

Evaluating an Alternative Risk Preference in Affine Term Structure Models

Jefferson Duarte

University of Washington and IBMEC

Dai and Singleton (2002) and Duffee (2002) show that there is a tension in affine term structure models between matching the mean and the volatility of interest rates. This article examines whether this tension can be solved by an alternative parametrization of the price of risk. The empirical evidence suggests that, first, the examined parametrization is not sufficient to solve the mean-volatility tension. Second, the usual result in the estimation of affine models, indicating that some of the state variables are extremely persistent, may have been caused by the lack of flexibility in the parametrization of the price of risk.

Term structure models have several uses, including pricing fixed-income derivatives, managing the risk of fixed-income portfolios, and detecting relationships between the term structure of interest rates and macro-variables such as inflation and consumption. To perform well in these tasks, term structure models must be numerically and econometrically tractable while matching the empirical properties of the term structure movements.

At least two empirical properties of the term structure of interest rates have been well established by financial economists over the years [see Dai and Singleton (2003) for a survey]. First, the term premium, or the expected excess return of Treasury bonds, has a high time variability. Second, the volatility of interest rates is time varying. These two properties are so prominent in the data that they will be referred to as stylized facts.

While these two stylized facts are very well established in the empirical literature, affine term structure models are thoroughly discussed in the theoretical literature. Affine models are those in which the yield of zero coupon bonds are affine functions of the model state variables. Classic

I am especially grateful to George Constantinides for very useful comments and continuous support. I thank Federico Bandi, John Cochrane, Wayne Ferson, Alan Hess, Ravi Jagannathan, Avraham Kamara, Per Mykland, Jeffrey Russell, Kenneth Singleton, Stephen Schaefer, Pietro Veronesi, Eric Zivot, an anonymous referee, and seminar participants at the London Business School, London School of Economics, University of California–Irvine, University of Chicago, University of Rochester, University of Washington–Seattle, EFA and WFA meetings for their helpful comments. Financial support from the Graduate School of Business–University of Chicago and CAPES–Brazil is greatly appreciated. This article was previously entitled, “The Relevance of the Price of Risk in Affine Term Structure Models.” All errors are mine. Address correspondence to Jefferson Duarte, University of Washington Business School, 267 MacKenzie Hall, Box 353200, Seattle, WA 98195-3200, or e-mail: jduarte@u.washington.edu.

examples of affine models are Vasicek (1977) and Cox, Ingersoll, and Ross (1985; hereafter CIR).

The interest in affine models is understandable given their convenient numerical and econometric tractability. Aside from their tractability, however, there is evidence that current affine models do not match the two stylized facts verified by the empirical term structure literature. Specifically, Dai and Singleton (2002) and Duffee (2002) provide evidence that current affine models with sufficient flexibility to generate the observed variation in the term premium are incapable of producing any time variation in the volatility of interest rates. That is, there is a tension between matching the first and second moments of the data in the affine models.¹

Only a subset of affine models has been empirically rejected. The theoretical definition of affine models is not based on any parametrization for the price of any source of risk [see Duffie and Kan (1996)]. Conversely, the time-series estimation of affine models is based on maintained hypotheses about the prices of risk. Consequently, if the empirical rejection of affine models is driven only by the maintained assumption about the price of risk, then it may be possible to build highly tractable and accurate affine models by allowing more flexible parametrization for the price of any source of risk.

The affine term structure model proposed here is different from previous affine models because of its parametrization for the price of any source of risk. The proposed model is called the “Semi affine square-root” (SAS-R) model. The SAS-R model assumes a parametrization for the price of risk more flexible than the parametrizations assumed in affine models previously examined in the empirical literature.

To analyze the difference in performance of the SAS-R model in relation to other affine models in the current literature, a series of traditional affine models is compared with the corresponding SAS-R models. The power of each model to explain the time variation of the term premium is compared. The comparison between these models indicates that the SAS-R model improves in matching the time variability of the term premium. The SAS-R model improvement is caused by the fact that it allows the change in sign of any source of risk. The change in sign of the price of risk permits the SAS-R model to match the mean reversion that is in the level of the rates. In all the estimated SAS-R models, the level of the rates is more mean reverted than in the corresponding traditional affine models.

The mean reversion of the state variables in the SAS-R model explains some puzzling results of previous studies of affine models, and it suggests that a difference may exist in the performance of the proposed model in

¹ There is evidence that this mean-volatility tension is also present in other models outside the affine class [see Ahn, Dittmar, and Gallant (2002)].

detecting relationships between the term structure of interest rates and macrovariables. Extremely persistent state variables are a finding common to previous estimations of affine models.² Indeed, highly persistent state variables with half-lives of centuries are not unusual in the estimation of affine models. State variables with half-lives of centuries are puzzling because it would be economically sensible to find state variables with half-lives of the same magnitude as the average duration of business expansions or contractions. The faster mean reversion of the state variables in the SAS-R model suggests that the strong persistence in other affine models could have been partially driven by the strong restrictions in the price of risk. Consequently some of the relationships between the term structure of interest rates and macrovariables derived in previous studies may have been skewed by the price of risk restrictions.

Even though the SAS-R models perform better in matching the time variability of the term premium, the SAS-R improvement is not sufficient to solve the tension between matching the first and second moments of yields. To analyze this tension, a model that does not allow for stochastic volatility of yields is estimated and used as a benchmark of the performance to forecast the change in yields. This homoscedastic model performs better in forecasting changes in yields than any other estimated model with stochastic volatility.

The remainder of the article is organized as follows: Section 1 presents a general semiaffine square-root model. Section 2 empirically examines the proposed model. Section 3 summarizes the results. All proofs are in the appendix.

1. Model

The description of the model is divided into two sections. First, a general semiaffine square-root model is presented. Second, the models estimated in the empirical section of the article are presented.

