EUROPEAN CENTRAL BANK
EUROSYSTEM

WORKING PAPER SERIES
NO 1205 / JUNE 2010

NELSON-SIEGEL AFFINE AND QUADRATIC YIELD CURVE SPECIFICATIONS

WHICH ONE IS BETTER AT FORECASTING?

by Ken Nyholm
and Rositsa Vidova-Koleva

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1205 / JUNE 2010

NELSON-SIEGEL, AFFINE AND QUADRATIC YIELD CURVE SPECIFICATIONS

WHICH ONE IS BETTER AT FORECASTING?[1]

by Ken Nyholm[2]

and Rositsa Vidova-Koleva

In 2010 all ECB publications feature a motif taken from the €500 banknote.

1 We would like to thank the referee for providing helpful comments. Needless to say, any remaining errors are our own. The majority of the work on this paper was completed while Rositsa Vidova-Koleva was visiting the Risk Management Division of the European Central Bank.

2 Risk Management Division, European Central Bank, Kaiserstrasse 29, Frankfurt am Main, Germany, phone: +49 69 1344 7926, e-mail: Ken.Nyholm@ecb.int

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract 4
Non-technical summary 5
1 Introduction 6
2 Review of related literature 9
3 Discrete term structure models 13
4 Estimation results 17
4.1 In-sample fit 21
4.2 Out-of-sample fit 23
5 Conclusion 27
Appendix 28
References 32
Tables and figures 34

Abstract

In this paper we compare the in-sample fit and out-of-sample forecasting performance of no-arbitrage quadratic and essentially affine term structure models, as well as the dynamic Nelson-Siegel model. In total eleven model variants are evaluated, comprising five quadratic, four affine and two Nelson-Siegel models. Recursive re-estimation and out-of-sample one-, six- and twelve-months ahead forecasts are generated and evaluated using monthly US data for yields observed at maturities of 1, 6, 12, 24, 60 and 120 months. Our results indicate that quadratic models provide the best in-sample fit, while the best out-of-sample performance is generated by three-factor affine models and the dynamic Nelson-Siegel model variants. However, statistical tests fail to identify one single-best forecasting model class.

JEL classification codes: C14, C15, G12

Keywords *Nelson-Siegel model; affine term structure models; quadratic yield curve models; forecast performance*

Non technical summary

This paper presents an extensive comparative study of the forecasting performance of three main yield curve model classes, namely the affine, quadratic and the dynamic Nelson-Siegel models.

The affine and dynamic Nelson-Siegel model specifications have been investigated extensively in the literature, while relatively little attention has been paid to the quadratic class. In the current paper we strive to close this gap by conducting an extensive out-of-sample comparison of all three model classes. To this end we rely on US yield curve data covering the period from January 1970 to December 2000. Recursive re-estimations and out-of-sample forecasts are generated for each model at forecasting horizons of one, six and twelve months starting in January 1994 and ending in December 2000.

Our empirical results indicate that better in-sample fit is provided by the quadratic model variants. Out-of-sample forecasts for the tested models are compared to the random-walk forecasts, and here results indicate that all tested model specifications tend to perform better than random-walk forecasts. Judged only by the size of the mean squared forecast errors, we find that Nelson-Siegel and affine models perform better than their quadratic counterparts, while this conclusion is somewhat weaker when actual statistical tests are performed.

1 Introduction

In this paper we compare the forecasting performance of the three main classes of term structure models advocated by the financial literature: the affine models, originally introduced by Duffie and Kan (1996), classified by Dai and Singleton (2000) and extended to the 'essentially' affine specification by Duffee (2002); the class of quadratic yield curve models classified by Ahn, Dittmar and Gallant (2002) and Leippold and Wu (2002); and the dynamic Nelson-Siegel model, introduced by Diebold and Li (2006) and Diebold, Ji and Li (2006), which builds on Nelson and Siegel (1987). While the forecasting performance of the affine and dynamic Nelson-Siegel models have been investigated extensively in the literature, relatively little attention has been paid to the quadratic class of yield curve models. In the current paper we strive to close this gap by conducting an extensive out-of-sample comparison of all three model classes. To this end we rely on US yield curve data covering the period from January 1970 to December 2000. Recursive re-estimations and out-of-sample forecasts are generated for each model at forecasting horizons of one, six and twelve months starting in January 1994 and ending in December 2000. Our results indicate that quadratic models provide the best in-sample fit, while the best out-of-sample performance is generated by three-factor affine models and the dynamic Nelson-Siegel models. However, statistical tests fail to identify one single-best forecasting model class.

One reason for the scarce number of studies dealing with the forecasting performance of the quadratic class of yield curve models is probably that it is an arduous task to estimate such models. At least when compared to the estimation of dynamic Nelson-Siegel models (DNSMs), the quadratic yield curve models pose a serious econometric challenge. It is well-know that it can be difficult to obtain convergence of the likelihood function in affine models; it is equally, or more, challenging to reach convergence in quadratic models given their richer parametric structure.

As suggested by their name, in quadratic term structure models (QTSMs), the yield curve factors enter *quadratically* in the observation equation, and as such, estimation by the regular Kalman filter technique is invalidated. Instead, estimation can be carried out by an extended Kalman filter or by other non-linear techniques. Our estimation approach relies on the unscented Kalman filter (UKF), developed by Julier and Uhlmann (1997).

Estimation time and model complexity is of interest in academic research because it determines how long the researcher has to wait before the results are available. Especially in a study like ours, which hinges on multiple rounds of model re-estimation, the used computer-time is considerable.[1] While estimation-time and model complexity is a tedious fact of life for an academic researcher, it is of dire importance for a practitioner. On the one hand, it is not feasible to use a model, which requires several days of estimation time, if the results produced by the model are needed on a shorter frequency e.g. daily. Similarly, if parameter estimates vary 'too much' from re-estimation to re-estimation, or if convergence of the model is not obtained

[1] We re-estimate the included model variants (five quadratic, four affine, and two Nelson-Siegel specifications), and perform out-of-sample forecasts for ninety six periods for each model. Even when executing calculations on a high-speed computing network with multiple parallel sessions, it takes a considerable amount of time to obtain the final results.

easily, decision makers using the output of the model probably would (and should) be skeptical about how much reliance they can attach to conclusions drawn on the basis of such a model. On the other hand, it is also necessary to evaluate the added benefit to a decision making process, which a more complex model may bring. For example, quadratic yield curve models may be regarded ideal in a setting where long-term yield curve forecasts are required, because, due to the way they are specified, QTSMs facilitate easy incorporation of restrictions ensuring that simulated yields remain in the positive quadrant. To obtain non-negative yield simulations by construction, is, for example, much more difficult if one relies on affine or Nelson-Siegel type yield curve models.[2]

The main contribution of our paper lies in the systematic comparison of the forecasting performance of quadratic, affine and Nelson-Siegel yield curve model specifications. Using US data from January 1970 to December 2000 we estimate and evaluate the performance of five quadratic models (three three-factor and two two-factor models), four 'essentially' affine models (two three-factor and two two-factor models), and two dynamic Nelson-Siegel three-factor models. In addition, within each yield curve model class, we also introduce variations with respect to how parsimonious the specifications are. In the quadratic model class we include: a maximally flexible model, which has the largest possible number of parameters to be estimated, allowing for interactions among the factors governing yield curve's dynamics; a medium flexible model, which imposes zero restrictions on some of the parameters but still allowing for factors' interactions; and finally a minimal specification, which is the most parsimonious specification included and which imposes

[2] In the current paper, however, no attention is paid to these practical aspects of yield curve models: only the pure in-sample and out-of-sample forecast precision is assumed to be relevant.

independence on the yield factors. The same model variants (except for the medium flexible one) are estimated for the affine and the Nelson-Siegel model classes.

Our empirical results indicate that better in-sample fit is provided by the QTSMs. Out-of-sample forecasts for the tested models are compared to the random-walk forecasts, and here our results indicate that all tested model specifications tend to perform better than the random walk. Simply judged by the size of the mean squared forecast errors, we find that Nelson-Siegel and affine models perform better than their quadratic counterparts, while this conclusion is somewhat weaker when actual statistical tests are performed. Our results do not declare a clear winner among three-factor quadratic, affine and Nelson-Siegel models.

2 Review of Related Literature

The Nelson-Siegel model, first presented in Nelson and Siegel (1987), provides an intuitive description of the yield curve at each point in time. In contrast to the no-arbitrage term structure models, this model class is derived in an ad-hoc manner and does not, theoretically, preclude arbitrage opportunities. However, extensions of the Nelson-Siegel model that are arbitrage-free do exist, see Bjork and Christensen (1999), and Christensen, Diebold and Rudebusch (2008a) and (2008b). The model is easy to estimate and fits yield curve data well in-sample. Set in a dynamic context the model has proven to produce good out-of-sample forecasts, see among others Diebold and Li (2006) and Diebold et al. (2006). For example, using US data from 1994 to 2000 Diebold and Li (2006) show that the dynamic Nelson-Siegel model performs better out-of-sample than the random-walk, and a large number of time-series models on yields as well as slope-regression models.

Coroneo, Nyholm and Vidova-Koleva (2008) show that the DNSM is not statistically significantly different from the arbitrage-free ATSM. Although forecast performance is not the primary objective of interest in that paper, as a secondary objective, they demonstrate that the DNSM produces forecasts that are as good as a Gaussian affine arbitrage-free model on US data covering the period from 1994 to 2000.

Arbitrage-free models, which include the affine and the quadratic specifications considered here, derive the dynamics of the yield curve under a risk-neutral probability measure. The existence of the risk-neutral measure implies that bond prices are arbitrage-free. The observed yield curve evolution is a result of the yields behaviour under a data-generating (historical or physical) measure. The transition from the risk-neutral to the physical measure is established via a function called market price of risk. It determines the risk premium on bonds' returns.[3]

Affine arbitrage-free term structure models, as characterized by Duffie and Kan (1996), have been extensively studied in the financial literature both with respect to their theoretical underpinnings as well as to their predictive abilities.[4] Dai and Singleton (2000) provide the admissibility conditions and suggest a classification scheme for 'completely' affine term structure models. As noted by Duffee (2002), the 'completely' affine modeling scheme is restrictive in terms of the allowed functional form used to characterize the market price of risk, and as a result hereof, fails to match important features of observed yield curves. Duffee (2002) presents a broader class of affine models, which he terms 'essentially' affine, and where the market price of risk specification is more flexibly formulated. Using US

[3]In this respect the Nelson-Siegel model is not arbitrage-free and does not account for risk pricing.

[4]Excellent surveys of the literature on affine yield curve models is offered by Piazzesi (2004) and Dai and Singleton (2003).

data from 1952 to 1994 as estimation period, and data from 1995 to 1998 as out-of-sample evaluation period, he shows that the 'essentially' affine model specification performs better than its 'completely' affine counterpart in terms of out-of-sample forecasting ability for yields measured at 6, 24 and 120 month maturities, when evaluated at forecasting horizons of 3, 6 and 12 months.[5]

In contrast to the numerous studies published on affine term structure models, their in-sample fit and out-of-sample performance, only few studies are concerned with the empirical performance of quadratic models. Examples of such studies comprise Ahn et al. (2002), Leippold and Wu (2002), (2003) and (2007), Realdon (2006), Kim (2004) and Brandt and Chapman (2003). Quadratic models claim to remedy some of the deficiencies that pertain to the affine model class. For example, it is straightforward to guarantee positive interest rates in the quadratic specification, something that is not easily achieved by the class of affine models. Also, better in-sample fit is claimed by quadratic models due to the allowed more flexible interaction between yield curve factors. For example, Ahn et al. (2002) and Leippold and Wu (2002) point out that affine models imply a certain trade-off between modeling heteroscedasticity of yields and negative correlation between yield curve factors. This relationship is relaxed in quadratic models. Similarly, in quadratic models one can encompass both time-varying risk premia and

[5] Cheridito, Filipović and Kimmel (2007) relax further the affine modeling restrictions by proposing an 'extended' specification for the market price of risk, which smooths the tension between matching the time-series behavior of yields and their cross-sectional relationship at a given point in time, i.e. the yield curve's location and shape. This is achieved by specifying a more general market price of risk that allows the parameters governing the time-series behaviour of yields (under the objective measure) to differ substantially from the parameters governing the cross-sectional fit of the yield curve (under the risk-neutral measure). While no out-of-sample forecasting comparison is conducted, the paper shows that the suggested extension of the affine modeling framework improves the in-sample fit, using US zero-coupon bond prices data covering the period from January 1972 to December 2002.

conditional heteroscedasticity of yields, which is not possible in the affine class of yield curve models, see e.g. Kim (2004), Dai and Singleton (2002) and Duffee (2002).

Ahn et al. (2002) describe the classification and canonical representation of QTSMs analogously to the classification of affine models in Dai and Singleton (2000). They show that the quadratic model specification can capture the conditional volatility of yields better than the affine class. In addition they show that the projected yields derived from the quadratic models are closer to observed yields, when compared to similar projections made from an affine model.

While the main purpose of Kim (2004) is to investigate whether there is a trade-off between volatility modeling and risk-premia modeling in quadratic-Gaussian term structure models similar to the existing one in affine (pure-Gaussian) models, he also performs an out-of-sample experiment comparing the forecasting ability of the quadratic and affine models that he investigates. The author uses a factor-augmented version of the Kalman filter, where the state space is augmented by the squared state variables, to estimate three quadratic-Gaussian term structure specifications. In an in-sample analysis, the quadratic specifications, due to their flexibility, are able to capture better different features of the data compared to their affine counterpart. In an out-of-sample forecasting exercise covering two data periods, one from 1993 to 1995, and another from 1996 to 1998, he compares the performance of these quadratic models to the pure-Gaussian term structure model in predicting the 6-, 12-, 24-, 60- and 120-month maturities on a 3 month and 1 year horizon. While on the first forecasting period results are somewhat mixed and one of the quadratic specifications uniformly outperforms the affine specification but only on the shorter forecasting horizon, for the second evaluation period, the only clear conclusion seems to be that the random walk performs better than the affine and the three quadratic models. Overall, Kim (2004) concludes that there is not a clear winning model in terms of forecasting. Our conclusions are similar to his comparing different model specifications of the quadratic and affine classes.

