Dynamics of Interest Rate Curve by Functional Auto-regression

V. Kargin[1] (Cornerstone Research)
A. Onatski[2] (Columbia University)

March, 2004

Abstract
The paper applies methods of functional data analysis – functional auto-regression, principal components and canonical correlations – to the study of the dynamics of interest rate curve. In addition, it introduces a novel statistical tool based on the singular value decomposition of the functional cross-covariance operator. This tool is better suited for prediction purposes as opposed to either principal components or canonical correlations. Based on this tool, the paper provides a consistent method for estimating the functional auto-regression of interest rate curve. The theory is applied to estimating dynamics of Eurodollar futures rates. The results suggest that future movements of interest rates are predictable only at very short and very long horizons.

1. Introduction

Interest rates are both a barometer of the economy and an instrument for its control. In addition, evolution of interest rates enters as a vital input in valuation of many financial products. Consequently, the evolution of interest rates is of great interest to both macroeconomics scholars and financial economists.

It is a widespread opinion that interest rate dynamics can be completely described in terms of 3 factors, which are often modeled as principal components of the interest rate variation. Indeed, more than 95% percent of the variation can be decomposed in 3 factors. However, recent research (Cochrane and Piazzesi (2002)) has suggested that this opinion may have flaws.

The new evidence indicates that projecting interest rate curve to the 3 main principal components severely handicaps the ability to predict future interest rates. In particular, it appears that the best predictors of interest rate movements are among those factors that do not contribute much to the overall interest rate variation.

If not principal components, then what statistical tool is appropriate to the analysis of interest rates predictability? One possibility is the canonical correlation analysis, a tool that along with principal components was invented by Harold Hotelling in early 1930s. This tool should be modified, of course, to account for functional nature of interest rate data (Leurgans, Moyeed, Silverman (1993)). It turns out, however, that there is a statistical technique that is even better suited for the prediction purposes. The technique is based on the singular value decomposition of the cross-covariance operator. The technique seems to be relatively novel and we call it singular factor analysis. We place

[1] skarguine@cornerstone.com ;
Cornerstone Research, 599 Lexington Avenue floor 43, New York, NY 10022
[2] ao2027@columbia.edu;
Economics Department, Columbia University, 1007 International Affairs Building MC 3308, 420 West 118th Street , New York, NY 10027

this technique in framework of functional auto-regression (Bosq 1995), which is natural for modeling dynamic functional data.

As an application, we illustrate the method using ten years of data on Eurodollar futures contracts. Consistently with previous research, we find that the best predictors of future interest rates are not among the largest principal components but are hidden among the residual components. We also find that the best level of predictability is at the 1-day and 1-year ahead horizons.

Our analysis contribute to a long standing problem of whether the interest rates are predictable. Some research (Duffee (2002), Ang and Piazzesi (2003)) indicates that it is hard to predict better than simply assuming random walk evolution. This means that today's interest rate is the best predictor for tomorrow's interest rate, or, for that matter, for the interest rate one year from now. The subject, however, is torn with controversy. Cochrane and Piazzesi (2002) and Diebold and Lie (2002) report, on the contrary, that their methods improve over random walk prediction. We find that the controversy may be result of whether the researchers choose to restrict attention to 3 main principal components or not.

The limitation of our approach is that we do not attempt to use non-interest rate information such as the current level and innovation in inflation, GDP growth rate and other macro variables. Recently it was discovered (Ang and Piazzesi (2003)) that there is a significant correlation in the dynamic of macro variables and interest rates. We believe that after suitable modification our method can also be applied to include macro variables.

We also do not aim to derive implications of the interest rate predictability neither for the control of economy by interest rate targeting, nor for the management of financial portfolios. We believe that methods of functional data analysis may be as useful in the problems of control as in the problems of estimation. With respect to the financial applications, we want to point out that predictability of future interest rates does not necessarily imply arbitrage opportunities. Rather the relevant question is whether portfolios that correspond to the predictable combinations of interest rates generate excess returns that cannot be explained by traditional risk factors. This is a question for a separate research effort.

The rest of the paper is organized as follows. The model is described in Section 2. The principal component method of estimation is in Section 3. A modification of canonical correlation method is in Section 4. The data are described in Section 5. The results of estimation of principal components are briefly summarized in Section 6. The results of estimation of singular factors are in Section 7. And Section 8 concludes.

2. Model

Let $P_t(T)$ denote time t price of a coupon-free bond with maturity at time $t+T$. Assume this function is differentiable in T. Then we can define forward rates as

$$f_t(T) = -\frac{\partial \log P_t(T)}{\partial T}.$$

The interest rate can conceivably exist for every positive maturity, which is less than a certain bound $\bar{T}$. But in practice, only a discrete subset of these interest rates is observable at each moment of time. In addition, the maturities of observable interest rates are typically change from time to time and an interpolation procedure is usually used to infer the curve of the interest rates. These facts mean that a natural way to model the interest rates is as an imperfectly observed function of the continuous parameter T.

Let us abuse the notation and use $f_t(T)$ for the forward rate curve with the subtracted mean, $f_t(T) - \bar{f}(T)$. We will model the forward interest rates as evolving according to one of the following functional auto-regressions:

(A) $$f_{t+\delta}(T) = \rho[f_t(T)] + \varepsilon_t(T)$$

or

(B) $$f_{t+\delta}(T) - f_t(T) = \rho[f_t(T) - f_{t-\delta}(T)] + \varepsilon_t(T),$$

or

(C) $$f_{t+\delta}(T) - f_t(T) = \rho[f_t(T)] + \varepsilon_t(T),$$

where $f_t(T)$ is an random variable that takes values in the real Hilbert space of square-integrable functions on $[0, \bar{T}]$, ρ is a Hilbert-Schmidt integral operator, and $\varepsilon_t(T)$ is a strong H-white noise in the Hilbert space.