1.1 The SAS-R model

Let $X_t = (X_{1,t}, \dots, X_{n,t})'$ be a state variable vector following an $It\hat{o}$ process. Under the framework set out by Duffie and Kan (1996), an affine term structure model satisfies the following two conditions: First, the short-term interest rate is

$$r(X_t) = \delta_0 + \sum_{i=1}^n \delta_i \times X_{i,t} \quad (1)$$

² Some examples of estimations of affine models that resulted in state variables with extremely strong persistence are in Chen and Scott (1993), Pearson and Sun (1994), Campbell and Viceira (1997), Duffie and Singleton (1997), Jagannathan and Sun (1998), and Duffee (2002).

Second, under the equivalent martingale measure, Q , the state variables follow the diffusion

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

where δ_i , $i = 0$ to n are constant, θ^Q is an $n \times 1$ vector, and κ^Q is an $n \times n$ matrix. The notation $(\kappa\theta)^Q$ is used to denote the $n \times 1$ vector equal to $\kappa^Q \times \theta^Q$. The matrix Σ is an $n \times n$ matrix and S_t is a diagonal matrix with the i th diagonal element given by $\alpha_i + \beta_i' X_t$, α_i is a constant, and β_i is an $n \times 1$ vector. The following notation is also used, $\alpha = (\alpha_1, \dots, \alpha_n)'$ and β equal to the $n \times n$ matrix whose i th row is β_i' .

In addition to these conditions, sufficient technical conditions must be assumed to guarantee that the model is admissible. For a description of these technical conditions see Dai and Singleton (2000). Notice that no assumption is made about the parametrization of $\lambda(X_t)$ to define an affine model. The pricing formulas are independent of the parametrization of the price of risk vector.

The SAS-R model is an affine model where, in addition to the conditions of Equations (1) and (2), the price of risk has the following parametrization:

$$\lambda(X_t) = \Sigma^{-1} \lambda_0 + \sqrt{S_t} \lambda_1 + \sqrt{S_t^-} \lambda_2 X_t, \quad (3)$$

where λ_0 and λ_1 are $n \times 1$ vectors and λ_2 is an $n \times n$ matrix. The matrix S_t^- is an $n \times n$ diagonal matrix with the i th diagonal element given by $S_t^-(i, i) = (\alpha_i + \beta_i' X_t)^{-1}$ if $\inf(\alpha_i + \beta_i' X_t) > 0$ and $S_t^-(i, i) = 0$ otherwise.

Completely affine models are affine models where the vector λ_0 and the matrix λ_2 in Equation (3) are null. Essentially affine models were proposed by Duffee (2002), and they are affine models where the vector λ_0 in Equation (3) is null. The semiaffine model is an extension to the essentially affine models where λ_0 is not null.

An examination of Equation (3) provides some initial clues about the cause of the difference between the empirical performance of the SAS-R model and of the completely and essentially affine models in matching the time variability of the term premium. First, notice that a consequence of the parametrization for the price of risk in the completely affine models is that the sign of the i th element of the price of risk vector is the same as the sign of the i th element of the vector λ_1 . Therefore, in the completely affine models, the sign of any element of the price of risk vector cannot change. Second, the essentially affine models partially solve this limitation of the completely affine models. Indeed, an examination of Equation (3) reveals that in the essentially affine models, the sign of $\lambda_i(X)$ can change if $S_t^-(i, i) \neq 0$, and hence some of the elements of the price of risk vector can switch signs. Third, the SAS-R solves the limitation of the completely and

essentially affine models because the sign of any element in the price of risk vector $\lambda(X)$ can change.

In essentially affine models, the sign of $\lambda_i(X)$ can switch only if $S_t^-(i, i)$ is different from zero. By construction, $S_t^-(i, i)$ is different from zero only if X_i does not affect the volatility of yields. Consequently, essentially affine models allow for sign switching in the price of risk only at the expense of limiting the volatility dynamics. The additional term, $\Sigma^{-1}\lambda_0$, in the price of risk parametrization of the SAS-R model offers additional sign-switching flexibility at no expense of limiting the volatility dynamics. As opposed to essentially affine models, the SAS-R model can match the time variability of the term premium without the expense of not matching the time variability of the volatility of the rates.

The drift of the state variables under the physical probability measure in the SAS-R model is given by

$$\mu(X_t) = \kappa\theta + \Sigma\sqrt{S_t}\Sigma^{-1}\lambda_0 - \kappa \times X_t, \quad (4)$$

where $\kappa = \kappa^Q - \Sigma(\lambda_1(1)\beta'_1 \dots \lambda_1(n)\beta'_n)' - \Sigma\sqrt{S_t}\sqrt{S_t}^{-1}\lambda_2$ and $\kappa\theta = \kappa \times \theta = (\kappa\theta)^Q + \Sigma(\lambda_1(1)\alpha_1 \dots \lambda_1(n)\alpha_n)'$.

The drift of the state variables is not affine in the SAS-R model, and for this reason the model is called “semiaffine.” The drift in Equation (4) has an additional square-root term which motivates the name “Semiaffine square-root.” The nonlinearity of the drift in Equation (4) raises a question related to the existence of a solution to the state variables’ stochastic differential equation under the physical probability measure; this question is answered in Appendix A.1.

The half-lives of the state variables in the essentially and completely affine models are given by $\ln(2)/d_i$, where d_i is the i th eigenvalue of κ . The half-lives of X in the do not have simple expressions because of the nonlinearity of the drift. However, for the parameters estimated, the expected value of $X_{t+\Delta t}$ conditional to $X_t = x$ is accurately approximated by $\bar{\theta} - (\bar{\theta} - x)\exp[-\bar{\kappa} \times \Delta t]$. Hence, $\ln(2)/\bar{d}_i$ $i=1$ to n , is used as a measure of the mean reversion of the state variables X , where $\bar{d}_i$ is an eigenvalue of $\bar{\kappa}$.

