Journal of Statistical Software

MMMMMM YYYY, Volume VV, Issue II. *http://www.jstatsoft.org/*

Zero-Coupon Yield Curve Estimation with the Package termstrc

Robert Ferstl
University of Regensburg

Josef Hayden
University of Regensburg

Abstract

Zero-coupon yield curves and spread curves are important inputs for various financial models, e.g. pricing of securities, risk management, monetary policy issues. Since zero-coupon rates are rarely directly observable, they have to be estimated from market data~~, e.g. of existing coupon bonds~~. The literature broadly distinguishes between parametric and spline-based estimation methods for the zero-coupon yield curve. Our package consists of several widely-used approaches, i.e. the parametric Nelson and Siegel (1987) method with the Svensson (1994) extension, and the McCulloch (1975) cubic splines approach. Extensive summar~~ies~~ and plots are provided to compare the results of the different estimation methods. We illustrate the application of ~~our functions~~ by practical examples ~~with~~ market data from European government bonds.

Keywords: fixed income, term structure estimation, bond pricing, R.

Around the turn of the last century, a famous Austrian economist, Eugen von Böhm-Bawerk (1851-1914), declared that the cultural level of a nation is mirrored by its rate of interest: the higher a people's intelligence and moral strength, the lower the rate of interest.

A History of Interest Rates, Homer and Sylla (2005)

1. Introduction

The term structure of interest rates or the zero-coupon yield curve is the relationship between fixed income investments with only one payment at maturity, and the time to maturity of th~~ese~~ cashflow~~s~~. It is used in different areas of application, e.g. risk management, financial engineering, monetary policy issues. The zero-coupon yield curve is the basis to value other fixed income instruments. Basically, ~~we can calculate~~ the net present value (NPV) of any cash flow ~~with it~~. For example, the fair price of a bond is ~~just~~ the sum of its discounted future coupon and redemption payments. By comparing ~~it~~ to the price on the market, we can identify mispriced securities. The numerous areas of application for the term structure of interest rates ha~~s~~ lead to a fairly large amount of publications by researchers and practitioners.

1.1. Fixed income basics

Before we come to the problem of zero-coupon yield curve estimation, let us introduce the definitions of a few basic terms used in the fixed income literature. A *discount bond* or *zero-coupon bond* is a fixed income investment with only one payment at maturity. The *spot rate* or *zero-coupon rate* is the interest paid on a discount bond. With continuous compounding the fair price in the absence of arbitrage opportunities of a discount bond paying ~~1~~ Euro at maturity date m_t is given by

$$p(m_t) = e^{s(m_t)m_t}. \tag{1}$$

Therefore, the *spot curve* (or *zero-coupon yield curve*) shows the spot rates for different maturities. ~~The spot rate is the interest paid for investments that start today ($t = 0$).~~ The *forward rate* $f(t', T)$ is the interest contracted now to be paid for an future investment between the time t' and T. The forward rate as a function of maturity is the *forward curve*. With continuous compounding, we have the following relationship between spot rates and forward rates.

$$e^{s(m_{t'})m_{t'}} e^{f(t',T)(m_T - m_{t'})} = e^{s(m_T)m_T}$$

We can solve this for the forward rate.

$$f(t', T) = \frac{s(m_T)m_T - s(m_{t'})m_{t'}}{m_T - m_{t'}}$$

The *instantaneous forward rate* describes the return for an infinitesimal investment period after the date t'.

$$f(t') = \lim_{T \to t'} f(t', T)$$

Another interpretation is the marginal increase in the total return from a marginal increase in the length of the investment period. Thus, the spot rate can be seen as the average of the instantaneous forward rates.

$$s(m_t) = \frac{1}{m_t} \int_0^{m_t} f(s)ds \tag{2}$$

In practice, we can only obtain zero-coupon rates for a limited amount of maturities directly from the market. Therefore, we have to estimate them from observed prices of *coupon bonds.* The following information is typically available on the market: the clean price p^c, the cashflows c_t (coupons and redemption payment) and their maturity dates m_t. An investor who wants to buy the bond has to pay the dirty price p_d, which consists of the quoted market price (clean price) p^c and the accrued interest a. This is the amount of interest that has accumulated since the last coupon payment. Similar to (1), the bond pricing equation under continuous compounding is the present value of all cash flows.

$$p^c + a = \sum_{t=1}^{T} c_t e^{-s(m_t)m_t}$$

An equivalent formulation makes use of the *discount factors* $d_t = \delta(m_t) = e^{-s(m_t)m_t}$. The continuous *discount curve* $\delta(\cdot)$ is formed by interpolation of the discount factors.

$$p^c + a = \sum_{t=1}^{T} c_t \delta(m_t)$$

Each payment (coupons and redemption) has the structure of a discount bond. This makes it possible to relate the coupon bond prices to the spot and the forward curve. The usual way to compare coupon bonds with different maturities is to calculate the internal rate of return of the cash flows. The so-called *yield-to-maturity* (YTM) is the solution for y in the following equation.

$$p^c + a = \sum_{t=1}^{T} c_t e^{-y m_t} \tag{3}$$

As can be seen from (1), the YTM for a discount bond is equal to the spot rate. This does not hold for coupon bonds. Plotting just the yield-to-maturity for coupon bonds with different maturities does not result in a yield curve which can be used to discount cash flows or price any other fixed income security except, the bond from which it was calculated. Therefore, estimating the term structure of interest rates from a set of coupon bonds cannot be seen as a simple curve-fitting of the YTMs.

1.2. Literature review

We have seen before that the spot curve, the forward curve and the discount curve are implied by each other. The following estimation procedures try to approximate one of them, from which it is possible to calculate the others. The simplest method to obtain spot rates from a sample of coupon bonds is bootstrapping. This is an iterative technique based on the pricing equation for a coupon bond. It works only when all cashflows have the same maturity intervals (see, e.g. Hagan and West 2006). Therefore, other estimation procedures are needed, which should fulfill the following requirements. It should price the underlying bonds correctly and result in a continuous spot and forward curve.

Bank for International Settlements (2005) contains a survey about zero-coupon yield curve estimation procedures ~~among~~ central banks. It turns out, that the following two approaches are widely used. The first are spline-based methods for the discount function proposed by McCulloch (1971, 1975). The second approach is based on a parsimonious specification of the forward curve with a family of exponential polynomials developed by Nelson and Siegel (1987) and extended by Svensson (1994). Both methods minimize the price/yield errors, however, estimation procedures are different, e.g. they can ~~involve weightings of~~ the errors, the objective function can become nonlinear.

