A Note on the Relation Between Principal Components and Dynamic Factors in Affine Term Structure Models[*]

Caio Ibsen Rodrigues de Almeida[**]

Abstract

In econometric applications of the term structure, affine models are among the most used ones. Nevertheless, even presenting a closed form characteristic function, its estimation procedure still presents many points to be understood and difficulties to be removed. In this note, we address one of these points. Suppose we estimate an affine dynamic term structure model, and also apply principal component analysis to the interest rate database available. A very plausible question would inquire about the relation (if any) between the principal components obtained assuming no dynamic restrictions, and the dynamic factors estimated using the proposed term structure model. We answer this question when estimating a standard affine model using zero coupon data. We show that each principal component can be approximated by a linear transformation of the dynamic factors. Although simple, this is an important step to the understanding of the mechanics of dynamic affine term structure models. A numerical example using U.S. zero data illustrates the result.

Keywords: Term Structure of Interest Rates, Dynamic Affine Models, Principal Component Analysis, Maximum Likelihood.

JEL Codes: C5, C51.

1. Introduction

The term structure of interest rates is a fundamental element for the whole economy. It informs, for different maturities, the cost of borrowing money, being directly related to macroeconomic variables and central bank decisions. In addition, there is an enormous number of term structure related securities, making it an extremely important variable for market participants. For instance, in the U.S. market, some of the possible fixed income instruments traded are: Treasury zero coupon and coupon-bearing bonds, corporate bonds, mortgage-backed securities, swaps, FRAs, caps, floors and swaptions (for a mathematical description of these instruments see Brigo and Mercurio (2001)). These are some of the reasons why both academics and practitioners demonstrate enormous interest in understanding the sources that drive the term structure movements.

Affine term structure models have been intensively used to model the evolution of the term structure over time. In Vasicek (1977) seminal work, based in no-arbitrage conditions, he proposes a Gaussian model for the short term rate dynamics. Cox et al. (1985) proposed an equilibrium model and derived the short term rate dynamics as being a square root process, which is also one of the basic examples of affine processes. Some years later, empirical financial econometrists proposed and estimated many multi-factor versions of the basic Vasicek, CIR and combinations of these models, with the purpose of explaining stylized facts supported by data: Chen and Scott (1993) and Pearson and Sun (1994) estimate Multi-factor CIR models by the direct Maximum Likelihood method; Pennacchi (1991), interested in explaining the joint dynamics of interest rates and inflation, estimates a multi-factor Gaussian model by Kalman filtering methods; Dai and Singleton (2000) and Dai and Singleton (2002) classify the family of affine models, and estimate general three-factor affine models using Generalized Methods of Moments and Approximated Maximum Likelihood Estimation. DS (2002) show, through empirical implementation, that affine models are compatible with a time-varying risk premium and,as a result, can capture the failure of the expectation hypothesis (EH);[1] Duffie et al. (2003b) estimate a mixed Gaussian/CIR model to explain returns on Russian Brady bonds which are subject to default risk; and Duarte (2004) implements a multi-factor CIR model with a general market price of risk, using a combination of filtering techniques and Quasi-Maximum Likelihood, with the intention of simultaneously explaining the first and second moments of yields on U.S. swaps and treasury bonds. All these are dynamic models which present predetermined stochastic differential equations describing the dynamics of the factors driving the term structure movements, and which, in addition, impose restrictions that rule out arbitrages in the market: There should exist a risk-neutral

---

[1]The failure of the EH has been largely documented in the literature. Some examples are: Fama and Bliss (1987), Campbell and Shiller (1991) and Backus et al. (2001).

measure $\mathcal{Q}$, under which discounted bond prices are martingales.

On the other hand, principal component analysis (PCA) (Flury, 1988) has been traditionally applied to term structures of interest rates to identify the main factors driving the term structure movements. Since Litterman and Scheinkman (1991) found that essentially three factors were enough to describe the movements of the U.S. treasury term structure, PCA has been applied to many problems in financial engineering: risk management as in Singh (1997), portfolio immunization as in Barber and Copper (1996), identification of main driving forces of term structures as in Heidari and Wu (2003), Collin and Goldstein (2002), and Almeida et al. (2003), besides being a benchmark used to define the number of factors in dynamic models. Whenever PCA is applied to yield levels,[2] principal components inherit the qualitative characteristics of yields, including autoregressive behavior with near-unit roots (see Backus et al. (1999) or Diebold and Li (2003) for a discussion on stylized facts of the term structure of interest rates). One justification given by authors who apply PCA to the level of the term structure instead of to its first difference or percent return, is that in the level case, factors are significantly more persistent than errors, making their identification more precise (see Heidari and Wu (2003), page 77). In particular, by contrasting PCA with dynamic term structure models, no dynamic restriction to rule out arbitrages is imposed when estimating the time series of the principal components, which can be considered a relaxation of what is done in dynamic term structure models. Then a natural question arises: is there any clear relation between principal components (unrestricted estimated factors) and dynamic factors (restricted)?

In this note, we answer this question when the considered dynamic model is affine. We show that the significant principal components of the term structure can be well approximated by a linear transformation of the dynamic factors obtained in affine models. This result proves to be interesting and practical, at least for two reasons. First, it helps in the understanding of the operational mechanics of affine models. For instance, consider the case where we compare the standard Maximum Likelihood methodology[3] to the new methodology proposed in Collin et al. (2003). Collin et al. (2003) first calculate the principal components of the term structure, and then they estimate three-factor dynamic affine models which match without errors the first two or three principal components.[4] In terms of fitting principal components, how can we compare this methodology to the standard one? The

---

[2]See, for instance, Dai and Singleton (2000) and Heidari and Wu (2003). In addition, Dai and Singleton (2002) **qualitatively** relate, for affine models, dynamic factors and principal components, when PCA was applied directly to yield levels.

[3]It assumes that a specific set of yields, in a number equal to the number of factors, is measured without error. Chen and Scott (1993) were the first to apply it.

[4]In the estimation process, when they match two principal components exactly, more weight is given to the dynamics of yields, while when they match three principal components, more weight is given to the cross-sectional fitting of yields.

result in this note shows that when we estimate an affine process by the standard methodology we are implicitly fitting the significant principal components with error. Second, the results in this note are useful to inform differences and similarities of dynamic affine models and simple PCA methods in the daily procedures of financial institutions, such as risk management and portfolio optimization. We can, for instance, propose upper bounds to the difference obtained for the risk measured[5] when in one case the principal components are used to generate the portfolio probability density, while in the other case, the dynamic factors are used. An important variable to be considered in this context is the order of the errors in the approximation of the principal components by dynamic factors.[6]

The note is organized as follows. Section 2 briefly describes affine models and the estimation method used. Section 3 presents the general idea of the approximation proposed (for any parametric dynamic term structure model), and then introduces two different linear approximations of the principal components by dynamic factors. Section 4 presents an empirical example where we estimate a three-factor Gaussian model for the U.S. treasury term structure and implement the linear approximations proposed in section 3. Section 5 concludes.

