

Theoretical and Empirical Analysis of Common Factors in a Term Structure Model

Theoretical and Empirical Analysis of Common
Factors in a Term Structure Model

By

Ting Ting Huang

**CAMBRIDGE
SCHOLARS**

P U B L I S H I N G

Theoretical and Empirical Analysis of Common Factors in a Term Structure Model,
By Ting Ting Huang

This book first published 2009

Cambridge Scholars Publishing

12 Back Chapman Street, Newcastle upon Tyne, NE6 2XX, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

Copyright © 2009 by Ting Ting Huang

All rights for this book reserved. No part of this book may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the copyright owner.

ISBN (10): 1-4438-1311-7, ISBN (13): 978-1-4438-1311-2

Theoretical and Empirical Analysis of Common Factors in a Term Structure Model

This paper can be put into a part of the series in Mathematical Finance. Mathematical Finance as a discipline across mathematics and Finance has been growing rapidly. From the past decade or so, there are plenty of researchers in academy and industry emerging to discuss the Term Structure Models. The growth and diversity manifests itself in the mix of mathematical analysis and econometrics methods.

Titles in the series will be scholarly and professional papers, intended to be read by a mixed audience of economists, mathematicians, operations research scientists, financial engineers, and other investment professionals.

Introductory Note

This paper demonstrates that the concepts and tools necessary for understanding and implementing models with the empirical copula can be more intuitive than those involved in the Black-Scholes pricing and diffusion models. It explores the interplay among financial economic theory, the availability of relevant data, and the choice of econometric methodology in the empirical study of common factors in term structure models. It has been prepared in my work as a PhD Candidate at the Department of Economics at the University of Pittsburgh in USA. It was accepted as a part of PhD dissertation thesis titled "Theoretical and Empirical Analysis of Common Factors in a Term Structure Model" at the Department of Economics of the University of Pittsburgh in October 2008.

I am grateful to my advisor Professor Xinfu Chen and Professor Steven Husted, not only for their guidance and advice throughout the completion of my thesis but also for giving me four wonderful years of teaching and researching at the University of Pittsburgh. I also want to thank to my other advisors, Professor James Cassing and Professor James Feigenbaum for joining the committee for my overview. Many thanks go to my dear husband Dr. Bruce Qiang Sun for fruitful and encouraging discussions. I am also grateful to my department colleagues for many valuable comments on my research and for sharing their knowledge in finance and economic theory. Finally, I appreciated the support and encouragement from my parents throughout the entire four years of working on this project.

Pittsburgh, August 2009

Ting Ting Huang

THEORETICAL AND EMPIRICAL ANALYSIS OF COMMON FACTORS IN A TERM STRUCTURE MODEL

TING TING HUANG

ABSTRACT. This paper studies dynamical and cross-sectional structures of bonds, typically used as risk-free assets in mathematical finance. After reviewing a mathematical theory on common factors, also known as principal components, we compute empirical common factors for 10 US government bonds from the daily data for the period 1993–2006 (data for earlier period is not complete) obtained from the official web site *www.treas.gov*. We find that the principal common factor contains 91% of total variance and the first two common-factors contain 99.4% of total variance. Regarding the first three common factors as stochastic processes, we find that the simple AR(1) models produce sample paths that look almost indistinguishable (in characteristic) from the empirical ones, although the AR(1) models do not seem to pass the normality based Portmanteau statistical test. Slightly more complicated ARMA(1,1) models pass the test. To see the independence of the first two common factors, we calculate the empirical copula (the joint distribution of transformed random variables by their marginal distribution functions) of the first two common-factors. Among many commonly used copulas (Gaussian, Frank, Clayton, FGM, Gumbel), the copula that corresponds to independent random variables is found to fit the best to our empirical copula. Loading coefficients (that of the linear combinations of common factors for various individual bonds) are briefly discussed. We conclude from our empirical analysis that yield-to-maturity curves of US government bonds from 1993 to 2006 can be simply modelled by two independent common factors which, in turn, can be modelled by ARMA(1,1) processes.

JEL classification: C32; G12; C13; C16; G10; G21.

Keywords: Term Structure, Common Factor, Principal Component, Copula, Yield-to-Maturity Curve

Date: September 5, 2009.

This work is partially supported by the National Science Foundation grant DMS-0504691.

CONTENTS

List of Tables	x
List of Figures	xi
1. Introduction	1
2. Common Factors of Random Variables	4
3. Common Factors of Stochastic Processes	10
3.1. The Setting	10
3.2. The Principal Components	11
3.3. The Numerical Procedure	11
3.4. Using Empirical Formula as the Theoretical One	13
4. Modelling the US Treasury Bonds	13
4.1. Data	13
4.2. The Common Factor Model	14
4.3. The Empirical Error in Using (4.1)	21
4.4. Modeling The Common Factors	21
4.5. Loading Coefficients	26
5. The Independency of the First Two Common Factors	30
5.1. Correlation	30
5.2. The Kendall's τ	30
5.3. Copula	31
5.4. The Empirical Copula	33
6. Conclusion	36
References	37

LIST OF TABLES

1	Historical Data for US Daily Treasury Yield Curve	17
2	Common Factor Analysis for US Treasury Bond Yield	20
3	Analysis of ε_t^j in (4.1) for US Treasury Bond	22
4	Moments of Common Factors and the Corresponding AR(1) Residuals	26

LIST OF FIGURES

1	Historical Plots for US Treasury Bond Rates	15
2	Surface of 10 Treasury Bond Yield Rates	16
3	The actual yields and the fitted curves	19
4	Empirical Common Factors	23
5	Sample paths of $\{X_t^1\}$ modelled by (4.3)	26
6	Sample paths of $\{X_t^2\}$ modelled by (4.4)	27
7	Sample paths of $\{X_t^3\}$ modelled by (4.5)	28
8	The Mean and the Loading Coefficients	29
9	The choice of optimal dependence parameters	34
10	The difference between the empirical Copula and the product copula	35

1. INTRODUCTION

Term structure models deal with dynamical and cross sectional behavior of bonds of various maturities. In many classical and modern theories of finance, bonds are regarded as risk free assets used for hedging and pricing financial security derivatives. Thus, term structure models are the foundations of many modern theories and have been attracted tremendous amount of attention in the last two decades.

