\tau \Sigma \Sigma' B_\tau - A_1, \quad (3.2.7)$$

$$B'_{\tau+1} = B'_\tau (\Phi - \Sigma \lambda_1) - B'_1, \quad (3.2.8)$$

with boundary conditions $A_0 = 0$ and $B_0 = 0$. The parameters λ_0 and λ_1 govern the time-varying market price of risk, specified as an affine function of the yield

curve factors

$$\Lambda_t = \lambda_0 + \lambda_1 X_t^{NA}.$$

The coefficients $A_1 = -a_1^{NA}$ and $B_1 = -b_1^{NA}$ in equations (3.2.7) to (3.2.8) refer to the short rate equation

$$r_t = a_1^{NA} + b_1^{NA} X_t^{NA} + v_t,$$

where usually r_t is approximated by the one-month yield.

If the factors X_t^{NA} driving the dynamics of the yield curve are assumed to be unobservable, the estimation of affine term structure models requires a joint procedure to extract the factors and to estimate the parameters of the model. This is a difficult task, given the non-linearity of the model and that the number of parameters grows with the number of included factors. As the factors are latent, identifying restrictions have to be imposed. Moreover, as mentioned by Ang and Piazzesi (2003), the likelihood function is flat in the market-price-of-risk parameters and this further complicates the numerical estimation process.

The most common procedure to estimate affine term structure models is described by Chen and Scott (1993). It relies on the assumption that as many yields, as factors, are observed without measurement error. Hence, it allows for recovering the latent factors from the observed yields by inverting the yield curve equation. Unfortunately, the estimation results will depend on which yields are assumed to be measured without error and will vary according to the choice made. Alternatively, to reduce the degree of arbitrariness, observable factor can be used. For example, Ang et al. (2006) use the short rate, the spread and the quarterly GDP growth rate as yield curve factors. It is also possible to rely on pure statistical techniques in the determination of yield curve factors, as e.g. De Pooter et al. (2007) who use extracted principal components as yield curve factors.

3.2.3 Motivation

The affine no-arbitrage term structure models impose a structure on the loadings a^{NA} and b^{NA} , presented in equations (3.2.6) to (3.2.8), such that the resulting yield curves, in the maturity dimension, are compatible with the estimated time-series dynamics for the yield curve factors. This hard-coded internal consistency between the dynamic evolution of the yield curve factors, and hence the yields at different maturity segments of the curve, is what ensures the absence of arbitrage opportunities. A similar constraint is not integrated in the setup of the Nelson-Siegel model (see, Bjork and Christensen (1999)).

However, in practice, when the Nelson-Siegel model is estimated, it is possible that the no-arbitrage constraints are approximately fulfilled, i.e. fulfilled in a statistical sense, while not being explicitly imposed on the model. It cannot be excluded that the functional form of the yield curve, as it is imposed by the Nelson and Siegel factor loading structure in equations (3.2.2) and (3.2.3), fulfils the no-arbitrage constraints most of the times.

As a preliminary check for the comparability of the Nelson-Siegel model and the no-arbitrage model, Figure 3.2 compares extracted yield curve factors i.e. \hat{X}_t^{NA} and \hat{X}_t^{NS} for US data from 1970 to 2000 (the data is presented in Section 3.3). We estimate the Nelson-Siegel factors as in Diebold and Li (2006), and the no-arbitrage model as in Ang and Piazzesi (2003) using the Chen and Scott (1993) method, and assuming that yields at maturities 3, 24, 120 months are observed without error.

[FIGURE 3.2 AROUND HERE]

Although the two models have different theoretical backgrounds and use different estimation procedures, the extracted factors are highly correlated. Indeed, the estimated correlation between the Nelson-Siegel level factor and the first latent factor from the no-arbitrage model is 0.95. The correlation between the slope and the second latent factor is 0.96 and between the curvature and the third latent factor is 0.65.⁷

⁷Correlations are reported in absolute value.

On the basis of these results and in order to properly investigate whether the Nelson-Siegel model is compatible with arbitrage-freeness, we conduct a test for the equality of the Nelson-Siegel factor loadings to the implied no-arbitrage ones obtained from an arbitrage-free model. To ensure correspondence between the Nelson-Siegel model and its arbitrage-free counterpart, we use extracted Nelson-Siegel factors as exogenous factors in the no-arbitrage setup. The model that we estimate is the following

$$y_t = a^{NA} + b^{NA} \widehat{X}_t^{NS} + \epsilon_t^{NA}, \quad \epsilon_t^{NA} \sim (0, \Omega), \quad (3.2.9)$$

where \widehat{X}_t^{NS} are the estimated Nelson-Siegel factors from equations (3.2.2) to (3.2.4), the observation errors ϵ_t^{NA} are not assumed to be normally distributed and a^{NA} and b^{NA} satisfy the no-arbitrage restrictions presented in equations (3.2.6) to (3.2.8). In order to impose these no-arbitrage restrictions we have to fit a VAR(1) on the estimated Nelson-Siegel factors

$$\widehat{X}_t^{NS} = \mu + \Phi \widehat{X}_{t-1}^{NS} + u_t, \quad (3.2.10)$$

with $u_t \sim N(0, \Sigma \Sigma')$, to specify the market price of risk as an affine function of the estimated Nelson-Siegel factors

$$\Lambda_t = \lambda_0 + \lambda_1 \widehat{X}_t^{NS}, \quad (3.2.11)$$

and the short rate equation as

$$r_t = a_1^{NA} + b_1^{NA} \widehat{X}_t^{NS} + v_t. \quad (3.2.12)$$

In this way, we estimate the no-arbitrage factor loading structure that emerges when the underlying yield curve factors are identical to the Nelson-Siegel yield curve factors. The test is then formulated in terms of the equality between the intercepts of the two models, a^{NS} and a^{NA} , and the relative loadings, b^{NA} and b^{NS} .

3.3 Data

We use U.S. Treasury zero-coupon yield curve data covering the period from January 1970 to December 2000 constructed by Diebold and Li (2006), based on end-of-month CRSP government bond files.⁸ The data is sampled at a monthly frequency providing a total of 372 observations for each of the maturities observed at the (1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, 120) month segments.

[FIGURE 3.3 AROUND HERE]

The data is presented in Figure 3.3. The surface plot illustrates how the yield curve evolves over time. Table 3.1 reports the mean, standard deviation and autocorrelations to further illustrate the properties of the data.

[TABLE 3.1 AROUND HERE]

The estimated autocorrelation coefficients are significantly different from zero at a 95 percent level of confidence for lag one through twelve, across all maturities.⁹ Such high autocorrelations could suggest that the underlying yield series are integrated of order one. If this is the case, we would need to take first-differences to make the variables stationary before valid statistical inference could be drawn, or we would have to resort to co-integration analysis. However, economic theory tells us that nominal yield series cannot be integrated, since they have a lower bound support at zero and an upper bound support lower than infinity. Consequently, and in accordance with the yield-curve literature, we model yields in levels and thus disregard that their in-sample properties could indicate otherwise.¹⁰

⁸The data can be downloaded from <http://www.ssc.upenn.edu/fdiebold/papers/paper49/FBFITTED.txt> and Diebold and Li (2006, pp. 344-345) give a detailed description of the data treatment methodology applied.

⁹A similar degree of persistence in yield curve data is also noted by Diebold and Li (2006).

¹⁰It is often the case in yield-curve modeling that yields are in levels. See, among others, Nelson and Siegel (1987), Diebold and Li (2006), Diebold, Rudebusch and Aruoba (2006), Ang and Piazzesi (2003) and Dai and Singleton (2000).

3.4 Estimation Procedure

To estimate the Nelson-Siegel factors \widehat{X}_t^{NS} in equation (3.2.4), we follow Diebold and Li (2006) by fixing the decay parameter $\lambda = 0.0609$ in equation (3.2.3) and by using OLS.¹¹ We treat the obtained Nelson-Siegel factors as observable in the estimation of the no-arbitrage model presented in equations (3.2.6) to (3.2.12).

To estimate the parameters of the arbitrage-free model we standardize the Nelson and Siegel factors and use the two-step procedure proposed by Ang et al. (2006). In the first step, we fit a VAR(1) for the standardized Nelson-Siegel factors to estimate $\widehat{\mu}$, $\widehat{\Phi}$ and $\widehat{\Sigma}$ from equation (3.2.10). And, we project the short rate (one-month yield) on the standardized Nelson-Siegel yield curve factors, to estimate the parameters in the short rate equation (3.2.12). In the second step, we minimize the sum of squared residuals between observed yields and fitted yields to estimate the market-price-of-risk parameters $\widehat{\lambda}_0$ and $\widehat{\lambda}_1$ of equation (3.2.11). Finally, we un-standardize the Nelson-Siegel factors and compute \widehat{a}^{NA} and \widehat{b}^{NA} .

Our goal is to test whether the Nelson-Siegel model in equations (3.2.2) to (3.2.4) is statistically different from the no-arbitrage model in equations (3.2.6) to (3.2.12). Since the estimated factors, \widehat{X}_t^{NS} are the same for both models we can formulate our hypotheses in the following way:

$$\begin{aligned} H_0^1 &: a_\tau^{NA} = a_\tau^{NS} \equiv 0, \\ H_0^2 &: b_\tau^{NA}(1) = b_\tau^{NS}(1), \\ H_0^3 &: b_\tau^{NA}(2) = b_\tau^{NS}(2), \\ H_0^4 &: b_\tau^{NA}(3) = b_\tau^{NS}(3), \end{aligned}$$

where $b_\tau^{NA}(k)$ denotes the loadings on the k-th factor in the no-arbitrage model at maturity τ , and $b_\tau^{NS}(k)$ denotes the corresponding variable from the Nelson-Siegel model.

¹¹This value of λ maximizes the loading on the curvature at 30 months maturity as shown by Diebold and Li (2006).