The assumption that coefficient operator, ρ, is a Hilbert-Schmidt integral operator means that ρ acts as follows

$$\rho : x(T) \to \int_0^\infty \rho(S,T)x(T)dT,$$

where $\int_0^\infty \int_0^\infty \rho(S,T)^2 dSdT < \infty$.

Because of this assumption, model (C) is not reducible to model (A): If operator ρ is Hilbert-Schmidt, operator $I + \rho$ is not, and vice versa.

All these models are particular cases of the general model

(M) $$f_{t+\delta}(T) = \sum_{i=0}^{n} (D_i + \rho_i) L^i [f_t(T)] + \varepsilon_{t+\delta}(T),$$

where L is the time-shift operator:

$$L : f_t(T) \to f_{t-\delta}(T),$$

and D_i and ρ_i are differential and integral operators. However, we will not work in this generality.

In intuitive terms, in model (A) the expectation of future values of the forward rate curve is determined by the current values. In model (B) the expectation of the future changes is determined by the past changes. In model (C) the expectation of the future changes of the forward rate curve is determined by the current values of the curve

We aim to estimate operator ρ using a finite sample of imperfectly observed curves and predict the future curve using this estimate. In the following sections we describe several approaches to the estimation.

We conclude this section by briefly describing the formalism of Hilbert space valued random variables and explaining how it relates to a more familiar language of random processes.

Consider an abstract real Hilbert space H. Let function f_n map a probability space $(\Omega,\mathbb{A},P)$ to H. We call this function an H-valued random variable if the scalar product (g, f_n) is a standard random variable for every g from H.[3]

Definition 1. If $E\|f\| < \infty$, then there exists an element of H, denoted as Ef and called ***expectation*** of f, such that

$$E(g, f) = (g, Ef), \text{ for any } g \in H .$$

Definition 2. Let f be an H-valued random variable, such that $E\|f\|^2 < \infty$ and $Ef = 0$. The ***covariance operator*** of f is the bounded linear operator on H, defined by

$$C_f(g) = E[(g, f)f], \quad g \in H .$$

If $Ef \neq 0$, one sets $C_f = C_{f-Ef}$.

Definition 3. Let f_1 and f_2 be two H-valued random variables, such that $E\|f_1\|^2 < \infty, E\|f_2\|^2 < \infty$ and $Ef_1 = Ef_2 = 0$. Then the ***cross-covariance operators*** of f_1 and f_2 are bounded linear operators on H defined by

$$C_{f_1,f_2}(g) = E[(g, f_1)f_2], \quad g \in H,$$
$$C_{f_2,f_1}(g) = E[(g, f_2)f_1], \quad g \in H.$$

If $Ef_1 \neq 0$ or/and $Ef_2 \neq 0$, one sets

$$C_{f_1,f_2} = C_{f_1-Ef_1, f_2-Ef_2},$$
$$C_{f_2,f_1} = C_{f_2-Ef_2, f_1-Ef_1}.$$

Definition 4. A sequence $\{\eta_n, \quad n \in Z\}$ of H-valued random variables is said to be ***H-white noise*** if

1) $0 < E\|\eta_n\|^2 = \sigma^2 < \infty, \quad E\eta_n = 0, \quad C_{\eta_n}$ do not depend on n and

[3] The definitions that follow are slight modifications of those in Chapters 2,3 of Bosq (2000).

2) η_n is orthogonal to η_m; $n,m \in Z, \quad n \neq m$; i.e.,

$$E\{(x,\eta_n)(y,\eta_m)\} = 0, \text{ for any } x, y \in H.$$

$\{\eta_n, \quad n \in Z\}$ is said to be a **strong H-white noise** if it satisfies 1) and

2') $\{\eta_n, \quad n \in Z\}$ is a sequence of i.i.d. H-valued random variables.

Example: Stochastic Processes

Consider a set of stochastic processes, $f_i(T)$, on interval $[0,\bar{T}]$. Let $Ef_i(T) = 0$ for each T. Let the covariance function of each process be $Ef_i(S)f_i(T) = \Gamma_{ii}(S,T)$, and cross-covariance function between two processes be $Ef_i(S)f_j(T) = \Gamma_{ij}(S,T)$. Assume that with probability 1 the sample paths of the processes are in $L^2[0,\bar{T}]$. Each stochastic process defines an H-valued random variable with zero mean. The covariance operator of f_i is the integral operator with kernel $\Gamma_{ii}(S,T)$:

$$C_{f_i} : g(T) \mapsto \int_0^{\bar{T}} \Gamma_{ii}(S,T) g(T)\, dT,$$

and the cross-covariance operator of f_i and f_j is the integral operator with kernel $\Gamma_{ij}(S,T)$. We will abuse notation and identify variance and covariance operators with their kernels.

3. Estimation by Principal Components

The natural estimator for the mean forward curve, $\bar{f}(T)$, is

$$\hat{f}(T) = \frac{1}{n}\sum_{i=1}^{n} f_{i\delta}(T).$$

To define an estimator of ρ, we relate it to covariance and cross-covariance operators.

Consider for definiteness model (C). Let Γ_{11} be variance operator of random curve f_t and Γ_{12} be the cross-covariance operators for curves $f_t(T)$ and $f_{t+\delta}(T) - f_t(T)$. From the model we know that

$$\begin{aligned}\Gamma_{12}(S,T) &= E\{[f_{t+\delta}(S) - f_t(S)]f_t[T]\} \\ &= E\left\{\int \rho(S,S')f_t(S')f_t(T)dS'\right\} \\ &= \int \rho(S,S')\Gamma_{11}(S',T)dS',\end{aligned}$$

or, in operator form

$$\Gamma_{12} = \rho\Gamma_{11}$$

Assuming that Γ_{11} is invertible, we have $\rho = \Gamma_{12}\Gamma_{11}^{-1}$. The natural procedure would be to substitute estimates of the covariance and cross-covariance operators into this formula. As we will see shortly, substituting the empirical covariance and cross-covariance operators will not work properly, so we will need to modify this procedure. In this section we focus on the estimation procedure that uses principal components.

