Non-parametric confidence intervals of instantaneous forward rates

Jacques F. Carriere*

Department of Mathematical Sciences, University of Alberta, 632 Central Academic Building, Edmonton, Alta., Canada T6G 2G1

Received 17 March 1998; accepted 19 January 2000

Abstract

Using the price of US Treasury Strips, we show how to estimate forward rates with spline models. Confidence intervals on these rates are constructed with bootstrap methods. An unusual feature of the data is the heteroscedastic and correlated error structure. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Term structure of interest rates; Instantaneous forward rates; Strip bonds; Bootstrapping; Splines; Correlation; Heteroscedasticity

1. Introduction

Let \( b_m \) denote the price of a default-free zero-coupon bond at a fixed time that matures \( m \) years in the future with a redemption value of $1. Next, let \( f_m \) denote the instantaneous forward rate for a maturity of \( m \), thus

\[
\frac{\partial}{\partial m} \log e [b_m].
\] (1.1)

In this paper we will focus on the estimation and the construction of confidence intervals for \( f_m \). A large body of research exists on methods of estimation for \( f_m \), \( b_m \), and the yield rate \( -\log_e (b_m)/m \). Consult Anderson et al. (1996) for a comprehensive book on this topic. The estimation methods can be grouped into parametric and non-parametric techniques. An example of a good-fitting parametric formula is the formula presented by Svensson (1994). However, our view is that the best-fitting formulas are non-parametric formulas like splines. Some examples of papers that have used splines are: Shea (1985), Vasicek and Fong (1982) and Delbaen and Lorimier (1992). In all these papers special attention was given to the estimation problem but little attention was given to the inference issues. The estimation of \( f_m \) is a relatively easy exercise, however, the construction of confidence intervals and the characterization of the distribution of the estimator is not trivial because of heteroscedasticity and correlation in the observed values.

We will estimate \( f_m \) by using price data on Treasury Strips from the *Wall Street Journal* (WSJ). We prefer to use the prices for US Treasury Strips because the ask and bid prices on a strip is approximately equal to the price function \( b_m \). The WSJ gives the bid and asked prices for “stripped coupon interest”, “stripped Treasury Bond principal”,

*Tel.: +1-403-492-3396; fax: +1-403-492-6826.
E-mail address: jacques.carriere@math.ualberta.ca (J.F. Carriere)
and “stripped Treasury Note principal” at various maturities. Let \( N \) denote the total number of distinct maturities in the published data at a particular time and let \( m_k > 0 \) for \( k = 1, 2, \ldots, N \) denote the observed maturities at that time. Observed maturities were always <30 years. Also, let \( B_k \) denote the average of the bid and asked prices for all strips with an observed maturity of \( m_k \), thus \( B_k \approx b_{m_k} \). It will be convenient to define \( B_0 = 1 \) and \( m_0 = 0 \).

Using \( B_k \), we can calculate the observed forward rates as follows:

\[
F_k = -\frac{\log_e[B_k] - \log_e[B_{k-1}]}{m_k - m_{k-1}}
\]  

(1.2)

for \( k = 1, 2, \ldots, N \). This observed rate corresponds to a maturity of

\[
m_k = \frac{m_k + m_{k-1}}{2}.
\]  

(1.3)

Fig. 1 is a plot of \( F_k \) versus \( m_k \) for data taken from October 11, 1990 issue of the WSJ. In this case, there were \( N = 121 \) observations. Theoretically, \( f_m > 0 \) but note that \( F_k < 0 \) at two maturities. Also note that the variance does not look constant as a function of maturity. All the graphs and calculations in this paper were done with GAUSS.
2. A spline model