We claim that the Nelson-Siegel model is statistically compatible with arbitrage-freeness if H_0^1 to H_0^4 are not rejected at traditional levels of confidence. Notice that to test for H_0^1 to H_0^4 we only need to estimate a^{NA} and b^{NA} , since the Nelson-Siegel loading structure is fixed from the model. To account for the two-step estimation procedure of the no-arbitrage model and for the generated regressor problem, we construct confidence intervals around \hat{a}^{NA} and \hat{b}^{NA} using the resampling procedure described in the next section.

3.4.1 Resampling procedure

To recover the empirical distributions of the estimated parameters we conduct block resampling and reconstruct multiple yield curve data samples from the original yield curve data in the following way. We denote with G the matrix of observed yield ratios with elements $y_{t,\tau}/y_{t-1,\tau}$ where $t = (2, \dots, T)$ and $\tau = (1, \dots, N)$.

We first randomly select a starting yield curve y_k , where the index k is an integer drawn randomly from a discrete uniform distribution $[1, \dots, T]$. The resulting k marks the random index value at which the starting yield curve is taken.

In a second step, blocks of length w are sampled from the matrix of yield ratios G . The generic i -th block can be denoted by $\tilde{g}_{z,i}$ where z is a random number from $[2, \dots, T - w + 1]$ denoting the first observation of the block and I is the maximum number of blocks drawn, $i = 1 \dots I$.¹² A full data-sample of regenerated yield curve ratios \tilde{G} can then be constructed by vertical concatenation of the drawn data blocks $\tilde{g}_{z,i}$ for $i = 1 \dots I$.

Finally, a new data set of resampled yields can be constructed via:

$$\begin{cases} \tilde{y}_1 &= y_k \\ \tilde{y}_s &= \tilde{y}_{s-1} \odot \{\tilde{G}\}_s, \quad s = 2, \dots, S, \end{cases} \quad (3.4.1)$$

where $\{\tilde{G}\}_s$ denotes the s^{th} row of the matrix of resampled ratios \tilde{G} , and \odot denotes

¹²We use \sim to indicate the re-sampled variables.

element by element multiplication.

We choose to resample from yield ratios for two reasons. First, it ensures positivity of the resampled yields. Second, as reported in Table 3.1, yields are highly autocorrelated and close to $I(1)$. Therefore, one could resample from first differences, but as reported in Table 3.2, first differences of yields are highly autocorrelated and not variance-stationary. Yield ratios display better statistical properties regarding variance-stationarity, as can be seen by comparing the correlation coefficients for squared differences and ratios in Table 3.2. Block-bootstrapping is used to account for serial correlation in the yield curve ratios.

[TABLE 3.2 AROUND HERE]

A similar resampling technique has been proposed by Rebonato et al. (2005). They provide a detailed account for the desirable statistical features of this approach. In the present context we recall that the method ensures: (i) the exact asymptotic recovery of all the eigenvalues and eigenvectors of yields; (ii) the correct reproduction of the distribution of curvatures of the yield curve across maturities; (iii) the correct qualitative recovery of the transition from super- to sub-linearity as the yield maturity is increased in the variance of n-day changes, and (iv) satisfactory accounting of the empirically-observed positive serial correlations in the yields.

To test hypotheses H_0^1 to H_0^4 we employ the following scheme:

1. Construct a yield curve sample \tilde{y} following equation (3.4.1);
2. Estimate the Nelson-Siegel yield curve factors \tilde{X}_t^{NS} on \tilde{y} ;
3. Use \tilde{X}_t^{NS} to estimate the parameters \tilde{a}^{NA} and \tilde{b}^{NA} from the arbitrage-free model given in equations (3.2.6) - (3.2.12);
4. Repeat steps 1 to 3, 1000 times to build a distribution for the parameter estimates \hat{a}^{NA} and \hat{b}^{NA} ;

5. Construct confidence intervals for \hat{a}^{NA} and \hat{b}^{NA} using the sample quantiles of the empirical distribution of the estimated parameters.

Note that by fixing λ in step 2, the Nelson-Siegel factor loading structure remains unchanged from repetition to repetition. We set the block length equal to 50 observations, i.e. $w = 50$, and generate a total of 370 yield curve observations for each replication, i.e. $S = 370$.¹³

3.5 Results

This section presents three sets of results to help assess whether the Nelson-Siegel model is compatible with arbitrage-freeness when applied to US zero-coupon data. Our main result is a test of equality of the factor loadings on the basis of the resampling technique outlined in section 3.4. In addition we compare the in-sample and out-of-sample performance of the Nelson-Siegel model, equations (3.2.2) - (3.2.4), to the no-arbitrage model based on exogenous Nelson-Siegel yield curve factors, equations (3.2.6) - (3.2.12).

3.5.1 Testing results

Using the resampling methodology outlined in section 3.4, we generate empirical distributions for each factor loading of the no-arbitrage yield curve model in equation (3.2.9). Results are presented for each maturity covered by the original data sample. The Nelson-Siegel factor loading structure, in equations (3.2.2) and (3.2.3), is constant across all bootstrapped data sampled because λ is treated as a known parameter.¹⁴ Hence, only the extracted Nelson-Siegel factors vary across the bootstrap samples.

¹³The last block is drawn to contain 20 observations as to obtain a total number of observations for each regenerated sample close to the number of observations of the original sample, 372.

¹⁴The results presented in the paper are robust to changes in λ . We have performed the calculations for other values of λ , namely $\lambda = 0.08$, $\lambda = 0.045$, and $\lambda = 0.0996$, and the results for these values of λ are qualitatively the same as the ones presented in the paper.

Parameter estimates and corresponding empirical confidence intervals for the no-arbitrage model, equations (3.2.6) - (3.2.12), are shown in Table 3.3. The diagonal elements of the matrices holding the estimated autoregressive coefficients $\widehat{\Phi}$ and the covariance matrix of the VAR residuals $\widehat{\Sigma}$, in equation (3.2.10), are significantly different from zero at a 95 percent level of confidence. In addition, the estimates of a_1^{NA} , and the two first elements of the (3×1) vector b_1^{NA} in equation (3.2.12), are also different from zero, judged at the same level of confidence.

[TABLE 3.3 AROUND HERE]

The estimated intercepts of the no-arbitrage model \hat{a}^{NA} , computed as in equations (3.2.6)- (3.2.7), are presented in Table 3.4, for each maturity covered by the original data. This table reports also the 95 percent confidence intervals, obtained from the resampling, and the Nelson-Siegel intercepts, a^{NS} . Therefore, results in Table 3.4 allow for testing H_0^1 for the equality between the intercepts in the yield curve equations for the no-arbitrage and the Nelson-Siegel models. Tables 3.5 to 3.7 present the corresponding results that allow us to test H_0^2 , H_0^3 , and H_0^4 , i.e. whether the corresponding yield curve factor loadings are identical, in a statistical sense.

[TABLE 3.4 to 3.7 AROUND HERE]

Figure 3.4 gives a visual representation of the results contained in Tables 3.4 to 3.7. The figure shows the estimated no-arbitrage loadings, \hat{a}^{NA} and \hat{b}^{NA} , with the relative 50 percent and 95 percent empirical confidence intervals obtained from resampling, as well as the parameter values for the Nelson-Siegel model, b^{NS} , for comparison.

It is clear from Figure 3.4 that the empirical distributions are highly skewed for most of the maturities. Consider, for example, the plot for the intercept estimates (the top left plot in Figure 3.4) at maturity 120. It is evident that the distribution of the no-arbitrage coefficient is highly right skewed.

[FIGURE 3.4 AROUND HERE]

This non-normality of the distributions for the estimated no-arbitrage parameters, is further analyzed in Table 3.8. This table shows that all distributions display skewness, excess kurtosis, or both. Selected maturities are shown in Table 3.8, however, this result holds for all maturities included in the sample. We also perform the Jarque-Bera test for normality, and reject normality at a 95 percent confidence level for all maturities.

[TABLE 3.8 AROUND HERE]

Visual confirmation of the documented non-normality is provided by Figures 3.5 to 3.8. For a representative selection of maturities, these figures show the empirical distribution of the estimated no-arbitrage loadings, and a normal distribution approximation. In addition, the figures show the 95 percent confidence intervals derived from the empirical distribution and the normal approximation.

[FIGURE 3.5 to 3.8 AROUND HERE]

The non-normality of the empirical distributions for the bootstrapped intercepts \hat{a}^{NA} , and factor loadings \hat{b}^{NA} , indicates that the confidence intervals should be constructed using the sample quantiles of the empirical distribution. The empirical 95 percent confidence intervals are included in Tables 3.4, 3.5, 3.6 and 3.7. The lower bound of the confidence intervals is denoted by a subscript L , and the upper bound by a U .

By inspecting the tables, we reach the following conclusions for the tested hypotheses:

$H_0^1 : a_\tau^{NA} = a_\tau^{NS} \equiv 0$	not rejected at a 95% level of confidence,
$H_0^2 : b_\tau^{NA}(1) = b_\tau^{NS}(1)$	not rejected at a 95% level of confidence,
$H_0^3 : b_\tau^{NA}(2) = b_\tau^{NS}(2)$	not rejected at a 95% level of confidence,
$H_0^4 : b_\tau^{NA}(3) = b_\tau^{NS}(3)$	not rejected at a 95% level of confidence.

For the test of the curvature parameter in H_0^4 an additional comment is warranted. As can be seen from Figure 3.4, the curvature parameter, at middle maturities, is the closest to violating the 95 percent confidence band, and this parameter thus constitutes the “weak point” of the Nelson-Siegel model in relation to the no-arbitrage constraints. This finding is in line with Bjork and Christensen (1999) who prove that a Nelson-Siegel type model with two additional curvature factors, each with its own λ , theoretically would be arbitrage-free. However, when acknowledging that Litterman and Scheinkman (1991) find that the curvature factor only accounts for approximately 2 percent of the variation of yields, and in the light of our results, one can question the significance of imposing constraints on parameters that have an explanatory power in the range of 2 percent. Our empirical finding is also supported by the theoretical results in Christensen et al. (2007) who show that adding an additional term at very long maturities reconciles the dynamic Nelson-Siegel model with the affine arbitrage-free term structure models.