Since the pioneer work of Vasicek [28] in 1977, there has been significant amount of progress towards term structure models; see, for instance, Cox, Ingersoll, and Ross [11] (1985), Ho and Lee [18] (1986), Black, Derman, and Troy [5] (1990), Heath, Jarrow and Morton [17](1992). In a typical term structure model, yields or prices of zero coupon bond are modelled by state variables which can be either observable or unobservable.

In 1996, Duffie and Kan [15] systematically studied a special class of term structures, the Affine Term Structure Model (ATSM); here the term “affine” mainly refers to the assumption that yields are liner combinations of state variables. An affine model gives a full description of the cross sectional and dynamical behavior of interest rates. At any point of time, the simple linear span (with deterministic loading coefficients) by the state variables determines the cross-section of interest rates. The dynamic properties of yields are inherited from the dynamics of the state process. For any set of maturities, the model guarantees the corresponding family of bond price processes satisfies the arbitrage-free conditions. Recently Junker, Szimayer, and Wagner [20] (2006) proposed a nonlinear cross-sectional dependence in the term structure of US treasury bonds by certain copula functions, and pointed out risk management implications.

Quite often the stable variables are modelled by common factors, also known as principal components. With the help of common factors,

one can, with as little loss of information as possible, reduce large-dimensional data to a limited number of factors. In this direction, Litterman and Scheinkman [22] in 1991 used a principal component analysis on a covariance matrix of three US treasury bonds of 1986 to estimate the common factors. Pagan, Hall, and Martin [23] (1995) used a set of stylized factors that pertain to the nature for the term structure modelling. Baum and Bekdache [4] (1996) also applied stylized facts as common factor to the dynamics of short-, medium-, and long-term interest rates, and explained the factors by incorporating asymmetric GARCH representations. Connor [10] (1995) and Campbell, Lo, and MacKinlay [8] (1997) summarized that there are three types of factor models available for examining the stochastic behavior of multiple assets and returns. The first one is the known-factor model which uses observable factors by linear regression to describe the common behavior of multiple returns. The second type is the fundamental factor model which use some micro-attributes of assets to construct common factors and explain assets returns. The third one is the statistical factor model which treats the factors as the latent variables or the unobservables which could be estimated from historical returns, and can capture the stochastic behavior of the multiple returns. The statistical factors, as discussed by Alexander [1] (2001) and Zivot and Wang [29] (2003), can be modelled by principal components which are linear combinations of returns; these factors accurately reflect the structure of the covariance of multivariate time series and the sources of variations of multiple asset returns. Cochrane [9] (2001) demonstrated that pricing kernels can be linear in the factors both in the economic time series and in the pricing models.

The aim of this paper is to perform empirical examination for a common factor model and report our new findings. Our work is similar to that of Piazzesi [24], but we focused on the empirical part, instead of the almost impossible data fitting about the cross sectional

behavior (loading coefficients) of actual yields to that derived from consistent (arbitrage-free) term structure models such as ATSM. We shall use principal components, which here we also call common factors, as state variables to construct a general affine term structure model. We compute empirical common factors for 10 US government bonds (the 3month, 6month, 1year, 2year, 3year, 5year, 7year, 10year, 20year, and 30year) from the daily data from 1993 to 2007, the only period from which we can obtain a complete official data set from the web site *www.treas.gov*. We find that the principal common factor contains 91% of total variance and the first two common-factors contains 99.4% of total variance. Regarding the first three common factors as stochastic processes, we find that the simple AR(1) models produce sample paths that look almost indistinguishable (in characteristic) from the empirical ones, although the AR(1) models do not seem to pass certain normality based statistical tests. A slightly more complicated ARMA(1,1) model passes the test. Also we verify the independency of the first two common factors by copulas. Among many commonly used copulas (Gaussian, Frank, Clayton) we found that the copula that corresponds to independent random variables fits the best to our empirical copula. Factors loading (the coefficients of the linear combinations of common factors for various individual bonds) are briefly discussed. We conclude from our empirical analysis that yield-to-maturity curve of US government bond from 1993 to 2006 can be simply modelled by two independent common factors, which in turn, can be modelled by ARMA(1,1) processes.

The rest of the paper is organized as follows. Section 2 is devoted to a review on theory of common factors of random variables. The theory is then applied in Section 3 to a set of several times series. In Section 4, we present an empirical formula that describes yields of US government bonds by a two-factor model, and demonstrate how the empirical common factors can be modelled by AR(1) and ARMA(1,1)

processes. In Section 5 we demonstrate our new discovery: the two empirical factors can be considered as independent in the sense that their joint distribution is very close to the product of the two marginal distributions. The conclusion is confirmed with the help of a theory of copulas. Section 6 concludes the paper.

2. COMMON FACTORS OF RANDOM VARIABLES

We begin with a review of the theoretical analysis on the common factors, also known as principal components [27].

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For an L^2 random variable f , we denote

$$\mathbb{E}[f] := \int_{\Omega} f(x) \mathbb{P}(dx), \quad \text{Var}[f] := \int_{\Omega} \left(f(x) - \mathbb{E}[f] \right)^2 \mathbb{P}(dx).$$

For L^2 random variables f and g , their covariance are denoted by

$$\text{Cov}[f, g] := \mathbb{E} \left[(f - \mathbb{E}[f])(g - \mathbb{E}[g]) \right] = \int_{\Omega} (f - \mathbb{E}[f])(g - \mathbb{E}[g]) \mathbb{P}(dx).$$

Also, for random variables g_1, \dots, g_k we denote

$$\text{span}\{g_1, \dots, g_k\} = \left\{ \sum_{i=1}^k c^i g_i \mid (c^1, \dots, c^k) \in \mathbb{R}^k \right\}.$$

Given random variables ξ, g_1, \dots, g_k , the **best linear indicator** of ξ by g_1, \dots, g_k is the projection of ξ on $\text{span}\{\mathbf{1}, g_1, \dots, g_k\}$ where $\mathbf{1}$ is the constant function: $\mathbf{1}(x) = 1 \ \forall x \in \Omega$. That is, the best linear indicator is a linear combination $c_0 \mathbf{1} + \sum_{i=1}^k c_i g_i$ such that

$$\left\| \xi - (c_0 + \sum_{i=1}^k c_i g_i) \right\|_{L^2}^2 = \min_{g \in \text{span}\{\mathbf{1}, g_1, \dots, g_k\}} \text{Var}[\xi, g].$$

We call $\varepsilon := \xi - (c_0 + \sum_{i=1}^k c_i g_i)$ the **remainder**. Roughly speaking, common factors of a given set of random variables are those spacial normalized random variables $\{g_1, \dots, g_k\}$ such that the sum of the variances of all remainders is the smallest possible. Mathematically, we formulate them as follows.