In this section, we present a spline model of the forward rates. Let $0 = \kappa_0 < \kappa_1 < \cdots < \kappa_{n-1} < \kappa_n = 30$ denote the fixed knots of the spline and let $\kappa = (\kappa_1, \ldots, \kappa_{n-1})'$ denote the interior knots. Thus, the formula for a $q$-polynomial spline is

\[
\text{Spline}(m|q, \kappa) = \sum_{i=0}^{q} \phi_i m^i + \sum_{j=1}^{n-1} \xi_j (m - \kappa_j)^q 1(m > \kappa_j),
\]

where $1(e)$ is an indicator function that is equal to 1 if the event $e$ is true and 0 otherwise. This spline is a piecewise polynomial of degree $q$ with $q - 1$ continuous derivatives at the interior knots. The parameter $q = 1, 2, \ldots$ is fixed and it controls the smoothness. The parameter $n = 1, 2, \ldots$ controls the fit and it is also fixed. In this paper, we will let $q = 2$ and $n = 12$. The unknown parameters of this spline are $\phi_0, \ldots, \phi_q$ and $\xi_1, \ldots, \xi_{n-1}$. Consult Seber and Wild (1989) for more information about splines in regression analysis.

At this point, it will be convenient to introduce some matrix notation. Let $p = q + n$ denote the total number of parameters in the spline model and let $\beta$ denote a $p \times 1$ parameter vector that is defined as

\[
\beta = [\phi_0, \ldots, \phi_q, \xi_1, \ldots, \xi_{n-1}]'.
\]

Next, define

\[
x(m) = [1, m^1, \ldots, m^q, (m - \kappa_1)^q 1(m > \kappa_1), \ldots, (m - \kappa_{n-1})^q 1(m > \kappa_{n-1})]'.
\]

Thus, we can write $\text{Spline}(m|q, \kappa) = x'(m)\beta$. We are now ready to specify the statistical model for the observed data. Our model is

\[
F_k = x'(\tilde{m}_k)\beta + \varepsilon_k,
\]

where $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ are random variables that are not necessarily independent and identically distributed. Let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_N)'$, let $F = (F_1, \ldots, F_N)'$ and let

\[
X = \begin{bmatrix}
x'(\tilde{m}_1) \\
x'(\tilde{m}_2) \\
\vdots \\
x'(\tilde{m}_N)
\end{bmatrix}.
\]

These definitions allow us to write our linear model in matrix notation. Specifically,

\[
F = X\beta + \varepsilon.
\]

In this paper, we will estimate $\beta$ by a weighted least squares (WLS) method. Let $W$ denote an $N \times N$ diagonal matrix where all the off-diagonal elements are equal to zero. Using our matrix notation, we can write the WLS loss function as

\[
L(\beta) = (F - X\beta)' W (F - X\beta).
\]

Note that in this definition, $W$ is assumed to be fixed. Usually, we want to minimize $L(\beta)$ subjected to a linear constraint of $A\beta = \mu$. An example of a desirable linear constraint is $1'X\beta = 1'F$ where $1$ is a column vector of ones. In this case $A = 1'X$ and $\mu = 1'F$. The constrained WLS estimator is that value that minimizes $L(\beta)$ subject to the constraint. That value is equal to

\[
\hat{\beta} = (X'WX)^{-1} [X'WF + A'\lambda].
\]
where
\[
\lambda = [A(X'WX)^{-1}A']^{-1} [\mu - A(X'WX)^{-1}X'WF].
\] (2.9)

Using this estimator, we find that we can estimate the forward rates with \( \hat{f}_m = x'(m)\hat{\beta} \). We found a very good-fitting curve by using the constraint \( 1'X\beta = 1'F \) and a weighting matrix equal to the identity matrix. The fixed parameters were \( q = 2, n = 12 \) and \( \kappa_j = 30j/n \) for \( j = 1, 2, \ldots, n - 1 \). The residuals are defined as
\[
e_k = F_k - \hat{f}_{\tilde{m}_k}.
\] (2.10)

These residuals were unusual because they exhibited some heteroscedasticity and correlation.