Using yield curve modeling for purposes other than relative pricing, as for example central bankers and fixed-income strategists do, one might be tempted to use the Nelson-Siegel model on the basis of its arbitrage-freeness compatibility.

The hypothesis H_0^1 through H_0^4 test the equality between each no-arbitrage factor loading and the corresponding Nelson-Siegel factor loading separately. The results reported above are confirmed by a joint F test. To perform the test we use the empirical variance-covariance matrix of the estimates. The test statistic is 0.22 and the 95 percent critical F-value with 72 and 300 degrees of freedom is 1.34. Therefore, we also cannot reject the hypothesis that the loading structures of the two models are equal in a statistical sense.

3.5.2 In-sample comparison

To conduct an in-sample comparison of the two models, we estimate the Nelson-Siegel model in equations (3.2.2) - (3.2.4) and the no-arbitrage model in equations (3.2.6) - (3.2.12), where the latter model uses the yield curve factors extracted from

the former. Measures of fit are displayed in Table 3.9.

A general observation is that both models fit data well: the means of the residuals for all maturities are close to zero and show low standard deviations. The root mean squared error, RMSE, and the mean absolute deviation, MAD, are also low and similar for both models.

More specifically, Table 3.9 shows that the averages of the residuals from the fitted Nelson-Siegel model, $\hat{\epsilon}^{NS}$, for the included maturities, are all lower than 16 basis points, in absolute value. In fact, the mean of the absolute residuals across maturities is 5 basis points, while the corresponding number for $\hat{\epsilon}^{NA}$ is 3 basis points. The 3 months maturity is the worst fitted maturity for the no-arbitrage model with a mean of the residuals of 8 basis points. For the Nelson-Siegel model the worst fitted maturity is the 1 month segment with a mean of the residuals close to -16 bp. Furthermore, the two models have the same amount of autocorrelation in the residuals. A similar observation is made for the Nelson-Siegel model alone by Diebold and Li (2006).

[TABLE 3.9 AROUND HERE]

Drawing a comparison on the basis of RMSE and MAD figures gives the conclusion that both models fit data equally well.

3.5.3 Out-of-sample comparison

As a last comparison-check of the equivalence of the Nelson-Siegel model and the no-arbitrage counterpart, we perform an out-of-sample forecast experiment. In particular, we generate h -steps ahead iterative forecasts in the following way. First, the yield curve factors are projected forward using the estimated VAR parameters from equation (3.2.10)

$$\hat{X}_{t+h|t}^{NS} = \sum_{s=0}^{h-1} \hat{\Phi}^s \hat{\mu} + \hat{\Phi}^h \hat{X}_t^{NS},$$

where $h \in \{1, 6, 12\}$ is the forecasting horizon in months. Second, out-of-sample forecasts are calculated for the two models, given the projected factors,

$$\begin{aligned}\hat{y}_{t+h|t}^{NS} &= b^{NS} \hat{X}_{t+h|t}^{NS}, \\ \hat{y}_{t+h|t}^{NA} &= \hat{a}_t^{NA} + \hat{b}_t^{NA} \hat{X}_{t+h|t}^{NS},\end{aligned}$$

where subscripts t on \hat{a}_t^{NA} and \hat{b}_t^{NA} indicate that parameters are estimated using data until time t . To evaluate the prediction accuracy at a given forecasting horizon, we use the mean squared forecast error, MSFE, the average squared error over the evaluation period, between t_0 and t_1 , for the h -months ahead forecast of the yield with maturity τ

$$MSFE(\tau, h, m) = \frac{1}{t_1 - t_0 + 1} \sum_{t=t_0}^{t_1} (\hat{y}_{t+h,\tau|t}^m - y_{t+h,\tau})^2, \quad (3.5.1)$$

where $m \in \{NA, NS\}$ denotes the model.

The results presented are expressed as ratios of the MSFEs of the two models against the MSFE of a random walk. The random walk represents a naïve forecasting model that historically has proven very difficult to outperform. The success of the random walk model in the area of yield curve forecasting is due to the high degree of persistence exhibited by observed yields. The random walk h -step ahead prediction, at time t , of the yield with maturity τ is

$$\hat{y}_{t+h,\tau|t} = y_{t,\tau}.$$

To produce the first set of forecasts, the model parameters are estimated on a sample defined from 1970:01 to 1993:01, and yields are forecasted for the chosen horizons, h . The data sample is then increased by one month and the parameters are re-estimated on the new data covering 1970:01 to 1993:02. Again, forecasts are produced for the forecasting horizons. This procedure is repeated for the full sample, generating forecasts on successively increasing data samples. The forecasting

performances are then evaluated over the period 1994:01 to 2000:12 using the mean squared forecast error, as shown in equation (3.5.1).

Table 3.10 reports on the out-of-sample forecast performance of the Nelson-Siegel and the implied no-arbitrage model evaluated against the random walk forecasts.

[FIGURE 3.10 AROUND HERE]

The well-known phenomenon of the good forecasting performance of the random walk model is observed for the 1 month forecasting horizon. For the 6 and 12 month forecasting horizons, the Nelson-Siegel model and the no-arbitrage counterpart generally perform better than the random walk model, as shown by ratios being less than one.

Turning now to the relative comparison of the no-arbitrage model against the Nelson-Siegel model, it can be concluded that they exhibit very similar forecasting performances. If we consider every maturity for each forecasting horizon as an individual observation, then there are in total 54 observations. In 18 of these cases the Nelson-Siegel model is better, in 24 cases the no-arbitrage model is better, and in the remaining 12 cases the models perform equally well. Even when one model is judged to be better than its competitor, the differences in the performance ratios are very small. Typically, a difference is only seen at the second decimal with a magnitude of 1 to 3 basis points.

In summary, it can be concluded that there is no systematic pattern across maturities and forecasting horizons showing when one model is better than its competitor. Indeed, to formally compare the forecasting performance of the two models we calculate the Diebold-Mariano statistic for each maturity and forecasting horizon. At a 5 percent level we do not reject the hypothesis that the no-arbitrage model and the Nelson-Siegel model forecast equally well, see Table 3.11.

[TABLE 3.11 AROUND HERE]

3.6 Conclusion

In this paper we show that the model proposed by Nelson and Siegel (1987) is compatible with arbitrage-freeness, in the sense that the factor loadings from the model are not statistically different from those derived from an arbitrage-free model which uses the Nelson-Siegel factors as exogenous factors, at a 95 percent level of confidence.

In theory, the Nelson-Siegel model is not arbitrage-free as shown by Bjork and Christensen (1999). However, using US zero-coupon data from 1970 to 2000, a yield curve bootstrapping approach and the implied arbitrage-free factor loadings, we cannot reject the hypothesis that Nelson-Siegel factor loadings fulfill the no-arbitrage constraints, at a 95 percent confidence level. Furthermore, we show that the Nelson-Siegel model performs as well as the no-arbitrage counterpart in an out-of-sample forecasting experiment. Based on these empirical observations, we conclude that the Nelson-Siegel model is compatible with arbitrage-freeness.

This conclusion is of relevance to fixed-income money managers and central banks in particular, since such organizations traditionally rely heavily on the Nelson-Siegel model for policy and strategic investment decisions.

Table 3.1: Summary statistics of the US zero-coupon data

τ	mean	std dev	min	max	$\rho(1)$	$\rho(2)$	$\rho(3)$	$\rho(12)$
1	6.44	2.58	2.69	16.16	0.97*	0.93*	0.89*	0.69*
3	6.75	2.66	2.73	16.02	0.97*	0.94*	0.91*	0.71*
6	6.98	2.66	2.89	16.48	0.97*	0.94*	0.91*	0.73*
9	7.10	2.64	2.98	16.39	0.97*	0.94*	0.91*	0.73*
12	7.20	2.57	3.11	15.82	0.97*	0.94*	0.91*	0.74*
15	7.31	2.52	3.29	16.04	0.97*	0.94*	0.91*	0.75*
18	7.38	2.50	3.48	16.23	0.98*	0.94*	0.92*	0.75*
21	7.44	2.49	3.64	16.18	0.98*	0.95*	0.92*	0.76*
24	7.46	2.44	3.78	15.65	0.98*	0.94*	0.92*	0.75*
30	7.55	2.36	4.04	15.40	0.98*	0.95*	0.92*	0.76*
36	7.63	2.34	4.20	15.77	0.98*	0.95*	0.93*	0.77*
48	7.77	2.28	4.31	15.82	0.98*	0.95*	0.93*	0.78*
60	7.84	2.25	4.35	15.01	0.98*	0.96*	0.94*	0.79*
72	7.96	2.22	4.38	14.98	0.98*	0.96*	0.94*	0.80*
84	7.99	2.18	4.35	14.98	0.98*	0.96*	0.94*	0.78*
96	8.05	2.17	4.43	14.94	0.98*	0.96*	0.95*	0.81*
108	8.08	2.18	4.43	15.02	0.98*	0.96*	0.95*	0.81*
120	8.05	2.14	4.44	14.93	0.98*	0.96*	0.94*	0.78*

Descriptive statistics of monthly yields at different maturities, τ , for the sample from January 1970 to December 2000. $\rho(p)$ refers to the sample autocorrelation of the series at lag p and * denotes significance at 95 percent confidence level. Confidence intervals are computed according to Box and Jenkins (1976).