The main point of Kim (2005) is to show that there is evidence of nonlinearity in the term structure of yields and that in general nonlinear models perform better that affine models. In an out-of-sample experiment comparing ATSMs and QTSMs, using zero-coupon bond yields of 3-, 6-, 12-, 60- and 120-month maturities from 1959 to 1995 to estimate the models, and data from 1996 to 1999 to produce the forecasts, he finds that the quadratic model generates smaller root mean squared forecast errors than the affine model class.

3 Discrete Term Structure Models

The model specifications we consider are set in a state-space framework. The observed yields are assumed to depend on several unobserved factors. A certain dynamic process is hypothesized for the evolution of the underlying yield curve factors in the state equation. The translation of these factors into a yield curve at each point in time, is obtained via an observation equation. The functional form of the observation equation is dictated by the specific yield curve model under investigation. In affine models this 'translation' of factors into yields is achieved through a linear function in the yield curve factors. Quadratic models, in addition to the linear term also include a term which is quadratic in the yield curve factors. Arbitrage-free versions of affine and quadratic models impose additional constraints on the functional relation between the coefficients of the observation equation and the parameters that govern the law of motion of the yield curve factors. Since the entries in the factor-loading matrices in the observation equation depend on maturity, the imposed structure (the no-arbitrage constraints) ensures that the model is internally consistent, i.e. that the dynamic evolution of the factors driving yield curve changes over time is appropriately reflected in the shape and the location of the yield curve observed at any time. As is clear from below, the class of dynamic Nelson-Siegel models does not, by construction, impose such a no-arbitrage consistency and is in this sense not arbitrage-free.

In line with Dai, Le and Singleton (2006) we formulate our modeling framework in discrete time. The affine model can be seen as a restricted version of the quadratic one where the parameters corresponding to the quadratic term in the observation equation are equal to zero. The dynamic Nelson-Siegel model, although keeping the linear functional form of the observation equation, differs from the affine model by imposing a different (ad-hoc) structure on the functional form of the yield curve factor loadings in the observation equation, that does not conform with the no-arbitrage restrictions. Below we rely on this relationship between the three classes of yield curve models under investigation. First, we show the observation equation for the quadratic model and then we impose the appropriate restrictions that allow us to obtain the affine and the dynamic Nelson-Siegel models from the specification of the quadratic model. In the Appendix we show the formal derivation of the discrete-time version of the quadratic yield curve model. For more details on discrete QTSMs, see Realdon (2006), and on quadratic models in continuous time, see Ahn et al. (2002) and Leippold and Wu (2002). For a detailed derivation of the discrete ATSM see for example Ang and Piazzesi (2003) and Ang, Piazzesi and Wei (2006).

It is assumed that observed yields at time t, for the relevant maturities τ, are a function of a vector of N unobservable state variables labeled by X_t. It is further assumed that the vector X_t follows a first order Gaussian VAR process under the objective measure:

$$X_{t+1} = \mu + \Phi X_t + \Sigma \eta_{t+1}, \tag{1}$$

where η_{t+1} is an $N \times 1$ vector of $i.i.d.$ $N(0, I)$ errors, Φ is an $N \times N$ autoregressive matrix, μ is an $N \times 1$ vector and Σ is an $N \times N$ matrix.
It is also assumed that the short rate is a quadratic function of the factors:

$$r(X_t) = c_r + b_r' X_t + X_t' A_r X_t, \tag{2}$$

where A_r is an $N \times N$ matrix, b_r is an $N \times 1$ vector and c_r is a constant.

The market price of risk is assumed to be a linear function of the state variables:

$$\Lambda(X_t) = \lambda_0 + \lambda_1 X_t, \tag{3}$$

with λ_0 being an $N \times 1$ vector and λ_1 - a matrix of dimension $N \times N$. This representation of the market price of risk is in the spirit of Duffee (2002) i.e. it is 'essentially' affine.

The observed zero-coupon bond yield $Y(X_t, \tau)$, at time t for maturity τ, is then written as a quadratic function of the state variables,

$$Y(X_t, \tau) = -\frac{c_\tau}{\tau} - \frac{b_\tau'}{\tau} X_t - X_t' \frac{A_\tau}{\tau} X_t + \epsilon_{t,\tau}, \tag{4}$$

where $\epsilon_{t,\tau} \sim N(0, R)$ and R is assumed to be a diagonal matrix, i.e. it is assumed that the observation errors are not correlated across maturities ($cov(\epsilon_{t,\tau_i}, \epsilon_{t,\tau_j}) = 0$, $\tau_i \neq \tau_j$ for all i, j) and also across time ($cov(\epsilon_{t,\tau}, \epsilon_{s,\tau}) = 0$

for $t \neq s$). For a given maturity τ, c_τ is a constant, b_τ is an $N \times 1$ vector and A_τ is an $N \times N$ matrix. These are found as solutions to the recursive difference equations shown below

$$A_\tau = -A_r + (\Phi - \Sigma\lambda_1)' \left[A_{\tau-1} + 2\, A_{\tau-1}\, \Psi^{-1} A_{\tau-1}\right] (\Phi - \Sigma\lambda_1), \tag{5}$$

$$\begin{aligned} b_\tau = &- b_r + (\Phi - \Sigma\lambda_1)' b_{\tau-1} + 2\,(\Phi - \Sigma\lambda_1)' A_{\tau-1}\, \Psi^{-1} b_{\tau-1} + \\ &2\,(\Phi - \Sigma\lambda_1)' \left(A_{\tau-1} + 2\, A_{\tau-1}\, \Psi^{-1} A_{\tau-1}\right)(\mu - \Sigma\lambda_0)\,, \end{aligned} \tag{6}$$

$$\begin{aligned} c_\tau = &- c_r + c_{\tau-1} + b'_{\tau-1}\, (\mu - \Sigma\lambda_0) - \frac{1}{2}\ln\left|\Psi\right| - \frac{1}{2}\ln\left|\Sigma\,\Sigma'\right| \\ &+ (\mu - \Sigma\lambda_0)' \left(A_{\tau-1} + 2\, A_{\tau-1}\, \Psi^{-1} A_{\tau-1}\right)(\mu - \Sigma\lambda_0) \\ &+ \frac{1}{2}\, b'_{\tau-1} \Psi^{-1} b_{\tau-1} + 2\, b'_{\tau-1} \Psi^{-1} A_{\tau-1}\,(\mu - \Sigma\lambda_0)\,, \end{aligned} \tag{7}$$

with boundary conditions $c_0 = 0$, $b_0 = 0_{(N\times 1)}$, $A_0 = 0_{(N\times N)}$ and therefore $c_1 = -c_r$, $b_1 = -b_r$, $A_1 = -A_r$.[6] We define $\Psi \equiv (\Sigma\Sigma')^{-1} - 2A_{\tau-1}$.

The corresponding no-arbitrage affine model can then be obtained by setting $A_\tau = 0$ in the recursive difference equations (5) - (7) and $A_r = 0$ in equation (2).

The dynamic Nelson-Siegel model in its state-space form, as in Diebold and Li (2006), is in our application assumed to have the same state dynamics as those of the arbitrage-free models shown in equation (1). The observation equation is, however, a special case of equation (4), where $A_\tau = 0$, $c_\tau = 0$ and where the vector of factor loadings (corresponding to $-\frac{b_\tau}{\tau}$ in equation (4)) has the following specific functional form:

[6] The details of the derivation can be found in the Appendix.

$$-\frac{b_\tau}{\tau} \equiv \left(1 \qquad \frac{1-e^{-\gamma\tau}}{\gamma\tau} \qquad \frac{1-e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau}\right), \tag{8}$$

which does not necessarily fulfill the no-arbitrage restrictions presented in equations (5) - (7). The parameter γ is the so-called time-decay parameter.[7]

4 Estimation Results

We use U.S. Treasury zero-coupon yield curve data covering the period from January 1970 to December 2000. The sample consists of monthly yield observations for maturities of 1, 6, 12, 24, 60 and 120 months. These data are also used in Diebold and Li (2006), and are based on end-of-month CRSP government bond files.[8]

Similar to Leippold and Wu (2007) we rely on the unscented Kalman filter, developed by Julier and Uhlmann (1997), to estimate all models. Alternatively, the estimation of quadratic term structure models could be accomplished using the extended Kalman filter (EKF) or the method of moments (MM): for example, simulated method of moments (SMM) like Brandt and Chapman (2003); the efficient method of moments (EMM) as Ahn et al. (2002); or the general method of moments (GMM) as Leippold and Wu (2003).[9] It is well-know that the EKF implies a significant amount of approximation error, while in the case of the MM one needs to specify which are the most important statistical and economical moments of the data that should be matched. For example, Leippold and Wu (2003) define three categories of properties of interest rates: general statistical properties (means of

[7]Naturally, no risk premium is specified for the Nelson-Siegel model.

[8]The data can be downloaded from Francis Diebold's webpage: http://www.ssc.upenn.edu/ fdiebold/papers/paper49/FBFITTED.txt.

[9]Kim (2004) uses the linear Kalman filter but augments the state space with the quadratic function of the factors.

the sample yields and first order autocorrelation of the short rate); forecasting relations (the forward regression slope) and conditional dynamics (the hump-shaped dynamics of the conditional volatility of bond yields). The statistical properties they choose aim at matching the on-average upward sloping yield curve, the large persistence of bond yields and the positive skewness of the interest rate distribution. Brandt and Chapman (2003) base their choice of moments primarily on economic relations. They use the unconditional means and the residual standard deviations from a first-order autoregressions of the level, slope and curvature; their contemporaneous and first-order lagged correlations; the slope coefficients from linear projection of yields (LPY) regression and from a conditional volatility (LPV) regression.[10]

Instead of using a particular MM technique, one could rely on the Kalman filter in the estimation procedure as we do in the current setup. The employment of the UKF is not necessary for the estimation of the affine and the Nelson-Siegel models, since the state and the measurement equations are linear in the state vector. However, for comparison purposes, and to avoid differences stemming from the estimation procedure, we apply the UKF to all models. In this context it is also noted that the UKF has been shown to produce more accurate results than linear techniques even in the estimation of linear systems (see Wan and Merwe (2001)).

The UKF methodology is based on the idea that it is easier to approximate a distribution than it is to approximate a nonlinear function, see for example Julier and Uhlmann (2004), Julier and Uhlmann (1997) and Wan and Merwe (2001) and the references therein. As mentioned in the introduction, we estimate eleven different models, which fall in the categories of affine,

[10]They define the six-month yield, $Y_t(6)$, to be the 'level' factor, the difference between the ten-year and the six-month yield, $Y_t(120) - Y_t(6)$ to be the 'slope' factor and $Y_t(6) + Y_t(120) - 2Y_t(24)$ to be the 'curvature' factor.

quadratic and dynamic Nelson-Siegel models. Within each model class we differentiate the estimated models with respect to their degree of parsimony. In particular, when referring to the included model variants we use the notation $M_n(k)$, where: M refers to the model class, $M = \{Q, A, NS\}$, corresponding to quadratic, affine and Nelson-Siegel models respectively; $n = \{1, 2, 3\}$ refers to the model variant, where 1 stands for the maximally flexible representation of a model, 2 for the independent-factors model specification, and 3 is used only in the case of the quadratic model where it denotes the 'triangular' specification - the market price of risk matrix in equation (3) is triangular; finally, k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 1 summarizes the necessary identification restrictions for the different model classes. In addition to a characterization based on the number of included factors, the model specifications are also differentiated by their parametrization. Table 2 displays the parametrization of the investigated model variants. As can be observed from that table, we estimate five specifications of quadratic models, four affine specifications and two Nelson-Siegel. In this respect we use the three levels of parsimony referred to above: maximally flexible; minimal (independent-factors); and triangular (only in the case of 3-factor QTSM), as mentioned above.

Table 1 AROUND HERE

Table 2 AROUND HERE

Table 2 reports the corresponding imposed parameter restrictions that define the selected specifications. All possible model permutations are however not investigated. For example, we do not include two-factor versions of the Nelson-Siegel model, and we do not investigate 'triangular' affine models. We have chosen the included model variants on the basis of the trade-off between generality of results and computation time. Factors' interactions have two transmission channels. One is through the direct covariances which are accounted for by the autoregressive matrix Φ in the state equation (1). The other is through the market prices of risk, i.e. matrix λ_1 in the market-price-of-risk equation (3). Table 2 shows that for the quadratic model class, the maximally flexible specification, for example, is based on the identifying restrictions from Table 1 (i.e. the matrix of autoregressive parameters, Φ, in the state equation (1) is triangular and the error-term variance matrix Σ, is diagonal,) without imposing any further constraints on the parameters. The flexibility of this model hence stems from the specification of a full (unrestricted) market price of risk, λ_1 matrix, in equation (3). The independent-factors quadratic model deviates from the maximally flexible specification by imposing a diagonal structure on λ_1 (in addition to the diagonality imposed on the factors' autoregressive matrix Φ in equation (1)), whereas the 'triangular' model variant imposes a triangular structure on λ_1.[11] In defining the affine model variants we follow the pattern used for the quadratic models, however we do not consider a 'triangular' affine model specification. Since the Nelson-Siegel model class is formulated directly under the empirical measure, it does not require a characterization of the market price of risk.[12] In effect, the maximally flexible Nelson-Siegel specification imposes only the appropriate identification restriction of a triangular structure on the error-term variance Σ. The minimal flexible version of this model class assumes a diagonal structure for both Φ and Σ as in the quadratic and affine independent-factors cases.