The price of risk can be parametrized in different ways. However, arbitrary choices of the price of risk can lead to arbitrage opportunities [see Ingersoll (1987, p. 400)]. The price of risk parametrization must satisfy technical conditions to make the model arbitrage free. Formally the model is arbitrage free if it admits an equivalent martingale measure Q . The SAS-R model is arbitrage free because it admits an equivalent martingale measure (for proof, see Appendix A.1). Note that the SAS-R model is different from completely and essentially affine models because it simply adds a vector of constants to the price of risk specification.

1.2 The estimated models

In the empirical work presented herein, some essentially and completely affine models are compared with their corresponding SAS-R models. To limit the size of the article, only some SAS-R models are compared with their corresponding affine models. The completely and essentially affine models chosen for comparison are the preferred models in Duffee (2002). All the estimated models have three state variables ($n = 3$) because of the usual characterization of term structure movements as changes in three factors [see Litterman and Scheinkman (1988)].

To identify the estimated models, I use a notation similar to the one in Dai and Singleton (2000). The symbol $CA_m(n)$ is used to represent an n -factor completely affine model with only m state variables causing the changes in the instantaneous covariance matrix S_t . In addition, the term *EA* represents an essentially affine model and the term *SAS-R* represents an SAS-R model. For instance, the term $EA_1(3)$ represents a three-factor essentially affine model with only one factor causing the changes in the instantaneous covariance matrix. The term $SAS-R_2(3)$ represents a three-factor SAS-R model with only two factors driving the changes in the instantaneous covariance matrix.

The estimated models are the $EA_1(3)$ and its corresponding $SAS-R_1(3)$, the $CA_2(3)$ and its corresponding $SAS-R_2(3)$, and the CIR and its corresponding $SAS-R_3(3)$ model. I estimate the model $CA_2(3)$ instead of estimating the model $EA_2(3)$ because Duffee (2002) did not find any evidence that the model $EA_2(3)$ has better performance than the $CA_2(3)$ model. There is no semiaffine generalization for the model $EA_0(3)$. The model $EA_0(3)$ is estimated because Dai and Singleton (2002) and Duffee (2002) present evidence that the model $EA_0(3)$ is the one that better matches the time variability of the term premium, therefore the model $EA_0(3)$ is used as a benchmark for the power to forecast yield changes.

Restrictions are imposed on the parameters of the estimated models. Some of these restrictions result from the canonical form presented in Dai and Singleton (2000) and in Duffee (2002), other restrictions are imposed to keep the models parsimonious. An estimation of unrestricted models indicates that the relaxation of the restrictions imposed for parsimony would not result in a significant difference in the log-likelihood function. In all estimated models, the eigenvalues of κ are constrained to be positive to guarantee stationarity of the state variables and the matrix Σ is assumed equal to the identity matrix $I_{3,3}$. This restriction imposed on Σ results from the canonical form in Dai and Singleton (2000). Table 1 displays all parameters of the estimated models.

1.2.1 The estimated $EA_0(3)$ model. In addition to the constraints imposed on β , θ , α , and κ by the canonical form in Dai and Singleton (2000), some elements of the matrices κ and λ_2 are constrained to be equal

Table 1
Estimated parameters in each model

	$EA_0(3)$	$EA_1(3)$	$SAS-R_1(3)$	$CA_2(3)$	$SAS-R_2(3)$	CIR	$SAS-R_3(3)$
$\kappa(1, 1)$	0.558(21.1)	0.003	0.183(1.9)	0.158(5.3)	0.383(3.2)	2.95(25.8)	2.924(26.2)
$\kappa(1, 2)$	0	0	0	-0.286(-7.6)	-0.287(-8.1)	0	0
$\kappa(1, 3)$	0	0	0	0	0	0	0
$\kappa\theta(1)$	0	0.158(2.4)	0.150(2.3)	0	0	0.679(2.5)	0.697(5.6)
δ_1	0.019(63.7)	0.001(4.4)	0.001(4.4)	0.001(3.3)	0.001(3.3)	1	1
$\beta(1, 1)$	0	1	1	1	1	0.005(2.3)	0.005(5.2)
$\beta(1, 2)$	0	0	0	0	0	0	0
$\beta(1, 3)$	0	0	0	0	0	0	0
$\alpha(1)$	1	0	0	0	0	0	0
$\lambda_0(1)$	0	0	1.029(2.1)	0	1.166(2.3)	0	0
$\lambda_1(1)$	-0.543(-2.6)	-0.014(-7.3)	-0.194(-2.0)	-0.017(-6.4)	-0.230(-2.0)	-13.52(-2.2)	-13.35(-2.1)
$\lambda(1, 1)$	0	0	0	0	0	0	0
$\lambda(1, 2)$	1.755(24.1)	0	0	0	0	0	0
$\lambda_2(1, 3)$	0	0	0	0	0	0	0
$\kappa(2, 1)$	0	-0.375(-2.4)	-0.380(-1.5)	-0.210(-5.9)	-0.218(-5.8)	0	0
$\kappa(2, 2)$	3.304(9.1)	0.587(20.5)	0.585(20.5)	0.390(13.1)	0.378(12.1)	0.455(6.8)	1.424(4.1)
$\kappa(2, 3)$	0	5.207(1.3)	4.911(1.1)	0	0	0	0
$\kappa\theta(2)$	0	-18.787	-0.311	0.210(2.1)	0.191(5.9)	0.037(9.0)	0.036(8.3)
δ_2	0.008(10.6)	0.001(1.7)	0.001(1.2)	0.001(6.9)	0.001(7.0)	1	1
$\beta(2, 1)$	0	10.619(0.8)	10.473(0.6)	0	0	0	0
$\beta(2, 2)$	0	0	0	1	1	0.008(9.3)	0.009(8.1)
$\beta(2, 3)$	0	0	0	0	0	0	0
$\alpha(2)$	1	1	1	0	0	0	0
$\lambda_0(2)$	0	0	0	0	0	0	2.934(2.6)
$\lambda_1(2)$	-0.217(2.6)	-3.731(-1.0)	-3.771(-1.0)	0	0	-6.73(-0.9)	-115.3(-2.9)
$\lambda_2(2, 1)$	0	39.471(0.6)	39.346(0.4)	0	0	0	0
$\lambda_2(2, 2)$	-1.701(-4.5)	0	0	0	0	0	0
$\lambda_2(2, 3)$	0	5.743(1.3)	5.698(1.1)	0	0	0	0
$\kappa(3, 1)$	-0.603(-3.4)	0	0	0.634(5.0)	0.666(4.7)	0	0
$\kappa(3, 2)$	0	0	0	-1.67(-14.9)	-1.63(-14.6)	0	0
$\kappa(3, 3)$	0.066(1.2)	2.911(7.1)	2.888(7.1)	1.791(19.5)	1.763(19.1)	9×10^{-6}	0.180(1.8)
$\kappa\theta(3)$	0	0	0	-11.387	-1.0008	0	0
δ_3	0.010(33.2)	0.003(4.6)	0.003(4.6)	0.006(20.0)	0.006(19.4)	1	1
$\beta(3, 1)$	0	0.286(2.1)	0.300(2.1)	0	0	0	0
$\beta(3, 2)$	0	0	0	1	1	0	0
$\beta(3, 3)$	0	0	0	0	0	0.002(19.0)	0.002(19.4)
$\alpha(3)$	1	1	1	1	1	0	0
$\lambda_0(3)$	0	0	0	0	0	0	1.041(2.0)
$\lambda_1(3)$	-0.185(-3.8)	0	0	-0.196(-4.1)	-0.200(-4.2)	-7.18(-14.9)	-100(-2.0)
$\lambda_2(3, 1)$	0.705(3.9)	0	0	0	0	0	0
$\lambda_2(3, 2)$	0.294(2.9)	0	0	0	0	0	0
$\lambda_2(3, 3)$	-0.065(-1.1)	-1.334(-3.3)	-1.308(-3.2)	0	0	0	0
δ_0	0.062(1.3)	0.049(2.3)	0.012(1.7)	0.049(16.4)	0.015(11.3)	-0.307(-3.3)	-0.312(-8.2)