Table 3.2: Autocorrelations

Yield differences						
τ	$\rho(1)$	$\rho(3)$	$\rho(12)$	$\rho^2(1)$	$\rho^2(3)$	$\rho^2(12)$
1	0.06	-0.07	-0.06	0.23*	0.08	0.08
3	0.12*	-0.05	-0.13*	0.34*	0.07	0.22*
6	0.16*	-0.09	-0.08	0.32*	0.09	0.20*
12	0.15*	-0.10	-0.05	0.16*	0.11*	0.13*
24	0.18*	-0.11*	0.00	0.21*	0.13*	0.13*
36	0.14*	-0.11*	0.03	0.12*	0.14*	0.14*
60	0.13*	-0.07	0.03	0.09	0.13*	0.13*
84	0.10	-0.09	-0.03	0.17*	0.22*	0.18*
120	0.10	-0.05	-0.03	0.15*	0.19*	0.23*

Yield ratios						
τ	$\rho(1)$	$\rho(3)$	$\rho(12)$	$\rho^2(1)$	$\rho^2(3)$	$\rho^2(12)$
1	0.07	-0.05	0.10	0.23*	0.12*	0.02
3	0.11*	0.00	0.01	0.34*	0.10	0.16*
6	0.16*	0.00	0.04	0.25*	0.13*	0.13*
12	0.16*	-0.04	0.04	0.10	0.13*	0.07
24	0.16*	-0.07	0.03	0.06	0.12*	0.03
36	0.13*	-0.09	0.06	0.01	0.06	0.05
60	0.12*	-0.04	0.05	0.01	0.01	0.01
84	0.11*	-0.04	0.00	0.04	0.07	0.03
120	0.08	-0.03	0.00	0.03	0.06	0.06

Sample autocorrelations of first yield differences Δy , squared first yield differences Δy^2 , yield ratios $\frac{y_t}{y_{t-1}}$ and squared demeaned yield ratios $\left(\frac{y_t}{y_{t-1}} - \bar{\mu}\right)^2$, for selected maturities τ , at lags 1, 3 and 12. * denotes significance at 95 percent confidence level. Confidence intervals are computed according to Box and Jenkins (1976). $\rho(p)$ and $\rho^2(p)$ denote, respectively, the correlation of the variables and their squares, at lag p .

Table 3.3: Parameter estimates

Parameter	Estimated value	$Q_{2.5}$	$Q_{97.5}$
$\hat{\mu}_1$	-0.247	-1.170	0.911
$\hat{\mu}_2$	-0.006	-0.992	1.158
$\hat{\mu}_3$	-0.408	-1.164	0.895
$\hat{\Phi}_{11}$	0.991*	0.926	1.021
$\hat{\Phi}_{21}$	-0.031	-0.094	0.032
$\hat{\Phi}_{31}$	0.070	-0.102	0.154
$\hat{\Phi}_{12}$	0.024	-0.037	0.068
$\hat{\Phi}_{22}$	0.933*	0.888	1.013
$\hat{\Phi}_{32}$	0.036	-0.140	0.185
$\hat{\Phi}_{13}$	0.000	-0.035	0.062
$\hat{\Phi}_{23}$	0.038	-0.015	0.082
$\hat{\Phi}_{33}$	0.771*	0.755	0.975
$\hat{\Sigma}_{11}$	0.162*	0.086	0.306
$\hat{\Sigma}_{21}$	-0.051	-0.192	0.042
$\hat{\Sigma}_{31}$	-0.110	-0.302	0.014
$\hat{\Sigma}_{22}$	0.324*	0.067	0.305
$\hat{\Sigma}_{32}$	0.009	-0.170	0.071
$\hat{\Sigma}_{33}$	0.596*	0.150	0.532

Parameter estimates (continued)

Parameter	Estimated value	$Q_{2.5}$	$Q_{97.5}$
$\hat{\lambda}_{0,1}$	-0.215	-3.672	1.967
$\hat{\lambda}_{0,2}$	-0.354	-3.043	1.995
$\hat{\lambda}_{0,3}$	0.297	-2.390	3.053
$\hat{\lambda}_{1,11}$	-0.062	-0.470	1.262
$\hat{\lambda}_{1,21}$	-0.123	-0.799	0.523
$\hat{\lambda}_{1,31}$	0.124	-1.098	0.728
$\hat{\lambda}_{1,12}$	0.117	-2.734	1.051
$\hat{\lambda}_{1,22}$	-0.049	-0.633	1.343
$\hat{\lambda}_{1,32}$	0.150	-1.080	1.378
$\hat{\lambda}_{1,13}$	-0.187	-4.208	0.209
$\hat{\lambda}_{1,23}$	-0.169	-2.238	-0.019
$\hat{\lambda}_{1,33}$	-0.024	-0.399	3.209
\hat{a}_1^{NA}	0.537*	0.115	1.202
$\hat{b}_1^{NA}(1)$	0.168*	0.064	0.390
$\hat{b}_1^{NA}(2)$	0.146*	0.061	0.623
$\hat{b}_1^{NA}(3)$	0.000	-0.039	0.023

Estimated parameters from the no-arbitrage model in equations (3.2.6) to (3.2.12) with the 95 percent confidence intervals obtained by resampling. The confidence intervals $[Q_{2.5} Q_{97.5}]$ refer to the empirical 2.5 percent and 97.5 percent quantiles of the distributions of the parameters. A star * is used to indicate when a parameter estimate is significantly different from zero at a 95 percent level of confidence.

Table 3.4: Estimation results for a^{NA}

τ	a^{NS}	\hat{a}^{NA}	\tilde{a}_L^{NA}	\tilde{a}_U^{NA}
1	0.00	0.00	-0.10	0.05
3	0.00	0.00	-0.04	0.05
6	0.00	0.00	-0.02	0.06
9	0.00	0.01	-0.02	0.05
12	0.00	0.01	-0.02	0.05
15	0.00	0.00	-0.02	0.04
18	0.00	0.00	-0.02	0.03
21	0.00	0.00	-0.03	0.02
24	0.00	0.00	-0.04	0.01
30	0.00	0.00	-0.05	0.01
36	0.00	-0.01	-0.06	0.02
48	0.00	-0.01	-0.07	0.03
60	0.00	-0.01	-0.06	0.03
72	0.00	0.00	-0.04	0.03
84	0.00	0.00	-0.02	0.02
96	0.00	0.00	-0.01	0.04
108	0.00	0.01	-0.02	0.07
120	0.00	0.01	-0.04	0.10

Estimated intercepts from the no-arbitrage model \hat{a}^{NA} with the 95 percent confidence intervals obtained from the resampling $[\tilde{a}_L^{NA}, \tilde{a}_U^{NA}]$. The confidence intervals refer to the empirical 2.5 percent and 97.5 percent quantiles of the distribution of the parameters. The second column of the Table reports the Nelson-Siegel loadings.

Table 3.5: Estimation results for $b^{NA}(1)$

τ	$b^{NS}(1)$	$\hat{b}^{NA}(1)$	$\tilde{b}_L^{NA}(1)$	$\tilde{b}_U^{NA}(1)$
1	1.00	0.98	0.87	1.16
3	1.00	0.99	0.90	1.06
6	1.00	0.99	0.89	1.04
9	1.00	1.00	0.92	1.04
12	1.00	1.00	0.93	1.04
15	1.00	1.00	0.94	1.04
18	1.00	1.00	0.96	1.05
21	1.00	1.00	0.97	1.06
24	1.00	1.00	0.98	1.06
30	1.00	1.01	0.98	1.08
36	1.00	1.01	0.96	1.10
48	1.00	1.00	0.95	1.10
60	1.00	1.00	0.95	1.09
72	1.00	1.00	0.95	1.06
84	1.00	1.00	0.96	1.03
96	1.00	1.00	0.92	1.01
108	1.00	0.99	0.88	1.04
120	1.00	0.99	0.82	1.08

Estimated loadings of the level factor from the no-arbitrage model $\hat{b}^{NA}(1)$ with the 95 percent confidence intervals obtained from the resampling $[\tilde{b}_L^{NA}(1) \tilde{b}_U^{NA}(1)]$. The confidence intervals refer to the empirical 2.5 percent and 97.5 percent quantiles of the distribution of the parameters. The second column of the Table reports the Nelson-Siegel loadings on the level.

Table 3.6: Estimation results for $b^{NA}(2)$

τ	$b^{NS}(2)$	$\hat{b}^{NA}(2)$	$\tilde{b}_L^{NA}(2)$	$\tilde{b}_U^{NA}(2)$
1	0.97	0.93	0.83	1.08
3	0.91	0.89	0.83	0.98
6	0.84	0.83	0.77	0.92
9	0.77	0.77	0.71	0.84
12	0.71	0.72	0.66	0.76
15	0.66	0.66	0.62	0.70
18	0.61	0.62	0.57	0.64
21	0.56	0.57	0.52	0.59
24	0.53	0.53	0.48	0.56
30	0.46	0.46	0.40	0.50
36	0.41	0.41	0.35	0.45
48	0.32	0.32	0.27	0.38
60	0.27	0.26	0.23	0.32
72	0.23	0.22	0.20	0.26
84	0.19	0.19	0.18	0.22
96	0.17	0.17	0.15	0.21
108	0.15	0.15	0.11	0.20
120	0.14	0.13	0.07	0.19

Estimated loadings of the slope factor from the no-arbitrage model $\hat{b}^{NA}(2)$ with the 95 percent confidence intervals obtained from the resampling $[\tilde{b}_L^{NA}(2) \tilde{b}_U^{NA}(2)]$. The confidence intervals refer to the empirical 2.5 percent and 97.5 percent quantiles of the distribution of the parameters. The second column of the Table reports the Nelson-Siegel loadings on the slope.