3. Estimation of the variance structure

Traditionally, statisticians will assume that \( \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N \) are independent and identically distributed random variables. However, the residuals in our case do not exhibit that behaviour. We will assume that \( \text{Var}(e_k) = \sigma_k^2 = g(k) \). In this definition, \( g(k) \) is a map of \( [1, 2, \ldots, N] \) onto \( [1, 2, \ldots, n] \), that is defined as
\[
g(k) = \sum_{j=1}^{n} j 1(\kappa_{j-1} < \tilde{m}_k \leq \kappa_j).
\] (3.1)

Thus, the variance is \( \sigma_k^2 \) on the interval \( (\kappa_{j-1}, \kappa_j] \) and we get an estimable heteroscedastic error structure. It will be useful to define the standard deviation matrix \( \tilde{S} \). This is a diagonal matrix where the diagonal elements are equal to \( \sigma_k \) and all the off-diagonal elements are zero. Now let us describe how \( S \) can be estimated. The number of observations in the interval \( (\kappa_{j-1}, \kappa_j] \) is
\[
N_j = \sum_{k=1}^{N} 1(\kappa_{j-1} < \tilde{m}_k \leq \kappa_j).
\]

Let \( e_1, e_2, \ldots, e_N \) denote the residuals and let \( e_k \) be associated with the maturity of \( \tilde{m}_k \). A standard non-parametric estimator of \( \sigma_k^2 \) is
\[
\hat{\sigma}_k^2 = \frac{1}{N_j} \sum_{k=1}^{N} e_k^2 1(\kappa_{j-1} < \tilde{m}_k \leq \kappa_j) - \left[ \frac{1}{N_j} \sum_{k=1}^{N} e_k 1(\kappa_{j-1} < \tilde{m}_k \leq \kappa_j) \right]^2.
\] (3.2)

Using \( \hat{\sigma}_k \) we can construct an estimator of \( S \), denoted as \( \hat{S} \). In Section 5, we showed how a WLS estimator of \( \beta \) can be constructed for a fixed weighting matrix \( W \), without discussing what \( W \) should be. Ideally, we would let
\[
W = [S \ S]^{-1}.
\] (3.3)

However, \( S \) is not known. Thus, we recommend that \( \beta \) and \( S \) be estimated by iterated weighted least squares. This method will generate a sequence of estimators, denoted as \( \hat{S}_1, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \ldots \) and \( \hat{S}_2, \hat{S}_3, \ldots \). As a starting value, we suggest that \( \hat{S}_1 \) be set to the identity matrix. Let \( \hat{\beta}_1 \) be the estimator in this case. Using \( \hat{\beta}_1 \), we can calculate the residuals and estimate \( \hat{S}_2 \). In turn, \( \hat{\beta}_2 \) is found by letting \( W = [\hat{S}_2 \ \hat{S}_2]^{-1} \). The process is repeated until we have convergence in successive estimates of the parameters. Table 1 shows \( \hat{\sigma}_j \) for \( j = 1, 2, \ldots, n = 12 \). Note that the standard deviations have stabilized after five iterations. Also note that the standard deviations in the first iteration are very close to those in the last iteration.
Table 1
Estimates of the standard deviations

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>$\hat{\delta}_1$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_2$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_3$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_4$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_5$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_7$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_6$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_8$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_9$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_{10}$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_{11}$</td>
<td>1.00</td>
</tr>
<tr>
<td>$\hat{\delta}_{12}$</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Fig. 2. Spline estimates of the forward rates. This is a plot of $F_k$, $\hat{f}_{\delta_{10}}^1$, and $\hat{f}_{\delta_{10}}^4$ versus the maturity $m_k$. 
Let $\hat{f}_{m_k}^i = x'(\hat{m}_k^i)\hat{\theta}_i$ denote the estimated forward rates at iteration $i = 1, 2, \ldots$. We suggest that the iterations should stop whenever the average relative error (ARE) is less than 0.001, i.e.,

$$\text{ARE}_{i+1} = \frac{1}{N} \sum_{k=1}^{N} \left| \frac{\hat{f}_{m_k}^{i+1}}{\hat{f}_{m_k}^i} - 1 \right| < 0.001,$$

where $N = 121$. Note that the ARE of 0.001 could be smaller but the effect on the estimated values would be minimal. In our application we found that $\text{ARE}_2 = 0.0274$, $\text{ARE}_3 = 0.0047$, $\text{ARE}_4 = 0.0007$. Thus, we stopped after the fourth iteration. Fig. 2 is a plot of the observed rates $F_k$ against the estimated rates from the first and fourth iterations.

