Essays on Affine Term Structure Models

Bovorn Vichiansin

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2006

Program Authorized to Offer Degree: Economics

University of Washington
Graduate School

This is to certify that I have examined this copy of a doctoral dissertation by

Bovorn Vichiansin

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Chair of the Supervisory Committee:

Eric Zivot

Reading Committee:

Eric Zivot

Jefferson Duarte

Richard Startz

Date: 4/21/06

University of Washington

Abstract

Essays on Affine Term Structure Models

Bovorn Vichiansin

Chair of the Supervisory Committee:
Professor Eric Zivot
Department of Economics

Recent studies by Dai and Singleton (2002), Duffee (2002), and Duarte (2004) show that affine term structure models that match the time variability of the expected returns of bond yields do not generate time variation in the volatility of interest rates. This failure indicates that affine models fail to match one stylized fact of the term structure of U.S. interest rates. However, in this paper I show that the empirical limitation can be solved by allowing a more flexible specification of the risk-premia. I find that the affine models perform poorly at short forecasting horizons, but perform very well at longer horizons. Some affine models produce more accurate out-of-sample forecast than the random walk and other benchmark models. My empirical work also shows that the Kalman filter has the ability to filter discrepancy in zero-coupon bond yields occurring from different choices of splines, whereas the factor inversion method does not. When applied to different spline techniques, the factor inversion method gives larger differences in in-sample and out-of-sample bond yield forecasts than the Kalman filter method. Thus, I recommend estimating affine models by the Kalman filter rather than the factor inversion method. Lastly, I present an example of how the estimates of risk premia of GEA model can be used as an aid in duration decisions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I would like to thank my Reading Committee. Particularly, I am greatly indebted to Professor Eric Zivot who advised me throughout the years and provided me with profusely detailed and disciplined feedback. His support made my time at the University of Washington very pleasant. Just as inspirational and enthusiastic was Professor Jefferson Duarte. I owe him a lot in respect to term structure modeling insights and I very much enjoyed his contagious optimism throughout the dissertation process. I also want to thank Professor Richard Startz for his very useful and pointed comments and his great responsiveness.

I would like to thank Fred Nick for especially facilitating me to use his high speed computers in CSSCR to aid the completion of this dissertation. I appreciate many people at the World Bank Treasury for interesting conversations while I was completing this dissertation. My appreciation also extends to Gregory Duffee and Don Kim for their valuable comments while a part of this work was presented at International Conference on Computing in Economics and Finance.

I am not mentioning for sake of conciseness many other people who all taught me a lesson or a proof, at some point in time, but I hope they will forgive me. I wish I had consulted with some of them more often.

I would like to thank my family and numerous friends (too many to name here) for their continued moral support. I would like to thank my grandfather and grandmother for the role models they are in getting me to this point. I profoundly thank Siriluck Watewai for her continued support and encouragement.

Last but not least, this thesis is dedicated to my parents. I am most thankful for their unconditional support and enduring love over the years. I am greatly indebted to my father, for being such a father, and to my mother, for being such a mother!

I hope as I finish writing these lines, the fantastic time I had at the University of Washington will turn out to have been not an end in itself, but a beginning and a gate to the end of education which is creation.

DEDICATION

For my parents

CHAPTER I

Introduction

A term structure of interest rates is a set of yields on zero coupon bonds with a sequence of maturing dates. Term structure models are used extensively for many purposes, including forecasting future interest rates, risk management of portfolio containing bonds, the valuation of interest rate derivatives such as swaptions, caps and floors, the implementation of hedging strategies, the conduct of monetary policy, the financing of public debt, and the study of expectations of real economic activity and inflation.

One of the most popular classes of term structure models is the affine class[1]. The main advantage of this special class is its analytical tractability. In particular, it provides closed-form solutions for bond and bond-option prices. For models outside this class, bond prices generally have to be computed by performing Monte Carlo simulation or solving Feyman-Kac partial differential equations (PDF), which is computationally costly.

Affine term structure models (ATSM) require specification of state variables in continuous time under both the equivalent martingale measure and the physical probability measure. The physical probability measure is used to forecast future interest rates, derivative prices, and the future bond returns. The equivalent martingale measure is needed to fit the cross-section of bond yields, and to determine the current value of yields and interest rate derivatives. There are some benefits in defining the process of state variables to be affine[2] under both measures. If the dynamics of state variables are affine under the equivalent martingale measure (Q-measure), the bond price will take a simple form, which is an exponential affine function of unobservable factors. The disadvantage

[1] Examples of famous affine models are Vasicek (1977), Cox, Ingersoll and Ross (1985), Duffee and Kan (1996), Duffie and Singleton (1997), Dai and Singleton (2000), Collin-Dufresne and Goldstein (2001), Duffee (2002)

[2] Duffie, Filipovic and Schachermayer (2003) present a general characterization of affine processes.

of having non-affine dynamics under the physical measure (P-measure) is that the estimation will be much less simple.

There is a problem with the price of risk specification in existing affine models. A problem with the price of risk specification derived from the Gisanov theorem is that if at least one state variable goes to zero, the corresponding volatility of the state variables will go to infinity. This will cause the price of risk to explode and will lead to arbitrage opportunities. However, from the published paper of Cox, Ingersoll, and Ross (1985; hereafter CIR) to work by Duffee (2002), the price of risk specifications in affine term structure models have been deliberately designed to avoid the form derived from Gisanov theorem and prevent this problem as one of state variables going to zero. Examples of stable market price of risk representations in the affine term structure model literature are the Dai and Singleton (2000) completely affine model; the Duffee (2002) essentially affine model; and the Duarte (2004) semi-affine square-root model. Appendix A gives details of the price of risk specifications in these models.

Duffee (2002) improves a failure of completely affine models, which cannot match the observed relationship between the slope of the term structure and the expected returns on bonds, by proposing a more flexible form of the price of risk. His models are known as essentially affine models. Duarte (2004) demonstrates that his semi-affine square-root models produce better term premium forecasts than the corresponding essentially affine models. However, both Duffee (2002) and Duarte (2004) find that three-factor affine models cannot simultaneously match both the expected future yields and the time-variation in conditional variances of interest rates.

The first aim of this dissertation is to seek a term structure model in the framework of the affine class that can solve the mean-volatility tension described above. More specifically, the first question that I attempt to answer is which stochastic volatility model produces the change in yields more accurately than the essentially Gaussian affine model. I find that when I use a more flexible form of the price of risk than used by Duffee (2002), the affine models have more degrees of freedom and can outperform the essentially Gaussian affine in terms of forecasting bond yields.

In terms of calibrating and estimating the models, there are many papers that employ a special technique, called factor inversion. For this technique, first one needs to arbitrarily choose a set of yields to be equal to the number of state variables, and assume that those yields are measured without error. Then, one can invert these bond yields into the unobservable factors as if they were directly observable and estimate the models by the method of maximum likelihood.[3]

An obvious criticism with this factor extracting approach is that different choices of the reference bonds imply different state variable realizations, which is inconsistent with the model. Therefore, to not reject the common structure of the model, one must assume that all bonds are priced with error. The assumption that yields are observed without error is not strictly plausible. The error may represent market microstructure effects or measurement errors, which exist due to interpolation processes, rounding off prices, bid-ask spreads, asynchronous trading, temporary deviations appearing, and other market imperfections. In large samples, these errors should be small. However in small samples, different results obtain when different maturities for the bonds measured without error are used in model calibration. Whatever measurement error structure is chosen, given that one cannot simply invert the model to find the unknown factors, the Kalman filter[4] is a more preferred estimation method as compared to the factor inversion technique.

The basic principle of the Kalman filter is the use of temporal series of observable variables to reconstruct the value of the unobservable variables. First, the method requires that the model is expressed in linear Gaussian state-space form. A state-space model is characterized by a measurement equation and a transition equation. Once this has been made, an iteration process can begin. The Kalman filter can be used to calibrate affine models in which all yields are measured with error, but the factor inversion technique cannot. In this dissertation, I show obvious drawbacks of factor inversion technique with

---

[3] Some examples of studies in the term structure model literature that make use of this technique include Chen and Scott (1993), Pearson and Sun (1994), Duffie and Singleton (1997), Duffee (2002), Duarte (2004), Dai, Singleton, and Yang (2003), Ang and Bekaert (2003).

[4] See, for example, Duan and Simonato (1999), De Jong (2000), Han (2002), Chen and Scott (2003), and Duffee and Stanton (2004).

the poor assumption. I find that this problem is significantly improved by the Kalman filter.

The remainder of the dissertation is organized as follows. Chapter II presents the structure of generalized essentially affine models. Chapter III presents estimated three-factor affine term structure models. Data and estimation methods as well as Monte Carlo Study of a generalized essentially affine model are presented in this chapter. In addition, this chapter reports bond yield forecasting performance of affine models compared to their corresponding models and model-free benchmarks. In Chapter IV, I discuss the methods used to extract zero-coupon bond yields from coupon bond yields and illustrate a serious problem in estimating affine models by the factor inversion technique. Chapter V presents an application of a generalized essentially affine model in making investment decision. Chapter VI concludes.

CHAPTER II

The Generalized Essentially Affine Term Structure Model

The Generalized Essentially Affine (GEA) Model is defined by the following equations. First, the instantaneous short rate at time t is assumed to be an affine function of the n unobservable state variables X_t:

$$r_t = \delta_0 + \sum_{i=1}^{N} \delta_i X_{i,t} \tag{1}$$

where δ_i is an n-dimensional vector. Second, under the equivalent martingale measure, the state variables follow the affine diffusion:

$$dX_t = \kappa^Q(\theta^Q - X_t)dt + \Sigma\sqrt{S_t}\,dW_t^Q \tag{2}$$

where $\kappa^Q(\theta^Q - X_t)dt$ is an instantaneous drift and $\Sigma\sqrt{S_t}$ is a matrix of volatilities. The term W_t^Q is an n-dimensional Wiener process under the equivalent martingale measure Q, κ^Q and Σ are $(n \times n)$ matrices, and S_t is an $(n \times n)$ diagonal matrix with i^{th} diagonal element

$$S_{i,t} = \alpha_i + \beta_i' X_{i,t} \tag{3}$$

with α_i a constant and β_i an n-dimensional vector[5]. Third, the state variables follow the following stochastic differential equation (SDE) under the physical probability measure:

---

[5] α_i and β_i must be positive to ensure that $S_{i,t}$ is positive.

$$dX_t = \kappa^P(\theta^P - X_t)dt + \Sigma\sqrt{S_t}dW_t^P \tag{4}$$

where W_t^P is an n-dimensional vector of independent standard Brownian Motions under the physical measure, κ^P and Σ are (n × n) matrices. The dynamics of the state variables under both measures are characterized by drift and diffusion terms. The model allows for mean-reversion and stochastic volatility. The state variables always have a tendency to move towards the long run mean θ^P. The speed of adjustment is measured by the parameter κ^P. The speed of adjustment cannot be zero, otherwise the dynamic behavior of the state variables is nonstationary. The price of risk process derived from Gisanov theorem is

$$\lambda(X_t) = \left(\Sigma\sqrt{S_t}\right)^{-1}\left(\kappa^P\left(\theta^P - X_t\right) - \kappa^Q\left(\theta^Q - X_t\right)\right) \tag{5}$$

The problem with this price of risk[6] specification is that it may become extremely large if some of the elements of S_t get close to zero, and the Novikov condition may not be satisfied. The Navikov condition is a sufficient condition for absence of arbitrage, and it requires that under the physical probability measure for the time interval [0,T]

$$E\left[\exp\left\{\int_o^T \lambda(X_u)\lambda(X_u)du\right\}\right] < \infty \tag{6}$$

In simple terms it means that the variation in λ_u must be finite. To avoid this problem one may limit and make the price of risk equal to:

[6] The price of risk is an amount that investors require in order to take an additional unit of risk.

$$\lambda(X_t) = \min\left(\left(\Sigma\sqrt{S_t}\right)^{-1}\left(\kappa^P\left(\theta^P - X_t\right) - \kappa^Q\left(\theta^Q - X_t\right)\right), C\right) \tag{7}$$

where C is a vector upper bound consisting of large numbers. When the bound is not hit, the dynamics of the state variables are affine under both measures. When the bound is hit, the dynamics of the state variables under the physical measure will be non-linear. However, for estimation purposes, one may work as if the dynamics are affine under both measures because the bound C can be so far away from being hit. As a result, the moment conditions used to estimate this model will be very close to the moment conditions implied by the affine diffusions above.