Table 3.7: Estimation results for $b^{NA}(3)$

τ	$b^{NS}(3)$	$\hat{b}^{NA}(3)$	$\tilde{b}_L^{NA}(3)$	$\tilde{b}_U^{NA}(3)$
1	0.03	0.00	-0.10	0.06
3	0.08	0.10	0.05	0.18
6	0.14	0.19	0.13	0.26
9	0.19	0.24	0.17	0.27
12	0.23	0.26	0.21	0.28
15	0.25	0.27	0.23	0.29
18	0.27	0.28	0.24	0.30
21	0.29	0.28	0.23	0.30
24	0.29	0.27	0.24	0.30
30	0.30	0.26	0.23	0.31
36	0.29	0.25	0.23	0.31
48	0.27	0.23	0.22	0.29
60	0.24	0.21	0.20	0.27
72	0.21	0.20	0.19	0.23
84	0.19	0.19	0.18	0.22
96	0.17	0.19	0.16	0.21
108	0.15	0.18	0.13	0.21
120	0.14	0.18	0.11	0.21

Estimated loadings of the curvature factor from the no-arbitrage model $\hat{b}^{NA}(3)$ with the 95 percent confidence intervals obtained from the resampling $[\tilde{b}_L^{NA}(3) \tilde{b}_U^{NA}(3)]$. The confidence intervals refer to the empirical 2.5 percent and 97.5 percent quantiles of the distribution of the parameters. The second column of the Table reports the Nelson-Siegel loadings on the curvature.

Table 3.8: Summary statistics for the resampled parameters

Intercept \tilde{a}^{NA}				
τ	mean	st.dev.	skewness	kurtosis
3	0.00	0.02	0.11	9.66
12	0.01	0.02	-0.24	8.91
24	0.00	0.01	-3.11	18.77
60	-0.01	0.02	0.34	9.25
84	0.00	0.01	5.49	57.71
120	0.02	0.04	1.06	7.71

Loading of the level $\tilde{b}^{NA}(1)$				
τ	mean	st.dev.	skewness	kurtosis
3	0.99	0.04	0.28	9.39
12	0.99	0.03	0.76	9.02
24	1.01	0.02	2.85	17.25
60	1.01	0.04	-0.88	10.97
84	1.00	0.02	-5.66	60.42
120	0.97	0.06	-1.03	8.17

Loading of the slope $\tilde{b}^{NA}(2)$				
τ	mean	st.dev.	skewness	kurtosis
3	0.91	0.03	0.47	5.56
12	0.71	0.02	-0.08	3.45
24	0.53	0.02	-0.99	6.67
60	0.27	0.02	0.52	5.01
84	0.20	0.01	3.00	34.43
120	0.14	0.03	-0.10	3.97

Loading of the curvature $\tilde{b}^{NA}(3)$				
τ	mean	st.dev.	skewness	kurtosis
3	0.10	0.03	0.93	3.39
12	0.25	0.02	-0.52	4.59
24	0.28	0.02	-0.73	2.71
60	0.22	0.02	1.72	8.99
84	0.19	0.01	1.05	5.42
120	0.16	0.02	-0.85	6.80

Summary statistics of the empirical distributions of the estimated parameters obtained using resampled data.

Table 3.9: Measures of fit

Residuals from the Nelson-Siegel model

τ	mean	st dev	min	max	RMSE	MAD	$\rho(1)$	$\rho(6)$	$\rho(12)$
1	-0.159	0.200	-1.046	0.387	0.200	0.040	0.513	0.332	0.443
3	0.027	0.114	-0.496	0.584	0.114	0.013	0.274	0.159	0.326
6	0.091	0.135	-0.412	0.680	0.135	0.018	0.543	0.346	0.471
12	0.046	0.122	-0.279	0.483	0.122	0.015	0.586	0.127	0.289
24	-0.040	0.073	-0.398	0.261	0.073	0.005	0.493	0.044	0.153
36	-0.066	0.090	-0.432	0.339	0.089	0.008	0.417	0.256	0.183
60	-0.053	0.096	-0.520	0.292	0.096	0.009	0.655	0.312	-0.037
84	0.006	0.097	-0.446	0.337	0.096	0.009	0.518	0.159	-0.083
120	0.002	0.140	-0.763	0.436	0.140	0.020	0.699	0.345	0.091

Residuals from no-arbitrage model

τ	Mean	st dev	min	max	RMSE	MAD	$\rho(1)$	$\rho(6)$	$\rho(12)$
1	0.000	0.168	-0.730	0.752	0.168	0.028	0.361	0.197	0.363
3	0.080	0.132	-0.508	0.817	0.132	0.018	0.448	0.219	0.312
6	0.060	0.135	-0.295	0.795	0.134	0.018	0.579	0.361	0.432
12	-0.019	0.109	-0.355	0.439	0.109	0.012	0.514	0.147	0.306
24	-0.041	0.071	-0.323	0.217	0.071	0.005	0.491	0.134	0.096
36	-0.018	0.088	-0.286	0.405	0.088	0.008	0.474	0.320	0.263
60	0.004	0.100	-0.332	0.379	0.100	0.010	0.688	0.350	0.101
84	0.019	0.097	-0.479	0.343	0.097	0.009	0.527	0.157	-0.070
120	-0.060	0.144	-0.801	0.375	0.144	0.021	0.705	0.464	0.249

Summary statistics of residuals of the Nelson-Siegel and the no-arbitrage models. The Nelson-Siegel model is estimated according to equations (3.2.2) - (3.2.4). The no-arbitrage yield curve model is estimated according to equations (3.2.6) - (3.2.12). Statistics are shown for selected maturities, τ . RMSE is the root mean squared error and MAD is the mean absolute deviation. Autocorrelations are denoted by $\rho(p)$, where p is the lag.

Table 3.10: Out-of-sample performance

τ	1-m ahead		6-m ahead		12-m ahead	
	NS	NA	NS	NA	NS	NA
1	0.82	0.67	0.67	0.56	0.66	0.59
3	0.91	0.89	0.72	0.70	0.64	0.63
6	1.08	1.03	0.81	0.82	0.65	0.67
9	1.06	1.21	0.80	0.83	0.64	0.66
12	1.01	1.00	0.80	0.81	0.64	0.65
15	1.06	0.98	0.79	0.79	0.64	0.65
18	1.04	1.03	0.80	0.80	0.65	0.65
21	1.06	1.07	0.80	0.80	0.66	0.66
24	1.09	1.11	0.80	0.80	0.67	0.67
30	1.04	1.04	0.80	0.78	0.68	0.67
36	0.99	0.98	0.80	0.78	0.70	0.69
48	0.98	0.98	0.84	0.81	0.76	0.73
60	1.10	1.04	0.88	0.85	0.81	0.79
72	1.02	1.01	0.90	0.88	0.85	0.84
84	1.08	1.08	0.91	0.91	0.87	0.86
96	1.03	1.03	0.93	0.94	0.91	0.92
108	1.04	1.08	0.95	0.98	0.93	0.96
120	1.08	1.32	1.02	1.08	1.00	1.05

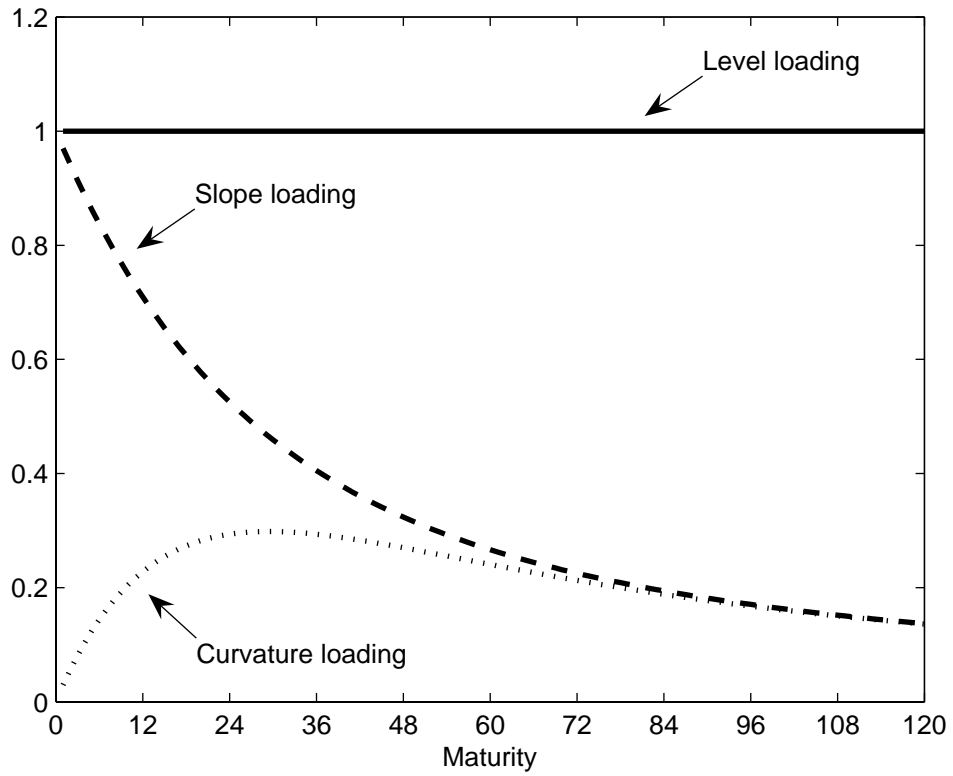
Ratios of the Mean Squared Forecast Error (MSFE) of the no-arbitrage model (NA) and the Nelson-Siegel model (NS) both measured against the performance of the random walk model. A ratio lower than 1 means that the MSFE for the respective model is lower than the forecast error generated by the random walk, and hence that the model performs better than the random walk model. The models are estimated on successively increasing data samples starting 1970:1 until the time the forecast is made, and expanded by one month each time a new set of forecasts are generated. Forecasts for horizons of 1, 6 and 12 months ahead are evaluated on the sample from 1994:1 to 2000:12. Bold entries in the table indicate superior performance of one model (NA or NS) against the other model.

