Forecasting the Yield Curve with S-Plus

Dario Cziráky, PhD
Senior Consultant, Finance Sector, Insightful Corporation

1 Introduction

Methods capable of forecasting the entire yield curve based on a time series extension of the Nelson-Siegel model Nelson and Siegel (1987) were suggested in the literature and compared to the non-parametric alternatives Diebold and Li (2004). While relatively successful in forecasting the yield curve, the Nelson-Siegel model tends to have poor fit to highly non-linear yield curves and at the long end of the term structure, although this can be improved by considering the Svensson (1994) model.

However, we find that Nelson-Siegel and Svensson models have poor forecasting performance around the points of non-parallel shifts, hence making them potentially problematic in interest rate risk management.

In this paper we show how to implement these models using non-linear least squares and how to obtain standard errors and confidence intervals for the parameters, which proves to be useful in assessing the goodness-of-fit at specific points in the term structure, such as at the events of non-parallel shifts.

Furthermore, we consider an alternative model based on principal components and smoothing splines, which gives improved forecasting performance, particularly for the highly non-linear changes in the term structure curvature.

For empirical illustrations we use the Euribor short-term interest rates with maturities ranging from one week to one year and the Government bond yields with longer maturities of 2 to 10 years, which gives a fairly complete European term structure data (Figure 1).

Considering a widely spread term structure presents a modelling challenge, especially when attempting to forecast the yield curve. The EU term structure has notably changed both slope and curvature at several points from the 2001–2006 period covered by our data. Several of the changes can be viewed as non-parallel shifts in the term structure, hence causing non-linear shifts in the shape of the yield curve at a number of time points.

In this paper we carry out an array of backtests comparing out-of sample performance of various multivariate linear models such as PCA-VAR and multivariate GARCH models, along with a combinations of Nelson-Siegel-Svensson and MGARCH methods. We adopt a hands-on approach and demonstrating how the suggested methods can be easily implemented in S or R programming languages, illustrating the methods on the EU terms structure data.

2 Nelson–Siegel and Svensson functions

The Nelson and Siegel (1987) model specifies the yield curve of interest rate vector $r_t = (r_{1t}, \ldots, r_{Mt})$ equation for a given time t as a non-linear function of the maturities vector $m = (m_1, \ldots, m_M)$ and can be written for a given rate (maturity) i at time t as

$$Y_{it}^{NS}(m, \ \theta_t) = \beta_{t0} - \beta_{t1} \left(\frac{1 - \exp\left(-m_i/\tau_t^{-1}\right)}{m_i/\tau_t^{-1}} \right)$$
$$+ \beta_{t2} \left(\frac{1 - \exp\left(-m_i/\tau_t^{-1}\right)}{m_i/\tau_t^{-1}} - \exp\left(-m_i/\tau_t^{-1}\right) \right). \tag{1}$$

The parameters in (1) to be estimated for each t are thus $\theta_t = (\beta_{t0}, \beta_{t1}, \beta_{t2}, \tau_t)$. Fitting (1) to each time point $t = 1, \ldots, T$ results in a multivariate time series Θ of the form

$$\Theta = \begin{pmatrix} \beta_{10} & \beta_{11} & \beta_{12} & \tau_1 \\ \beta_{20} & \beta_{11} & \beta_{22} & \tau_2 \\ \vdots & \vdots & \vdots & \vdots \\ \beta_{T0} & \beta_{T1} & \beta_{T2} & \tau_T \end{pmatrix}. \tag{2}$$
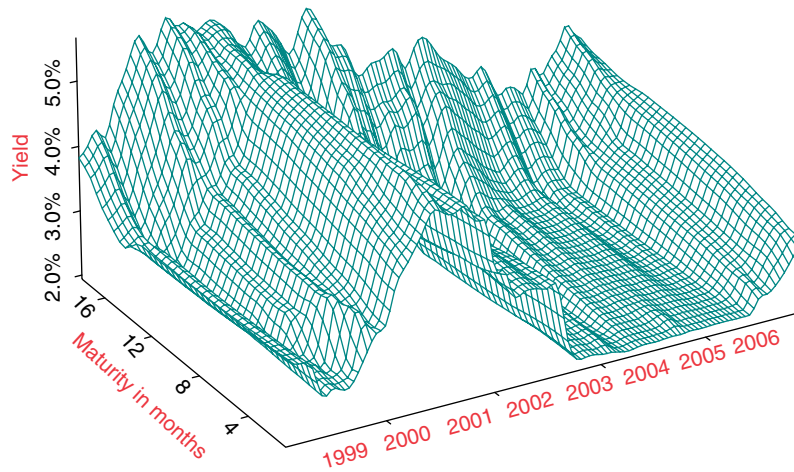
Figure 1: EU interest rates.

