Forecasting the Yield Curve in a Data-Rich Environment using the Factor-Augmented Nelson-Siegel Model

Peter Exterkate[*]   Dick van Dijk   Christiaan Heij

Patrick J.F. Groenen

Econometric Institute, Erasmus University Rotterdam

Abstract

Various ways of extracting macroeconomic information from a data-rich environment are compared with the objective of forecasting yield curves using the Nelson-Siegel model. Five issues in factor extraction are addressed, namely, selection of a subset of the available information, incorporation of the forecast objective in constructing factors, specification of a multivariate forecast objective, data grouping before constructing factors, and selection of the number of factors in a data-driven way. Our empirical results show that each of these features helps to improve forecast accuracy, especially for the shortest and longest maturities. The data-driven methods perform well in relatively volatile periods, when simpler models do not suffice.

Keywords: yield curve prediction, Nelson-Siegel model, factor extraction, variable selection

JEL Classification: C53, E43, E44.

---

[*]Corresponding author. Address: Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands; email: exterkate@ese.eur.nl; phone: +31-10-4081264; fax: +31-10-4089162.

1 Introduction

Forecasting the yield curve is of much practical interest, not only for individual investors, but also for pension funds, central banks, and policy makers. Despite the relevance of this issue, surprisingly little research effort has been spent on this issue until recently. This may be partly due to negative results obtained by early studies on yield curve forecasting, such as Duffee (2002). That study investigates the forecasting performance of affine term structure models, which postulate that yields evolve as affine functions of a limited number of latent risk factors, see Vašíček (1977), Cox et al. (1985), Duffie and Kan (1996), and Dai and Singleton (2000). Duffee (2002) dismisses this entire class of models for forecasting purposes by showing that the forecasts obtained from affine models are inferior to random walk (no-change) forecasts.

More positive results have emerged recently based on the framework of Nelson and Siegel (1987). Originally intended to describe cross-sectional aspects of yield curves, the Nelson-Siegel model imposes a parsimonious three-factor structure on the links between yields of different maturities, where the factors can be interpreted as level, slope and curvature. Diebold and Li (2006) find that a dynamic reformulation of this model provides forecasts that outperform the random walk and various alternative forecasting approaches, see also Christensen et al. (2007).

Both the Nelson-Siegel and affine models are essentially purely statistical models of the yield curve. At the same time, it is widely believed that yield curve dynamics are closely linked to macroeconomic developments, for various reasons. For example, central banks around the world use short-term interest rates as their policy instruments, and it is widely recognized that their actions respond to macroeconomic aggregates such as inflation and output; see Taylor (1993). Because longer-term interest rates can be regarded as a weighted average of expected future short-term rates, it is plausible

that the entire yield curve responds to macroeconomic shocks. Such a link does also exist in the reverse direction. Economic agents respond to changing interest rates by altering their investment plans and by adjusting their inflation expectations. Not surprisingly then, several recent studies have developed extensions of yield curve models that incorporate macro variables, in an attempt to capture their interaction, see Ang and Piazzesi (2003), Dewachter and Lyrio (2006), Hördahl et al. (2006), Rudebusch and Wu (2008), among others. Diebold et al. (2006) propose a way to include macroeconomic factors in the Nelson-Siegel model, and they find clear evidence that macro aggregates have a statistically significant effect on yields. The analysis in Diebold et al. (2006) (and in the other studies cited above) is purely based on in-sample fit, however, and does not consider out-of-sample forecasting.

In this paper we examine several important aspects related to the inclusion of macroeconomic variables in the Nelson-Siegel model from a forecasting perspective.

First, in today's data-rich environment, a natural question is which macro factors to include in the model. Diebold et al. (2006) use three specific variables, intended to represent the level of real economic activity, monetary policy, and inflation, respectively. Arguably, many more macro variables may influence the evolution of the yield curve. However, including a large number of individual variables leads to an abundance of parameters to be estimated. A natural possibility is to use factors extracted from a large panel of specific macro variables. Indeed, De Pooter et al. (2007) find that including a small number of principal components leads to an improvement in forecast accuracy, compared to the use of specific individual variables.[1] Here we examine this issue in more detail by comparing a number of data-based variable selection methods with several approaches to construct factors from a large panel of variables. The methods for selecting specific macroeconomic variables in a data-driven way are

_____

[1]Mönch (2008) finds similar results when augmenting an affine model with such principal components.

3

based on least angle regression (LARS) or multiresponse sparse regression (MRSR), as proposed by Efron et al. (2004) and Similä and Tikka (2006), respectively. The factor construction methods include Principal Component Analysis (PCA), next to a number of alternative approaches discussed below.

Second, we investigate whether it is useful to take the forecast objective explicitly into account when constructing the macroeconomic factors. While this is a natural thing to do when selecting specific individual variables, this is not the case for methods that construct factors from a large panel of variables. In particular, principal component analysis, which is by far the most popular factor construction method, renders the same macro factors regardless of which series we wish to forecast. We examine three alternative approaches that do take the series to be forecasted into consideration in the construction of the macro factors, namely Partial Least Squares (PLS), as introduced by Wold (1966), Principal Covariate Regression (PCovR) proposed by Heij et al. (2007), and PCA based on variables selected by thresholding rules, proposed by Bai and Ng (2008). In the first two approaches, factors are constructed using all available macro variables, but with weights depending on their degree of comovement with the forecast objective. In the latter approach, principal components are taken only from those variables that are most correlated with the variable that we aim to predict. Hence, with this method we also address the issue whether or not it is desirable to include all available data in PCA, see also Boivin and Ng (2006).

Third, while there may be little doubt that the yield curve is linked to the macroeconomy, it is plausible that different characteristics of the yield curve are related to different macroeconomic variables. We examine this issue by using the variable selection and factor construction methods in two different ways. Specifically, we select and construct macro factors for the three Nelson-Siegel factors jointly and for the level, slope and curvature factors separately.

Fourth, we explore whether it pays off to construct factors from groups of related macro variables, instead of one large pool of all available variables. Given that real economic activity, monetary policy, and inflation appear to be the most relevant macroeconomic dimensions for the yield curve, it may be worthwhile to construct factors that are explicitly related to these characteristics. We implement this idea for both PCA and PCovR.

Fifth, in addition to the question which macro factors to include in a yield curve model, it is also relevant to ask how many factors should be included. Here we compare the predictive accuracy of models with a fixed number of three factors to models with a varying number of macro factors based on historical forecasting performance.

We address the issues listed above empirically, by examining the out-of-sample forecasting performance of the Factor-Augmented Nelson-Siegel (FANS) model for the US yield curve over the period from January 1994 until December 2003, for forecast horizons of one, three, six and twelve months ahead. Our results show that data-driven methods are particularly useful in volatile times, when simpler methods fail. The preselected variables suggested by Diebold et al. (2006) provide the best forecasts only for predicting medium-term yields (15 to 60 months) at the longest horizon considered in this study (12 months). For longer maturities, it is better to form groups of related variables and then extract factors from these groups, preferably by using principal covariate regression — thus, explicitly considering the forecast objective when constructing factors. For shorter maturities, and for medium maturities at shorter horizons, it is even better to extract principal components or covariates from all available information. Methods that treat the three Nelson-Siegel factors jointly generally outperform methods that treat these factors separately. Finally, varying the number of macro factors based on recent historical performance leads to an additional improvement in forecast accuracy.

5

In the remainder of this article, Section 2 begins with a description of the Nelson-Siegel model and the extension by Diebold et al. (2006). Furthermore, the techniques of least angle regression, multiresponse sparse regression, principal covariate regression, partial least squares, and hard and soft thresholding are discussed. Section 3 describes the data on U.S. zero-coupon yields and macroeconomic aggregates, as well as details of our forecasting procedure. Section 4 contains the empirical forecasting results, and Section 5 concludes.

2 Methods

This section reviews the Nelson-Siegel model and the techniques that we employ to construct macroeconomic factors to be included in this model.

2.1 Nelson-Siegel model

Nelson and Siegel (1987) propose a parsimonious model for describing the yield curve. With minor modifications as explained in Diebold and Li (2006), the model is

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

where $y_t\left(\tau\right)$ is the yield at time t for a maturity of τ months. As discussed in Diebold and Li (2006), the β_{1t}, β_{2t}, and β_{3t} parameters can be interpreted as level, slope, and curvature factors, respectively. Further, the parameter λ_t determines the rate of decay of the loading for the slope factor β_{2t} and the maturity at which the loading for the curvature factor β_{3t} attains its maximum value.