The kernels of the empirical covariance and cross-covariance operators are

$$\hat{\Gamma}_{11}(T_1,T_2)=\frac{1}{n}\sum_{i=1}^{n} f_{i\delta}(T_1)f_{i\delta}(T_2)-\hat{f}(T_1)\hat{f}(T_2)\,,$$

$$\hat{\Gamma}_{12}(T_1,T_2)=\frac{1}{n-1}\sum_{i=1}^{n-1} f_{i\delta}(T_1)f_{(i+1)\delta}(T_2)-\hat{f}(T_1)\hat{f}(T_2)\,.$$

Unfortunately, empirical covariance operator $\hat{\Gamma}_{11}$ is singular and cannot be inverted. We must use a regularization method to obtain a consistent estimate of ρ. The regularization method, advocated by Bosq, uses the principal components of the empirical covariance operator. The idea is to determine how the operator acts on the components of $f_t(T)$ that contribute most of the variation.

Let π_{k_n} is the projector on a set of k_n eigenvectors of $\hat{\Gamma}_{11}$, associated with the largest eigenvalues. Denote the span of this eigenvectors as H_{k_n} and define the regularized covariance and cross-covariance estimates $\widetilde{\Gamma}_{11}=\pi_{k_n}\hat{\Gamma}_{11}\pi_{k_n}^{'}$ and $\widetilde{\Gamma}_{12}=\pi_{k_n}\hat{\Gamma}_{12}\pi_{k_n}^{'}$. These are simply the empirical covariance and cross-covariance operators restricted to H_{k_n}. Then define

$$\widetilde{\rho}=\pi_{k_n}^{'}\widetilde{\Gamma}_{12}\widetilde{\Gamma}_{11}^{-1}\pi_{k_n}\,.$$

Note that $\widetilde{\rho}$ is $\widetilde{\Gamma}_{12}\widetilde{\Gamma}_{11}^{-1}$ on H_{k_n}, and zero on the orthogonal complement to H_{k_n}. The claim is that under certain conditions this estimator is consistent.

Assume that ρ is a Hilbert-Schmidt operator . Let $a_1=(\lambda_1-\lambda_2)^{-1}$, and

$$a_i=\max\left\{(\lambda_{i-1}-\lambda_i)^{-1},(\lambda_i-\lambda_{i+1})^{-1}\right\} \text{ for } i>1\,,$$

where λ_i are eigenvalues of the covariance function $\Gamma_{11}(T_1,T_2)$ ordered in the decreasing order.

Theorem 1. *If for some* $\beta>1$

$$\lambda_{k_n}^{-1}\sum_{1}^{k_n} a_j=O\left(n^{1/4}(\log n)^{-\beta}\right),$$

then $\widetilde{\rho}_n$. *is consistent in operator norm induced by* L^2 *norm:*

$$\left\|\widetilde{\rho}_n-\rho\right\|_{L^2}\to 0 \text{ a.s.}$$

Proof: This is a restatement of Theorem 8.7 in Bosq.

The condition of the theorem requires that the eigenvalues of the covariance matrix do not approach zero too fast, and that the eigenvalues are not too close to each other. The efficiency of the estimation procedure depends on the judicious choice of parameter k_n. A possible approach is to select k_n so that it minimized a norm of empirical prediction error on a sub-sample of data.

4. Canonical Correlations and Singular Factors

The estimation method outlined in the previous section may perform very badly if it happens that the best predictors for future evolution of interest rates have nothing to do with the largest principal components. In this case we would better off by searching for predictors directly without first projecting interest rates on the largest principal components.

Statisticians long ago recognized the need for a method of finding the most important relations between two random vectors. In early 1930s Harold Hotelling, who previously invented the principal components, suggested a new method of data analysis, the canonical correlation analysis (CCA). This method is described in detail in Anderson (1984). In the classical situation we are given two random vectors, x_i and y_i. We are looking for vectors of coefficients, u_i and v_i that maximize correlations $Cov(ux, vu)$ subject to constraints $Var(ux) = Var(vy) = 1$. Here ux and vx mean the scalar products. If no other constraints are imposed, then the maximized covariance is called ***the first canonical correlation*** of x and y, and the maximizing vectors $u^{(1)}$ and $v^{(1)}$ are called ***the first canonical variates***.

The second canonical correlation is determined by the same maximization problem with added constraints that ux and vy must be uncorrelated with $u^{(1)}x$ and $v^{(1)}y$. The following canonical correlations and variates are defined in a similar fashion.

The definition suggests finding the solutions by a sequence of constrained maximizations. Another approach is to note (by writing out the Lagrange multiplier conditions) that this problem is equivalent to finding such λ that guarantees the existence of nontrivial solutions of the following equation:

$$\begin{pmatrix} -\lambda\Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & -\lambda\Gamma_{22} \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix} = 0,$$

where Γ_{ij} are variance-covariance matrices of vectors x and y.

This set-up can be generalized in a straightforward way to the functional case. But in estimation a difficulty arises. The classical method of estimation is by substituting the sample variances and covariances for the population variances and covariances. This procedure breaks down and gives inconsistent estimates if the data are functional. Intuitively the reason is that the number of degrees of freedom is larger than the number of curves in the sample, so there is a potential for finding spurious high correlations.

A consistent method for estimating the canonical correlations was suggested by Leurgans, Moyeeed and Silverman (1993). Its essence is that functions that enter the maximization problem are penalized for roughness. They called this method smoothed canonical correlation analysis and proved it consistency.

