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. 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 summary statistics and plots are provided to compare the results of the different estimation methods. We illustrate the application of the package by practical examples using market data from European government bonds.

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

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 this cashflow. 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. It can be used to calculate the net present value (NPV) of any cash flow with it. For example, the fair price of a bond is the sum of its discounted future coupon and redemption payments. By comparing this fair price to the price on the market, we can identify mispriced securities. The numerous areas of application for the term structure of interest rates have 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 one Euro at maturity date \( m_t \) is given by

\[
p(m_t) = e^{s(m_t)m_t}.
\]
Therefore, the *spot curve* (or *zero-coupon yield curve*) shows the spot rates for different maturities. The *forward rate* \( f(t', T) \) is the interest contracted now to be paid for a 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
\[
s(m_t) = \frac{1}{m_t} \int_{0}^{m_t} f(s) ds.
\] (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* which provide periodical coupon payments and a redemption payment at maturity. 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 a coupon 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
\[
p_c + a = \sum_{t=1}^{T} c_t e^{-y(m_t)m_t}.
\] (3)

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, such that
\[
p_c + a = \sum_{t=1}^{T} c_t \delta(m_t).
\]

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
\[
p_c + a = \sum_{t=1}^{T} c_t e^{-ym_t}.
\] (4)

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 can not be seen as a simple curve-fitting of the YTM.

1.2. Literature review

We have shown 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 in (3). 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. They should price the underlying bonds correctly and result in a continuous spot and forward curve.

The Bank for International Settlements (2005) contains a survey about zero-coupon yield curve estimation procedures at 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 assign weights to the errors, or 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 weightings 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 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 rate modeling based on 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 particularly 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 \( (\cdot)' \) is the transpose of a matrix. \( \iota \) defines a column vector filled with ones.

**Maturity matrix**

\[
M_{[n \times m]} = \{m_{ij}\}
\]

The number of rows \( n \) is determined by the number of cashflows of the \( j \)-th bond with the longest maturity. Dates after the maturity of the bond \( j \) are filled up with zeros until the maturity date of the bond with the longest maturity. One element \( m_{ij} \) of this matrix refers, therefore, to the time to occurrence (in days) 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.

\[
m_{[1 \times m]} = \{m_j\}
\]

**Cashflow matrix**

\[
C_{[n \times m]} = \{c_{ij}\}
\]

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 each bond includes the redemption payment.

**Discount factor matrix**

\[
D_{[n \times m]} = \{d_{ij}\}
\]

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 positive interest rates 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**

\[
p_{c_{[1 \times m]}} = \{p_j\}
\]

\( p_j \) 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 obtains the right to 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.
\[ a_{1 \times m} = \{a_j\}. \]

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.

\[ p = p^c + a \]

The elements are denoted by

\[ p_{1 \times m} = \{p_j\}. \]

**Yield-to-maturity vector**

This vector contains the yield-to-maturity described in (4).

\[ y_{1 \times m} = \{y_j\} \]

**Duration vector**

To obtain an indicator for the sensitivity of a bond’s price against changes in the interest rate, one needs to account for the fact that coupons are paid during the lifetime of a bond. A standard measure of risk is the (Macaulay) duration which computes the average maturity of a bond using the present values of its cash flows as weights.

\[ d_{1 \times m} = \frac{\iota'(C \cdot M \cdot D)}{\iota'(C \cdot D)} \]

Here, the discount matrix \( D \) contains the discount factors calculated with the yield-to-maturity of each bond as in (4).