Definition 1. Let ξ_1, \dots, ξ_m be L^2 random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

- (1) A random variable f is called a **principal common factor** of ξ_1, \dots, ξ_m if

$$\mathbb{E}[f] = 0, \quad \text{Var}[f] = 1,$$

$$\min_{f_1, \dots, f_m \in \text{span}\{f\}} \sum_{j=1}^m \text{Var}[\xi_j - f_j] \leq \min_{g_1, \dots, g_m \in \text{span}\{g\}} \sum_{j=1}^m \text{Var}[\xi_j - g_j] \quad \forall g \in L^2.$$

- (2) An ordered set $\{f_1, \dots, f_k\}$ is called a set of **common factors** of ξ_1, \dots, ξ_m if for each $i, l = 1, \dots, k$,

$$\mathbb{E}[f_i] = 0, \quad \text{Cov}[f_i, f_l] = \delta_{il},$$

$$\min_{\eta_1, \dots, \eta_i \in \text{span}\{f_1, \dots, f_i\}} \sum_{j=1}^m \text{Var}[\xi_j - \eta_j] \leq \min_{\eta_1, \dots, \eta_i \in \text{span}\{g_1, \dots, g_i\}} \sum_{j=1}^m \text{Var}[\xi_j - \eta_j] \quad \forall g_1, \dots, g_i \in L^2.$$

Note that if $\{f_1, \dots, f_k\}$ is a set common factors of ξ_1, \dots, ξ_m ($k \leq m$), then f_1 is a principal common factor. In the sequel, we shall derive an algorithm that computes common factors. First we study the best linear indicator.

Lemma 2.1. (i) For any L^2 random variables ξ, g_1, \dots, g_i ,

$$\min_{\eta \in \text{span}\{g_1, \dots, g_i\}} \text{Var}[\xi - \eta] = \min_{b^1, \dots, b^i \in \mathbb{R}} \text{Var}[\xi - (b^1 g_1 + \dots + b^i g_i)].$$

(ii) Suppose $\{f_1, \dots, f_k\}$ is a set of random variables so normalized that $\mathbb{E}[f_i] = 0, \mathbb{E}[f_i, f_l] = \delta_{il}$ for $i, l = 1, \dots, k$. Then for every random variables ξ ,

$$\min_{\eta \in \text{span}\{f_1, \dots, f_i\}} \text{Var}[\xi - \eta] = \text{Var}[\xi - (\beta^1 f_1 + \dots + \beta^i f_i)] \quad \forall i = 1, \dots, k$$

where

$$\beta^i = \text{Cov}[\xi, f_i] \quad \forall i = 1, \dots, k.$$

Proof. The first assertion (i) follows from the definition of $\text{span}\{g_1, \dots, g_k\}$. The second assertion follows from the fact that $\{\mathbf{1}, f_1, \dots, f_k\}$ is an orthonormal set of L^2 and the best linear indicator is the orthogonal projection of ξ onto $\text{span}\{\mathbf{1}, f_1, \dots, f_k\}$. \square

Note that if $\{f_1, \dots, f_k\}$ is a set of common factors, then we can write

$$\xi_j = \beta_j^0 \mathbf{1} + \beta_j^1 f_1 + \dots + \beta_j^k f_k + \varepsilon_j \quad \forall j = 1, \dots, m$$

where $\beta_j^0 = \mathbb{E}[f_j]$, $\beta_j^i = \text{Cov}[\xi_j, f_i]$ ($i = 1, \dots, k$) and ε_j is a random variable that is not correlated to any of f_1, \dots, f_k : $\text{Cov}[\varepsilon_j, f_i] = 0$ for $i = 1, \dots, k$, $j = 1, \dots, m$.

To find common factors, we recall a well-known result from linear algebra.

Lemma 2.2. *Assume that \mathbf{A} is a semi-positive definite non-trivial matrix and let λ be the maximum eigenvalue of \mathbf{A} . Then*

$$\max_{\mathbf{w} \in \mathbb{R}^n} \frac{\mathbf{w} \mathbf{A} \mathbf{w}^T}{\mathbf{w} \mathbf{w}^T} = \lambda.$$

In addition, the maximum is obtained at and only at eigenvectors of \mathbf{A} associated with λ .

The following theorem characterizes principal common factors.

Theorem 1. *Assume that ξ_1, \dots, ξ_m are random variables, not all of them are constants. Then a random variable is a principal common factor if and only if it is a linear combination of $\{\xi_j - \mathbb{E}[\xi_j]\}_{j=1}^m$ with a weight being an eigenvector of the covariance matrix $\mathbf{A} := (\text{Cov}[\xi_i, \xi_j])_{m \times m}$ associated the maximum eigenvalue; more precisely, f is a principal common factor of ξ_1, \dots, ξ_m if and only if*

$$(2.1) \quad f(x) = \sum_{j=1}^m \frac{e^j}{\sqrt{\lambda}} \left(\xi_j(x) - \mathbb{E}[\xi_j] \right) \quad \forall x \in \Omega,$$

where λ is the maximum eigenvalue of \mathbf{A} and $\mathbf{e} = (e^1, \dots, e^m)$ satisfies $\mathbf{e} \mathbf{A} = \lambda \mathbf{e}$, $|\mathbf{e}|^2 = 1$.

Moreover, if f given by (2.1) is a principal common factor of ξ_1, \dots, ξ_m , then

$$\begin{aligned} \text{inf}_{\eta \in \text{span}\{f\}} \text{Var}[\xi_j - \eta] &= \text{Var}[\xi_j - \beta_j f], \quad \beta_j = \sqrt{\lambda} e^j \quad \forall j = 1, \dots, m, \\ \min_{\eta \in \text{span}\{f\}} \text{Var}[\xi_j - \eta] &= \sum_{j=1}^m \left\{ \text{Var}[\xi_j] - \beta_j^2 \right\} = \sum_{j=1}^m \text{Var}[\xi_j] - \lambda. \end{aligned}$$

Proof. Let λ be the maximum eigenvalue of the covariance matrix $\mathbf{A} = (\text{Cov}[\xi_i, \xi_j])_{m \times m}$.

