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

by

Ting Ting Huang

M.A., Economics, 2003

Submitted to the Graduate Faculty of

the Department of Mathematics in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

MATHEMATICS DEPARTMENT

This thesis was presented

by

Ting Ting Huang

It was defended on

April 16th 2008

and approved by

Xinfu Chen, Dept. Mathematics

James Feigenbaum, Dept. Economics

Dehua Wang, Dept. Mathematics

David Stoffer, Dept. Statistics

Thesis Advisor: Xinfu Chen, Dept. Mathematics

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

Ting Ting Huang, M.S.

University of Pittsburgh, 2008

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 (3month, 6month, 1year, 2year, 3year, 5year, 7year, 10year, 20year, and 30year) 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.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 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 [27] in 1977, there has been significant amount of progress towards term structure models; see, for instance, Cox, Ingersoll, and Ross [10] (1985), Ho and Lee [17] (1986), Black, Derman, and Troy [5] (1990), Heath, Jarrow and Morton [16](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 [14] 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 [19] (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 [21] 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 [22] (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 [9] (1995) and Campbell, Lo, and MacKinlay [7] (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 [28] (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 [8] (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 [23], 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.

2.0 COMMON FACTORS OF RANDOM VARIABLES

We begin with a review of the theoretical analysis on the common factors, also known as principal components [26].

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

$$E[f] := \int_\Omega f(x)\, P(dx), \qquad \mathrm{Var}[f] := \int_\Omega \Big(f(x) - E[f]\Big)^2 P(dx).$$

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

$$\mathrm{Cov}[f,g] \ := \ E\Big[(f - E[f])(g - E[g])\Big] = \int_\Omega (f - E[f])(g - E[g]) P(dx).$$

Also, for random variables $g_1, \cdots, g_k$ we denote

$$\mathrm{span}\{g_1, \cdots, g_k\} = \Big\{ \sum_{i=1}^k c^i g_i \mid (c^1, \cdots, c^k) \in R^k \Big\}.$$

Given random variables $\xi, g_1, \cdots, g_k$, the **best linear indicator** of ξ by $g_1, \cdots, g_k$ is the projection of ξ on $\mathrm{span}\{\mathbf{1}, g_1, \cdots, g_k\}$ 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

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

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 $\{g_1, \cdots, g_k\}$ such that the sum of the variances of all remainders is the smallest possible. Mathematically, we formulate them as follows.

Let $\xi_1, \cdots, \xi_m$ be L^2 random variables on a probability space $(\Omega, \mathcal{F}, P)$.

1. A random variable f is called a **principal common factor** of $\xi_1, \cdots, \xi_m$ if

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

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

2. An ordered set $\{f_1, \cdots, f_k\}$ is called a set of **common factors** of $\xi_1, \cdots, \xi_m$ if for each $i, l = 1, \cdots, k$,

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

$$\min_{\eta_1, \cdots, \eta_i \in \mathrm{span}\{f_1, \cdots, f_i\}} \sum_{j=1}^{m} \mathrm{Var}\Big[\xi_j - \eta_j\Big] \qquad \min_{\eta_1, \cdots, \eta_i \in \mathrm{span}\{g_1, \cdots, g_i\}} \sum_{j=1}^{m} \mathrm{Var}\Big[\xi_j - \eta_j\Big]$$
$$\forall\, g_1, \cdots, g_i \in L^2.$$

Note that if $\{f_1, \cdots, f_k\}$ is a set common factors of $\xi_1, \cdots, \xi_m$ (km), 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.

(i) For any L^2 random variables $\xi, g_1, \cdots, g_i$,

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

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

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

where

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

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

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

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

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

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

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

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

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

The following theorem characterizes principal common factors.

Assume that $\xi_1, \cdots, \xi_m$ 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 $\{\xi_j - E[\xi_j]\}_{j=1}^m$ with a weight being an eigenvector of the covariance matrix $\mathbf{A} := (\mathrm{Cov}[\xi_i, \xi_j])_{m\times m}$ associated the maximum eigenvalue; more precisely, f is a principal common factor of $\xi_1, \cdots, \xi_m$ if and only if

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

where λ is the maximum eigenvalue of $\mathbf{A}$ and $\mathbf{e} = (e^1, \cdots, e^m)$ 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 $\xi_1, \cdots, \xi_m$, then

$$\min_{\eta \in \mathrm{span}\{f\}} \mathrm{Var}[\xi_j - \eta] = \mathrm{Var}[\xi_j - \beta_j f], \quad \beta_j = \sqrt{\lambda} e^j \qquad \forall\, j = 1, \cdots, m, \tag{2.2}$$

$$\sum_{j=1}^m \min_{\eta \in \mathrm{span}\{f\}} \mathrm{Var}[\xi_j - \eta] = \sum_{j=1}^m \Big\{ \mathrm{Var}[\xi_j] - \beta_j^2 \Big\} = \sum_{j=1}^m \mathrm{Var}[\xi_j] - \lambda. \tag{2.3}$$

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

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,

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

Now decompose g by

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

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

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

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

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

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

Also, for each j,

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

where

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

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

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

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

Let K be the dimension of the space $\mathrm{span}\{\xi_1 - E[\xi_1], \cdots, \xi_m - E[\xi_m]\}$ and $\{\lambda_i\}_{i=1}^m$ be the complete set of eigenvalues of the covariance matrix $\mathbf{A} = (\mathrm{Cov}[\xi_i, \xi_j])_{m \times m}$, arranged in the order $\lambda_1 \lambda_2 \cdots \lambda_m 0$.

Then for each $k \in \{1, \cdots, K\}$, a set $\{f_1, \cdots, f_k\}$ of random variables is a set of common factors of $\xi_1, \cdots, \xi_m$ if and only if there exist vectors $\mathbf{e}_1, \cdots, \mathbf{e}_k$ in R^m, $\mathbf{e}_i = (e_i^1, \cdots, e_i^m)$, such that

$$\begin{aligned} \mathbf{e}_i\ \mathbf{A} = \lambda_i \mathbf{e}_i, \quad \mathbf{e}_i \cdot \mathbf{e}_l = \delta_{il} \quad \forall\, i, l = 1, \cdots, k, \\ f_i = \sum_{j=1}^m \frac{e_i^j}{\sqrt{\lambda_i}} \Big(\xi_j - E[\xi_j] \Big) \quad \forall\, i = 1, \cdots, k. \end{aligned}$$

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

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

where

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

In particular,

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

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

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

$$\begin{aligned} \xi_i &= \beta_j^0 + \sum_{i=1}^{K} \beta_j^i f_i, \quad (\beta_j^0 := E[\xi_j]) \quad \forall\, j = 1, \cdots, m, \\ \sum_{j=1}^{m} \mathrm{Var}[\xi_j] &= \sum_{j=1}^{m}\sum_{i=1}^{K} \beta_j^{i2} = \sum_{i=1}^{K} \lambda_i = \sum_{j=1}^{m} \lambda_j. \end{aligned}$$

Hence, if we use a set $\{f_1, \cdots, f_k\}$ 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

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

3.0 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 $\{Y_t^1\}_{t\in\mathbf{T}}, \cdots, \{Y_t^m\}_{t\in\mathbf{T}}$ 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 [14] with k factors, these yields are described by

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

Here $\beta_j^0, \beta_j^1, \cdots, \beta_j^k$ are constants, and $\{X_t^1\}_{t\in\mathbf{T}}, \cdots, \{X_t^k\}_{t\in\mathbf{T}}$ 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 $\mathbf{Y}_t = (Y_t^1, \cdots, Y_t^m)$ 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 [26] and Stoffer [24]. 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.