If we consider a fixed set of maturities $\left(\tau_1, \tau_2, \ldots, \tau_m\right)$ and denote the corresponding vector of observed yields at time t by $y_t = \left(y_t\left(\tau_1\right), y_t\left(\tau_2\right), \ldots, y_t\left(\tau_m\right)\right)'$ and the

parameter vector by $\beta_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$, and if we add an error term to Equation (1), we find

$$y_t = \Lambda_t \beta_t + \varepsilon_t, \tag{2}$$

where Λ_t depends on λ_t only, for given τ.

Diebold and Li (2006) interpret Equation (2) as the measurement equation of a state space model. The state variables β_t are assumed to evolve according to a first-order vector autoregressive process (with mean μ):

$$\beta_t - \mu = A (\beta_{t-1} - \mu) + \eta_t. \tag{3}$$

The disturbances ε_t and η_t are assumed to be zero-mean white noise and to be mutually uncorrelated. The covariance matrix of ε_t is commonly assumed to be diagonal, see, for example, Diebold and Li (2006) and Diebold et al. (2006). The covariance matrix Q of η_t is left unrestricted in those studies. As Diebold and Li (2006) find that one lag is sufficient to describe the dynamics of β_t, no further lags are included in the transition equation (3). Assuming normally distributed error terms, maximum likelihood estimates and forecasts are obtained using the Kalman filter.

In most studies, λ_t is assumed constant and its value is fixed by the researcher, as in Diebold and Li (2006). However, a constant λ_t can also be estimated along with the other model parameters using the Kalman filter, as in Diebold et al. (2006) and De Pooter et al. (2007). Koopman et al. (2010) propose ways of allowing for time-varying λ_t. For simplicity, we will use a time-invariant λ, which is estimated along with the other model parameters.[2]

Now assume that, at every time t, a large number of k macroeconomic variables are

---

[2]That is, λ is kept fixed for the estimation period. As we discuss below, the Nelson-Siegel model is estimated using a ten-year rolling window, so that the estimated value of λ will actually vary from one window to the next.

available, denoted by x_t. For reasons of parsimony, we wish to summarize this large amount of information by a limited number of p factors, f_t. These factors can, for example, be obtained by preselection of a subset of variables from x_t, as in Diebold et al. (2006). Alternatively, the factors can be extracted from x_t using principal component analysis as in De Pooter et al. (2007). We describe alternative methods for constructing these factors below, in Sections 2.2—2.6.

For now, assume that f_t is available and that it is normalized to have mean zero. We follow Diebold et al. (2006) in their procedure for incorporating this information into the Nelson-Siegel framework. The observation equation (2) remains unchanged. In the state equation (3), f_t is appended to the state vector β_t and the dimensions of A, η_t, and Q are increased as appropriate.

Introducing macroeconomic information into the model in this manner leads to a substantial increase in the number of parameters. For example, the inclusion of three macro factors increases the dimension of both A and Q from 3×3 to 6×6. As Q is symmetric, the factor-augmented model has $(36 - 9) + (21 - 6) = 42$ additional parameters. To avoid problems of overfitting and numerical difficulties, we impose the following two restrictions. First, the VAR transition matrix is restricted to have the following structure:

$$A = \left(\begin{array}{c|c} \text{diagonal} & \text{unrestricted} \\ \hline \text{zero} & \text{diagonal} \end{array} \right),$$

where the blocking corresponds to the partitioning of the state vector into β_t and f_t. In particular, this restriction implies that we do not model any yields-to-macro feedbacks. Second, the covariance matrix Q is restricted to be diagonal. Diebold et al. (2006) tested and rejected both of these restrictions. However, as mentioned before, their analysis was strictly based on in-sample criteria. As our focus is on out-of-sample

8

forecasting, the notion of parsimony is an important consideration. As we show in Section 4, imposing these restrictions improves the forecasting performance.

For convenience, the model considered in this study is shown below:

$$
\begin{aligned}
y_t &= \Lambda\beta_t + \varepsilon_t, & \varepsilon_t &\sim \mathcal{NID}\left(0, \operatorname{diag}\left(\sigma_1^2, \ldots, \sigma_m^2\right)\right), \\
\begin{pmatrix} \beta_t - \mu \\ f_t \end{pmatrix} &= A \begin{pmatrix} \beta_{t-1} - \mu \\ f_{t-1} \end{pmatrix} + \eta_t, & \eta_t &\sim \mathcal{NID}\left(0, Q\right).
\end{aligned}
$$

(4)

We shall refer to this model as the Factor-Augmented Nelson-Siegel (FANS) model.

Below, we describe methods to extract the factors f_t from the macroeconomic variables x_t. To facilitate the discussion, we define

$$
B^+ = \begin{pmatrix} \beta_2' \\ \beta_3' \\ \vdots \\ \beta_T' \end{pmatrix}, \quad B = \begin{pmatrix} \beta_1' \\ \beta_2' \\ \vdots \\ \beta_{T-1}' \end{pmatrix}, \quad X = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_{T-1}' \end{pmatrix}, \quad F = \begin{pmatrix} f_1' \\ f_2' \\ \vdots \\ f_{T-1}' \end{pmatrix},
$$

where T is the length of the estimation sample. To rule out scale effects, every column of X (that is, the time series of observations on each variable separately) is normalized to have mean zero and unit variance over the estimation window. We denote the number of original macro variables (columns of X) by k, and the number of macro factors (columns of F) by p.

In essence, all of the methods for constructing factors considered in this study boil down to choosing a $k \times p$ matrix W and defining $F = XW$. That is, the factors that we use in the FANS model are linear combinations of observed macroeconomic aggregates. The following sections describe three methods for choosing these combinations. Common features are that they achieve a dimension reduction (f_t contains far

9

fewer elements than x_t; that is, $p \ll k$) and that f_{t-1}, together with β_{t-1}, should have predictive power for β_t.

One more technical comment is in place at this stage. Each of the factor construction methods described below requires, in principle, full observability of B, whereas B is latent in our application. We choose to overcome this problem by the following straightforward approach. First, we estimate the FANS model without any macro factors, by maximum likelihood using the Kalman filter. The resulting estimates of β_t (in Kalman filtering terminology: the smoothed state vectors) form the matrices B and B^+, which are used to build response variables in constructing the macro factors f_t. The FANS model is then reestimated using these factors. Although this procedure could be iterated, this approach is not pursued here for computational considerations.

2.2   Least angle regression

The Least Angle Regression (LARS) methodology, originating with Efron et al. (2004), can be used to select a limited number of informative variables out of a large group. The main idea is to "add" predictors to the model one at a time, starting with the predictor that correlates most with the response variable. This predictor's coefficient is increased from its starting value zero, up to the point where the residual is equally correlated with the predictor chosen initially and a second predictor. This second predictor is added to the "most correlated" set, and the coefficients on both predictors in this set are now simultaneously increased in such a way as to keep the residual equally correlated with the two predictors. As soon as a third predictor shows equal correlation, it also enters the "most correlated" set, and so on, until either the residual is zero or all predictors have entered.

In any stage of this procedure, only the predictors in the "most correlated" set have nonzero coefficients. Hence, LARS can be used as a variable selection method by

10

stopping the algorithm after a prespecified number of predictors have been selected. This procedure closely approximates the more well-known Lasso method proposed by Tibshirani (1996); see Efron et al. (2004) for a discussion of this similarity.

In our setting, we wish to select a small number of variables (columns) from X which, together with B, have predictive power for B^+. To this end, we employ a simple two-step procedure: we first perform an auxiliary OLS regression in $B^+ = \iota\alpha' + B\Delta + E$. (We use the symbol ι to denote a vector of ones, α is a vector of constants, Δ is a matrix of regression coefficients, and E is a matrix of disturbance terms.) The residuals R from this regression are used as response variables in the LARS algorithm, in order to explain those features of the yield factors that remain after correcting for autoregressive effects.

As R is multivariate in our setup, the most obvious application of the LARS algorithm is to feed the columns of R into the algorithm one at a time. An alternative approach is "multiresponse sparse regression" (MRSR), proposed by Similä and Tikka (2006) as an extension of LARS that allows for a multivariate response variable.