The maximum-likelihood estimates of the parameters. T -statistics are displayed in parentheses. The displayed parameters with values zero or one and without t -statistics are constrained. Some of the parameter constraints result from the Dai and Singleton (2000) canonical form and others are imposed to keep the models parsimonious. For a description of the constraints see Section 1.2. The t -statistics of $\kappa(1, 1)$ in the $EA_1(3)$ model and of $\kappa(3, 3)$ in the CIR model are not displayed because they are in the frontier of the parameter space since the state variables are constrained to be stationary. The values of $\kappa\theta(2)$ in the $EA_1(3)$ and $SAS-R_1(3)$ models and $\kappa\theta(3)$ in the $CA_2(3)$ and $SAS-R_2(3)$ models result from the constraints $\theta(2)=0$ and $\theta(3)=0$ respectively, from the canonical form in Dai and Singleton (2000).

to zero to keep the model parsimonious. They are $\kappa(2, 1)$, $\kappa(3, 2)$, $\lambda_2(1, 1)$, $\lambda_2(1, 3)$, $\lambda_2(2, 1)$, and $\lambda_2(2, 3)$.

1.2.2 The estimated $EA_1(3)$ and its corresponding $SAS-R_1(3)$ model

Restrictions are imposed on κ , α , β , and θ by the canonical form in Dai and Singleton (2000). The definition of essentially affine models in

Duffee (2002) implies that the first row of the matrix λ_2 is equal to zero. The additional parsimony restrictions on the estimated $EA_1(3)$ and $SAS-R_1(3)$ models are $\kappa(3, 1) = \kappa(3, 2) = 0$, $\lambda_1(3) = 0$ and $\lambda_2(2, 2) = \lambda_2(3, 1) = \lambda_2(3, 2) = 0$. The parameters $\lambda_0(2)$ and $\lambda_0(3)$ are restricted to be zero in the $SAS-R_1(3)$ model.

The major difference between the estimated $SAS-R_1(3)$ model and the estimated $EA_1(3)$ model is that λ_0 is a null vector in the $EA_1(3)$ model, while λ_0 is not constrained to be zero in the estimated $SAS-R_1(3)$. The number of free parameters in the estimated $EA_1(3)$ model is 17 and in its corresponding $SAS-R_1(3)$ is 18.

1.2.3 The estimated $CA_2(3)$ and its corresponding $SAS-R_2(3)$ model.

Additional parsimony restrictions are imposed on some parameters of the Dai and Singleton (2000) canonical form. The additional restrictions on the estimated $CA_2(3)$ and $SAS-R_2(3)$ models are $\lambda_1(2) = 0$, $\kappa\theta(1) = 0$, $\beta(3, 1) = 0$, and $\beta(3, 2) = 1$. The parameters $\lambda_0(2)$ and $\lambda_0(3)$ in the estimated $SAS-R_2(3)$ model are assumed to be zero. The matrix λ_2 is null in both models because Duffee (2002) did not find any evidence that the model $EA_2(3)$ has better performance than the $CA_2(3)$ model.

The major difference between the estimated $SAS-R_2(3)$ and the estimated $CA_2(3)$ models is that λ_0 is a null vector in the $CA_2(3)$ model, while λ_0 is not constrained to be zero in the estimated $SAS-R_2(3)$ model. The number of free parameters in the estimated $CA_2(3)$ model is 14 and in its corresponding $SAS-R_2(3)$ is 15.