3.0.1 The Setting

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

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

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 $Y^j : \Omega \to R$ by

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

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

Under the above setting, we see that

$$\mu^j \;:=\; E[Y^j] = \int_\Omega Y^j(t)P(dt) = \frac{1}{|\mathbf{T}_h|}\sum_{t\in\mathbf{T}_h} y_t^j, \tag{3.4}$$

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

3.0.2 The Principal Components

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

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

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

we have

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

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

3.0.3 The Numerical Procedure

According to Theorem 2, 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 $\mathbf{A}$:

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

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

$$X^i(t) = \frac{1}{\sqrt{\lambda^i}} \sum_{j=1}^m e_i^j (Y^j(t) - \mu^j) \quad \forall\, t \in \mathbf{T}_h, \qquad i = 1, 2, \cdots, m, \tag{3.6}$$

$$\mathbf{e}_0 = (\beta_0^1, \cdots, \beta_0^m) := (\mu^1, \cdots, \mu^m),$$

$$\beta_i^j = \sqrt{\lambda^i} e_i^j, \quad i, j = 1, \cdots, m. \tag{3.7}$$

We then obtain

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

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

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

where

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

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

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

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

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

The relative variance contributed by all the "negligible" terms is

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

3.0.4 Using Empirical Formula as the Theoretical One

Once we obtained the empirical formula (3.9), we can regarded 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

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

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

4.0 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:

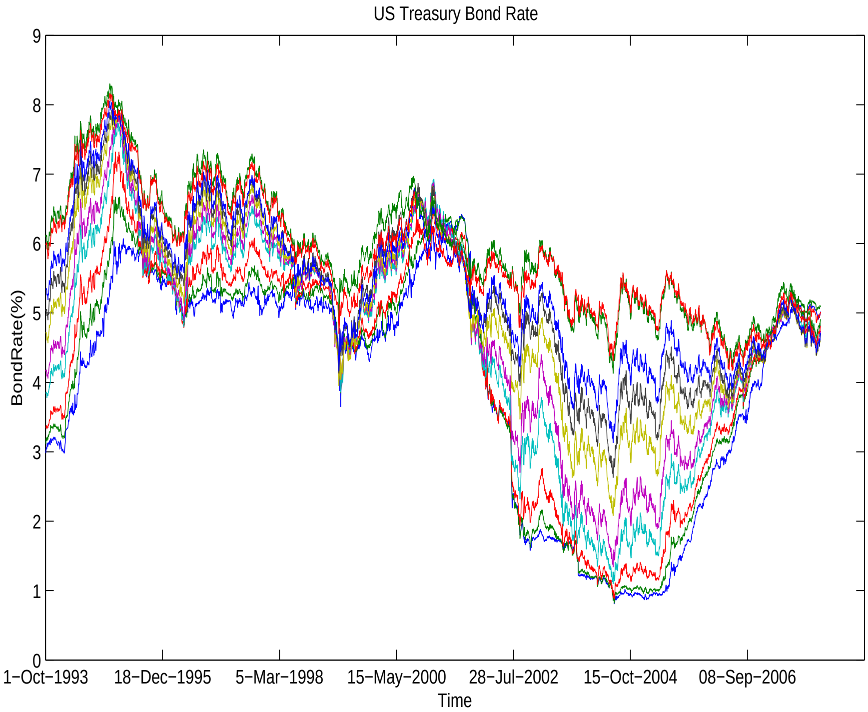
1. What is the "optimal" number of factors that balances simplicity and accuracy ?
2. Regarding the common factors as state variables, can we model them by simple stochastic processes ?
3. Regarding $\{\beta_i^j\}$ as loading coefficients, how much can we say about them ?

4.0.5 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 $\tau_j, j = 1, \cdots, m := 10$ 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.

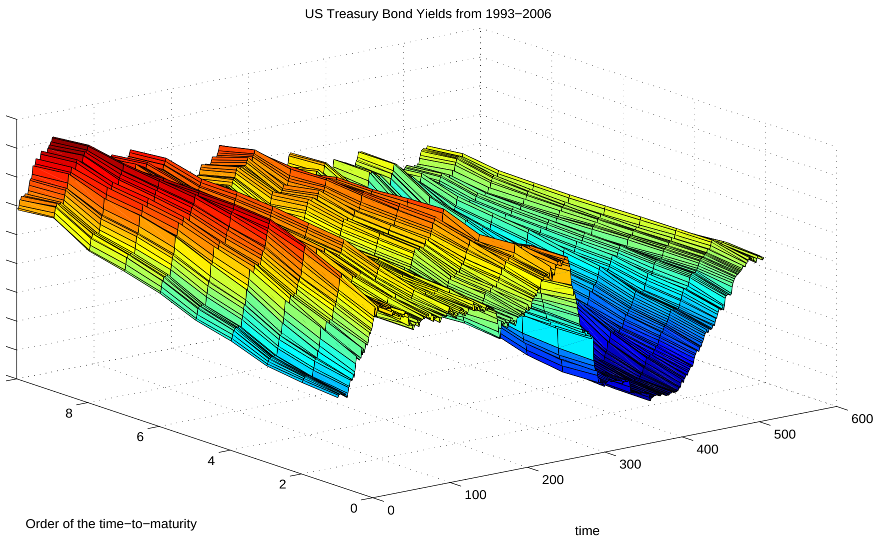
Figure 1: Historical Plots for US Treasury Bond Rates During 10/1/1993–12/29/2006



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

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

Figure 2: Surface of 10 Treasury Bond Yield Rates



The surface $y = y_t^j$, $1j10, 0t3313$ of the yields of 10 US Government Bonds on 3313 trading days from 1993 to 2006. Units of time in figure is week.

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.

4.0.6 The Common Factor Model

Now we let $\mathbf{T}$ $=\{1, 2, ..., 3313\}$ be trading times and $\{Y_t^1\}_{t\in\mathbf{T}}, \cdots, \{Y_t^m\}_{t\in\mathbf{T}}$ be daily yield rate where m=10. Thus, each column in the matrix $\{\mathbf{Y}_t\}_{t=1}^{3313} = \{(Y_t^1, Y_t^2, ..., Y_t^{10})\}_{t=1}^{3313}$ 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.

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

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

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 ith factor and the cumulative proportion of total variance explained by the first ith 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

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

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.

It is quite clear from Table 2 that it is reasonable to use two common factors to balance simplicity and accuracy. The two factor model (4.1) explains quite well the performance of all the US government bonds from 1993 to 2006. We point out that the original data contains only two decimal points, so except the first three common factors, all the rest factors do not have much actual meaning. Thus, a maximum of only three common factors could be used for any model that based on the empirical data that we gathered.