**Weights matrix**

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

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

whereas \( \omega_j \) is the weight for bond \( j \) with duration \( d_j \).

\[ \omega_j = \frac{1}{\sum_{i=1}^{m} \frac{1}{d_i}} \] (5)
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 economically implausible. Therefore, we use indirect estimation procedures that postulate a specific form for the spot rate function \( s(m_{ij}, \beta) \) or the discount function \( \delta(m_{ij}, \beta) \), where \( \beta \) 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

\[
\hat{p} = t'(C \cdot D).
\]

The pricing errors

\[ \epsilon_p = p - \hat{p} \]

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

\[ \epsilon_y = y - \hat{y}, \]

where \( \hat{y} \) are the yield-to-maturities based on the theoretical bond prices. The errors satisfy

\[
\begin{align*}
E(\epsilon) &= 0, \\
\text{VAR}(\epsilon) &= \sigma^2 \Omega^2, \\
\text{COV}(\epsilon_i, \epsilon_j) &= 0 \quad \text{for } i \neq j.
\end{align*}
\]

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

\[
\hat{\beta} = \arg \min_{\beta} \epsilon'(\epsilon^2 \Omega)' \tag{7}
\]

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

\[ \text{RMSE} = \sqrt{\frac{1}{m} \epsilon^2 \epsilon}, \]

or the mean absolute error

\[ \text{MAE} = \frac{1}{m} |\epsilon| \mu. \]

The next two sections present popular ways to specify a functional form for the spot rate or the discount function and solve the optimization problem given in (7).

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

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

with a parameter vector \( \beta = (\beta_0, \beta_1, \beta_2, \tau_1) \). As in (2), the spot rate is the average of the instantaneous forward rates.
\[ s(m_{ij}, \beta) = \frac{1}{m_{ij}} \int_0^{m_{ij}} f(m_{ij}, \beta) \, dm_{ij}, \]

resulting in

\[ s(m_{ij}, \beta) = \beta_0 + \beta_1 \frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{m_{ij}} + \beta_2 \left( \frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{m_{ij}} - \exp\left(-\frac{m_{ij}}{\tau_1}\right) \right) + \beta_3 \left( \frac{1 - \exp(-\frac{m_{ij}}{\tau_2})}{m_{ij}} - \exp\left(-\frac{m_{ij}}{\tau_2}\right) \right), \quad (9) \]

This specification can produce a wide range of possible curve shapes, including monotonic, humped, U-shapes or S-shapes. Svensson (1994) adds 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

\[ s(m_{ij}, \beta) = \beta_0 + \beta_1 \frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{m_{ij}} + \beta_2 \left( \frac{1 - \exp(-\frac{m_{ij}}{\tau_1})}{m_{ij}} - \exp\left(-\frac{m_{ij}}{\tau_1}\right) \right) + \beta_3 \left( \frac{1 - \exp(-\frac{m_{ij}}{\tau_2})}{m_{ij}} - \exp\left(-\frac{m_{ij}}{\tau_2}\right) \right) + \beta_4 \left( \frac{1 - \exp(-\frac{m_{ij}}{\tau_3})}{m_{ij}} - \exp\left(-\frac{m_{ij}}{\tau_3}\right) \right), \quad (10) \]

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

- \( \beta_0 > 0 \) is the asymptotic value of the spot rate function \( \lim_{m_{ij} \to \infty} s(m_{ij}, \beta) \), which can be seen as the long-term interest rate.
\[ \beta_1 \text{ determines the rate of convergence with which the spot rate function approaches its long-term trend, and } \beta_0 + \beta_1 \text{ is the starting value of the curve at the short end. The slope will be negative if } \beta_1 > 0 \text{ and vice versa.} \]

\[ \beta_2 \text{ determines the size and the form of the hump. } \beta_2 > 0 \text{ results in a hump at } \tau_1, \text{ whereas } \beta_2 < 0 \text{ produces a U-shape.} \]

\[ \tau_1 > 0 \text{ specifies the location of the first hump or the U-shape on the curve.} \]

\[ \beta_3, \text{ analogously to } \beta_2, \text{ determines the size and form of the second hump.} \]

\[ \tau_2 > 0 \text{ specifies the position of the second hump.} \]

The discount factor for any maturity \( m_{ij} \) can be calculated as follows:

\[ \delta(m_{ij}, \beta) = e^{-m_{ij}s(m_{ij}, \beta)}, \]

where \( s(m_{ij}, \beta) \) is the Nelson/Siegel or Svensson spot rate function defined in (9) and (10). We optimize the objective function in (7). The above specification of the discount function leads to a nonlinear optimization problem. Good starting values for the parameter vector are important to find a global minimum. Instead of minimizing pricing errors, it is common to minimize the yield errors. The yield-to-maturities of the theoretical bond prices can easily be calculated numerically, because (4) has only one real root.