2.   The Affine Mechanism

By precluding market arbitrages, assume the existence of an Equivalent Martingale Measure $\mathcal{Q}$ under which bond prices discounted by the money market account are martingales (see Duffie (2001) for details). Let Y denote the N-dimensional state space vector characterizing the probabilistic uncertainty of the term structure of interest rates. Affine models are the ones whose short rate process is an affine function of the state vector, $r_t = \rho_0 + \rho_1 Y_t$, and the risk-neutral dynamics ($\mathcal{Q}$ dynamics) of the state vector also presents its drift and covariance matrix written as affine functions of the state vector:

$$dY_t = \kappa^{\mathcal{Q}}(\theta^{\mathcal{Q}} - Y_t)dt + \Sigma\sqrt{S_t}dW_t^* \tag{1}$$

where W_t^* is an N-dimensional Brownian Motion under $\mathcal{Q}$, $\kappa^{\mathcal{Q}}$ is the $N \times N$ mean reversion matrix, $\theta^{\mathcal{Q}}$ is an $N \times 1$ vector representing long-run mean, and S is a diagonal matrix with $S_{ii} = \alpha_i + \beta_i^T Y_t$, $1 \leq i \leq N$, $\alpha_i \in \mathcal{R}$, $\beta_i \in \mathcal{R}^N$.[7] Duffie and Kan (1996) showed that the time t price of a zero coupon bond with maturity at T is given by:

$$P(t,T) = e^{\gamma(\tau)+\delta(\tau)'Y_t} \tag{2}$$

[5]Where we assume risk measured by the Value at Risk method (Jorion, 2000).
[6]We leave to the future, more to be explored in terms of the relation between principal components, dynamic factors, and applications in the financial market.
[7]Let α and β respectively denote the row vector containing all αs and the matrix whose i_{th} row is β_i^T.

where Y_t is the state vector at time t, $\tau = T - t$, and $\gamma(.)$ and $\delta(.)$ solve the following ODEs:

$$\delta'(t) = \rho_1 - \kappa^{\mathcal{Q}}\delta(t) - \tfrac{1}{2}\delta(t)^T H_1 \delta(t), \quad \delta(T) = 0; \tag{3}$$

$$\gamma'(t) = \rho_0 - \kappa^{\mathcal{Q}}\theta^{\mathcal{Q}}\delta(t) - \tfrac{1}{2}\delta(t)^T H_0 \delta(t), \quad \gamma(T) = 0; \tag{4}$$

where $\Sigma S_t \Sigma^T = H_0 + H_1 Y_t$.

We can directly see that the yield of these zero coupon bonds will be given by:

$$R(t,T) = -\frac{\gamma(\tau)}{\tau} - \frac{\delta(\tau)'}{\tau}Y_t = A(\tau) + B(\tau)'Y_t \tag{5}$$

A general theoretical description of the affine family is beyond the scope of this note and is given in Duffie et al. (2003a).

2.1 What happens under the physical measure? The importance of risk premia

So far we have only mentioned the behavior of affine processes under the risk-neutral measure $\mathcal{Q}$. On the other hand, $\mathcal{Q}$ should be seen just as a mathematical instrument used to price derivatives while the "real world" movements happen under the original (or physical) measure which we denote by P. Then, in order to propose a full model we must specify a parametric form for the risk premia demanded by investors who are holding "risky" bonds in the real world, where the risk comes from the uncertainty about the interest rates. Risk premia are completely captured by the market price of risk[8] Λ, which is the volatility of the state-price deflator.[9] Λ can also be seen as the process which defines the Radon-Nykodin derivative that directly relates the physical measure P to the risk-neutral measure $\mathcal{Q}$. The practical aspect which attracts our interest here is how the Brownian Motion vectors under both measures are related:

$$W_t^* = W_t + \int_0^t \Lambda_s ds \tag{6}$$

where W_t is an N-dimensional Brownian Motion under P.

In general, econometrists have preferred to restrict the parametric form of Λ so as to maintain the P-dynamics of the state vector also affine. In general, that does not need to be true for an affine model (see Duarte (2004) for an example

---

[8]Its interpretation is that, for a fixed maturity, it represents units of excess return earned per unit of risk for a zero coupon bond with that maturity.

[9]The state price deflator (spd) is a positive process with the property that the price of any market instrument deflated by spd is a martingale under the physical measure P.

of semi-affine dynamics for the state vector under P). The major advantage of maintaining the P-dynamics affine is the availability of nice approximations for transition densities of affine processes due to their closed form characteristic function (as in Duffie et al. (2003b) and Singleton (2001)). In an attempt to maintain the affine dynamics under P, the first available affine models of Vasicek (1977), CIR (1985) and variations, proposed the market price of risk as a multiple of volatility, $\Lambda_t = \sqrt{S_t}\lambda$, where λ is an $N \times 1$ vector, and S_t is defined above. Then from equations (1) and (6) we see that the drift under P would depend on the following extra term $\Sigma S_t \lambda$, which is also affine in Y_t, maintaining the affine structure under P. However, the parameterization of the risk premia as a multiple of volatility turned to be a problem when trying to match the empirical characteristics of expected bond returns, basically because risk premia were not allowed to change signs. More recently, Duffee (2002) has overcome this problem by proposing a more general parameterization for the market price of risk. It divides prices of risk into two sets: prices of risk for factors that drive the instantaneous volatility S_t, and for factors that do not drive volatility. For factors that do not drive stochastic volatility of the state vector, Duffee's parameterization allows dependence of their prices of risk on the whole state vector Y. Nonetheless, it still maintains the previous rigid dependence of the prices of risk of volatility factors only in the instantaneous volatility S_t itself. In this note, we adopt Duffee's parameterization:

$$\Lambda_t = \sqrt{S_t}\lambda_0 + \sqrt{S_t^-}\lambda_Y Y_t, \qquad (7)$$

where λ_0 is an $N \times 1$ vector, λ_Y is an $N \times N$ matrix, and S_t^- is a diagonal matrix defined by:

$$S_t^{ii-} = \left\{ \begin{array}{ll} \frac{1}{S_t^{ii}}, & \text{if } \inf(\alpha_i + \beta_i^t Y_t) > 0. \\ 0, & \text{otherwise.} \end{array} \right\} \qquad (8)$$

Note that with the extra term in the parameterization, risk premia can depend on more general linear combinations of elements in Y, namely $\Sigma \lambda_Y Y_t$ and not on simple combinations which model the instantaneous volatility $\Sigma S_t \lambda_0$.