4.0.7 The Empirical Error in Using (4.1)

In Table 3, we examine the empirical error terms ε_t^j in the formula (4.1). By our derivation, the sample mean of each ε_t^j is zero. For all ten bonds, the maximum size of the error term is about 0.5 (percent). The variances of the error terms are very small, ranging from 0.004 to 0.02. The most important quantities, listed in the last column, are the proportion of the variance of the error with respect to the variance of data. We can see that the variances of error terms range from 0.17 to 2.68 percent of the variances of the corresponding original data. The minimum 0.17% is attained at the 6 month bond and the maximum 2.68% is attained at the 30-year bond. Although the residuals coming from the cross-sectional estimation capture little explanation for the data, the relative large 0.95% residual of 3-month bond and the 2.68% residual of the 30-year bond may be explained by the relative independency of short-term and long term bond rates.

In conclusion, the factor model (4.1) with or without the ε_t term is a sound model for US government bond for the period 1993–2006.

4.0.8 Modeling The Common Factors

Now we investigate the common factor, or "state variable", $\{X_t^1\}$ and $\{X_t^2\}$. Empirical data are obtained from linear combinations of $Y_t^i - \mu^j \mathbf{1}$ by (3.6) and plotted in Figure 4. We regard $\{X_t^1\}_{t\in T}$ and $\{X_t^2\}_{t\in T}$ as stochastic process and would like to see if we can use simple models to describe them.

4.0.8.1 The AR(1) Model An AR(1) model for a stochastic process $\{x_t\}_{t=-\infty}^{\infty}$ is described by the auto-regression

$$x_t = \phi x_{t-1} + \sigma\, u_t \qquad \forall\, t = 0, \pm 1, \pm 2, \cdots \tag{4.2}$$

where $\phi \in (-1,1)$ and $\sigma 0$ are constants estimated and $\cdots, u_{-1}, u_0, u_1, \cdots$ are i.i.d N(0,1) distributed random variables. Using maximum likelihood for our empirical data displayed in Figure 4, we find that

$$X_t^1 = 0.9992\, X_{t-1}^1 + 0.0348\, u_t^1\ , \tag{4.3}$$

$$X_t^2 = 0.9980\, X_{t-1}^2 + 0.0665\, u_t^2\ , \tag{4.4}$$

$$X_t^3 = 0.9917\, X_{t-1}^3 + 0.1269\, u_t^3\ . \tag{4.5}$$

Note that all the coefficients $0.9992, 0.9980, 0.9917$ are very close to 1, so X_{t-1}^i and X_t^i are strongly correlated. This is expected since out unit of time is one trading day.

When $\Delta t = 1$ refers to as one day, we can rewrite (4.2) as

$$x_t - x_{t-\Delta t} = -\frac{1-\phi}{\Delta t}\, x_{t-\Delta t}\Delta t + \frac{\sigma}{\sqrt{\Delta t}}\sqrt{\Delta t}\, u_t\ .$$

This equation can be considered as a time discretization of the stochastic differential equation:

$$dx_t = -k x_t\, dt + \nu\, dB_t$$

where $\{B_t\}$ is the standard Brownian motion process and

$$k = \frac{1-\phi}{\Delta t}, \qquad \nu = \frac{\sigma}{\sqrt{\Delta t}}.$$

Using annual units, $\Delta t (= 1(\text{day})) = 1/250$ (year), we see that (4.3)–(4.5) can be considered as the time discretization of the stochastic differential equations, in annual units,

$$dX_t^1 = -0.20\ X_t^1\, dt + 0.55\ dB_t^1$$

$$dX_t^2 = -0.50\ X_t^2\, dt + 1.05\ dB_t^2$$

$$dX_t^3 = -2.08\ X_t^3\, dt + 2.01\ dB_t^3$$

where $\{B_t^1, B_t^2, B_t^3\}$ are standard Brownian motions. Here coefficients 0.55, 1.05, 2.01 are sizes of innovations related directly to the different sizes of "local wiggles" of the three paths in Figure 4.

The validity of (4.3)–(4.5) relies on the independency of $\{u_t\}$. For this we calculate

$$Q(d) := T\sum_{l=1}^{d} \rho_l^2,$$

where $1d < T-1$ and

$$\rho_l := \frac{\sum_{t=l+1}^{T}(u_t - \overline{u_t})(u_{t-l} - \overline{u_{t-l}})}{\sum_{t=1}^{T}(u_t - \overline{u_t})^2}, \qquad \overline{u_{t-l}} := \frac{1}{T-l}\sum_{t=l+1}^{T} u_{t-l}.$$

Taking a standard $d = \ln(T) = 8$ we find, for the first three common factors, that $Q(8) = 17.4,\ 25.4,$ and 43, respectively. Under the assumption that $\{u_t\}$ are i.i.d and normally distributed, these large Q values correspond to tail probabilities of 2.6%, 0.13% and 0.00% respectively, from which we conclude that the AR(1) processes could not pass the Portmanteau test if we assume that $\{u_t\}$ are i.i.d and normally distributed.

4.0.8.2 The ARMA(1,1) Model Next we use ARMA(1,1) to model the common factors. The equation for an ARMA(1,1) process reads

$$x_t = \phi x_{t-1} + \sigma\,(u_t + \theta u_{t-1})$$

where $\phi \in (-1,1)$, $\sigma > 0, \theta \in R$ are constants, $\cdots, u_{-1}, u_0, u_1, \cdots$ are i.i.d. N(0,1) distributed random variables. Using maximum likelihood and a standard package (Stata) we find that

$$\begin{aligned}
X_t^1 &= 0.9991\,X_{t-1}^1 + 0.0347\,(u_t^1 + 0.0572\,u_{t-1}^1),\\
X_t^2 &= 0.9978\,X_{t-1}^2 + 0.0664\,(u_t^2 + 0.0579\,u_{t-1}^2),\\
X_t^3 &= 0.9910\,X_{t-1}^3 + 0.1268\,(u_t^3 + 0.0386\,u_{t-1}^3).
\end{aligned}$$

Here again, the validity of the model relied on the independency of $\cdots, u_{-1}, u_0, u_1, \cdots$. Applying the Q statistics for the Portmanteau test described in the previous subsubsection, we find that for the first two common factors, $Q(8) = 8.7$ and $Q(8) = 15.7$ which corresponds to tail probability of 37% and 5% respectively. This indicate a clear pass for the principal factor and a marginal pass for the second factor. For the third common factor, its $Q(8)$ value is 37, indicating a clear fail of the test.

4.0.8.3 Simulations Here we simulate the first three common factors by the model equation (4.3)–(4.5) respectively where $\{u_t^i\}$ are assumed to be i.i.d N(0,1) distributed. For each equation, we simulated 15 sample paths. Together with the empirical one, the total 16 paths are plotted in Figures 5, 6, and 7, respectively. From these figures, we can see that AR(1) models (4.3)–(4.5) are indeed excellent models in the sense that they produce sample paths that are almost indistinguishable in characteristics from the empirical ones.

4.0.8.4 Normality Issue In a generic AR(1) model (4.2), the innovation term $\{u_t\}$ are only required to be independent. In standard packages normal distributions are used. Here we find that the empirical innovation u_t^i defined in (4.3)–(4.5) are not normally distributed; see the empirical evidence in Table 4. For example, the empirical set $\{u_t^1\}_{t=2}^{3313}$ has a kurtosis 7.5, which is above the acceptance level of normality. Hence, the Q test discussed earlier may not be used as a rejection of AR(1) model for the first two factors. We feel pretty content from Figures 5–7 that either AR(1) or ARMA(1,1) are legitimate models for the first two common factors.