Assuming that κ^Q, θ^Q, κ^P and θ^P satisfy sufficient conditions[7] to guarantee a unique solution to the SDEs in equations (2) and (4), bond prices take the form

$$P_t(\tau) = e^{A(\tau) - B(\tau)' X_t} \tag{8}$$

where $\tau = T - t$. The parameters $A(\tau)$ and $B(\tau)$ satisfy the following ordinary differential equations (ODEs):

$$\frac{dA(\tau)}{dt} = -(\kappa\theta)^{Q'} B(\tau) + \frac{1}{2}\sum_{i=1}^{n}\left[\Sigma' B(\tau)\right]_i^2 \alpha_i - \delta_0 \tag{9}$$

$$\frac{dB(\tau)}{dt} = -\kappa^{Q'} B(\tau) - \frac{1}{2}\sum_{i=1}^{n}\left[\Sigma' B(\tau)\right]_i^2 \beta_i + \delta_x \tag{10}$$

These ODEs can be solved simply by numerical integration with the initial conditions: $A(0) = 0$ and $B(0) = 0_{N\times 1}$. The notation $(\kappa\theta)^Q$ is used to denote an n-dimensional vector equal to $\kappa^Q \times \theta^Q$. The bond yield $Y_t(\tau)$ can be derived from $P_t(\tau) = e^{-\tau Y_t(\tau)}$ and equation (9):

[7] The conditions are given in Duffie and Kan (1996).

$$Y_t(\tau) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)'}{\tau} X_t \tag{11}$$

Dai and Singleton (2000) introduced the canonical form $A_m(n)$ to denote an n-factor model with the first m factors affecting the S_t matrix. I will adopt $CA_m(n)$ to denote the completely ATSM, $EA_m(n)$ to denote the essentially ATSM, and $GEA_m(n)$ to denote the generalized essentially ATSM.

II.1 Three Factor GEA and corresponding models

A primary empirical question is whether the proposed generalized essentially affine models can simultaneously improve the predictability of bond yields and generate sufficient time variation in the volatility of interest rates. I compare the predictability of bond yields using the following three-factor models: the $GEA_0(3)$, $GEA_1(3)$ and $GEA_2(3)$ models, and the $EA_0(3)$, $EA_1(3)$ and $EA_2(3)$ models. However, the $GEA_0(3)$ coincides with the $EA_0(3)$ model. Therefore, there are five total models to be evaluated.

A choice of three factors is motivated by Litterman and Scheinkman (1991). They show that the first three principal components in yield changes account for more than 95% of variation in yield changes across the term structure of U.S. treasury zero-coupon bonds. I do not consider the completely affine models because Dai and Singleton (2002) and Duffee (2002) find that these models are not successful in forecasting changes in bond yields.

I start by discussing the specification of the estimated models and then discuss the data, empirical methodology, and results. Finally, I will include a brief explanation of interpolation methods to extract zero-coupon bond prices from coupon bond prices and demonstrate that robustness of some estimation techniques to the method to generate discount yields.

The three-factor generalized essentially affine and essentially affine models share the following structures for the instantaneous interest rate, the dynamics of state variables under the P and Q measures, and the market price of risk:

$$r_t = \delta_0 + \delta_1 X_{1t} + \delta_2 X_{2t} + \delta_3 X_{3t} \tag{12}$$

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \left[\begin{bmatrix} (\kappa\theta)_1^P \\ (\kappa\theta)_2^P \\ (\kappa\theta)_3^P \end{bmatrix} - \begin{bmatrix} \kappa_{11}^P & \kappa_{12}^P & \kappa_{13}^P \\ \kappa_{21}^P & \kappa_{22}^P & \kappa_{23}^P \\ \kappa_{31}^P & \kappa_{32}^P & \kappa_{33}^P \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix}\right] dt + \Sigma\sqrt{S_t}\, dW_t^P \tag{13}$$

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \left[\begin{bmatrix} (\kappa\theta)_1^Q \\ (\kappa\theta)_2^Q \\ (\kappa\theta)_3^Q \end{bmatrix} - \begin{bmatrix} \kappa_{11}^Q & \kappa_{12}^Q & \kappa_{13}^Q \\ \kappa_{21}^Q & \kappa_{22}^Q & \kappa_{23}^Q \\ \kappa_{31}^Q & \kappa_{32}^Q & \kappa_{33}^Q \end{bmatrix}\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix}\right] dt + \Sigma\sqrt{S_t}\, dW_t^Q \tag{14}$$

$$\lambda(X_t) = \min\left(\left(\Sigma\sqrt{S_t}\right)^{-1}\left(\kappa^P\left(\theta^P - X_t\right) - \kappa^Q\left(\theta^Q - X_t\right)\right), C\right) \tag{15}$$

where $C = \left(c_1, c_2, c_3\right)'$. Restrictions are placed on each model in equations (12) - (15) depending on the number of state variables affecting the conditional volatility.

II.1.1 GEA$_0$(3) and EA$_0$(3)

In these models, none of the state variables enters the conditional volatility, so the models cannot produce time variation in the volatility of interest rates. The dynamics of state variables follow a three-dimensional Gaussian diffusion. The GEA$_0$(3) model coincides with the EA$_0$(3) model. The matrix Σ is assumed to be the identity matrix $I_{3\times3}$. The vector α has all elements equal one. The vector θ^P and the matrix β are null. The canonical form imposes a lower triangular structure on κ^P. In addition, κ_{11}, κ_{22}, and κ_{33} are strictly positive. The elements of the matrix κ^Q and the vector $(\kappa\theta)^Q$ are given by:

$$\kappa^Q = \begin{bmatrix} \kappa_{11}^Q & \kappa_{12}^Q & \kappa_{13}^Q \\ \kappa_{21}^Q & \kappa_{22}^Q & \kappa_{23}^Q \\ \kappa_{31}^Q & \kappa_{32}^Q & \kappa_{33}^Q \end{bmatrix}, (\kappa\theta)^Q = \begin{bmatrix} (\kappa\theta)_1^Q \\ (\kappa\theta)_2^Q \\ (\kappa\theta)_3^Q \end{bmatrix} \tag{16}$$

To keep the model parsimonious, the following parameters are fixed at zero: κ_{31}, κ_{32}, $(\kappa\theta)_2^Q$, κ_{11}^Q, κ_{31}^Q, and κ_{12}^Q. There are 16 free parameters in these models. Note that, for the $CA_0(3)$ model, the parameters κ_{12}^Q, κ_{13}^Q, and κ_{23}^Q are imposed to be zero. In addition, $\kappa_{11}^Q = \kappa_{11}^P$, $\kappa_{21}^Q = \kappa_{21}^P$, $\kappa_{22}^Q = \kappa_{22}^P$, $\kappa_{31}^Q = \kappa_{31}^P$, $\kappa_{32}^Q = \kappa_{32}^P$, and $\kappa_{33}^Q = \kappa_{33}^P$.

II.1.2 $GEA_1(3)$ and $EA_1(3)$

These models have one state variable determining the volatility structure. The canonical representation of the $GEA_1(3)$ model has the following parameters:

$$\kappa^Q = \begin{bmatrix} \kappa_{11}^Q & 0 & 0 \\ \kappa_{21}^Q & \kappa_{22}^Q & \kappa_{23}^Q \\ \kappa_{31}^Q & \kappa_{32}^Q & \kappa_{33}^Q \end{bmatrix}, (\kappa\theta)^Q = \begin{bmatrix} (\kappa\theta)_1^Q \\ (\kappa\theta)_2^Q \\ (\kappa\theta)_3^Q \end{bmatrix} \tag{17}$$

$\alpha_2 = \alpha_3 = \beta_{11} = 1$, $\beta_{ij} = 0$ for $i \geq 1, j > 1$, $\theta_2^P = \theta_3^P = 0$, $\Sigma = I_{3\times3}$ and $\kappa_{12}^P = \kappa_{13}^P = 0$. For parsimoniousness, additional constraints are imposed on some parameters. These restrictions are $\kappa_{31}^P = \kappa_{32}^P = 0$, $(\kappa\theta)_3^Q = 0$, $\kappa_{11}^Q = \kappa_{31}^Q = \kappa_{32}^Q = 0$. The $EA_1(3)$ model has similar structure to the $GEA_1(3)$ model but with the additional restriction, $(\kappa\theta)_1^Q = (\kappa\theta)_1^P$. Parsimonious restrictions for the estimated $EA_1(3)$ model are the same as that for the estimated $GEA_1(3)$ model. The number of free parameters in the estimated $GEA_1(3)$ model is 18, and in the estimated $EA_1(3)$ model is 17. Note that, the $CA_1(3)$ model imposes four additional constraints on the matrix κ^Q, $\kappa_{22}^Q = \kappa_{22}^P$, $\kappa_{23}^Q = \kappa_{23}^P$, $\kappa_{32}^Q = \kappa_{32}^P$, and $\kappa_{33}^Q = \kappa_{33}^P$.

II.1.3 GEA$_2$(3) and EA$_2$(3)

Among all estimated models, these models can generate the highest stochastic volatility because of the first two state variables driving volatility. The matrix κ^Q and the vector $(\kappa\theta)^Q$ in the EA$_2$(3) model are given by:

$$\kappa^Q = \begin{bmatrix} \kappa_{11}^Q & \kappa_{12}^Q & 0 \\ \kappa_{21}^Q & \kappa_{22}^Q & 0 \\ \kappa_{31}^Q & \kappa_{32}^Q & \kappa_{33}^Q \end{bmatrix}, (\kappa\theta)^Q = \begin{bmatrix} (\kappa\theta)_1^Q \\ (\kappa\theta)_2^Q \\ (\kappa\theta)_3^Q \end{bmatrix} \tag{18}$$

In the EA$_2$(3) model, the following restrictions are imposed $\kappa_{12}^Q = \kappa_{12}^P$, $\kappa_{21}^Q = \kappa_{21}^P$, $(\kappa\theta)_1^Q = (\kappa\theta)_1^P$, and $(\kappa\theta)_2^Q = (\kappa\theta)_2^P$, while these parameters can vary independently in the GEA$_2$(3) model. The GEA$_2$(3) model gives us four more degrees of freedom, as compared to the EA$_2$(3) model. In the estimated EA$_2$(3) model, κ_{31}^P, κ_{12}^P, κ_{32}^P, and κ_{11}^Q are fixed at zero to keep the model parsimonious. The parameters κ_{31}^P, κ_{12}^P, κ_{32}^P, κ_{11}^Q, and κ_{21}^Q are imposed to be zero to make the estimated GEA$_2$(3) model parsimonious. The number of free parameters in the estimated GEA$_2$(3) and EA$_2$(3) models is 19 and 16, respectively. Note that, the CA$_2$(3) model imposes three additional constraints on the matrix κ^Q. They are $\kappa_{12}^Q = \kappa_{12}^P$, $\kappa_{21}^Q = \kappa_{21}^P$, and $\kappa_{33}^Q = \kappa_{33}^P$.

CHAPTER III

Empirical Analysis

III.1 The Data

For the empirical analysis summarized in the first five tables, I use 356 monthly observations on U.S. Treasury zero-coupon bond yields for maturities of three and six months and one, three, five, and ten years extracted from coupon bond prices from the CRSP Government Bond files using the Nelson-Siegel-Bliss spline method (Bliss (1997)). The yields of all maturities are assumed to be measured with error and ordered in a vector Y according to increasing maturity (Y_1 is the six-month zero-coupon yield). The prices of coupon bonds with a callable feature and special liquidity problems are removed before the spline smoothing is applied.[8] The sample period ranges from November 1971 to September 2001[9]. The length of the sample period is determined by the unavailability of data for years prior 1971. The sample period includes the period of historically high interest rate volatility associated with the monetary experiment of the Federal Reserve between 1979 and 1982.

[8] See Bliss (1997) for more details

[9] I only have bond CRSP data available up to 2001

III.2 Estimation Methods

III.2.1 Factor Inversion Method

Factor inversion is one of the most applied techniques in the affine model literature due to its simplicity in implementation.[10] For three factor models, one can invert six different observed bond yields for three time-series of unobservable factors. Consider a set of six yields for 3-month, 6-month, 1-year, 3-year, 5-year, and 10-year maturities. It is assumed that the yields of 6-month, 3-year, and 10-year zero coupon bonds are perfectly observed. The choice of yields is chosen in order to span yields of the term structure as wide as possible. The yields of 3-month, 1-year, and 5-year zero coupon bonds are assumed to be measured with serially uncorrelated, zero-mean measurement errors. The first set of yields are stacked in the vector Y_t and the other in the vector $\hat{Y}_t$. The difference between the perfectly observed yields and imperfectly observed yields is the measurement error ε_t. At each month end t, a vector X_t of latent variables, $X_{1,t}$, $X_{2,t}$, and $X_{3,t}$, can be derived from equation (11) as follows

$$\begin{bmatrix} \hat{X}_{1,t} \\ \hat{X}_{2,t} \\ \hat{X}_{3,t} \end{bmatrix} = \begin{bmatrix} \frac{B_1(\tau_1)}{\tau_1} & \frac{B_2(\tau_1)}{\tau_1} & \frac{B_3(\tau_1)}{\tau_1} \\ \frac{B_1(\tau_2)}{\tau_2} & \frac{B_2(\tau_2)}{\tau_2} & \frac{B_3(\tau_2)}{\tau_2} \\ \frac{B_1(\tau_3)}{\tau_3} & \frac{B_2(\tau_3)}{\tau_3} & \frac{B_3(\tau_3)}{\tau_3} \end{bmatrix}^{-1} \left(\begin{bmatrix} Y_{\tau_1,t} \\ Y_{\tau_2,t} \\ Y_{\tau_3,t} \end{bmatrix} - \begin{bmatrix} -\frac{A(\tau_1)}{\tau_1} \\ -\frac{A(\tau_2)}{\tau_2} \\ -\frac{A(\tau_3)}{\tau_3} \end{bmatrix} \right)$$

or

$$\hat{X}_t = Z_1^{-1}\left(Y_t - Z_0\right) \tag{19}$$

[10] See, for example, Pearson and Sun (1994), Fisher and Gilles (1996), Duffie and Singleton (1997), Dai and Singleton (2002), Duffee (2002), and Duarte (2004)

where τ_1, τ_2, τ_3 are 6-month, 3-year, and 10-year periods before the maturity date. $B_i(\tau_j)$ is the i^{th} element of a vector $B(\tau_j)$. The log likelihood of yields at time t is the sum of the log likelihoods of the perfectly observed yields and the log likelihood of the error term as

$$L_t(\psi) = \ln f_y\left(Y_t \mid Y_{t-1}\right) + \ln f_\varepsilon\left(\varepsilon_t\right) \tag{20}$$

The measurement error vector, ε_t, is assumed to be jointly normally distributed with zero mean and $\omega'\omega$ variance-covariance matrix, where ω is a matrix derived from the Cholesky decomposition of $E(\varepsilon_t\varepsilon_t')$. The conditional density of the perfectly observed yields, $f_y(Y_t \mid Y_{t-1})$, can be written as

$$f_y\left(Y_t \mid Y_{t-1}\right) = \left|\frac{dX_t}{dY_t}\right| f_x\left(X_t \mid X_{t-1}\right) \tag{21}$$

The transition density of the state variables from time t-1 to t, $f_x\left(X_t \mid X_{t-1}\right)$, is multivariate normal with known mean and variance-covariance matrix given in Appendix C. The vector of estimated parameters ψ is chosen to maximize the sum of the log likelihood of yields over all observations.