The estimation of functional autoregression can then proceed as in the case of principal components, by using the canonical variates as the basis for dimension reduction. If u_i

and v_i are the canonical variates and λ_i is a canonical correlation, then the coefficient operator of the auto-regression, ρ, simply maps u_i to $\lambda_i v_i$. Substituting the consistent estimates for the canonical variates and correlations in this expression we can expect to get the consistent estimate of ρ.

While canonical correlations are useful for exploring the relationships in the data, they may, however, have deficiencies from the point of view of prediction. The reason is that it may easily happen that 2 linear combinations have a large correlation but the correlation will explain only little of future variation. At the same time, linear combinations with lower correlation may help more in explaining the future variation. Consequently, focusing attention on finding the largest correlations may be not the best strategy for prediction.

As an analogy, consider 3 agricultural crops and a single predictor, the amount of precipitation during the previous winter. Harvests of the first crop are very volatile and unrelated to the predictor, harvests of the second are a bit less volatile and have a correlation of about 80% with the predictor. Finally, the third crop has very steady harvests and their variation can be almost perfectly predicted by the precipitation amount. Assuming that only one crop can be chosen for prediction, which one to choose?

The answer depends on the preferences of the forecaster but it seems likely that in many situations the second crop will be chosen for prediction. For example, this is a likely outcome if the purpose is to predict the market price of the crop and the demand for all the crops is equally elastic.

Our situation is similar. Often it is reasonable to aim for non-perfect but economically most significant prediction.

In a sense, we need something in between the method of principal components that pays most attention to the largest components of variance and the method of canonical correlations that pays most attention to the largest correlations. A suitable alternative criterion appears to be the expectation of the scalar product of the predicted curve with the realized curve subject to a certain normalization. In the remainder of this section we will develop methods based on this criterion.

Let g_t and f_t are two curves in H and our task is to predict g_t using f_t. Let R be a linear bounded real functional, $R : H \mapsto \Re$, where $\Re$ is the space of real numbers. By the Riesz theorem it can be represented by a function, $R(T)$:

$$R : f(T) \mapsto \int_0^{\bar{T}} f(T)R(T)dT .$$

Let F be a linear H-valued function, $F : \Re \mapsto H$. By linearity, it also can be represented by a function, $F(T)$:

$$F : x \mapsto F(T)x .$$

We will look for R and F that maximize

(P) $E(g_t, FRf_t)$ subject to $\|R\| \le 1$ and $\|F\| \le 1$.

We will say that the solution of this problem is ***the first singular value,*** and the maximizing function R is ***the first singular factor.*** We will also say that function F is ***the curve of first singular factor loadings***.

By way of comparison, the problem of finding the first canonical correlation is the problem of maximizing

(P1) $E(g_t, FRf_t)$ subject to $E\|Rf_t\| \le 1$ and $E\|Fg_t\| \le 1$.

Thus the problem is different only by a constraint imposed on functions R and F. In comparison with problem P1, the constraints in problem P essentially prevents these functions from exaggerating importance of those linear combinations of f_t and g_t that have low variance.

In functional form problem (P) can be written as

$$\max E\left\{\int_0^{\bar{T}} g_t(T)F(T)\left(\int_0^{T} R(S)f_t(S)\,dS\right)dT\right\} = \iint F(T)\Gamma_{12}(T,S)R(S)\,dT\,dS$$

subject to

$$\int_0^{\bar{T}} [R(T)]^2\,dT \le 1 \text{ and } \int_0^{\bar{T}} [F(T)]^2\,dT \le 1.$$

Essentially we are looking for a function R that has the unit norm and whose image $\Gamma_{12}R$ has the maximal possible norm. This is the problem of finding the largest singular value and the corresponding singular vector of operator Γ_{12}. We can solve it by finding eigenvalues and eigenvectors of operator $\Gamma_{12}^*\Gamma_{12}$ that has kernel

$$\Gamma_{12}^*\Gamma_{12}(S_1,S_2) = \int_0^{\bar{T}} \Gamma_{12}(S_1,T)\Gamma_{12}(T,S_2)\,dT.$$

Function R is the eigenfunction that corresponds to the largest eigenvalue of this operator. When R is found, $F = \Gamma_{12}R/\|\Gamma_{12}R\|$.

Finding the next factor is essentially the same maximization problem, subject to the additional constraint that R must be orthogonal to the first factor R_1. This implies that the solution of the maximization problem is the second singular value and the corresponding singular vector of operator Γ_{12}. We can continue this operation and define the set of factors, R_i, that we will call ***singular factors***. Corresponding functions F_i will be called ***singular factor loadings.***

We can improve the method using the idea of smoothing from Leurgans, Moyeed and Silverman (1993) and use the maximization with adjusted constraints:

$\|R\| + \alpha[R] \le 1$ and $\|F\| + \alpha[F] \le 1$,

where $[f]$ penalizes function f for roughness.

Given that the singular factors and loadings are estimated, we can address the problem of finding the best predictor. First thing to note is that it is easy to estimate the coefficient matrix in the auto-regression in terms of singular factors and loadings.

Indeed, singular factors and loadings give a convenient representation of cross-covariance operator. Assume for simplicity, that there is only a finite number, k, of non-zero singular factors. Take them as basis vectors and complete them to an orthogonal basis in H. Operator Γ_{12} in this basis acts as follows

$$\Gamma_{12} : R_i \mapsto \begin{cases} \lambda_i F_i & if \quad i = 1\ldots, k, \\ 0 & otherwise. \end{cases}$$

Define $G_i = \Gamma_{11} R_i$. Since $\rho = \Gamma_{12}\Gamma_{11}^{-1}$, operator ρ acts as follows:

$$\rho : G_i \mapsto \begin{cases} \lambda_i F_i & if \quad i = 1\ldots, k, \\ 0 & otherwise. \end{cases}$$