First, we show that

$$\min_{g \in L^2(\Omega)} \sum_{j=1}^m \min_{\eta \in \text{span}\{g\}} \text{Var}[\xi_j - \eta] = \sum_{j=1}^m \text{Var}[\xi_j] - \lambda.$$

For this, let g be an arbitrary non-constant random variable. Then,

$$\begin{aligned} \sum_{j=1}^m \min_{\eta \in \text{span}\{g\}} \text{Var}[\xi_j - \eta] &= \sum_{j=1}^m \min_{b \in \mathbb{R}} \text{Var}[\xi_j - bg] \\ &= \sum_{j=1}^m \text{Var}[\xi_j - b_j g]^2 \Big|_{b_j = \frac{\text{Cov}[\xi_j, g]}{\text{Var}[g]}} \\ &= \sum_{j=1}^m \text{Var}[\xi_j] - \sum_{j=1}^m \frac{\text{Cov}^2[\xi_j, g]}{\text{Var}^2[g]}. \end{aligned}$$

Now decompose g by

$$g = \sum_{l=1}^m w^l (\xi_l - \mathbb{E}[\xi_l]) + \zeta, \quad \zeta \perp \xi_j - \mathbb{E}[\xi_j] \quad \forall j = 1, \dots, m.$$

Then $\text{Cov}[\xi_j, g] = \sum_{l=1}^m \text{Cov}[\xi_l, \xi_j] w^l$, so that

$$\begin{aligned} \frac{\sum_{j=1}^m \text{Cov}^2[\xi_j, g]}{\text{Var}^2[g]} &= \frac{\sum_{j=1}^m \left(\sum_{l=1}^m w^l \text{Cov}[\xi_l, \xi_j] \right) \left(\sum_{s=1}^m \text{Cov}[\xi_j, \xi_s] w^s \right)}{\sum_{l=1}^m \sum_{s=1}^m w^l \text{Cov}[\xi_l, \xi_s] w^s + \|\zeta\|_{L^2(\Omega)}^2} \\ &= \frac{\mathbf{w} \mathbf{A} \mathbf{w}^T}{\mathbf{w} \mathbf{A} \mathbf{w}^T + \|\zeta\|_{L^2(\Omega)}^2} \leq \lambda. \end{aligned}$$

Here the equal sign holds if and only if $\zeta = \mathbf{0}$ and \mathbf{w} is an eigenvector, associated with λ , of \mathbf{A} . Hence, (1.4) holds.

Now suppose f is a principal common factor. Then $f = \sum_{j=1}^m w^j(\xi_j - \mathbb{E}[\xi_j]) + \zeta$ where $\zeta \equiv 0$ and $\mathbf{w} = (w^1, \dots, w^m)$ is an eigenvector of \mathbf{A} . Using $\text{Var}[f] = 1$ one can derive that $\lambda|\mathbf{w}|^2 = 1$. Hence, the vector $\mathbf{e} = \sqrt{\lambda} \mathbf{w} = (e^1, \dots, e^m)$ is a unit eigenvector of \mathbf{A} associated with λ and $f = \sum_{j=1}^m e^j(\xi_j - \mathbb{E}[\xi_j])/\sqrt{\lambda}$.

Now let $\mathbf{e} = (e^1, \dots, e^m)$ be an arbitrary unit eigenvector of \mathbf{A} associated with the eigenvalue λ . Set $f = \sum_{j=1}^m e^j(\xi_j - \mathbb{E}[\xi_j])/\sqrt{\lambda}$. We show that f is a principal common factor. First, we can calculate

$$\begin{aligned} \mathbb{E}[f] &= \sum_{j=1}^m \frac{e^j}{\sqrt{\lambda}} \mathbb{E}[\xi_j - \mathbb{E}[\xi_j]] = 0, \\ \text{Var}[f] &= \sum_{s,j=1}^m \frac{e^s e^j}{\lambda} \text{Cov}[\xi_s, \xi_j] = \frac{\mathbf{e} \mathbf{A} \mathbf{e}^T}{\lambda} = |\mathbf{e}|^2 = 1. \end{aligned}$$

Also, for each j ,

$$\min_{\eta \in \text{span}\{f\}} \text{Var}[\xi_j - \eta] = \text{Var}[\xi_j - \beta_j f] = \text{Var}[\xi_j] - \beta_j^2$$

where

$$\beta_j = \text{Cov}[\xi_j, f] = \sum_{s=1}^m \frac{e^s}{\sqrt{\lambda}} \text{Cov}[\xi_s, \xi_j] = \sqrt{\lambda} e^j$$

since $\mathbf{e} \mathbf{A} = \lambda \mathbf{e}$. Thus, $(\beta_1, \dots, \beta_m) = \sqrt{\lambda} \mathbf{e}$. It then follows that

$$\begin{aligned} \sum_{j=1}^m \min_{\eta_j \in \text{span}\{f\}} \text{Var}[\xi_j - \eta_j] &= \sum_{j=1}^m \text{Var}[\xi_j] - \sum_{j=1}^m \beta_j^2 \\ &= \sum_{j=1}^m \text{Var}[\xi_j] - \lambda = \min_{g \in L^2} \sum_{j=1}^m \min_{\eta_j \in \text{span}\{g\}} \text{Var}[\xi_j - \eta_j]. \end{aligned}$$

Thus, by definition, f is a principal common factor. This completes the proof. \square

Theorem 2. Let K be the dimension of the space $\text{span}\{\xi_1 - \mathbb{E}[\xi_1], \dots, \xi_m - \mathbb{E}[\xi_m]\}$ and $\{\lambda_i\}_{i=1}^m$ be the complete set of eigenvalues of the covariance

matrix $\mathbf{A} = (\text{Cov}[\xi_i, \xi_j])_{m \times m}$, arranged in the order $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_m \geq 0$.