Table 3.11: Diebold-Mariano test statistics

τ	1-m ahead	6-m ahead	12-m ahead
1	-0.080	-0.214	-0.250
3	-0.037	-0.129	-0.146
6	-0.051	0.132	0.262
9	0.147	0.159	0.222
12	-0.015	0.085	0.154
15	-0.117	0.021	0.098
18	-0.040	0.017	0.086
21	0.048	-0.025	0.046
24	0.070	-0.318	- 0.165
30	-0.003	-0.174	-0.290
36	-0.022	-0.149	-0.239
48	0.002	-0.128	- 0.215
60	-0.082	-0.153	-0.233
72	-0.025	-0.121	-0.215
84	-0.007	-0.047	- 0.166
96	-0.016	0.315	0.447
108	0.069	0.231	0.322
120	0.266	0.290	0.366

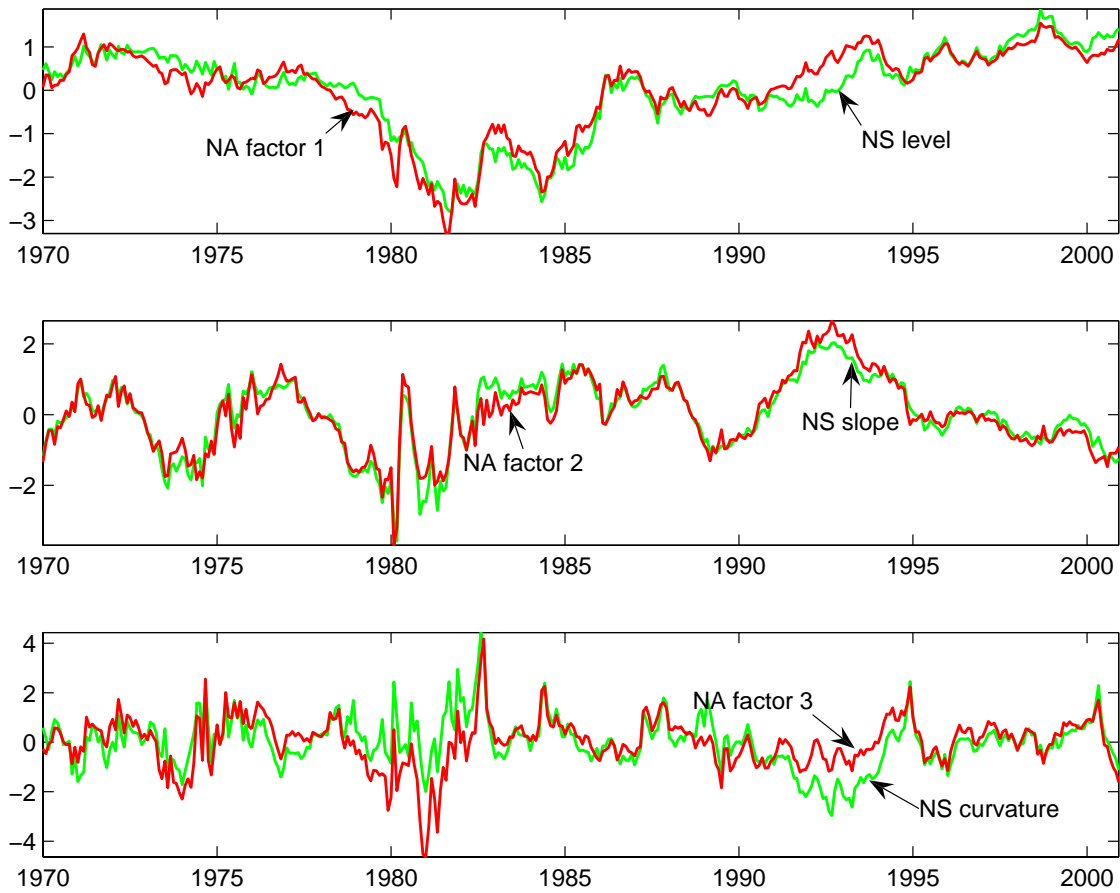
Diebold-Mariano test statistic to compare forecast accuracy of two models. We compare the no-arbitrage model against the Nelson-Siegel model. Negative numbers reflect superiority of the no-arbitrage model, and positive numbers indicate that the Nelson-Siegel model performs better. The null hypothesis is that the mean squared forecast error of the two models is identical. A number larger than 1.96 in absolute terms indicates that the forecasts produced by the models are significantly different at a 5 percent level.

Figure 3.1: Nelson-Siegel factor loadings



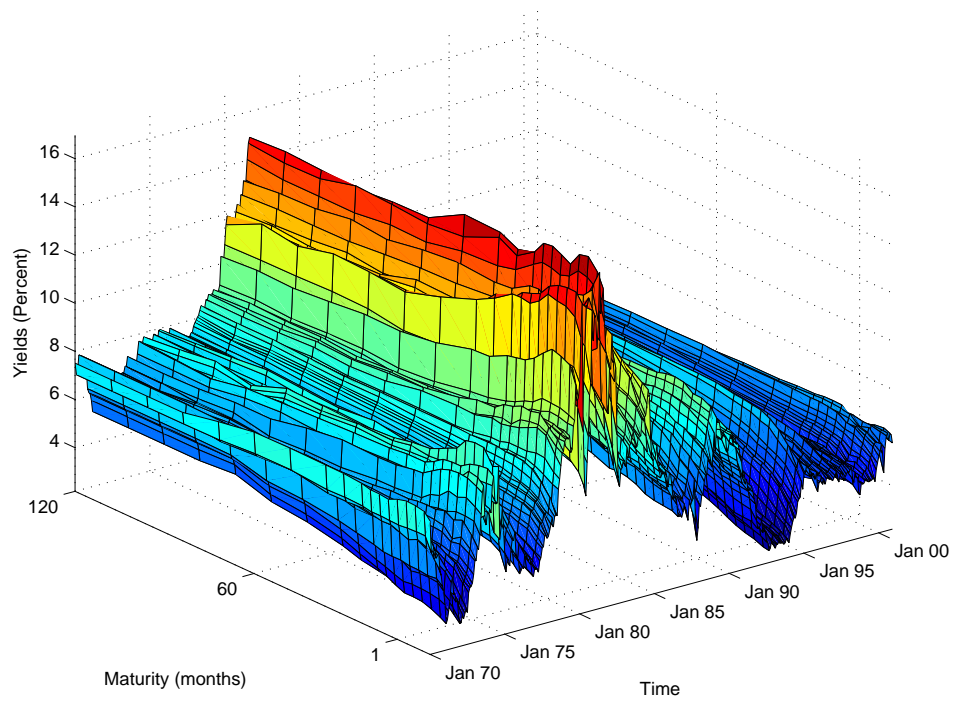
Nelson and Siegel (1987) factor loadings using the re-parameterized version of the model as presented by Diebold and Li (2006). The factor loadings b^{NS} are computed using $\lambda = 0.0609$ and equation (3.2.3).

Figure 3.2: No-Arbitrage Latent factors and Nelson and Siegel factors



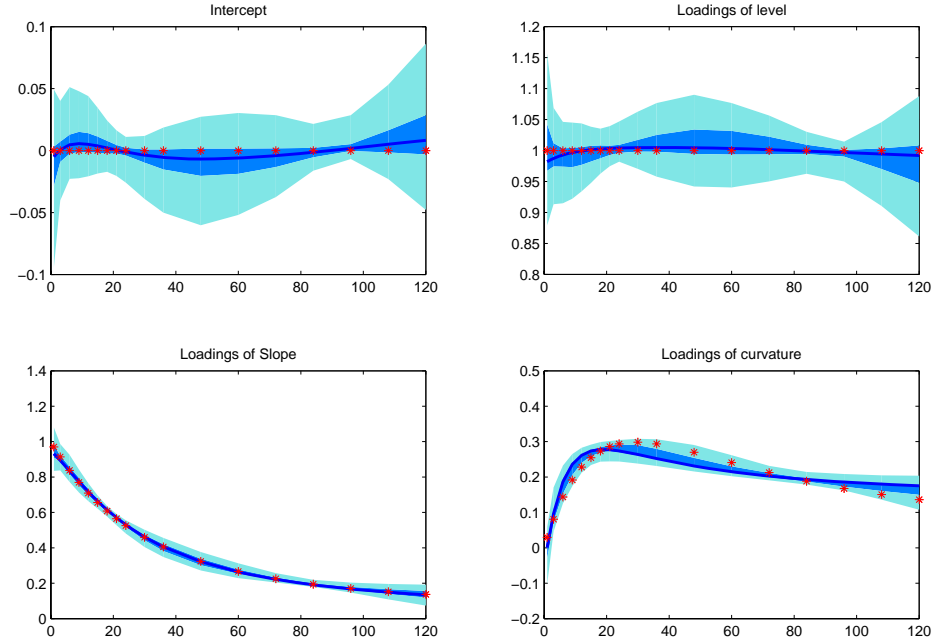
Extracted yield curve factors using US zero-coupon data observed at a monthly frequency and covering the period from 1970:1 to 2000:12. Factors are extracted from the Nelson-Siegel model and from the no-arbitrage model. “NS level” and “NA factor 1” refer to the first extracted factor from each model. The second and third extracted factors are correspondingly labeled “NS slope”, “NA factor 2” and “NS curvature”, “NA factor 3”.