1.2.4 The estimated CIR and its corresponding $SAS-R_3(3)$ model.

The matrices κ and β are diagonal in the estimated $SAS-R_3(3)$ and CIR models. In addition, the following restrictions are imposed: $\delta_1 = \delta_2 = \delta_3 = 1$, $\theta(3) = 0$, $\lambda_0(1) = 0$, α is a null vector, and λ_2 is a null matrix. The major difference between the estimated $SAS-R_3(3)$ model and the estimated CIR model is that λ_0 is a null vector in the CIR model, while λ_0 is not constrained to be zero in the estimated $SAS-R_3(3)$ model. The number of free parameters in the estimated CIR model is 12 and in its corresponding $SAS-R_3(3)$ is 14.

2. Empirical Analysis

2.1 Description of the data

The data are composed of monthly observations of yields of zero coupon bonds with maturities equal to 3 and 6 months, and 1, 2, 5, and 10 years. The yields are calculated by applying the McCulloch cubic spline method on month-end price quotes for treasury issues. Price quotes of callable treasury issues and of bonds with special liquidity problems were not used

in the calculation. [see Bliss (1997) for a detailed description of the calculation.]

The period analyzed is from January 1952 to December 1998. There are 564 observations for each yield. The data from 1952 to 1991 are from McCulloch and Kwon (1993), the data from 1992 to 1998 are based on Bliss (1997). The data used are the same as the data used by Duffee (2002). As noticed by Campbell and Viceira (1997) and others, there is strong evidence of changes in interest rate behavior between 1979 and 1982. Interest rates were unusually high and volatile between 1979 and 1982. Even though there is this apparent change in behavior during this period, I consider the whole time series. The reason for considering the whole sample period and not considering data only from 1983 to 1998 is that the slope of the term structure is used in the empirical tests and the power of the slope to predict changes in yields is smaller over the period of 1983 to 1998 than over the period 1952 to 1998.

Both in-sample and out-of-sample analysis are performed. To estimate the model, the data between January 1952 and December 1993 are used. The out-of-sample analysis is performed using the remainder of the data from January 1994 to December 1998.

2.2 Estimation method

A common assumption in the affine literature is that prices of some zero-coupon bonds are exactly observed. This assumption permits the inversion of the pricing equations to obtain a time series of the latent state variables, which are used to estimate the model parameters by maximum likelihood [see, for instance, Pearson and Sun (1994)]. Herein it is assumed that the yields of the 6-month, 2-year and 10-year zero-coupon bonds are observed without errors. The maturities of the perfectly observed rates are the same as those in Duffee (2002).

The log-likelihood for the exactly observed rates is

$$\ln L = \sum_{t=1}^{T-1} \ln f(y_{t+\Delta t}^{6\text{-month}}, y_{t+\Delta t}^{2\text{-year}}, y_{t+\Delta t}^{10\text{-year}} | y_t^{6\text{-month}}, y_t^{2\text{-year}}, y_t^{10\text{-year}}), \quad (5)$$

where $y_{t+\Delta t}^{6\text{-month}}$, $y_{t+\Delta t}^{2\text{-year}}$, and $y_{t+\Delta t}^{10\text{-year}}$ are the yields of the 6-month, 2-year and 10-year zero-coupon bonds at time $t + \Delta t$. The function $f(y_{t+\Delta t}^{6\text{-month}}, y_{t+\Delta t}^{2\text{-year}}, y_{t+\Delta t}^{10\text{-year}} | y_t^{6\text{-month}}, y_t^{2\text{-year}}, y_t^{10\text{-year}})$ is the density for $y_{t+\Delta t}^{6\text{-month}}$, $y_{t+\Delta t}^{2\text{-year}}$, $y_{t+\Delta t}^{10\text{-year}}$ conditional on $y_t^{6\text{-month}}$, $y_t^{2\text{-year}}$, $y_t^{10\text{-year}}$, and T is the number of term structure observations. Herein T is equal to 504 and Δt is equal to one month.

Through a change of variable, the conditional density $f(\cdot | \cdot)$ of the exactly observed yields can be written as

$$f(y_{t+\Delta t}^{6\text{-month}}, y_{t+\Delta t}^{2\text{-year}}, y_{t+\Delta t}^{10\text{-year}} | y_t^{6\text{-month}}, y_t^{2\text{-year}}, y_t^{10\text{-year}}) = |J| \times f(X_{t+\Delta t} | X_t), \quad (6)$$

Table 2
Likelihood ratio test of nested models

	<i>EA</i> ₀ (3)	<i>EA</i> ₁ (3)	<i>SAS-R</i> ₁ (3)	<i>CA</i> ₂ (3)	<i>SAS-R</i> ₂ (3)	<i>CIR</i>	<i>SAS-R</i> ₃ (3)
Log-likelihood	14771.7	14966.5	14968.8	14971.3	14973.9	14668.5	14673.6
$2 \times (\text{log-likelihood difference})$			4.6		5.1		10.2
$p\text{-value } \chi^2[1]$			3.2%		2.3%		0.1%

P-values of the likelihood ratio test of the SAS-R model against the corresponding completely and essentially affine models. The tests do not reject the null hypotheses of the SAS-R model at usual significance levels.

where $|J|$ is the Jacobian of the transformation of the yields $y_t^{6\text{-month}}$, $y_t^{2\text{-year}}$, $y_t^{10\text{-year}}$ to the state variables $X_{1,t}$, $X_{2,t}$, $X_{3,t}$, and $f(X_{t+\Delta t} | X_t)$ is the density for $X_{t+\Delta t}$ conditional on X_t .

The yields of the 3-month, 1-year, and 5-year bonds are assumed to have measurement errors, which are i.i.d. normal with mean zero and possibly a nonzero correlation. The choice of normally distributed errors is made for simplicity, as in Chen and Scott (1993). In principle, all yields could be measured with errors. Nevertheless, as noted by Duffie and Singleton (1997), the approach taken here has advantages for pricing because it forces the model to perfectly fit some yields.