III.2.2 Kalman Filter

There are several criticisms of the factor inversion technique previously described. Instead of assuming that some yields are observed without error, it is more reasonable to assume that all yields are observed with error. When all yields are observed with error, the state variables are no longer invertible from observed yields. They have to be extracted from observed yields by some kind of filtering process. One approach is to use the Kalman filter.[11] The Kalman filter is a recursive algorithm for computing the mathematical expectation of unobservable state variables, conditional on observing a history of noisy data on the hidden states, $E[X(t) \mid Y_t(\tau), Y_{t-1}(\tau), ..., Y_0(\tau)]$. The unobservable state variables can be filtered by applying the Kalman filter to the linear Gaussian state space representation of the model, characterized by the measurement equation and the transition equation. The measurement equation, a function of the parameter vector ψ of the term structure model, is

$$\begin{bmatrix} Y_t(\tau_1) \\ Y_t(\tau_2) \\ \vdots \\ Y_t(\tau_6) \end{bmatrix} = \begin{bmatrix} -\frac{A(\tau_1)}{\tau_1} \\ -\frac{A(\tau_2)}{\tau_2} \\ \vdots \\ -\frac{A(\tau_6)}{\tau_6} \end{bmatrix} + \begin{bmatrix} \frac{B(\tau_1)}{\tau_1} \\ \frac{B(\tau_2)}{\tau_2} \\ \vdots \\ \frac{B(\tau_6)}{\tau_6} \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ X_{3,t} \end{bmatrix} + \varepsilon_t$$

or

$$Y_t(\tau) = Z_0 + Z_1 X_t + \varepsilon_t \tag{22}$$

The measurement error ε_t is assumed to be normally distributed with zero mean. The noise in each bond yield is independent across time and other bond yields, and has a maturity-independent variance $h_{3/12}$, $h_{6/12}$, h_1, h_3, h_5 and h_{10}.

Since ATSMs are defined in continuous time but we only observe discrete time data, we need to work with the discrete-time dynamics of the continuous time process in equation (4). This discrete time process, or the transition equation, for the time interval equal to 1 is

$$X_{t+1} = m(X_t) + v_{t+1} \tag{23}$$

[11] The Kalman filter is commonly employed by control engineers and other physical scientists and has been successfully used in such diverse areas as signal processing in telecommunications, aerospace tracking, and underwater sonar systems. Since the 1980s, the Kalman filter has been applied to non-engineering purposes such as forecasting.

where $m(X_t)$ is the first conditional moment of state variables. The mean and variance-covariance matrix of v_{t+1} are zero and $Q(X_t)$, respectively. For the form of $Q(X_t)$, see Appendix C. In $EA_0(3)$ models the transition equation is a the first order markov and X_t is Gaussian. Therefore, the usual Kalman filter is applicable. In $EA_1(3)$ and higher models, X_t is no longer Gaussian because the X_t also enters into the variance $S_{i,t}$. Then, the usual Kalman filter is not appropriate and needs to be modified (see Duan and Simonato (1999) and De Jong (2000)).

Quasi-maximum likelihood (QML) is applied with the Kalman filter to estimate all models in this paper. If the dynamics of the state variables are discretized at a small interval, the Kalman filter-based QML estimates (based on the conditional variance of state vector and the corresponding non-Gaussian likelihood function) would be close to the true ML estimates. This concept leads to the idea of QML estimation. Under this approach, the density of the state vector, conditional on the previous observation, is assumed to have a multivariate Gaussian distribution. The mean vector and variance-covariance matrix of the state variables are assumed to be proportional to a small time interval between observations. For an analytical expression of the first and second conditional moments of the state variables, see Appendix C.

Another estimation method commonly applied to term structure models is the Efficient Method of Moments (EMM) (Gallant and Tauchen (1996)), combined with the semi-nonparametric (SNP) model. Estimates using this method are consistent and efficient, however, Duffee and Stanton (2004) compare the small sample performance of the Kalman filter based QML and EMM/SNP estimation and find that the parameter estimates from the EMM/SNP technique severely underestimate their true values. They conclude that for reasonable sample sizes, the results strongly support the Kalman filter. EMM is also very difficult to implement relative to factor inversion and kalman filter methods. Therefore, the Kalman filter is chosen over EMM in this study.

III.3 Estimation Results

Table 1 reports the estimated parameters and standard errors for all considered models, which are calibrated with monthly zero coupon bond yields running from November 1971 to December 1997. The standard errors are computed using the sample Hessian with the sandwich covariance matrix estimator. In practice, this estimator produces results comparable to the outer product of the gradient of the log-likelihood function. All models are first estimated with restrictions only from the canonical representation presented in Appendix D. To reduce concerns that these models are over-parameterized, they are re-estimated after setting some parameters, specifically those that have relatively large standard errors, of unconstrained models to zero. This procedure does not result in a significant difference in the value of the likelihood function, so the constrained models are studied.

In all estimated models, the eigenvalues of the matrix κ^P are constrained to be positive to guarantee stationarity of the state variables. In Table 1, we see that one factor has very slow mean reversion for every model, and then this factor causes almost parallel movements. The other factors revert much quicker to their long-run means and largely influence short term bond yields. Table 2 reports the log-likelihood value for each model. The $EA_1(3)$ model is nested with the $GEA_1(3)$ model, and the $EA_2(3)$ model is nested with the $GEA_2(3)$ model. Log-likelihood ratio tests are performed under the null hypotheses that additional parameters in the GEA model compared to its corresponding model are significant. The results indicate that one cannot reject the null hypotheses at 95 percent confidence levels.

III.4 Monte Carlo Simulation

It is well known that the QML estimates could be conditionally biased and inconsistent because the conditional variance in the transition equation depends on the state variables. I perform 400 Monte-Carlo simulations[12] of the Kalman filter estimation with antithetic paths and Euler discretization interval equal to 1/252 year on the generalized essentially $A_2(3)$ model to investigate the seriousness of potential biases from the Kalman filter based QML in small-sample estimation. The reason that this model is chosen for testing is that this model has two state variables affecting the volatility and more degrees of freedom than its corresponding essentially affine model. We can then expect that this model could generate the highest time variation in the conditional volatility and the largest biases among the estimated models analyzed in this paper.

Specifically, the true parameters, which are reported in Table 1, for the $GEA_2(3)$ model are based on the results of fitting the model with the implied zero-coupon month-end bond yields extracted from coupon bond by the Nelson-Siegel-Bliss curve fitting method during November 1971 and December 1997. The simulated data contains three hundred and fourteen months of six bond yields with maturities of three, six months, one, three, five, and ten years. The bond yields are simulated by adding normally distributed noise to the model-implied bond yields. The noise in each bond yield is independent across time and other bond yields and has a maturity-independent variance $h_{3/12}$, $h_{6/12}$, h_1, h_3, h_5 and h_{10}. A standard deviation of each noise is set at 10 basis points.

Table 3 reports the true parameters, and summarizes the simulation results based on four hundred Monte Carlo simulations. Biases are not worse than Monte Carlo results of multifactor models estimated by the Kalman filter reported in Duffee and Stanton (2004). Many mean estimates of the parameters that are identified under the equivalent martingale measure are close to their true values. Estimates of the parameters identified under the equivalent martingale measure exhibit small-sample bias. Most mean asymptotic standard errors are close to the Monte Carlo standard deviations of the parameter estimates. This result suggests that the Kalman filter should work well for all estimated models despite its suboptimal asymptotic properties under the small sample size chosen.

---

[12] Similar analyses have been performed by Duan and Simonato (1999), De Jong (2000), Duffee and Stanton (2004).

III.5 Bond Yield Predictability

The bond yield predictability of the GEA models and their corresponding models is presented in section III.5.1. Section III.5.2 compares the bond yield predictability of affine models to those of model-free benchmarks.

III.5.1 Among Affine Models: In-Sample Fit and Out-of-Sample Test

I examine the forecasting performance of the five affine models using both in-sample and out-of-sample tests. I focus on six month and one year forecasting horizons for five different bond maturities. All forecasts are constructed using the parameters reported in Table 1. The parameters are estimated with data between November 1971 and December 1997. The out-of-sample analysis is performed using the data from January 1998 to September 2001. After first two years of observations are left to initialize the Kalman filter, there are 290 months of observations in the in-sample period.

Tables 4 through 7 report the bond yield predictability of various affine models for six month and one year horizons, respectively. Tables 4 and 5 focus on in-sample forecasts, and Tables 6 and 7 focus on out-of-sample bond yield forecasts. The forecast errors are defined as the actual yield minus the predicted yield. The predicted yield is computed from the conditional expectation, $E_t[Y_{t+\Delta t}]$, where Δt denotes a forecasting horizon.

Statistical evaluation of root mean squared forecast errors for each maturity is assessed by Diebold and Mariano (DM) (1995) statistics. Diebold and Mariano (1995) develop a test of the null hypothesis that the accuracy of forecasts from two competing models does not differ. In Tables 4 through 7, I only report the DM statistic for the comparisons between $GEA_2(3)$ and $EA_0(3)$ because I would like to know whether the additional degrees of freedom allow the $GEA_2(3)$ model to generate more accurate bond yield forecasts than the $EA_0(3)$ model. Statistical significance related to the null hypothesis that the forecast error for each estimated model is zero is based on Newey-

West (1987) standard errors. The lag length is chosen by calculating the optimal lag length for each series individually using the method of Newey and West (1994) and averaging those optimal lags across series. Forecast errors with the 1% and 5% significance levels are marked with ‡ and †, respectively.

For in-sample yield forecasts, reported in Tables 4 and 5, the $GEA_2(3)$ model has the best performance in terms of RMSE. Out-of-sample results are reported in Tables 6 and 7. The $GEA_2(3)$ model still performs well for all maturities at six months and one year horizons but is not quite as successfully as the $GEA_0(3)$ model at forecasting. We also observe that the generalized essentially affine models outperform their corresponding essentially affine models in terms of forecasting future change in yields. The results from the $GEA_2(3)$ model are more clear than the $GEA_1(3)$ model. The averages of the forecast errors from the $GEA_2(3)$ model for all maturities and at both horizons are not statistically different from zero. Low autocorrelation of the forecast errors in Table 4 through Table 7 indicates that the residuals from all estimated models have almost no serial correlation.

A good forecasting performance of the $GEA_2(3)$ model indicate that this model could solve the tension between matching the first two moments of the interest rates. This is a stochastic volatility model that improves the matching of the affine models to the time-varying term premium.