There are several extensions available for the two methods mentioned above. Vasicek and Fong (1982) fit the discount function with exponential splines. Shea (1985) points out that the estimates are no more stable than the ones from a polynomial model. Fisher, Nychka, and Zervos (1995) proposed a smoothing spline for which Waggoner (1997) introduced a roughness penalty varying across maturities to decrease possible oscillation in the forward rate curve. Different weight~~ing~~s for the objective function of the exponential polynomial families can be found in Söderlind and Svensson (1997). Several works compare the performance of term structure estimation methods, (see, e.g. Bliss 1997; Bolder and Streliski 1999; Ioannides 2003).

In practice, new data for the yield curve is available everyday, and it is obvious to recalibrate the estimation in a dynamic way or even try to forecast the future parameters. Diebold and Li (2006) propose an approach that is based on the Nelson/Siegel model and where they interpret the parameters as factors for level, slope and curvature. Term structure estimation procedures do not have to be consistent with intertemporal interest modeling ~~by~~ diffusion processes (see, e.g. Björk and Christensen 1999; Filipovic 1999). For a consistent and arbitrage-free version of the Nelson/Siegel model, which can be used for pricing fixed income derivatives, see Christensen, Diebold, and Rudebusch (2007).

In this paper, we give a short overview about the topic of term structure estimation methods and introduce the package **termstrc**, which is written in the R system for statistical computing (R Development Core Team 2008). It is available from the Comprehensive R Archive Network at `http://CRAN.R-project.org/` and from the R-Forge development platform at `http://r-forge.r-project.org/projects/termstrc/`. The package provides an implementation of the two most widely-used methods for zero-coupon yield curve estimation from market data of coupon bonds, i.e. the parametric Nelson and Siegel (1987) method with the Svensson (1994) extension, and the McCulloch (1975) cubic splines approach. The software offers detailed summaries about the estimation results as well as graphical outputs of spot, forward, discount and credit spread curves. The code is highly vectorized and is useful for estimations with large data sets.

2. Zero-coupon yield curve estimation

2.1. Notation

Let us establish the necessary notation for a market data set of coupon bonds. We denote an element-wise multiplication with "$\cdot$" and $()'$ is the transpose of a matrix. $\boldsymbol{\iota}$ defines a column vector filled with ones.

Maturity matrix

$$\boldsymbol{M}_{[n\times m]} = \{m_{ij}\}$$

The number of rows n is determined ~~through~~ the number of cashflows of the j-th bond with the longest maturity. ~~For each bond j exists a column with the corresponding cashflow dates.~~ Dates after the maturity of the bond j are filled up with zeros till the maturity date of the bond with the longest maturity. One element m_{ij} of this matrix refers, therefore, to the ~~maturity date~~ of the i-th cashflow of the j-th bond.

Maturity vector

We denote with m_j the maturity of the last cashflow, i.e. the maturity of the j-th bond.

$$\boldsymbol{m}_{[1\times m]} = \{m_j\}$$

Cashflow matrix

$$\boldsymbol{C}_{[n\times m]} = \{c_{ij}\}$$

The cashflow matrix is defined analogously to the maturity matrix. One element c_{ij} refers to the i-th cashflow of the j-th bond. Note, that the last cashflow of a each bond includes the redemption payment.

Discount factor matrix

$$\boldsymbol{D}_{[n\times m]} = \{d_{ij}\}$$

One element d_{ij} of the matrix refers to the discount factor associated with the i-th cashflow of the j-th bond. The discount function $\delta(m_{ij})$ returns the discount factor for a given maturity. In the following sections we present several methods how to estimate it. From an economic point of view only positve interest are appropriate. This implies that the discount factors are nonnegative where the entries in the maturity matrix are greater than zero. Remember, zero entries in the maturity matrix mean that for these points in time now cash flows exist.

Clean price vector

$$\boldsymbol{p}^c_{[1\times m]} = \{p^c_j\}$$

p^c_j is the quoted market price of the j-th bond. It is given as percentage of the nominal value.

Accrued interest vector

When an investor buys a bond, he ~~will~~ receive all its future cash flows. If the purchase occurs between two coupon dates, the seller must be compensated for the fraction of the next coupon, the so-called *accrued interest*.

$$\boldsymbol{a}_{[1\times m]} = \{a_j\}$$

In practice, the calculation depends on the used day-count convention, e.g. 30/360, Actual/360. A basic form for the j-th bond is as follows.

$$a_j = \frac{\text{number of days since last coupon payment}}{\text{number of days in current coupon period}} \cdot \text{coupon}_j$$

Dirty price vector

The dirty price vector is the sum of the clean price and the accrued interest.

$$\boldsymbol{p} = \boldsymbol{p}^c + \boldsymbol{a}$$

The elements are denoted by

$$\boldsymbol{p}_{[1\times m]} = \{p_j\}\,.$$

Yield-to-maturity vector

This vector contains the yield-to-maturity described in (3).

$$\boldsymbol{y}_{[1\times m]} = \{y_j\}\,.$$

Duration vector

~~The time to maturity of a coupon bond should not be used as~~ an indicator for the sensitivity of a bond's price against changes in the interest rate~~. O~~ne needs to account for the fact that coupons are paid during the life~~-~~time of a bond. ~~Therefore, we calculate the average maturity weighted by the present values of its cash flows.~~ ~~This concept is called~~ (Macaulay) *duration*.

$$\boldsymbol{d}_{[1\times m]} = \frac{\boldsymbol{\iota}'(\boldsymbol{C}\cdot\boldsymbol{M}\cdot\boldsymbol{D})}{\boldsymbol{\iota}'(\boldsymbol{C}\cdot\boldsymbol{D})}$$

Here, the discount matrix $\boldsymbol{D}$ contains the disount factors calulated with the yield-to-maturity of each bond as in (3).

