A comparative study of principal component analysis on term structure of interest rates

Nien-Lin Liu

1 Department of Mathematical Sciences, Ritsumeikan University, 1-1-1 Noji Higashi, Kusatsu, Shiga 525-8577, Japan
E-mail gr012087@ed.ritsumei.ac.jp
Received January 10, 2010, Accepted March 31, 2010

Abstract

In this paper, principal component analysis (PCA) is applied to three different parametrization of interest rates: zero rates, yield curve, and forward rates. This comparative study is complementary to Akahori, Aoki, and Nagata [1] where they claimed that, under the no-arbitrage principle, yield curve cannot be a random walk. Conversely the forward curve could be a random walk. In our result of PCA, however, we observed that of the general beliefs. Our empirical results on the number of factors for the zero rates and the yield curve align with the general beliefs. This is a puzzle.

Keywords forward rates, principal component analysis, term structure of interest rates, zero rates

Research Activity Group Mathematical Finance

1. Introduction

With regard to the studies of the term structure of interest rates, a principal component analysis (PCA) is a relevant method. Fase [2] used the PCA to account for the variance among interest rates. Litterman and Scheinkman [3] used PCA to study the volatility of U.S. government bonds. In their empirical analyses, they found that there are three principal factors in the yield curve: the level, the steepness and the curvature. Bühler and Zimmermann [4], and Hiraki, Shiraishi and Takezawa [5] also indicated similar conclusions. Other than these, there have been plenty of studies applying PCA to interest rates. Here we referred only a few.

The previous analyses were performed, whether implicitly or not, on the basis of a random walk hypothesis (RWH) on the yield curve. The hypothesis is, however, fragile. There could be many alternative random walk hypotheses depending on the parametrizations since the yield curve is a stochastic process in an infinite dimensional space. For example, RWH’s on \( y_t(x) \) and \( r_t(T) := y(t, T - t) \) cannot be compatible (To make a distinction, we call the former the yield curve and the latter zero rate). Also, there could be RWH on forward rates \( \delta r_T \), etc. Then one is naturally led to ask: which parameterization is consistent with RWH?

No-arbitrage principle might be a one good criterion. It imposes a restriction on the drift. In Akahori, Aoki, and Nagata [1] (AAN model, hereafter), it is shown mathematically that the restriction is not consistent with RWH on yield curve; while the forward rate model they proposed is consistent with no-arbitrage condition.

The object of the present paper is to perform PCA on the real market data (the daily data of Japanese bonds and American bonds from 2007/6/20 to 2008/3/31 and from 2007/5/15 to 2008/3/31 respectively) in three different ways: to the increments of (i) spot rates (ii) yield curves, and (iii) forward rates on the basis of the AAN model. As a result, we obtain the following striking observation:

- The number of factors are two or three in the cases of (i) and (ii) while it is very large when the PCA is applied on the basis of AAN model in (iii).

Since we do not adopt any hypothesis testing style, we will not make any rigorous remark on the observation. However, the implication of the result could be the following: we need to construct a no-arbitrage model other than random walk model to explain the result.

In the rest of the paper, after giving the setting in Section 2, we present the results of our PCA in Section 3. Some concluding remarks will be made in Section 5.

2. Setting

2.1 Notations and definitions

We first explain the zero rates \( r_t(T) \), the yield curve \( y_t(x) \), and the forward rates \( F_t(T_1, T_2) \) in terms of the zero-coupon bond price \( P(t, T) \). The zero rates \( r_t(T) \), by which we mean the spot interest rate during \([t, T]\), is given by

\[
r_t(T) = -\frac{\log P(t, T)}{T - t}, \quad t \leq T,
\]

and the yield curve \( y_t(x) \) is given by

\[
y_t(x) = r_t(t + x), \quad x \geq 0.
\]

Here \( x \) represents the time to maturity. The forward rates are determined in terms of the zero rates as follows:

\[
F_t(T_1, T_2) := \frac{r_t(T_2)(T_2 - t) - r_t(T_1)(T_1 - t)}{T_2 - T_1}.
\]
Here the forward rate $F_i(T_1, T_2)$ represents the interest rate of the period $[T_1, T_2]$ pre-agreed at time $t$.

### 2.2 The data

We investigate the daily data of Japanese bonds and American bonds from 2007/6/20 to 2008/3/31 and from 2007/5/15 to 2008/3/31, respectively, which were obtained from Bloomberg. The zero rates $r(T)$ of Japanese bonds and American bonds are shown in Tables 1 and 2. The maturities of Japanese bonds are $T_1 = 2009/6/20, T_2 = 2009/12/20, T_3 = 2010/6/20, \ldots, T_{40} = 2028/12/20$, and the maturities of American bonds are $T_1 = 2009/5/15, T_2 = 2009/11/15, T_3 = 2010/5/15, \ldots, T_{40} = 2028/11/15$.

Since we use daily data, we set $\Delta t = 1/365$ year, and PCA’s are applied to $\Delta y(t) := y(t + \Delta t) - y(t)$, for $j = 1, 2, \ldots$ (precise description of $x_j$’s will be given later), $\Delta r_i(T) := r_i(t + \Delta t) - r_i(t)$ for $i = 1, 2, \ldots$, and $\Delta F_i(T_1, T_{i+1}) := F_i(t + \Delta t, T_{i+1}) - F_i(T_i, T_{i+1})$ for $i = 1, 2, \ldots$, respectively. Here $t$ runs through “2007/5/15 to 2008/3/31” and “2007/6/20 to 2008/8/31” but only the order is important.

### 3. The results of PCA’s

#### 3.1 Zero rates

The results of PCA applied to the increments of the zero rates in Tables 1 and 2 are given in Table 3. The results in Table 3 show that both Japanese and American cases need just two factors to reach the 95% level.