III.5.2 With Model-free Benchmarks: Out-of-Sample Test

Duffee (2002) and Duarte (2004) find that stochastic volatility affine models are typically unable to beat the random walk at horizon shorter than one year. I follow the standard procedure and compare out-of-sample forecasts from several affine models and their competitors at horizons up to two and a half years. Based on estimated parameters in Table 1, I calculate the root mean squared error of the h-month step ahead out-of-sample forecast from the $EA_1(3)$, $EA_0(3)$, and $GEA_2(3)$ models and compare them to that of the five competing forecasting models. The forecasting horizons are 6 months ($h = 6$) and 2 years (h = 24). The five competing models used to compare the forecast accuracy with affine models are as follows:

a. The driftless random walk model

$$Y^{\tau}_{f,t+h} = Y^{\tau}_{t} \tag{24}$$

The random-walk model predicts that the current bond yield is the best estimate of the future bond yield. This comparison will determine whether time-series forecasting from affine models is worthwhile.

b. Slope regression

$$Y^{\tau}_{f,t+h} - Y^{\tau}_{t} = \hat{\alpha}^{\tau} + \hat{\beta}^{\tau}(Y^{\tau}_{t} - Y^{3m}_{t}) \tag{25}$$

where the slope of the term structure at time t is defined as the difference between the current bond yield and three month bond yield.

c. Fama-Bliss regression

$$Y^{\tau}_{f,t+h} - Y^{\tau}_{t} = \hat{\alpha}^{\tau} + \hat{\beta}^{\tau}(F^{h,\tau}_{t} - Y^{\tau}_{t}) \tag{26}$$

The forecast yield change is regressed on the forward spread. This regression is a restricted version of Cochrane-Piazzesi[13] forward regression. $F^{h,\tau}_{t}$ is the forward rate at time t for loans between t + h to time t + h + τ.

d. Cochrane-Piazzei regression

[13] See Cochrane and Piazzesi (2002)

$$Y^{\tau}_{f,t+h} - Y^{\tau}_{t} = \hat{\alpha}^{\tau} + \hat{\beta}^{\tau}_{0} Y^{1y}_{t} + \hat{\beta}^{\tau}_{1} F^{12,1y}_{t} + \ldots + \hat{\beta}^{\tau}_{9} F^{12,9y}_{t} \tag{27}$$

This regression is a generalized version of the Fama-Bliss regression.

e. VAR(1) model

$$A^{\tau}_{f,t+h} = \hat{\alpha} + \hat{\mathrm{B}} A^{\tau}_{t} \tag{28}$$

where $A^{\tau}_{t} = [Y^{3m}_{t}, Y^{1y}_{t}, Y^{3y}_{t}, Y^{5y}_{t}, Y^{10y}_{t}]$

Forecasting results are reported in Tables 8 and 9. In Table 8 all affine models seem to predict bond yields poorly and forecasts show serial correlation for forecasting horizon up to one year. It is likely that affine models predict bond yields more accurately than do random walk model and other benchmarks as the forecast horizon lengthens as shown in Table 9. Serial correlation of the forecast errors of the affine models is mostly smaller than those reported in De Jong (2000).

One way to check whether the predictive power of affine models increases with the forecasting horizon is to compare the MSE ratio between each affine model and the random walk model at different forecasting horizons. Results are shown in Table 10. We see that all affine models outperform the random walk predictor at longer time horizons especially twenty four months. The impressive predictability of affine models in Table 10 carries over to Table 11 when the Cochrane-Piazzesi regression is used as a benchmark.

TABLE 1. Estimated parameters

	$EA_0(3)$	$GEA_1(3)$	$EA_1(3)$	$GEA_2(3)$	$EA_2(3)$
δ_0	0.024 (0.004)	0.034 (0.01)	0.036 (0.009)	0.019 (0.009)	0.011 (0.006)
$\delta_1 \times 100$	0.12 (0.03)	0.19 (0.05)	0.20 (0.01)	0.55 (0.3)	0.40 (0.2)
$\delta_2 \times 100$	0.246 (0.01)	0.31 (0.1)	0.16 (0.2)	0.19 (0.1)	0.36 (0.08)
$\delta_3 \times 100$	0.13 (0.1)	0.63 (0.14)	0.91 (0.08)	0.68 (0.2)	0.77 (0.7)
β_{21}	0	4.94 (0.90)	1.99 (0.29)	0	0
β_{31}	0	0.911 (0.23)	0.754 (0.69)	1	1
β_{32}	0	0	0	0.976 (0.51)	0.983 (0.34)
θ_1	0	1.01 (0.55)	1.00 (0.89)	0.907 (0.4)	1.001 (0.6)
θ_2	0	0	0	0.991 (0.4)	0.958 (0.3)
θ_3	0	0	0	0	0
κ_{11}	0.0068 (0.0007)	0.057 (0.07)	0.332 (0.24)	0.0027 (0.002)	0.351 (0.08)
κ_{21}	-0.091 (0.01)	-0.998 (0.63)	-0.988 (0.19)	-0.042 (0.04)	-0.007 (0.002)
κ_{31}	0	0	0	0	0
κ_{12}	0	0	0	0	0
κ_{22}	0.518 (0.069)	0.953 (0.30)	1.285 (0.89)	0.872 (0.8)	0.872 (0.1)
κ_{32}	0	0	0	0	0
κ_{13}	0	0	0	0	0
κ_{23}	0	0.888 (0.28)	0.937 (0.90)	0	0
κ_{33}	0.751 (0.09)	0.722 (0.14)	1.765 (0.85)	0.010 (0.01)	0.095 (0.05)
$\kappa\theta_1^Q$	0.988 (0.02)	0.662 (0.11)	0.332 (-)	2.515 (2.33)	0.351 (-)
$\kappa\theta_2^Q$	0	0.999 (0.26)	0.977 (0.54)	1.805 (0.29)	0.829 (-)
$\kappa\theta_3^Q$	0.995 (0.09)	0	0	9.571 (0.29)	7.333 (1.51)
κ_{11}^Q	0	0	0	0	0
κ_{21}^Q	-0.284 (0.004)	-1.057 (0.21)	-0.730 (0.06)	0	-0.007 (-)
κ_{31}^Q	0	0	0	0.0004 (0.0002)	0.0006 (0.0004)
κ_{12}^Q	0	0	0	1.125 (0.08)	0
κ_{22}^Q	0.681 (0.01)	1.000 (0.74)	1.285 (0.22)	1.003 (0.09)	0.997 (0.1)
κ_{32}^Q	-0.041 (0.001)	0	0	2.495 (1.59)	3.478 (1.70)
κ_{13}^Q	1.068 (0.26)	0	0	0	0
κ_{23}^Q	1.340 (0.01)	2.740 (0.19)	0.937 (0.51)	0	0
κ_{33}^Q	1.383 (0.9)	0.780 (0.21)	1.765 (0.88)	0.671 (0.004)	0.750 (0.2)

This table presents the estimated parameters of all constrained models estimated by the Kalman filter for Nelson-Siegel-Bliss data set. The models are estimated with monthly zero-coupon bond yields running from November 1971 to December 1997. The asymptotic standard errors are given in parentheses. (-) indicates a parameter that has a restriction imposed as described on pages 10 – 11. For the sake of conciseness, the estimated measurement error variances are not reported.

TABLE 2 Likelihood analysis

	$EA_0(3)$	$GEA_1(3)$	$EA_1(3)$	$GEA_2(3)$	$EA_2(3)$
Log Likelihood	9170.3	9312.4	9309.7	9338.7	9332.3
Likelihood ratio		5.4		12.8	
Critical value χ^2 (number of parameter restrictions)		3.8		9.5	

This table reports log likelihood, likelihood ratio test and their p-values between generalized essentially affine models and their corresponding models.

TABLE 3 Monte Carlo Simulation: The GEA$_2$(3) model

Parameters	True value	Mean	Std.dev	MC std.error
$\delta_0 \times 100$	1.9	1.3	0.9	0.5
$\delta_1 \times 1000$	5.5	10.0	2.8	1.7
$\delta_2 \times 1000$	1.9	2.4	1.0	0.5
$\delta_3 \times 1000$	6.8	10.1	2.2	1.3
$\beta_{32} \times 10$	9.8	9.8	5.1	1.6
$\theta_1 \times 10$	9.1	9.1	4.0	1.6
$\theta_2 \times 10$	9.9	9.9	4.1	2.0
$\kappa_{11} \times 1000$	2.7	1.8	2.1	1.4
$\kappa_{21} \times 100$	-4.2	-8.2	4.0	1.9
$\kappa_{22} \times 10$	8.7	8.8	8.0	5.6
$\kappa_{33} \times 100$	1.0	2.2	1.0	1.8
$\kappa\theta_1^Q$	2.5	2.2	2.3	1.0
$\kappa\theta_2^Q$	1.8	2.0	0.3	0.2
$\kappa\theta_3^Q$	9.6	10.0	0.3	0.5
$\kappa_{31}^Q \times 1000$	0.4	0.6	0.2	0.4
$\kappa_{12}^Q \times 10$	11.3	11.2	0.8	0.4
$\kappa_{22}^Q \times 10$	10.0	10.0	0.9	0.5
κ_{32}^Q	2.5	3.0	1.6	1.1
$\kappa_{33}^Q \times 10$	6.7	6.6	0.4	0.2
$h_{6/12} \times 1000$	1.0	1.0	0.6	0.2
$h_{6/12} \times 1000$	1.0	1.0	0.6	0.3
$h_1 \times 1000$	1.0	1.0	0.3	0.5
$h_3 \times 1000$	1.0	1.1	0.3	0.2
$h_5 \times 1000$	1.0	1.1	0.2	0.3
$h_{10} \times 1000$	1.0	1.1	0.2	0.1

This table summarizes the results of 400 Monte Carlo simulations. Twenty six years of monthly observations of 3- and 6-month, 1-, 3-, 5- and 10-year zero coupon bond yields are generated by the GEA$_2$(3) model. The data is observed with iid measurement error (standard deviation of the error $h_{3/12}$, $h_{6/12}$, h_1, h_3, h_5, h_{10}).

TABLE 4 In-sample yield fit

	Maturity	$EA_0(3)$	$GEA_1(3)$	$EA_1(3)$	$GEA_2(3)$	$EA_2(3)$
RMSE	3m	1.80	1.81	1.81	1.78	1.80
	6m	1.63	1.64	1.64	1.61	1.64
	1y	1.48	1.49	1.49	1.47	1.55
	3y	1.25	1.32	1.35	1.22	1.28
	5y	1.13	1.17	1.25	1.06	1.17
	10y	1.00	1.05	1.15	0.97	1.10
Mean error	3m	-0.08	-0.25	-0.09	-0.11	0.35
(×100)	6m	-0.10	-0.30	-0.10	-0.13	0.36
	1y	0.10	-0.08	0.10	-0.07	0.43†
	3y	0.35	0.15	0.35	-0.07	0.40†
	5y	0.44	0.12	0.49‡	-0.06	0.37†
	10y	0.51	-0.18	0.51‡	-0.05	0.30†
ρ(6)	3m	0.10	0.21	0.30	-0.03	-0.02
	6m	0.11	0.21	0.33	-0.03	-0.01
	1y	0.16	0.25	0.37	0.03	0.06
	3y	0.23	0.26	0.39	0.12	0.18
	5y	0.26	0.25	0.39	0.17	0.25
	10y	0.28	0.25	0.39	0.24	0.33

This table presents root mean-squared errors (RMSE) of in-sample 6-month forecasts from affine models estimated by the Kalman filter for Nelson-Siegel-Bliss data set. All models are estimated from November 1971 to December 1997. The symbols ‡ and † denote statistical significance related to a null hypothesis that forecast error is zero at the 1% and 5% levels where the standard errors are computed using the Newey and West (1987) method. ρ(6) is sample autocorrelation at displacement of six months.

TABLE 5 In-sample yield fit

	Maturity	$EA_0(3)$	$GEA_1(3)$	$EA_1(3)$	$GEA_2(3)$	$EA_2(3)$
RMSE	3m	2.17	2.32	2.34	2.15	2.18
	6m	2.05	2.23	2.25	2.04	2.06
	1y	1.98	2.13	2.16	1.93	2.01
	3y	1.78	1.81	1.91	1.61^{*}	1.79
	5y	1.60	1.64	1.81	1.49^{*}	1.63
	10y	1.45	1.46	1.71	1.36^{*}	1.49
Mean error	3m	-0.05	-0.10	0.07	-0.12	$0.69^{\dagger}$
(×100)	6m	0.08	-0.13	0.07	-0.20	$0.71^{\dagger}$
	1y	0.31	0.06	0.31	-0.14	$0.77^{\ddagger}$
	3y	0.66	0.23	$0.65^{\dagger}$	-0.16	$0.71^{\ddagger}$
	5y	0.79	0.18	$0.79^{\ddagger}$	-0.15	$0.65^{\ddagger}$
	10y	0.90	-0.13	$0.89^{\ddagger}$	-0.12	$0.54^{\ddagger}$
ρ(12)	3m	0.30	0.35	0.41	0.10	0.11
	6m	0.26	0.30	0.40	0.08	0.12
	1y	0.25	0.27	0.38	0.07	0.14
	3y	0.16	0.16	0.30	0.03	0.15
	5y	0.12	0.10	0.27	0.02	0.17
	10y	0.07	0.05	0.23	0.01	0.20

This table presents root mean-squared errors (RMSE) of in-sample 1-year forecasts from affine models estimated by the Kalman filter for Nelson-Siegel-Bliss data set. All models are estimated from November 1971 to December 1997. The asterisks, *, denote modified Diebold-Mariano statistical significance at 10% levels for which one can reject the null hypothesis that the difference between RMSE from the $GEA_2(3)$ model and the $EA_0(3)$ model is zero. The symbols ‡ and † denote statistical significance related to a null hypothesis that forecast error is zero at the 1% and 5% levels where the standard errors are computed using the Newey and West (1987) method. ρ(12) is sample autocorrelation at displacement of twelve months.