It is natural to estimate operator ρ by substituting the empirical counterparts of singular factors and loadings into this formula. Let functions $\hat{R}_i$ be k first eigenfunctions of $\hat{\Gamma}_{12}^{*}\hat{\Gamma}_{12}$, functions $\hat{F}_i$ be eigenfunctions of $\hat{\Gamma}_{12}\hat{\Gamma}_{12}^{*}$, variables $\hat{\lambda}_i$ be corresponding singular values and $\hat{G}_i = \hat{\Gamma}_{11}\hat{R}_i$. Define $\hat{\rho}$ as the linear extension of the following action:

$$\hat{\rho} : \hat{G}_i \mapsto \begin{cases} \hat{\lambda}_i \hat{F}_i & if \quad i = 1\ldots, k, \\ 0 & otherwise. \end{cases}$$

Theorem 2 (Consistency) *As n approaches ∞, operator $\hat{\rho}$ converges in norm to ρ with probability 1.*

Proof is based on the following lemmas

Lemma 1 *Operators $\hat{\Gamma}_{12}$ and $\hat{\Gamma}_{11}$ are consistent (in norm) estimates of Γ_{12} and Γ_{11}.*

Lemma 2 Let *A be a self-adjoint Hilbert-Shcmidt operator with eigenvalues $\lambda_1 > \ldots > \lambda_k > 0$. If $\hat{A}$ is a consistent estimate of A, then its first k eigenvectors and eigenvalues are consistent estimates of the first k eigenvectors and eigenvectors of operator A.*

Lemma 3 *If vector e is estimated consistently by $\hat{e}$, and operator Γ is estimated consistently in norm by $\hat{\Gamma}$, then vector Γe is estimated consistently by $\hat{\Gamma}\hat{e}$.*

Lemma 4 *If a linear operator, A, has finite rank and is given by its action on a basis as $A : e_i \mapsto \lambda_i f_i$, and if $\hat{\lambda}_i, \hat{e}_i, \hat{f}_i$ are consistent estimates for λ_i, e_i, f_i, then operator $\hat{A}$ given by action $\hat{A} : \hat{e}_i \mapsto \hat{\lambda}_i \hat{f}$, is a consistent in norm estimate of A.*

Given the estimate of the operator ρ, we can compute predictions as $\hat{\rho} f_t$.

5. Data

We use daily settlement data on the Eurodollar that we obtained from Commodity Research Bureau. The Eurodollar futures are traded on Chicago Mercantile Exchange. Each contract is an obligation to deliver a 3-month deposit of $1,000,000 in a bank account outside of the United States at a specified time. The available contracts has delivery dates that starts from several first months after the current date and then go each quarter up to 10 years into the future. The settlement is in cash.

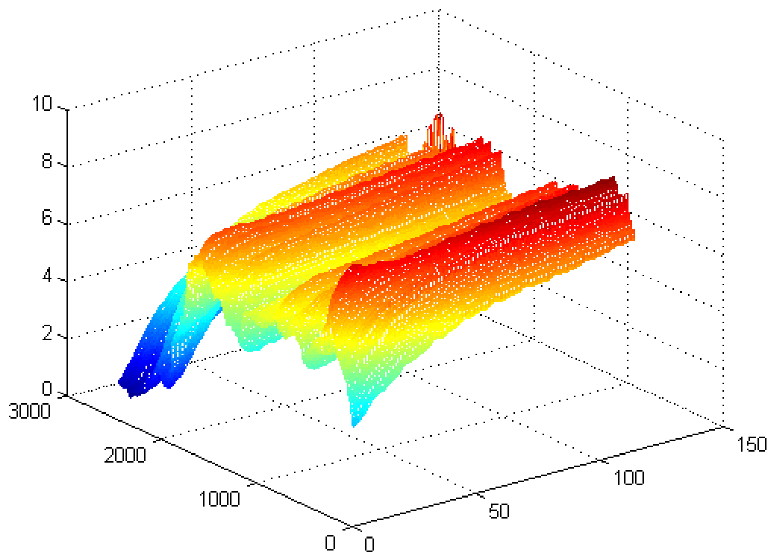
The available data start in 1982, however, we use only the data starting in 1994 when the trading on 10-year contract appeared. We interpolated available datapoints by cubic splines to obtain smooth forward rate curves. We restricted the curve to points that are 30 days from each other to speed up the estimation. We also removed datapoints with less than 90 or more than 3480 days to expirations. That left us with 114 points per curve and 2507 valid dates.

The main difference of futures contract from the forward contract is that it settled during the entire life of the contract, while in the forward contract the payment is made only at the settlement date. This difference and variability of short-term interest rates make the values of the contracts different (see Hull for an explanation). While the difference in values is small for short maturities, it can be significant for long maturities. There exists methods to adjust for this difference but for our illustrative purposes we will simply ignore this difference.

The rate on the forward contract is approximately the forward rate that we defined above. Indeed, the buyer of the contract expects to have a negative cash flow (forward price) on the settlement date and a positive cash flow ($100,000) 3 months after the settlement date. He has the following alternative investment: he buys a discount bond that will pay $100,000 three months after the settlement date. This costs $\$100{,}000 \times P_t(T+\delta)$, where δ denotes 3 months. He complements this by selling a discount bond that matures on the settlement day. If his overall investment is zero, then he is sure that on the settlement day he can make a payment of at least $\$100{,}000 \times P_t(T+\delta)/P_t(T)$. By arbitrage considerations we see that the price of the forward contract is $100,000 times the forward rate.

The next chart illustrate the evolution of Eurodollar forward rate curves.

Figure 1. Forward Curve Evolution.



Note: The forward curves correspond to Eurodollar rates from January 1994 to December 2003. The calendar time on the left axis and the time to maturity is on the right axis. Both are measured in working days.