Then for each $k \in \{1, \dots, K\}$, a set $\{f_1, \dots, f_k\}$ of random variables is a set of common factors of ξ_1, \dots, ξ_m if and only if there exist vectors $\mathbf{e}_1, \dots, \mathbf{e}_k$ in \mathbb{R}^m , $\mathbf{e}_i = (e_i^1, \dots, e_i^m)$, such that

$$\mathbf{e}_i \mathbf{A} = \lambda_i \mathbf{e}_i, \quad \mathbf{e}_i \cdot \mathbf{e}_l = \delta_{il} \quad \forall i, l = 1, \dots, k,$$

$$f_i = \sum_{j=1}^m \frac{e_i^j}{\sqrt{\lambda_i}} (\xi_j - \mathbb{E}[\xi_j]) \quad \forall i = 1, \dots, k.$$

Moreover, if $\{f_1, \dots, f_k\}$ is a set of common factors, then for each $i = 1, \dots, k$,

$$\sum_{j=1}^m \min_{\eta \in \text{span}\{f_1, \dots, f_i\}} \text{Var}[\xi_j - \eta] = \sum_{j=1}^m \text{Var}[\xi_j - (\beta_j^1 f_1 + \dots + \beta_j^i f_i)] = \sum_{j=i+1}^m \lambda_j,$$

where

$$(\beta_1^i, \dots, \beta_m^i) = \sqrt{\lambda_i} \mathbf{e}_i \quad \forall i = 1, \dots, K.$$

In particular,

$$(\xi_1, \dots, \xi_m) = (\mathbb{E}[\xi_1], \dots, \mathbb{E}[\xi_m]) + \sqrt{\lambda_1} \mathbf{e}_1 f_1 + \sqrt{\lambda_2} \mathbf{e}_2 f_2 + \dots + \sqrt{\lambda_K} \mathbf{e}_K f_K.$$

The proof is analogous to the $k = 1$ case and is omitted.

Note that $\xi_i \in \text{span}\{f_1, \dots, f_K\}$ for each $i = 1, \dots, m$ so that

$$\begin{aligned} \xi_i &= \beta_j^0 + \sum_{i=1}^K \beta_j^i f_i, \quad (\beta_j^0 := \mathbb{E}[\xi_j]) \quad \forall j = 1, \dots, m, \\ \sum_{j=1}^m \text{Var}[\xi_j] &= \sum_{j=1}^m \sum_{i=1}^K \beta_j^i{}^2 = \sum_{i=1}^K \lambda_i = \sum_{j=1}^m \lambda_j. \end{aligned}$$

Hence, if we use a set $\{f_1, \dots, f_k\}$ of k common factors to describe all random variables $\{\xi_j\}_{j=1}^m$, the percentage of the total variances of remainders over total variances is

$$(2.4) \quad \frac{\sum_{j=1}^m \text{Var}[\xi_j - \sum_{i=0}^k \beta_j^i f_i]}{\sum_{j=1}^m \text{Var}[\xi_j]} = 1 - \frac{\sum_{i=k+1}^m \lambda_i}{\sum_{i=1}^m \lambda_i}.$$

3. COMMON FACTORS OF STOCHASTIC PROCESSES

In this section we describe our way of specifying the cross sectional behavior of a term structure model by common factors. Let \mathbf{T} be a set of time moments (trading time) and $\{Y_t^1\}_{t \in \mathbf{T}}, \dots, \{Y_t^m\}_{t \in \mathbf{T}}$ be stochastic processes believed to be strongly correlated. An example in our mind is the case where Y_t^j is the yield at time t of the zero-coupon bond with fixed maturity of τ_j , i.e., the yield of the zero-coupon bond bought at time t and to be matured at time $t + \tau_j$. In a generic affine term structure model [15] with k factors, these yields are described by

$$(3.1) \quad Y_t^j = \beta_j^0 + \beta_j^1 X_t^1 + \dots + \beta_j^k X_t^k + \varepsilon_t^j \quad \forall t \in \mathbf{T}, \quad j = 1, \dots, m.$$

Here $\beta_j^0, \beta_j^1, \dots, \beta_j^k$ are constants, and $\{X_t^1\}_{t \in \mathbf{T}}, \dots, \{X_t^k\}_{t \in \mathbf{T}}$ are stochastic processes. In the terminology of finance, $\{Y_t^j\}$ are called observable variables and $\{X_t^i\}$ are called state variables, typically modelled as latent or unobservable variables. The term $\{\varepsilon_t^j\}$ are called individual (non-system) errors. Our purpose is to model $\{X_t^i\}$ by common factors of $\{Y_t^j\}$.

If we regard $\mathbf{Y}_t = (Y_t^1, \dots, Y_t^m)$ as a vector random variable, then since historical data are correlated, a standard sample covariance matrix may not accurately reflect the true covariance matrix, and therefore, complicated procedures are needed; see, for example, Tsay [27] and Stoffer [25]. Here, we shall pay attention on the cross sectional behavior of the yields and focus on the fitting of historical data by (3.1). We shall use the sample covariance matrix in our calculation and also provide our new point of view in performing such a calculation.

3.1. The Setting. Let $\mathbf{T}_h = \{t_i\}_{i=1}^n$ be historical trading dates. We use the following probability space:

$$(3.2) \quad \Omega = \mathbf{T}, \quad \mathcal{F} = 2^\Omega, \quad \mathbb{P}(\{t\}) = \frac{1}{|\mathbf{T}_h|} \quad \forall t \in \mathbf{T}_h.$$

Here 2^Ω is the collection of all subsets of Ω and $|\mathbf{T}_h|$ is the number of dates in the set \mathbf{T}_h .