4.0.9 Loading Coefficients

The vector $(\beta_1^i, \cdots, \beta_m^i)$ is commonly referred to as **loadings** of the common factor X_t^i. In Figure 8, we plot the loading coefficients for $i = 0, 1, 2, 3$ ($i = 0$ corresponding to the mean). Here the index j has been replaced by

$$\log\left(1 + \frac{\tau}{\text{month}}\right)$$

where τ is time to maturity.

The loadings of the first common factor are all positive. This means that changes in the principal common factor produce an overall uplift or an overall down-push of the yield-to-maturity curve. This common factor is therefore referred in the literature as the **level**

factor. The loadings of the second common factor is negative for short term bonds and positive for long term bonds; they monotonically increase as time to maturity increases. Changes in the second common factor thus rotate the yield curve. In the literature, the second factor is referred to as a **slope factor**. The curve of loading coefficients of the third common factor is convex in the $\log(1+\tau/\text{month})$ scale, negative in the middle and positive on both ends. The third common factor (if we include it in our model) therefore affects the curvature of the yield curve, and therefore it is called the **curvature factor** [23].

In an affine term structure model, the yields are assumed to be linear combinations of state variables (in our case common factors). If the model is complete (arbitrage free), the loadings are uniquely determined by specifications of the dynamical behavior of state variables, typically by the Itó lemma based stochastic differential equations. As mentioned by Piazzessi [23], currently studied affine term structure models do not seem to provide satisfactory loadings that match empirical ones. We shall leave this study in our future research.

Table 1: Historical Data for US Daily Treasury Yield Curve Rates (10/01/1993-12/26/2006)

Date	3mo	6mo	1yr	2yr	3yr	5yr	7yr	10yr	20yr	30yr
10/1/1993	2.98	3.11	3.35	3.84	4.18	4.72	5.03	5.34	6.12	5.98
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/3/1994	5.05	5.61	6.06	6.69	7.01	7.35	7.52	7.66	8.02	7.86
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/2/1995	5.53	5.64	5.65	5.82	5.89	5.98	6.10	6.15	6.61	6.48
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/1/1996	5.10	5.35	5.65	6.03	6.22	6.39	6.54	6.65	6.99	6.88
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/1/1997	5.10	5.27	5.44	5.75	5.83	5.93	6.05	6.04	6.38	6.33
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/1/1998	4.23	4.36	4.28	4.17	4.10	4.10	4.26	4.33	5.09	4.90
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/1/1999	5.16	5.32	5.47	5.83	5.93	6.00	6.23	6.06	6.55	6.19
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/2/2000	6.27	6.33	6.06	5.98	5.92	5.86	5.95	5.83	6.18	5.93
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/1/2001	2.37	2.37	2.47	2.82	3.18	3.90	4.33	4.55	5.39	5.38
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/1/2002	1.59	1.54	1.56	1.80	2.11	2.75	3.34	3.72	4.81	4.93
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/3/2003	0.95	1.00	1.13	1.47	1.93	2.84	3.40	3.96	4.92	5.00
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/3/2004	1.71	2.00	2.21	2.63	2.92	3.44	3.85	4.21	4.95	5.06
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10/3/2005	3.61	4.02	4.09	4.21	4.23	4.25	4.31	4.39	4.67	4.58
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
12/29/2006	5.02	5.09	5.00	4.82	4.74	4.70	4.70	4.71	4.91	4.81
Mean	3.9608	4.1262	4.2823	4.6017	4.7826	5.0683	5.2958	5.4197	5.9159	5.8126
Variance	1.6682	1.7003	1.6668	1.5957	1.4810	1.2754	1.1619	1.0495	0.9245	0.8838
Skewness	−0.6212	−0.6341	−0.5938	−0.4993	−0.3850	−0.1179	0.0301	0.2847	0.3690	0.4946
Kurtosis	1.9471	1.9949	2.0410	2.1662	2.1925	2.1665	2.0982	2.2449	2.3388	2.5164

Last four rows are statistics of daily data of the Treasury (zero-coupon) bond yields.

Figure 3: The actual yields and the fitted curves

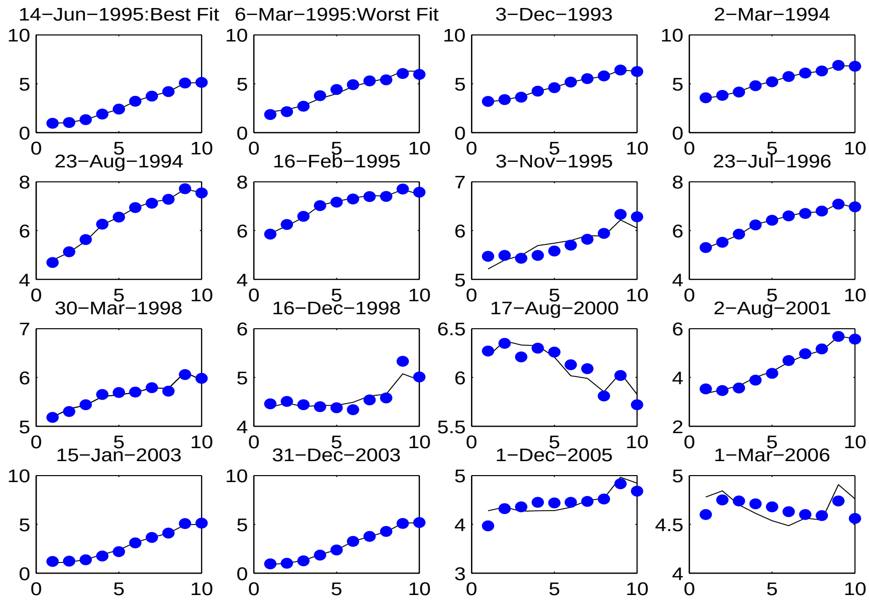


Table 2: Common Factor Analysis for US Treasury Bond Yield (10/01/1993-12/26/2006)

Eigen Value	17.2217	1.5741	0.0858	0.0133	0.0050	0.0029	0.0011	0.0009	0.0005	0.0003
Proportion	0.9109	0.0833	0.0045	0.0007	0.0003	0.0002	0.0001	0.0000	0.0000	0.0000
Cumulative	0.9109	0.9942	0.9987	0.9994	0.9997	0.9998	0.9999	1.0000	1.0000	1.0000
Eigenvector	0.3811	−0.4023	0.5167	0.4735	−0.0126	0.3917	−0.1712	0.1328	0.0338	0.0032
	0.3947	−0.3588	0.1996	−0.1514	0.1288	−0.5104	0.5809	−0.1623	−0.1089	−0.0044
	0.3954	−0.2261	−0.0734	−0.5362	0.0643	−0.1468	−0.6801	−0.0642	0.0243	−0.0812
	0.3832	−0.0455	−0.3898	−0.1234	−0.3747	0.2227	0.2358	0.2087	0.4369	0.4516
	0.3551	0.0704	−0.3922	0.0303	−0.1560	0.2831	0.1733	0.1106	−0.4330	−0.6164
	0.2990	0.2256	−0.2463	0.2510	0.3741	0.0753	−0.1010	−0.3293	−0.4392	0.5281
	0.2649	0.2955	−0.0962	0.4264	−0.0864	−0.2862	−0.1194	−0.4492	0.5015	−0.3074
	0.2277	0.3620	0.0432	0.0172	0.5832	−0.1273	0.0343	0.6361	0.2131	−0.0861
	0.1798	0.4265	0.3265	0.0455	−0.5694	−0.3626	−0.1396	0.2588	−0.3292	0.1627
	0.1608	0.4474	0.4512	−0.4495	0.0505	0.4466	0.1999	−0.3381	0.0994	−0.0474