To apply LARS in a multivariate setting, an extension of the correlation concept is required to make the condition "equally correlated" meaningful. Denote the fitted value of R, based on the first m regressors chosen, by $\hat{R}_m$ (with $\hat{R}_0 = 0$), and denote the j-th column of X by $x_{(j)}$. Following Similä and Tikka (2006), the role of correlations in the description of the univariate case above is now played by $\left|\left|\left(R - \hat{R}_m\right)' x_{(j)}\right|\right|$, where $||v||$ represents the L_2 vector norm $(\sum_i v_i^2)^{1/2}$. No other changes to the procedure are needed. An efficient algorithm to find the order in which variables are added is presented in Similä and Tikka (2006).

We use both the univariate and the multivariate variant of the LARS algorithm to select a predefined number of explanatory variables from a large panel of macro data. The selected predictors are used as the macro factors f_t in the FANS model.

2.3 Principal component regression and principal covariate regression

Heij et al. (2007) propose principal covariate regression (PCovR) as an alternative to principal component regression (PCR).

In PCR, principal components are first extracted from a group of predictors and then used as regressors. To obtain p principal components, the k predictors in the matrix X are "summarized" in $p \ll k$ factors by minimizing $||X - XUV||$ over the $k \times p$ matrix U and the $p \times k$ matrix V. The desired factors are the columns of XU. The method then proceeds as a standard OLS regression of the (univariate) dependent variable z on a constant and XU. That is, the objective function $||z - \alpha\iota - XU\gamma||$ is minimized over the scalar α and the vector γ.

Heij et al. (2007) argue that the failure to take the prediction objective into account when constructing the factors is a drawback of PCR. To overcome this problem, they combine the two steps of PCR into one objective function: in the same notation as above, they minimize

$$ w\,||z - \alpha\iota - XU\gamma||^2 / ||z||^2 + (1-w)\,||X - XUV||^2 / ||X||^2 \,, \qquad (5) $$

where $w \in [0,1]$ is a tuning parameter that governs the relative weight placed on each of the two objectives. Thus, the aims of good prediction and adequate use of the data are balanced in the PCovR objective (5); setting w at a higher value means that more weight is placed on predicting z relative to summarizing X, whereas choosing $w = 0$ amounts to standard PCR.

An obvious multivariate extension of Objective (5) to our problem is to minimize

$$ w\,||R - XU\Gamma||^2 / ||R||^2 + (1-w)\,||X - XUV||^2 / ||X||^2 \,, \qquad (6) $$

where Γ is now a matrix and R is, as before, the matrix of residuals from the OLS regression of B^+ on B and a constant.

Direct minimization of Objective (6) can be done using two singular value decompositions, as outlined in Heij et al. (2007). To operationalize this procedure, we need to specify a value for w. The high-dimensional nature of X leads to overfitting if w is chosen too large; see Heij et al. (2006) for a discussion of this aspect. The upper bound that they propose evaluates to about $w \leq 0.1$ for our problem. In order to make PCovR sufficiently different from standard PCR, we do not want to set w too small either; therefore, we fix $w = 0.1$.

2.4 Partial least squares

Wold (1966) proposed partial least squares (PLS) as a technique for finding orthogonal linear combinations of the variables in X that have predictive power for a univariate response, say, z. That is, in contrast to PCR, the PLS factors are constructed to explain the variance of z rather than X. Following Garthwaite (1994), we describe PLS in terms of sequential regressions.

Assume that X and z have mean zero. We first regress z on each column $x_{(j)}$ of X, yielding fitted vectors $\hat{z}_{(j)}$. The first factor is then constructed as $f_1 = \sum_j w_j \hat{z}_{(j)}$, where the weights are proportional to the variances of the columns of X, $w_j = x'_{(j)} x_{(j)}$, such that $f_1 = \sum_j \mathrm{cov}\left(x_{(j),z}\right) x_{(j)}$. Hence the first factor is a weighted average of the macro variables x_t, with weights depending on their covariance with the variable we aim to forecast. Now, replace both z and X by their residuals from regressions on f_1. The second factor f_2 is then found by applying the same procedure to the "new" z and X. Continuing in this manner, we can sequentially construct $f_3, \ldots, f_p$.

This procedure can be copied almost verbatim for a multivariate response Z; details can be found in Garthwaite (1994).

To apply this method in our context, we use a similar two-step approach as for the LARS algorithm. That is, we first find the residuals R from a regression of B^+ on B and a constant. The three columns of R are then used, either one at a time or all together, as response vectors z in the PLS algorithm. The constructed factors are used together as f_t in the FANS model.

2.5 Hard thresholding

Bai and Ng (2008) propose hard thresholding as a simple method for variable selection. In their univariate setting, consider forecasting a univariate response z using its own lag z^- and the columns of X. To select the most relevant columns, estimate the linear model

$$z = \alpha \iota + \delta z^- + \gamma_j x_{(j)} + \varepsilon \tag{7}$$

with OLS, for all columns of X, $j = 1, 2, \ldots, k$. We then select those columns of X for which the t test rejects $\gamma_j = 0$, at a prespecified significance level.

Following the suggestion in Bai and Ng (2008), to construct p predictors, we start by selecting p^* variables, with $p^* > p$. The first p principal components extracted from this subset of selected variables are then used as predictors.

In our multivariate context, we estimate

$$B^+ = \iota \alpha' + B \Delta + x_{(j)} \gamma_j' + E \tag{8}$$

with OLS, and compute the Wald statistic W_j for the test $\gamma_j = 0$.[3] We select the p^* variables for which this null hypothesis is rejected at 5% significance, and we use p principal components extracted from this set as macro factors in the FANS model.

---

[3]Note that this Wald test involves a zero restriction on one parameter in each of multiple (in our case, three) interdependent univariate regression equations. The computation of the Wald statistic in this nonstandard situation is outlined in Roy (1957).

2.6 Soft thresholding

Soft thresholding is also proposed by Bai and Ng (2008). They argue that a drawback of a hard thresholding rule is that it can easily lead to the selection of many similar predictors, because in deciding whether or not to include a certain predictor, the information contained in the other predictors is not considered.

They propose to use a sequential method, such as LARS or the Lasso, to select variables, and then to find the number of variables that minimizes an information criterion, such as the BIC. In our multivariate setting, we use MRSR as the selection method. The number of selected variables is chosen by minimizing the BIC in the system of linear equations

$$B^+ = \iota\alpha' + B\Delta + X_m\Gamma_m + E, \tag{9}$$

where X_m consists of the first m selected columns of X. The p macro factors to be used in the FANS model are again principal components extracted from the p^* variables that minimize the BIC.

3   Data and forecasting procedure

3.1   Data

Our yield data are unsmoothed Fama-Bliss U.S. Treasury yields for maturities of $3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108,$ and 120 months, from January 1979 to December 2003. For more information on the construction of these data, we refer to Bliss (1997). Time series plots of a short-term (three months), a medium-term (two years), and a long-term (ten years) yield are shown in Figure 1. We observe that these yields vary over a wide range of values over the sample period. Moreover, although the yields for different maturities show a large degree of comovement, it can

be seen that the spreads between yields of different maturities also vary considerably over time.

(Include Figure 1 around here.)

The different possible shapes of the yield curve are illustrated in Figure 2, where we show all seventeen yields in four selected months. A "typical" yield curve is increasing and slightly concave, as is the case in January 1992. However, other shapes, such as humped (1979), decreasing (1981), and almost flat (1990), occur fairly regularly. The Nelson and Siegel (1987) model can accurately approximate all these different shapes.

(Include Figure 2 around here.)

The macro factors that we include in the FANS models are extracted from a set of 132 monthly variables. These variables are categorized into eleven groups in an economically meaningful way. (For example, one group contains various price indices, one contains aggregates related to real output, and so on.) A previous version of this data set is described in Stock and Watson (2002). They also describe the transformations to stationarity and the grouping of variables.[4]

3.2   Forecasting procedure

We estimate the parameters of the FANS model over a rolling window with a length of 120 months. At the end of each window (say, at time t), predictions are made for times $t + h$, with forecast horizon $h = 1, 3, 6,$ or 12. Thus, for forecast horizon $h = 1$, the first prediction that we make is for January 1989. For $h = 3$, it is for March 1989, et cetera.