For each $t \in \mathbf{T}_h$, let y_t^j be the historical data of the yield of type j bond at time t . We define $Y^j : \Omega \rightarrow \mathbb{R}$ by

$$(3.3) \quad Y^j(t) = y_t^j \quad \forall t \in \mathbf{T}_h.$$

Then each $Y^j, j = 1, \dots, m$, is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

Under the above setting, we see that

$$(3.4) \quad \begin{aligned} \mu^j &:= \mathbb{E}[Y^j] = \int_{\Omega} Y^j(t) \mathbb{P}(dt) = \frac{1}{|\mathbf{T}_h|} \sum_{t \in \mathbf{T}_h} y_t^j, \\ \sigma^{ij} &:= \text{Cov}[Y^i, Y^j] = \int_{\Omega} [Y^i(t) - \mu^i][Y^j(t) - \mu^j] \mathbb{P}(dt) \\ (3.5) \quad &= \frac{1}{|\mathbf{T}_h|} \sum_{t \in \mathbf{T}_h} (y_t^i - \mu^i)(y_t^j - \mu^j). \end{aligned}$$

3.2. The Principal Components. According to the theory of principal component presented in the previous section, we seek random variables X^1, \dots, X^m on $(\Omega, \mathcal{F}, \mathbb{P})$ such that the following holds:

- (1) $\mathbf{1}, X^1, \dots, X^m$ form an orthonormal set in $L^2(\Omega)$;
- (2) For each $k = 1, \dots, m$, when we decompose Y^j as

$$Y^j = \hat{Y}^j + \varepsilon^j, \quad \hat{Y}^j := \mu^j \mathbf{1} + \sum_{i=1}^k \beta_i^j X^i + \varepsilon^j \quad \text{in } \Omega, \quad \beta_i^j := (Y^j, X^i)_{L^2(\Omega)},$$

we have

$$\sum_{j=1}^m \|\varepsilon^j\|_{L^2(\Omega)}^2 = \min_{\dim\{\mathbf{1}, \tilde{Y}^1, \dots, \tilde{Y}^m\} = k+1} \sum_{j=1}^m \|Y^j - \tilde{Y}^j\|_{L^2(\Omega)}^2.$$

Here $\dim\{\mathbf{1}, \tilde{Y}^1, \dots, \tilde{Y}^m\}$ denotes the linear dimension of the space $\text{span}\{\mathbf{1}, \tilde{f}, \dots, \tilde{Y}^m\}$.

3.3. The Numerical Procedure. According to Theorem 1, we can find the principal components as follows:

1. Let $\mathbf{A} = (\sigma^{ij})_{m \times m}$ where σ^{ij} is as (3.4). Find a complete eigen set $\{\lambda_i, \mathbf{e}_i\}_{i=1}^m$ of \mathbf{A} :

$$\lambda^1 \geq \lambda^2 \geq \dots \geq \lambda^m, \quad \mathbf{e}^i \mathbf{A} = \lambda^i \mathbf{e}^i, \quad \mathbf{e}^i \cdot \mathbf{e}^j = \delta_{ij} \quad i, j = 1, \dots, m.$$

2. For each i , write $\mathbf{e}_i = (e_i^1, \dots, e_i^m)$. Set

$$\begin{aligned} (3.6) \quad \mathbf{X}^i(t) &= \frac{1}{\sqrt{\lambda^i}} \sum_{j=1}^m e_i^j (Y^j(t) - \mu^j) \quad \forall t \in \mathbf{T}_h, \quad i = 1, 2, \dots, m, \\ \mathbf{e}_0 &= (\beta_0^1, \dots, \beta_0^m) := (\mu^1, \dots, \mu^m), \\ (3.7) \quad \beta_i^j &= \sqrt{\lambda^i} e_i^j, \quad i, j = 1, \dots, m. \end{aligned}$$

We then obtain

$$(3.8) \quad (Y^1, \dots, Y^m) = \mathbf{e}_0 + \sqrt{\lambda_1} \mathbf{e}_1 X^1 + \sqrt{\lambda_2} \mathbf{e}_2 X^2 + \dots + \sqrt{\lambda_m} \mathbf{e}_m X^m \quad \text{on } \Omega.$$

3. Now as far as historical data are concerned, we can write

$$(3.9) \quad (Y^1, \dots, Y^m) = \mathbf{e}_0 + \sqrt{\lambda_1} \mathbf{e}_1 X^1 + \dots + \sqrt{\lambda_k} \mathbf{e}_k X^k + (\varepsilon^1, \dots, \varepsilon^m) \quad \text{on } \mathbf{T}_h.$$

where

$$(\varepsilon^1, \dots, \varepsilon^m) = \sum_{i=k+1}^m \sqrt{\lambda_i} \mathbf{e}_i X^i.$$

4. Under the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ defined as in (3.3), we have, since $\mathbb{E}[X^i] = 0$,

$$\text{Cov}[X^j, X^l] := \int_{\Omega} X^i(t) X^j(t) \mathbb{P}(dt) = \frac{1}{|\mathbf{T}_h|} \sum_{t \in \mathbf{T}_h} X^i X^j = \delta_{ij}.$$

Also, we find the total variances of the “negligible” term $\varepsilon := (\varepsilon^1, \dots, \varepsilon^m)$ to be

$$(3.10) \quad \sum_{i=1}^m \text{Var}[\varepsilon^i] = \sum_{j=1}^m \frac{1}{|\mathbf{T}_h|} \sum_{t \in \mathbf{T}_h} (\varepsilon(t)^j)^2 = \sum_{j=k+1}^m \lambda_j.$$

The relative variance contributed by all the “negligible” terms is

$$(3.11) \quad \delta_k = \frac{\sum_{j=1}^m \text{Var}[\varepsilon_t^j]}{\sum_{i=1}^m \text{Var}[Y_t^i]} = \frac{\sum_{i=1}^m \frac{1}{|\mathbf{T}_h|} \sum_{t \in \mathbf{T}_h} (\varepsilon^i(t))^2}{\sum_{i=1}^m \frac{1}{|\mathbf{T}_h|} \sum_{t \in \mathbf{T}_h} (Y^i(t) - \mu^i)^2} = \frac{\sum_{i=k+1}^m \lambda_i}{\sum_{i=1}^m \lambda_i} \times 100\%.$$

3.4. Using Empirical Formula as the Theoretical One. Once we obtained the empirical formula (3.9), we can regard it as an assumption of a term structure model. More precisely, write $Y^j(t)$, $X^i(t)$, $\varepsilon^j(t)$ as Y_t^j , X_t^i , and ε_t^j respectively, we can assume in our model that

$$(3.12) \quad Y_t^j = \mu^j + \sum_{i=1}^k \beta_i^j X_t^i + \varepsilon_t^j \quad \forall j = 1, \dots, m, t \in \mathbf{T}.$$

In this expression, we regard each $\{Y_t^j\}_{t \in \mathbf{T}}$, $\{X_t^i\}$, $\{\varepsilon_t^j\}$ as a stochastic process.

4. MODELLING THE US TREASURY BONDS

Based on the above theoretical framework, in this section we use historical data of US Government bonds of various maturities to model its term structure by common factors. We focus on the following questions:

- (1) What is the “optimal” number of factors that balances simplicity and accuracy ?
- (2) Regarding the common factors as state variables, can we model them by simple stochastic processes ?
- (3) Regarding $\{\beta_i^j\}$ as loading coefficients, how much can we say about them ?