TABLE 6 Out-of-sample yield forecasts

	Maturity	$EA_0(3)$	$GEA_1(3)$	$EA_1(3)$	$GEA_2(3)$	$EA_2(3)$
RMSE	3m	0.80	0.82	0.85	0.80	1.06
	6m	0.80	0.81	0.83	0.79	0.98
	1y	0.81	0.82	0.84	0.80	0.90
	3y	0.88	0.89	0.89	0.88	0.89
	5y	0.83	0.85	0.83	0.82	0.86
	10y	0.70	0.73	0.73	0.68	0.88
Mean error	3m	0.32	0.10	0.35	0.02	-0.25
(×100)	6m	0.39	0.02	0.33	-0.10	-0.39
	1y	0.29	0.18	0.39	-0.11	-0.17
	3y	0.20	0.19	0.29	-0.20	-0.16
	5y	0.08	0.01	0.20	-0.19	-0.34
	10y	0.07	-0.49	0.08	-0.17	-0.64‡
ρ(6)	3m	-0.05	0.05	0.21	-0.07	0.04
	6m	-0.10	0.05	0.34	-0.10	0.13
	1y	0.08	0.07	0.05	0.11	0.10
	3y	-0.03	0.09	-0.03	0.03	0.09
	5y	-0.08	0.09	-0.06	-0.01	0.09
	10y	-0.14	0.09	-0.11	-0.01	0.11

This table presents root mean-squared errors (RMSE) of out-of-sample 6-month forecasts from affine models estimated by the Kalman filter for Nelson-Siegel-Bliss data set. All models are estimated from November 1971 to December 1997 and the forecast is made from January 1998 to September 2001. The symbols ‡ and † denote statistical significance related to a null hypothesis that forecast error is zero at the 1% and 5% levels where the standard errors are computed using the Newey and West (1987) method. ρ(6) is sample autocorrelation at displacement of six months.

TABLE 7 Out-of-sample yield forecasts

	Maturity	$EA_0(3)$	$GEA_1(3)$	$EA_1(3)$	$GEA_2(3)$	$EA_2(3)$
RMSE	3m	1.01	1.31	1.38	0.98	1.04
	6m	0.98	1.22	1.28	0.95	1.00
	1y	0.94	1.37	1.38	0.92	0.97
	3y	0.83	1.33	1.31	0.79	0.84
	5y	0.82	1.13	1.19	0.70	0.82
	10y	0.77	0.75	0.97	0.55^{*}	0.78
Mean error (×100)	3m	0.90	1.02	1.09	0.36	0.07
	6m	1.01	$0.91^{\ddagger}$	$1.01^{\ddagger}$	0.32	0.02
	1y	1.12	$1.08^{\ddagger}$	$1.12^{\ddagger}$	0.39	0.31
	3y	1.06	$1.04^{\ddagger}$	$1.06^{\ddagger}$	0.35	0.33
	5y	0.93	$0.79^{\dagger}$	$0.93^{\ddagger}$	0.32	0.09
	10y	0.71	0.15	$0.71^{\ddagger}$	0.23	-0.33
$\rho(12)$	3m	-0.47	-0.44	-0.45	-0.46	-0.44
	6m	-0.46	-0.42	-0.44	-0.45	-0.43
	1y	-0.44	-0.41	-0.42	-0.45	-0.46
	3y	-0.37	-0.36	-0.38	-0.41	-0.42
	5y	-0.33	-0.33	-0.34	-0.38	-0.39
	10y	-0.26	-0.28	-0.28	-0.31	-0.33

This table presents root mean-squared errors (RMSE) of out-of-sample 1-year forecasts from affine models estimated by the Kalman filter for Nelson-Siegel-Bliss data set. All models are estimated from November 1971 to December 1997 and the forecast is made from January 1998 to September 2001. The asterisk, * , denotes modified Diebold-Mariano statistical significance at 10% levels for which one can reject the null hypothesis that the difference between RMSE from the $GEA_2(3)$ model and the $EA_0(3)$ model is zero. The symbols ‡ and † denote statistical significance related to a null hypothesis that forecast error is zero at the 1% and 5% levels where the standard errors are computed using the Newey and West (1987) method. $\rho(12)$ is sample autocorrelation at displacement of twelve months.

TABLE 8 Out-of-sample 6 month forecasting horizon

Maturity	RMSE	Mean (x 100)	ρ (6)	RMSE	Mean (x 100)	ρ (6)
		$EA_1(3)$			Slope regression	
3m	0.85	0.35	0.21	-	-	-
6m	0.83	0.33	0.34	1.05	-0.33	0.53
1y	0.84	0.39	0.05	1.08	-0.39	0.56
3y	0.89	0.29	-0.03	0.97	-0.35	0.42
5y	0.83	0.20	-0.06	0.86	-0.28	0.32
10y	0.73	0.08	-0.11	0.66	-0.18	0.26
		$EA_0(3)$			Fama-Bliss regression	
3m	0.80	0.32	-0.05	1.05	-0.29	0.44
6m	0.80	0.39	-0.10	0.98	-0.27	0.52
1y	0.81	0.29	0.08	1.04	-0.34	0.56
3y	0.88	0.20	-0.03	0.90	-0.24	0.37
5y	0.83	0.08	-0.08	0.80	-0.16	0.26
10y	0.70	0.07	-0.14	0.62	-0.05	0.19
		$GEA_2(3)$			Cochrane-Piazzesi regression	
3m	0.80	0.02	-0.07	1.25	-0.84	0.43
6m	0.79	-0.10	-0.10	1.32	-0.90	0.51
1y	0.80	-0.11	0.11	1.30	-0.87	0.59
3y	0.88	-0.20	0.03	1.09	-0.71	0.56
5y	0.82	-0.19	-0.01	0.94	-0.62	0.49
10y	0.68	-0.17	-0.01	0.72	-0.49	0.40
		RW			VAR(1) on yield level	
3m	0.96	-0.27	0.44	1.17	-0.75	0.44
6m	0.83	-0.28	0.53	1.23	-0.79	0.52
1y	0.88	-0.28	0.55	1.22	-0.75	0.55
3y	0.81	-0.20	0.36	1.07	-0.64	0.47
5y	0.79	-0.13	0.25	0.95	-0.60	0.38
10y	0.61	-0.04	0.18	0.77	-0.52	0.31

This table presents the results of out-of-sample 6-month ahead forecasting using eight models, as described in section IV.2. Forecast error is defined as actual yield – forecasted yield. I report the root mean-squared and mean of forecasting errors, and sixth sample autocorrelation coefficients.

TABLE 9 Out-of-sample 24 month forecasting horizon

Maturity	RMSE	Mean (x 100)	ρ (18)	RMSE	Mean (x 100)	ρ (18)
		$EA_1(3)$			Slope regression	
3m	1.83	1.19	-0.16	-	-	-
6m	1.31	1.28	-0.57	1.42	0.51	0.75
1y	1.45	138	-0.96	1.37	0.29	0.61
3y	1.00	1.66	-0.09	1.12	0.05	0.50
5y	0.97	1.86	0.49	0.88	0.02	0.73
10y	0.75	2.16	0.86	0.54	-0.04	0.94
		$EA_0(3)$			Fama-Bliss regression	
3m	1.17	0.14	-0.52	1.22	0.32	-0.88
6m	1.17	0.17	-0.88	1.42	0.34	0.19
1y	1.15	0.17	-0.98	1.38	0.26	0.51
3y	0.89	0.11	-0.68	1.16	0.10	0.65
5y	0.64	0.08	0.13	0.97	0.14	0.81
10y	0.60	0.07	0.73	0.68	0.17	0.96
		$GEA_2(3)$			Cochrane-Piazzesi regression	
3m	1.21	0.20	-0.27	2.30	-1.02	0.52
6m	1.17	0.25	-0.72	2.39	-1.13	0.53
1y	1.22	0.28	-0.90	2.10	-1.15	0.43
3y	1.03	0.32	-0.14	1.70	-1.97	0.39
5y	0.85	0.37	0.40	1.48	-1.83	0.64
10y	0.61	0.46	0.81	1.00	-1.68	0.92
		RW			VAR(1) on yield level	
3m	1.21	0.30	-0.88	1.84	-1.44	-0.55
6m	1.24	0.23	-0.74	1.93	-1.54	-0.76
1y	1.28	0.14	-0.15	1.94	-1.55	-0.64
3y	1.19	0.14	-0.65	1.77	-1.47	-0.03
5y	1.01	0.21	0.81	1.62	-1.44	-0.54
10y	0.68	0.20	0.96	1.46	-1.39	-0.89

This table presents the results of out-of-sample 24-month ahead forecasting using eight models, as described in section IV.2. Forecast error is defined as actual yield – forecasted yield. I report the root mean-squared and mean of forecasting errors, and eighteenth sample autocorrelation coefficients.

TABLE 10 Out-of-sample tests for bond yields

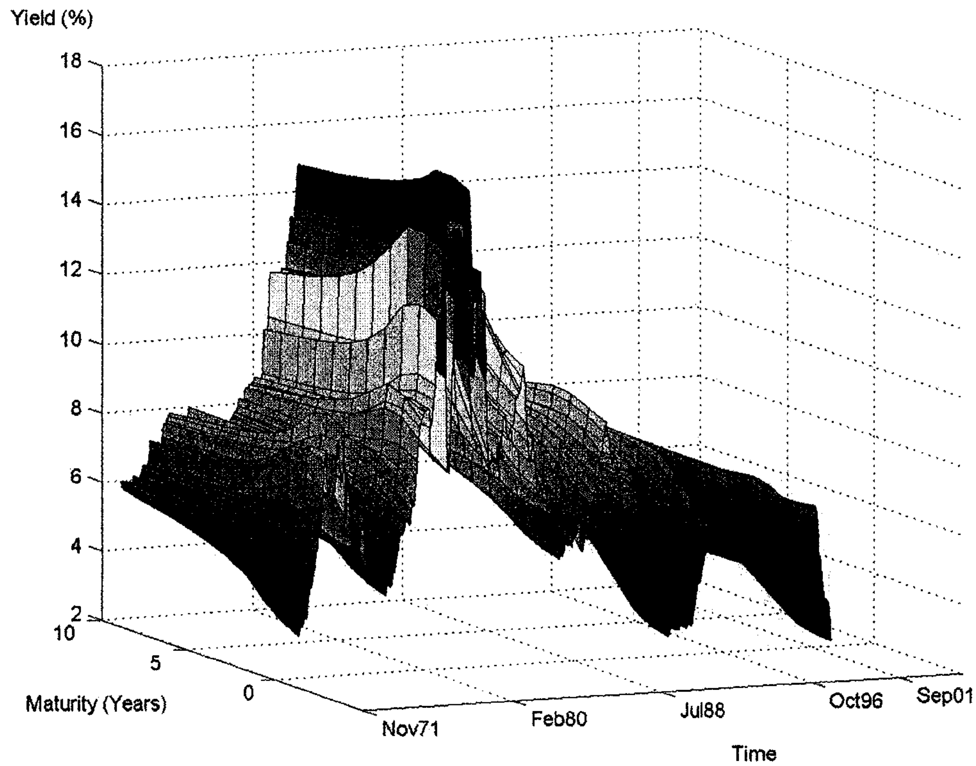
	Horizon (months)	Panel A: 6 month maturity			Panel B: 10 year maturity		
		$EA_1(3)$	$EA_0(3)$	$GEA_2(3)$	$EA_1(3)$	$EA_0(3)$	$GEA_2(3)$
MSE_{affine}/MSE_{rw}	3	1.35	1.27	1.28	1.58	1.40	1.49
	6	1.00	0.93	0.91	1.43	1.32	1.24
	12	1.68	0.99	1.03	1.79	1.00	0.95
	18	1.21	0.96	0.90	1.38	$0.83^{\dagger}$	0.89
	24	1.12	0.89	$0.89^{\dagger}$	1.22	$0.78^{\dagger}$	$0.80^{\dagger}$
	30	1.07	$0.87^{\dagger}$	$0.81^{\dagger}$	1.03	$0.78^{\dagger}$	$0.70^{\ddagger}$
DM (20)	3	-1.45	-1.83	-2.02	-1.81	-3.32	-2.55
	6	1.62	-0.64	-1.55	-1.56	-2.54	-0.67
	12	1.31	1.68	0.48	-2.15	-0.01	0.28
	18	-0.68	1.85	1.07	-2.60	2.84	0.62
	24	-2.29	1.20	0.35	-2.03	3.71	0.09
	30	-1.61	-0.46	-0.07	-2.15	0.43	-0.85

This table reports the ratio between the mean squared error (mse) of affine models and the mse of the random walk for the 6 months and 10 years yield curve segments. The ratio that is less than one signifies better forecasting performances of affine models. The symbols ‡ and † denote statistical significance related to a null hypothesis that two forecasts have the same mean squared error at the 5% and 10% levels.