Table 3: Analysis of ε_t^j in (4.1) for US Treasury Bond (10/01/1993-12/26/2006)

τ_j	min	max	Variance	$\text{Var}[\varepsilon_t^{\tau_j}]/\text{Var}[y_t^{\tau_j}]$
3 month	−0.5066	0.5163	0.0264	0.95%
6 month	−0.2178	0.2146	0.0050	0.17%
1 year	−0.2495	0.3121	0.0049	0.18%
2 year	−0.3301	0.3381	0.0143	0.56%
3 year	−0.2858	0.4443	0.0138	0.63%
5 year	−0.2357	0.2249	0.0070	0.43%
7 year	−0.1376	0.2085	0.0038	0.28%
10 year	−0.1582	0.1014	0.0023	0.21%
20 year	−0.2604	0.3033	0.0113	1.32%
30 year	−0.4531	0.3556	0.0209	2.68%

Figure 4: Empirical Common Factors

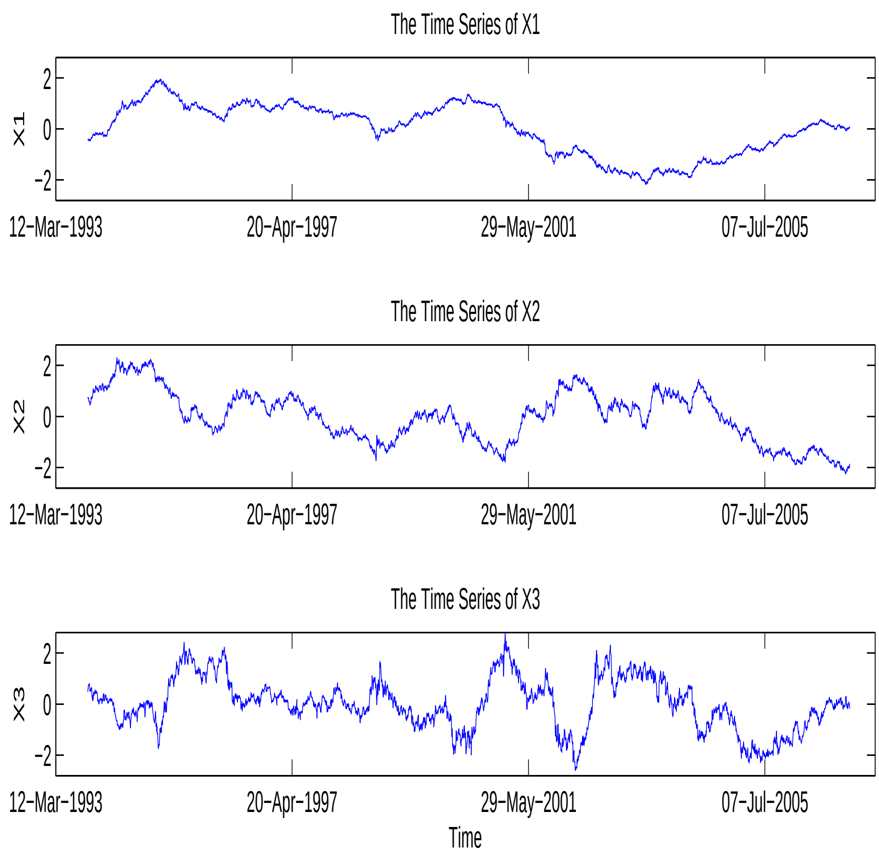


Figure 5: Sample paths of $\{X_t^1\}$ modelled by (4.3)

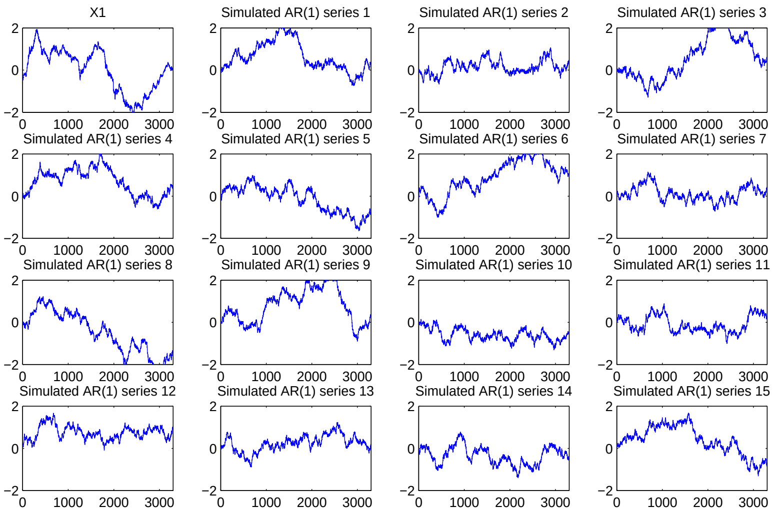


Figure 6: Sample paths of $\{X_t^2\}$ modelled by (4.4)

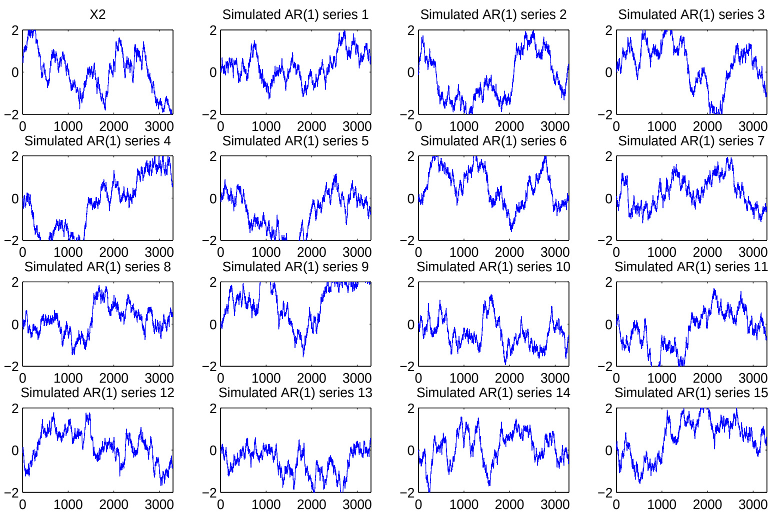


Figure 7: Sample paths of $\{X_t^3\}$ modelled by (4.5)