2.2   Model estimation: maximum likelihood

In theory, according to equation (5), conditioned on knowing the parameters of the affine model, the vector of observed yields and the state vector are related through a linear transformation. In practice, however, we usually observe more yields than the number of dynamic factors that we propose and implement in the model. Usually, we have a two- or three-dimensional state vector (dynamic factors) and observe at least six yields. That is the usual number of yields adopted in academic work using the U.S. term structure (see for instance, Duffie and Singleton (1997), Duffee (2002), and Duarte (2004). In contrast, Almeida (2004a) fits a

three-factor Gaussian model to the Brazilian swap term structure, using a set of eight observed yields). It is clear then that at least one of the yields should be measured with error, and more generally, we should make assumptions for the error structure of all yields. One possibility is to assume that all yields are measured with error. In this case, the state vector cannot be directly inverted from the observed yields, and filtering techniques are mandatory. We refer the reader to Pennacchi (1991), Lund (1997) and Duan and Simonato (1999) for term structure estimation using the Kalman filter. Another possibility is to assume that there is a one-to-one correspondence between the state vector and a subset of the observed yields, which are assumed to be measured exactly. Once we fix the parameters, the state vector can be inverted from that subset of yields by solving a linear system obtained through equation (5). The majority of empirical works done under this approach, restricted the market price of risk so that the state vector also presents affine dynamics under the physical measure P. Just as suggested in the previous subsection, approximation techniques for the likelihood function make the Maximum Likelihood estimator a natural choice as estimation process. The parameters are obtained by maximizing the likelihood of the observed yields, which is a function of the likelihood of the state vector. Empirical work using this approach include: Chen and Scott (1993), Pearson and Sun (1994), Duffie and Singleton (1997), and Duffie et al. (2003b), among others.

In this note, we assume that estimation is accomplished by the Maximum Likelihood method, precisely as described above. The important point to be noted is that once we obtain parameter estimates we are able to solve the ODE's (3) and (4) numerically and functions γ, δ will be readily available.[10] So, **after we obtain the model parameters, yields are a linear function of the state**. Moreover, we see that affine processes essentially do not differ from consistent parametric models of the term structure in the sense of Filipovic (1999) and Bjork and Christensen (1999) (see De-Rossi (2004) for an application). With fixed parameters, affine processes simply offer parametric functions A and B of the maturities, which relate the evolution of the term structure to the evolution of the state space Y in a consistent way where discounted bond prices are $\mathcal{Q}$-martingales.

For a detailed description of the implementation of the Maximum Likelihood Estimator for a multi-factor Gaussian model, see the appendix in Almeida (2004a). For details on the approximation of the transition density of affine processes with stochastic volatility, see appendix B in Duffie et al. (2003b).

---

[10]In particular, note that we only need the risk-neutral drift, the short rate and the volatility parameters represented by vector $\{\kappa^{\mathcal{Q}}, \theta^{\mathcal{Q}}, \rho_0, \rho_1, \Sigma, \alpha, \beta\}$ to solve the ODE's. In some particular cases, as in the Legendre Dynamic Model, we have a previous assessment of the ODE solutions without having specified the volatility parameters, which are free to be chosen to better fit the dynamics of the term structure under P (see (Almeida, 2004b)).

3.   Principal Components and Dynamic Factors

The relation between principal components and dynamic factors has been studied before this paper in at least two different contexts. First, Frachot et al. (1992) presented a theoretical description of this relation under an HJM model.[11] Second, Duffie and Kan (1996) presented a brief description of the relation between latent factors and observed yields under their multi-factor affine model. Although they have not mentioned principal components, they talk about factor rotations. However, highlighting the importance of this paper is the fact that Frachot et al. (1992) directly concentrate on HJM models and not on affine models, while Duffie and Kan (1996) do not present any kind of empirical analysis of their work.

Another interesting point is that relations between principal components and dynamic factors are not restricted to affine dynamic models.[12] For this reason, before getting to the details of the approximation presented in this paper, we propose a general description of the relation between principal components and dynamic factors in a general parametric model.

Later in this section, we show that, under the Maximum Likelihood estimation approach, the non-negligible principal components for the term structure can be approximately obtained by a specific linear transformation of the state vector.

3.1   Relation under a general parametric model

Suppose that we observe the yields of n zero coupon bonds with time to maturity $\tau_1, \tau_2, ..., \tau_n$, and that we intend to estimate a dynamic term structure model with k dynamic factors. In order to be able to directly invert the state vector from the observed data, we assume that a subset of k yields are measured without error, while the others are measured with i.i.d. zero-mean Gaussian errors. Suppose in addition that the parametric relation between the observed data and the state vector is given by the following equation:

$$R_t^{obs}(\tau_i) = f(\phi, Y_t, \tau_i) + \epsilon_t(\tau_i), i = 1, 2, ..., k. \tag{9}$$

where f is a generic invertible function which might be non-linear and which characterizes the measurement equation, $R_t^{obs}(\tau_i)$ represents the τ_i-maturity observed yield at time t, and ϵ_t is a vector of measurement errors which for all t is identically null in the positions of the yields measured exactly, while in the other positions, it presents i.i.d. Gaussian random variables.

The application of Principal Component Analysis to the historical set of observed yields reveals the following relation between yields and principal compo-

---

[11]I am grateful to an anonymous referee for this reference.
[12]I thank an anonymous referee for pointing me in this direction.

nents:

$$R_t^{obs} - E[R^{obs}] = \Psi pc_t \qquad (10)$$

where R_t^{obs} is a vector with all the time t observed yields stacked, Ψ is a matrix containing the eigenvectors of the covariance matrix of R^{obs}, and pc_t is a vector containing the scores of the principal components at time t.

By fixing the vector of parameters ϕ in equation (9), and using equation (10) to substitute vector R_t^{obs}, we obtain:

$$f(\phi, Y_t) + \epsilon_t = \Psi pc_t + E[R^{obs}] \qquad (11)$$

Extracting the vector pc_t from the last equation yields:

$$pc_t = \Psi'(f(\phi, Y_t) + \epsilon_t - E[R^{obs}]) = H + \Psi' f(\phi, Y_t) + \text{noise} \qquad (12)$$

where noise $= \Psi' \epsilon_t(\hat{\phi})$, $H = -\Psi' E[R^{obs}]$. If we neglect the noise term, we obtain the following approximation of the principal components as a function of the dynamic factors:

$$pc_t = H + \Psi' f(\phi, Y_t) \qquad (13)$$

The trick to obtain this nice theoretical relation consists in neglecting the noise vector. For these results to have any validity we have to give empirical evidence that confirms the relative small importance of the noise vector in the process. At this point, the affine models present two direct appeals. First, the function f is linear, giving a nice linear approximation of the principal components by dynamic factors. Second, we are able to provide evidence that the noise is negligible when the model is affine (as presented in the empirical section). The only limitation that prevents anyone from obtaining general empirical results, regarding these approximations for more general dynamic term structure models, is the effective implementation of such models.