Next, let $\hat{\sigma}_{(k)}$ denote the estimated standard deviations and let $\hat{\Sigma}$ denote the associated matrix from the last iteration. Also, let $e = (e_1, e_2, \ldots, e_N)'$ denote the residuals from the last iteration. Our statistical model suggests that the standardized errors

$$\tilde{e} = \hat{\Sigma}^{-1} e$$

should not exhibit any heteroscedasticity. If $\tilde{e}_k$ is the $k$th element of $\tilde{e}$, then $\tilde{e}_k = e_k/\hat{\sigma}_{(k)}$. Fig. 3 is a plot of $\tilde{e}_k$ versus $\hat{m}_k$. Note the lack of heteroscedasticity.

---

**Fig. 3.** This is a plot of the standardized errors $\tilde{e}_k$ versus the maturity $\hat{m}_k$ for $k = 1, 2, \ldots, N = 121$. 
4. Estimation of the correlation structure

Traditionally, statisticians will assume that $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$ are uncorrelated random variables. As we will show, the residuals in our case do not exhibit that behaviour. Let $R = \{\rho_{|i-j|}, i, j = 1, \ldots, N\}$ denote a positive definite correlation matrix where the diagonal elements are equal to $\rho_0 = 1$ and the off-diagonal elements are in the interval $(-1, 1)$, i.e., $-1 < \rho_{|i-j|} < 1$. We assume that there exists a lag, denoted as $l$, such that $\rho_{|i-j|} = 0$ if $|i-j| > l + 1$. Thus, $R$ is a banded matrix with elements in the left-lower and right-upper corners equal to zero. In this definition, the correlation structure corresponds to a moving-average process with a lag of $l$. We will assume that

$$\text{Var}(\varepsilon) = SRS, \quad (4.1)$$

Next, define $\hat{\varepsilon} = S^{-1} \varepsilon$. Thus $\text{Var}(\hat{\varepsilon}) = R$. Next, let $R_{ch}$ be the lower triangular matrix from a Choleski decomposition of $R$, thus $R = R_{ch} R'_{ch}$. It is instructive to note that $\text{Var}(R_{ch}^{-1} \hat{\varepsilon})$, is simply equal to the identity matrix.

Now let us describe how $R$ can be estimated. Let $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_N)'$ denote the standardized residuals from our last iteration. Then, a standard non-parametric estimator of $\rho_u$ for $u \geq 1$ is

$$\hat{\rho}_u = \frac{\sum_{k=1}^{N-u} \hat{\varepsilon}_k \hat{\varepsilon}_{u+k} - (1/(N-u)) \left( \sum_{k=1}^{N-u} \hat{\varepsilon}_k \right) \left( \sum_{k=1}^{N-u} \hat{\varepsilon}_{u+k} \right)}{\sqrt{\left( \sum_{k=1}^{N-u} \hat{\varepsilon}_k^2 - (1/(N-u)) \left( \sum_{k=1}^{N-u} \hat{\varepsilon}_k \right)^2 \right) \left( \sum_{k=1}^{N-u} \hat{\varepsilon}_{u+k}^2 - (1/(N-u)) \left( \sum_{k=1}^{N-u} \hat{\varepsilon}_{u+k} \right)^2 \right)}}. \quad (4.2)$$

Under the null hypothesis, the test statistic

$$T = \sqrt{\frac{N-u-3}{4} \log_e \left[ \frac{1 + \hat{\rho}_u}{1 - \hat{\rho}_u} \right]} \quad (4.3)$$

is approximately normal with a mean of zero and a variance of one. With a 99% level of significance, we would reject the null hypothesis if $|T| > 2.57$. Table 2 is a summary of $\hat{\rho}_u$ and $T$ for $u = 1, 2, \ldots, 10$ and $N = 121$. Note that we cannot reject the null hypothesis of no correlation for lags of two or more. However, we can reject the hypothesis that $\rho_1 = 0$. Therefore, we will assume that $l = 1$ throughout the rest of the paper.