Figure 3.3: Zero-coupon yields data



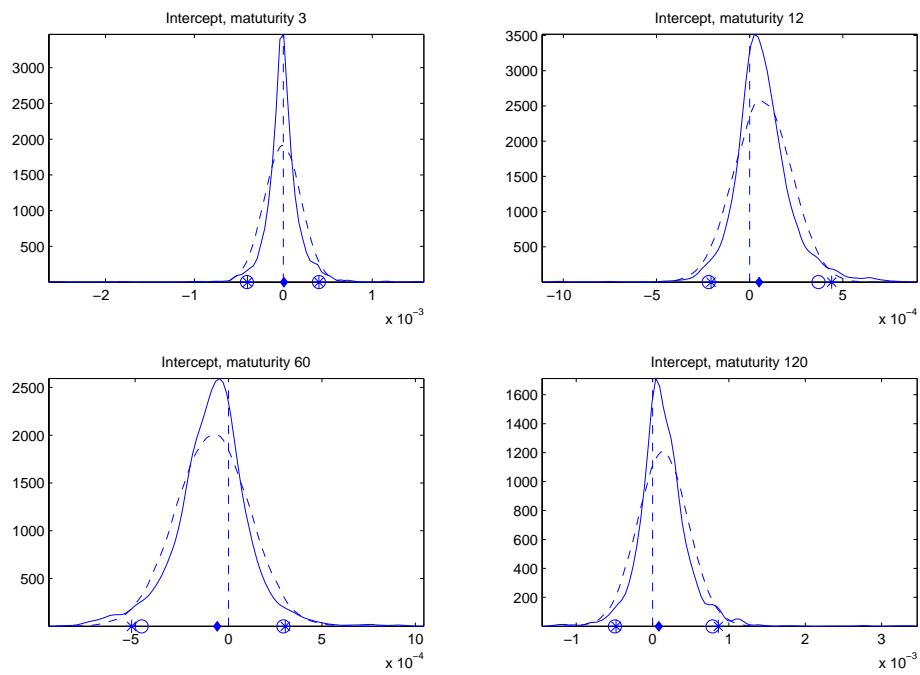
U.S. zero-coupon yield curve data observed at monthly frequency from 1970:1 to 2000:12 at maturities 1, 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108 and 120 months.

Figure 3.4: No-Arbitrage loadings of the Nelson and Siegel factors



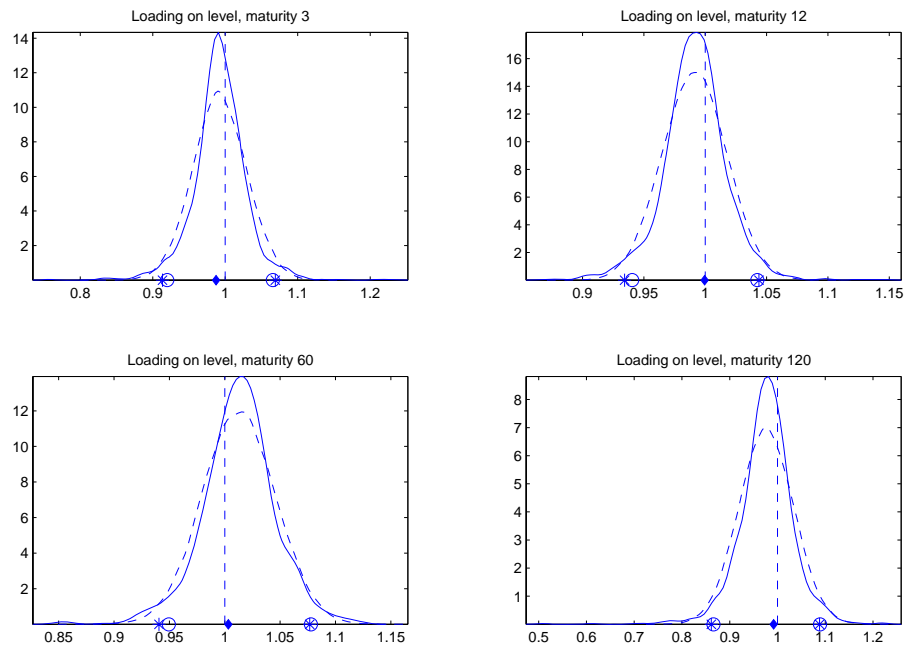
Estimated factor loadings and empirical 50 and 95 percent confidence intervals. Star * indicate the factor loadings from the Nelson-Siegel model, i.e. a^{NS} and b^{NS} in equations (3.2.2) and (3.2.3), while the continuous lines indicate the corresponding factor loadings estimated from the no-arbitrage model, i.e. \hat{a}^{NA} and \hat{b}^{NA} in equations (3.2.6) to (3.2.8). The distributions of the latter are obtained through resampling. The dark-shaded areas are the 50 percent confidence intervals, while the light-shaded areas show the 95 percent confidence intervals. These are computed as empirical quantiles.

Figure 3.5: Distribution of the estimated loadings for a^{NA}



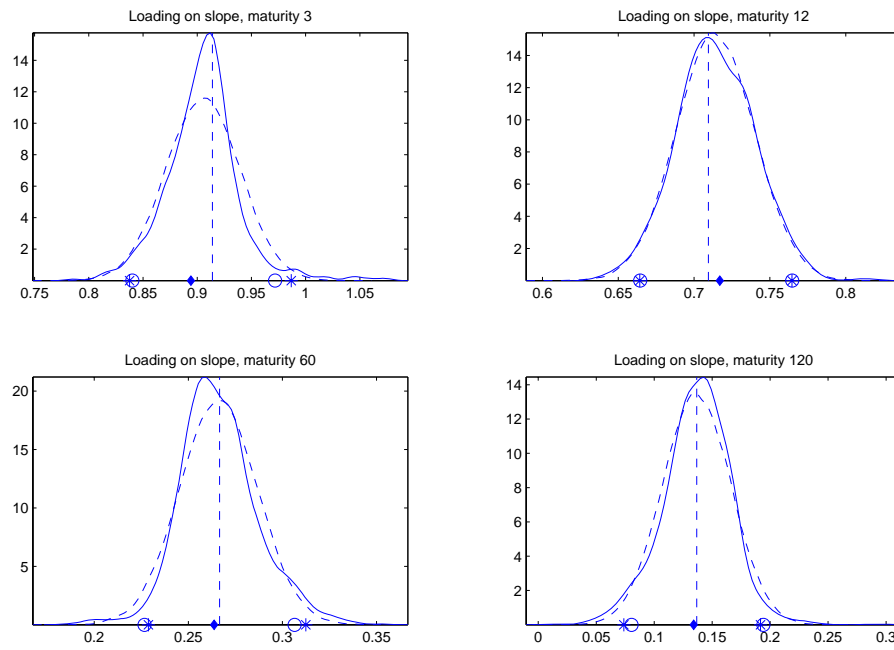
Empirical distributions, for selected maturities, of the no-arbitrage intercepts obtained from the resampling (continuous line), with the relative 95 percent confidence interval (asterisks). The dashed line is the Gaussian approximation with the relative 95 percent confidence intervals (circles). The diamonds are the estimated no-arbitrage intercepts and the dashed vertical line indicates the Nelson and Siegel intercepts.

Figure 3.6: Distribution of the estimated loadings for $b^{NA}(1)$



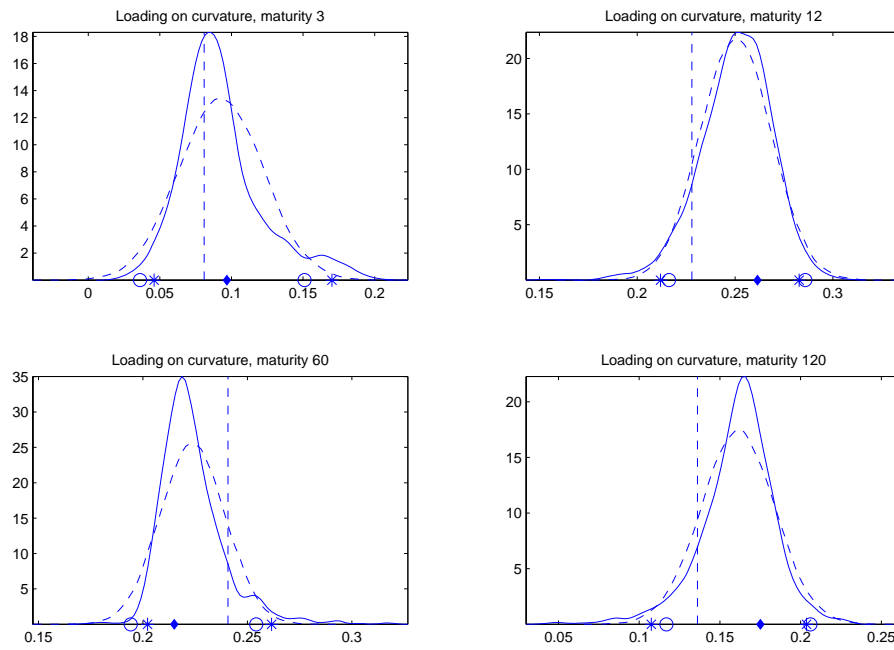
Empirical distributions, for selected maturities, of the no-arbitrage loadings of the level obtained from the re-sampling (continuous line), with the relative 95 percent confidence interval (asterisks). The dashed line is the Gaussian approximation with the relative 95 percent confidence intervals (circles). The diamonds are the estimated no-arbitrage loadings of the level and the dashed vertical line indicates the relative Nelson and Siegel loadings.

Figure 3.7: Distribution of the estimated loadings for $b^{NA}(2)$



Empirical distributions, for selected maturities, of the no-arbitrage loadings of the slope obtained from the re-sampling (continuous line), with the relative 95 percent confidence interval (asterisks). The dashed line is the Gaussian approximation with the relative 95 percent confidence intervals (circles). The diamonds are the estimated no-arbitrage loadings of the slope and the dashed vertical line indicates the relative Nelson and Siegel loadings.

Figure 3.8: Distribution of the estimated loadings for $b^{NA}(3)$



Empirical distributions, for selected maturities, of the no-arbitrage loadings of the curvature obtained from the re-sampling (continuous line), with the relative 95 percent confidence interval (stars). The dashed line is the Gaussian approximation with the relative 95 percent confidence intervals (circles). The diamonds are the estimated no-arbitrage loadings of the curvature and the dashed vertical line indicates the relative Nelson and Siegel loadings.