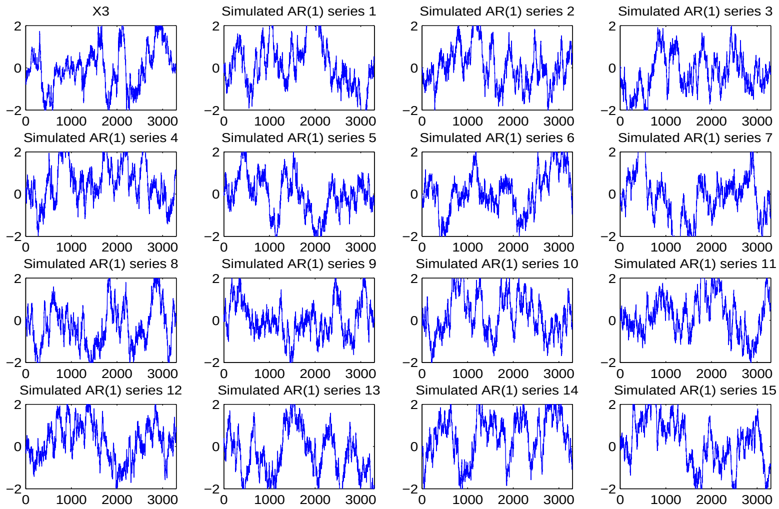
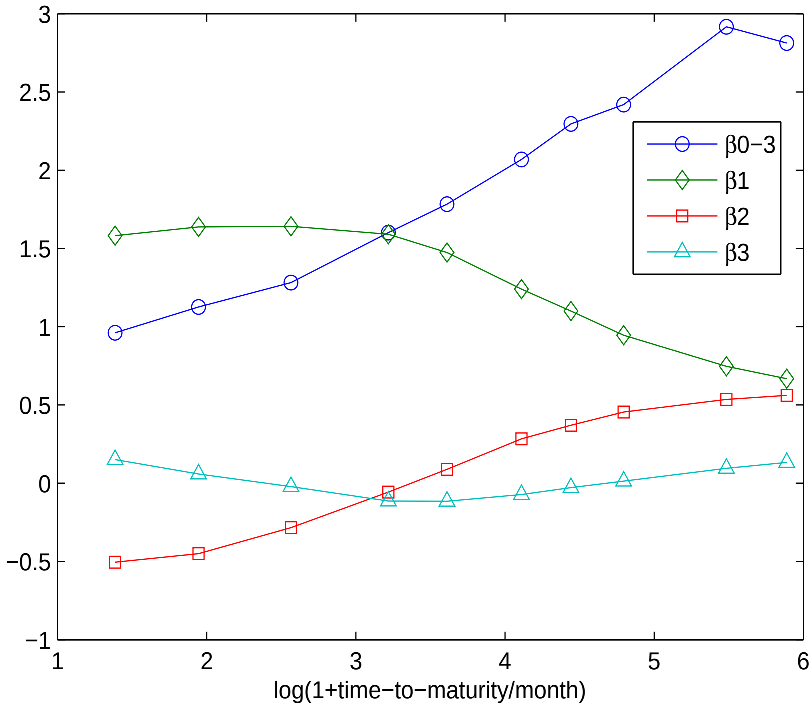


Table 4: Moments of Common Factors and the Corresponding AR(1) Residuals

	X_t^1	X_t^2	X_t^3	u_t^1	u_t^2	u_t^3
Variance	1.000	1.000	1.000	0.001	0.004	0.016
Skewness	−0.411	−0.027	−0.061	−0.008	0.509	0.314
Kurtosis	2.014	2.271	2.587	7.486	5.170	8.063

Figure 8: The Mean and the Loading Coefficients



The first three loading coefficients of each of the principal common factors are plotted as a function of the logarithm of the maturity of the yields. Here β^0, represents the mean of yields, has unit of 1% and is deducted by 3 for fitting in. Notice that (i) β^1 is positive, (ii) β^2 is increasing and changes sign at the middle, and (iii) the β^3 curve is convex, positive on both ends and negative in the middle.

5.0 THE INDEPENDENCY OF THE FIRST TWO COMMON FACTORS

In this section, we shall use copula theory to demonstrate empirically that the two common factors in (4.1) are independent.

5.0.10 Correlation

The most commonly used measure of dependence of random variables X and Y is the correlation coefficient

$$\rho(X,Y) = \frac{\text{Cov}[X,Y]}{\sqrt{\text{Var}[X]\text{Var}[Y]}}.$$

We use $\bar{\rho}$ to denote the sample correlation coefficient. By our derivation of the formula (4.1), we see that $\bar{\rho}(X_t^1, X_t^2) = 0$. Hence, if (X_t^1, X_t^2) are Gaussian distributed, we see that X_t^1 and X_t^2 are independent.

Our empirical data shows that (X_t^1, X_t^2) are not Gaussian, since the hypothesis that their marginal distributions are normal can be simply rejected. Hence, the single equation $\bar{\rho}(X_t^1, X_t^2) = 0$ alone is far from sufficient to say that X_t^1 and X_t^2 are independent.

5.0.11 The Kendall's τ

For empirical observation $\{X_t\}_{t=1}^T$ and $\{Y_t\}_{t=1}^T$ of time series $\{x_t\}$ and $\{y_t\}$, the Kendall's τ is defined as

$$\tau = \frac{2}{T(T-1)} \sum_{0t<\tau T} \text{sign}[(X_t - X_\tau)(Y_t - Y_\tau)].$$

It is an indicator to measure the difference between the probabilities of discordance and concordance. When X and Y are independent, the theoretical expectation of τ is zero.

The empirical value of τ is 0.07. This can be statistically regarded as zero.

5.0.12 Copula

To study independency of random variables, a very powerful tool is the theory of copula. In the sequel, we turn our attention to the copula approach for the dependence analysis of the common factors $\{X_t^1\}$ and $\{X_t^2\}$. The earliest paper explicitly relating copulas to the study of dependence among random variables appeared in Schweizer and Wolff (1981) [25]. Copulas allow us to separate the effect of the dependence from the effects of the marginal distributions.

In studying dependency of random variables, say X and Y with marginal cdf $F(\cdot)$ and $G(\cdot)$ respectively, one first normalized X and Y by the non-linear transformation

$$U = F(X), \qquad V = G(Y).$$

Then U and V are uniformly distributed on $[0, 1]$. In addition, X and Y are independent if and only if U and V are independent. Let $C(\cdot, \cdot)$ be the joint cumulative distribution function of (U, V): for every $(u, v) \in [0, 1]^2$,

$$C(u, v) = \text{Probability}(Uu, Vv) = \text{Probability}(F(X)u, G(Y)v).$$

Then C is called the copula of U and V. In general, a copula is defined as follows.

A 2-D copula is a function C defined on $[0, 1]^2$ that satisfies the following:

1. For every u, v in $[0, 1]$, $C(u, 0) = 0 = C(0, v)$ and $C(u, 1) = u, C(1, v) = v$;
2. For every u_1, u_2, v_1, v_2 in $[0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

We list a few commonly used copulas.

1. The **produce copula** models independent random variables and is given by

$$C_{\mathrm{I}}(u, v) = u\, v.$$

2. The Frechet-Hoeffding **upper bound copula**

$$C_{\mathrm{U}}(u, v) = \min(u, v).$$

3. The Frechet-Hoeffding **lower bound copula**

$$C_{\mathrm{L}}(u,v) = \max(u+v-1,0)$$

For any copula C, there holds

$$C_{\mathrm{L}}(u,v) \leq C(u,v) \leq C_{\mathrm{U}}(u,v).$$

4. The **Gaussian copula**

$$C_{\mathrm{Gaussian}}(u,v;\theta) = \Phi_G(\Phi^{-1}(u),\Phi^{-1}(v);\theta)$$

where $\theta \in [-1,1]$, $\Phi(\cdot)$ is the cdf of an N(0,1) random variable and Φ_G is the cdf of a 2-D Normal variable with zero mean and covariance matrix $[1,\theta;\theta,1]$.