#### 3.2 Yield curves

To apply PCA for the yield curves $y(t)$, for $j = 1, 2, \ldots$, we need to interpolate the above data since we do not have the data of $r_j(t + x_j)$ for $t + x_j \notin [T_i]$. The formulas for our interpolations are the following: For $(t, x_j)$ with $t + x_j \notin T_i$, one can choose some $i$ such that $T_i < t + x_j < T_{i+1}$. We then set

$$y(t) = r_i(T_i) + \frac{t + x_j - T_i}{T_{i+1} - T_i}[r_i(T_{i+1}) - r_i(T_i)].$$

The results of PCA applied to the increments of the yield curves are presented in Table 4.

### 3.3 Forward rates

First, we calculate the forward curve for the period $(T_i, T_{i+1})$ at each time $t$ using formula (1). The results of the computation are shown in Tables 5 and 6.
4. Unit root tests

In modern time-series econometrics, the first question of interest is whether the time series under consideration is stationary or not. A Dickey-Fuller test, as constructed by Dickey and Fuller [6] (DF statistics, hereafter), is used to test the null hypothesis for a simple unit root. There are several different versions of the Dickey-Fuller test. The models we select are as follows:

(i) no intercept case:

\[ Y_t = \rho Y_{t-1} + \epsilon_t \]

(ii) intercept case:

\[ Y_t = \mu + \rho Y_{t-1} + \epsilon_t \]

for \( t = 1, 2, \ldots \), where \( Y_0 = 0 \), \( \rho \) is a real number, and \( \{ \epsilon_t \} \) is a sequence of independent normal random variables with mean zero and variance \( \sigma^2 \).

Under the null hypothesis that \( \rho = 1 \), the time series \( Y_t \) converges to a stationary time series as \( t \to \infty \), if \( |\rho| < 1 \). If \( |\rho| = 1 \), the time series is not stationary and prone to transform by differencing. The time series with \( \rho = 1 \) is sometimes called a random walk.

Table 10 shows DF statistics for each variable of the yield curves. Note that Symbols \(*\) and \(*\) denote rejections of the hypothesis of a unit root at the 0.05 and 0.01 significance levels, respectively.

Tables 11 and 12 show DF statistics for spot rates. According to the DF statistics, the null hypothesis cannot be rejected for the yield curves and the spot rates models. But it was shown in [1] that under the no-arbitrage principle, yield curve cannot be a random walk.

Next, Tables 13 and 14 show DF statistics for forward rates. It also cannot reject the null hypothesis for forward rates model around no intercept (zero mean). This is consistent with what we expect.

5. Conclusion

The empirical results we obtained in this paper shows that the number of factors of the forward rates are much greater than that of the general beliefs (two or three factors). Our empirical results on the number of factors for the zero rates and the yield curve align with the general beliefs.

The result implies that one cannot stick to both RWH and no arbitrage principle. In the modern finance, however, the latter is indispensable. Then, we should search for an alternative model other than random walk model to explain the observed reduction of the factors in PCA’s.

References

[5] T. Hiraki, N. Shiraiishi and N. Takezawa, Cointegration, com-
### Table 11. Unit root tests for Japanese spot rates.

<table>
<thead>
<tr>
<th>Maturities</th>
<th>intercept</th>
<th>no intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>no intercept</td>
<td>$T_1$</td>
<td>$T_2$</td>
</tr>
<tr>
<td>intercept</td>
<td>$-0.72039$</td>
<td>$-0.7703$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturities</th>
<th>intercept</th>
<th>no intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>no intercept</td>
<td>$T_{11}$</td>
<td>$T_{12}$</td>
</tr>
<tr>
<td>intercept</td>
<td>$-0.77766$</td>
<td>$-0.76561$</td>
</tr>
</tbody>
</table>

### Table 12. Unit root tests for American spot rates.

<table>
<thead>
<tr>
<th>Maturities</th>
<th>intercept</th>
<th>no intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>no intercept</td>
<td>$T_{31}$</td>
<td>$T_{22}$</td>
</tr>
<tr>
<td>intercept</td>
<td>$-0.37225$</td>
<td>$-0.35691$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturities</th>
<th>intercept</th>
<th>no intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>no intercept</td>
<td>$T_{31}$</td>
<td>$T_{32}$</td>
</tr>
<tr>
<td>intercept</td>
<td>$-0.19068$</td>
<td>$-0.17159$</td>
</tr>
</tbody>
</table>

### Table 13. Unit root tests for Japanese forward rates.

<table>
<thead>
<tr>
<th>Maturities</th>
<th>intercept</th>
<th>no intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>no intercept</td>
<td>$F(T_1, T_2)$</td>
<td>$F(T_2, T_3)$</td>
</tr>
<tr>
<td>intercept</td>
<td>$-1.00665$</td>
<td>$-0.97636$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Maturities</th>
<th>intercept</th>
<th>no intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>no intercept</td>
<td>$F(T_{11}, T_{12})$</td>
<td>$F(T_{12}, T_{13})$</td>
</tr>
<tr>
<td>intercept</td>
<td>$-0.72556$</td>
<td>$-0.59346$</td>
</tr>
</tbody>
</table>

### Table 14. Unit root tests for American forward rates.

<table>
<thead>
<tr>
<th>Maturities</th>
<th>intercept</th>
<th>no intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>no intercept</td>
<td>$F(T_{31}, T_{32})$</td>
<td>$F(T_{32}, T_{33})$</td>
</tr>
<tr>
<td>intercept</td>
<td>$0.242504$</td>
<td>$0.23077$</td>
</tr>
</tbody>
</table>

**Notes:**