Bibliography

- Ahn, Dong-Hyun, Robert F. Dittmar, and Ronald A. Gallant (2002) ‘Quadratic Term Structure Models: Theory and Evidence.’ *Review of Financial Studies* 15(1), 243–288
- 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(4), 745–787
- Ang, A., G. Bekaert, and M. Wei (2007) ‘The term structure of real rates and expected inflation.’ *Journal of Finance*, *forthcoming*
- Ang, A., M. Piazzesi, and M. Wei (2006) ‘What Does the Yield Curve Tell us about GDP Growth.’ *Journal of Econometrics* 131, 359–403
- Bai, J., and S. Ng (2002) ‘Determining the Number of Factors in Approximate Factor Models.’ *Econometrica* 70(1), 191–221
- Bernadell, C., J. Coche, and K. Nyholm (2005) ‘Yield Curve Prediction for the Strategic Investor.’ ECB Working Paper 472
- Bernanke, B. S., and J. Boivin (2003) ‘Monetary policy in a data-rich environment.’ *Journal of Monetary Economics* 50(3), 525–546
- BIS (2005) *Zero-Coupon Yield Curves: Technical Documentation* (Bank for International Settlements, Basle)

- Bjork, T., and B.J. Christensen (1999) ‘Interest Rate Dynamics and Consistent Forward Rate Curves.’ *Mathematical Finance* 9, 323–348
- Black, F., and P. Karasinski (1993) ‘Bond and option pricing when the short rates are log-normal.’ *Financial Analyst Journal* 47, 52–59
- Black, F., E. Derman, and E. Toy (1990) ‘A one-factor model of interest rates and its application to treasury bond options.’ *Financial Analyst Journal* 46, 33–39
- Bofinger, E. (1975) ‘Estimation of a density function using order statistics.’ *Australian Journal of Statistics* 17(1), 1–7
- Bouye, E., and M. Salmon (2002) ‘Dynamic Copula Quantile Regression and Tail Area Dynamic Dependence in Forex Markets.’ *Manuscript, Financial Econometrics Research Center*
- Box, G.E.P., and G.M. Jenkins (1976) *Time Series Analysis: Forecasting and Control* (San Francisco: Holden Day)
- Chen, R.R., and L. Scott (1993) ‘Maximum likelihood estimation for a multi-factor equilibrium model of the term structure of interest rates.’ *Journal of Fixed Income* 3, 14–31
- Christensen, Jens, Francis Diebold, and Glenn Rudebusch (2007) ‘The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models.’ FRB of San Francisco Working Paper No. 2007-20
- Christoffersen, P.F. (1998) ‘Evaluating Interval Forecasts.’ *International Economic Review* 39(4), 841–862
- Cox, J.C., J.E. Ingersoll, and S.A. Ross (1985) ‘A theory of the term structure of interest rates.’ *Econometrica* 53, 385–407
- Dai, Q., and K. Singleton (2000) ‘Specification analysis of affine term structure models.’ *Journal of Finance* 55, 1943–1978

- Dai, Q., and T. Philippon (2005) ‘Fiscal Policy and the Term Structure of Interest Rates.’ NBER working paper
- De Pooter, M., F. Ravazzolo, and D. van Dijk (2007) ‘Predicting the Term Structure of Interest Rates: Incorporating Parameter Uncertainty and Macroeconomic Information.’ *Tinbergen Institute Discussion Papers*
- Dewachter, H., and M. Lyrio (2006) ‘Macro Factors and the Term Structure of Interest Rates.’ *Journal of Money, Credit, and Banking* 38(1), 119–140
- Diebold, F. X., and C. Li (2006) ‘Forecasting the term structure of government bond yields.’ *Journal of Econometrics* 130, 337–364
- Diebold, F.X., G. D. Rudebusch, and S. B. Aruoba (2006) ‘The macroeconomy and the yield curve: A dynamic latent factor approach.’ *Journal of Econometrics* 131, 309–338
- Diebold, F.X., L. Ji, and C. Li (2004) ‘A Three-Factor Yield Curve Model: Non-Affine Structure, Systematic Risk Sources, and Generalized Duration.’ Working paper, University of Pennsylvania
- Diebold, F.X., M. Piazzesi, and G. D. Rudebusch (2005) ‘Modeling Bond Yields in Finance and Macroeconomics.’ *American Economic Review* 95, 415–420
- Dionne, G., P. Duchesne, M. Pacurar, and Montréal (2005) *Intraday Value at Risk (IVaR) Using Tick-by-tick Data with Application to the Toronto Stock Exchange* (HEC Montréal, Centre de recherche en e-finance)
- Doz, C., D. Giannone, and L. Reichlin (2006) *A Quasi Maximum Likelihood Approach for Large Approximate Dynamic Factor Models* (Centre for Economic Policy Research)
- Duffee, G.R. (2002) ‘Term premia and interest rate forecasts in affine models.’ *Journal of Finance* 57, 405–443

- Duffie, D., and R. Kan (1996) ‘A yield-factor model of interest rates.’ *Mathematical Finance* 6, 379–406
- Engle, R.F., and S. Manganelli (2004) ‘CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles.’ *Journal of Business & Economic Statistics* 22(4), 367–382
- Fama, E.F., and J.D. MacBeth (1973) ‘Risk, Return, and Equilibrium: Empirical Tests.’ *The Journal of Political Economy* 81(3), 607–636
- Fama, E.F., and R.R. Bliss (1987) ‘The Information in Long-Maturity Forward Rates.’ *The American Economic Review* 77(4), 680–692
- Favero, C.A., L. Niu, and L. Sala (2007) ‘Term Structure Forecasting: No-Arbitrage Restrictions vs. Large Information Set.’ *CEPR Discussion Paper*
- Fleming, M.J., E.M. Remolona, Monetary, Economic Dept, and Bank for International Settlements (1999) *The Term Structure of Announcement Effects* (Bank for International Settlements, Monetary and Economic Dept.)
- Furfine, C. (2001) ‘Do macroeconomic announcements still drive the Treasury market.’ *BIS Quarterly Review* pp. 49–57
- Giannone, D., L. Reichlin, and L. Sala (2005) ‘Monetary Policy in Real Time.’ *Nber Macroeconomics Annual 2004*
- Giot, P. (2005) ‘Market risk models for intraday data.’ *The European Journal of Finance* 11(4), 309–324
- Gonedes, N.J. (1973) ‘Evidence on the Information Content of Accounting Numbers: Accounting-Based and Market-Based Estimates of Systematic Risk.’ *The Journal of Financial and Quantitative Analysis* 8(3), 407–443
- Hansen, B.E. (1994) ‘Autoregressive Conditional Density Estimation.’ *International Economic Review* 35(3), 705–730

- Harvey, C.R., and A. Siddique (1999) ‘Autoregressive Conditional Skewness.’ *The Journal of Financial and Quantitative Analysis* 34(4), 465–487
- Hendricks, W., and R. Koenker (1992) ‘Hierarchical Spline Models for Conditional Quantiles and the Demand for Electricity.’ *Journal of the American Statistical Association*
- Hördahl, P., O. Tristani, and D. Vestin (2006) ‘A joint econometric model of macroeconomic and term-structure dynamics.’ *Journal of Econometrics* 131(1-2), 405–444
- Jasiak, J, and C. Gourieroux (2006) ‘Dynamic quantile models.’ Working Papers 2006 (4), York University
- Koenker, R. (2005) *Quantile Regression* (Cambridge University Press)
- Koenker, R., and G. Bassett Jr (1978) ‘Regression Quantiles.’ *Econometrica* 46(1), 33–50
- Kozicki, S., and PA Tinsley (2001) ‘Shifting endpoints in the term structure of interest rates.’ *Journal of Monetary Economics* 47(3), 613–652
- Kupiec, P. (1995) ‘Techniques for verifying the accuracy of risk management models.’ *Journal of Derivatives* 3(2), 73–84
- Law, P. (2006) ‘Macro factors and the yield curve.’ PhD dissertation, STANFORD UNIVERSITY
- Litterman, R., and J. Scheinkman (1991) ‘Common factors affecting bond returns.’ *Journal of Fixed Income* 47, 129–1282
- Ludvigson, S.C., and S. Ng (2005) ‘Macro Factors in Bond Risk Premia.’ *NBER Working Paper*
- Merton, R. C. (1973) ‘Theory of rational option pricing.’ *Bell Journal of Economics and Management Science* 4, 141–183

- Mönch, E. (2005) 'Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach.' ECB Working Paper series No. 544
- (2006) 'Term structure surprises: The predictive content of curvature, level, and slope.' Working Paper
- Nelson, C.R., and A.F. Siegel (1987) 'Parsimonious modeling of yield curves.' *Journal of Business* 60, 473–89
- Portnoy, S., and R. Koenker (1997) 'The Gaussian Hare and the Laplacian Tortoise: Computability of Squared-Error versus Absolute-Error Estimators.' *Statistical Science* 12(4), 279–296
- Rebonato, R., S. Mahal, M.S. Joshi, L. Bucholz, and K. Nyholm (2005) 'Evolving Yield Curves in the Real-World Measures: A Semi-Parametric Approach.' *Journal of Risk* 7, 29–62
- Rudebusch, G.D., and T. Wu (2004) 'A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy.' *Federal Reserve Bank of San Francisco Working Paper*
- Soderlind, P., and L.O.E. Svensson (1997) 'New Techniques to Extract Market Expectations from Financial Instruments.' *Journal of Monetary Economics* 40, 383–429
- Vasicek, O. (1977) 'An equilibrium characterization of the term structure.' *Journal of Financial Economics* 5, 177–188
- Wu, T. (2006) 'Macro Factors and the Affine Term Structure of Interest Rates.' *Journal of Money, Credit, and Banking* 38(7), 1847–1875