Weights matrix

In the next section, we will use the following matrix for weighting the estimation errors:

$$\boldsymbol{\Omega}_{[m\times m]} = \begin{pmatrix} \omega_1 & 0 & \cdots & 0 \\ 0 & \omega_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \omega_m \end{pmatrix},$$

whereas ω_j is the weight for bond j with duration d_j.

$$\omega_j = \frac{\frac{1}{d_j}}{\sum_{i=1}^{m}\frac{1}{d_i}} \tag{4}$$

2.2. Estimation procedure

The simplest smoothing technique is linear interpolation. In many cases it is not appropriate for zero-coupon yield curve estimation, because it can lead to spikes in the forward rate curve, which is ~~a problem from an economic point of view~~. Therefore, we use indirect estimation procedures that postulate a specific form for the spot rate function $s(m_{ij}, \boldsymbol{\beta})$ or the discount function $\delta(m_{ij}, \boldsymbol{\beta})$, where $\boldsymbol{\beta}$ is a vector of parameters (see, e.g. Martellini, Priaulet, and Priaulet 2003). This allows us to construct a discount matrix. The *theoretical bond prices* are then defined as sums of the discounted cash flows of each bond.

$$\hat{\boldsymbol{p}} = \boldsymbol{\iota}'(\boldsymbol{C} \cdot \boldsymbol{D}) \tag{5}$$

The *pricing errors*

$$\boldsymbol{\epsilon}_p = \boldsymbol{p} - \hat{\boldsymbol{p}}$$

are the deviation of the theoretical prices from the dirty prices observed on the market. Analogously, we can define the *yield errors*

$$\boldsymbol{\epsilon}_y = \boldsymbol{y} - \hat{\boldsymbol{y}},$$

where~~as~~ $\hat{\boldsymbol{y}}$ are the yield-to-maturities ~~calculated with~~ the theoretical bond prices. The errors satisfy

$$\begin{aligned}
\mathsf{E}(\boldsymbol{\epsilon}) &= \mathbf{0}\,, \\
\mathsf{VAR}(\boldsymbol{\epsilon}) &= \sigma^2 \boldsymbol{\Omega}^2\,, \\
\mathsf{COV}(\epsilon_i, \epsilon_j) &= 0 \quad \text{for}\, i \neq j\,.
\end{aligned}$$

The goal is to find the parameters that minimize the weighted squared errors.

$$\hat{\boldsymbol{\beta}} = \arg\min_{\hat{\boldsymbol{\beta}}} \boldsymbol{\iota}'(\boldsymbol{\epsilon}^2 \boldsymbol{\Omega})' \tag{6}$$

The goodness of fit can be measured for example with the *root mean squared error*

$$\text{RMSE} = \sqrt{\frac{1}{m} \boldsymbol{\epsilon}^2 \boldsymbol{\iota}}\,,$$

or the *mean absolute error*

$$\text{MAE} = \frac{1}{m} |\boldsymbol{\epsilon}| \boldsymbol{\iota}\,.$$

The next two sections present popular ways to specify a form for the spot rate or the discount function and solve the optimization problem given in (6).

3. Nelson/Siegel and Svensson method

Nelson and Siegel (1987) propose a parsimonious model of the instantaneous forward rate as a solution to a second-order differential equation for the case of equal roots

$$f(m_{ij}, \boldsymbol{\beta}) = \beta_0 + \beta_1 \exp\left(-\frac{m_{ij}}{\tau_1}\right) + \beta_2 \left[\left(\frac{m_{ij}}{\tau_1}\right) \exp\left(-\frac{m_{ij}}{\tau_1}\right)\right], \tag{7}$$

with a parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \tau_1)$. As in (2), the spot rate is the average of the instantaneous forward rates

$$s(m_{ij}, \boldsymbol{\beta}) = \frac{1}{m_{ij}} \int_0^{m_{ij}} f(m_{ij}, \boldsymbol{\beta})\, dm_{ij}\,,$$

resulting in

$$s(m_{ij}, \boldsymbol{\beta}) = \beta_0 + \beta_1 \frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{\frac{m_{ij}}{\tau_1}} + \beta_2 \left(\frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{\frac{m_{ij}}{\tau_1}} - \exp(-\frac{m_{ij}}{\tau_1})\right). \tag{8}$$

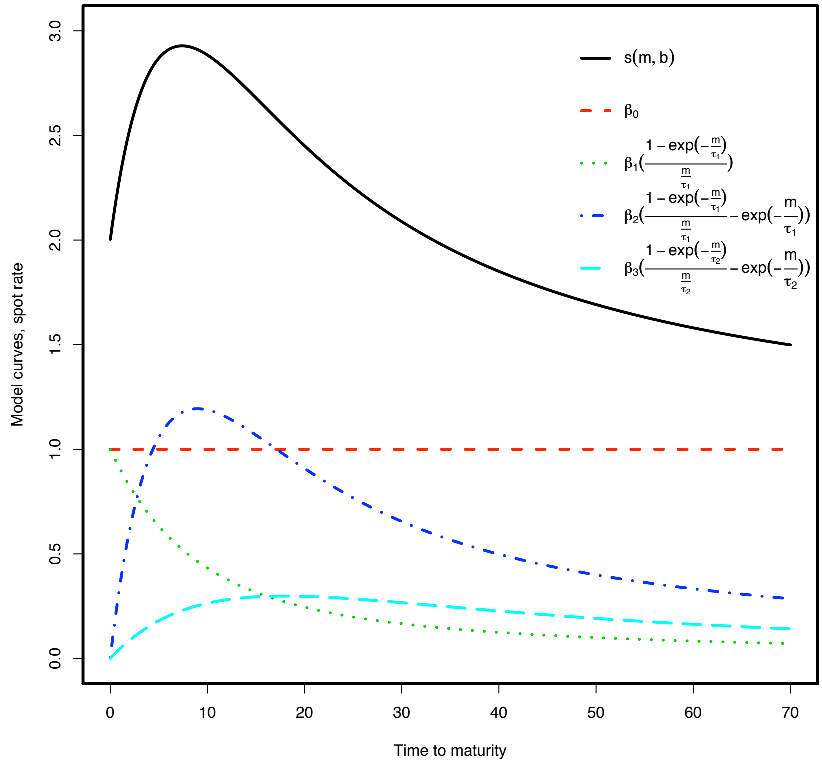
This specification can produce a wide range of possible curve shapes, including monotonic, humped, U-shapes or S-shapes. Svensson (1994) add~~ed~~ another term with two new parameters to increase the flexibility. It allows for a second hump in the curve. The spot rate function is then defined as

$$\begin{aligned} s(m_{ij}, \boldsymbol{\beta}) = \beta_0 + \beta_1 \frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{\frac{m_{ij}}{\tau_1}} + \beta_2 \left(\frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{\frac{m_{ij}}{\tau_1}} - \exp(-\frac{m_{ij}}{\tau_1})\right) \\ + \beta_3 \left(\frac{1 - \exp(-\frac{m_{ij}}{\tau_2})}{\frac{m_{ij}}{\tau_2}} - \exp(-\frac{m_{ij}}{\tau_2})\right), \end{aligned} \tag{9}$$

with a parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \tau_1, \beta_3, \tau_2)$. Figure 1 shows the Svensson (1994) spot rate function and the impact of the different components ~~with~~ $\boldsymbol{\beta} = (1, 1, 4, 5, 1, 10)$. The parameters have the following interpretations:

- $\beta_0 > 0$ is the asymptotic value of the spot rate function $\lim_{m_{ij}\to\infty} s(m_{ij}, \boldsymbol{\beta})$, which can be seen as long-term interest rate.
- β_1 determines the rate of convergence which with the spot rate function approaches its long-term trend, and $\beta_0 + \beta_1$ is the starting value of the curve at the short end. The slope will be negative if $\beta_1 > 0$ and vice versa.
- β_2 determines the size and the form of the hump. $\beta_2 > 0$ results in a hump at τ_1, whereas $\beta_2 < 0$ produces a U-shape.
- $\tau_1 > 0$ specifies the location of the first hump or the U-shape on the curve.
- β_3, analogously to β_2, determines the size and form of the second hump.
- $\tau_2 > 0$ specifies the position of the second hump.

Figure 1: Model curves



The discount factor for any maturity can be calculated as follows:

$$\delta(m_{ij}, \boldsymbol{\beta}) = e^{-m_{ij} s(m_{ij}, \boldsymbol{\beta})} ,$$

where $s(m_{ij}, \boldsymbol{\beta})$ is the Nelson/Siegel or Svensson spot rate function defined in (8) and (9). We optimize the objective function in (6). The above specification of the discount function leads to a nonlinear optimization. Good starting values for the parameter vector are important to find a global minimum. Instead of ~~the~~ pricing errors~~,~~ it is common to minimize the yield errors. ~~Calculating the~~ yield-to-maturities of the theoretical bond prices ~~is easy to solve~~ numerically, because (3) has only one real root.

When minimizing the unweighted pric~~e~~ errors, bonds with a longer maturity get a higher weighting, because they are more sensitive to changes in prices, which leads to a less accurate fit at the short end. Solutions to this heteroskedasticity problem are to use weights for the price errors, or to minimize the yield errors. A possible specification for the weights is based on the inverse of the duration as in (4) (see Bliss 1997).

4. Cubic splines

McCulloch (1971, 1975) use the following definition of the discount function.

$$\delta(m_{ij}, \boldsymbol{\beta}) = 1 + \sum_{l=1}^{k} \beta^l g^l(m_{ij}) \tag{10}$$

It is a linear combination of functions satisfying $g^l(0) = 0$, and the unknown parameter vector $\boldsymbol{\beta}$ will be estimated with ordinary least squares (OLS). This piecewise function is twice-differentable at each knot point, which results in a smooth curve.

4.1. Knot point selection

McCulloch (1975) defines a k-parameter spline with $k-1$ knot points q_l. We sort the cashflow matrix $\boldsymbol{C}$ and the maturity matrix $\boldsymbol{M}$ such that the m bonds are arranged in ascending order by their maturity dates $\boldsymbol{m}$. The following specification places an approximately equal number of bonds between adjacent knots. It sets $q_1 = 0$ and $q_{k-1} = m_m$. m_j is the maturity date of the j-th bond. For $1 < l < k-1$ we find the further knot points:

$$q_l = m_h + \theta(m_{h+1} - m_h)$$

where

$$h = \left\lceil \frac{(l-1)m}{k-2} \right\rceil$$

and

$$\theta = \frac{(l-1)m}{k-2} - h.$$

McCulloch (1971) sets the number of basis functions k to the integer nearest to the square root of the number of observed bonds. This allows a smooth fit of the discount function.

$$k = \lfloor \sqrt{m} + 0.5 \rfloor$$

4.2. Basis functions for cubic splines

In order to generate the family of cubic splines relative to these knots, we define for $m_{ij} < q_{l-1}$

$$g^l(m_{ij}) = 0.$$

For $q_{l-1} \leq m_{ij} < q_l$, we define

$$g^l(m_{ij}) = \frac{(m_{ij} - q_{l-1})^3}{6(q_l - q_{l-1})}$$

When $q_l \leq m_{ij} < q_{l+1}$, we define

$$g^l(m_{ij}) = \frac{c^2}{6} + \frac{ce}{2} + \frac{e^2}{2} - \frac{e^3}{6(q_{l+1} - q_l)},$$

where

$$c = q_l - q_{l-1}$$

and e is

$$e = m_{ij} - q_l .$$

For $q_{l+1} \leq m_{ij}$, we define

$$g^l(m_{ij}) = (q_{l+1} - q_{l-1}) \left[\frac{2q_{l+1} - q_l - q_{l-1}}{6} + \frac{m_{ij} - q_{l+1}}{2} \right] .$$

(Set $q_{l-1} = q_l = 0$ when $l = 1$.)
The above formulas apply when $l < k$. When $l = k$, we define

$$g^l(m_{ij}) = m_{ij},$$

regardless of m_{ij}.
The basis functions are calulated for the maturities of each cashflow $m_{i,j}$ and summarized in the matrix

$$\boldsymbol{G}^l_{[n \times m]} = \{g^l_{ij}\} ,$$

where $g^l_{ij} = g^l(m_{ij})$.

4.3. Regression fitting of the discount function

The dirty prices are again expressed as the sum of the discounted cashflows plus an idiosyncratic error.

$$\boldsymbol{p} = \boldsymbol{\iota}' \left(\boldsymbol{C} \cdot \boldsymbol{D} \right) + \boldsymbol{\epsilon} \tag{11}$$

The discount factor matrix is defined as the weighted sum of the $l = 1 \dots k$ basis functions.

$$\boldsymbol{D} = 1 + \beta^1 \boldsymbol{G}^1 + \cdots + \beta^k \boldsymbol{G}^k \tag{12}$$

We substitute (12) in (11) and get an expression which is linear in the parameter vector $\boldsymbol{\beta} = (\beta^1, \dots, \beta^k)$.