Diebold and Li (2004) interpret the first three columns of Θ as latent variables representing the *level*, *slope*, and *curvature* of the term structure. London (2005) gives similar interpretation to the first three principal components of the interests rates time series, hence the Nelson-Siegel model can be interpreted as a latent variable model that can be used to estimate the underlying latent factors driving the term structure. The consequence is that we can now consider forecasting latent factors. In the Nelson-Siegel model we already know the functional relationship between the factors and the individual interest rates, and we can infer the future shape of the yield curve.

Diebold and Li (2004) fit the Nelson-Siegel function (1) by fixing τ_t to a constant and estimate the beta coefficients by ordinary least squares (OLS) applied sequentially to each time point.

However, the Nelson-Siegel function is known to fit poorly to highly non-linear term structures, especially at the longer end of the term structure. To overcome this problem Svensson (1994) generalized the (Nelson and Siegel 1987) model by including additional parameters leading to the model given by

$$Y_{it}^{NSS}(\boldsymbol{m}, \theta_t) = \beta_{t0} - \beta_{t1}\left(\frac{1 - \exp\left(-m_{it}/\tau_{t1}^{-1}\right)}{m_{it}/\tau_{t1}^{-1}}\right)$$

$$+ \beta_{t2}\left(\frac{1 - \exp\left(-m_i/\tau_{t1}^{-1}\right)}{m_{it}/\tau_{t1}^{-1}} - \exp\left(-m_{it}/\tau_{t1}^{-1}\right)\right) \quad (3)$$

$$+ \beta_{t3}\left(\frac{1 - \exp\left(-m_i/\tau_{t2}^{-1}\right)}{m_{it}/\tau_{t2}^{-1}} - \exp\left(-m_{it}/\tau_{t2}^{-1}\right)\right)$$

hence the time-varying parameters are now $\theta_t = (\beta_{t0}, \beta_{t1}, \beta_{t2}, \beta_{t3}, \tau_{t1}, \tau_{t2})$, which apparently loses some of the interpretability of the simpler "three-factor" Nelson-Siegel model. A similar parametrization was considered by El Karoui *et al.* (2006).

The θ_t parameters in (1) and (3) can be estimated by nonlinear least squares by minimizing

$$S_t = \min_{\theta_t} \sum_{i=1}^{M} (\boldsymbol{Y}_{ti}(\boldsymbol{m}, \theta_t) - \boldsymbol{r}_{ti})^2, \quad (4)$$

for each time *t*. The challenge in estimating (4) is not as much in obtaining unconstrained parameters, but in obtaining standard errors or confidence intervals along with allowing for bound constraints on the parameters or latent factors. Zivot and Wang (2006) describe estimation methods in S+FinMetrics capable of estimating unconstrained parameters in Nelson-Siegel and Svensson models. Scherer and Martin (2005) briefly illustrate derivative free estimation of El Karoui et al. (2006) version of the Svensson model without obtaining the confidence intervals for the coefficients.

We will show who how to implement the Svensson model (and hence with minor modification of the objective function the Nelson-Siegel model) in the S programming language by obtaining coefficient estimates along with standard errors, while allowing for upper and lower bound constrains on the coefficients.

The Svensson function (3) can be implemented in S as a simple function of a data vector `data` and a generic parameter vector `p` as follows

```
NSS.obj <- function(p, data)
{
zero.fit <-(p[1]-p[2]*((1-exp(-(Mtr/p[3])))/(Mtr/p[3]))+
           p[4]*((1-exp(-(Mtr/p[3])))^2/(-(Mtr/p[3]))
            -exp(-(Mtr/p[3])))+
           p[5]*((1-exp(-(Mtr/p[6])))^2/(-(Mtr/p[6]))
            -exp(-(Mtr/p[6]))))
sum((zero.fit - data)^2)
}
```

which gives us the necessary objective function for numerical optimization. In order to obtain standard errors of the parameters we can use a Newton algorithm, which requires expressions for the first and second derivatives (gradient and Hessian). These expressions can be obtained automatically by the `deriv3` function from the MASS library (Venables and Ripley 2001), the necessary S code for the Svensson function can be written as

```
NSS.der <- deriv3(~((p1-p2*((1-exp(-(Mtr/p3)))/(Mtr/p3))+
    p4*((1-exp(-(Mtr/p3)))^2/(-(Mtr/p3))-exp(-(Mtr/p3)))+
    p5*((1-exp(-(Mtr/p6)))^2/(-(Mtr/p6))
     -exp(-(Mtr/p6)))-data)^2,
    c("p1", "p2", "p3", "p4", "p5", "p6"),
    function(data, p1, p2, p3, p4, p5, p6) NULL)

NSS.GradHess <- function(p, data)
{
    e <- NSS.der(data, p[1], p[2], p[3], p[4], p[5], p[6])
    g <- colSums(attr(e, "gradient"))
    H <- colSums(attr(e, "hessian"), 2)
    list(gradient = g, hessian = H[row(H) <= col(H)])
}
```

With `NSS.obj`, `NSS.der`, and `NSS.GradHess` at hand we can make use of a general purpose optimization routine such as `nlminb` (non-linear minimization with box constrains) in S-Plus.