When minimizing the unweighted price 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 pricing errors, or to minimize the yield errors. A possible specification for the weights is based on the inverse of the duration as in (5) (see Bliss 1997).

4. Cubic splines

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

\[ \delta(m_{ij}, \beta) = 1 + \sum_{l=1}^{k} \beta_l g_l(m_{ij}). \]

(11)

It is a linear combination of functions satisfying \( g_l'(0) = 0 \), and the unknown parameter vector \( \beta \) 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 \( C \) and the maturity matrix \( M \) such that the \( m \) bonds are arranged in ascending order by their maturity dates \( m \). 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.

\[ k = \lfloor \sqrt{m} + 0.5 \rfloor \]

This allows a smooth fit of the discount function.

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 calculated for the maturities of each cashflow \( m_{ij} \) and summarized in the matrix

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

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

\[ p = \iota' (C \cdot D) + \epsilon. \]  

The discount factor matrix is defined as the weighted sum of the \( l = 1 \ldots k \) basis functions

\[ D = 1 + \beta^1 G^1 + \cdots + \beta^k G^k. \]  

We substitute (13) in (12) and get an expression which is linear in the parameter vector \( \beta = (\beta^1, \ldots, \beta^k) \).

\[
p = \iota' (C \cdot (1 + \beta^1 G^1 + \cdots + \beta^k G^k)) + \epsilon
\]

\[
p = \iota' (C + C \cdot (\beta^1 G^1 + \cdots + \beta^k G^k)) + \epsilon
\]

\[
p = \iota' C + \iota' C \cdot (\beta^1 G^1 + \cdots + \beta^k G^k) + \epsilon
\]

\[
p - \iota' C = \beta^1 \iota' C \cdot G^1 + \cdots + \beta^k \iota' C \cdot G^k + \epsilon
\]

We summarize the terms on both sides as follows:

\[
X_{[m \times k]} = \{x_{[m \times 1]}\} \quad x_{[m \times 1]} = (\iota' C \cdot G^l)'
\]

\[
z_{[m \times 1]} = (p - \iota' C)'
\]

\[
z = X\beta + \epsilon \quad (14)
\]

The unknown parameters can now be estimated with OLS.

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

We can use the resulting parameters to calculate the discount function in (11) for any given maturity \( m_{ij} \) between the first and the last knot point, which can then be converted to the spot rate function

\[
s(m_{ij}, \beta) = -\ln \delta(m_{ij}, \beta) / m
\]

4.4. Confidence intervals for the discount function

McCulloch (1975) plots error bands one standard error above and below of the best estimate. We derive a confidence interval for the predicted discount function. Under the assumption of normally distributed disturbances \( \epsilon \), the ordinary least squares coefficient estimator of (14) is normally distributed with mean \( \beta \) and variance-covariance matrix \( \sigma^2 X'X^{-1} \).

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 (11) is normally distributed with mean

\[
\mu = 1 + g(m_{ij})'\beta
\]

and variance
\[ \sigma^2 = g(m_{ij})' \left( \sigma^2 (X'X)^{-1} \right) g(m_{ij}), \]

where \( g(m_{ij}) = (g_1(m_{ij}), \ldots, g_l(m_{ij})) \).