$$
\begin{aligned}
\boldsymbol{p} &= \boldsymbol{\iota}' \left(\boldsymbol{C} \cdot \left(1 + \beta^1 \boldsymbol{G}^1 + \cdots + \beta^k \boldsymbol{G}^k\right)\right) + \boldsymbol{\epsilon} \\
\boldsymbol{p} &= \boldsymbol{\iota}' \left(\boldsymbol{C} + \boldsymbol{C} \cdot \left(\beta^1 \boldsymbol{G}^1 + \cdots + \beta^k \boldsymbol{G}^k\right)\right) + \boldsymbol{\epsilon} \\
\boldsymbol{p} &= \boldsymbol{\iota}'\boldsymbol{C} + \boldsymbol{\iota}'\boldsymbol{C} \cdot \left(\beta^1 \boldsymbol{G}^1 + \cdots + \beta^k \boldsymbol{G}^k\right) + \boldsymbol{\epsilon} \\
\boldsymbol{p} - \boldsymbol{\iota}'\boldsymbol{C} &= \beta^1 \boldsymbol{\iota}'\boldsymbol{C} \cdot \boldsymbol{G}^1 + \cdots + \beta^k \boldsymbol{\iota}'\boldsymbol{C} \cdot \boldsymbol{G}^k + \boldsymbol{\epsilon}
\end{aligned}
$$

We summarize the terms on both sides as follows:

$$
\boldsymbol{X}_{[m \times k]} = \{\boldsymbol{x}_{[m \times 1]}\} \qquad \boldsymbol{x}_{[m \times 1]} = \left(\boldsymbol{\iota}'\boldsymbol{C} \cdot \boldsymbol{G}^l\right)'
$$

$$
\boldsymbol{z}_{[m \times 1]} = \left(\boldsymbol{p} - \boldsymbol{\iota}'\boldsymbol{C}\right)'
$$

$$
\boldsymbol{z} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{13}
$$

The unknown parameters can now be estimated with OLS.

$$
\hat{\boldsymbol{\beta}}_{[k \times 1]} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}'\boldsymbol{z}
$$

We can use the resulting parameters to calculate the discount function in (10) for any given maturity m_{ij} between the first and the last knot point, which can then be converted to the spot rate function.

$$
s(m_{ij}, \boldsymbol{\beta}) = \frac{-\ln \delta(m_{ij}, \boldsymbol{\beta})}{m}
$$

4.4. Confidence intervals for the discount function

McCulloch (1975) plots error bands ~~above and below~~ one standard error of the best estimate. We derive a confidence interval for the predicted discount function. Under the assumption of normally distributed disturbances $\boldsymbol{\epsilon}$, the ordinary least squares coefficient estimator of (13) is normally distributed with mean $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma^2 \boldsymbol{X}'\boldsymbol{X}^{-1}$.

Following Greene (2002), the confidence interval for a linear combination of coefficients can be obtained by applying the decomposition of Oaxaca (1973). Therefore, the discount function $\delta(m_{ij})$ in (10) is normally distributed with mean

$$
\mu = 1 + \boldsymbol{g}(m_{ij})'\boldsymbol{\beta}
$$

and variance

$$
\sigma^2 = \boldsymbol{g}(m_{ij})' \left(\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \boldsymbol{g}(m_{ij})
$$

whereas $\boldsymbol{g}(m_{ij}) = \left(g^1(m_{ij}), \ldots, g^l(m_{ij})\right)$.

The $1-\alpha$ confidence interval for μ of the discount function can now be constructed the usual way

$$\mathrm{P}\left[\delta(m_{ij}) - t_{\alpha/2}s \leq \mu \leq \delta(m_{ij}) + t_{\alpha/2}s\right] = 1-\alpha,$$

where s is the estimate for σ, $t_{\alpha/2}$ the appropriate critical value from the t-distribution with $m-k$ degrees of freedom and $1-\alpha$ the desired level of confidence.

5. Practical application

The goverment debt market is the common data source for estimating a zero-coupon yield curve of a country. Government bonds are usually the most liquid securities, and can be considered default-free, provided the issuing country has a good rating. We demonstrate the application of our package with a data set of European government bonds obtained from Thomson Financial Datastream. The following examples of the two mentioned procedures, as well as a dynamic estimation of the zero-coupon yield curve.

5.1. Parametric methods

We load the package with the following command.

```
R> library("termstrc")
```

In the next steps we load the dataset govbonds and explore its structure.

```
R> data(govbonds)
R> summary(govbonds)
```

```
        Length Class  Mode
GERMANY 8      -none- list
AUSTRIA 8      -none- list
BELGIUM 8      -none- list
FINLAND 8      -none- list
FRANCE  8      -none- list
SPAIN   8      -none- list
```

It includes data for government bonds of eight European countries. The bonds are classified by their *International Securities Identifying Number (ISIN)*, and all the necessary information on the future cash flows is given.

```
R> str(govbonds$GERMANY)
```

```
List of 8
 $ ISIN        : chr [1:52] "DE0001141414" "DE0001137131" "DE0001141422" "DE0001137149" ...
 $ MATURITYDATE:Class 'Date'  num [1:52] 13924 13952 13980 14043 14064 ...
 $ ISSUEDATE   :Class 'Date'  num [1:52] 11913 13215 12153 13298 10411 ...
 $ COUPONRATE  : num [1:52] 0.0425 0.0300 0.0300 0.0325 0.0413 ...
 $ PRICE       : num [1:52] 100.0  99.9  99.8  99.8 100.1 ...
 $ ACCRUED     : num [1:52] 4.09 2.66 2.43 2.07 2.39 ...
 $ CASHFLOWS   :List of 3
  ..$ ISIN: chr [1:384] "DE0001141414" "DE0001137131" "DE0001141422" "DE0001137149" ...
  ..$ CF  : num [1:384] 104 103 103 103 104 ...
  ..$ DATE:Class 'Date'  num [1:384] 13924 13952 13980 14043 14064 ...
 $ TODAY       :Class 'Date'  num 13908
```

Suppose, we want to perform a zero-coupon yield curve estimation for several countries with the Nelson and Siegel (1987) method minimizing the duration weighted price errors. The sample of bonds ~~should be~~ restricted to a maximum maturity of 30 years.

```
R> group <- c("GERMANY", "FRANCE", "BELGIUM", "SPAIN")
R> bonddata <- govbonds
R> matrange <- c(0, 30)
R> method <- "Nelson/Siegel"
R> fit <- "prices"
R> weights <- "duration"
R> b <- matrix(rep(c(0, 0, 0, 1), 4), nrow = 4, byrow = TRUE)
R> rownames(b) <- group
R> colnames(b) <- c("beta0", "beta1", "beta2", "tau1")
R> x <- nelson_estim(group, bonddata, matrange, method, fit, weights,
...    b)
```