As an illustration, consider estimating the yield curve for the monthly EURIBOR terms structure date on February 2000

```
results <- nlminb(start=rep(0.5,6),
                  objective=NSS.obj,
                  gradient=NSS.GradHess,
                  hessian=T,
                  data=EURIBOR.Feb2000,
                  lower=c(0.0001,-0.5,0.0001,
                          0.0001,0.0001,0.0001),
                  upper=c(1,1,5,1,1,1),
            control=nlminb.control(iter.max=50,
                  eval.max=100))
```

Hessian matrix can be easily extracted as `results$hessian`. Moreover, we can compute the coefficient correlation matrix by `round(cov2cor (results$hessian),5)`.

Now, lets consider an empirical application. Figure 2 shows Svensson model estimates for selected months at times of pronounced shifts in the term structure. Recalling Figure 1, we know that in late 2000 early 2001 the term structure changed both direction and shape, thus inducing a non-parallel shift with pronounced changes in the yield curve. Along the path the shape changed from concave to convex which caused very poor fit of the Svensson model (we obtained similar results using the simpler Nelson-Siegel model).

The obvious issue is how reliable would Nelson-Siegel or Svensson models be in yield curve forecasting around the inversion points, at times when the term structure is shocked by non-parallel shifts? Before trying to answer this question, we can take a closer look at the coefficient estimates at selected points in the terms structure. Table 1 shows estimates for February and November 2000, at the times when the yield curve had concave shape, and before it assumed convex shape.

Clearly, we can notice changes in the estimated coefficients, including a sign change in β_1, however, it is evident that none of the beta parameters (i.e. latent factors) were estimated precisely, the estimates are insignificantly different from zero. On the other hand the τ parameters that control the curvature of the function are significant. Here we can also notice that τ_1 significantly increased from 2.18 to 3.22 between February and November 2000, which suggests the importance of estimating τ parameters from the data rather then fixing them to a constant as is often done in practice.

In addition, coefficients seem to be highly correlated, and notably correlations between τ_1 and all other coefficients changed signs between two time points (Table 2). With these estimates at hand, we have slightly better insight in the miss-fit of the Nelson-Siegel and Svensson models at
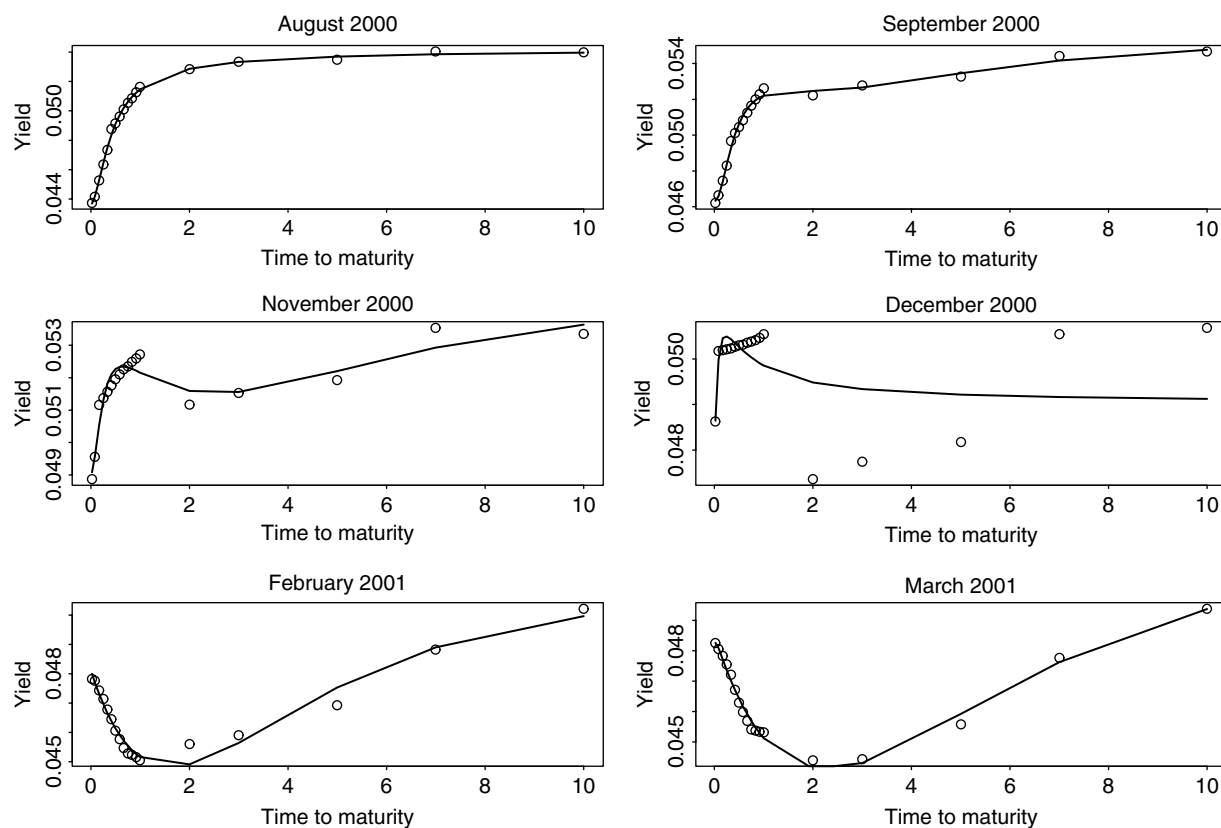