[11]The canonical formulation of quadratic models in Ahn et al. (2002) defines λ_1 as a triangular matrix. However, as Kim (2004) points out this is not necessary for identification purposes. Instead it guarantees that the autoregressive matrix in the factors' law of motion is triangular under both the physical and the risk-neutral measures.

[12]For this reason Table 2 contains 'n.a.' entries for λ_0 and λ_1 under the Nelson-Siegel model variants.

4.1 In-sample fit

Tables 3 to 7 report the parameter estimates of the analyzed model variants. Table 3 shows the estimates of the state equation (1), Table 4 displays the estimates of the market price of risk equation (3), Table 5 contains parameter estimates of the equation for the short rate (2). The parameter describing the time-decay of the loading structure in the Nelson-Siegel model from equation (8) is presented in Table 6. The estimated standard deviations of the error terms in the observation equation (4), for each of the estimated model variants, are shown in Table 7.

Table 3 AROUND HERE

Table 4 AROUND HERE

Table 5 AROUND HERE

Table 6 AROUND HERE

Table 7 AROUND HERE

To facilitate in-sample fit comparisons of the estimated models, Table 8 displays statistics on the error-terms from the yield curve observation equation (4) at maturities of 1, 6, 12, 24, 60 and 120 months. The mean, standard deviation, min, max, autocorrelation of first, second and twelfth order, and mean absolute deviation (MAD) of the errors are shown for each of the estimated model variants.

Table 8 AROUND HERE

Figure 1 AROUND HERE

Figure 2 AROUND HERE

Figure 3 AROUND HERE

Table 8 demonstrates that all models in general fit the data well, as also confirmed by Figures 1 to 3. The well-known phenomenon, stemming from the near integration of the time-series of yields, of relatively high error-term autocorrelation, is found for all models. At lag one, the autocorrelation ranges from 0.3 to 0.7, approximately, and it almost disappears at lag 12. When investigating how the mean errors and the MAD depend on maturity across the models, it seems that better fits are provided for the medium part of the maturity spectrum, whereas yields for short and long maturities are fitted slightly worse, with the worst fit produced for the shortest maturities. Another pattern that emerges from Table 8 is that, as expected, the more flexible models fit the data better than the less flexible models do. Judging the in-sample fit by the MAD, the overall best fitting model is the maximally flexible three factor quadratic model Q1(3), which produces the smallest MAD for all maturities. The best fitting model class is the quadratic model that uses three factors, followed by the affine three-factor model, the Nelson-Siegel model class, the quadratic two-factor model, and the worst fitting, in relative terms, is the affine two-factor model.

Table 9 AROUND HERE

Table 9 displays the characteristics of the estimated latent factors and Figure 4 presents time series plots of these factors. It is difficult to give an economic interpretation to latent factors directly. Thus we report in Table 9 the correlations between the latent factors and the principal components as well as their correlation with the level, the slope and the curvature. In the latter case we consider as level the long end of the curve, $Y_t(120)$, as slope the difference between the long and the short end, $[Y_t(120) - Y_t(1)]$, and as curvature - $[Y_t(120) + Y_t(1) - 2Y_t(24)]$. In general, the most persistent factor is most highly correlated with the first principal component, and with the yield curve level. The least persistent factor is most highly correlated with the third (second) principal component and with the curvature (slope) in the case of three- (two-) factor models. In the case of quadratic models the less and the least persistent factors are cross correlated and in this case the correlation with the second and the third PC (and with the slope and the curvature) is not so high and clearly distinguished as in the case of the other model specifications.

Figure 4 AROUND HERE

Figure 5 AROUND HERE

Figure 6 AROUND HERE

4.2 Out-of-sample fit

To compare the out-of-sample forecast performance of the investigated model classes we reestimate the model recursively and produce forecasts using expanding data samples starting in January 1994 and ending in December 2000. We first estimate the models on a sub-sample covering January 1970 to January 1993, and produce forecasts for the 1, 6 and 12 month horizons; then, one observation is added to the data sample, and the models are re-estimated, after which a new set of forecasts is generated, again for the horizons of 1, 6 and 12 months. This process is repeated until the full data sample is covered and a total of 96 forecasts are generated for each model.

As Kim (2004) suggests, the forecasting performance of a given model could be highly sensitive to the chosen out-of-sample forecasting period, the forecasting horizon and even the method of estimation. We estimate all the models with the UKF, although this is not necessary for the affine models. We reserved the last seven years of data for the forecasting exercise, in order to perform tests on the produced forecast statistics.

As a gauge to compare the out-of-sample performance of the models we rely on the mean squared prediction errors (MSPEs) of each model, divided by the MSPEs produced by the random walk. In order to perform statistical test to identify which model(s) out-perform other models, at a given level of confidence, the Diebold-Mariano (DM) test (see Diebold and Mariano (1995)) and the Clark-West (CW) test (see, Clark and West (2007) and Clark and West (2006)) are used. The latter test is needed when the tested models are nested, and the former test is used when models are non-nested.[13] Table 10 documents when one or the other test is used. The null hypothesis of the DM test is that the models have equal MSPEs. The null hypothesis under the CW test is that the more general model has a MSPE greater than or equal to the one of the more parsimonious model (the nested model), while the alternative is that the larger model has a smaller MSPE than the parsimonious one.

Table 10 AROUND HERE

Table 11 AROUND HERE

Table 11 contains the main results of our analysis. It documents the ratios of MSPE ratios of the model under consideration to the random walk for each of the investigated models. The bold entries in the table show the 'best' model in terms of the smallest MSPE for a given forecasting horizon at a given maturity segment. For example, the first bold entry in Table 11 is 0.701, observed at the one-month forecasting horizon for the one-month segment of the yield curve. This entry signifies that the more parsimonious Nelson-Siegel model (NS2(3)) performs best at this forecasting horizon for

[13]The quadratic class of models is more general than the affine, and one could be tempted to think that the affine model class is fully nested by the quadratic one. However, this is not necessarily so. Looking at the maximally admissible specifications defined by Dai and Singleton (2000) and Ahn et al. (2002) for affine and quadratic models, respectively, and consulting Table 1 it can be seen that the estimated affine models specifications cannot be obtained from the quadratic model class simply by setting the qudratic terms' coefficients equal to zero. For example, while b_r in equation (2) is estimated in the affine specifications, it is set equal to zero in the qudratic ones for identification purposes.

that yield curve segment, judging by the ratio of MSPEs of NS2(3) to the RW. Identifying the 'best' (smallest) MSPE ratios for each maturity and forecasting horizon, we perform the corresponding test (see Table 10) to determine whether the MSPE of the referenced model is statistically different from the others. We also test whether the MSPE of each model is significanlty different from the MSPE of the random walk with the Diebold-Mariano test. A star '*' in Table 11 indicates that a given model's MSPE ratio with the random walk's MSPE is significantly different from unity. Meaning that the model under consideration performs significantly better (worse) than the random walk, if the ratio is lower (higher) than unity. The applied level of confidence for all tests performed in this analysis is 95%.

Based on the results presented in Table 11 no clear winner of the forecasting experiment emerges. It is also not possible to find a model that dominates other models at a given set of maturities or at a given forecasting horizon. However, some tendencies seem apparent. First, all models perform better as the forecasting horizon is extended. For example, all three-factor quadratic specifications, realize a higher number of performance ratios below unity as the forecasting horizon is increased from 1 to 6 months. Similarly, the Nelson-Siegel model class presents a noticeable improvement in the forecast ratios when extending the forecast horizon from 1 month to 6 months and from 6 to 12 months. Second, at the one-month forecast horizon, the performance of the quadratic three-factor models and the affine and quadratic two-factor models, exhibits a U-shaped pattern, indicating that these model classes, judged in isolation, are relatively better at forecasting yields from the medium maturity spectrum. Third, affine three-factor models and the Nelson-Siegel models show a generally better performance at forecasting short maturities than forecasting medium-term maturities, and the forecasting performance further deteriorates for longer maturities. Forth, the affine three-factor models and the Nelson-Siegel based models overall seem to perform better than the quadratic three-factor models and the quadratic and affine two-factor models. Fifth, it seems that the Nelson-Siegel model class produces slightly better forecasts than all competing models for the longest maturity, regardless of the forecasting horizon. However, when judged across all tested forecasting horizons and included maturities the performance of the three-factor affine model class and the Nelson-Siegel model class is indistinguishable.

Table 12 AROUND HERE

Table 13 AROUND HERE

Table 14 AROUND HERE

Table 15 AROUND HERE

Tables 12 to 15 display all forecasts that the performed statistical tests fail to reject as equally good. For a given maturity and forecasting horizon, we test each model's forecast against the 'best', i.e. the one with the smallest MSPE. In the cases where we apply the Diebold-Mariano test, we report in the tables only the values for which the zero hypothesis of equal MSPEs cannot be rejected. In the cases where the appropriate test to apply is Clark-West we keep in the tables only the values for which the test is rejected, i.e. the zero hypothesis that the larger model has also larger MSPE than the more parsimonious model is rejected (the alternative is that the larger model has a smaller MSPE). All test results we report are at the 95% level of significance. Table 12 presents the statistical test for all models and confirms the conclusions highlighted above. The rest of the tables slices-and-dices the model forecasts according to the imposed model specification. Table 13 shows the equally good forecasts among three-factor models. Table

14 presents similar results among the maximally flexible models and Table 15 among the diagonal models.

5 Conclusion

An extensive out-of-sample forecasting experiment is conducted among quadratic, affine and dynamic Nelson-Siegel models. Using US data covering the period from January 1970 to December 2000 a recursive re-estimation and out-of-sample forecasting methodology is implemented for eleven model specifications falling in the three main yield curve modeling categories. Forecasts are generated on the basis of the estimated models at forecasting horizons of 1, 6 and 12 months, for each model specification.

Our results show that while quadratic three-factor models provide the best in-sample fit, the conclusion as regards the out-of-sample comparison of the tested models is less clear. A tendency emerges, showing that all models perform better the longer the forecasting horizon; and that the dynamic Nelson-Siegel models seem to perform best among the tested models, for longer maturity segments of the yield curve, especially at longer forecasting horizons.

The main qualitative conclusion of the model comparison conducted in the current study is that affine three-factor model and the dynamic Nelson-Siegel models perform equally well in the out-of-sample forecasting experiment, and they perform better than the quadratic three- and two-factor models.

APPENDIX

Assume that the dynamic evolution of the vector of N state variables, X_t, under the risk-neutral measure, Q, is described by

$$X_{t+1} = \mu^Q(X_t) + \Sigma\xi_{t+1}, \tag{A-1}$$

where $\xi_{t+1} \sim N(0, I)$, $\mu^Q(X_t)$ is a vector of $N \times 1$ functions of the state variables and Σ is an $N \times N$ matrix. Assume also that the state variables follow a VAR(1) process under the objective measure

$$X_{t+1} = \mu + \Phi X_t + \Sigma\eta_{t+1}, \tag{A-2}$$

with $\eta_{t+1} \sim N(0, I)$. Note that the variance-covariance matrix, Σ, is the same under both measures. Further we specify the market price of risk as a linear function of the state

$$\Lambda(X_t) = \lambda_0 + \lambda_1 X_t, \tag{A-3}$$

where λ_0 is an $N \times 1$ vector and λ_1 is $N \times N$ matrix. Then

$$\begin{aligned}\mu^Q(X_t) =& \mu + \Phi X_t - \Sigma\Lambda(X_t)\\ =& \Phi X_t + \mu - \Sigma\lambda_0 - \Sigma\lambda_1 X_t\\ =& (\Phi - \Sigma\lambda_1)X_t + \mu - \Sigma\lambda_0\end{aligned} \tag{A-4}$$

The price of a zero-coupon bond is an exponential quadratic function of the state variables

$$P_{t,\tau}(X_t) = \exp\left[c_\tau + b'_\tau X_t + X'_t A_\tau X_t\right], \tag{A-5}$$

and the yield on a zero-coupon bond with τ periods to maturity is then

$$\begin{aligned}Y_{t,\tau}(X_t) =& -\frac{\ln P_{t,\tau}(X_t)}{\tau}\\ =& -\frac{c_\tau}{\tau} - \frac{1}{\tau}b'_\tau X_t - \frac{1}{\tau}X'_t A_\tau X_t.\end{aligned}$$

In quadratic models the short rate is a quadratic function of the state

$$r(X_t) = c_r + b'_r X_t + X'_t A_r X_t, \tag{A-6}$$

with A_r an $N \times N$ matrix, b_r $N \times 1$ vector and c_r is a constant.

The price of a zero-coupon bond at time t with τ periods to maturity satisfies

$$P_{t,\tau} = E_t^Q\left[\exp\left(-\sum_{i=t}^{t+\tau-1} r_i\right)\right] = E_t^Q\left[\exp\left(-r_t\right)P_{t+1,\tau-1}\right], \tag{A-7}$$

where E_t^Q denotes the expectation under the risk-neutral probability measure.