5. The **Clayton copula**: $C_{\mathrm{Clayton}}(u,v;\theta) = (u^{-\theta}+v^{-\theta}-1)^{-\frac{1}{\theta}}$, $\theta > 0$.
6. The **Frank copula**: $C_{\mathrm{Frank}}(u,v;\theta) = -\frac{1}{\theta}\log(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$, $\theta \in R$.
7. The **Gumbel copula**: $C_{\mathrm{Gumbel}}(u,v;\theta) = \exp(-((-\log u)^{\theta}+(-\log v)^{\theta}))^{\frac{1}{\theta}}$, $\theta 1$.
8. The **FGM copula**: $C_{\mathrm{FGM}}(u,v;\theta) = uv(1+\theta(1-u)(1-v))$, $|\theta| 1$.

The development of copula is based on the following Sklar Theorem.

[Sklar's Theorem] Let H be a joint distribution function with margins F and G. Then there exists a copula C such that for all x,y in R, $H(x,y) = C(F(x),G(y))$.

If F and G are continuous, then C is unique; otherwise, C is uniquely determined on Range(F)×Range(G). Conversely, if C is a copula and F and G are distribution functions, then H defined above is a joint distribution function with margins F and G.

Sklar's Theorem tells us that for any two random variables with joint and marginal distribution known, the copula function will be determined if the marginal distributions are continuous. On the other hand, if we know the marginal distributions, different copulas can applied to approximate the joint distributions. This will give us an elastic framework to do the analysis of dependence relation between two random variables.

5.0.13 The Empirical Copula

To demonstrate that the common factors we obtained in (4.1) are independent, we measure the distance between the empirical copula derived from the common factors $\{X_t^1\}$ and $\{X_t^2\}$ and the copulas describes in the examples above. We shall show that empirical copula is close to the product copula, i.e. $\{X_t^1\}$ and $\{X_t^2\}$ can be statistically regarded as independent.

Given observations $\{X_t\}_{t=1}^T$ and $\{Y_t\}_{t=1}^T$ of random variables, a commonly used empirical distribution of X and Y are defined by

$$\bar{F}(x) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(xX_t\}, \qquad \bar{G}(y) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}(yY_t\},$$

where $\mathbf{1}$ is the indicator function that takes value 1 when its logical argument is true and zero when its logical argument is false.

Now let $U = F(X)$ and $V = G(Y)$. The empirical copula for U and V is first defined on the lattice

$$\Sigma = \left\{ \left(\frac{t_1}{T}, \frac{t_2}{T}\right) : t_1, t_2 = 0, 1...T \right\}$$

by

$$C_{\text{Empirical}}(u, v) = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\Big(\bar{F}(X_t)u\Big)\, \mathbf{1}\Big(\bar{G}(Y_t)v\Big) \quad \forall\, u, v \in \Sigma,$$

and then extended over $[0, 1]^2$ by a linear interpolation.

We will use the following norms to measure the difference between the empirical copula and the target copulas:

1. The L^1 distance $\|\cdot\|_{L_1}$:

$$\|f - g\|_{L_1} = \int_0^1 \int_0^1 |f(u, v) - g(u, v)| dudv.$$

2. The L^2 distance:

$$\|f - g\|_{L_2} = \Big(\int_0^1 \int_0^1 [f(u, v) - g(u, v)]^2 dudv \Big)^{\frac{1}{2}}.$$

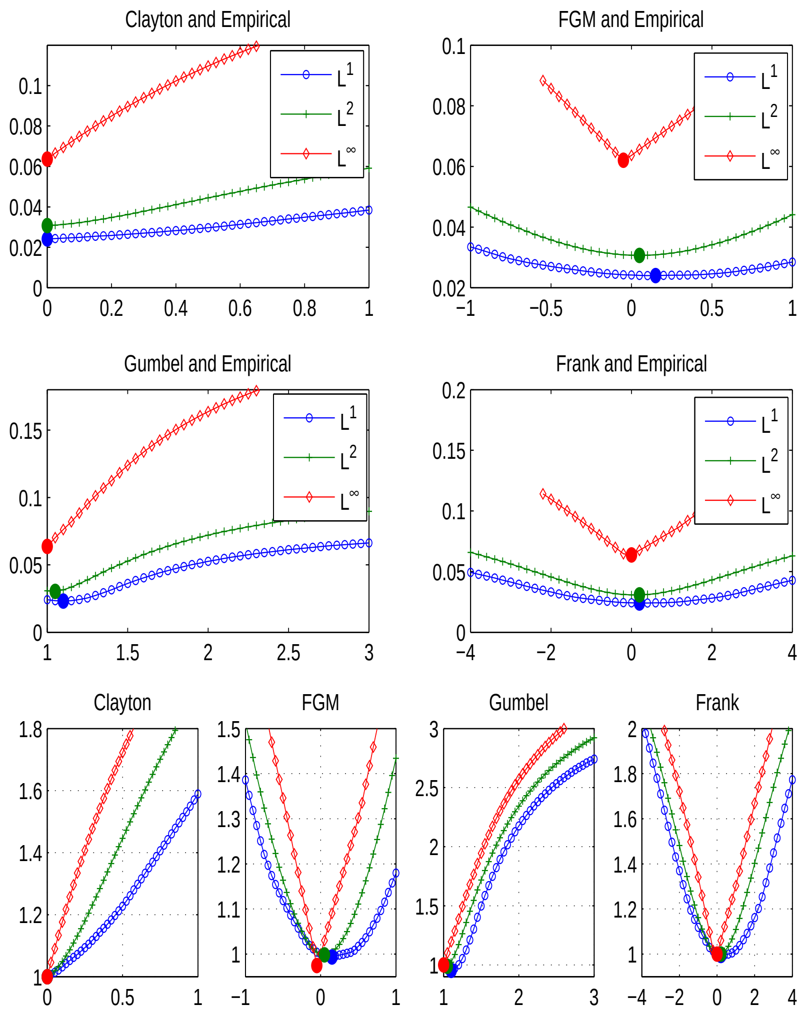
3. The L^∞ distance:

$$\|f - g\|_\infty = \sup_{0u,v1} |f(u, v) - g(u, v)|\,.$$

Based on these measures, we calculate the difference between the empirical copula and several standard copulas. Figure 9 shows the difference between the empirical copula and several parametric copulas. From these figures, we see that the minimum distance between empirical distance and the parametric copulas are attained near those values ($\theta = 0$ for Gaussian, Clayton, and FGM, and $\theta = 1$ for for Gumbel) that the parametric copula becomes the produce copula