Figure 2: Yield curves from August 2000 till March 2001

Table 1: NSS model coefficient estimates

	Estimate	SD	t-ratio	p-value
February 2000 estimates				
β_0	0.0620	6.0000	0.0103	0.4959
β_1	0.0206	4.7970	0.0043	0.4983
τ_1	2.1830	0.0105	207.0776	0.0000
β_2	0.0024	5.2406	0.0005	0.4998
β_3	0.0061	3.7436	0.0016	0.4994
τ_2	0.3950	0.0256	15.4273	0.0000
November 2000 estimates:				
β_0	0.0539	6.0000	0.0090	0.4964
β_1	−0.0250	5.0398	−0.0050	0.5020
τ_1	3.2181	0.0025	1313.0597	0.0000
β_2	0.0232	5.4330	0.0043	0.4983
β_3	0.0066	2.4684	0.0027	0.4989
τ_2	0.1309	0.0659	1.9867	0.0235

the points of curvature changes in the term structure. Given the imprecision in the estimated coefficients, we should not be surprised the fit was poor.

3 Smoothed splines and latent variable models

While Nelson-Siegel and Svensson functions are still great tools for yield curve forecasting in the "normal" times, we need methods that are less sensitive to non-parallel shifts in the term structure. Fisher et al. (1995) proposed estimating yield curve at any given point *t* by using the *smoothing splines* method that minimizes the penalized sum of squares

$$\sum_{i=1}^{M} (r_{it} - f(m))^2 + \lambda \int \left(f''(t)\right)^2 dt, \tag{5}$$

where the second part is the penalty term with penalty parameter λ. Smoothing splines can be estimated with the S+FinMetrics function term.struc() with the method="smooth" argument. Consider a smoothed spline estimate of the yield curve for December 2000 shown in Figure 3. Compared to the Nelson-Siegel and Svensson fit, this is a notable improvement. Unlike Nelson-Siegel type of models, the smoothing splines don't have an immediate applicability in yield curve forecasting. In the next section, however, we will show that smoothing splines can indeed be used in forecasting with an improved performance around curvature shifts.

Consider an affine latent variable model written for time *t* as

$$\boldsymbol{r}_t = \boldsymbol{A}\boldsymbol{f}_t + \boldsymbol{e}_t, \tag{6}$$

where $\boldsymbol{r}_t \equiv (r_t^{(1)}, r_t^{(2)}, \ldots, r_t^{(M)})$ is the vector of interest rate data with *M* different maturities, $\boldsymbol{f}_t \equiv (f_t^{(1)}, f_t^{(2)}, \ldots, f_t^{(K)})$ is a vector of latent factors driving the term structure. Finally, $\boldsymbol{e}_t \equiv (e_t^{(1)}, e_t^{(2)}, \ldots, e_t^{(M)})$ is the vector of idiosyncratic terms reflecting the fraction of the term structure not explained by the latent variables. We generally allow for $K < M$. Note that we could easily interpret the first three factors as the level, slope, and curvature of the term structure.

By collecting the equations for all time points into matrices we obtain $\boldsymbol{R}_T = (\boldsymbol{r}_1', \boldsymbol{r}_2', \ldots, \boldsymbol{r}_T')'$, $\boldsymbol{F}_T = (\boldsymbol{f}_1', \boldsymbol{f}_2', \ldots, \boldsymbol{f}_T')'$, and $\boldsymbol{E}_T = (\boldsymbol{e}_1', \boldsymbol{e}_2', \ldots, \boldsymbol{e}_T')'$. We can thus re-write the factor model (6) for the entire term structure as $\boldsymbol{R}_T = \boldsymbol{A}\boldsymbol{F}_T + \boldsymbol{E}_T$, hence we can obtain the closed-form expression for