The system of difference equations (5) - (7) is obtained in the following way. Substitute (A-5) and (A-6) in (A-7):

$$
\begin{aligned}
&\exp\left(c_\tau + b_\tau' X_t + X_t' A_\tau X_t\right) \qquad \text{(A-8)}\\
&= E_t^Q\left[\exp\left(-c_r - b_r' X_t - X_t' A_r X_t\right)\exp\left(c_{\tau-1} + b_{\tau-1}' X_{t+1} + X_{t+1}' A_{\tau-1} X_{t+1}\right)\right]\\
&= \exp\left(-c_r - b_r' X_t - X_t' A_r X_t\right) E_t^Q\Big[\exp\left(c_{\tau-1} + b_{\tau-1}' X_{t+1} + X_{t+1}' A_{\tau-1} X_{t+1}\right)\Big].
\end{aligned}
$$

From the equation for the state process under the risk-neutral measure, (A-1), we obtain

$$
\begin{aligned}
X_{t+1}' A_{\tau-1} X_{t+1} &= \left[\mu^Q(X_t) + \Sigma\xi_{t+1}\right]' A_{\tau-1}\left[\mu^Q(X_t) + \Sigma\xi_{t+1}\right]\\
&= \mu^Q(X_t)' A_{\tau-1}\mu^Q(X_t) + \xi_{t+1}'\Sigma' A_{\tau-1}\Sigma\xi_{t+1} + 2\mu^Q(X_t)' A_{\tau-1}\Sigma\xi_{t+1}\\
&= M + K'\Sigma\xi_{t+1} + \xi_{t+1}'\Sigma' A_{\tau-1}\Sigma\xi_{t+1}
\end{aligned}
$$

where $M \equiv \mu^Q(X_t)' A_{\tau-1}\mu^Q(X_t)$ and $K' \equiv 2\mu^Q(X_t)' A_{\tau-1}$. Next from (A-1) it follows also that

$$
b_{\tau-1}'\, X_{t+1} = b_{\tau-1}'\left[\mu^Q(X_t) + \Sigma\,\xi_{t+1}\right] = b_{\tau-1}'\mu^Q(X_t) + b_{\tau-1}'\Sigma\,\xi_{t+1}.
$$

Substituting these in (A-8) and taking logs it follows that

$$
\begin{aligned}
&c_\tau + b_\tau' X_t + X_t' A_\tau X_t\\
&= - c_r - b_r' X_t - X_t' A_r X_t + c_{\tau-1} + b_{\tau-1}'\mu^Q(X_t) + M\\
&\quad + \ln\left\{E_t^Q\left[\exp\left(b_{\tau-1}'\Sigma\xi_{t+1} + K'\Sigma\xi_{t+1} + \xi_{t+1}'\Sigma' A_{\tau-1}\Sigma\xi_{t+1}\right)\right]\right\}\\
&= - c_r - b_r' X_t - X_t' A_r X_t + c_{\tau-1} + b_{\tau-1}'\mu^Q(X_t) + \mu^Q(X_t)' A_{\tau-1}\mu^Q(X_t)\\
&\quad + \ln\left\{E_t^Q\left[\exp\left(b_{\tau-1}'\Sigma\xi_{t+1} + K'\Sigma\xi_{t+1} + \xi_{t+1}'\Sigma' A_{\tau-1}\Sigma\xi_{t+1}\right)\right]\right\}. \qquad \text{(A-9)}
\end{aligned}
$$

Then notice that

$$
\begin{aligned}
&\ln\left\{E_t^Q\left[\exp\left(b_{\tau-1}'\Sigma\xi_{t+1} + K'\Sigma\xi_{t+1} + \xi_{t+1}'\Sigma' A_{\tau-1}\Sigma\xi_{t+1}\right)\right]\right\}\\
&= \ln\frac{\exp\left\{\frac{1}{2}\left(b_{\tau-1}' + K'\right)\left[\left(\Sigma\Sigma'\right)^{-1} - 2A_{\tau-1}\right]^{-1}\left(b_{\tau-1} + K\right)\right\}}{\left|\Sigma\Sigma'\right|^{\frac{1}{2}}\left|\left(\Sigma\Sigma'\right)^{-1} - 2A_{\tau-1}\right|^{\frac{1}{2}}}\\
&= \frac{1}{2}\left(b_{\tau-1} + K\right)'\left[\left(\Sigma\Sigma'\right)^{-1} - 2A_{\tau-1}\right]^{-1}\left(b_{\tau-1} + K\right)\\
&\quad + \ln\left|\left(\Sigma\Sigma'\right)^{-1} - 2A_{\tau-1}\right|^{-\frac{1}{2}} - \ln\left|\Sigma\Sigma'\right|^{\frac{1}{2}}\\
&= \frac{1}{2}\left[b_{\tau-1}' + 2\mu^Q(X_t)' A_{\tau-1}\right]\left[\left(\Sigma\Sigma'\right)^{-1} - 2A_{\tau-1}\right]^{-1}\left[b_{\tau-1} + 2A_{\tau-1}\mu^Q(X_t)\right]\\
&\quad - \ln\left|\left(\Sigma\Sigma'\right)^{-1} - 2A_{\tau-1}\right|^{\frac{1}{2}} - \ln\left|\Sigma\Sigma'\right|^{\frac{1}{2}}\\
&= \frac{1}{2} b_{\tau-1}'\Psi^{-1} b_{\tau-1} + 2b_{\tau-1}'\Psi^{-1} A_{\tau-1}\mu^Q(X_t) + 2\mu^Q(X_t)' A_{\tau-1}\Psi^{-1} A_{\tau-1}\mu^Q(X_t)\\
&\quad + \ln\left|\Psi\right|^{-\frac{1}{2}} - \ln\left|\Sigma\Sigma'\right|^{\frac{1}{2}}, \qquad \text{(A-10)}
\end{aligned}
$$

where $\Psi \equiv (\Sigma\Sigma')^{-1} - 2A_{\tau-1}$.[14] Substitute this in (A-9)

$$
\begin{aligned}
&c_\tau + b'_\tau X_t + X'_t A_\tau X_t \\
&= - c_r - b'_r X_t - X'_t A_r X_t + c_{\tau-1} + b'_{\tau-1}\mu^Q(X_t) + \mu^Q(X_t)' A_{\tau-1}\mu^Q(X_t) \\
&\quad + \frac{1}{2} b'_{\tau-1}\Psi^{-1} b_{\tau-1} + 2b'_{\tau-1}\Psi^{-1}A_{\tau-1}\mu^Q(X_t) + 2\mu^Q(X_t)' A_{\tau-1}\Psi^{-1}A_{\tau-1}\mu^Q(X_t) \\
&\quad + \ln\left|\Psi\right|^{-\frac{1}{2}} - \ln\left|\Sigma\Sigma'\right|^{\frac{1}{2}}. \qquad \text{(A-11)}
\end{aligned}
$$

Next substitute for $\mu^Q(X_t)$ in (A-11) from (A-4) and group the terms corresponding to the different degrees of X_t

$$
\begin{aligned}
&c_\tau + b'_\tau X_t + X'_t A_\tau X_t \\
&= - c_r - b'_r X_t - X'_t A_r X_t + c_{\tau-1} + b'_{\tau-1}(\Phi - \Sigma\lambda_1) X_t + b'_{\tau-1}(\mu - \Sigma\lambda_0) \\
&\quad + X'_t (\Phi - \Sigma\lambda_1)' \left(A_{\tau-1} + 2A_{\tau-1}\Psi^{-1}A_{\tau-1}\right)(\Phi - \Sigma\lambda_1) X_t \\
&\quad + X'_t (\Phi - \Sigma\lambda_1)' \left(A_{\tau-1} + 2A_{\tau-1}\Psi^{-1}A_{\tau-1}\right)(\mu - \Sigma\lambda_0) \\
&\quad + (\mu - \Sigma\lambda_0)' \left(A_{\tau-1} + 2A_{\tau-1}\Psi^{-1}A_{\tau-1}\right)(\Phi - \Sigma\lambda_1) X_t \\
&\quad + (\mu - \Sigma\lambda_0)' \left(A_{\tau-1} + 2A_{\tau-1}\Psi^{-1}A_{\tau-1}\right)(\mu - \Sigma\lambda_0) \\
&\quad + \frac{1}{2} b'_{\tau-1}\Psi^{-1} b_{\tau-1} + 2b'_{\tau-1}\Psi^{-1}A_{\tau-1}(\Phi - \Sigma\lambda_1) X_t \\
&\quad + 2b'_{\tau-1}\Psi^{-1}A_{\tau-1}(\mu - \Sigma\lambda_0) - \frac{1}{2}\ln\left|\Psi\right| - \frac{1}{2}\ln\left|\Sigma\Sigma'\right|.
\end{aligned}
$$

Finally the difference equations become

$$
\begin{aligned}
A_\tau &= - A_r + (\Phi - \Sigma\lambda_1)' \left[A_{\tau-1} + 2A_{\tau-1}\Psi^{-1}A_{\tau-1}\right](\Phi - \Sigma\lambda_1), \\
b_\tau &= - b_r + (\Phi - \Sigma\lambda_1)' b_{\tau-1} + 2(\Phi - \Sigma\lambda_1)' A_{\tau-1}\Psi^{-1} b_{\tau-1} + \\
&\qquad 2(\Phi - \Sigma\lambda_1)' \left(A_{\tau-1} + 2A_{\tau-1}\Psi^{-1}A_{\tau-1}\right)(\mu - \Sigma\lambda_0), \\
c_\tau &= - c_r + c_{\tau-1} + b'_{\tau-1}(\mu - \Sigma\lambda_0) - \frac{1}{2}\ln\left|\Psi\right| - \frac{1}{2}\ln\left|\Sigma\Sigma'\right| \\
&\quad + (\mu - \Sigma\lambda_0)' \left(A_{\tau-1} + 2A_{\tau-1}\Psi^{-1}A_{\tau-1}\right)(\mu - \Sigma\lambda_0) \\
&\quad + \frac{1}{2} b'_{\tau-1}\Psi^{-1} b_{\tau-1} + 2b'_{\tau-1}\Psi^{-1}A_{\tau-1}(\mu - \Sigma\lambda_0).
\end{aligned}
$$

with the boundary conditions $c_0 = 0$, $b_0 = 0_{(N\times 1)}$ and $A_0 = 0_{(N\times N)}$ and therefore $c_1 = -c_r$, $b_1 = -b_r$, $A_1 = -A_r$.

The one-period yield

[14] The expression after the first equality sign in equation (A-10) follows from the fact that if If $\xi \sim N(0, I)$, i.e. $\Sigma\xi \sim N(0, \Sigma\Sigma')$, then

$$
E\left[\exp\left(\xi'\Sigma' A \Sigma\xi + b'\Sigma\xi\right)\right] = \frac{\exp\left\{\frac{1}{2} b' \left[(\Sigma\Sigma')^{-1} - 2A\right]^{-1} b\right\}}{\left|\Sigma\Sigma'\right|^{\frac{1}{2}}\left|(\Sigma\Sigma')^{-1} - 2A\right|^{\frac{1}{2}}}.
$$

$$
\begin{aligned}
Y_t(1) &= -\ln P_{t,1}(X_t) = -c(1) - b(1)' X_t - X_t' A(1) X_t \\
&= c_r + b_r' X_t + X_t' A_r X_t = r(X_t)
\end{aligned}
$$

The corresponding difference equations of the affine model can be easily obtained by substituting for $A_\tau = A_r = 0$.

Following Julier and Uhlmann (1997) we rely on the UKF method to estimate the parameters for the tested models. The general idea of the UKF is to chose some points of the distribution of the state variable (called sigma points), e.g. the mean and points spread around it, where the spread is a scaling of the standard deviation. The points are propagated then through the non-linear system. First through the dynamic state function, the state variable is updated and then its new weighted mean and covariance estimates are computed. The sigma points are redrawn and then propagated through the measurement function. The observed variable that corresponds to each of these sigma points is computed using the specified non-linear measurement function. The mean and the covariance of the observed/measured variable are then approximated as weighted sample mean and covariance of the posterior sigma points. The weights depend on exogenous parameters which adapt according to the distribution of the state variable. The exogenous weights can shrink or expand the points about the mean thus decreasing or increasing respectively the effect of higher order moments. The advantages of the UKF over the EKF are that the unscented filter does not require derivative computations and calculates the mean to a higher order of accuracy than the extended one, whereas the covariance is calculated to the same order of accuracy in the case of the UKF, as in the EKF. At the same time the UKF is not computationally more complex that the EKF.