Now, let us have a look at the results.

```
R> x
```

```
---------------------------------------------------
Parameters for Nelson/Siegel, Svensson estimation:

Method: Nelson/Siegel
Fitted: prices
Weights: duration

---------------------------------------------------

           GERMANY      FRANCE      BELGIUM        SPAIN
beta_0  0.05130485  0.05118613  0.052087336  0.052729208
beta_1 -0.01268934 -0.01242461 -0.008902787  0.002793828
beta_2 -0.03215030 -0.03036739 -0.036863081 -0.047882039
tau_1   2.68940959  2.54290443  2.185687965  1.863315577
```

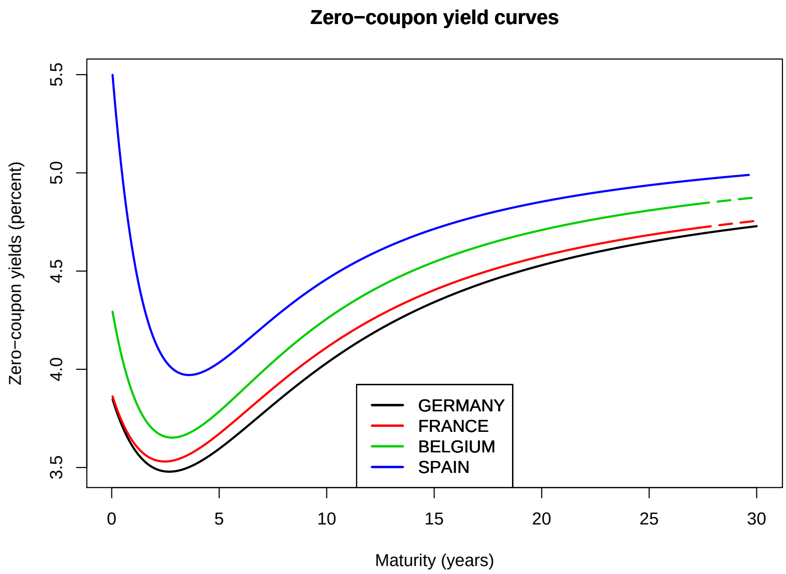
The summary method gives goodness of fit measures for the price and the yield errors. Moreover, it shows the convergence information from the solver~~,~~ to check whether a solution to the nonlinear optimization problem has been found.

```
R> summary(x)
```

```
---------------------------------------------------
Goodness of fit:
---------------------------------------------------

                  GERMANY       FRANCE     BELGIUM       SPAIN
RMSE-Prices  0.3578726990 0.2214576932 2.222911540 2.011818325
AABSE-Prices 0.2030132843 0.1184783548 0.776049451 1.724348860
RMSE-Yields  0.0008413222 0.0003923194 0.005007131 0.008045454
AABSE-Yields 0.0005305512 0.0002735921 0.002114202 0.005710166


---------------------------------------------------
Convergence information:
---------------------------------------------------

        Convergence ()
GERMANY "converged"
FRANCE  "converged"
BELGIUM "converged"
SPAIN   "converged"

        Solver message
GERMANY "relative convergence (4)"
FRANCE  "relative convergence (4)"
BELGIUM "relative convergence (4)"
SPAIN   "relative convergence (4)"
```

Our package offers several options to plot the results, e.g. spot rate, forward rate, discount and spread curves. Figure 2 shows the estimated zero-coupon yield curves. The dashed lines indicate that the curve was extrapolated, which is possible with the Nelson and Siegel (1987) and Svensson (1994) approach.

Figure 2: Zero-coupon yield curves estimated with Nelson/Siegel



5.2. Spline-based methods

In this section, we to demonstrate how to estimate the term structure of interest rates with the McCulloch (1975) cubic splines approach applied to French governement bonds.

```
R> y <- splines_estim(c("FRANCE"), govbonds, c(0, 30))
R> y


---------------------------------------------------
Parameters for Cubic splines estimation:

[1] "FRANCE:"
      alpha 1       alpha 2       alpha 3       alpha 4       alpha 5
 0.0136620572 -0.0018022601 -0.0003772906  0.0001867318  0.0010633259
      alpha 6       alpha 7
 0.0014693772 -0.0408879155
```

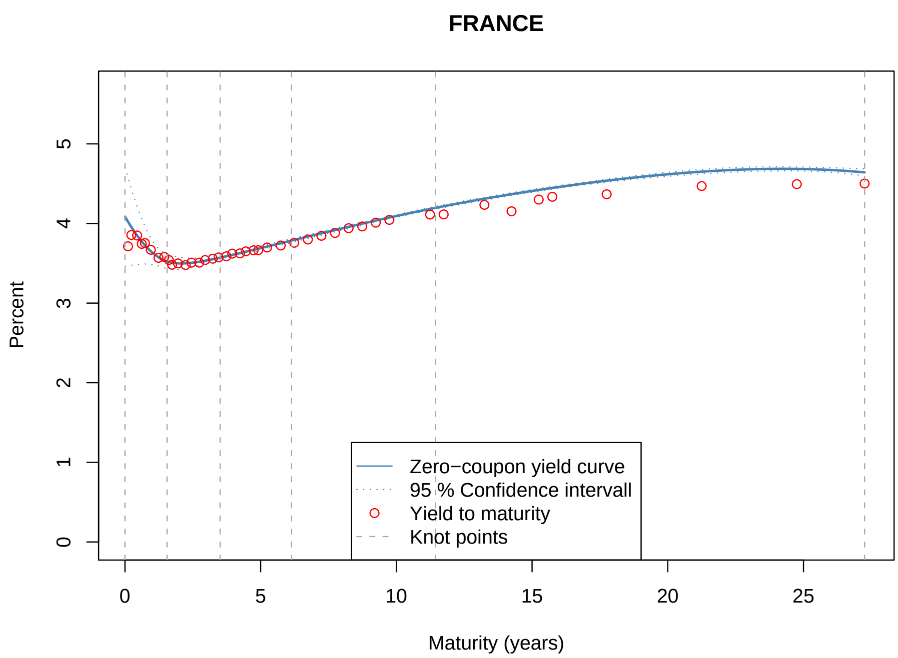
The summary method shows details from the OLS estimation of the parameters and the goodness of fit measures.

```
R> summary(y)


---------------------------------------------------
Goodness of fit:
---------------------------------------------------
```

```
                   FRANCE
RMSE-Prices   0.1819767236
AABSE-Prices  0.0820766103
RMSE-Yields   0.0005212926
AABSE-Yields  0.0002514331

---------------------------------------------------
Summary statistics for the fitted models:
---------------------------------------------------

$FRANCE

Call:
lm(formula = -Y[[k]] ~ X[[k]] - 1)

Residuals:
      Min        1Q    Median        3Q       Max
-0.663324 -0.030923 -0.007992  0.041845  0.908719

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
alpha 1  0.0136621  0.0072072   1.896    0.066 .
alpha 2 -0.0018023  0.0020768  -0.868    0.391
alpha 3 -0.0003773  0.0008139  -0.464    0.646
alpha 4  0.0001867  0.0002862   0.652    0.518
alpha 5  0.0010633  0.0001208   8.804 1.66e-10 ***
alpha 6  0.0014694  0.0002102   6.989 3.39e-08 ***
alpha 7 -0.0408879  0.0032007 -12.774 6.14e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1989 on 36 degrees of freedom
Multiple R-Squared:     1,        Adjusted R-squared:     1
F-statistic: 3.211e+05 on 7 and 36 DF,  p-value: < 2.2e-16
```