TABLE 11 Out-of-sample tests for bond yields

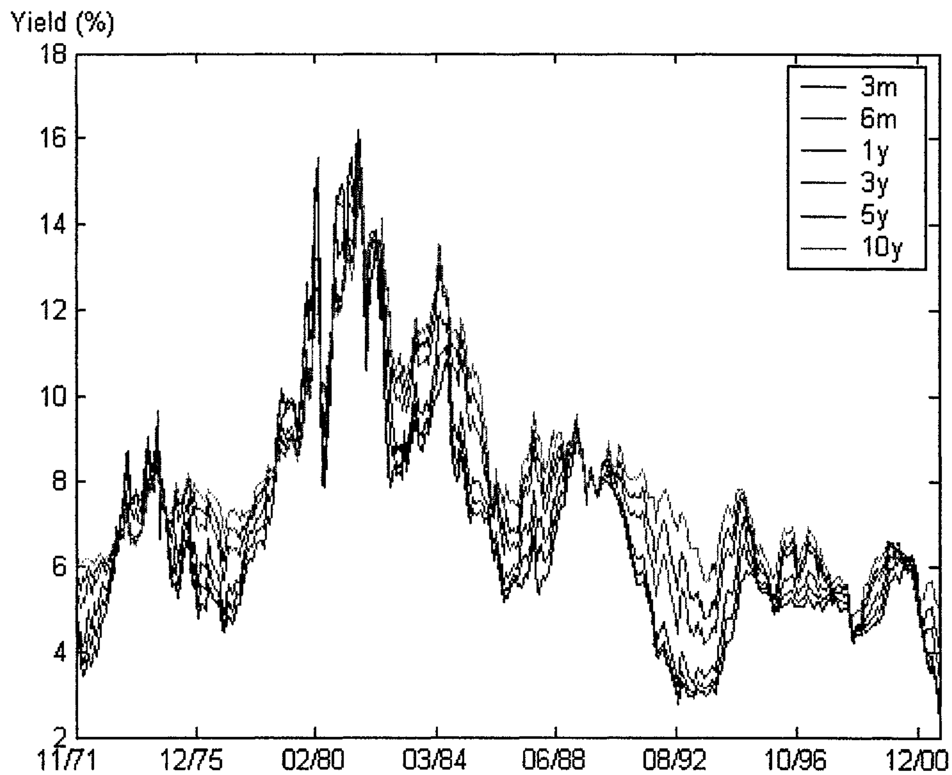
	Horizon (months)	**Panel A: 6 month maturity**			**Panel B: 10 year maturity**		
		$EA_1(3)$	$EA_0(3)$	$GEA_2(3)$	$EA_1(3)$	$EA_0(3)$	$GEA_2(3)$
MSE_{affine}/MSE_{cp}	3	0.67	$0.73^{\dagger}$	0.79	1.50	1.63	1.01
	6	$0.40^{\dagger}$	$0.37^{\dagger}$	$0.36^{\dagger}$	1.03	0.95	0.89
	12	$0.61^{\dagger}$	$0.68^{\dagger}$	$0.60^{\dagger}$	1.10	$0.61^{\dagger}$	0.59
	18	$0.58^{\dagger}$	$0.55^{\dagger}$	$0.45^{\ddagger}$	0.97	$0.21^{\ddagger}$	$0.35^{\ddagger}$
	24	$0.39^{\dagger}$	$0.31^{\dagger}$	$0.31^{\ddagger}$	0.56	$0.36^{\ddagger}$	$0.37^{\ddagger}$
	30	$0.48^{\dagger}$	$0.16^{\dagger}$	$0.15^{\ddagger}$	0.78	$0.21^{\ddagger}$	$0.15^{\ddagger}$
DM (20)	3	-0.41	-1.36	-1.20	-0.77	-2.42	-0.04
	6	2.94	2.08	1.88	-0.66	-0.81	0.48
	12	1.95	2.05	1.82	-0.88	1.41	1.03
	18	1.07	2.22	2.03	-1.37	4.93	2.86
	24	0.90	3.11	2.89	-1.84	3.53	2.81
	30	0.98	2.73	2.68	-2.20	2.83	2.88

This table reports the ratio between the mean squared error (mse) of affine models and the mse of the Cochrane-Piazzesi regression for the 6 months and 10 years yield curve segments. The ratio that is less than one signifies better forecasting performances of affine models. The symbols ‡ and † denote statistical significance related to a null hypothesis that two forecasts have the same mean squared error at the 5% and 10% levels.



Yield to maturity of zero bonds with time to maturity of three months to ten years. Monthly data from November 1971 and September 2001 (359 observations)

FIGURE 1 Zero-coupon bond yields of the U.S. treasury data



This figure shows the time-series evolution of the data used in the study for the observed maturity segments (3 months, 6 months, 1 year, 3 years, 5 years, 10 years).

FIGURE 2 Time series of the term structure of interest rates

CHAPTER IV

Comparison of Kalman Filter and Factor Inversion Estimation Methods to Curve Fitting Methods

IV.1 Zero-Coupon Bond Yields and Curve Fitting Methods

In practice, zero-coupon bond yields at most maturities are unobserved. I use five curve-fitting methods to extract zero-coupon bond yields from the same traded coupon-bond dataset: Unsmoothed Fama-Bliss (UFB), Fisher-Waggoner (FW), Smoothed Fama-Bliss (SFB), Nelson-Siegel-Bliss (NSB), and McCulloch (MC).

The McCulloch (1975) method uses a cubic spline[14], estimated by ordinary least squares, to approximate the discount function. Fama-Bliss's (1987) unsmoothed method, known as the bootstrap method, is an iterative method for piecewise-linear fitting of forward rate curves. The Nelson-Siegel's (1987) extended method, known as the Nelson-Siegel-Bliss method, approximates the yield curve, rather than the discount function, with a five parameter exponential polynomial that improves fit for long maturities[15]. The Fama and Bliss's smoothed method uses the Nelson-Siegel (1987) function to smooth out the unsmoothed Fama-Bliss piecewise linear curve. The Fisher-Waggoner (1997) method uses a cubic spline with variable roughness penalty (VRP) for different maturities to extract the forward rate curve. This last method improves the results of Fisher *et al.*(1995) on short maturity securities and performs well on medium and long maturity bonds. The most common interpolation techniques employed in the large financial institutions are techniques that generate smoothed yield curves[16].

---

[14] A spline can be thought of as a number of polynomials joined smoothly at the point where they join. These join points are called knot points and smooth means that at these points the first and second derivatives of the curve exist.
[15] See Bliss (1997).
[16] See Nymand, A. P., M. Golding, and A. Moreno (2004).

Figure 3 plots five zero-coupon yield curves at the end of August 1972 when the yield curves are upward-sloping. It is obvious that these zero-coupon yield curves do not lie on top of one another. The magnitude of these differences increases with maturity and becomes large for maturities longer than five years, regardless of the spline. The zero-coupon yield curves from Nelson-Siegel-Bliss and Smoothed Fama-Bliss methods are much smoother than that from Unsmoothed Fama-Bliss, Fisher-Waggoner, and McCulloch cubic spline methods. The standard deviations of yields among the five yield curves for long-term maturities could be as large as seven percent.

Table 12 shows statistics of the sample means and standard deviations of all spline-implied yield curves for the two different sub-periods: November 1971 - December 1997 and January 1998 – September 2001. These periods correspond to the periods used when performing in-sample and out-of-sample forecasting tests. In this table, one can see that short-term bonds give lower average returns than long-term bonds for all five data sets over both sub-periods. As one would expect, the sample standard deviations show that short-maturity discount bonds are more volatile than long-maturity discount bonds. The standard deviations in the first period are higher than the second period, primarily due to the fact that the first period includes the Federal Reserve experimental period during the early 1980s.

A main problem in estimating affine models with zero-coupon bond yields is that there is no generally accepted best practice for extracting the zero-coupon bond price from coupon bonds. It is only known that some spline methods generate smoother zero curves than others. From the discussion of the pros and cons of the factor inversion method of estimation, one can expect that estimated results from this technique would depend on the choice of splines. The next section shows how different the estimated results are when applying different splines to generate zero coupon bond yields.

IV.2 Comparison of Kalman Filter and Factor Inversion Estimation Methods

Many studies in the term structure model literature employ maximum likelihood and factor inversion techniques to calibrate affine models, and they make conclusions based on one or two discount bond data sets. It would be problematic if the results depended on which discount bond data set one uses. For instance, the best model from running a horse-race comparison would be different if we use different discount bond data sets. One may now wonder just how different the estimated results are across discount bond data sets if we estimate an affine model by the factor inversion and the Kalman filter. In this section, I compare the robustness of the two methods to the technique of generating the discount bonds. To present results in a more meaningful way, in-sample and out-of-sample forecasts are performed and then compared to the root mean squared forecasting errors. The standard deviations are used as a measure of dispersion of RMSEs across discount bond data sets.

Table 13 begins by examining the ability of the two estimation methods to yield forecasts. In this table, in-sample root mean squared errors for maturities six months, three years and ten years at three, six, and twelve-month forecasting horizons are reported. The $GEA_2(3)$ model is chosen to produce forecasts because of its superior forecasting performance, illustrated in chapter III. Table 13 shows that if the affine model is estimated by factor inversion, the standard deviations of RMSEs among the different implied zero-yield data sets increase as maturities decrease. Hence we observe a large dispersion of RMSEs in column 2 through column 6. This result is not surprising because the short term bond yields fluctuate much more than long term bond yields as shown in Table 12. We do not observe large dispersion under the Kalman filter columns. Therefore, the factor inversion technique is not robust to a choice of yields, a choice of methods to generate zero-coupon bond yields, nor is it robust to measurement error.

The excellent filtering performance of the Kalman filter carries over to the out-of-sample period. The magnitude of dispersion does not depend on forecasting horizon and bond maturity. Table 14 also documents that the factor inversion technique is not a robust estimation of the method of generating the discount bonds. The RMSEs are very different across data sets if we estimate an affine model with this technique. The largest dispersion appears on six-month yields and one-year forecast horizons under the factor inversion columns.

We observe these results because the values of state variables extracted from the Kalman filter are fairly close to that of other data sets. Also, the forecasted bond yields from the same model are similar across data sets. However, the state variables extracted by factor inversion differ across data sets, and we see large dispersion in the yield forecasts. These formal comparisons suggest that the Kalman filter is a much better alternative to factor inversion when estimating affine term structure models. Therefore, if policy makers want to make a decision based on affine models, they should avoid the factor inversion technique.

TABLE 12 Summary statistics for zero-coupon bonds

	Maturity	Data set				
	(year)	UFB	SFB	NSB	FW	MC
11/71-12/97	0.25	7.02	7.00	7.00	6.94	7.01
	0.50	7.27	7.26	7.26	7.29	7.27
	1.00	7.50	7.51	7.52	7.75	7.50
Mean	3.00	7.95	7.97	7.98	8.02	7.97
	5.00	8.19	8.21	8.21	8.20	8.23
	10.00	8.43	8.49	8.49	8.45	8.49
	0.25	2.78	2.77	2.77	2.76	2.77
	0.50	2.78	2.77	2.77	2.79	2.78
	1.00	2.67	2.66	2.67	2.77	2.68
St dev	3.00	2.38	2.38	2.38	2.41	2.39
	5.00	2.26	2.24	2.25	2.24	2.26
	10.00	2.09	2.09	2.10	2.08	2.07
1/98-9/01	0.25	4.92	4.88	4.84	4.88	4.89
	0.50	4.99	5.01	5.00	5.01	4.98
	1.00	5.12	5.13	5.13	5.32	5.08
Mean	3.00	5.33	5.32	5.32	5.39	5.31
	5.00	5.42	5.45	5.44	5.41	5.45
	10.00	5.46	5.69	5.69	5.57	5.60
	0.25	0.85	0.86	0.86	0.85	0.86
	0.50	0.91	0.89	0.89	0.91	0.90
	1.00	0.91	0.90	0.90	0.93	0.92
St dev	3.00	0.76	0.80	0.80	0.79	0.76
	5.00	0.68	0.67	0.67	0.67	0.68
	10.00	0.51	0.47	0.48	0.51	0.51

This table presents the sample means and sample standard deviations in percentage per year of the yields to maturity of five zero-coupon bonds extracted from traded coupon bonds by five spline methods. The sample contains 359 monthly observations in the period between November 1971 and September 2001. All maturities are expressed in years.

TABLE 13 In-sample yield forecasts

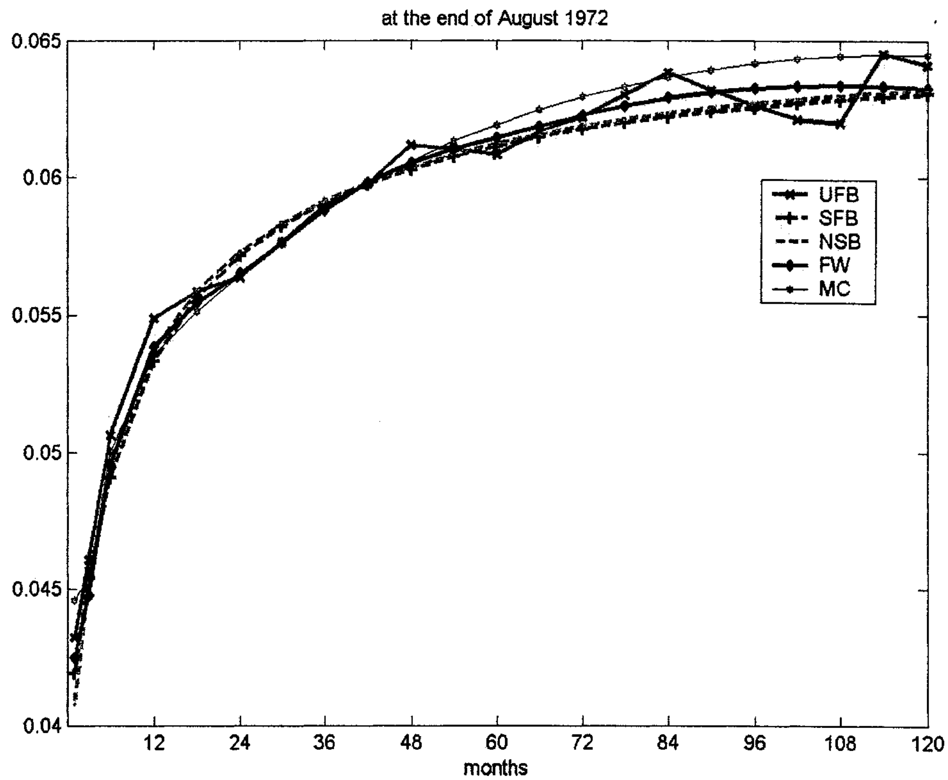
Forecast		Factor Inversion						Kalman Filter					
Horizon	Maturity	UFB	SFB	FW	NSB	MC	sd(%)	UFB	SFB	FW	NSB	MC	sd(%)
3m	6m	1.31	1.29	1.40	1.19	1.33	7.6	1.36	1.37	1.36	1.38	1.36	0.9
	3y	0.97	0.97	0.99	0.90	0.97	3.5	0.99	1.00	0.99	0.99	1.00	0.5
	10y	0.73	0.70	0.71	0.68	0.71	1.8	0.78	0.75	0.76	0.75	0.78	1.5
6m	6m	1.80	1.78	1.99	1.54	1.86	16.4	1.63	1.64	1.62	1.61	1.63	1.1
	3y	1.30	1.31	1.34	1.17	1.30	6.6	1.22	1.22	1.22	1.22	1.23	0.4
	10y	1.02	0.97	0.97	0.94	0.98	2.9	1.01	0.97	1.00	0.97	1.00	1.9
1y	6m	2.44	2.40	2.80	1.97	2.55	30.2	2.09	2.08	2.06	2.04	2.07	1.9
	3y	1.80	1.82	1.87	1.61	1.81	10.0	1.62	1.61	1.62	1.61	1.63	0.8
	10y	1.50	1.42	1.43	1.37	1.43	4.6	1.44	1.36	1.40	1.36	1.40	3.3

This table presents root mean-squared errors (RMSE) of in-sample forecasts from the $GEA_2(3)$ model. The model is estimated with zero-coupon yields between November 1971 and December 1997. All numbers are expressed in percentage.