---

[4]We thank Cem Cakmakli for making a more recent version of these data available to us.

We report forecasting results for the subsamples January 1994 until December 1998 and January 1999 until December 2003. The first subsample closely corresponds to the forecasting period considered by Diebold and Li (2006). (The forecasts for January 1989 until December 1993 are only used for selecting the number of macro factors, as we describe below.) As is clear from Figure 1, the yields exhibit a lower level and more volatility in the latter of these two subsamples. However, Figure 3 shows that the average yield curves (defined as simple arithmetic averages) for both subsamples have roughly similar shapes.

(Include Figure 3 around here.)

If we denote the point forecast for the yield for maturity τ at time $t + h$, made at time t using method M, by $\hat{y}_{t+h|t}(M, \tau)$, the predictive accuracy is evaluated as the mean squared forecast error, defined as

$$\text{MSE}(M, h, \tau) = \frac{1}{T} \sum_t \left[\hat{y}_{t+h|t}(M, \tau) - y_{t+h}(\tau) \right]^2, \tag{10}$$

where the summation runs over the period with length T over which we wish to evaluate the performance.

The forecasting methods that we consider differ in the procedure for constructing macroeconomic factors. Table 1 gives an overview and shows the abbreviations by which we refer to each method. In all but the first of these methods, the number of macro factors is fixed at three to make our results comparable to those in Diebold et al. (2006) and De Pooter et al. (2007).

(Include Table 1 around here.)

The groups of variables used in the PCAG and PCOVRG methods are three of the eleven groups identified by Stock and Watson (2002), chosen such that each group

contains one of the variables used by Diebold et al. (2006). This choice is intended to make our results comparable to theirs.

We consider the first six methods listed in Table 1 as our basic methods. The second set of six are multivariate methods, each selecting three macro factors. A final set of six methods contains variants of each of these multivariate methods, in which the number of factors is not fixed at three. Instead, we estimate the FANS model with either 1, 2, 3, or 4 macro factors, after which the actual number of factors used in forecasting is selected based on past performance; that is, the number of factors leading to the smallest MSE over the last 60 months is used. The abbreviations used to refer to these methods are equal to those for their three-factor counterparts, plus a B for "Best number of factors".

4 Forecasting results

In this section, we evaluate the forecasting performance of the various methods of factor construction listed in Table 1. As our focus is not on the parameter estimates, a short discussion on that topic is deferred to Appendix A.

We divide the seventeen maturities that we study into three groups: short (one year or less), medium (one to five years), and long (more than five years). For each of these groups, we report results only for one representative maturity (three months, two years, and ten years, respectively).[5] The results are qualitatively similar for all maturities within the three groups. The values of MSE (M, h, τ), as defined in Equation (10), are shown in Tables 2 and 3.

(Include Tables 2 and 3 around here.)

---

[5]Results for the other maturities are available upon request.

For all methods, the general pattern can be summarized as follows. Forecasts are more accurate if the horizon is shorter and if the maturity is longer, as can be seen from the rows labeled "NO" in Tables 2 and 3. Comparing these two rows, we can also see that forecasting is easier in the first (less volatile) period. Note that, as Diebold and Li (2006) reported before, simple forecasts made without any macro information are already of reasonable quality. Consider, for example, the MSE for maturity $\tau = 3$ months at a forecast horizon of $h = 1$ month in Table 2, which is equal to 697. This means that when forecasting the short-term yield one month ahead, the average absolute error is $\sqrt{697} \approx 26$ basis points. In fact, one-month-ahead prediction errors are around 0.3 percentage points across all maturities and for both subperiods.

As discussed in Section 2.1, we have imposed some restrictions on the parameter matrices in the state equation of the FANS model. We have also constructed yield forecasts with the FANS model without these restrictions, and we find that the restricted version outperforms the unrestricted model in almost all cases, whether macro information is included or not. The unrestricted model is superior only for forecasting short-term yields over short horizons. For all other cases, dropping the restrictions leads to a dramatic increase in MSE of more than 40% on average, possibly due to overfitting. Therefore, in what follows we only report results obtained from the restricted FANS model.

Considering Table 2 further, it becomes clear that including macro information does not improve much on the no-macro benchmark in the first subperiod. The yield curve was not very volatile in 1994-1998, see Figure 1, and the Nelson-Siegel model without any macro factors appears to be sufficient as a forecasting tool in this period.

In 1999—2003 (Table 3), however, gains in predictive accuracy range from 3% to as much as 34% for the basic methods. For both the long and the short ends of the yield curve, the best-performing basic methods are generally PCAG and PCOVRG.

The latter method even beats the benchmark in all twelve situations listed in the table. For medium-term yields (15 to 60 months) the DRA method, based on the variables selected by Diebold et al. (2006), performs very well. Note that PCAG, PCOVRG and DRA have in common that a major part of the macro information is excluded from the model: PCAG and PCOVRG use only three out of the eleven groups identified by Stock and Watson (2002) (amounting to 58 out of 132 variables), whereas DRA uses only 3 out of 132 variables. On the other hand, the LARSU and PLSU procedures have all macro variables as inputs; their performance is not impressive, suggesting that these methods have difficulties in coping with the abundance of macro variables.

Turning to the multivariate methods with the number of factors fixed at three in the middle six rows of Table 3, we find that it is possible to include all macro information without obstructing forecast accuracy. PCA and (especially) PCOVR perform better than their basic counterparts in many cases. The most notable example is the forecast for the short yield over a 6-month horizon in 1999—2003: PCOVR results in a mean squared forecast error that is 16% (12 percentage points) smaller than the best-performing basic method, and 39% smaller than the no-macro benchmark. Comparing the multivariate methods to the basic methods, a similar picture emerges in 1994—1998; however, it is difficult even for PCA and PCOVR to beat the no-macro benchmark during this period.

Two of these multivariate methods have univariate counterparts: we may compare LARSM to LARSU and PLSM to PLSU. In general, the multivariate methods outperform their univariate counterparts. Thus, it is profitable to consider the yield factors β as a group rather than as three distinct variables. (Incidently, for PLS, Garthwaite (1994) argues that "in most situations, the univariate method is likely to give the better prediction equations." This does not seem to be the case in our application.)

Considering the last six rows of Tables 2 and 3, we observe that data-driven selec-

tion of the number of macroeconomic factors leads to further improvements (compared to fixing this number at three) in some cases, most notably hard thresholding (HTB) and partial least squares (PLSMB). For principal component analysis (PCAB), however, the results deteriorate slightly: it appears that the additional model uncertainty is not worth the additional degree of freedom in this case, so that it seems better to simply work with three factors, as advocated by De Pooter et al. (2007). For the other three methods, the predictive accuracy does not change much. As shown in Table 4, the methods that gain most from selecting the number of factors (hard thresholding and PLS) generally require fewer than three factors, suggesting that the first two factors constructed by these methods already summarize most relevant information. This result illustrates the importance of considering the forecast objective in factor construction: as the first two factors have sufficient predictive power, it is often better to neglect the third factor.

(Include Table 4 around here.)

Table 5 enables a closer look at the relative importance that each method assigns to each group of macroeconomic variables. It shows the average squared weight of all variables in each group, normalized to sum to one for each method. For example, the first number in this table means that 22% of the variation in the LARSU factors comes from variables connected to real output and income. We observe that all methods give large weights to the "Employment and Hours" group; however, the best-performing methods (PCA and PCOVR) put more weight on price indices.

Given the relatively poor performance of both thresholding methods (ST and HT), it is perhaps surprising that the weights they assign to the eleven groups of variables are not very different from those assigned by other methods. This similarity means that the thresholding methods do not select variables from the wrong groups; rather, they

21

seem to pick "the wrong variables from the right groups". As it is hard for thresholding methods to differentiate between variables with large in-sample correlations, it seems preferable to use methods that summarize all information in the data set, instead of selecting a smaller number of variables.

(Include Table 5 around here.)

To gain more insight in the variability of the predictive performance over time, we plot a five-year rolling mean squared forecast error in Figure 4. Each point shows the mean squared forecast error measured over the past 60 months. Thus, the circled points (December 1998 and December 2003) correspond to the numbers used in Tables 2 and 3, respectively. We plot the errors only for two methods, one maturity, and one forecast horizon, but the picture is similar for the other cases. The best predictability of yields is realized around the turn of the century, while forecasting accuracy was better in 2003 than in the years preceding it.