4.1. Data. On the official web site www.trea.gov, various information on US government debt is published by the requirement of the law. The Daily Treasury Yield Curve Rates is downloaded from www.treas.gov/offices/domestic-finance/debt-management/interest-rate/yield.shtml. Maximizing the number of bonds whose historical data are available, we found a complete set of data from 10/1/1993 to 12/29/2006 (the date that this research starts with) for 10 different (zero-coupon) bonds with maturity 3month, 6month, 1year, 2year, 3year, 5year, 7year, 10year, 20year, and 30year, respectively. We denote these times to maturity by $\tau_j, j = 1, \dots, m := 10$ and the yields of the bonds by Y_t^j .

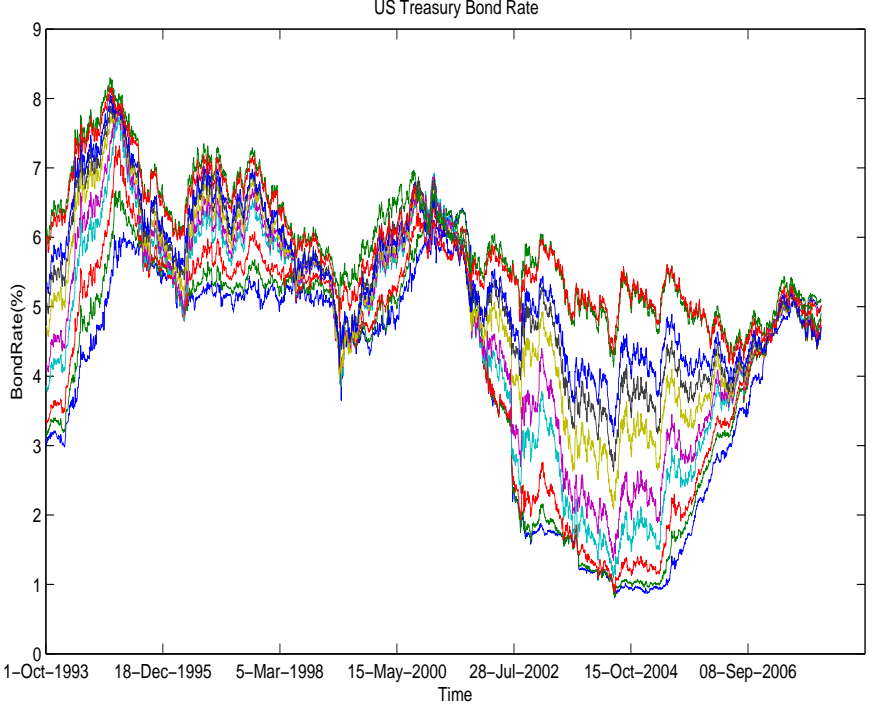
We pick the time window 10/1/1993–12/29/2006 since data on certain bonds are missing for earlier periods and at the current stage we do not want to use any theoretical interpolations to enter our empirical study.

The dynamical (in time) behavior of the 10 yields of different maturities is plotted in Figure 1. The surface describing the dynamical and cross sectional (in time-to-maturity) behavior of the 10 yields are plotted in Figure 2. The yields on every first trading day of October are listed in Table 1, along with their statistics.

From Figure 1, we see the time trends of the bond yields. More precisely, in Figure 2, we can see that the short-term rates decreases as time approach to 2003, partly due to the depression at that time, and bounces back quickly in 2005. The long-term rates follow more or less the same fashion but in a smaller magnitude. In Table 1, the yields are reported in percentages. The statistics are calculated from a total of 3313 observations. From the statistics, we can see that yields do not seem to be normally distributed. In particular, the kurtosis of all yields are all well below 3 (the kurtosis of normal distribution). Since we have 3133 data, these statistics is significant enough to exclude the hypothesis that yields are normally distributed or have fat tails (i.e. kurtosis ≥ 3). Also, benchmark normal distributions are symmetric around the mean, so that the skewness is 0. The distribution of short-term and medium-term yields show negative skewness, which means the distribution of yields is skewed to the left, so, intuitively, the distribution has a long left tail. But The distribution of long-term yields show positive skewness, which means the distribution of yields is skewed to the right, so, intuitively, the distribution has a long right tail. The smaller than normal kurtosis shows that its tails are thinner compared to the normal distribution.

4.2. The Common Factor Model. Now we let $\mathbf{T} = \{1, 2, \dots, 3313\}$ be trading times and $\{Y_t^1\}_{t \in \mathbf{T}}, \dots, \{Y_t^m\}_{t \in \mathbf{T}}$ be daily yield rate where

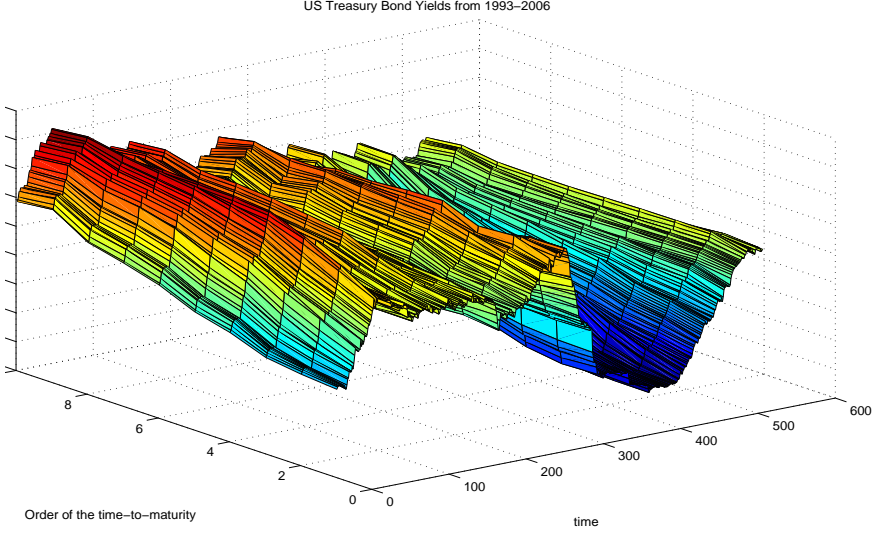
FIGURE 1. Historical Plots for US Treasury Bond Rates



Each curve represents $y = y_t^j$ where y_t^j is the yield at time t of the bond with maturity τ_j .