Table 2: Coefficient correlations

	β_0	β_1	τ_1	β_2	β_3	τ_2
February 2000 estimates						
β_0	1.00000	−0.95598	−0.83066	−0.96056	−0.85922	−0.83961
β_1	−0.95598	1.00000	0.65465	0.99866	0.95890	0.86459
τ_1	−0.83066	0.65465	1.00000	0.67984	0.43794	0.67064
β_2	−0.96056	0.99866	0.67984	1.00000	0.94404	0.88359
β_3	−0.85922	0.95890	0.43794	0.94404	1.00000	0.73999
τ_2	−0.83961	0.86459	0.67064	0.88359	0.73999	1.00000
November 2000 estimates						
β_0	1.00000	−0.97048	0.41694	−0.97647	−0.71620	−0.82744
β_1	−0.97048	1.00000	−0.57941	0.99873	0.80569	0.90240
τ_1	0.41694	−0.57941	1.00000	−0.57441	−0.39409	−0.64945
β_2	−0.97647	0.99873	−0.57441	1.00000	0.77895	0.89117
β_3	−0.71620	0.80569	−0.39409	0.77895	1.00000	0.74432
τ_2	−0.82744	0.90240	−0.64945	0.89117	0.74432	1.00000
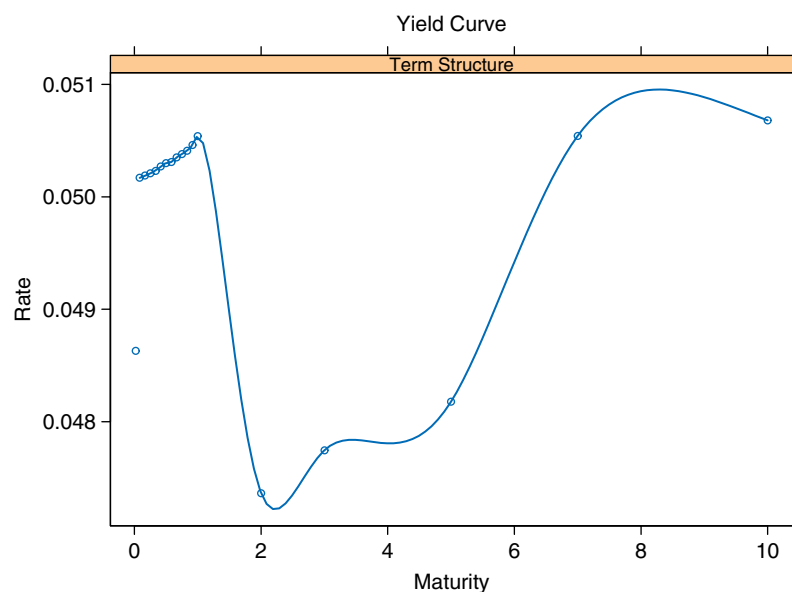
Figure 3: December 2000 yield curve.

A as $\hat{A} = \left(F_T' F_T\right)^{-1} F_T' R_T$. Consequently, the fitted term structure, expressed as an affine function of the latent factors can be obtained as $\hat{R}_T = \left(\left(F_T' F_T\right)^{-1} F_T'\right) F_T$.

It follows, that given the estimates of the future realizations of the latent factors we can predict, say step-ahead, term structure by

$$\hat{r}_{T+1} = \left(\left(F_T' F_T\right)^{-1} F_T'\right) \hat{f}_{T+1}. \tag{7}$$

To use (7) in forecasting we need an estimate of the F_T matrix, as well as the forecast $\hat{f}_{T+1}$. The simplest way of obtaining the scores is to compute K principal components and use the components scores as an estimate of F_T. Subsequently, we can estimate a forecasting model such as MGARCH or VAR on the principal components thereby obtaining an estimate of f_{T+1}.

Before considering alternative forecasting models, we need to briefly check how good are the principal component estimates of the latent factors, i.e., assess the in-sample goodness-of-fit. Retaining $K = 3$ principal components (hence mimicking the three substantive components of the term structure), we compute the predicted (fitted) term structure as $\hat{A}R_T$. Figure 4 shows the fit at the short and long end of the term structure indicating very good fit, therefore we are indeed able to summarize the term structure time series with only three latent factors.