(Include Figure 4 around here.)

Figure 5 illustrates the relative performance of univariate PLS, multivariate PLS, and PLS with data-driven selection of the number of factors. (Again, the circled points correspond to the numbers in Tables 2 and 3, and the picture for other methods is roughly similar.) The multivariate method almost always dominates the univariate method, and not fixing the number of factors leads to an additional improvement. As we can see by comparing Figures 4 and 5, adding macroeconomic information helps most when forecasting without such information is difficult, particularly from 2001 onwards.

(Include Figure 5 around here.)

22

Summarizing our empirical results, we find substantial evidence that including macroeconomic information in the Nelson-Siegel model improves predictive accuracy the most in volatile periods, when the basic yields-only model is insufficient. Concerning the specific issues related to the use of macro factors as discussed in the introduction, we also find several interesting conclusions. First, including factors extracted from a large panel of macro variables generally renders more accurate forecasts than pre-selecting specific individual variables in a data-driven way. Second, it is better to use the target variable in constructing macro factors than to ignore it. Third, it is better to construct a single set of factors for all three yield factors jointly than to treat each yield factor separately. Fourth, for long-term yields it is better to construct factors from groups of related macro variables, instead of one large pool of all available variables, but the opposite holds for short- and medium-term yields. Fifth, it is better to select the number of macro factors based on past performance than to fix it in advance.

5 Conclusion

This study investigates various ways of incorporating macroeconomic information in the Nelson-Siegel framework for forecasting the yield curve. By comparing the predictive performance of these techniques with methods found in the recent literature, we find that alternative methods perform better in important cases. It is not easy to improve upon forecasts made without any macro information in stable times, like the years 1994—1998. When there is little volatility in yields, the dynamic Nelson-Siegel model without macro factors is sufficient for forecasting, although gains of around 10% can still be attained in several cases. On the other hand, when volatility is relatively high (as in 1999—2003), macroeconomic variables are of substantive help in forecasting the yield curve. Gains of around 30% are attainable in this case.

The gains obtained by including macroeconomic information depend crucially on the way in which macro information is incorporated in the model. Selecting variables, as in Diebold et al. (2006), is useful only for forecasting yields with medium-term maturities (more than one, but less than five years). In all other cases, it is beneficial to extract information from a larger pool of available variables: from predefined groups of variables for longer maturities (using 58 variables in this study), or even from all available information (132 variables in our case) for shorter maturities.

Principal component analysis (PCA), as advocated by De Pooter et al. (2007), yields acceptable results: using principal components generally improves forecast accuracy by around 10%, relative to the no-macro benchmark. Still, principal covariate regression (PCovR) improves on simple PCA under most circumstances, often gaining another 10%. Other factor construction methods, based on selecting variables (using least angle regression, or a thresholding rule), are dominated by methods that use all available information, such as PCovR. We also find that it is helpful to consider the three yield factors as a group, rather than as three separate variables.

We also investigated the benefits of selecting the number of factors in a data-driven way, based on predictive performance in the recent past. The results indicate that such a selection procedure leads to a small decrease in forecast errors for some methods. There is, however, a trade-off between additional modeling freedom and model uncertainty.

To conclude, incorporating macroeconomic information appears to be most useful when it is needed the most, when yields are highly volatile, and it is beneficial to explicitly consider the forecasting objective when constructing factors. We expect that the factor construction rules studied here may also prove useful in other models and contexts. An interesting example may be the use of augmented affine models for the yield curve, in the style of Mönch (2008).

24

A   Parameter estimates

This appendix illustrates some of the estimates of the parameters of the FANS model. Figure 6 shows a time series plot of the estimates of λ. All estimates are dated in the last month of the estimation window. We only show results for the model without macro factors; for the methods with macro factors, the estimates are very similar.

(Include Figure 6 around here.)

Typical values of λ found in the literature, such as Diebold and Li (2006), Diebold et al. (2006), and De Pooter et al. (2007), are around 0.06 to 0.08. Our estimates lie, on average, around these values. However, the λ estimates show great variability over time, casting some doubt on studies where λ is kept fixed over long time periods.

For the parameter estimates of μ, our results are similar to those found in the papers cited above. The mean of the level parameter (the hypothetical yield of infinitely long maturity) is estimated to be around 600 basis points. The mean slope parameter is around -200, corresponding to an upward-sloping yield curve. Finally, the mean of the curvature parameter is small, but larger than zero, which indicates that the yield curve is slightly concave.

For the other parameters in the FANS model, our estimates are comparable to those reported by Diebold et al. (2006). The standard deviations σ_i are largest for the shortest maturities: typical values are around 15 basis points for short maturities, 2 to 4 for medium-term yields, and again slightly higher for longer maturities (over five years): roughly between 3 and 10 basis points.

The estimated transition matrix A in the FANS model (see Equation (4)) is always close to nonstationarity, with large own-lag coefficients for the level and slope parameters, illustrating the high persistence of yield curves. The block of A governing the macro-to-yields feedback relations is not negligible compared to the diagonal

25

elements, confirming our economic intuition about such links and the result that these parameters are significantly nonzero, obtained by Diebold et al. (2006). Finally, the estimates of the state variance matrix Q generally show a pattern with much larger shocks to the curvature factor than to the other factors.

As an illustrative example, we provide the parameter estimates found by the PCOVR method over the last estimation window, January 1994—December 2003. The vector of maturities τ is shown for ease of interpretation of the vector of standard deviations σ. Note that all elements of A and Q shown as "0" were actually imposed to be zero, as discussed in Section 2.1.

$$\lambda = 0.06, \qquad \mu = \left(\begin{array}{ccc} 619.86, & -333.59, & -172.69 \end{array}\right)',$$

$$\tau = \left(\begin{array}{ccccccccc} 3, & 6, & 9, & 12, & 15, & 18, & 21, & 24, & 30, \\ & 36, & 48, & 60, & 72, & 84, & 96, & 108, & 120 \end{array}\right)',$$

$$\sigma = \left(\begin{array}{ccccccccc} 35.04, & 21.84, & 16.17, & 8.59, & 2.74, & 1.93, & 2.32, & 3.92, & 2.08, \\ & 2.35, & 2.37, & 5.29, & 4.39, & 3.85, & 2.85, & 4.04, & 9.54 \end{array}\right)',$$

$$A = \left(\begin{array}{cccccc} 0.93 & 0 & 0 & 0.02 & 0.05 & 0.01 \\ 0 & 0.98 & 0 & 0.36 & 0.38 & 0.16 \\ 0 & 0 & 0.93 & 0.03 & 0.09 & 0.14 \\ 0 & 0 & 0 & 0.79 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.74 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.72 \end{array}\right),$$

$$Q = \left(\begin{array}{cccccc} 670 & 0 & 0 & 0 & 0 & 0 \\ & 1364 & 0 & 0 & 0 & 0 \\ & & 5273 & 0 & 0 & 0 \\ & & & 3780 & 0 & 0 \\ & & & & 4638 & 0 \\ & & & & & 5076 \end{array}\right).$$

References

Ang, A., and M. Piazzesi, 2003, A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, Journal of Monetary Economics 50, 745–787.

Bai, J., and S. Ng, 2008, Forecasting economic time series using targeted predictors, Journal of Econometrics 146, 304–317.

Bliss, R.R., 1997, Testing term structure estimation methods, Advances in futures and options research 9, 197–231.

Boivin, J., and S. Ng, 2006, Are more data always better for factor analysis?, Journal of Econometrics 132, 169–194.

Christensen, J.H.E., F.X. Diebold, and G.D. Rudebusch, 2007, The affine arbitrage-free class of Nelson-Siegel term structure models, NBER Working Paper 13611.

Cox, J.C., J.E. Ingersoll, and S.A. Ross, 1985, A theory of the term structure of interest rates, Econometrica 53, 385–408.

Dai, Q., and K.J. Singleton, 2000, Specification analysis of affine term structure models, Journal of Finance 55, 1943–1978.

De Pooter, M., F. Ravazzolo, and D. van Dijk, 2007, Predicting the term structure of interest rates: Incorporating parameter uncertainty, model uncertainty and macroeonomic information, Tinbergen Institute Discussion Paper 2007-028/4.

Dewachter, H., and M. Lyrio, 2006, Macro factors and the term structure of interest rates, Journal of Money, Credit, and Banking 38, 119–140.