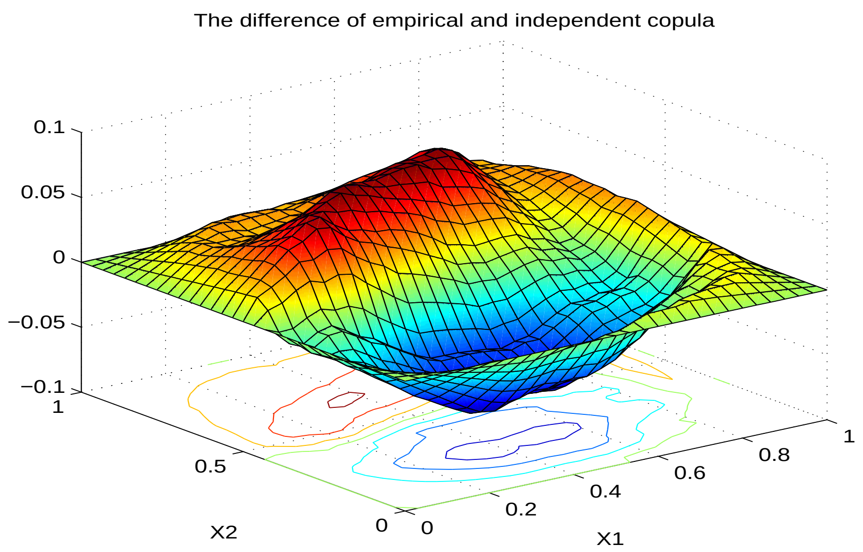
In Figure 9, we also list ratios of these distances to that from the product copula being 0.024, 0.031, and 0.064 for the L^1, L^2 and L^∞ distances, respectively. It shows that the empirical copula for our common factors $\{X_t^1\}$ and $\{X_t^2\}$ has the minimum distance to the product copula in L_1, L_2 and L_∞ norms. And in Figure 10, we plot the 3-D difference between empirical copula and the produce copula. From the smallness of the distance between empirical copula and the produce copula, we can reasonably conclude that the two common factors in (4.1) are independent.

Figure 9: The choice of optimal dependence parameters



The top four plots are the L^1, L^2 and L^∞ distances between the empirical copula and several parametric copulas. The four plots in last row are ratios of these distances to that from the product copula being 0.024, 0.031, and 0.064 for the L^1, L^2 and L^∞ distances, respectively.

Figure 10: The difference between the empirical Copula and the product copula



The difference between the empirical Copula and the product copula.

6.0 CONCLUSIONS

By utilizing common factors, we have been able to show that all US Government bonds from 1993 to 2006 can be described by a two factor model (4.1), where the remainder term ε_t is composed of about 0.4% of the total variance of the yields. Also, we have shown empirically that the two factors X_t^1 and X_t^2 can be regarded as independent. The model (4.1) with remain term ε_t deleted thus provides a simple yet effective model for term structure.

In the near future, we shall study the following:

1. the marginal distribution of common factors;
2. the dynamics of the state variable $\{X_t^i\}$, $i = 1, 2$, based on simple AR(1) and ARMA(1,1) models; our philosophy is the simple the better;
3. the structure of the load coefficients;
4. in a framework similar to Black–Scholes no arbitrage argument, establish a complete theoretical description of the term model based on (4.1) and our findings on empirical experiments.

BIBLIOGRAPHY

[1] ALEXANDER, C., "Market Models: A Guide to Financial Data Analysis", Wiley, Hoboken, NJ. 2001.

[2] BABBS, S. H., AND NOWMAN, B. K., "Kalman Filtering of Generalized Vasicek Term Structure Models", *Journal of Financial and Quantative Analysis*, **34(1)**, 115-129, March, 1999.

[3] BABBS, S. H., AND NOWMAN, B. K., "Econometric Analysis of a Continuous Time Multi-Factor Generalized Vasicek Term Structure Model: International Evidence", *Asia-Pacific Financial Markets*, **5**,159-183, 1998.

[4] BAUM, C., AND BEKDACHE, B., "Factor-GARCH Modeling of the Treasury Term Structure", working paper, Boston College, June, 1996.

[5] BLACK, F., DERMAN, E., AND TROY, W., "A One-Factor Model on Interest Rates and its Ajpplication to Treasury Bond Options", *Financial Analysis Journal*, **46**, 33-39, 1990.

[6] BLACK, F., KARASINSKI, P., "Bond and Option Pricing when Short Rates are Lognormal", *Financial Analysis Journal*, 52-59, July-August, 1991.

[7] CAMPBELL, J. Y., LO, A. W., AND MACKINLAY, A. C., *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ, 1997.

[8] COCHRANE, J.H., *Asset Pricing*, Princeton University Press, Princeton, NJ,2001.

[9] CONNOR, G. , "The Three Types of Factor Models: A comparison of Their explanatory Power", *Financial Analysts Journal*, **51**, 42-46, 1995.

[10] COX, J. C., INGERSOLL, J. E., AND ROSS, S. A., "A Theory of The Term Structure of Interest Rates", *Econometrica*, **53(2)**, 385-407, Mar., 1985.

[11] DAI, Q., SINGLETON, K. J., "Specification Analysis of Affine term Structure Models", *Journal of Finance*, **55**, 1943-1978, 2000.

[12] DUFFIE, D., FILIPOVIC, D., AND SCHACHERMAYER, W., "Affine Processes and Applications in Finance", *The Annals of Applied Probability*, **13**, 984-1053, 2003.

[13] Duffie, D., Kan, R., "Multi-Factor Term Structure Models", Phil. Trans. R. Soc. Lond. A, **347**, 577-586, 1994.

[14] Duffie, D., Kan, R., "A Yield-Factor Model of Interest Rates", *Mathematical Finance*, **6**, 379-406, 1996.

[15] Duffie, D., Pan, J., Singleton, K. J., "Transform Analysis and Asset Pricing for Affine Jump-Diffusions", *Econometrica*, **6**, 1343-1376, Oct.,2000.

[16] Heath, D., Jarrow, R., & Morton, A., "Bond Pricing and the Term Structrue of Interest Rates: A New Methodology for Contingent Claims Valuation", *Econometrica*, **60**, 77-106, 1992.

[17] Ho, T., & Lee, S., "Term Structure movements an pricing interest rate contingent claims", *Journal of finance*, **42**, 1129-1142, 1986.

[18] Jong, F. D., "Time Series and Cross-Section Information in Affine Term-structure Models", *Journal of Business & Economic Statistics*, **18(3)**, 300-314, July, 2000.

[19] Junker, M., Szimayer, A., and Wagner, N., "Nonlinear term structure dependence: Copula Functions, Empirics, and Risk Implications", *Journal of Banking & Finance*, **30**(2006), 1171-1199.

[20] Lemke, W., "Term Structure Modeling and Estimation in a State Space Framework", Springer, 2006.

[21] Litterman, R., Scheinkman, J., (1991), "Common factors affecting bond returns", *Journal of Fixed Income*, **1**, 62–74.

[22] Pagan, A. R., Hall, A. D., and Martin, V., "Modeling the term structure", working paper, Bond University, 1995.

[23] Piazzesi, M., "Affine Term Structure Models", working paper, University of Chicago, 2003.

[24] Stoffer, D. S., Time Series Analysis and Its Applications with R Examples, *2nd Ed*, Springer, 2006.

[25] Schweizer, B., and Wolff, E. F., "On nonparametric Measures of Dependence for Random Variables", *Annals of Statistics*, **9**, 870-885, 1981.

[26] Tsay, R. S., *Analysis of Financial Time Series*, 2nd Edition, Wiley, 2005.

[27] Vasicek, O. A., "An equilibrium characteristization of the Term Structure", *J. Financial Economics*, **5**, 177-188, 1977.

[28] Zivot, E. and Wang, J., *Modeling Financial Time series with S-Plus*, Springer-Verlag, New York, 2003.