An S function that returns step-ahead forecasts from a vector autoregressive model can be written as

```
Forecast.fun <- function(data)
{
PreDict <- predict(VAR(data,na.rm=T,max.ar=2))[[1]]
return(list(Forecast = PreDict))
}
```
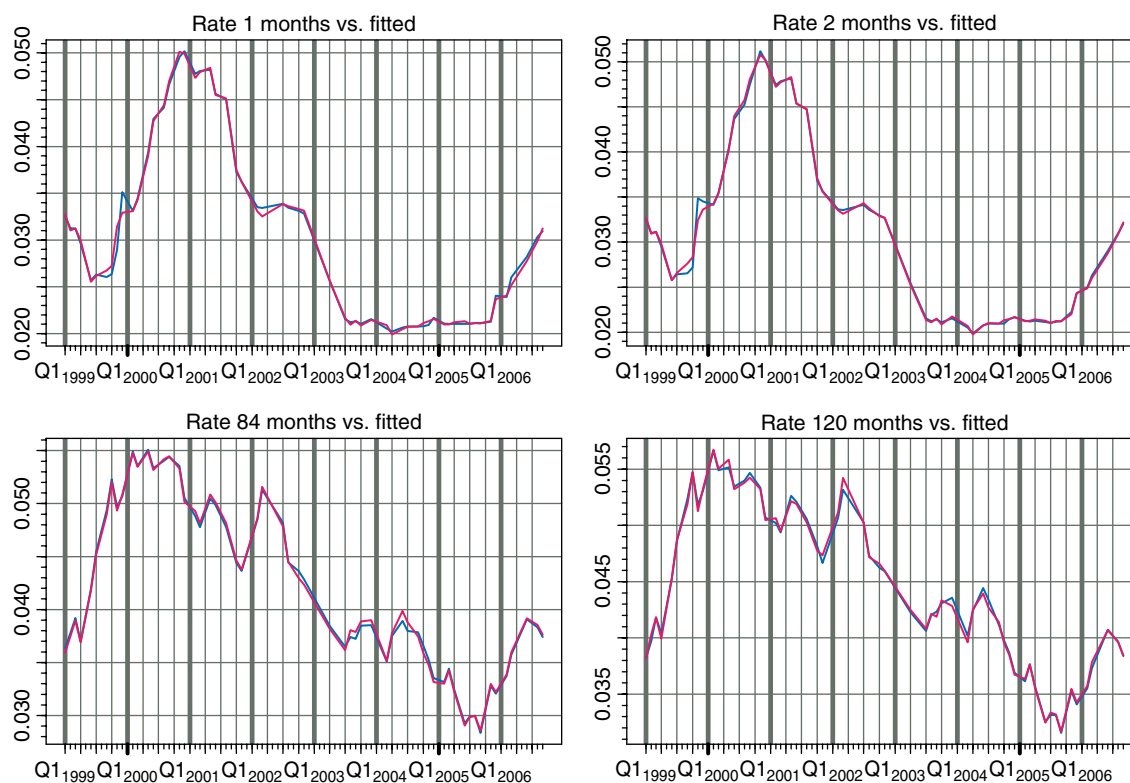

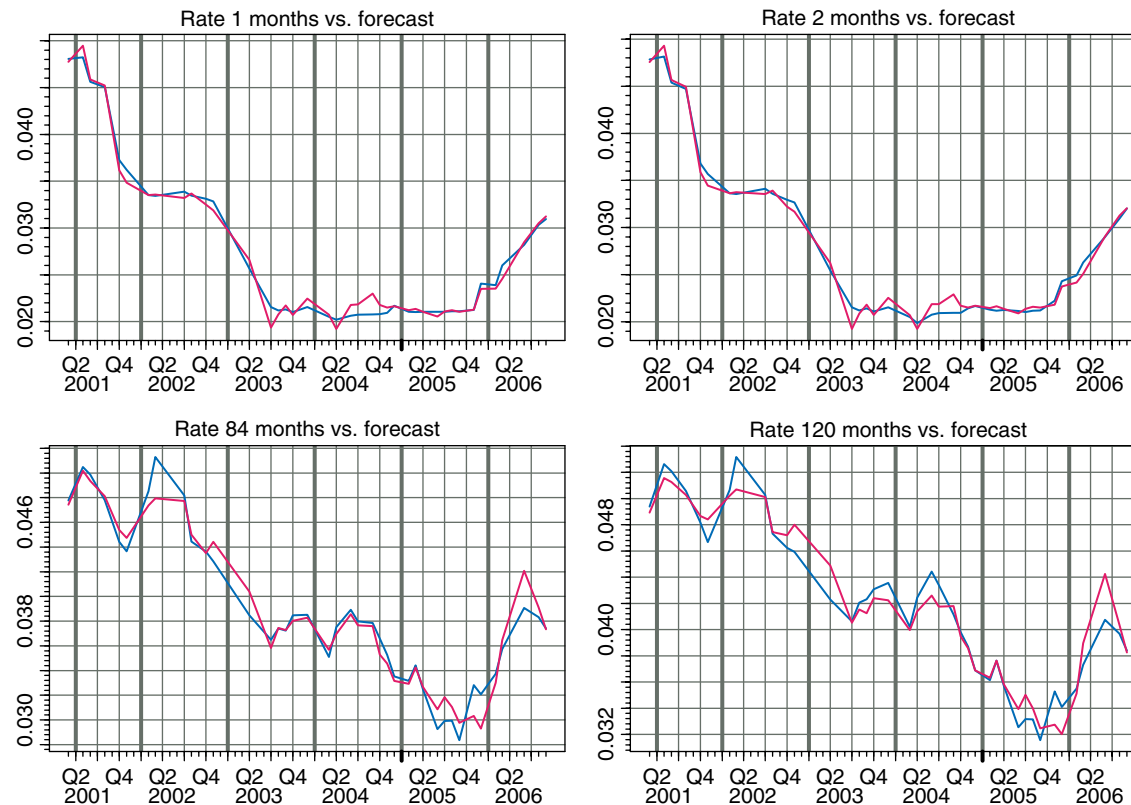**Figure 4: In-sample fit of PCA model.**