3.1.1 Relation when the dynamic model is quadratic

Before concentrating on the affine model, it is interesting to sketch the approximation for a second particular case of dynamic term structure model: The quadratic term structure model. Longstaff (1989) and Beaglehole and Tenney (1991) were pioneers in the exploration of quadratic term structure models. In these models the term structure of interest rates is parameterized as a quadratic function of the state vector (for a complete theoretical characterization, see Leippold and Wu (2002); for practical applications, see Leippold and Wu (2003). One of its advantages is the capability to naturally generate positive interest rates, in

contrast to some affine models as, for instance, the multi-factor affine Gaussian models that present positive probability of attaining negative interest rates.

In Beaglehole and Tenney (1991), the model measurement equation is given by:[13]

$$R_t^{obs} = A(\phi) + B(\phi)'.Y_t + Y_t'C(\phi).Y_t + \epsilon_t \tag{14}$$

For this model, the particularization of equation (13) becomes:

$$pc_t = H + \Psi'(A(\phi) + B(\phi)'.Y_t + Y_t'C(\phi).Y_t) \tag{15}$$

which gives the principal components as a quadratic function of the latent variables, once the model has been estimated.

3.2 Relation when the dynamic model is affine

Let $A(\phi) = [A(\tau_1), A(\tau_2), ..., A(\tau_n)]'$ and $B(\phi) = [B(\tau_1)\ B(\tau_2)\ B(\tau_n)]$ be respectively an $n \times 1$ vector and $n \times n$ matrix coming from equation (5) applied to each maturity. The measurement equation for the affine model is given by:

$$R_t^{obs} = A(\phi) + B(\phi)'Y_t + \epsilon_t \tag{16}$$

where exactly as before, R_t^{obs} represents the vector of observed yields at time t, and ϵ_t is a vector of measurement errors which for all t is identically null in the positions of the yields measured exactly, while in the other positions, it presents i.i.d gaussian random variables.

In the estimation process, the optimizer chooses $\hat{\phi}$ so as to maximize the log-likelihood function, in equation (16). Once fixed $\hat{\phi}$, the estimated value for the parameter vector ϕ, $A(\hat{\phi})$ and $B(\hat{\phi})$ generate functions of maturities whose shape is specifically defined by vector $\hat{\phi}$. Moreover, the residuals in the term structure fit (estimated errors) can be directly obtained from:

$$R_t^{obs} = A(\hat{\phi}) + B(\hat{\phi})'Y_t + \hat{\epsilon}_t(\hat{\phi}) \tag{17}$$

3.3 First linear approximation: discarding the residuals of the model fitting procedure

At this point we can substitute equation (17) in (10) to obtain:

$$pc_t = \Psi'(A(\hat{\phi}) - E[R^{obs}] + B(\hat{\phi})'Y_t + \epsilon_t(\hat{\phi})) = H + GY_t + \text{noise} \tag{18}$$

---

[13]I kindly thank an anonymous referee for this example.

where noise $= \Psi' \epsilon_t(\hat{\phi})$, $H = \Psi'(A(\hat{\phi}) - E[R^{obs}])$ and $G = \Psi' B(\hat{\phi})'$. We obtain our first approximation by discarding the noise term thus getting:

$$\hat{pc}_t = H + GY_t \qquad (19)$$

We will see below that typical error fits $\epsilon_t(\hat{\phi})$ when estimating affine processes oscillate between 1% and 5% of the original yields. As the noise is obtained through a linear transformation of the error fits, where the linear transformation presents weights coming from normalized vectors (namely the eigenvectors of $cov(R^{obs})$), typical noise will have the same order of the error fit.
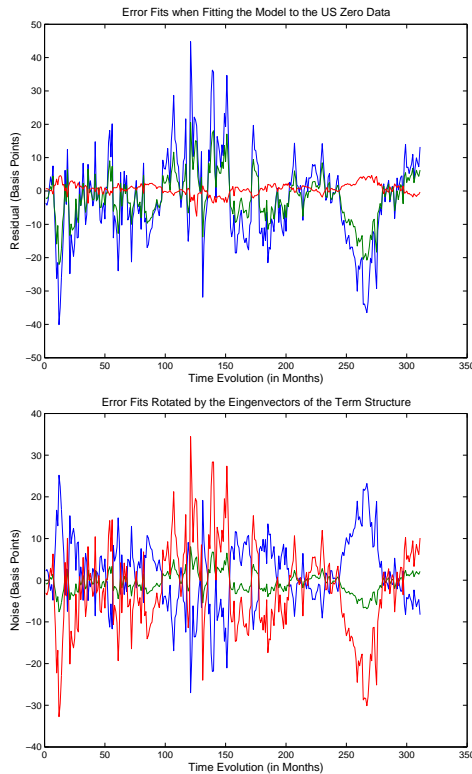

Figure 1
Comparing the magnitude of error fits and noise for the U.S. treasury term structure

Just as an informative illustration before the empirical section, figure 1 shows respectively the error fits and transformed error fits (noise) for the three-factor

Gaussian model estimated for the U.S. treasury term structure. Compare the order of the error fits and noise to the order of the yield values which appear in figure 2 to see that the percent range proposed for the error fits of 1-5% applies in this case.
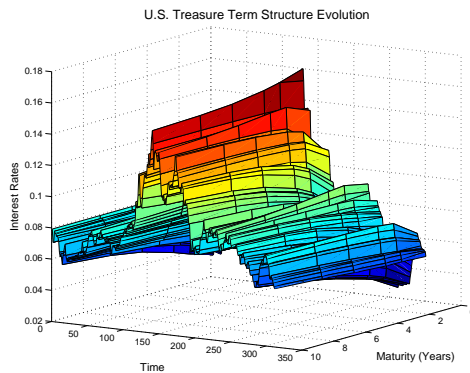

Figure 2
Temporal evolution if the U.S. term structure of interest rates

3.4 Second linear approximation: discarding non-significant principal components

Assume that although we observe n yields, there are only k significant principal components driving the movements of the term structure. In this subsection, we discard the remaining $n - k$ principal components and propose another way of relating principal components and dynamic factors.