6. Principal Components

Principal component factors

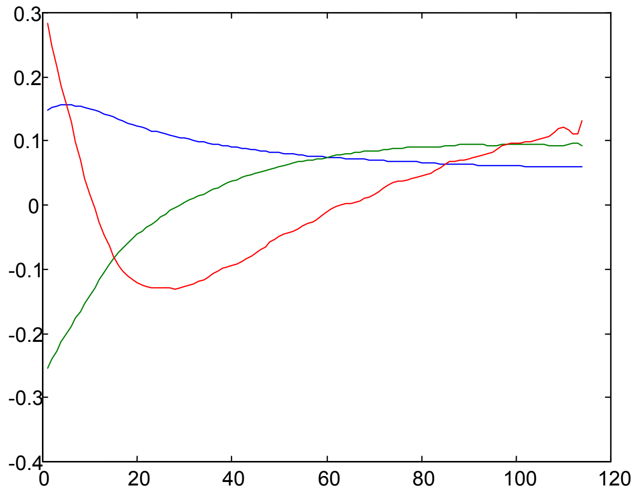
Table 1. Eigenvalues of the forward rates variance matrix.

Eigenvalue
113.02
10.96
0.815
0.1253
0.0648
0.024

Note: The estimates are from the daily Eurodollar forward rates data.

The results of estimation suggest that the rank of operator ρ is around 3. This corresponds well to the findings in the previous empirical literature that suggest that the evolution of the term structure can be decomposed into evolution of 3 main factors: level, slope and curvature. The next Chart shows the eigenvectors of the covariance matrix of forward rates that corresponds to the highest eigenvalues of the operator.

Figure 2. Eigenvectors of the variance matrix.



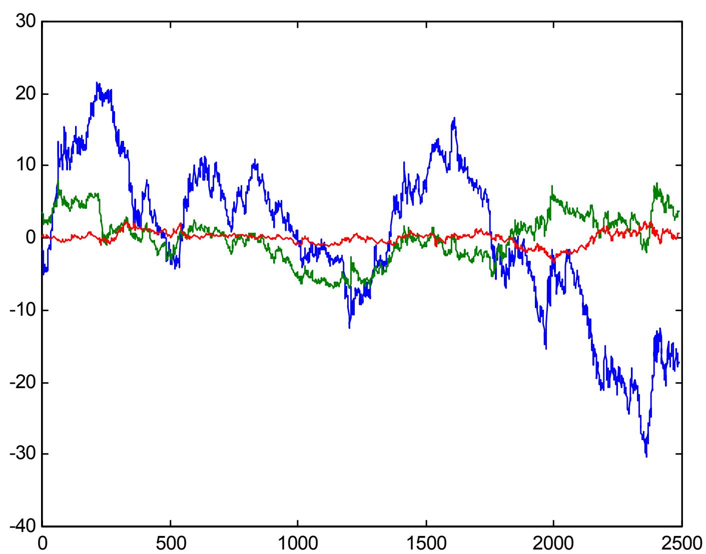
Note: The eigenvectors are estimated using the daily data on the Eurodollar forward rates.

Clearly the eigenvectors correspond to the usual factors of the term structure dynamics. The blue line corresponds to the "level" factor. It has the largest eigenvalue. The green line is the "slope" factor. It has the second largest eigenvalue. And the red line is the "curvature" factor.

Dynamics of principal component factor loadings

The picture below illustrates how loadings on the principal component factors changed over time.

Figure 3. Time Evolution of Factor Loadings



Note: On the horizontal axis the calendar time in working days since January 1994. On the vertical line is the value of the factor loading. The blue line is for the loading on the "level", the green is for the "slope", and the red is for the "curvature". The principal component curves were estimated using the entire data period from January 1994 till December 2003.

It is clear from this picture that the “level” is the most important contributor to the interest rates, followed by the “slope” and the “curvature”. Some mean reversion in the dynamics is also apparent.

Estimates of coefficient operator

Operator ρ maps the term structure curve to a linear combination of the factor curves. The action of operator $\rho - I$ on the factor curves themselves is given by the matrix in Table 2:

Table 2. Estimate of operator $\rho - I$ in the basis of eigenvectors.

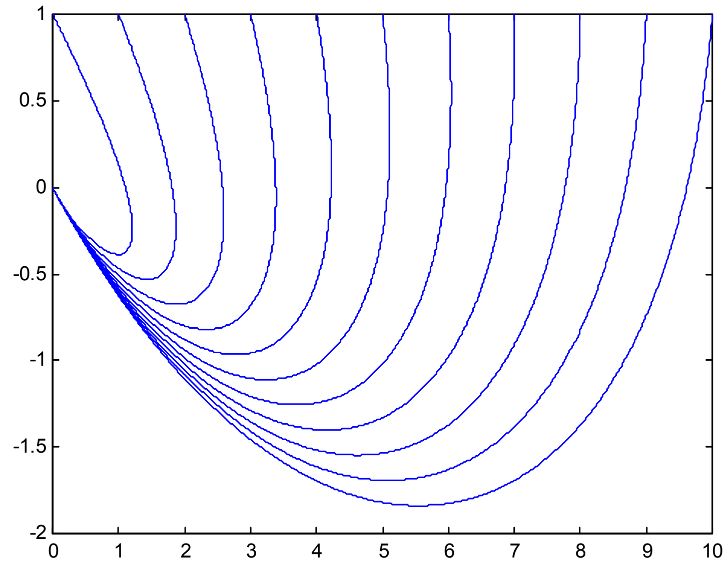
-0.21	0.1	1.48
-0.12	-0.43	-0.77
0.00	-0.05	-0.74

Note: The coefficients are estimated from the daily data and all coefficients are multiplied by 100 for convenience.

What is surprising about this matrix is that it is non-symmetric and that the off-diagonal elements are quite large compared with diagonal elements. This observation suggests two possible explanations. The first one is that the dynamic of the model is more complex than the simple mean reversion. The second one is that the estimates of the off-diagonal elements are not sufficiently precise. First, we address the issue of dynamics.