$m=10$. Thus, each column in the matrix $\{\mathbf{Y}_t\}_{t=1}^{3313} = \{(Y_t^1, Y_t^2, \dots, Y_t^{10})\}_{t=1}^{3313}$ corresponds to the yield of 3 month, 6 month, 1 year, 2 year, 3 year, 5 year, 7 year, 10 year, 20 year, and 30 year bond, respectively. The sample mean vector, $\{\beta_0^j\}_{j=1}^m$, is shown in the row with heading "Mean" in Table 1.

FIGURE 2. Surface of 10 Treasury Bond Yield Rates



The surface $y = y_t^j$, $1 \leq j \leq 10, 0 \leq t \leq 3313$ of the yields of 10 US Government Bonds on 3313 trading days from 1993 to 2006. Units of time in figure is week.

According to the formulas (3.6)–(3.8), we obtain the following term structure model:

$$(4.1) \quad \begin{pmatrix} Y_t^{3m} \\ Y_t^{6m} \\ Y_t^{1y} \\ Y_t^{2y} \\ Y_t^{3y} \\ Y_t^{5y} \\ Y_t^{7y} \\ Y_t^{10y} \\ Y_t^{20y} \\ Y_t^{30y} \end{pmatrix} = \begin{pmatrix} 3.96 \\ 4.13 \\ 4.28 \\ 4.60 \\ 4.78 \\ 5.07 \\ 5.30 \\ 5.42 \\ 5.92 \\ 5.81 \end{pmatrix} + 4.15 \begin{pmatrix} 0.38 \\ 0.39 \\ 0.40 \\ 0.38 \\ 0.36 \\ 0.30 \\ 0.26 \\ 0.23 \\ 0.18 \\ 0.16 \end{pmatrix} X_t^{1+1.25} \begin{pmatrix} -0.40 \\ -0.36 \\ -0.23 \\ -0.05 \\ 0.07 \\ 0.23 \\ 0.30 \\ 0.36 \\ 0.43 \\ 0.45 \end{pmatrix} X_t^2 + \varepsilon_t.$$

TABLE 1. Historical Data for US Daily Treasury Yield Curve

Date	3mo	6mo	1yr	2yr	3yr	5yr	7yr	10yr	20yr	30yr
93	2.98	3.11	3.35	3.84	4.18	4.72	5.03	5.34	6.12	5.98
.
.
.
94	5.05	5.61	6.06	6.69	7.01	7.35	7.52	7.66	8.02	7.86
.
.
.
95	5.53	5.64	5.65	5.82	5.89	5.98	6.10	6.15	6.61	6.48
.
.
.
96	5.10	5.35	5.65	6.03	6.22	6.39	6.54	6.65	6.99	6.88
.
.
.
97	5.10	5.27	5.44	5.75	5.83	5.93	6.05	6.04	6.38	6.33
.
.
.
98	4.23	4.36	4.28	4.17	4.10	4.10	4.26	4.33	5.09	4.90
.
.
.
99	5.16	5.32	5.47	5.83	5.93	6.00	6.23	6.06	6.55	6.19
.
.
.
00	6.27	6.33	6.06	5.98	5.92	5.86	5.95	5.83	6.18	5.93
.
.
.
01	2.37	2.37	2.47	2.82	3.18	3.90	4.33	4.55	5.39	5.38
.
.
.
02	1.59	1.54	1.56	1.80	2.11	2.75	3.34	3.72	4.81	4.93
.
.
.
03	0.95	1.00	1.13	1.47	1.93	2.84	3.40	3.96	4.92	5.00
.
.
.
04	1.71	2.00	2.21	2.63	2.92	3.44	3.85	4.21	4.95	5.06
.
.
.
05	3.61	4.02	4.09	4.21	4.23	4.25	4.31	4.39	4.67	4.58
.
.
.
06	5.02	5.09	5.00	4.82	4.74	4.70	4.70	4.71	4.91	4.81
.
.
.
Mean	3.96	4.13	4.28	4.60	4.78	5.07	5.30	5.42	5.92	5.81
Variance	1.67	1.70	1.67	1.60	1.48	1.28	1.16	1.05	0.92	0.88
Skewness	-0.62	-0.63	-0.59	-0.50	-0.39	-0.12	0.03	0.28	0.37	0.49
Kurtosis	1.95	1.99	2.04	2.17	2.19	2.17	2.10	2.24	2.34	2.52

We pick the first trading date of October for each year to show this historical daily data Last four rows are statistics of daily data of the Treasury (zero-coupon) bond yields, with round-up 2 digits for space limit.

Here we have replaced j in Y_t^j by the true time to maturity for easy reading. Here the column vector coefficients of X_t^1 and X_t^2 are unit

vectors (sum of the square of entries equal to 1), being the eigenvectors of the sample covariance matrix of $\{Y_t^j\}_{j=1}^{10}$. The scalar multiples are the square root of eigenvalues. The empirical mean and variance of each X_t^i are zero and one respectively.

The accuracy of the two factor model can be seen from Figure 3. In each plot in Figure 3, the dots are the actual yield and the curve is the fitted yield-to-maturity curve obtained by right-hand side of (4.1) with the ε_t term dropped. The first two plots are the best and worst fit respectively; the time t of the rest 14 plots is randomly picked from our historical date set. Here the vertical axis has unit of percentage ¹, and the horizontal axis is the index j of the maturity τ_j . One can see that for historical data, the two term factor model, represented by the curve, fits the actual data, represented by the dots, very well.

The accuracy of the model can also be seen from Table 2 where the proportion of total variance explained by the i th factor and the cumulative proportion of total variance explained by the first i th factors are calculated by formula (3.11) and displayed. From Table 2, we can see that the principal common factor explains about 91 percent of the total variance of the US treasury bonds from 1993 to 2006. Combined with the second common factor, the linear regression could explain more than 99.4 percent of the total variance of all the bonds. Thus, using the common factor model (4.1), the remainder term ε_t^i contains only about 0.6% of the total variance.

¹Notice that the scale is different for each plot, for the yield rates are changing over time.