Let τ_{exact} denote the maturities of the k bonds priced without error, and A_{exact} and B_{exact} respectively the coefficients for these maturities in equation (17). Using this equation, noting that residuals are zero by construction, we obtain:

$$R^{\tau_{\text{exact}}} = A_{\text{exact}} + B_{\text{exact}} Y_t \tag{20}$$

On the other hand, when we discard the negligible principal components,[14] in equation (10), the relation between the significant principal components and all

---

[14]Negligible here has the following qualitative meaning: we only keep principal components attached to eigenvalues of $Cov(R)$ that explain more than a predetermined fixed threshold of the term structure variance: 0.1% of the variance is used as the threshold in this work.

the observed yields is given by:

$$R_t^{obs} - E[R^{obs}] = \Psi^{sub} pc_t^{sub} + \theta_t \tag{21}$$

where Ψ^{sub} and pc^{sub} respectively represent the subsets of the first k eigenvectors and k first principal components of the term structure R, ordered by the importance in explaining its variance, and θ is an error term introduced by discarding the last $n - k$ principal components. In particular, note that to restrict equation (21) to the subset of yields measured without error, we only need to select the rows of Ψ^{sub} corresponding to those yields, and stack them in a new matrix, which we name Ψ_{exact}^{sub}. Rewriting equation (21) for the yields priced without error, we get:

$$R^{\tau_{exact}} - E[R^{\tau_{exact}}] = \Psi_{exact}^{sub} pc_t^{sub} + \theta_t(exact) \tag{22}$$

where $\theta(exact)$ represents the subset of the error vector θ related to the variables priced without error. Note that the error term is not zero for the subset of exact yields because it comes from discarding non-significant principal components and not from the model measurement equation (20), which presents zero error by construction.

Combining equations (20) and (22) we obtain:

$$pc_t^{sub} = (\Psi_{exact}^{sub})'(A_{exact} - E[R^{\tau_{exact}}] + B_{exact}'Y_t - \theta_t(exact)) = I + JY_t + \text{noise}_2 \tag{23}$$

where $\text{noise}_2 = -(\Psi_{exact}^{sub})'\theta_t(exact)$, $I = (\Psi_{exact}^{sub})'(A_{exact} - E[R^{\tau_{exact}}])$ and $J = (\Psi_{exact}^{sub})'B_{exact}'$.

Again, we discard the noise and obtain the approximation:

$$\hat{pc_t^{sub}} = I + JY_t \tag{24}$$

Here we artificially introduce errors when we assume that we are going to approximate the movements of the term structure by using fewer principal components than are available (in our applications, we discard three principal components and use three). This error might have a slightly different order from the error introduced by the model when pricing yields were measured with error. For this reason, testing both approximations might be worth.

4. Empirical Example

In this empirical exercise, data basically consist of the same database used in Dai and Singleton (2002): 312 monthly observations on U.S. treasury zero-coupon bond yields for maturities of 2, 3, 5, 7, and 10 years, together with the 6-month

LIBOR, covering the 1970-1995 period.[15] Figure 2 presents the U.S. term structure evolution. On the cross-sectional side, it is flat (a little inverted) throughout the sample period. From a time series perspective, it shows a lot of variation, achieving both its minimum and maximum values on the LIBOR maturity (six months): minimum in 1981 when the 6-month LIBOR was 2.88% and maximum in 1992 when it was 16.2%. Figure 3 presents the first three principal component loadings for this curve, when PCA is applied to the yield levels. Unsurprisingly, they respectively represent level, slope, and curvature factors. Together, they explain 99.99% of the term structure variation. The principal component related to the level explains 96.4%, the one related to the slope explains 3.4%, and the curvature factor explains merely 0.19% of the variation.
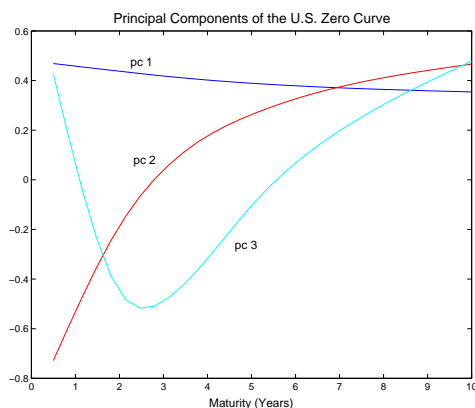

Figure 3
Loadings of the significant principal components for the U.S. term structure of interest rates

We estimate a three-factor Gaussian model using the Maximum Likelihood method, assuming that the 6-month LIBOR rate and zero coupon treasury rates with maturities of 5 and 10 years are priced without error, while the remaining zero coupon, with maturities of 2, 3, and 7 years are assumed to be priced with i.i.d. zero-mean Gaussian errors. The maximum value achieved by the log-likelihood function was of 37.46. Table 1 presents the estimated parameters as well as their standard errors, obtained by the Outer Product (BHHH) method. Note that all the parameters are significant at a 95% confidence level. Table 2 presents mean

---

[15]It is usual to construct the U.S. term structure of interest rates by a bootstrap procedure that simultaneously makes use of short-term LIBOR rates and U.S. treasury bond data (see Brigo and Mercurio (2001)).

and standard deviation of the residuals of the cross-sectional fits for the maturities assumed to be priced with error. Residuals present acceptable standard errors when compared to other empirical studies (for instance, see Duffie and Singleton (1997) or Dai and Singleton (2002)). Figure 4 presents, for each maturity whose yields were measured with error, the correspondent daily yield variations jointly with the daily residuals obtained by the dynamic model when fitting that yield. For maturities 2, 3, and 7 the ratio $\frac{\text{standard error of residual}}{\text{standard error of data in first difference}}$, was, respectively, 23.5%, 14% and 4%. Although we have applied PCA to the yield levels, these values give qualitative indication of the order of the error in the approximation of the principal components by a linear transformation of the state vector, if we had applied PCA to the first difference of yields. Figures 5 and 6 respectively present functions B and A which appear in equation (5) and directly express the relation between yields and state variables. Note that, in the nomenclature of Litterman and Scheinkman (1991), the dynamic factor Y_1, whose loadings are represented by function B_1, works as a level factor, while factors Y_2 and Y_3 with respective loadings B_2 and B_3 work as different slope factors.