TABLE 14 Out-of-sample yield forecasts

Forecast		Factor Inversion						Kalman Filter					
Horizon	Maturity	UFB	SFB	FW	NSB	MC	sd(%)	UFB	SFB	FW	NSB	MC	sd(%)
3m	6m	0.52	0.48	0.51	0.58	0.48	4.1	0.78	0.75	0.77	0.72	0.79	2.8
	3y	0.52	0.53	0.55	0.54	0.52	1.3	0.67	0.68	0.67	0.68	0.67	0.5
	10y	0.43	0.40	0.42	0.40	0.42	1.3	0.54	0.53	0.53	0.53	0.56	1.3
6m	6m	0.82	0.85	0.89	0.74	0.82	5.5	0.85	0.86	0.83	0.79	0.85	2.8
	3y	0.90	0.92	0.96	0.84	0.90	4.3	0.90	0.91	0.91	0.88	0.93	1.8
	10y	0.71	0.66	0.72	0.66	0.70	2.8	0.71	0.69	0.71	0.68	0.73	1.9
1y	6m	0.86	1.67	0.96	0.94	1.18	32.9	0.97	0.96	0.99	0.95	0.98	1.6
	3y	0.96	1.35	0.95	0.93	1.06	17.5	0.76	0.78	0.82	0.79	0.75	2.7
	10y	0.89	0.94	0.89	0.87	0.92	2.8	0.56	0.51	0.55	0.55	0.54	1.9

This table presents root mean-squared errors (RMSE) of out-of-sample forecasts from the $GEA_2(3)$ model. The model is estimated with zero-coupon yields between November 1971 and December 1997 and the forecast is generated from January 1998 to September 2001. All numbers are expressed in percentage.



This figures plot upward and humped shape yield curves at the end of August 1972 for five datasets. All discount curves are extracted from traded coupon bond prices by five interpolation methods in section IV.1

FIGURE 3 Yield curves

CHAPTER V

An application of the GEA model in investment strategies

Extracting risk premia for various maturity sectors of a yield curve is one of the applications for the GEA model. Using estimated parameters of a model and implied state variables, it is possible to compute the model implied expected excess returns for various sectors in the bond market. In this chapter, I demonstrated an example of how investors can take advantage of time varying risk premia extracted from the GEA model. Figure 4 below shows the $GEA_2(3)$ model implied ex-ante expected excess return and the ex-post realized excess return for a five-year zero coupon bond over a 3-month bill. The purpose for choosing the $GEA_2(3)$ model is because of its superior bond yield predictability. The model expected return captures the dynamics of a realized return very well and it is likely that the model's expected return for the five year bond leads the realized return.

In order to illustrate the benefits of the model with more data, I recalibrated the $GEA_2(3)$ model using bond data from November 1971 to September 2001. If the extracted risk premium decreases or the change in risk premium becomes negative, investors may expect that the yield of a five-year bond to fall or the price of a five year bond to be raise in the near future. Therefore, investors should go long a five-year bond. In contrast, if the extracted risk premium increases, investors may expect the yield of a five year bond to increase or the price of a five-year bond to decrease. Therefore, investors should short a five-year bond.

If we know how many months an ex-ante risk premium leads an ex-post excess return, then we can implement certain investment strategies. Let X_t and Y_t be an ex-ante risk premium extracted from the model and an ex-post excess return (realized), respectively. We can compute $Corr(\Delta X_t, \Delta Y_{t-i})$ for i = 0, 1, 2, and 3 months to see the dynamic relationship between changes in the risk premium and changes in the excess return. These correlations are $Corr(\Delta X_t, \Delta Y_t)$ = 52.6%, $Corr(\Delta X_t, \Delta Y_{t-1})$ = 80.8%,

$Corr(\Delta X_t, \Delta Y_{t-2}) = 50.0\%$, $Corr(\Delta X_t, \Delta Y_{t-3}) = 48.1\%$. The correlation of a change in an ex-ante risk premium and a change in an ex-post excess return is highest when i =1 month, and the correlation is much higher than 50%.

I use the following example to demonstrate how investors can take advantage of a five-year bond ex-ante risk premium leading a five year bond ex-post excess return for one-month for most of the time between 1995 to 2001. Suppose an investor starts investing in January 1995 and invests the same amount of money into a bond for one-month at the beginning of each month[17]. He can choose the following three investment strategies. The first strategy: he can buy and hold a short-term bill, such as a three month bill for a month at the beginning of each month. The second strategy: he can be more aggressive by taking more interest rate risk. He can buy a long term bond with a constant maturity, for example, he can invest in a long-term bond, such as five-year bond, and roll over to the same maturity bond after holding it for one-month.

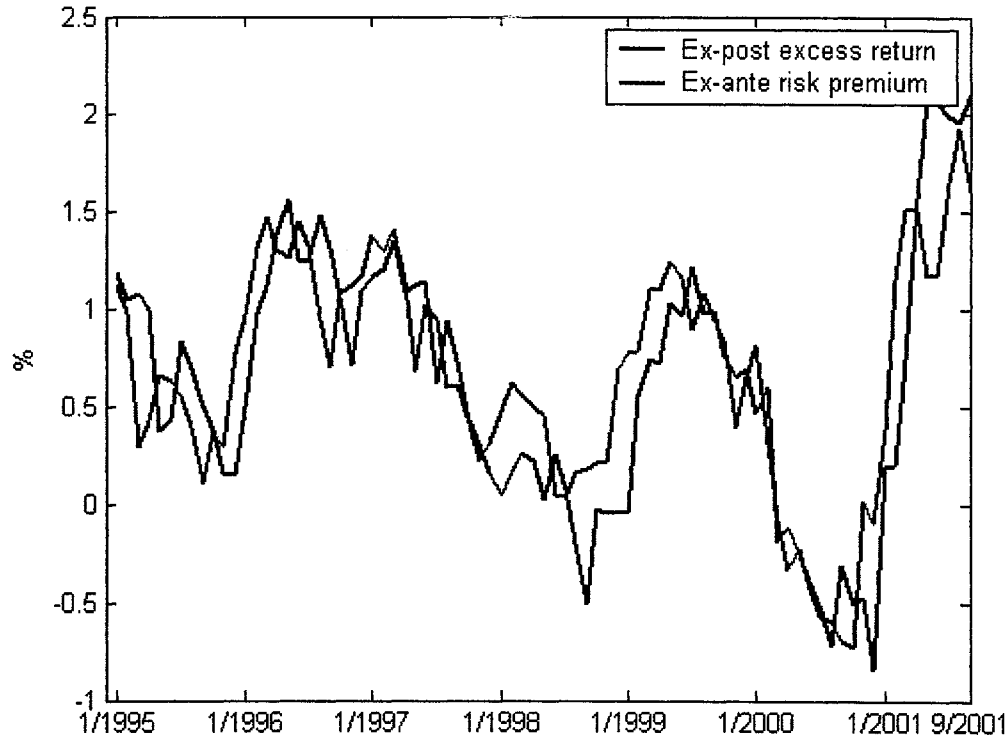
Unlike the second strategy, the third strategy makes use of time varying risk premium for the five-year bond. For this strategy, the investor longs and shorts the long term bond based on the estimates of risk premium instead of only going long in the long term bond. Specifically, he can long (short) the long term bond when the estimate of a change in risk premium is negative (positive).

Figure 5 illustrates the cumulative returns of the second and the third strategies over the first strategy between 1995 and 2001. Compared to the first strategy, the investor earns a cumulative excess return of approximately 10% using the second strategy while he earns a cumulative excess return of 70% with the third strategy over the 7 year period. Notice that he loses money between 1996 and 1999 for the second strategy but he does not loose money using the third strategy.

This example demonstrates how investors can take advantage of the time variation in risk premium extracted by the GEA model to help them making investment decision. Even though in real-life, one would require much more sophisticated

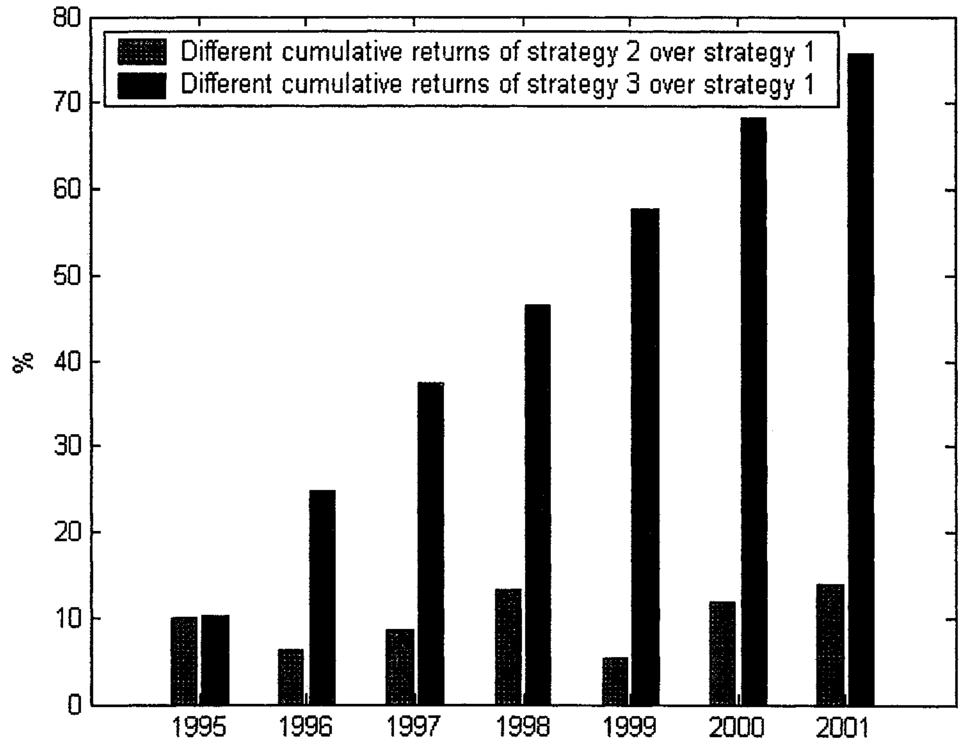
[17] Transaction cost of buying and selling US treasuries is very low according to the Brokerage Commission and Fee Schedule of Fidelity Investment.

investment strategies and signals from many different models. This example, however, does show that the GEA model can be very useful in making investment decision.



This figure plots realized excess returns and the $GEA_2(3)$ model-implied estimates of risk premia for five year zero coupon bond between 1995 and 2001

FIGURE 4 Model-implied Risk Premia



This figure shows cumulative returns of the second strategy and the third strategy over the first strategy conditional on the $GEA_2(3)$ model estimates of risk premia over seven years (1995-2001).

FIGURE 5 Cumulative Returns of Investment Strategies

CHAPTER VI

Conclusions

In this paper I specify a canonical representation for a family of admissible generalized essentially affine models which nest essentially and completely affine models. I use a state space framework and the Kalman filter to estimate five versions of affine models by allowing all zero-coupon bond yields to be observed with error. I first compare the ability of affine models, with the new and the traditional price of risk specifications to forecast future changes in the U.S. term structure, and then I find that the new form of the price of risk considerably helps affine models in producing more accurate yield forecasts. In addition, one sub-family of the stochastic volatility models with the new price of risk specification outperforms an essentially affine model with homoscedastic interest rates. This implies that a failure of affine models to match one of the stylized facts of the term structures of interest rates is solved. The paper also presents some evidence that these models perform poorly at short forecasting horizons, but perform very well at longer horizons.