The log-likelihood function is the sum of the log-likelihood of the exactly observed rates as given by Equation (5) and the log-likelihood of the model disturbances. The evaluation of the density $f(X_{t+\Delta t} | X_t)$ is made using the quasi-likelihood method. For a detailed description of the estimation method, see Appendix A.2.

2.3 Estimation results

The estimated parameters and their *t*-values are displayed in Table 1. The *t*-values are given in parentheses. The displayed parameters with values zero or one and without *t*-values are restricted. The log-likelihood function for each model is displayed in Table 2. The likelihood ratio tests are performed and they indicate that the null hypotheses of the SAS-R models cannot be rejected at the usual confidence levels. The observed ranges of the state variables are displayed in Table 3.

Term structure movements are usually interpreted as changes in the “level,” “slope,” and “curvature.” In the analyzed models, term structure movements are easily interpreted because yields are given by an affine function of the state variables, that is, $y_t^r = \frac{A(\tau)}{\tau} + \frac{B'(\tau) \times X_t}{\tau}$, and hence a movement in one state variable, $X_{i,t}$, moves the term structure in a way consistent with the factor loading function $\frac{B_i(\tau)}{\tau}$. To control for the fact that some of the estimated models allow for feedback in the drift of the state variables, for instance, the estimated *CA*₂(3) model, I make the following change of state variables: $X^* = (N^Q)^{-1} \times X$, where N^Q is the eigenvector of the matrix κ^Q . The yields of zero-coupon bonds as

Table 3
Range of state variables and prices of risk

		$EA_0(3)$	$EA_1(3)$	$SAS-R_1(3)$	$CA_2(3)$	$SAS-R_2(3)$	CIR	$SAS-R_3(3)$
X_1	Min	-2.184	2.980	2.777	3.877	3.749	0.174	0.182
	Max	3.277	63.730	63.126	47.810	46.712	0.292	0.299
X_2	Min	-1.547	-45.923	-14.102	0.346	0.424	0.020	0.019
	Max	1.578	65.130	95.423	30.716	31.959	0.152	0.150
X_3	Min	-5.202	-4.806	-4.891	-9.284	-3.606	0.007	0.006
	Max	7.025	4.028	4.098	11.023	17.263	0.119	0.118
$\lambda_1(X)$	Min	-3.258	-0.110	-0.511	-0.115	-0.408	-0.511	-0.501
	Max	2.227	-0.024	0.706	-0.033	0.720	-0.395	-0.391
$\lambda_2(X)$	Min	-2.901	96.616	88.909	0.000	0.000	-0.240	-1.279
	Max	2.413	2399.416	2367.673	0.000	0.000	-0.088	1.448
$\lambda_3(X)$	Min	-1.594	-5.371	-5.359	-1.103	-1.147	-0.108	-0.473
	Max	2.013	6.409	6.396	-0.227	-0.238	-0.027	0.688

The maximum and minimum observed values of the state variables and of the price of risk. The SAS-R has a parametrization for the price of risk that allows all the terms of the price of risk vector to change sign. This additional flexibility of the SAS-R model is shown through the signs of the maximum and minimum values of $\lambda_i(X)$.

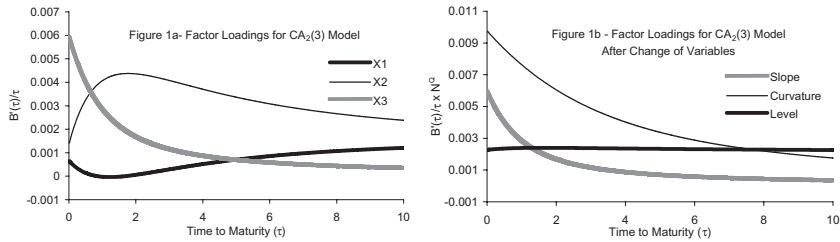


Figure 1

Figure 1a plots the factor loadings $B'(\tau)/\tau$ as a function of the time to maturity τ for the estimated $CA_2(3)$ model. Figure 1b plots the factor loadings $B'(\tau)/\tau \times N^Q$, which are the factor loading functions on the state variables $X^* = (N^Q)^{-1}X$, where N^Q is the eigenvector of κ^Q . Notice that one of the X^* 's has a flat factor loading, which indicates that a change on this state variable changes the level of rates. The factor loading functions for the other estimated models are not plotted herein for reasons of space. However, all estimated models have a state variable X^* that has a flat factor loading function, $B'(\tau)/\tau \times N^Q$. The state variable X^* with a flat factor loading function $B'(\tau)/\tau \times N^Q$ is identified as the “level” state variable.

functions of X^* are $y_t^\tau = \frac{(-A(\tau) + B'(\tau) \times N^Q \times X_t^*)}{\tau}$, and hence the factor loading function on the new state variable $X_{i,t}^*$ is $(\frac{B'(\tau) \times N^Q}{\tau})_i$. Figure 1 plots the factor loading functions for the $CA_2(3)$ model on the original state variables and on the transformed state variables X^* . The factor loading functions on the transformed variables for the other models are similar to the ones plotted for the $CA_2(3)$ model, and for reasons of space are not given here.

An examination of the displayed factor loading function on the transformed state variables X^* reveals that, first, one state variable controls the “slope” of the term structure because a change in the slope state variable greatly affects the difference between the short-term yields and long-term yields; second, another state variable controls the “level” of the term

structure because a change in the level state variable equally affects the yields of all maturities, and hence shocks in the level state variable result in parallel shifts in the term structure; and third, the third state variable is related to shocks in the curvature of the term structure because it explains the movements of the yield of the 5-year bond that are not correlated with the movements of the yield of the 10-year bond and of the short term interest rate. The changes in the yields with longer time to maturity are mostly explained by the level and the curvature state variables in all estimated models.