Table 1

Parameters and standard errors for the $A_0(3)$ on the U.S. treasury term structure

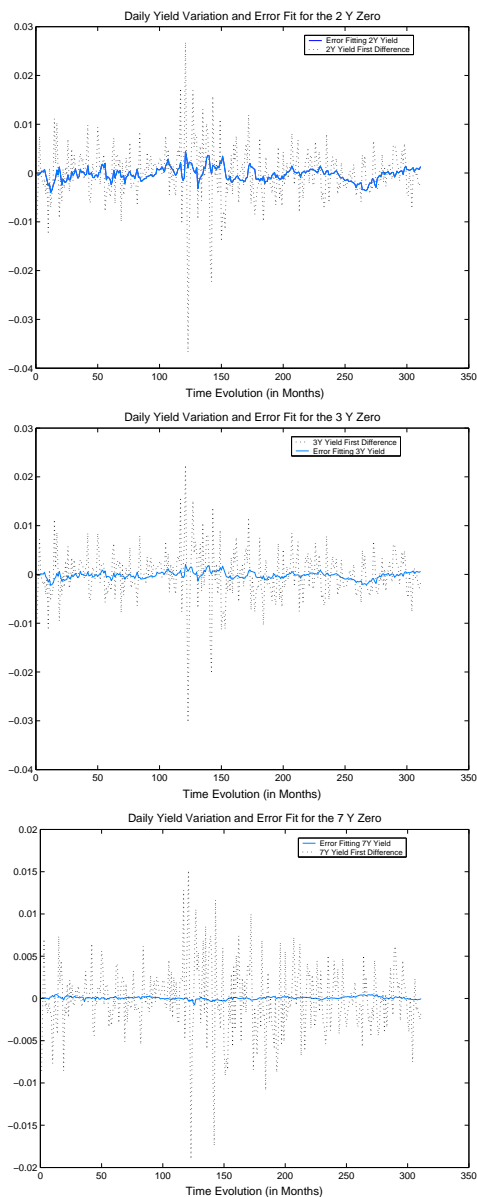| Parameter	Value	Standard error	Ratio $\frac{Value}{Std}$
κ_{11}	0.007295	0.002271	3.2
κ_{12}	-0.05655	0.01304	4.3
κ_{22}	0.6552	0.2435	2.7
κ_{33}	0.4684	0.01816	25.8
Σ_{11}	0.00487	0.0070	3.7
Σ_{22}	0.03867	0.0031	26.9
Σ_{33}	0.03361	0.0020	17.2
$\lambda_0(3)$	-1.411	0.3773	3.7
$\lambda_Y(3,1)$	-16.79	8.191	2.0
$\lambda_Y(1,2)$	25.67	5.72	4.5
$\lambda_Y(2,2)$	72.93	8.261	8.8
$\lambda_Y(2,3)$	5.113	2.309	2.2

Figure 4
Comparing the magnitude of yield variation to the error in fitting these yields
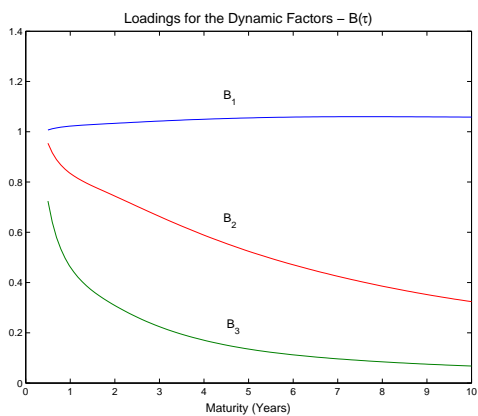
Figure 5
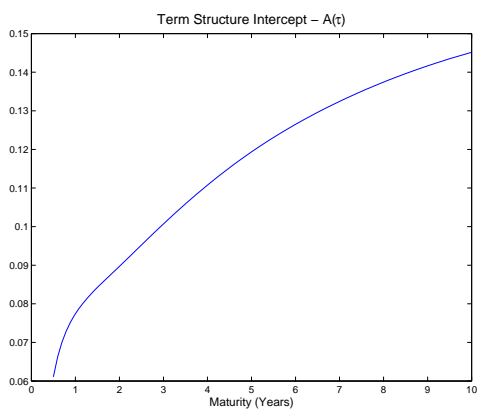Loadings of the affine model implied dynamic factors – function $B(\tau)$

Figure 6
Function $A(\tau)$ implied by the affine model

The approximations in writing the principal components as a linear transformation of the state vector work well for the 30 years of data from the U.S. zero curve. In order to see that, we first plot in figure 7 the time series of each principal component and the correspondent state variable[16] from the original state vector. Dashed lines represent state variables. We calculate correlation coefficients between the ith principal component and i^{th} state variable, and respectively obtained -0.884, 0.775 and 0.65. Looking at the pictures and also at these numbers we see that principal components and dynamic factors are very related to each other. Actually, we want to convince the reader that they are simply particular linear transformations of one another chosen by the optimization process when fixing model parameters to maximize the likelihood function, as proposed in section 3.

In order to empirically test the approximation from subsection 3.3, we plot in figure 8 each significant principal component (pc) together with its approximation, the linearly transformed variable obtained using equation (19), where dashed lines represent approximation. **Note how hard it is to distinguish pc from the approximated pc, for the first two principal components**. The correlation coefficients achieved between the first three principal components and their approximations were respectively 0.9999, 0.9998 and 0.9056. However, although we have a high correlation coefficient between the third principal component and its approximation, the approximation does not work so well for this principal component. The reason is simple: as this principal component explains only 0.19% of the variation in the term structure, its magnitude is not that different from the order of the residuals obtained from the yields measured with error. Then we see that although we cannot approximate the third principal component very well it does not play an important role in the dynamics of the term structure.[17] A clearer vision of the approximation of the first two principal components can be obtained in Figure 9 where we plot the relative error between the principal component and its approximation. For the first principal component, the approximation differs by more than 10% of its value only in 5.77% of the months. As for the second principal component, differences between the component and its approximation are bigger than 10% in 12.18% of the months. For any particular principal component, the biggest relative differences happen in months where its value is close to zero and the model fitting errors play a role.

---

[16]For pc_1 we plot it against minus the first state variable because they are negatively correlated.