Operator ρ have 1 real and 2 complex eigenvalues. All of them are less than 1 in absolute value so the operator is stable in the sense that it corresponds to a stable dynamical system: the deviation from the mean tends to the zero eventually The complexity of the dynamics can be seen from the following Figures.

Figure 4. Dynamics of factor loadings.

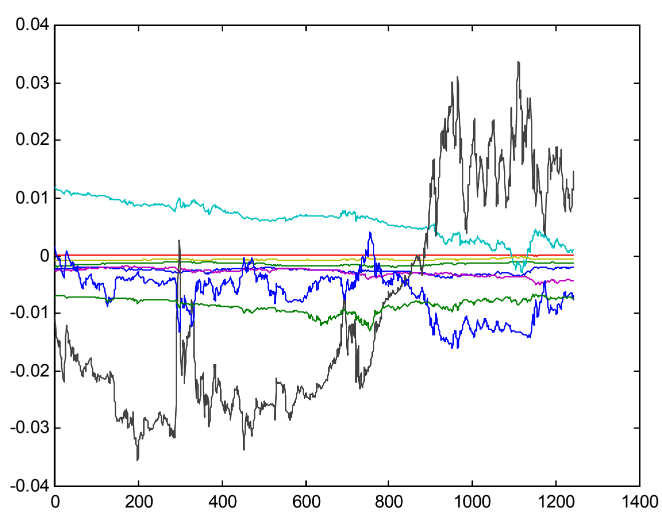


Note: The horizontal axis shows the loading on the "level" factor. The vertical axis shows the loading on the "slope" factor. The loading on the curvature factor is not shown. Its initial value was set equal to 1.

This picture shows evolution of the factor loadings with different initial values in the absence of external noise, so it illustrates impulse response function for a particular impulse. While the loadings are seen to converge to zero eventually, the convergence is not monotonic. The "level" initially increases before converging to zero, and the slope at some point becomes negative.

We can address the concern about the precision of the estimates by plotting the results of the estimation as the number of data increases:

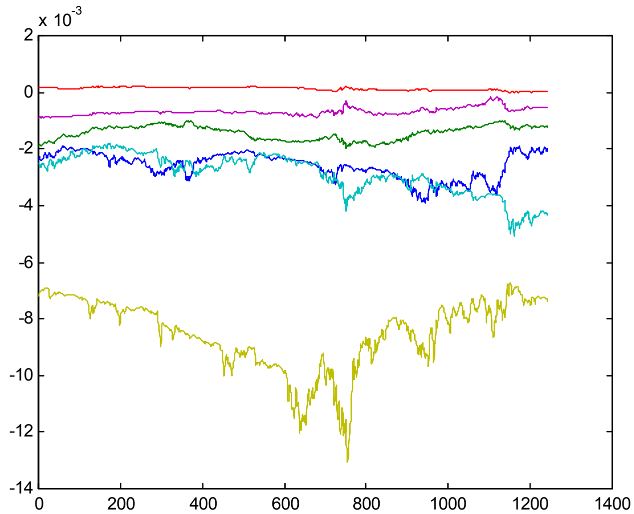
Figure 5. The evolution of matrix entries of the estimate of operator ρ.



Note: The operator ρ is estimated using the daily data on the Eurodollar forward rates. The estimation is on the rolling basis so it uses all the information available at the time of estimation.

This chart suggest that certain of the entries in the coefficient operator are indeed unstable over time. It turns out that these are all entries that are over the main diagonal. The entries on the main diagonal and below are sufficiently stable.

Figure 6. The evolution of lower-diagonal matrix entries of the estimate of operator ρ.



Note: The operator ρ is estimated using the daily data on the Eurodollar forward rates.

Forecasting

An important characteristic of a model is its forecasting performance out of sample. We compare the performance of our model with that of the simple random walk model and with the performance of the model by Diebold-Li. The following charts plot the performance of these models for different forecasting ranges. The performance in these charts is measured by the realized mean squared error (MSE) of the prediction.

7. Estimation by Singular Factors

Singular values and singular factors

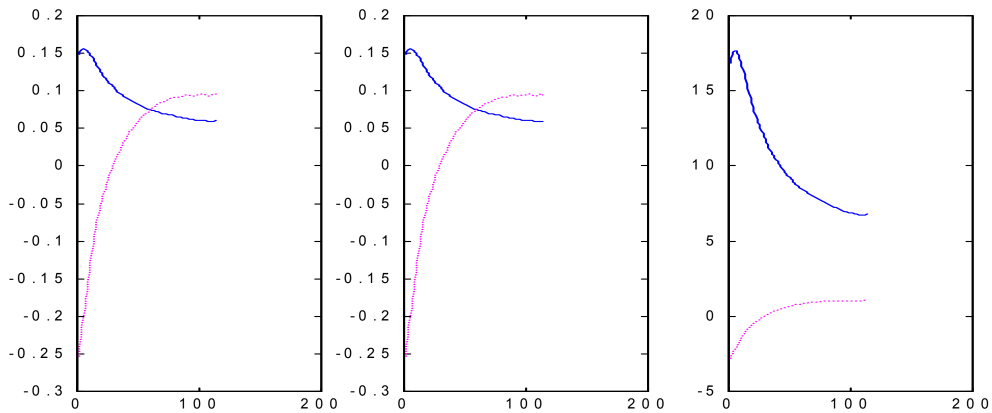
Table 3 Singular values for different time intervals and models

C daily	C yearly	B Daily	B Yearly	A Daily	A Yearly
0.0518	10203	0.0022	22242	12719	746.3225
0.0007	135	0.0001	144	119	30.4738
0.0001	0	0	0	1	0.0378