Using $\hat{\rho}_u$ we can construct an estimator of $R$, denoted as $\hat{R}$. A major weakness of the moment-type estimator $\hat{R}$ is that $\hat{R}$ is not positive definite in certain cases. Positive definiteness is a necessary property for doing a Choleski decomposition. Let us describe how to construct a positive definite estimator of $R$. A necessary and sufficient condition for $R$ to be a positive definite matrix is that the smallest eigenvalue of $R$ (denoted as $\lambda_1$) must be positive. Therefore, we suggest that $\hat{R}$ be equal to the matrix that minimizes

$$L(R) = \log_e [\det(R)] + \hat{\varepsilon}' R^{-1} \hat{\varepsilon}, \quad (4.4)$$

subject to the constraint that $\lambda_1 > 0$. To find $\hat{R}$ we need an iterative optimization program. We found that the moment-type estimator for $R$ is a good starting value for the iterations. Using this method we found that $\hat{\rho}_1 = -0.399$ when $l = 1$.

Next, let $R_{ch}$ be the lower triangular matrix from a Choleski decomposition of $\hat{R}$, thus $\hat{R} = R_{ch} \hat{R}_{ch}'$. Then, the standardized and uncorrelated residuals are denoted as $\hat{\varepsilon} = (\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_N)'$ and they are defined as

$$\hat{\varepsilon} = \hat{R}_{ch}^{-1} \hat{\varepsilon} = \hat{R}_{ch}^{-1} S^{-1} \varepsilon. \quad (4.5)$$

Table 2
Estimates of the correlations

<table>
<thead>
<tr>
<th>$u$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\rho}_u$</td>
<td>-0.360</td>
<td>0.043</td>
<td>-0.181</td>
<td>0.107</td>
<td>0.036</td>
<td>-0.190</td>
<td>-0.098</td>
<td>0.099</td>
<td>0.142</td>
<td>-0.067</td>
</tr>
<tr>
<td>$T$</td>
<td>-3.90</td>
<td>0.45</td>
<td>-1.90</td>
<td>1.11</td>
<td>0.37</td>
<td>-1.99</td>
<td>-1.02</td>
<td>1.03</td>
<td>1.48</td>
<td>-0.70</td>
</tr>
</tbody>
</table>
where \( e \) are the residuals from the last iteration. The residuals \( \hat{e} \) will be useful for constructing confidence intervals by the bootstrap method.

5. Confidence intervals by bootstrapping

Up to now, our focus has been on the estimation of the spline model without touching on the inference issues. In this section, we will show how to do inferences, like constructing confidence intervals, with the bootstrap method. For a detailed discussion of the bootstrap, the reader can consult the classical work of Efron (1982).

Let \( \hat{\beta} = R_{n1}^{1} S^{-1} e \) and let \( \hat{\beta}_k \) denote the \( k \)th element of \( \hat{\beta} \). We will assume that \( \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N \) are independent and identically distributed random variables with a common cumulative distribution function (CDF) of \( H(x) \). Usually, statisticians will assume that \( H(x) \) has a standard normal distribution. In our case, this assumption is not unreasonable. Using the standardized and uncorrelated residual errors of \( \hat{\beta}_k \) from the spline regression, we grouped the data and conducted a chi-squared goodness-of-fit test with 13 degrees of freedom. The test statistic was 14.62, therefore we cannot reject the null hypothesis that \( H(x) \) is a standard normal CDF at a 95% level of significance because 14.62 is <22.36, which is the 95th percentile of a chi-squared distribution with 13 degrees of freedom.

Fig. 4 is a plot of a standard normal density and a kernel density estimate based on the residual errors \( \hat{\beta}_k \). This density estimate used a normal kernel with an optimal smoothing parameter. Note how close the kernel

Fig. 4. The density of the the standardized and uncorrelated errors.
density estimate is to the normal density. See Simonoff (1996) for more information about kernel density estimates.