In the $EA_1(3)$, $SAS-R_1(3)$, $CA_2(3)$, $SAS-R_2(3)$, CIR, and $SAS-R_3(3)$ the level state variable is nonstationary under the risk-neutral measure and it is a linear combination of the state variables that affect the volatility of yields. The diffusion of the state variables X^* under the equivalent martingale measure has drift

$$(N^Q)^{-1} \times (\kappa\theta)^Q - \Lambda^Q X^*, \quad (7)$$

where Λ^Q is a matrix with the eigenvalues of κ^Q . It turns out that for estimated models with stochastic volatility, the level state variable is nonstationary under the equivalent martingale measure as implied by the estimated negative values for eigenvalues of κ^Q (see Table 4). Nonstationary or highly persistent state variables under the equivalent martingale measure are necessary because shocks in the level of the term structure largely affect yields of bonds with a long time to maturity. Hence the effect of shocks in this state variable must subsist for a long time under the pricing measure. In addition, an examination of the matrix $(N^Q)^{-1}$ (see Table 4) reveals that the level state variable is always a linear combination of the state variables X that affect the volatility of yields. Therefore, in all models with stochastic volatilities, the level state variable is closely related to the volatility of the yields.

2.4 The time variability of the term premium

Let $R_{t+\Delta t}^{n+\Delta t}$ be the log return of holding from t to $t + \Delta t$ a zero-coupon bond with time to maturity equal to $n + \Delta t$ years at time t . Let $y_t^{\Delta t}$ represent the annualized yield of a zero-coupon bond with time to maturity equal to Δt . The expected excess return or term premium at time t in the Δt return of the $(n + \Delta t)$ -year bond is defined as

$$E_t[R_{t+\Delta t}^{n+\Delta t}] - \Delta t \times y_t^{\Delta t}. \quad (8)$$

Consider the regression

$$R_{t+\Delta t}^{n+\Delta t} - \Delta t \times y_t^{\Delta t} - (E_t[R_{t+\Delta t}^{n+\Delta t}] - \Delta t \times y_t^{\Delta t}) = \gamma_0 + \gamma_1 s_t + \varepsilon_{t+\Delta t}, \quad (9)$$

where $E_t[R_{t+\Delta t}^{n+\Delta t}] - \Delta t \times y_t^{\Delta t}$ is the term premium inferred by a term structure model and s_t is the difference between the five-year yield and the

Table 4**Eigenvalues, eigenvectors of the matrices K and K^Q with the corresponding state variables half-lives**

	Eigenvalues		Eigenvectors						Half-lives (years)
	d_i	d_i^Q	N^Q			$(N^Q)^{-1}$			
$EA_0(3)$	3.304	1.603	1.679	1.000	0.000	0.000	1.000	0.000	0.21
	0.558	0.558	1.000	0.000	0.000	1.000	-1.679	0.000	1.24
	0.066	0.001	0.291	0.185	1.000	-0.185	0.019	1.000	10.43
$EA_1(3)$	2.911	1.577	0.000	0.000	1.000	0.000	0.000	1.000	0.24
	0.587	0.587	11.062	1.000	0.882	-0.882	1.000	-11.062	1.18
	0.003	-0.011	1.000	0.000	0.000	1.000	0.000	0.000	220
$SAS-R_1(3)$	2.888	1.581	0.000	0.000	1.000	0.000	0.000	1.000	0.23
	0.585	0.585	10.658	1.000	0.881	-0.881	1.000	-10.658	1.13
	0.183	-0.011	1.000	0.000	0.000	1.000	0.000	0.000	8.88
$CA_2(3)$	1.791	1.791	0.000	-0.583	0.893	0.570	-1.452	1.000	0.39
	0.546	0.541	0.000	0.813	0.471	-0.471	0.893	0.000	1.27
	0.003	-0.009	1.000	1.512	0.174	0.813	0.583	0.000	245
$SAS-R_2(3)$	1.763	1.763	0.000	-0.596	0.878	0.610	-1.447	1.000	0.38
	0.130	0.540	0.000	0.803	0.495	-0.495	0.878	0.000	1.16
	0.631	-0.009	1.000	1.525	0.181	0.803	0.5962	0.000	11.77
CIR	2.945	2.879	1.000	0.000	0.000	1.000	0.000	0.000	0.24
	0.455	0.399	0.000	1.000	0.000	0.000	0.000	0.000	1.52
	9.E-06	-0.014	0.000	0.000	1.000	0.000	1.000	1.000	78,588
$SAS-R_3(3)$	2.924	2.861	1.000	0.000	0.000	1.000	0.000	0.000	0.22
	1.424	0.396	0.000	1.000	0.000	0.000	1.000	0.000	0.72
	0.180	-0.014	0.000	0.000	1.000	0.000	0.000	1.000	9.37

Eigenvalues of $\kappa(d_i)$ and $\kappa^Q(d_i^Q)$, the matrix of eigenvectors of $\kappa^Q(N^Q)$ and its inverse $(N^Q)^{-1}$ and the half-lives of the state variables under the physical probability measure. The drift of the SAS-R model is nonlinear, therefore there is no simple expression for the half-lives of the state variables. For details on the calculation of the half-lives in the SAS-R model (see Section 1.1).

six-month yield, that is, s_t is a measure of the slope of the term structure. If a term structure model matches the time variability of the term premium, then the coefficients γ_0 and γ_1 in Equation (9) should not be statistically different from zero. Indeed, a term structure model that matches the time variability of the term premium should embody all the information available at time t useful to forecast the excess returns of zero-coupon bonds.

The excess return of a zero-coupon bond is by definition

$$R_{t+\Delta t}^{n+\Delta t} - \Delta t \times y_t^{\Delta t} = (n + \Delta t) \times y_{t+\Delta t}^{n+\Delta t} - n \times y_{t+\Delta t}^n - \Delta t \times y_t^{\Delta t}, \quad (10)$$

and hence, Equation (9) has the same information as the regression

$$y_{t+\Delta t}^n - E_t[y_{t+\Delta t}^n] = \alpha_0 + \alpha_1 \times s_t + \varepsilon_{t+\Delta t}, \quad (11)$$

where $E_t[y_{t+\Delta t}^n]$ is the expected value of $y_{t+\Delta t}^n$, conditional on the information at time t calculated using a term structure model. If the coefficient α_1 is statistically different from zero in Equation (11), then there is evidence that the time variability of the term premium is not matched by the analyzed term structure model.