Diebold, F.X., and C. Li, 2006, Forecasting the term structure of government bond yields, Journal of Econometrics 130, 337–364.

Diebold, F.X., 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.

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.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani, 2004, Least angle regression, Annals of Statistics 32, 407–499.

Garthwaite, P.H., 1994, An interpretation of partial least squares, Journal of the American Statistical Association 89, 122–127.

Heij, C., P.J.F. Groenen, and D. van Dijk, 2006, Time series forecasting by principal covariate regression, Econometric Institute Report 2006-37.

— 2007, Forecast comparison of principal component regression and principal covariate regression, Computational Statistics and Data Analysis 51, 3612–3625.

Hördahl, P., O. Tristani, and D. Vestin, 2006, A joint econometric model of macroeconomic and term-structure dynamics, Journal of Econometrics 131, 405–444.

Koopman, S.J., M.I.P. Mallee, and M. van der Wel, 2010, Analyzing the term structure of interest rates using the dynamic Nelson-Siegel model with time-varying parameters, Journal of Business and Economic Statistics, forthcoming.

Mönch, E., 2008, Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented VAR approach, Journal of Econometrics 146, 26–43.

Nelson, C.R., and A.F. Siegel, 1987, Parsimonious modeling of yield curves, Journal of Business 60, 473–489.

Roy, S.N., 1957, Some aspects of multivariate analysis (Wiley, New York).

Rudebusch, G.D., and T. Wu, 2008, A macro-finance model of the term structure, monetary policy, and the economy, Economic Journal 118, 906–926.

Similä, T., and J. Tikka, 2006, Common subset selection of inputs in multiresponse regression, in: Proceedings of the IEEE international joint conference on neural networks, 1908–1915.

Stock, J.H., and M.W. Watson, 2002, Macroeconomic forecasting using diffusion indexes, Journal of Business and Economic Statistics 20, 147–162.

Taylor, J.B., 1993, Discretion versus policy rules in practice, Carnegie-Rochester conference series on public policy 39, 195–214.

Tibshirani, R., 1996, Regression shrinkage and selection via the Lasso, Journal of the Royal Statistical Society: Series B 58, 267–288.

Vašíček, O., 1977, An equilibrium characterization of the term structure, Journal of Financial Economics 5, 177–188.

Wold, H., 1966, Nonlinear estimation by iterative least squares procedures, in: F. David, ed., Research papers in statistics: Festschrift for J. Neyman (Wiley, New York), 411–444.

Table 1: Macro factor construction methods.

Abbreviation	Name	Description
Basic methods		
NO	No macro factors	The Nelson-Siegel model without any macro factors, as in Diebold and Li (2006)
DRA	Diebold, Rudebusch, Aruoba	Macro factors are capacity utilization, the federal funds rate, and the price deflator for personal consumption expenditures, as in Diebold et al. (2006)
LARSU	Univariate LARS	The first macro variable selected by LARS for each of the three yield factors β
PLSU	Univariate PLS	The first PLS factor constructed for each of the three yield factors
PCAG	PCA on groups	One principal component from each of the three groups "Real Output and Income", "Interest Rates and Spreads", and "Price Indices", from Stock and Watson (2002)
PCOVRG	PCovR on groups	One principal covariate from each of the same three groups as in PCAG
Multivariate methods		
LARSM	Multivariate LARS	The first three macro variables selected by MRSR for the three yield factors jointly
ST	Soft thresholding	Three factors constructed by the soft thresholding rule
HT	Hard thresholding	Three factors constructed by the hard thresholding rule
PLSM	Multivariate PLS	The first three PLS factors constructed for the three yield factors jointly
PCA	Principal component analysis	Three principal components from the full macro data set, as in De Pooter et al. (2007)
PCOVR	Principal covariate regression	Three principal covariates from the full macro data set

Table 2: Mean squared forecast errors, 1994—1998.

Maturity		3				24				120			
Horizon		1	3	6	12	1	3	6	12	1	3	6	12
NO		697	1577	3626	8837	758	3214	6705	13208	715	2611	6188	12704
Basic methods													
DRA	(0)	1.35	1.33	1.32	1.78	1.09	1.26	1.48	2.06	1.05	1.18	1.29	1.65
LARSU	(0)	2.00	2.28	1.95	1.81	1.15	1.25	1.29	1.60	1.18	1.17	1.19	1.29
PLSU	(0)	1.82	2.17	2.04	1.74	1.20	1.18	1.16	1.15	1.34	1.25	1.22	1.26
PCAG	(3)	1.25	1.00	0.98	1.11	1.03	1.02	1.01	1.05	1.12	1.03	**0.96**	**0.95**
PCOVRG	(1)	1.06	**0.98**	1.19	1.76	1.00	1.14	1.40	1.74	1.04	1.16	1.27	1.39
Multivariate methods													
LARSM	(2)	1.86	1.95	1.54	1.56	1.19	1.16	1.09	1.54	1.03	0.98	0.98	1.42
ST	(1)	1.71	1.63	1.29	1.24	1.13	1.07	1.05	1.28	1.07	0.98	1.01	1.20
HT	(1)	1.73	1.75	1.35	1.07	1.06	1.04	1.02	0.95	1.09	1.08	1.11	1.09
PLSM	(1)	1.81	1.80	1.53	1.18	1.10	1.09	1.04	0.96	1.32	1.22	1.13	1.15
PCA	(3)	1.55	1.48	1.10	**0.93**	1.02	1.03	**0.98**	0.91	1.05	1.04	1.02	1.03
PCOVR	(1)	1.18	1.00	**0.88**	1.03	1.06	1.06	1.08	1.07	1.03	1.03	1.08	1.09
Selected number of factors													
LARSMB	(0)	1.90	1.95	1.66	1.54	1.10	1.17	1.09	1.49	1.01	1.03	1.05	1.37
STB	(2)	1.69	1.81	1.34	1.21	1.15	1.09	1.04	1.22	1.09	0.99	0.99	1.14
HTB	(0)	1.62	1.58	1.32	1.19	1.08	1.06	1.02	1.14	1.08	1.03	1.03	1.14
PLSMB	(3)	1.43	1.30	1.19	1.05	1.13	**0.94**	1.02	**0.88**	1.26	**0.97**	1.08	1.12
PCAB	(0)	1.58	1.56	1.28	1.09	1.01	1.09	1.11	1.06	1.04	1.09	1.09	1.07
PCOVRB	(0)	1.26	1.36	1.34	1.26	1.02	1.12	1.23	1.21	1.03	1.06	1.08	1.04

Notes: The first row reports the MSE (in squared basis points), as defined in Equation (10), for the method NO, see Table 1. For ease of comparison, relative values $\mathrm{MSE}(M, h, \tau) / \mathrm{MSE}(\mathrm{NO}, h, \tau)$ follow in the remaining rows, for all other methods described in Table 1. For each maturity and horizon (both measured in months), the lowest relative MSE is printed in bold. If a column contains no bold entry, this means that none of the methods outperforms the NO benchmark. The numbers in parentheses indicate in how many of the twelve cases the respective methods outperform this benchmark.

Table 3: Mean squared forecast errors, 1999—2003.