References

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., M. Piazzesi, and M. Wei (2006) 'What does the yield curve tell us about GDP growth.' *Journal of Econometrics* 131, 359–403

Bjork, T., and B.J. Christensen (1999) 'Interest rate dynamics and consistent forward rate curves.' *Mathematical Finance* 9(4), 323–348

Brandt, Michael W., and David A. Chapman (2003) 'Comparing multifactor models of the term structure.' Working paper

Cheridito, Patrick, Damir Filipović, and Robert L. Kimmel (2007) 'Market price of risk specifications for affine models: Theory and evidence.' *Journal of Financial Economics* 83, 123–170

Christensen, Jens, Francis X. Diebold, and Glenn D. Rudebusch (2008a) 'An arbitrage-free generalized Nelson-Siegel term structure model.' University of Pennsilvania and FRB of San Francisco Working Paper

Christensen, Jens H. E., Francis X. Diebold, and Glenn D. Rudebusch (2008b) 'The affine arbitrage-free class of Nelson-Siegel term structure models.' NBER Working paper No. 13611

Clark, Todd E., and Kenneth D. West (2006) 'Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis.' *Journal of Econometrics* 135, 155–186

——— (2007) 'Approximately normal tests for equal predictive accuracy in nested models.' *Journal of Econometrics* 138, 291–311

Coroneo, L., K. Nyholm, and R. Vidova-Koleva (2008) 'How arbitragefree is the Nelson-Siegel model?' ECB Working Paper no 874

Dai, Qiang, A. Le, and Kenneth Singleton (2006) 'Discrete-time dynamic term structure models with generalized market prices of risk.' Working paper

Dai, Qiang, and Kenneth Singleton (2000) 'Specification analysis of affine term structure models.' *Journal of Finance* 55, 1943–1978

——— (2002) 'Expectation puzzles, time-varying risk premia, and affine models of the term structure.' *Journal of Financial Economics* 63, 415–441

——— (2003) 'Term structure dynamics in theory and reality.' *Review of Financial Studies* 16(3), 631–678

Diebold, Francis X., and Canlin Li (2006) 'Forecasting the term structure of government bond yields.' *Journal of Econometrics* 130, 337–364

Diebold, Francis X., and Roberto S. Mariano (1995) 'Comparing predictive accuracy.' *Journal of Business and Economic Statistics* 13(3), 253–263

Diebold, Francis X., Lei Ji, and Canlin Li (2006) 'A three-factor yield curve model: Non-affine structure, systematic risk sources, and generalized duration.' In *Macroeconomics, Finance and Econometrics: Essays in Memory of Albert Ando,* ed. L. R. Klein (Cheltenham, U.K.: Edward Elgar) pp. 240–274

Duffee, Gregory 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

Julier, Simon J., and Jeffrey K. Uhlmann (1997) 'A new extension of the Kalman filter to nonlinear systems.' In 'Proceedings of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls'

—— (2004) 'Unscented filtering and nonlinear estimation.' In 'Proceedings of the IEEE,' vol. 92 pp. 401–422

Kim, Don H. (2004) 'Time-varying risk and return in the quadratic-Gaussian model of the term structure.' manuscript

Kim, Dong H. (2005) 'Nonlinearity in the term structure.' manuscript, University of Manchester

Leippold, M., and L. Wu (2002) 'Asset pricing under the quadratic class.' *Journal of Financial and Quantitative Analysis* 37, 271–295

—— (2003) 'Design and estimation of quadratic term structure models.' *European Finance Review* 7, 47–73

—— (2007) 'Design and estimation of multi-currency quadratic models.' *Review of Finance* 11, 167–207

Nelson, C.R., and A.F. Siegel (1987) 'Parsimonious modeling of yield curves.' *Journal of Business* 60, 473–89

Piazzesi, Monika (2004) 'Affine term structure models.' In *Handbook of Financial Econometrics,* ed. Yacine Ait-Sahalia and Lars Peter Hansen

Realdon, M. (2006) 'Quadratic YTerm Structure Models in Discrete Time.' *Finance Research Letters* 3, 277–289

Wan, Eric A., and Rudolph Van Der Merwe (2001) 'Chapter 7: The unscented kalman filter.' In 'Kalman Filtering and Neural Networks' Wiley pp. 221–280

Table 1: Conditions for Identification

model	Restrictions
quadratic	Φ - triangular, Σ - diagonal, $b_r = 0$, A_r - symmetric
affine	Φ - triangular, $\mu = 0$, Σ - diagonal

This table shows admissibility conditions for affine and quadratic term structure models. The variables μ, Φ and Σ are the vector of constants, the matrix of autoregressive parameters, and the matrix of error-term co-variances in the equation of the yield curve factor dynamics, $X_{t+1} = \mu + \Phi X_t + \Sigma \eta_{t+1}$. The variables, b_r and A_r are the constant and the parameter matrix corresponding to the quadratic term in the equation for the short rate process $r(X_t) = c_r + b_r' X_t + X_t' A_r X_t$.

Table 2: Estimated Models' Specifications

Model	Parameters to estimate		Restrictions		
	k = 3 factors	k = 2 factors	Φ	Σ	λ_1
Q1(k) (maximally flexible)	33	20	triang	diag	full
Q2(k) (minimal)	21	16	diag	diag	diag
Q3(k) (triangular)	30		triang	diag	triang
A1(k) (maximally flexible)	30	19	triang	diag	full
A2(k) (minimal)	21	16	diag	diag	diag
NS1(k) (maximally flexible)	25		full	triang	n.a.
NS2(k) (minimal)	16		diag	diag	n.a.

This table presents the evaluated model specifications and the imposed parameter restrictions. The variables Φ and Σ are the matrix of autoregressive parameters and the matrix of error-term co-variances in the equation for the yield curve factor dynamics, $X_{t+1} = \mu + \Phi X_t + \Sigma \eta_{t+1}$. The variable λ_1 is the matrix of parameters that load on the yield curve factors X_t in the equation for the market price of risk, $\Lambda(X_t) = \lambda_0 + \lambda_1 X_t$. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 3: Estimated parameters - state equation

	Φ			$\mu * 100$	$\Sigma * 100$		
Q1(3)	0.96			-0.69	0.21		
	-0.20	0.66		-2.18		0.18	
	-0.02	-0.04	0.99	-0.41			0.15
Q2(3)	0.96			-0.20	0.46		
		0.98		-0.13		0.22	
			0.63	41.53			0.01
Q3(3)	0.96			-0.35	0.21		
	0.04	0.65		0.25		0.04	
	0.04	-0.41	1.00	0.18			0.21
A1(3)	0.99				1.17		
	0.27	0.92				18.01	
	-0.02	0.00	0.93				1.68
A2(3)	0.98				0.59		
		0.93				0.78	
			0.91				5.07

Estimated parameters - state equation (continued)

	Φ			$\mu * 100$	$\Sigma * 100$		
NS1(3)	0.99	0.01	-0.01	0.01	0.02		
	-0.04	0.93	0.01	0.01	-0.01	0.04	
	-0.01	0.02	0.96	0.01	0.02	0.00	0.07
NS2(3)	0.98			0.01	0.02		
		0.93		-0.01		0.04	
			0.95	0.00			0.08
Q1(2)	0.98			0.20	0.23		
	0.03	0.94		0.07		0.37	
Q2(2)	0.94			0.28	0.49		
		0.99		0.09		0.22	
A1(2)	0.97				0.43		
	0.74	0.91				9.96	
A2(2)	0.97				0.20		
		0.92				1.51	

This table shows the parameter estimates of the state equation $X_{t+1} = \mu + \Phi X_t + \Sigma\eta_{t+1}$. The variables μ, Φ and Σ are the vector of constants, the matrix of autoregressive parameters, and the matrix of the error term co-variances. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$. Standard errors are not reported due to the computational burden.

Table 4: Estimated parameters - market price of risk

	λ_1			λ_0
Q1(3)	8.92	-5.85	6.97	3.02
	-1.21	4.46	3.61	-0.28
	3.94	-0.59	-2.25	0.18
Q2(3)	4.76			0.34
		-8.34		-0.55
			3.08	-3.87
Q3(3)	15.89			1.50
	-1.36	-1.96		0.55
	-8.26	2.39	-0.53	-0.55
A1(3)	-1.07	1.13	4.03	0.13
	1.50	0.87	1.91	-0.07
	-1.60	1.05	2.02	0.25
A2(3)	-3.09			0.03
		-1.16		0.19
			4.93	-0.11

Estimated parameters - market price of risk (continued)

	λ_1		λ_0
Q1(2)	-6.80	-2.06	0.88
	5.03	0.80	-0.47
Q2(2)	1.81		-0.20
		-5.68	0.37
A1(2)	-6.27	-0.15	0.06
	7.63	-0.31	-0.15
A2(2)	-11.84		0.02
		-1.63	0.19

This table presents the parameter estimates of the market price of risk equation $\Lambda(X_t) = \lambda_0 + \lambda_1 X_t$. The variables λ_0 and λ_1 are the constant and the matrix of parameters that load on the yield curve factors X_t. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$. Standard errors are not reported due to the computational burden.

Table 5: Estimated parameters - short rate

	A_r			b_r	c_r
Q1(3)	1	-2.45	-1.02		0.00
	-2.45	1	2.09		
	-1.02	2.09	1		
Q2(3)	1				-1.29
		1			
			1		
Q3(3)	1	10.19	1.01		0.00
	10.19	1	8.91		
	1.01	8.91	1		
A1(3)				-0.02	0.01
				0.00	
				-0.02	
A2(3)				-0.06	0.01
				-0.06	
				0.01	

Estimated parameters - short rate (continued)

	A_r		b_r	c_r
Q1(2)	1	-0.94		0.00
	-0.94	1		
Q2(2)	1			0.00
		1		
A1(2)			-0.07	0.01
			0.00	
A2(2)			-0.16	0.01
			-0.03	

This table shows the parameter estimates for the short rate equation $r(X_t) = c_r + b_r' X_t + X_t' A_r X_t$. The variables, c_r, b_r and A_r are the constant, the vector of parameters that load linearly on the yield curve factors X_t and the parameter matrix corresponding to the term that is quadratic in X_t. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$. Standard errors are not reported due to the computational burden.

Table 6: Estimated parameters - NS time-decay parameter

	γ_{NS}
NS1(3)	0.072
NS2(3)	0.066

This table shows the parameter estimates of the time-decay parameter in the Nelson-Siegel factor loading matrix, $-\frac{b_\tau}{\tau} \equiv \begin{pmatrix} 1 & \frac{1-e^{-\gamma\tau}}{\gamma\tau} & \frac{1-e^{-\gamma\tau}}{\gamma\tau} - e^{-\gamma\tau} \end{pmatrix}$. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 7: Standard Deviations of Observation Equation Errors

maturity	Q1(3)	Q2(3)	Q3(3)	A1(3)	A2(3)	NS1(3)	NS2(3)	Q1(2)	Q2(2)	A1(2)	A2(2)
1	0.05	0.19	0.06	0.28	0.26	0.54	0.54	0.68	0.69	0.71	0.71
6	0.16	0.15	0.16	0.15	0.16	0.02	0.04	0.20	0.21	0.23	0.23
12	0.06	0.07	0.06	0.07	0.07	0.12	0.12	0.08	0.07	0.06	0.06
24	0.11	0.11	0.11	0.11	0.11	0.05	0.05	0.12	0.12	0.12	0.12
60	0.05	0.07	0.05	0.00	0.05	0.13	0.13	0.07	0.08	0.04	0.06
120	0.24	0.22	0.24	0.25	0.25	0.13	0.12	0.23	0.22	0.26	0.25
logL func	15935	15747	15933	15793	15696	15748	15719	15512	15440	15385	15361

This table shows the parameter estimates of the diagonal matrix $\sqrt{R}$, i.e. the standard deviations of the error-term of the observation equation, $Y(X_t,\tau) = -\frac{c_\tau}{\tau} - \frac{b'_\tau}{\tau}X_t - X'_t\frac{A_\tau}{\tau}X_t + \epsilon_{t,\tau}$, where $\epsilon_{t,\tau} \sim N(0,R)$ and R is assumed to be diagonal, i.e. it is assumed that observation errors are not correlated across maturities ($cov(\epsilon_{t,\tau_i},\epsilon_{t,\tau_j}) = 0$, $\tau_i \neq \tau_j$ for all i,j) and also across time ($cov(\epsilon_{t,\tau},\epsilon_{s,\tau}) = 0$ for $t \neq s$) for maturities of 1, 6, 12, 24, 60 and 120 months. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1,2,3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2,3\}$.

Table 8: In-sample fit

	τ	mean	st.dev.	min	max	ρ_1	ρ_2	ρ_{12}	MAD
Q1(3)	1	-0.04	0.23	-1.47	0.68	0.56	0.38	0.02	0.05
	6	0.03	0.19	-0.73	0.96	0.53	0.38	0.16	0.04
	12	-0.01	0.09	-0.42	0.58	0.43	0.26	0.08	0.01
	24	-0.01	0.13	-0.50	0.70	0.48	0.34	0.13	0.02
	60	0.01	0.06	-0.20	0.63	0.41	0.31	0.20	0.00
	120	-0.01	0.23	-0.77	0.59	0.74	0.65	0.24	0.05
Q2(3)	1	-0.05	0.26	-1.76	0.41	0.64	0.37	0.09	0.07
	6	0.04	0.20	-0.60	0.95	0.48	0.36	0.13	0.04
	12	0.02	0.14	-0.80	1.25	0.40	0.10	0.14	0.02
	24	0.02	0.16	-0.57	1.38	0.48	0.22	0.15	0.03
	60	0.02	0.10	-0.31	0.58	0.55	0.33	0.34	0.01
	120	0.00	0.23	-0.68	0.76	0.72	0.62	0.20	0.05
Q3(3)	1	-0.03	0.31	-1.66	3.40	0.40	0.27	0.04	0.10
	6	0.02	0.27	-0.97	3.47	0.36	0.24	0.07	0.07
	12	-0.02	0.19	-0.53	3.27	0.13	0.09	0.02	0.04
	24	-0.03	0.20	-0.48	3.14	0.19	0.16	0.01	0.04
	60	0.00	0.17	-0.23	3.03	0.07	0.06	0.04	0.03
	120	-0.02	0.25	-0.79	2.25	0.57	0.50	0.18	0.06
A1(3)	1	-0.02	0.34	-1.95	1.88	0.44	0.11	0.00	0.12
	6	0.09	0.33	-1.55	1.78	0.61	0.35	0.22	0.11
	12	0.04	0.27	-1.51	1.70	0.55	0.18	0.08	0.07
	24	0.01	0.24	-1.35	1.52	0.59	0.21	0.05	0.06
	60	0.01	0.12	-0.56	0.75	0.56	0.23	0.12	0.01
	120	-0.03	0.27	-0.86	1.02	0.74	0.62	0.22	0.07
A2(3)	1	-0.02	0.24	-1.87	1.32	0.33	-0.01	0.01	0.06
	6	0.07	0.26	-1.32	1.50	0.58	0.32	0.22	0.07
	12	0.03	0.22	-1.69	1.74	0.50	0.08	0.10	0.05
	24	0.01	0.20	-1.21	1.62	0.54	0.13	0.08	0.04
	60	0.01	0.10	-0.71	0.68	0.54	0.15	0.23	0.01
	120	-0.02	0.25	-0.79	0.89	0.75	0.63	0.26	0.06
NS1(3)	1	-0.23	0.45	-2.57	2.36	0.49	0.23	-0.03	0.20
	6	0.08	0.31	-1.36	1.96	0.62	0.35	0.07	0.10
	12	0.09	0.28	-1.19	1.69	0.54	0.28	0.08	0.08
	24	0.04	0.20	-0.77	1.10	0.62	0.31	0.04	0.04
	60	0.03	0.17	-0.48	0.80	0.65	0.52	0.18	0.03
	120	0.04	0.14	-0.68	0.75	0.51	0.31	-0.03	0.02