Next, let us construct an empirical CDF of $H(x)$ by using the standardized and uncorrelated residual errors of $\hat{e}_k$ from the spline regression. The empirical CDF is

$$\hat{H}(x) = \frac{1}{N} \sum_{k=1}^{N} 1(\hat{e}_k \leq x). \quad (5.1)$$

The traditional bootstrap method requires that we re-sample from $\hat{H}(x)$. However, an alternate method is too simply re-sample from a standard normal distribution. This alternate method is valid because we could not reject the null hypothesis of normality. Let us proceed in the traditional manner. Thus, $\hat{e}_{i,\text{boot}}$ for $k = 1, 2, \ldots, N$ and $l = 1, 2, \ldots, B$ are $N \times B$ independent and identically distributed random variables with a common CDF of $\hat{H}(x)$. The value $B$ is the number of bootstrap simulations and it should be made as large as possible. Let $\hat{e}_{i,\text{boot}} = (\hat{e}_{i,1,\text{boot}}, \ldots, \hat{e}_{i,N,\text{boot}})$. Using this simulated data, we calculate the heteroscedastic and correlated errors

$$e_{i,\text{boot}} = \hat{S}_r \hat{R}_{\text{ch}} e_{i,\text{boot}}. \quad (5.2)$$

Next, we calculate

$$F_{i,\text{boot}} = X \hat{\beta} + e_{i,\text{boot}}, \quad (5.3)$$

where $\hat{\beta}$ is the estimate from the last iteration.

Using the simulated data, we apply the same estimation formulas as before and derive the bootstrap estimates $\hat{\beta}_{i,\text{boot}}, \hat{S}_{i,\text{boot}},$ and $\hat{R}_{i,\text{boot}}$ for $l = 1, 2, \ldots, B$. Specifically, we use estimates from the fourth iteration of the iterated WLS, while the estimates of $S$ and $R$ are based on the residuals from the last iteration and the estimate of $R$ is found by minimizing Eq. (4.4). Next, let $g(\hat{\beta}, \hat{S}, \hat{R})$ be any real-valued function of the parameters $\hat{\beta}, \hat{S}$ and $\hat{R}$. Also, let $G(x)$ denote the CDF of the estimator $g(\hat{\beta}, \hat{S}, \hat{R})$. Then the bootstrap estimate of the CDF $G(x)$ is

$$\hat{G}(x) = \frac{1}{B} \sum_{l=1}^{B} \left[ g(\hat{\beta}_{l,\text{boot}}, \hat{S}_{l,\text{boot}}, \hat{R}_{l,\text{boot}}) \leq x \right]. \quad (5.4)$$

Using $\hat{G}(x)$, we can now approximate any distributional property of the estimator $g(\hat{\beta}, \hat{S}, \hat{R})$. For example, the approximate 97.5th percentile is a value $\xi_{0.975}$ such that

$$\xi_{0.975} = \inf \left\{ \xi : \hat{G}(\xi) \geq 0.975 \right\}. \quad (5.5)$$

As another example, we can approximate $E[g(\hat{\beta}, \hat{S}, \hat{R})]$ with

$$\int_{-\infty}^{\infty} x \, d\hat{G}(x) = \frac{1}{B} \sum_{l=1}^{B} g(\hat{\beta}_{l,\text{boot}}, \hat{S}_{l,\text{boot}}, \hat{R}_{l,\text{boot}}). \quad (5.6)$$

Applying the bootstrap method with $B = 2000$ yielded confidence intervals for all our parameters. First, a 95% confidence interval for the correlation $\rho_1$ was $[-0.3588, -0.4501]$ which includes our estimate of $\hat{\rho}_1 = -0.399$. Next, a 95% confidence interval of $\sigma_{12}$ was $[0.0182, 0.0762]$ and again our estimate of $\hat{\sigma}_{12} = 0.04664$ lies in this confidence interval. Next, the 95% confidence intervals for the instantaneous forward rates are shown in Fig. 5. Also shown is $F_{l,\text{boot}}$ for some $l$, which is in one of the bootstrap simulations.
Fig. 5. Confidence intervals for the forward rates.

References