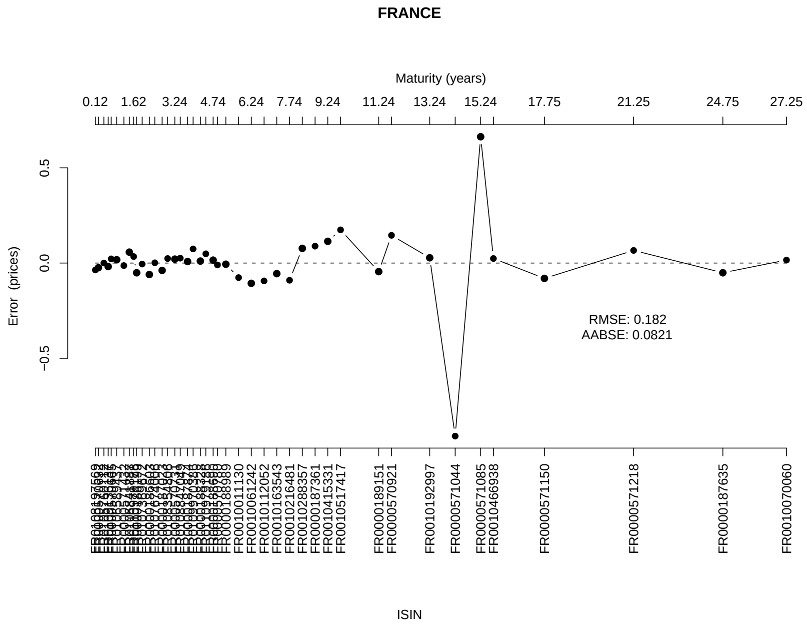
Figure 3 shows the yield-to-maturities and the estimated zero-coupon yield curve together with the automatically selected knot points.

Figure 3: Zero-coupon yield curve for French government bonds estimated with cubic splines



As we can see in Figure 4, there seems to be a miss~~-~~pricing of two bonds. They ~~will~~ be removed and the estimation redone.

Figure 4: Pricing errors for French government bonds



~~The result is,~~ as expected, ~~a better~~ goodness of fit

```
R> z <- splines_estim(c("FRANCE"), rm_bond(bonddata, c("FR0000571044",
...     "FR0000571085"), "FRANCE"), c(0, 30))
```

```
R> summary(z)
```

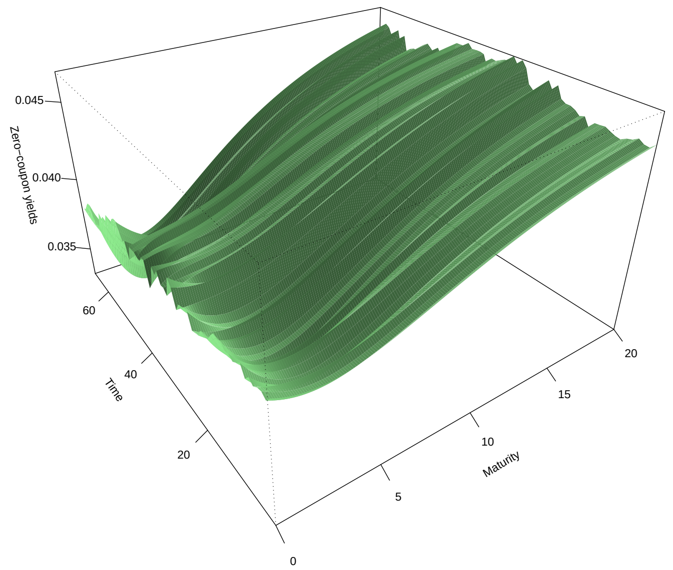
```
---------------------------------------------------
Goodness of fit:
---------------------------------------------------

                   FRANCE
RMSE-Prices  0.0615589340
AABSE-Prices 0.0515614078
RMSE-Yields  0.0003452360
AABSE-Yields 0.0002025576

---------------------------------------------------
Summary statistics for the fitted models:
---------------------------------------------------

$FRANCE

Call:
lm(formula = -Y[[k]] ~ X[[k]] - 1)

Residuals:
     Min       1Q   Median       3Q      Max
-0.10516 -0.03516 -0.01789  0.05302  0.14043

Coefficients:
          Estimate Std. Error t value Pr(>|t|)
alpha 1  8.886e-03  1.558e-03   5.704 1.89e-06 ***
alpha 2 -9.269e-04  3.869e-04  -2.396   0.0221 *
```

```
alpha 3 -5.500e-04  1.175e-04  -4.682 4.17e-05 ***
alpha 4  9.038e-04  3.307e-05  27.334  < 2e-16 ***
alpha 5  1.526e-03  5.085e-05  30.005  < 2e-16 ***
alpha 6 -3.921e-02  8.287e-04 -47.318  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06663 on 35 degrees of freedom
Multiple R-Squared:     1,        Adjusted R-squared:     1
F-statistic: 2.856e+06 on 6 and 35 DF,  p-value: < 2.2e-16
```

5.3. Dynamic estimation

~~In the following,~~ we provide results of a dynamic estimation of the zero-coupon yield curve for the time between November 30, 2007 and February 1, 2008. We use~~d~~ the Svensson (1994) method, together with duration weights and minimization of the price errors. Figure 5 shows ~~how~~ the estimated French yield curves ~~develop within~~ that time period.