In-sample fit (continued)

	τ	mean	std.	min	max	ρ_1	ρ_2	ρ_{12}	MAD
NS2(3)	1	-0.25	0.41	-3.02	1.79	0.45	0.23	0.04	0.17
	6	0.06	0.25	-1.53	1.38	0.54	0.29	0.07	0.06
	12	0.06	0.25	-1.63	1.29	0.50	0.26	0.13	0.06
	24	0.03	0.17	-1.04	0.86	0.55	0.25	0.05	0.03
	60	0.01	0.16	-0.84	0.74	0.63	0.53	0.17	0.03
	120	0.03	0.12	-0.35	0.69	0.48	0.29	0.02	0.01
Q1(2)	1	-0.37	0.51	-2.85	1.64	0.63	0.49	0.05	0.26
	6	0.00	0.22	-0.84	1.04	0.62	0.45	0.08	0.05
	12	0.04	0.14	-1.04	0.75	0.50	0.26	0.08	0.02
	24	0.02	0.15	-0.61	0.85	0.58	0.29	0.03	0.02
	60	0.00	0.06	-0.44	0.26	0.52	0.40	0.15	0.00
	120	0.04	0.23	-0.65	0.74	0.76	0.65	0.27	0.05
Q2(2)	1	-0.37	0.55	-3.01	1.24	0.66	0.52	0.10	0.30
	6	-0.01	0.21	-0.86	0.89	0.60	0.40	0.02	0.05
	12	0.03	0.10	-0.97	0.53	0.43	0.16	0.02	0.01
	24	0.02	0.13	-0.55	0.76	0.57	0.30	0.02	0.02
	60	0.00	0.06	-0.43	0.34	0.47	0.32	0.06	0.00
	120	0.04	0.21	-0.65	0.61	0.75	0.64	0.26	0.05
A1(2)	1	-0.36	0.56	-3.06	2.25	0.65	0.49	0.01	0.31
	6	-0.01	0.31	-1.04	1.48	0.66	0.45	0.12	0.10
	12	0.03	0.22	-1.21	1.36	0.57	0.27	0.10	0.05
	24	0.02	0.22	-0.90	1.35	0.62	0.30	0.07	0.05
	60	0.01	0.10	-0.44	0.47	0.59	0.34	0.14	0.01
	120	0.04	0.27	-0.69	1.02	0.77	0.67	0.30	0.07
A2(2)	1	-0.35	0.55	-2.90	1.56	0.64	0.52	0.08	0.30
	6	0.00	0.25	-1.16	1.07	0.61	0.44	0.07	0.06
	12	0.03	0.16	-1.33	1.08	0.49	0.20	0.12	0.03
	24	0.02	0.17	-0.82	1.18	0.61	0.30	0.11	0.03
	60	0.00	0.08	-0.60	0.36	0.52	0.28	0.21	0.01
	120	0.03	0.25	-0.68	0.85	0.78	0.68	0.30	0.06

The table reports summary statistics for the in-sample fit of the evaluated models. The in-sample fit refers to the properties of the error-term $\epsilon_{t,\tau}$ in the yield curve observation equation $Y(X_t, \tau) = -\frac{c_\tau}{\tau} - \frac{b'_\tau}{\tau} X_t - X'_t \frac{A_\tau}{\tau} X_t + \epsilon_{t,\tau}$, where $\epsilon_{t,\tau} \sim N(0, R)$ and R is assumed to be diagonal. 'mean' is the average, 'st.dev' is the standard deviation, 'min' is the minimum and 'max' is the maximum estimation error. ρ_p denotes the autocorrelation at lag p and 'MAD' is the mean absolute deviation. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 9: Estimated factors: summary statistics

		mean	std	ρ_1	ρ_{12}	correlations								
						F1	F2	F3	PC1	PC2	PC3	level	slope	curv
Q1(3)	F1	-0.17	0.01	0.96	0.59	1			-0.66	**0.61**	0.35	-0.38	**0.77**	0.49
	F2	0.03	0.00	0.91	0.42	-0.82	1		0.60	-0.66	**0.15**	0.33	-0.85	**-0.07**
	F3	-0.18	0.01	0.98	0.79	-0.20	0.20	1	**0.85**	0.50	0.13	**0.97**	0.16	-0.37
Q2(3)	F1	-0.05	0.02	0.97	0.56	1			-0.41	**0.79**	0.37	-0.08	**0.85**	0.36
	F2	-0.07	0.01	0.98	0.78	-0.06	1		**-0.87**	-0.48	-0.09	**-0.98**	-0.13	0.40
	F3	1.13	0.00	0.73	0.09	0.39	0.32	1	-0.43	-0.02	**0.76**	-0.37	0.05	**0.79**
Q3(3)	F1	-0.09	0.01	0.96	0.56	1			-0.44	**0.75**	0.40	-0.12	**0.82**	0.41
	F2	0.00	0.00	0.89	0.34	0.79	1		-0.41	0.77	**-0.16**	-0.12	0.89	**-0.04**
	F3	0.08	0.01	0.98	0.80	0.04	0.08	1	**-0.89**	-0.42	-0.10	**-0.98**	-0.06	0.40
A1(3)	F1	0.02	0.08	0.99	0.82	1			**-0.92**	-0.32	-0.20	**-0.98**	0.04	0.30
	F2	0.29	0.43	0.92	0.43	-0.04	1		0.40	-0.44	**-0.73**	0.17	-0.49	**-0.71**
	F3	-0.02	0.05	0.95	0.42	0.01	-0.56	1	-0.34	**0.92**	-0.01	0.00	**0.96**	-0.03
A2(3)	F1	0.01	0.03	0.98	0.79	1			**-0.88**	-0.47	-0.07	**-0.98**	-0.12	0.42
	F2	-0.01	0.03	0.96	0.55	-0.09	1		-0.39	**0.84**	0.31	-0.05	**0.88**	0.27
	F3	-0.06	0.10	0.88	0.21	0.26	0.54	1	-0.42	0.07	**0.84**	-0.31	0.15	**0.87**
NS1(3)	F1	0.01	0.00	0.99	0.80	1			**0.88**	0.45	0.14	**0.99**	0.11	-0.33
	F2	0.00	0.00	0.96	0.50	-0.08	1		0.39	**-0.89**	-0.14	0.04	**-0.95**	-0.13
	F3	0.00	0.00	0.85	0.33	0.22	0.28	1	0.43	-0.05	**-0.69**	0.32	-0.19	**-0.89**
NS2(3)	F1	0.01	0.00	0.99	0.80	1			**0.87**	0.46	0.15	**0.99**	0.12	-0.32
	F2	0.00	0.00	0.96	0.50	-0.02	1		0.46	**-0.87**	-0.14	0.11	**-0.95**	-0.16
	F3	0.00	0.00	0.83	0.29	0.23	0.18	1	0.40	0.08	**-0.69**	0.33	-0.06	**-0.90**

Estimated factors: summary statistics (continued)

		mean	std	ρ_1	ρ_{12}	correlations								
						F1	F2	F3	PC1	PC2	PC3	level	slope	curv
Q1(2)	F1	0.12	0.01	0.98	0.78	1			**0.90**	0.41		**0.99**	0.06	
	F2	0.06	0.01	0.97	0.55	0.05	1		-0.33	**0.83**		0.01	**0.88**	
Q2(2)	F1	0.05	0.02	0.96	0.52	1			0.34	**-0.86**		0.00	**-0.90**	
	F2	0.07	0.01	0.98	0.78	-0.13	1		**0.87**	0.47		**0.98**	0.13	
A1(2)	F1	0.00	0.03	0.98	0.79	1			**-0.94**	-0.34		**-0.99**	0.02	
	F2	0.04	0.33	0.95	0.51	0.09	1		0.25	**-0.93**		-0.10	**-0.92**	
A2(2)	F1	0.00	0.01	0.98	0.78	1			**-0.88**	-0.47		**-0.98**	-0.12	
	F2	-0.01	0.05	0.95	0.51	-0.19	1		-0.30	**0.92**		0.05	**0.93**	

The table reports summary statistics for the estimated yield curve factors, X_t, in the state-space system where the observation equation is given by $Y(X_t,\tau) = -\frac{c_\tau}{\tau} - \frac{b'_\tau}{\tau}X_t - X'_t\frac{A_\tau}{\tau}X_t + \epsilon_{t,\tau}$, and where the state equation is given by $X_{t+1} = \mu + \Phi X_t + \Sigma\eta_{t+1}$. 'mean' refers to the average, 'st.dev' is the standard deviation, and ρ_p refers to the autocorrelation at lag p. F1, F2 and F3 denote the first, second and third component of X_t. PC1, PC2 and PC3 refer to the first, second and third principle component. 'Level', 'slope', and 'curv' are defined as the long end of the yield curve, $Y(120)$, the difference between the long and the short end, $[Y(120) - Y(1)]$, and $[Y(120) + Y(1) - 2Y(24)]$ respectively. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 10: Tests applied between models: Clark-West (CW) or Diebold-Mariano (DM)

<table>
<tr><th></th><th>Q1(3)</th><th>Q2(3)</th><th>Q3(3)</th><th>A1(3)</th><th>A2(3)</th><th>NS1(3)</th><th>NS2(3)</th><th>Q1(2)</th><th>Q2(2)</th><th>A1(2)</th><th>A2(2)</th></tr>
<tr><td>Q1(3)</td><td></td><td colspan="2">CW</td><td colspan="4">DM</td><td colspan="2">CW</td><td colspan="2">DM</td></tr>
<tr><td>Q2(3)</td><td>CW</td><td></td><td>CW</td><td colspan="4">DM</td><td>DM</td><td>CW</td><td colspan="2">DM</td></tr>
<tr><td>Q3(3)</td><td colspan="2">CW</td><td></td><td colspan="5">DM</td><td>CW</td><td colspan="2">DM</td></tr>
<tr><td>A1(3)</td><td colspan="3">DM</td><td></td><td>CW</td><td colspan="4">DM</td><td colspan="2">CW</td></tr>
<tr><td>A2(3)</td><td colspan="3">DM</td><td>CW</td><td></td><td colspan="5">DM</td><td>CW</td></tr>
<tr><td>NS1(3)</td><td colspan="5">DM</td><td></td><td>CW</td><td colspan="4">DM</td></tr>
<tr><td>NS2(3)</td><td colspan="5">DM</td><td>CW</td><td></td><td colspan="4">DM</td></tr>
<tr><td>Q1(2)</td><td>CW</td><td colspan="6">DM</td><td></td><td>CW</td><td colspan="2">DM</td></tr>
<tr><td>Q2(2)</td><td colspan="3">CW</td><td colspan="4">DM</td><td>CW</td><td></td><td colspan="2">DM</td></tr>
<tr><td>A1(2)</td><td colspan="3">DM</td><td>CW</td><td colspan="5">DM</td><td></td><td>CW</td></tr>
<tr><td>A2(2)</td><td colspan="3">DM</td><td colspan="2">CW</td><td colspan="4">DM</td><td>CW</td><td></td></tr>
</table>

This table outlines which test is used when comparing the forecasting performance of the evaluated model specifications. ‘CW’ refers to Clark-West and ‘DM’ to Diebold-Mariano test. The CW test is applied when testing the forecasts produced by nested models. In all other cases (non-nested models) the DM test is applied.