Maturity		3				24				120			
Horizon		1	3	6	12	1	3	6	12	1	3	6	12
NO		1037	3155	8803	27194	1003	3551	8537	21445	903	2092	4077	7905
Basic methods													
DRA	(4)	1.41	1.30	1.24	1.44	0.95	0.95	0.87	**0.66**	1.15	1.36	1.36	1.18
LARSU	(4)	1.09	0.84	0.81	0.89	1.00	1.01	1.04	1.09	1.12	1.25	1.30	1.40
PLSU	(3)	1.38	1.05	0.93	1.05	0.96	0.95	1.01	1.10	1.28	1.27	1.37	1.42
PCAG	(9)	**0.87**	0.74	0.80	0.91	0.97	0.97	0.98	0.97	1.03	1.03	1.01	0.94
PCOVRG	(12)	0.95	0.76	0.73	0.90	1.00	0.97	0.96	0.94	**0.97**	**0.97**	**0.91**	**0.77**
Multivariate methods													
LARSM	(5)	1.13	0.91	0.84	0.87	0.88	0.97	1.02	1.11	1.05	1.21	1.30	1.52
ST	(7)	1.17	0.85	0.74	0.80	0.90	0.91	0.86	0.92	1.03	1.08	1.02	1.07
HT	(2)	1.22	0.93	0.95	1.08	1.02	1.01	1.07	1.11	1.22	1.20	1.22	1.24
PLSM	(4)	1.17	0.85	0.88	1.01	0.87	0.95	1.00	1.06	1.17	1.17	1.16	1.20
PCA	(8)	1.20	0.91	0.81	0.86	0.93	0.90	0.89	0.88	1.05	1.03	1.00	0.95
PCOVR	(8)	1.12	0.71	**0.61**	0.78	0.94	0.80	**0.84**	0.84	1.06	1.02	1.08	0.98
Selected number of factors													
LARSMB	(5)	1.13	0.88	0.81	0.86	0.91	0.96	1.02	1.09	1.07	1.20	1.33	1.51
STB	(7)	1.17	0.85	0.74	0.86	0.93	0.93	0.86	0.93	1.07	1.12	1.02	1.06
HTB	(6)	1.20	0.90	0.77	**0.74**	1.04	0.99	0.91	0.80	1.12	1.16	1.11	1.02
PLSMB	(7)	1.13	0.78	0.79	0.90	**0.84**	0.86	0.89	0.97	1.22	1.10	1.05	1.00
PCAB	(8)	1.19	0.89	0.77	0.82	0.89	0.95	0.91	0.88	1.03	1.11	1.04	0.95
PCOVRB	(7)	1.12	**0.71**	0.62	0.82	0.95	**0.80**	0.85	0.87	1.07	1.03	1.10	1.02

Notes: See Table 2.

Table 4: Average "best" number of factors, 1994—2003.

Horizon	1	3	6	12
LARSMB	2.38	3.50	2.93	2.53
STB	2.35	3.19	2.95	1.46
HTB	1.38	1.57	1.74	2.30
PLSMB	1.93	1.93	1.95	2.22
PCAB	1.63	2.28	2.27	2.20
PCOVRB	2.31	2.28	2.07	1.93

Table 5: Average weight of groups of macro variables, 1994—2003.

Group name	#	LARSU	PLSU	LARSM	ST	HT	PLSM	PCA	PCOVR
Real Output and Income	20	0.22	0.09	0.11	0.11	0.12	0.08	0.12	0.12
Employment and Hours	30	**0.39**	**0.51**	**0.29**	**0.29**	0.19	0.33	0.17	0.17
Housing	10	0.00	0.00	0.06	0.07	**0.28**	0.00	0.14	0.14
Orders and Inventories	10	0.03	0.02	0.08	0.13	0.05	0.01	0.05	0.05
Money and Credit	11	0.04	0.06	0.11	0.09	0.05	0.08	0.03	0.03
Stock Prices	4	0.01	0.00	0.01	0.02	0.02	0.00	0.02	0.02
Interest Rates and Spreads	17	0.17	0.00	0.18	0.14	0.24	0.00	0.20	0.20
Exchange Rates	5	0.03	0.00	0.01	0.01	0.02	0.00	0.01	0.01
Price Indices	21	0.08	0.30	0.10	0.09	0.03	**0.46**	**0.26**	**0.26**
Wages	3	0.00	0.01	0.00	0.00	0.00	0.03	0.00	0.00
Consumer Expectations	1	0.03	0.00	0.05	0.05	0.00	0.00	0.00	0.00

Notes: this table shows the relative importance of each of the eleven groups of macroeconomic variables in the factor construction methods. The numbers in the table are sums of squared weights, averaged over the 120 estimation windows ending in 1994—2003, and normalized to sum to one for each method. For each method, the largest weight is printed in bold. The column headed "#" lists the total number of variables per group.

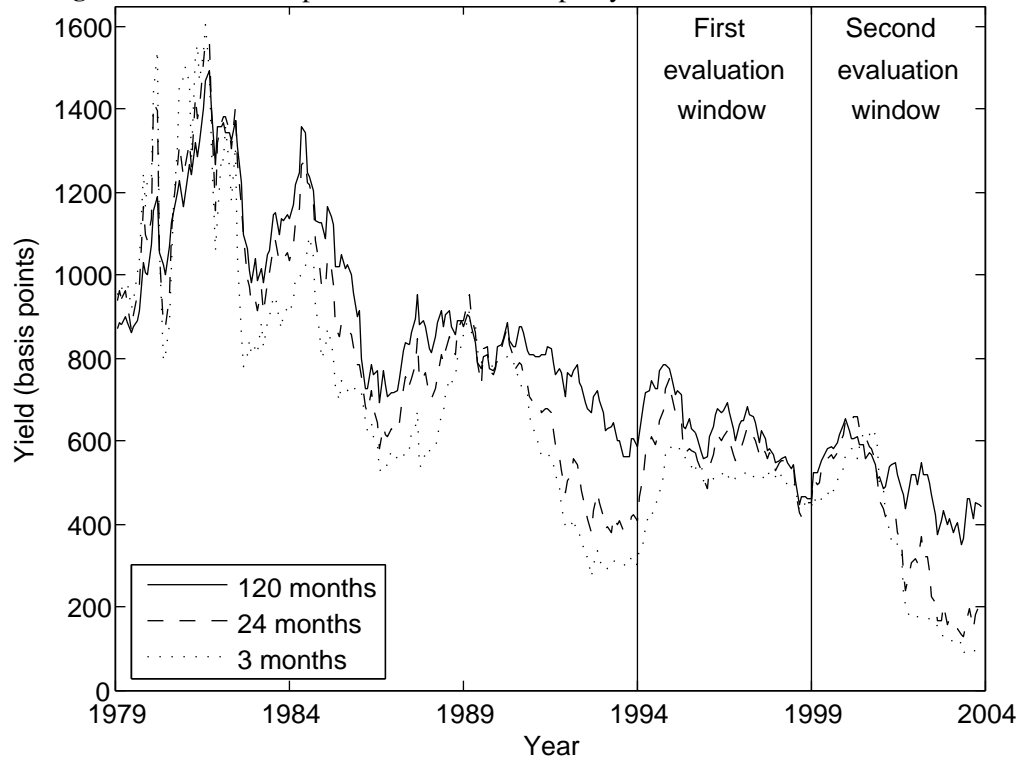**Figure 1:** Time series plots of U.S. zero-coupon yields for three selected maturities.

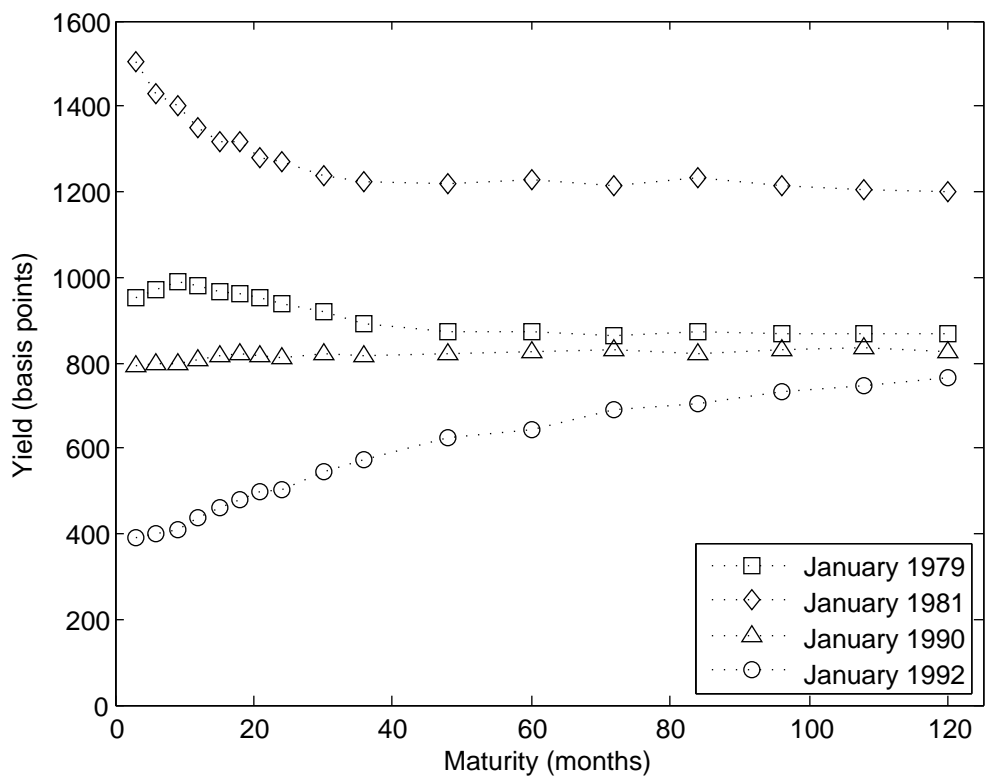**Figure 2:** The yield curve in four selected months.