The 1 - \( \alpha \) confidence interval for \( \mu \) of the discount function can now be constructed in the usual way

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

where \( s \) is the estimate for \( \sigma \), \( t_{\alpha/2} \) 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 government 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 show the application of the package using the before mentioned procedures, as well as a rolling 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 data set `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 six 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 pricing errors. The sample of bonds is restricted to a maximum maturity of 30 years.
Zero-Coupon Yield Curve Estimation with the Package termstrc

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, ...
R> b)

Now, let us have a look at the results.

R> x

Parameter estimation:
Method: Nelson/Siegel
Fitted: prices
Weights: duration

<table>
<thead>
<tr>
<th></th>
<th>GERMANY</th>
<th>FRANCE</th>
<th>BELGIUM</th>
<th>SPAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>beta_0</td>
<td>0.0513</td>
<td>0.0512</td>
<td>0.0521</td>
<td>0.0527</td>
</tr>
<tr>
<td>beta_1</td>
<td>-0.0127</td>
<td>-0.0124</td>
<td>-0.0089</td>
<td>0.0028</td>
</tr>
<tr>
<td>beta_2</td>
<td>-0.0322</td>
<td>-0.0304</td>
<td>-0.0369</td>
<td>-0.0479</td>
</tr>
<tr>
<td>tau_1</td>
<td>2.689</td>
<td>2.543</td>
<td>2.186</td>
<td>1.863</td>
</tr>
</tbody>
</table>

The summary method gives goodness of fit measures for the pricing 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:

<table>
<thead>
<tr>
<th></th>
<th>GERMANY</th>
<th>FRANCE</th>
<th>BELGIUM</th>
<th>SPAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE-Prices</td>
<td>0.358</td>
<td>0.224</td>
<td>2.214</td>
<td>2.012</td>
</tr>
<tr>
<td>AABSE-Prices</td>
<td>0.203</td>
<td>0.118</td>
<td>0.776</td>
<td>1.724</td>
</tr>
<tr>
<td>RMSE-Yields</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>AABSE-Yields</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Convergence information:

<table>
<thead>
<tr>
<th></th>
<th>GERMANY</th>
<th>FRANCE</th>
<th>BELGIUM</th>
<th>SPAIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convergence ()</td>
<td>converged</td>
<td>converged</td>
<td>converged</td>
<td>converged</td>
</tr>
</tbody>
</table>

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.

5.2. Spline-based methods

In this section, we demonstrate how to estimate the term structure of interest rates with the McCulloch (1975) cubic splines approach applied to French government bonds.
Figure 2: Zero-coupon yield curves estimated with Nelson/Siegel

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:
1m(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 0.0665
alpha 2 -0.0018023  0.0020768 -0.868 0.3913
alpha 3 -0.0003773  0.0003729  0.006 0.9968
alpha 4 0.0001867  0.0002862  0.652 0.5178
alpha 5 0.0010633  0.0010566  0.999 0.3222
alpha 6 0.0014694  0.0014694  1.000 0.3168
alpha 7 -0.0408879 -0.0408879 -1.000 0.3168
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 misspricing of two bonds. They can be removed and the estimation is redone.

As expected, the goodness of fit is improved.

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

R> summary(z)

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

<table>
<thead>
<tr>
<th>Call:</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{RMSE-Prices}$: 0.0615589340</td>
</tr>
<tr>
<td>$\text{AABSE-Prices}$: 0.0515614078</td>
</tr>
<tr>
<td>$\text{RMSE-Yields}$: 0.0003452360</td>
</tr>
<tr>
<td>$\text{AABSE-Yields}$: 0.0002025576</td>
</tr>
</tbody>
</table>

Summary statistics for the fitted models:

$\text{FRANCE}$

Call:
5.3. Rolling estimation procedure

We now provide results of a daily rolling estimation of the zero-coupon yield curve for the time between November 30, 2007 and February 1, 2008. We use the Svensson (1994) method, together with duration weights and minimization of the price errors. Figure 5 shows the estimated French yield curves during that time period.

The estimated parameters are presented 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 starting values for the next one.
Figure 5: Zero-coupon yield curves in France

Figure 6: Estimated parameters
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 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