Table 11: MSPE ratios with the random walk over the period 1/1994 - 12/2000

	maturity	Q1(3)	Q2(3)	Q3(3)	A1(3)	A2(3)	NS1(3)	NS2(3)	Q1(2)	Q2(2)	A1(2)	A2(2)
1m forecast	1	0.90	1.80	0.93	0.85	0.80	0.84	**0.70**	2.23*	1.97*	2.29*	2.30*
	6	1.18	0.95	1.10	0.90	0.90	**0.83**	1.02	0.99	0.97	0.99	1.03
	12	0.84	0.76*	0.85	0.80*	**0.75***	0.82*	0.91	0.78*	0.82*	0.84	0.77*
	24	0.97	0.94	1.02	0.99	0.97	1.02	**0.86***	0.99	1.01	1.02	0.95
	60	0.99	0.98	1.03	1.03	1.05	1.12	1.01	1.00	1.03	**0.90***	1.03
	120	1.90*	1.71*	1.97*	2.23*	2.35*	1.10	**1.02**	1.68*	1.70*	1.65*	2.15*
trace MSFE		0.44	0.49	0.45	0.45	0.45	0.38	**0.36**	0.54	0.52	0.53	0.57
6m forecast	1	0.53	0.44	0.50	0.60	**0.36***	0.50*	0.62	0.45	0.43*	0.38*	0.41
	6	0.72	0.62	0.77	0.59	0.56	0.62	0.73	0.58	0.67	0.55	**0.55**
	12	0.76	0.67	0.83	0.60	0.64	0.67	0.70	0.64	0.73	0.64	**0.60**
	24	0.77	0.71*	0.85*	0.65*	0.69*	0.74	0.69*	0.69*	0.75*	0.66*	**0.64***
	60	0.90	0.87	0.98	0.81	0.91	0.83	0.81	0.85	0.92	**0.81**	0.90
	120	1.21	1.18	1.26*	1.20	1.39	0.91	**0.90**	1.13	1.21	1.09	1.36
trace MSFE		2.93	2.71	3.13	2.64	2.77	2.57	2.63	2.62	2.84	**2.50**	2.70
12m forecast	1	0.58	0.50	0.65	0.54	0.48	0.49	0.65	**0.45**	0.54	0.63	0.51
	6	0.68	0.58	0.78	**0.47**	0.53	0.48	0.63	0.52	0.65	0.63	0.54
	12	0.70	0.60	0.80	**0.46**	0.55	0.49	0.59	0.54	0.66	0.63	0.55
	24	0.73	0.64	0.81	**0.51**	0.62	0.55	0.60	0.59	0.69	0.65	0.60
	60	0.84	0.81	0.90	0.69	0.86	**0.68**	0.74	0.78	0.85	0.81	0.89
	120	1.06	1.06	1.09	1.01	1.30	**0.77**	0.85	1.01	1.07	1.05	1.32
trace MSFE		5.85	5.34	6.43	4.64	5.49	**4.41**	5.14	4.96	5.70	5.57	5.56

This table presents the forecast performance of the evaluated model specifications in terms of the ratio of the mean squared prediction error (MSPE) of each model to the MSPE of the random walk. Three forecast horizons are evaluated 1, 6, and 12 months ahead, for yields observed at maturities of 1, 6, 12, 24, 60 and 120 months. Bold figures denote the 'best' model forecast (with the smallest MSPE ratio) for a given maturity at a given forecast horizon. Stars '*' denote ratios that are significantly different from unity according to the Diebold-Mariano test at the 95% level of significance. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 12: Equally good forecasts

	maturity	Q1(3)	Q2(3)	Q3(3)	A1(3)	A2(3)	NS1(3)	NS2(3)	Q1(2)	Q2(2)	A1(2)	A2(2)
1m forecast	1	0.90		0.93	0.85	0.80		**0.70**				
	6	1.18	0.95	1.10	0.90	0.90	**0.83**		0.99	0.97	0.99	1.03
	12	0.84	0.76	0.85		**0.75**	0.82	0.91	0.78	0.82	0.84	
	24	0.97	0.94		0.99			**0.86**				
	60	0.99						1.01			**0.90**	
	120							**1.02**				
6m forecast	1		0.44	0.50		**0.36**		0.62	0.45	0.43	0.38	
	6	0.72	0.62	0.77			0.62	0.73	0.58	0.67		**0.55**
	12	0.76	0.67				0.67	0.70	0.64	0.73		**0.60**
	24						0.74	0.69	0.69	0.75		**0.64**
	60	0.90	0.87	0.98		0.91	0.83	0.81	0.85		**0.81**	
	120							**0.90**				
12m forecast	1		0.50	0.65	0.54	0.48	0.49	0.65	**0.45**		0.63	0.51
	6		0.58	0.78	**0.47**		0.48	0.63	0.52	0.65		
	12		0.60	0.80	**0.46**		0.49	0.59	0.54	0.66		
	24		0.64	0.81	**0.51**		0.55	0.60	0.59	0.69		
	60			0.90	0.69		**0.68**		0.78		0.81	0.89
	120				1.01		**0.77**					

This table presents equally good forecasts among all tested models according to the performed Diebold-Mariano or Clark-West statistical tests. Three forecast horizons are evaluated - 1, 6, and 12 months ahead, for yields observed at maturities of 1, 6, 12, 24, 60, 120 months. For a given maturity and forecasting horizon, bold figures denote the 'best' model forecast, i.e. lowest ratio of mean squared prediction errors (MSPE) of a particular model to the MSPE of the random walk. The presented results are not worse, according to the tests at 95% significance level, than the 'best' performing model. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 13: Equally good forecasts among three-factor models

	maturity	Q1(3)	Q2(3)	Q3(3)	A1(3)	A2(3)	NS1(3)	NS2(3)	Q1(2)	Q2(2)	A1(2)	A2(2)
1m forecast	1	0.90		0.93	0.85	0.80		**0.70**				
	6	1.18	0.95	1.10	0.90	0.90	**0.83**					
	12	0.84	0.76	0.85		**0.75**	0.82	0.91				
	24	0.97	0.94		0.99			**0.86**				
	60		**0.98**	1.03	1.03	1.05		1.01				
	120							**1.02**				
6m forecast	1		0.44	0.50		**0.36**		0.62				
	6	0.72	0.62	0.77		**0.56**	0.62	0.73				
	12	0.76	0.67	0.83	**0.60**		0.67	0.70				
	24	0.77	0.71	0.85	**0.65**			0.69				
	60	0.90	0.87	0.98	0.81	0.91		**0.81**				
	120							**0.90**				
12m forecast	1	0.58	0.50	0.65		**0.48**	0.49	0.65				
	6		0.58	0.78	**0.47**		0.48	0.63				
	12		0.60	0.80	**0.46**		0.49	0.59				
	24		0.64	0.81	**0.51**		0.55	0.60				
	60			0.90	0.69		**0.68**					
	120				1.01		**0.77**					

This table presents equally good forecasts among all tested three-factor models according to the performed Diebold-Mariano or Clark-West statistical tests. Three forecast horizons are evaluated - 1, 6, and 12 months ahead, for yields observed at maturities of 1, 6, 12, 24, 60, 120 months. For a given maturity and forecasting horizon, bold figures denote the 'best' model forecast, i.e. lowest ratio of mean squared prediction errors (MSPE) of a particular model to the MSPE of the random walk. The presented results are not worse, according to the tests at 95% significance level, than the 'best' performing model. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 14: Equally good forecasts among maximally-flexible models

	maturity	Q1(3)	Q2(3)	Q3(3)	A1(3)	A2(3)	NS1(3)	NS2(3)	Q1(2)	Q2(2)	A1(2)	A2(2)
1m forecast	1	0.90			0.85		**0.84**					
	6	1.18			0.90		**0.83**		0.99		0.99	
	12				0.80		0.82		**0.78**		0.84	
	24	**0.97**			0.99		1.02				1.02	
	60	0.99									**0.90**	
	120						**1.10**					
6m forecast	1						0.50		0.45		**0.38**	
	6	0.72			0.59		0.62		0.58		**0.55**	
	12	0.76			**0.60**		0.67		0.64			
	24	0.77			**0.65**				0.69			
	60	0.90					0.83		0.85		**0.81**	
	120						**0.91**				1.09	
12m forecast	1				0.54		0.49		**0.45**		0.63	
	6				**0.47**		0.48		0.52			
	12				**0.46**		0.49		0.54			
	24				**0.51**		0.55		0.59			
	60				0.69		**0.68**		0.78		0.81	
	120				1.01		**0.77**					

This table presents equally good forecasts among all tested maximally-flexible models according to the performed Diebold-Mariano or Clark-West statistical tests. Three forecast horizons are evaluated - 1, 6, and 12 months ahead, for yields observed at maturities of 1, 6, 12, 24, 60, 120 months. For a given maturity and forecasting horizon, bold figures denote the 'best' model forecast, i.e. lowest ratio of mean squared prediction errors (MSPE) of a particular model to the MSPE of the random walk. The presented results are not worse, according to the tests at 95% significance level, than the 'best' performing model. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Table 15: Equally good forecasts among diagonal models

	maturity	Q1(3)	Q2(3)	Q3(3)	A1(3)	A2(3)	NS1(3)	NS2(3)	Q1(2)	Q2(2)	A1(2)	A2(2)
1m forecast	1					0.80		**0.70**				
	6		0.95			**0.90**		1.02		0.97		
	12		0.76			**0.75**		0.91		0.82		
	24		0.94					**0.86**				
	60		**0.98**			1.05		1.01				1.03
	120							**1.02**				
6m forecast	1		0.44			**0.36**		0.62		0.43		
	6		0.62					0.73		0.67		**0.55**
	12		0.67					0.70		0.73		**0.60**
	24							0.69		0.75		**0.64**
	60		0.87			0.91		**0.81**		0.92		0.90
	120							**0.90**				
12m forecast	1		0.50			**0.48**		0.65		0.54		
	6		0.58			**0.53**		0.63		0.65		0.54
	12		0.60					0.59		0.66		**0.55**
	24		0.64			0.62		**0.60**				0.60
	60					0.86		**0.74**				0.89
	120							**0.85**				

This table presents equally good forecasts among all tested diagonal models according to the performed Diebold-Mariano or Clark-West statistical tests. Three forecast horizons are evaluated - 1, 6, and 12 months ahead, for yields observed at maturities of 1, 6, 12, 24, 60, 120 months. For a given maturity and forecasting horizon, bold figures denote the 'best' model forecast, i.e. lowest ratio of mean squared prediction errors (MSPE) of a particular model to the MSPE of the random walk. The presented results are not worse, according to the tests at 95% significance level, than the 'best' performing model. The model classification scheme $Mn(k)$ denotes $M = \{Q, A, NS\}$ referring to **Q**uadratic, **A**ffine and **N**elson-**S**iegel models, with $n = \{1, 2, 3\}$ denoting the model variant: maximally flexible; independent-factors or 'triangular' specification, respectively. k counts the number of yield curve factors included in the examined model variants, i.e. $k = \{2, 3\}$.

Figure 1: Fitted and observed yield curves on randomly selected dates - quadratic three- and two-factor models

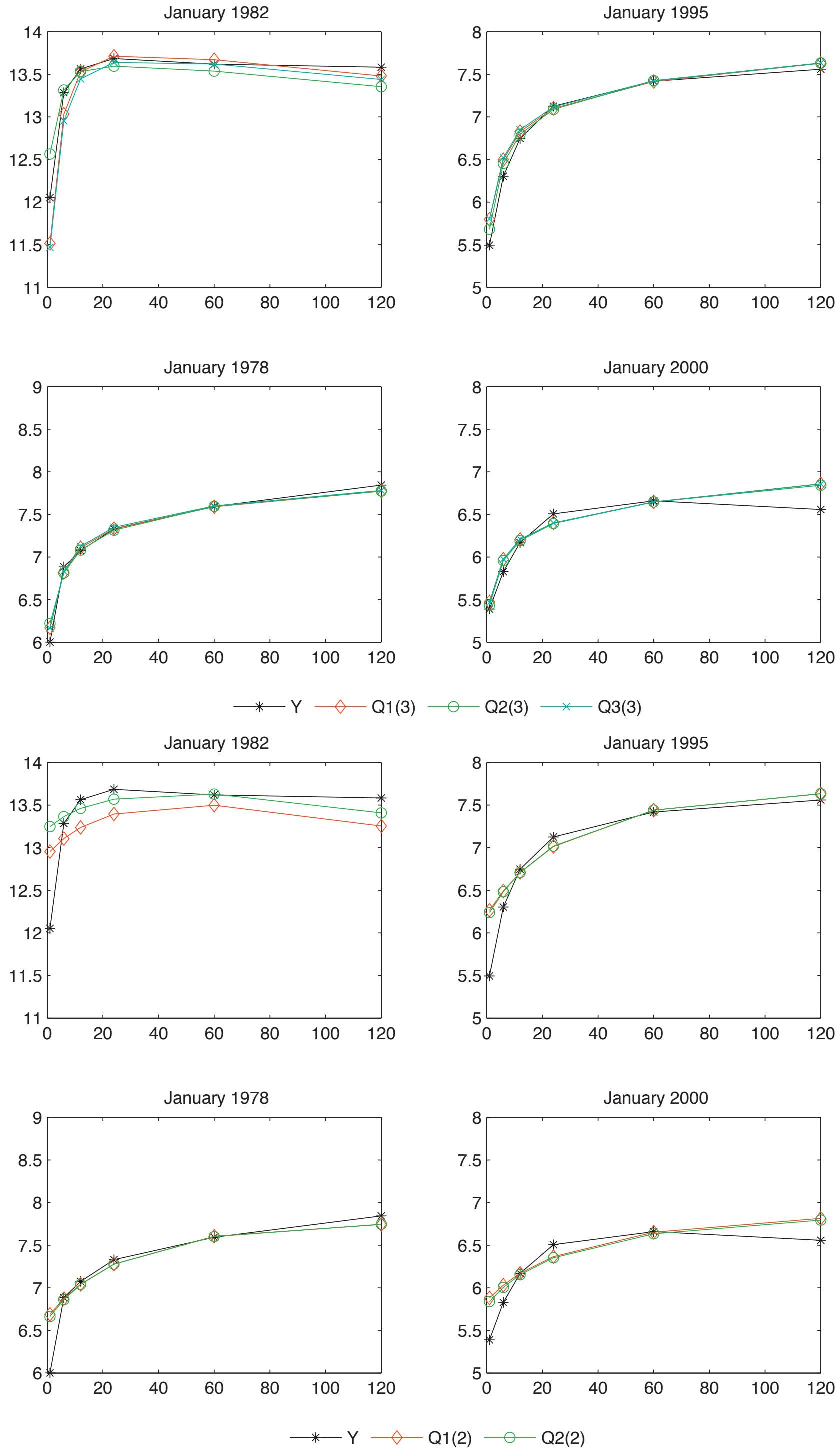

Figure 2: Fitted and observed yield curves on randomly selected dates - affine three- and two-factor models