Figure 5: Zero-coupon yield curves in France

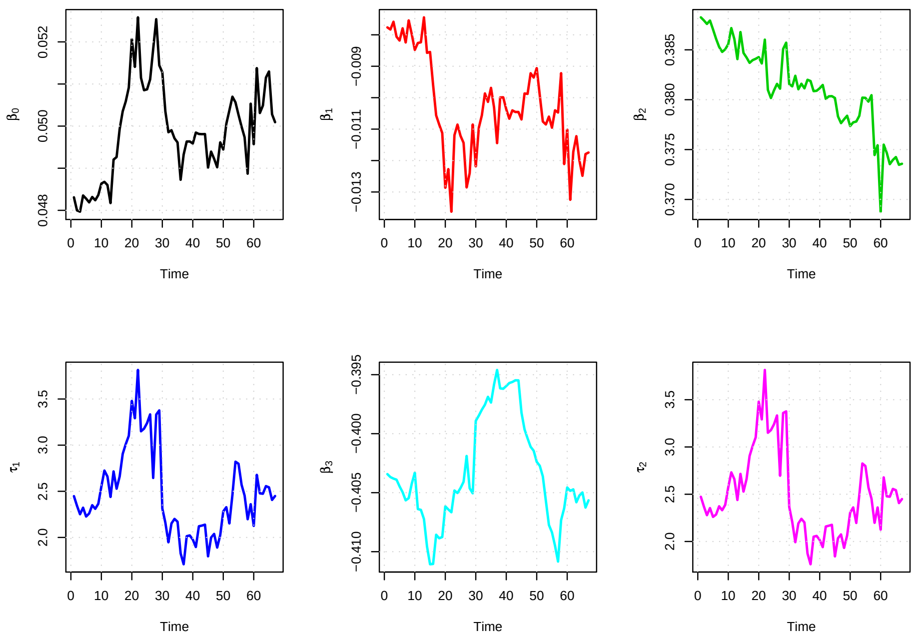


The ~~evolution of the~~ parameters ~~can be seen~~ in Figure 6. To speed up the estimation and ensure that the algorithm stays in a global minimum, the estimated parameters from the previous period were used as start ~~parameters~~ for the next one.

6. Conclusion

In this paper, we presented the R extension package **termstrc**. It provides functions for the estimation of zero-coupon yield curves from market data of coupon bonds. The package covers the two most widely-used approaches in practice and provides a simple interface to them. The

Figure 6: Estimated parameters



results contain detailed summaries about the estimation, as well as graphical outputs of spot, forward, discount and spread curves.

Acknowledgments

The authors want to thank Alois Geyer and Kurt Hornik for their comments on the package and the paper.

References

Bank for International Settlements (2005). "Zero-coupon Yield Curves: Technical Documentation." *BIS Papers 25*, Monetary and Economic Department.

Björk T, Christensen BJ (1999). "Interest Rate Dynamics and Consistent Forward Rate Curves." *Mathematical Finance*, **9**(4), 323–348.

Bliss RR (1997). "Testing Term Structure Estimation Methods." *Advances in Futures and Options Research*, **9**, 197–231.

Bolder D, Streliski D (1999). "Yield Curve Modelling at the Bank of Canada." *Technical Reports 84*, Bank of Canada.

Christensen JH, Diebold FX, Rudebusch GD (2007). "The Affine Arbitrage-Free Class of Nelson-Siegel Term Structure Models." *Technical report*, Federal Reserve Bank of San Francisco.

Diebold FX, Li C (2006). "Forecasting the Term Structure of Government Bond Yields." *Journal of Econometrics*, **130**(2), 337–364.

Filipovic D (1999). "A Note on the Nelson-Siegel Family." *Mathematical Finance*, **9**(4), 349–359.

Fisher M, Nychka D, Zervos D (1995). "Fitting the Term Structure of Interest Rates with Smoothing Splines." *Finance and Economics Discussion Series 95-1*, Board of Governors of the Federal Reserve System (U.S.).

Greene WH (2002). *Econometric Analysis (5th Edition)*. Prentice Hall.

Hagan PS, West G (2006). "Interpolation Methods for Curve Construction." *Applied Mathematical Finance*, **13**(2), 89–129.

Ioannides M (2003). "A Comparison of Yield Curve Estimation Techniques Using UK Data." *Journal of Banking & Finance*, **27**(1), 1–26.

Martellini L, Priaulet P, Priaulet S (2003). *Fixed-Income Securities: Valuation, Risk Management and Portfolio Strategies*. Wiley.

McCulloch JH (1971). "Measuring the Term Structure of Interest Rates." *Journal of Business*, **44**(1), 19–31.

McCulloch JH (1975). "The Tax-Adjusted Yield Curve." *Journal of Finance*, **30**(3), 811–830.

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

Oaxaca R (1973). "Male-Female Wage Differentials in Urban Labor Markets." *International Economic Review*, **14**, 693–709.

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Shea GS (1985). "Interest Rate Term Structure Estimation with Exponential Splines: A Note." *The Journal of Finance*, **40**(1), 319–325.

Svensson LE (1994). "Estimating and Interpreting Forward Interest Rates: Sweden 1992 -1994." *Technical Reports 4871*, National Bureau of Economic Research, Inc.

Söderlind P, Svensson L (1997). "New Techniques to Extract Market Expectations from Financial Instruments." *Journal of Monetary Economics*, **40**(2), 383–429.

Vasicek OA, Fong HG (1982). "Term Structure Modeling Using Exponential Splines." *Journal of Finance*, **37**(2), 339–348.

Waggoner DF (1997). "Spline Methods for Extracting Interest Rate Curves from Coupon Bond Prices." *Working Paper 97-10*, Federal Reserve Bank of Atlanta.

Affiliation:

Robert Ferstl
Department of Finance
University of Regensburg
93053 Regensburg, Germany
E-mail: robert.ferstl@wiwi.uni-regensburg.de
URL: http://www-finanzierung.uni-regensburg.de

Josef Hayden
Department of Finance
University of Regensburg
93053 Regensburg, Germany
E-mail: josef.hayden@wiwi.uni-regensburg.de
URL: http://www-finanzierung.uni-regensburg.de

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
MMMMMM YYYY *Accepted:* yyyy-mm-dd