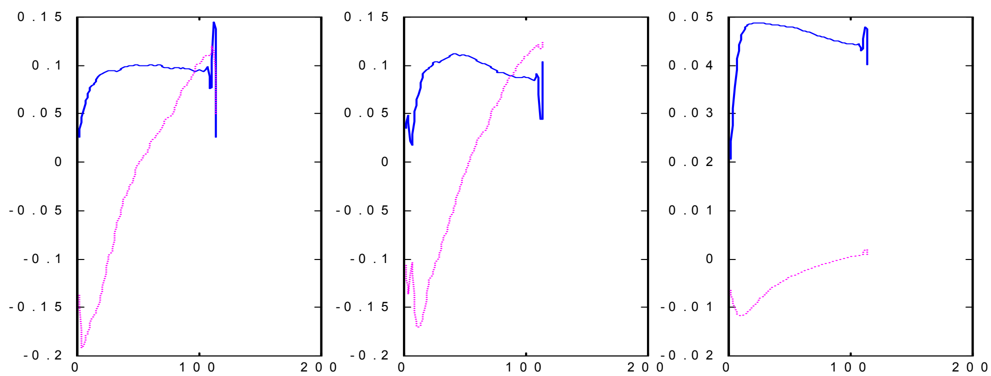
The next figure shows the two largest singular factor curves R, corresponding factor loading curves F and curves G for models A, B and C estimated on daily data.

Figure 7 Singular factors and loadings.

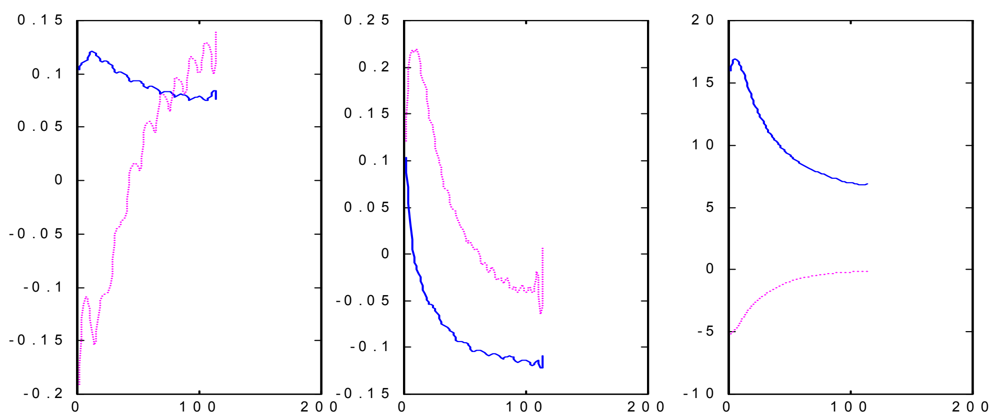
Model A



Model B



Model C



Note: The singular factors and loadings are estimated the daily data on the Eurodollar forward rates. The left panel shows singular factors, the middle panel shows the corresponding factor loadings, and the right panel shows functions G . The solid blue line is for the fist singular factor and the dotted magenta line is for the second singular factor.

Relation to Principal Components

Estimated Dynamics

Performance in Predictions

8. Conclusion

References:

A. Ang and M. Piazzesi (2003) "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables" *Journal of Monetary Economics*, **50**, 745-787

T. W. Anderson (1984) *An Introduction to Multivariate Statistical Analysis,* 2nd edition, John Wiley and Sons

D. Bosq (2000) *Linear Processes in Function Spaces: Theory and Applications,* Springer-Verlag

J. H. Cochrane and M. Piazzesi (2002) "Bond Risk Premia" *NBER Working paper* 9178

J. C. Cox, J. E. Ingersol, and S. A. Ross (1985) "A Theory of the Term Structure of the Interest Rates" *Econometrica*, **53**, 385-408.

G. Da Prato and J. Zabczyk (1992) *Stochastic equations in infinite dimensions,* Cambridge University Press,

F. X. Diebold and C. Li (2002) "Forecasting the Term Structure of Government Bond Yields" Working Paper (available at http://www.ssc.upenn.edu/~diebold)

G. R. Duffee (2002) "Term Premia and Interest Rate Forecasts in Affine Models" *Journal of Finance*, **57**, 405-443

D. Duffie and R. Kan (1996) "A Yield-Factor model of Interest Rates" *Mathematical Finance*, **6**, 379-406.

R. S. Goldstein (2000) "The Term Structure of Interest Rates as a Random Field" *Review of Financial Studies*, **13**, 365-384

D. Heath, R. Jarrow, and A. Morton (1990) "Bond Pricing and the Term Structure of Interest Rates: A Discrete Time Approximation" *The Journal of Financial and Quantitative Analysis*, **25**, 419-440.

D. Heath, R. Jarrow, and A. Morton (1992) "Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation" *Econometrica*, **60**, 77-105.

T. S. Y. Ho and S.-B. Lee (1986) "Term Structure Movements and Pricing Interest Rate Contingent Claims" *Journal of Finance*, **41**, 1011-1029

H. Hotelling (1936) "Relations between two sets of variates" *Biometrika*, **28**, 321-377

J. Hull and A. White (1990) “Pricing Interest Rate Derivative Securities” *Review of Financial Studies*, **3**, 573-592

R. A. Jarrow (2002) *Modeling Fixed-Income Securities and Interest Rate Options,* 2nd edition, Stanford University Press

D. P. Kennedy (1997) “Characterizing Gaussian Models of the Term Structure of Interest Rates” *Mathematical Finance*, **7**, 107-116.

S. E. Leurgans, R. A. Moyeed, and B. W. Silverman (1993) “Canonical Correlation Analysis when Data are Curves” *Journal of Royal Statistical Society B*, **55**, 725-740

M. Piazzesi (2003) “Bond Yields and the Federal Reserve” *Working paper*

P. Santa-Clara and D. Sornette (2001) “The Dynamics of the Forward Rate Curve with Stochastic String Shocks” *Review of Financial Studies*, **14**, 149-185

O. A. Vasicek (1977) “An Equilibrium Characterization of the Term Structure” *Journal of Financial Economics*, **5**, 177-188