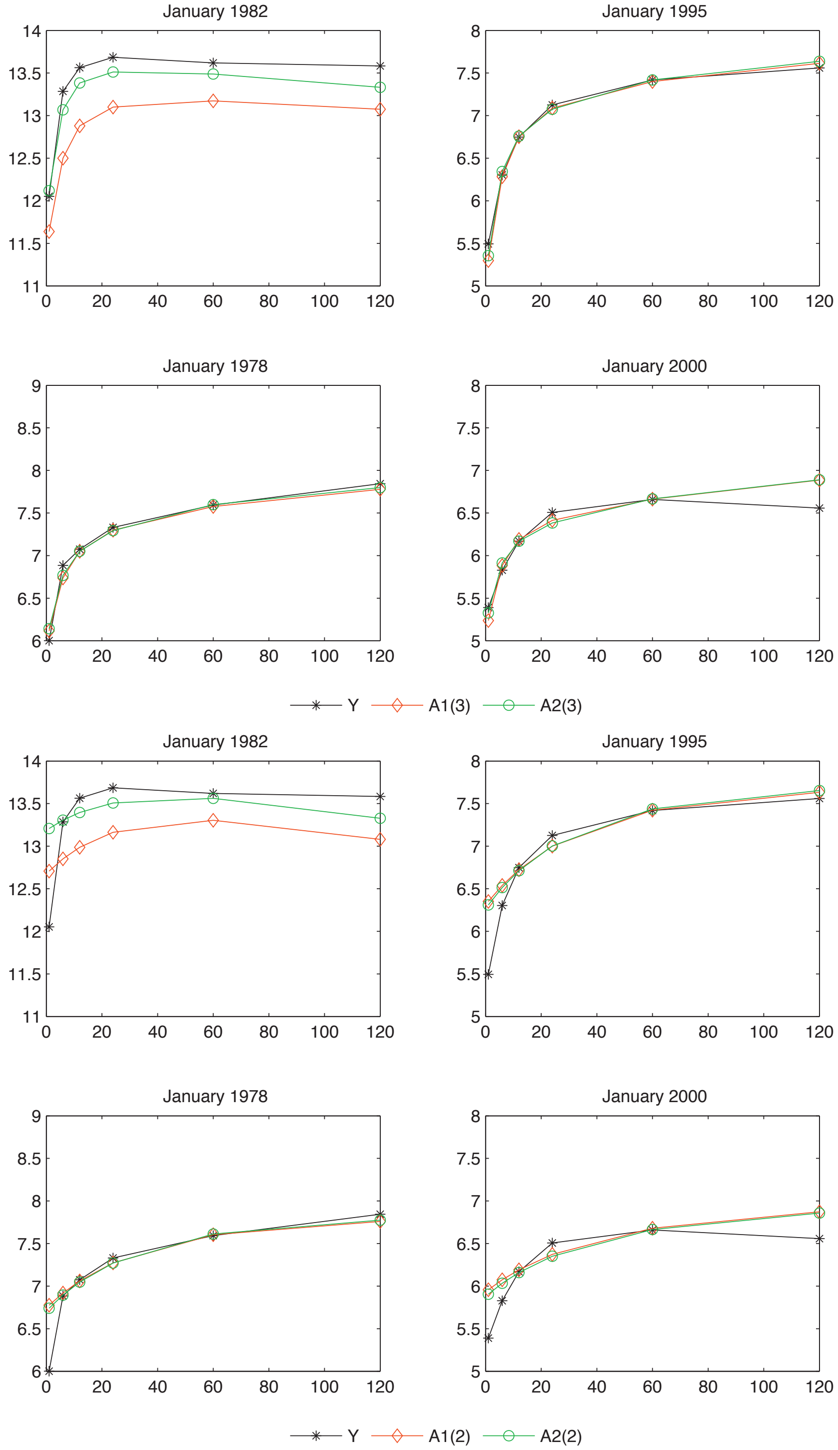

Figure 3: Fitted and observed yield curves on randomly selected dates - three-factor Nelson-Siegel models

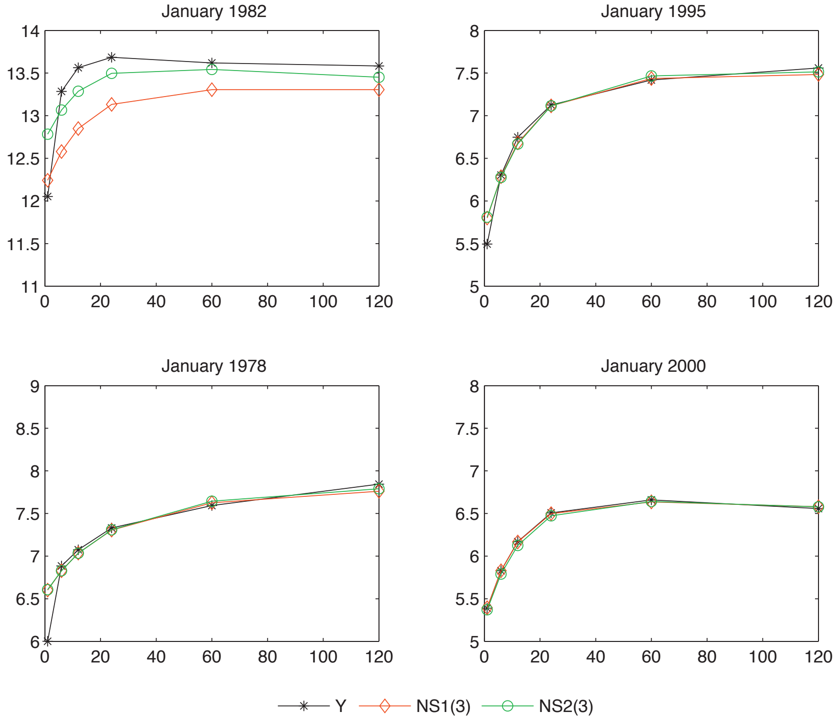

Figure 4: Standardized factors and principal components - quadratic models

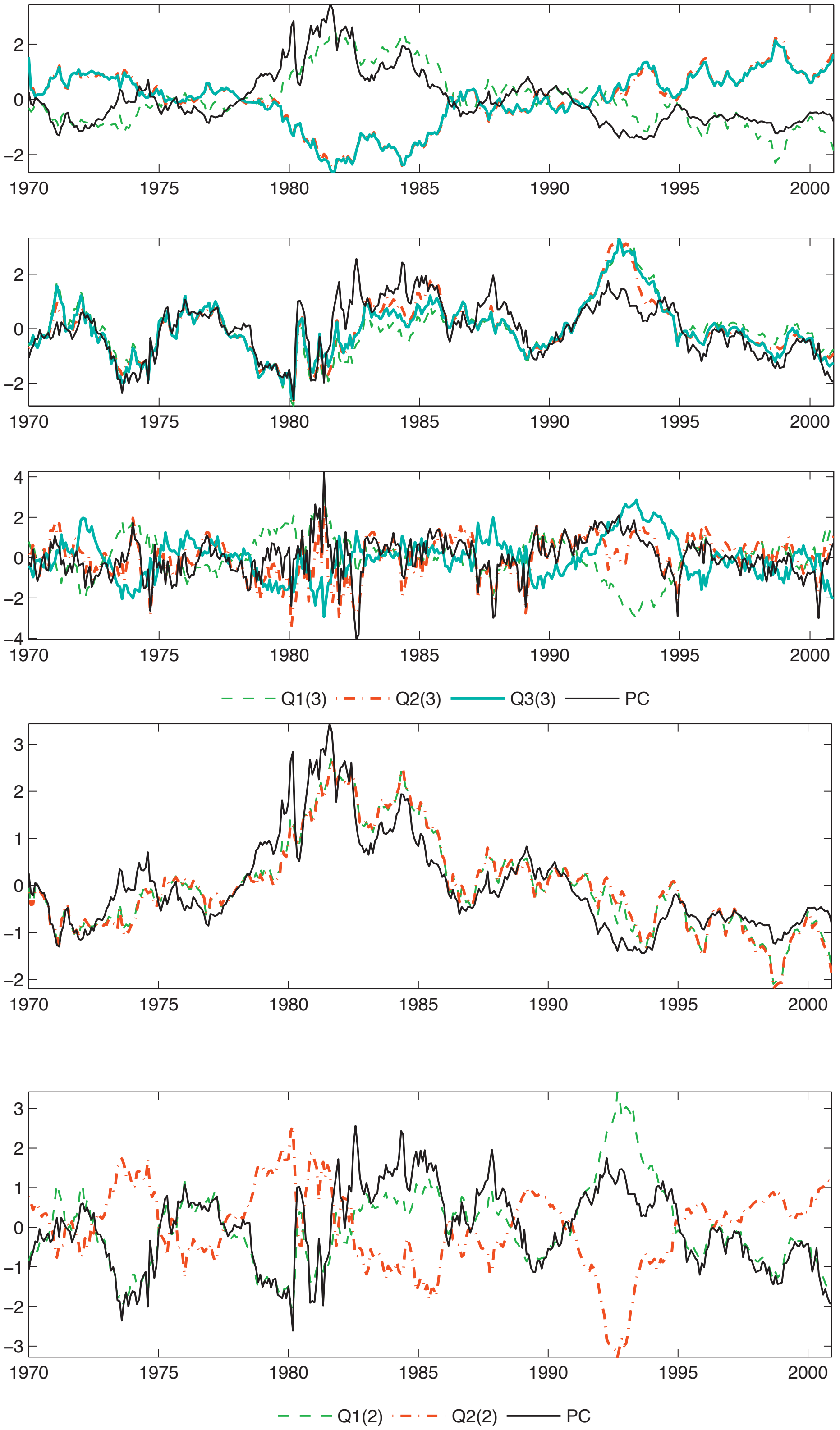

Figure 5: Standardized factors and principal components - affine models

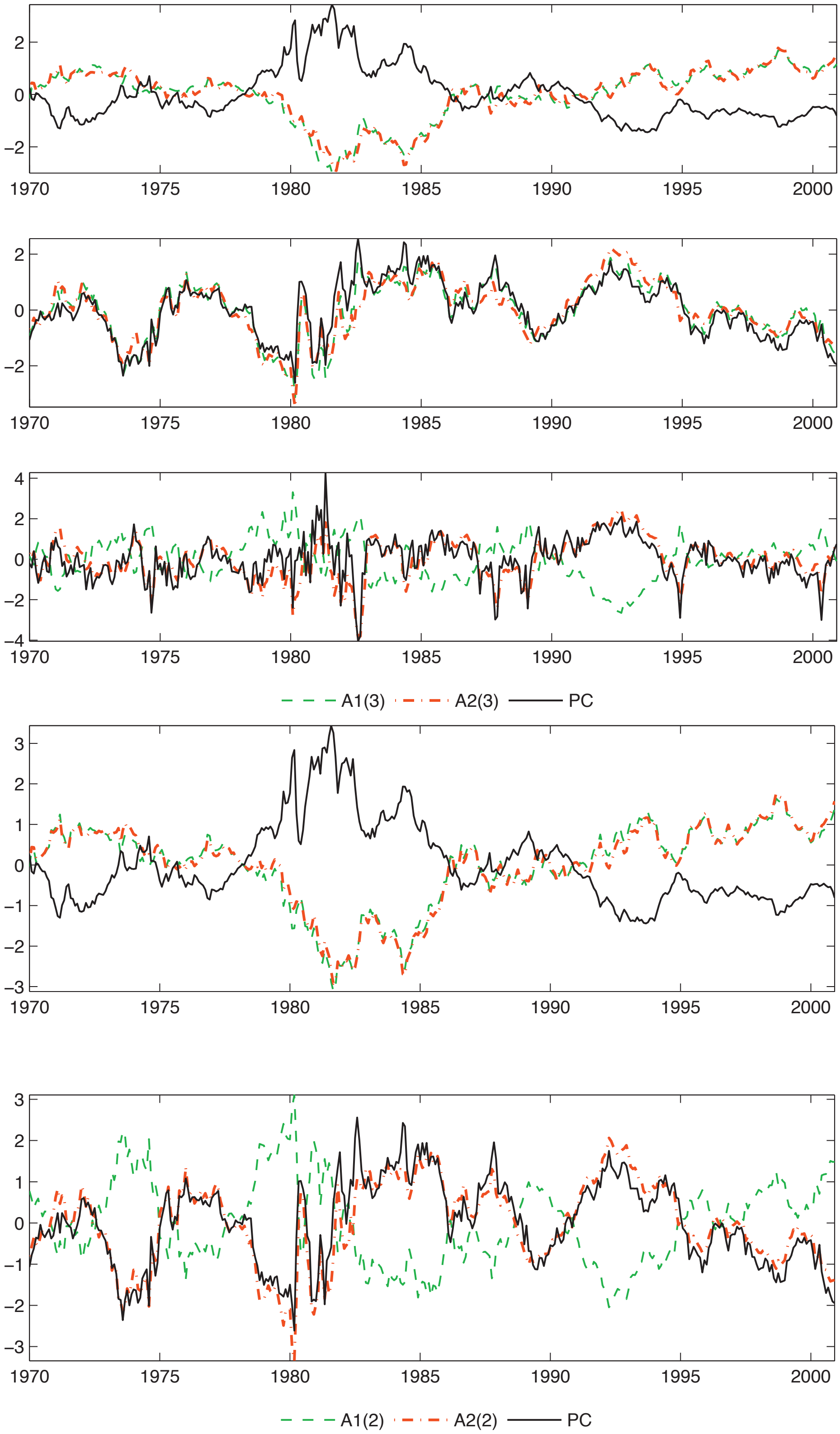

Figure 6: Standardized factors and principal components - Nelson-Siegel models