The last conclusion is that the Kalman filter is a reasonable alternative to the factor inversion method. Despite its simplicity in implementation, the factor inversion technique has a very serious drawback. The estimation results of affine models by this technique highly depend on the method of generating zero-coupon bond yields, whereas the Kalman filter gives similar results across methods. Therefore, the Kalman filter is a reasonably accurate and tractable estimation technique that is recommended to estimate affine models. Lastly, I have presented an example of how the GEA model can be used as an aid in making investment decisions.

Bibliography

[1] Ang, A. and G. Bekaert, 2003, The Term Structure of Real Rates and Expected Inflation, Working paper, Columbia University.

[2] Bliss, R., 1997, Testing term structure estimation methods, *Advances in Futures and Options Research*, 9, 197-231.

[3] Chen, R, and L. Scott, 1993, Maximum likelihood estimation of a multifactor equilibrium model of the term structure of interest rates, *Journal of Fixed Income* 3, 14–31.

[4] Chen, R, and L. Scott, 2003, Multi-factor Cox-Ingersoll-Ross models of the term structure: estimates and tests from a kalman filter, *Journal of Real Estate Finance and Economics*, vol 27(2), pp 143-172.

[5] Cochrane, J.H. and Piazzesi, M., 2002, Bond Risk Premia, Manuscript, University of Chicago and UCLA.

[6] Collin-Dufresne, P. and R. S. Goldstein, 2002, Do bond span the fixed income markets? Theory and evidence for unspanned stochastic volatility, *Journal of Finance* 57, 1685-1730.

[7] J. C. Cox, J. E. Ingersoll Jr., and S. A. Ross, 1985b, A theory of the term structure of interest rates, *Econometrica*, 53: 385–407.

[8] Dai, Q., and K. J. Singleton, 2000, Specification analysis of affine term structure models, *Journal of Finance* 55, 1943-1978.

[9] Dai, Q., and K. J. Singleton, 2002, Expectation puzzles, time-varying risk premia, and dynamic models of the term structure, *Journal of Financial Economics* 63, 415-441

[10] Dai, Q., and K.J. Singleton, 2003, Term structure dynamics in theory and reality, *Review of Financial Studies*, vol. 16, no. 3, pp. 631-678.

[11] Dai, Qiang, and K.J. Singleton, and W. Yang, 2003, Regime shifts in a dynamic term structure model of U.S. treasury bond yields, Working paper, New York University.

[12] Diebold, F. X. and R. S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.

[13] De Jong, F., 2000, Time-series and cross-section information in affine term structure models, *Journal of Economics and Business Statistics*, vol. 18, no. 3, 300-314.

[14] Duan, J.C., and J. Simonato, 1999, Estimating exponential-affine term structure models by Kalman filter, *Review of Quantitative Finance and Accounting,* 13(2), 111-135.

[15] Duarte, J., 2004, Evaluating and alternative risk preference in affine term structure models, *Review of financial studies,* vol. 17 no. 2, 379-404.

[16] Duffee, G. R., 2002, Term premia and interest rate forecasts in affine models. *Journal of Finance*, Vol-57, 405-444.

[17] Duffee, G. R. and R. H. Stanton, 2004, Estimation of dynamic term structure models, *Working Paper*, Haas School of Business.

[18] Duffie, D. and R. Kan, 1996, A yield factor model of interest rates, *Mathematical Finance*, vol. 6 no. 4, 379-406.

[19] Duffie, D. and K. J. Singleton, 1997, An econometric model of the term structure of interest-rate swap Yield, *Journal of Finance* 52, no. 4, 1287-1321.

[20] Duffie, D., D. Filipovic, and W. Schachermayer, 2003, Affine Processes and Applications in Finance, *Annals of Applied Probability*, 13, 984-1053.

[21] Fidelity investment, 2006, Account Information Brokerage Commission and Fee Schedule.

[22] Fisher, M., and C. Gilles, 1996, Estimating exponential-affine models of the term structure, Working paper, Federal Reserve Board.

[23] Fisher, M., D. Nychka, and D. Zervos, 1995, Fitting the term structure of interest rates with smoothing splines, *Finance and Economic Discussion Series*, 95-1, Federal Reserve Board.

[24] Gallant, A.R. and G.E. Tauchen, 1996, Which moments to match?, *Econometric Theory*, 12, 657-681.

[25] Han, B., 2002, Stochastic Volatilities and Correlations of Bond Yields, Working Paper, Ohio State University.

[26] Litterman, R. and J. Scheinkman, 1991, Common factor affecting bond returns, *Journal of fixed income* 1, 54-61.

[27] McCulloch, J. H., 1975, The tax-adjusted yield curve, *Journal of Finance* 30, 811–830.

[28] Nelson, C. and A. Siegel, 1987, Parsimonious modeling of yield curves, *Journal of Business*, 60, 473 - 489.

[29] West, Newey K. and K. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–708.

[30] West, Newey K. and K. West, 1994, Automatic lag selection in covariance matrix estimation, *The Review of Economic Studies*, 61:631–653.

[31] Nymand, A. P., M. Golding, and A. Moreno, 2004, Measuring prices form financial instruments yield curve modeling, Working Paper, European Central Bank.

[32] Pearson, N. D. and T.-S., 1994, Sun, Exploiting the condititional density in estimating the term structure: An application to the Cox, Ingersoll and Ross model, *Journal of Finance*, XLIX, no 4:1279–1304.

[33] Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics*, 5, 177 - 188.

[34] Waggoner, D., 1997, Spline methods for extracting interest rate curves from coupon bond prices, Working Paper, Federal Reserve Bank Atlanta.

Appendix A: Price of Risk Specifications

These are some famous price of risk specifications in the term structure model literature that make models remain affine under Q measure:

i) Multifactor extensions of CIR model (1985), Duffie and Kan (1996), and Dai and Singleton (2000)

$$\Lambda_t = \sqrt{S_t}\lambda_1 \tag{A.1}$$

$\lambda_1 \in \Re^N$ The first complete characterization of the general class of multifactor affine models is provided by Duffie and Kan (1996). The affine models with this price of risk specification and correlated factors are known as the completely affine models.

ii) Duffee (2002)

$$\Lambda_t = \sqrt{S_t}\lambda_1 + \sqrt{S_t^-}\lambda_2 X_t \tag{A.2}$$

where

$$S_{t(ii)}^- = \begin{cases} \left(\alpha_i + \beta_i^{'} X_t\right)^{-1} if & \inf(\alpha_i + \beta_i^{'} X_t) > 0 \\ 0 & otherwise \end{cases} \tag{A.3}$$

$\lambda_1 \in \Re^N$ and $\lambda_2 \in \Re^{N\times N}$ The definition S^- ensures that the price of risk does not explode as diagonal elements in $S(X_t)$ go to zero.

iii) Duarte (2004)

$$\Lambda_t = \Sigma^{-1}\lambda_0 + \sqrt{S_t}\lambda_1 + \sqrt{S_t^-}\lambda_2 X_t \tag{A.4}$$

where $\lambda_0 \in \Re^N$, $\lambda_1 \in \Re^N$, and $\lambda_2 \in \Re^{N\times N}$. In this model, the Q-dynamics are affine but the P-dynamics are non-affine.

Appendix B: Nelson-Siegel-Bliss Method

Nelson-Siegel (1987) considers three different factors for the forward rate:

$$f_t(\tau) = \beta_{1t} + \beta_{2t}e^{-\lambda_t\tau} + \beta_{3t}\lambda_t\tau e^{-\lambda_t\tau} \tag{B.1}$$

The relationship between forward rates and yield to maturity is

$$y_t(\tau) = \frac{1}{\tau}\int_0^\tau f_t(u)du \tag{B.2}$$

Integrating this forward curve using a concept that the zero-coupon yield is an equally-weighted average of forward rates, we obtain the Nelson-Siegel yield curve model

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) \tag{B.3}$$

For a fixed λ, the Nelson-Siegel yield curve can be interpreted as a linear model with three factors, which can be interpreted as level, slope and curvature. This model does a reasonable job in fitting the yield cure, but has difficulties fitting the short-end of the curve typically underestimating these yields.

Bliss (1997) extends the Nelson and Siegel model by adding additional exponential parameter (λ_2) or extra hump term. The Nelson-Siegel-Bliss (NSB) model can be interpreted similar to the Nelson-Siegel model. For fixed λ_1 and λ_2, it has three linear factors, level, slope and curvature.

The yield curve model implied by the forward rate curve of Bliss (1997) is

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda_{1t}\tau}}{\lambda_{1t}\tau}\right) + \beta_{3t}\left(\frac{1-e^{-\lambda_{2t}\tau}}{\lambda_{2t}\tau} - e^{-\lambda_{2t}\tau}\right) \quad \text{(B.4)}$$

Then, the parameters $\varphi = \{\beta_1, \beta_2, \beta_3, \lambda_1, \lambda_2\}$ are estimated by minimizing

$$\min_{\varphi} \sum_{j=1}^{N_t} \left(w_j e_j\right)^2 \quad \text{(B.5)}$$

subject to

$$0 \le y_t\left(\tau_{\min}\right)$$

$$0 \le y_t\left(\infty\right)$$

These constraints ensure that the short and long ends of the discount rates are positive and the forward rates are non-negative. The errors ε_j are defined as

$$\varepsilon_j = \begin{cases} P_j^A - \hat{P}_j & \text{if } \; P_j^A < \hat{P}_j \\ P_j^B - \hat{P}_j & \text{if } \; P_j^B > \hat{P}_j \\ 0 & \text{otherwise} \end{cases}$$

and the weights, w_j, are defined in terms of Macaulay duration d_j measured in days

$$w_j = \frac{\frac{1}{d_j}}{\sum_{j=1}^{N_t} \frac{1}{d_j}}$$

The short-rate and the long-rate are given by $\beta_{1t} + \beta_{2t}$ and β_{1t}, respectively.

Appendix C: Models Estimation via Kalman Filter

Implementation of the Kalman filter to estimate affine term structure models is complicated due to the fact that the state variables are unobserved and correlated. If the state variables are uncorrelated, their dynamics are Gaussian, and the errors in transition equation are normally distributed, we can use the standard Kalman filter to estimate the model. For all affine models considered in this paper, except the $EA_0(3)$ model, we have to deal with a non-Gaussian state-space model, and the likelihood is not available in analytical form. Therefore, I implement quasi-maximum likelihood (QML), which requires approximating non-Gaussian states by the Gaussian ones with the first two moments.[18] The idea is to find the first two moments of a linear transformation of the state vector X_t. More specifically, we assume that the matrix κ^P can be diagonalized:

$$\kappa^P = \Omega D \Omega^{-1} \tag{C.6}$$

where D is an N × N diagonal with elements given by the elgenvalues of κ^P and Ω is the N × N matrix with the eigenvector associated with the i^{th} eigenvalue in its i^{th} column. Define

$$X_t^* = \Omega^{-1} X_t \tag{C.7}$$

Hence, for T > t

$$E\left[X_T^* \mid X_t^*\right] = \theta^* + e^{-D(T-t)}\left(X_t^* - \theta^*\right) \tag{C.8}$$

$$V\left[X_T^* \mid X_t^*\right] = \left[D^j + D^k\right]^{-1}\left[1 - e^{-(T-t)\left(D^j + D^k\right)}\right]\left[G_0^{jk} + \sum_{i=1}^{N} \theta_i^* G_i^{jk}\right]$$

[18] See, for instance, Fisher and Gilles (1996), and Duffee (2002)

$$+\sum_{i=1}^{N} G_i^{jk}\left[D^j + D^k - D^i\right]^{-1}\left[e^{-(T-t)D^i} - e^{-(T-t)\left(D^j+D^k\right)}\right]\left(X_{i,t}^* - \theta_i^*\right) \qquad \text{(C.9)}$$

where the f(j,k) is the element (j,k) in the matrix F(j,k) and $G_0 = \Sigma^* diag(\alpha^*)\Sigma^{*'}$, $G_i = \Sigma^* diag(\beta_i^*)\Sigma^{*'}$, $\theta^* = \Omega^{-1}\theta$, $\beta^* = \beta\Omega$, $\Sigma^* = \Omega^{-1}\Sigma$.

The conditional mean and variance for state variables are

$$\begin{aligned} E\left[X_T \mid X_t\right] &= \Omega E\left[X_T^* \mid X_t^*\right] \\ &\equiv m(X_t) \end{aligned} \qquad \text{(C.10)}$$

$$\begin{aligned} V\left[X_T \mid X_t\right] &= \Omega V\left[X_T^* \mid X_t^*\right]\Omega' \\ &\equiv Q(X_t) \end{aligned} \qquad \text{(C.11)}$$

The Kalman filter state space model is characterized by the measurement equation and the transition equation. The affine models in equation (4) give the following transition equation for the state variables

$$\hat{X}_{t|t-1} = a_t + \Phi_t \hat{X}_{t-1} + v_t \qquad \text{(C.12)}$$

Under normality assumption of v_t, the conditional distribution of the state variables X_t is multivariate normal with the following variance-covariance matrix

$$\hat{P}_{t|t-1} = \Phi_t P_{t-1} \Phi_t' + \hat{Q}_t \qquad \text{(C.13)}$$

The second part of the Kalman filter state space form is the measurement equation. Zero-coupon bond yields are used to construct this equation. Let $Y_t = (Y_{1t},\dots,Y_{kt})'$ denote