Figure 3: The average yield curve in the two evaluation subsamples.
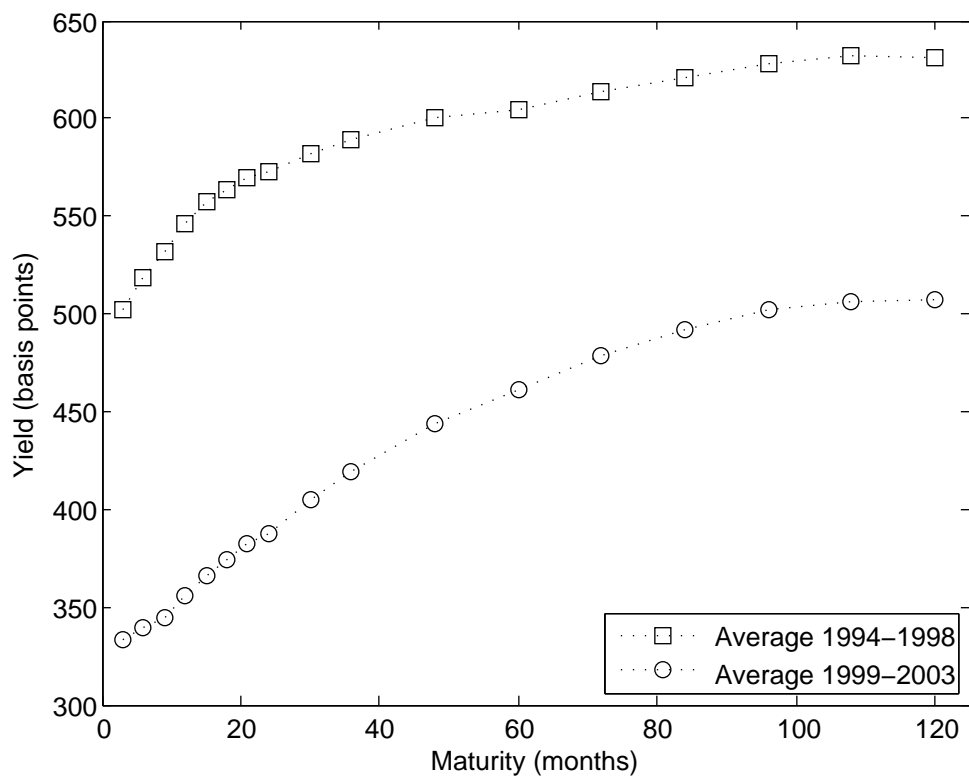
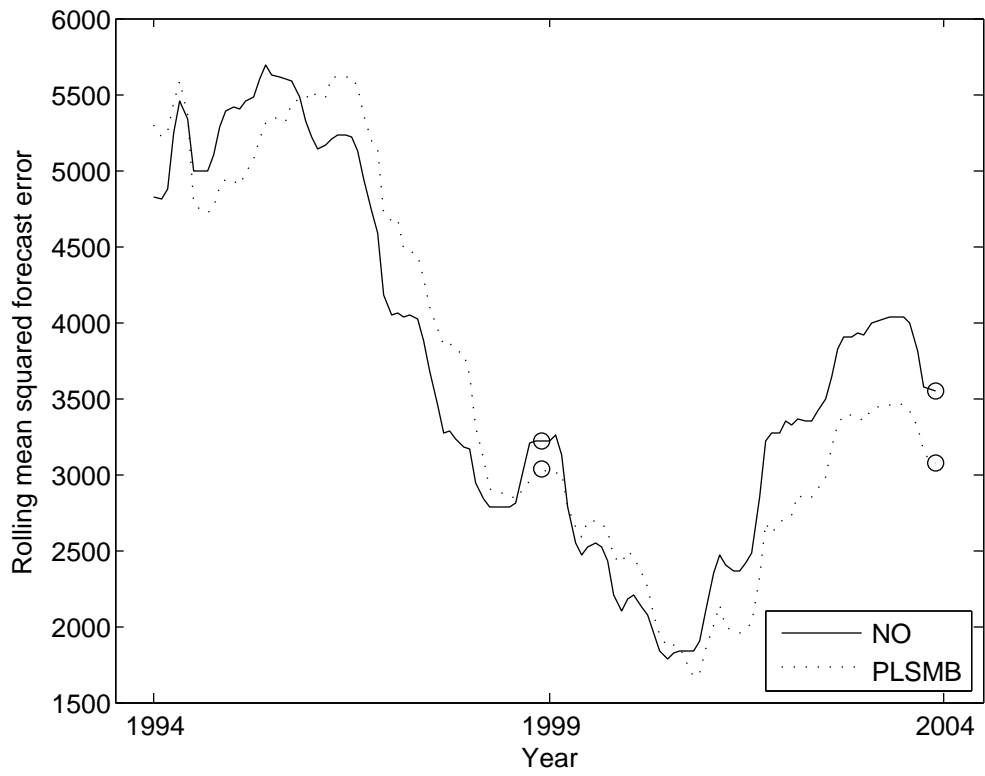**Figure 4:** Rolling means of squared forecast errors for forecast horizon $h = 3$, maturity $\tau = 24$ months.

Figure 5: Rolling means of squared forecast errors, forecast horizon $h = 3$, maturity $\tau = 24$ months, relative to the NO benchmark.
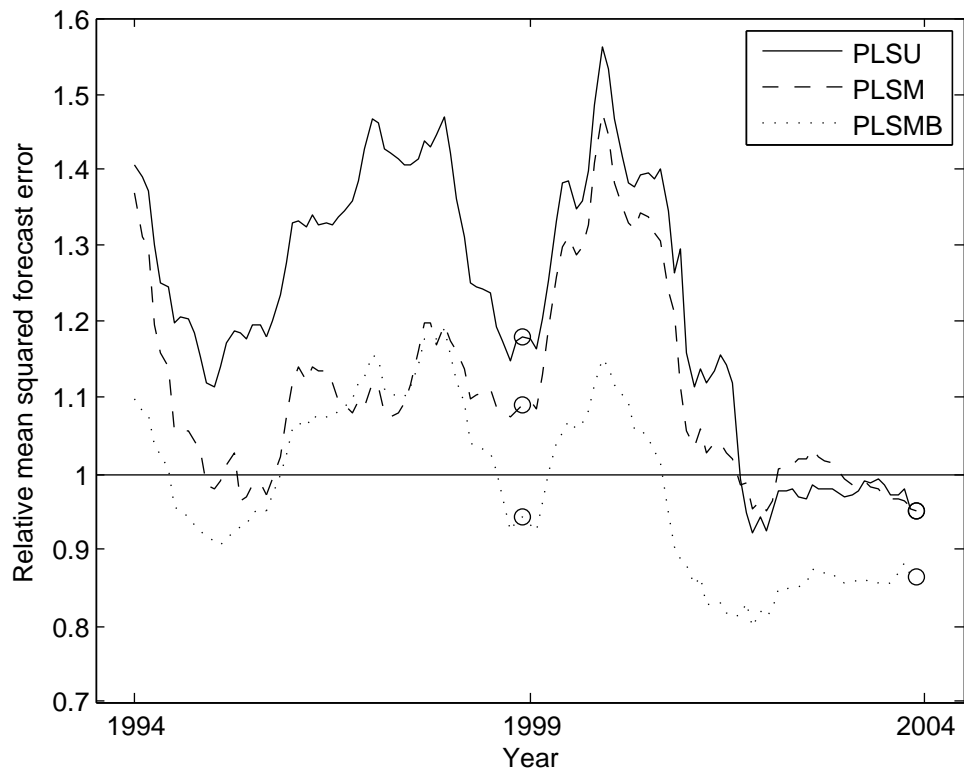
Figure 6: Estimates of the parameter λ in the FANS model.