Figure 5: Out-of-sample backtest of PCA–VAR model

where `VAR` is a built-in function for estimating vector autoregressive models in `S+FinMetrics` that can automatically select the best fitting VAR model (here we limit the choice to either first or second order VAR due to short time series of 59 monthly observations.) How well can a PCA-VAR model forecast the term structure? The results of a backtest of the step-ahead forecasting performance is shown in Figure 5—we can notice very close fit at the short end and slightly worst fit at the long end.

The backtest itself is relatively easy to implement, namely the following function will return the results of a rolling step-ahead forecasting model, where data is an S-Plus data object of class `timeSeries`, `width` denotes the width of the rolling window and `n` sets the number of principal components to be used in the forecasting model:

```
PCAbacktest <- function(data,width,n)
{
w <- width
PCA.model <- princomp(data,cor=T)
PCA.ts <- timeSeries(pos=positions(data),
          data=data.frame(PCA.model$scores[,1:n]))
MyPCA <- roll(Forecast.fun, data=PCA.ts,
              width=w, trace=T)
F.mat <- cbind(as.matrix(MyPCA$Forecast@data),1)
A.mat <- solve(t(F.mat)%*%F.mat)%*%t(F.mat)%*%
```

```
                as.matrix(data@data[w:nrow(data),])
R.fit     <- F.mat%*%A.mat
DataToReturn <- timeSeries(pos=positions(data)
                 [w:nrow(data)],
                 data=data.frame(data@data
                 [w:nrow(data),],R.fit))
colIds(DataToReturn) <- c(paste("Rate",colIds(data),
                     "months"),
                     paste("Forecast",colIds(data)))
return(DataToReturn)
}
```

Finally, we can apply smoothing splines (6) to the forecasted rates, which for selected months gives the plots in Figure 6, which also shows jackknife confidence intervals. The improvement over the Nelson-Siegel estimates in Figure 2 is notable. We can plot the smoothing estimates for a given time point using a function such as following:

```
SSplot <- function(data,maturity,i=1,MyTitle)
{
x <- maturity;
y <- as.numeric(as.matrix(data@data[i,]));
fit <- smooth.spline(x, y)
```
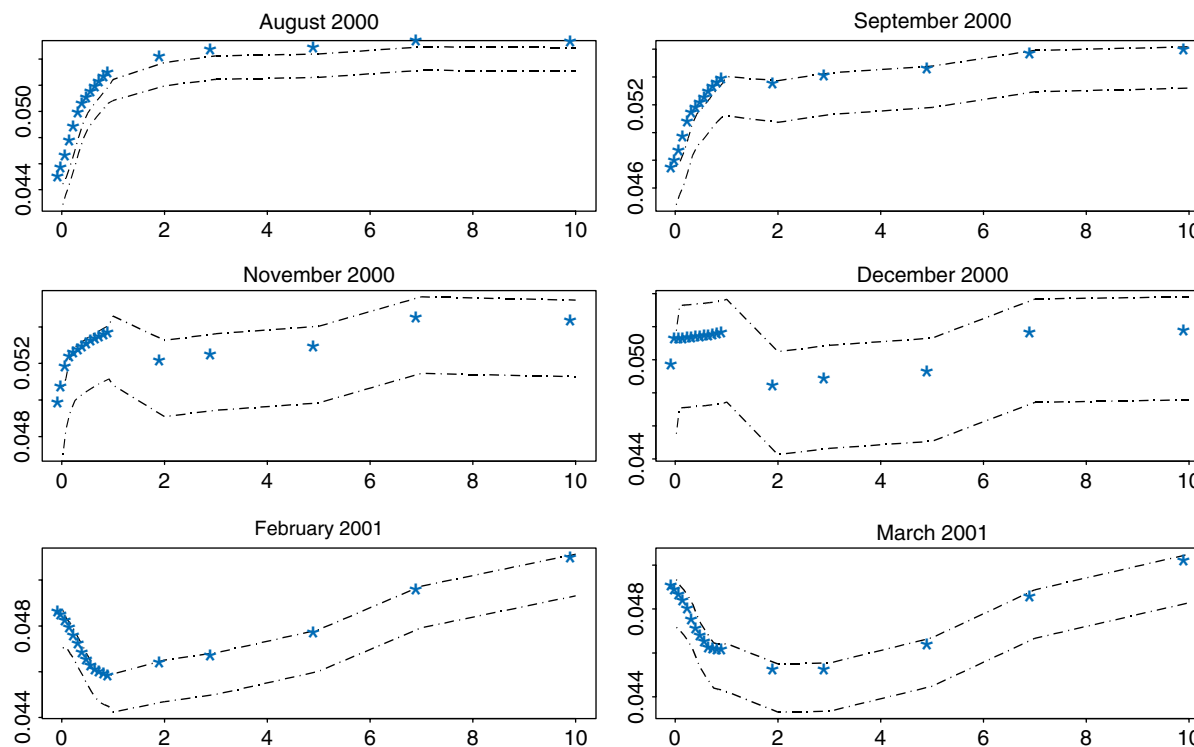
Figure 6: Smoothed spline estimates with confidence intervals.

```
res <- (fit$yin - fit$y)/(1-fit$lev)
sigma <- sqrt(var(res))
upper <- fit$y + 2.0*sigma*sqrt(fit$lev)
lower <- fit$y - 2.0*sigma*sqrt(fit$lev)
matplot(fit$x, cbind(upper,fit$y,lower),
        type="lpl",pch="*",lty=c(3,5),col=c(1,2,1),
        main=paste(MyTitle),cex=2)
invisible()
}
```