[17]The eigenvalue associated with the third principal component indicates that the U.S. zero curve, for the period analyzed, could have had almost all its movements captured by a two-factor model.
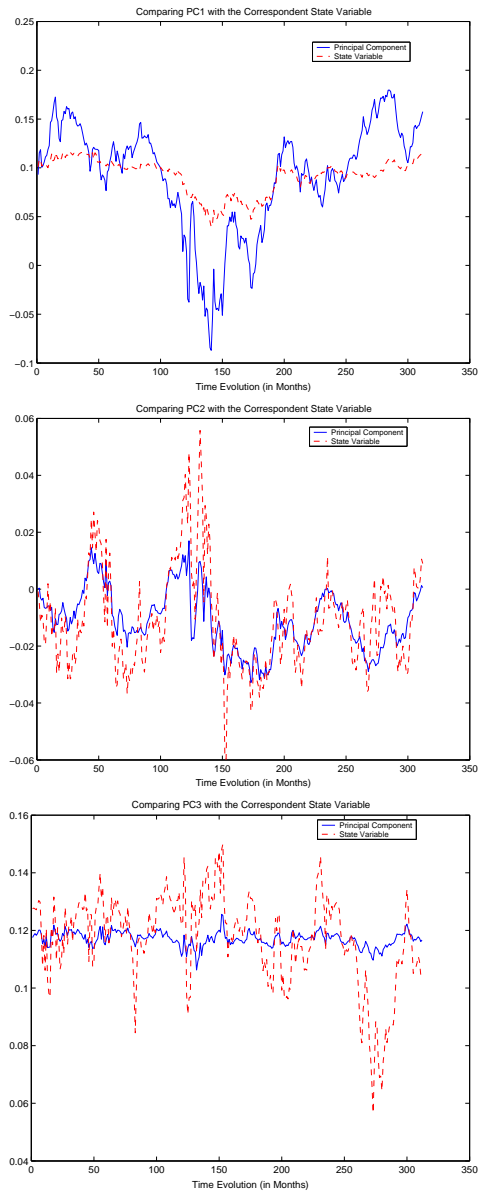
Figure 7
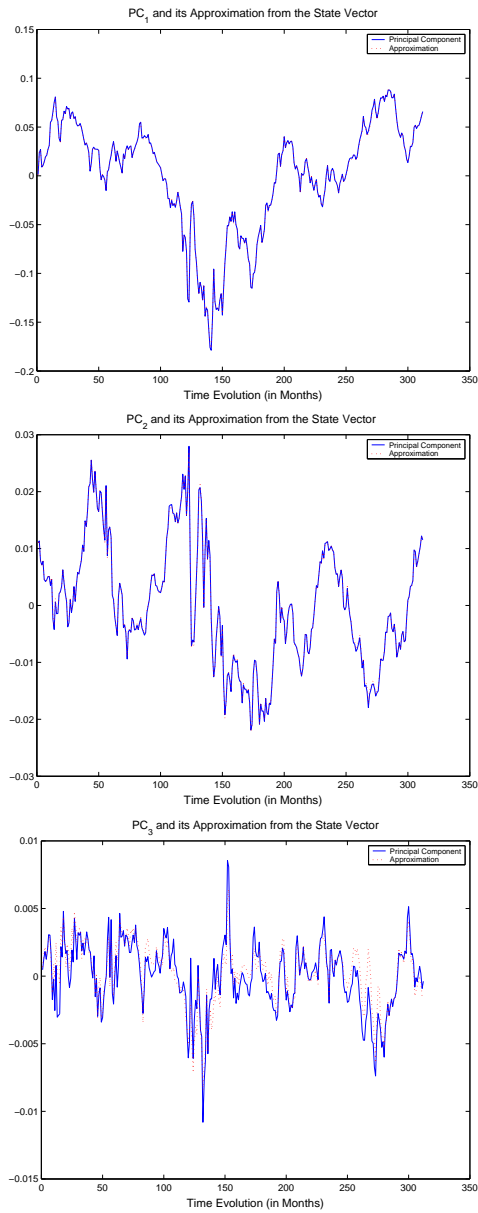Translated principal components and state variables

Figure 8
Translated principal components and linearly transformed state variables

For the approximation reported in subsection 3.4, the results were a little superior to the ones obtained using the approximation from subsection 3.3. Compare the relative error for the approximation of the first two principal components in figures 9 and 10, and note that errors present a slightly smaller magnitude in figure 10. In addition, correlation achieved between the first three principal components and their approximations were respectively 0.9999, 0.9998 and 0.9343, slightly outperforming the approximation for the third principal component from subsection 3.3. The reason for that is that when we discard principal components 4, 5 and 6 the order of the error θ in the approximation of the yield movements is of three basis points as can be seen in table 3, reasonably smaller than the 13 and 7 basis points standard errors of the residuals of the two and three year yields, obtained for ϵ, the dynamic model error discarded in the approximation of subsection 3.3.

Table 2

Statistics for the error fits of the dynamic affine model

Maturity (Years)	Mean (bp)	Std (bp)
2	2.5	13.0
3	2.0	7.0
7	0.5	2.0

Table 3

Statistics for the error θ in yields, obtained when discarding principal components 4, 5 and 6

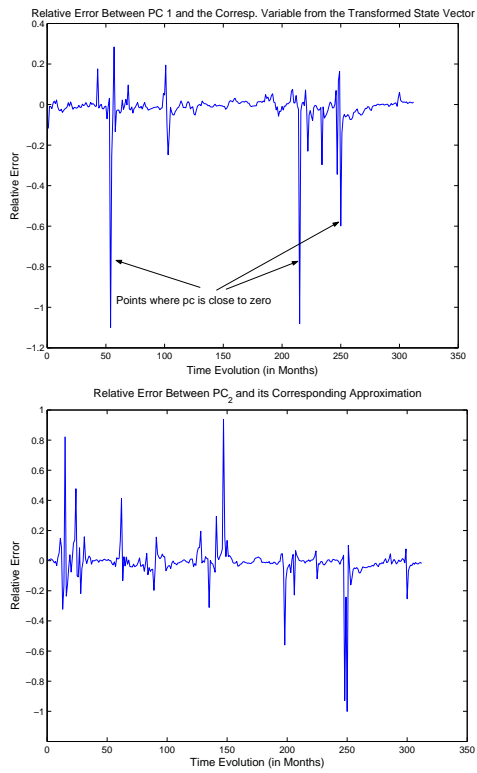| Maturity (Years)	Mean (bp)	Std (bp)
0.5	0	0.6
2	0	3.2
3	0	1.5
5	0	3.2
7	0	2.0
10	0	3.0

Figure 9
Relative error in the first linear approximation of the principal components

Relative Error Between PC$_1$ and Approximation When Discarding Some PCs

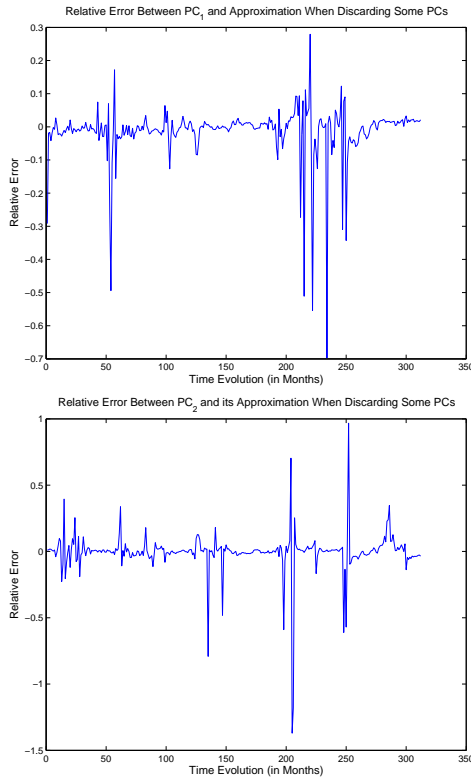Relative Error Between PC$_2$ and its Approximation When Discarding Some PCs

Figure 10
Relative error in the second linear approximation of the principal components.