3.1 Comparing alternative models

As final comparison, we consider 15 different yield curve forecasting methods and backtest it to evaluate their overall forecasting performance. We consider forecasting the Nelson-Siegel and Svensson model coefficients, following the idea in Diebold and Li (2004) as well as the latent factor models estimated as (6). With the latter, we obtain the smoothing spline estimates of the interest rates of the step-ahead forecast and take the difference between this and the observed rates. With the Nelson-Siegel-Svensson functions, we used the step-ahead coefficients forecasts to estimate the predicted yield curve, again taking the difference between this estimate and the observed rates.

We estimate four multivariate GARCH models (DVEC, EWMA1, EWMA2, and BEKK) along with the linear VAR model for the Nelson-Siegel, Svensson, and latent factor models, respectively. For details of S+FinMetrics implementation of the MGARCH and VAR models see Zivot and Wang (2006).

A simple backtest summary evaluation measure can be computed as

$$\sum_{t=s}^{T} \sum_{i=1}^{N} \left(Y_{ti}\left(m, \theta_t\right) - r_{ti}\right)^2, \tag{8}$$

which is the overall sum of squared forecasting errors, and s denotes the starting time of the backtest. We set $s = 20$ and run the backtests over the 2001–2006 period, again using the EURIBOR time series data. The results are summarized in Table 3. We have used AIC and BIC information criteria to select the best fitting models.

We find that forecasting latent factors estimated by principal components and then spline-smoothing the predicted rates generally gives better forecasts of the yield curve then applying the same forecasting methods to the Nelson-Siegel and Svensson coefficients.

Table 3: Forecasting performance backtest

	NS model	NSS model	PCA
DVEC	0.01627	0.02770	0.00515
EWMA1	0.01621	0.02380	0.00513
EWMA2	0.03627	0.06477	0.01672
BEKK	0.02011	0.02770	0.00516
VAR	0.00101	0.00138	0.00076

To conclude, we can note that changes in non-linear structure of the interest rates are easier to capture by principal components applied directly to the time series of interests rates then by applying forecasting models to the series of Nelson-Siegel-Svensson coefficients. This conclusion seems to hold regardless of the forecasting method used, though we find that VAR model performs better then more complex MGARCH models.

REFERENCES

Diebold, F.X. and Li, C. (2003), Forecasting the Term Structure of Government Bond Yields. *NBER Working Paper*, No. 10048.

El Karoui, N., Frachot, A. and Geman, H. (1998), On the Behavior of Long Zero Coupon Rates in a No Arbitrage Framework. *Revieew of Derivatives Research* **1**, 351–369.

Fisher, M., Nychka, D., and Zervos, D. (1995), Fitting the Term Structure of Interest Rates with Smoothing Splines. *Finance and Economics Discussion Series*, 1995-1. Board of Governors of the Federal Reserve System.

London, J. (2005), *Modeling Derivatives in C++*. Hoboken: John Wiley.

Nelson, C.R. and Siegel, A.F. (1987), Parsimonious Modeling of Yield Curves, *Journal of Business*, **60**(4), 473–489.

Scherer, B. and Martin, R.D. (2005), *Introduction to Modern Portfolio Optimization With NUOPT and S-Plus*. New York: Springer.

Svensson, L.E.O. (1994), Estimating and Interpreting Forward Interest Rates: Sweden 1992–1994. *NBER Working Paper* No. 4871.

Vebables, W.N. and Ripley, B.D. (2001), *S Programming*. New York: Springer.

Zivot, E. and Wang, J. (2006), *Modeling Financial Time Series with S-Plus*. New York: Springer.

W