5. Conclusion

In this note, we show that the linear structure embedded in dynamic affine term structure models directly translates into an approximation of the non-negligible principal components by linear transformations of the state vector. The smallest the model fitting errors, the better the approximation. A second approximation is also proposed and tested where we discard the non-significant principal components and use only the yields measured without error to linearly relate the significant principal components to the state vector. In spite of describing a small result, this note helps in the understanding of the operational structure of dynamic affine term structure models. Although such models have been intensively used by the empirical finance community due to their tractability, there are still many

details to be understood regarding their precise implementation. In a future exten-
sion, we intend to describe the role of unspanned stochastic volatility (see Collin
and Goldstein (2002) and Collin et al. (2003)) in breaking the linear dependence
between state vector and principal components, in dynamic affine models.

References

Almeida, C. I. R. (2004a). The dynamic legendre model: Separating the estimation
process between cross sectional fit and time series dynamics. Working Paper,
Department of Mathematics, Stanford University.

Almeida, C. I. R. (2004b). Time-varying risk premia in emerging markets: Expla-
nation by a multi-factor affine term structure model. *International Journal of
theoretical and Applied Finance*, 7(7):919–947.

Almeida, C. I. R., Duarte, A. M., & Fernandes, C. A. C. (2003). A generalization
of PCA for non-observable term structures. *International Journal of Theoretical
and Applied Finance*, 6(8):885–903.

Backus, D., Foresi, S., Mozumbar, A., & Wu, L. (2001). Predictable changes in
yields and forward rates. *Journal of Financial Economics*, 59(3):281–311.

Backus, D., Foresi, S., & Telmer, C. (1999). Discrete time models of bond pricing.
In *Advanced Fixed Income Valuation Tools*. John Wiley and Sons.

Barber, J. R. & Copper, M. L. (1996). Immunization using principal component
analysis. *Journal of Portfolio Management*, Fall:99–105.

Beaglehole, D. & Tenney, M. (1991). General solutions of some interest rate
contingent claim pricing equations. *Journal of Fixed Income*, 2:69–83.

Bjork, T. & Christensen, B. J. (1999). Interest rate dynamics and consistent
forward rate curves. *Mathematical Finance*, 9:323–348.

Brigo, D. & Mercurio, F. (2001). *Interest Rate Models: Theory and Practice.*
Springer Verlag.

Campbell, J. & Shiller, R. (1991). Yield spreads and interest rate movements: A
bird's eye view. *Review of Economic Studies*, 58:495–514.

Chen, R. R. & Scott, L. (1993). Maximum likelihood estimation for a multifactor
equilibrium model of the term structure of interest rates. *Journal of Fixed
Income*, 3:14–31.

Collin, D. P. & Goldstein, R. S. (2002). Do bonds span the fixed income markets?
theory and evidence for unspanned stochastic volatility. *Journal of Finance*,
LVII:1685–1730.

Collin, D. P., Goldstein, R. S., & Jones, C. S. (2003). Identification and estimation of 'maximal' affine term structure models: An application to stochastic volatility. Working paper, Graduate School of Industrial Administration, Carnegie Mellon University.

Cox, J. C., Ingersoll, J. E., & Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–407.

Dai, Q. & Singleton, K. (2000). Specification analysis of affine term structure models. *Journal of Finance*, LV(5):1943–1977.

Dai, Q. & Singleton, K. (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics*, 63:415–441.

De-Rossi, G. (2004). Kalman filtering of consistent forward rate curves: A tool to estimate and model dynamically the term structure. *Journal of Empirical Finance*, 11:277–308.

Diebold, F. & Li, C. (2003). Forecasting the term structure of government yields. Working Paper, University of Pennsylvania.

Duan, J. C. & Simonato, J. G. (1999). Estimating and testing exponential-affine term structure models by kalman filter. *Review of quantitative Finance and Accounting*, 13:111–135.

Duarte, J. (2004). Evaluating alternative risk preferences in affine term structure models. *Review of Financial Studies*, 17:379–404.

Duffee, G. R. (2002). Term premia and interest rates forecasts in affine models. *Journal of Finance*, 57:405–443.

Duffie, D. (2001). *Dynamic Asset Pricing Theory*. Princeton University Press.

Duffie, D., Filipovic, D., & Schachermayer, W. (2003a). Affine processes and applications in finance. *The Annals of Applied Probability*, 13(3):984–1053.

Duffie, D. & Kan, R. (1996). A yield factor model of interest rates. *Mathematical Finance*, 6(4):379–406.

Duffie, D., Pedersen, L., & Singleton, K. (2003b). Modelling sovereign yield spreads: A case study of russian debt. *Journal of Finance*, 58:119–160.

Duffie, D. & Singleton, K. (1997). An econometric model of the term structure of interest rates swap yields. *Journal of Finance*, 52:1287–1323.

Fama, E. & Bliss, R. (1987). The information in long maturity forward rates. *American Economic Review*, 77(4):680–692.

Filipovic, D. (1999). A note on the Nelson and Siegel family. *Mathematical Finance*, 9(4):349–359.

Flury, B. (1988). *Common Principal Components and Related Multivariate Models*. John Willey and Sons, New York.

Frachot, A., Janci, D., & Lacoste, V. (1992). Factor analysis of the term structure: A probabilistic approach. Working Paper, Direction Générale Des Études, Banque de France.

Heidari, M. & Wu, L. (2003). Are interest rates derivatives spanned by the term structure of interest rates? *Journal of Fixed Income*, 13(1):75–86.

Jorion, P. (2000). *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill.

Leippold, M. & Wu, L. (2002). Asset pricing under the quadratic class. *Journal of Financial and quantitative Analysis*, 37(2):271–295.

Leippold, M. & Wu, L. (2003). Design and estimation of quadratic term structure models. *European Finance Review*, 7:47–73.

Litterman, R. & Scheinkman, J. A. (1991). Common factors affecting bond returns. *Journal of Fixed Income*, 1:54–61.

Longstaff, F. (1989). A nonlinear general equilibrium model of the term structure of interest rates. *Journal of Finance*, 23:1259–1282.

Lund, J. (1997). Non-linear Kalman filtering techniques for term structure models. Working Paper Aarhus School of Business.

Pearson, N. D. & Sun, T. S. (1994). Exploiting the conditional density in estimating the term structure¿ An application of Cox, Ingersoll and Ross model. *Journal of Finance*, 49:1279–1304.

Pennacchi, G. G. (1991). Identifying the dynamics of real interest rates and inflation: Evidence using survey data. *Review of Financial Studies*, 4:53–86.

Singh, M. K. (1997). Value at risk using principal component analysis. *Journal of Portfolio Management*, 24:101–112.

Singleton, K. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics*, 102:111–141.

Vasicek, O. A. (1977). An equilibrium characterization of the TErm structure. *Journal of financial Economics*, 5:177–188.