To analyze whether the SAS-R models and the corresponding essentially affine models match the time variability of the term premium,

Table 5
Square root of the mean squared error of each model (%)

	$EA_0(3)$	$EA_1(3)$	$SAS-R_1(3)$	$CA_2(3)$	$SAS-R_2(3)$	CIR	$SAS-R_3(3)$
In-sample (1952–1993)							
y_{t+6m}^n	1.3450	1.3709	1.3604	1.4350	1.4336	1.4811	1.4555
y_{t+5y}^n	0.8348	0.8874	0.8749	0.9163	0.9112	0.9141	0.9088
y_{t+10y}^n	0.7079	0.7579	0.7443	0.7741	0.7648	0.7789	0.7718
Out-of-sample (1994–1998)							
y_{t+6m}^n	0.6128	0.5251	0.5311	0.6277	0.6512	0.6139	0.5522
y_{t+5y}^n	0.7858	0.7184	0.7325	0.8204	0.8620	0.7965	0.7883
y_{t+10y}^n	0.7256	0.6717	0.6879	0.7308	0.7830	0.7139	0.7204
Out-of-sample (1995–1998)							
y_{t+6m}^n	0.4095	0.3922	0.4091	0.5506	0.5340	0.5141	0.5120
y_{t+5y}^n	0.6961	0.6518	0.6764	0.7624	0.7796	0.7267	0.7485
y_{t+10y}^n	0.6585	0.6279	0.6542	0.6881	0.7170	0.6674	0.6931

The root mean squared error (RMSE) is the squareroot of the mean squared forecasting error $(y_{t+\Delta t}^n - E_t[y_{t+\Delta t}^n])^2$, where the conditional expectations, $E_t[y_{t+\Delta t}^n]$, are calculated with the parameters displayed in Table 1. The displayed values are percentages. The forecasting period (Δt) is six months. There is a clear increasing pattern in the RMSE's from the left to the right, indicating that models with time-varying yield volatilities have more difficulty in forecasting future yields. This is the tension between matching the first and second conditional moments of yields that has been described in the literature [see, for instance, Dai and Singleton (2002)].

Equation (11) with Δt equal to six months and n equal to 6 months, 5 years, and 10 years is used. The SAS-R and the essentially affine models, conditional expectations $E_t[y_{t+\Delta t}^n]$ are calculated with the parameters estimated for each model and displayed in Table 1. The conditional expectations in the SAS-R model do not have a closed-form solution, and hence they are calculated with Monte Carlo simulation. See Appendix 4.2 for details.

Table 5 displays the square root of the mean square forecasting error, $(y_{t+\Delta t}^n - E_t[y_{t+\Delta t}^n])^2$, of all estimated models. Table 6 displays the results of Equation (11). The p -values indicate that the SAS-R models match the time variability of the term premium better than the corresponding essentially and completely affine models. The semiaffine models not only capture the information on the slope of the term structure better than essentially affine models, but they also produce better in-sample forecasts of future yield changes, as indicated by the squareroot of the mean square error of the forecasts.

Even though the results indicate that the semiaffine extension improves the matching of the affine models to the time-varying term premium, the tension between matching the first and second moments of the data remains. The model with no stochastic volatility, $EA_0(3)$, performs better than all the models with stochastic volatility in terms of matching the expected changes in yields. The $EA_0(3)$ model not only captures all the information on the slope of the term structure but also produces better forecasts of changes in yields. The problem with the $EA_0(3)$ model is that by construction it does not produce any time variation on the volatility of

Table 6
Regression of the forecasting residuals on the slope of the term structure

	<i>EA</i> ₀ (3)	<i>EA</i> ₁ (3)	<i>SAS-R</i> ₁ (3)	<i>CA</i> ₂ (3)	<i>SAS-R</i> ₂ (3)	<i>CIR</i>	<i>SAS-R</i> ₃ (3)
<i>y</i> ^{6-month}							
Adj. <i>R</i> ²	−0.0018	0.0343	0.0301	0.0605	0.0592	0.0831	0.1497
<i>α</i> ₀	−0.0001 (0.0017)	0.0033 (0.0017)	0.0029 (0.0017)	0.0028 (0.0017)	0.0030 (0.0017)	0.0046 (0.0018)	0.0054 (0.0017)
<i>α</i> ₁	0.0157 (0.1223)	−0.2510 (0.1210)	−0.2344 (0.1191)	−0.3450 (0.1231)	−0.3414 (0.1233)	−0.4158 (0.1289)	−0.5459 (0.1181)
<i>p</i> -value	0.4490	0.0193	0.0248	0.0026	0.0029	0.0007	0.0000
<i>y</i> ^{5-year}							
Adj. <i>R</i> ²	−0.0010	0.1087	0.0958	0.1346	0.1283	0.1240	0.1684
<i>α</i> ₀	0.0004 (0.0009)	0.0040 (0.0009)	0.0035 (0.0009)	0.0031 (0.0009)	0.0028 (0.0009)	0.0037 (0.0009)	0.0037 (0.0009)
<i>α</i> ₁	−0.0248 (0.0744)	−0.2814 (0.0747)	−0.2620 (0.0750)	−0.3261 (0.0737)	−0.3166 (0.0733)	−0.3118 (0.0739)	−0.3612 (0.0733)
<i>p</i> -value	0.3695	0.0001	0.0003	0.0000	0.0000	0.0000	0.0000
<i>y</i> ^{10-year}							
Adj. <i>R</i> ²	−0.0002	0.1405	0.1240	0.1599	0.1480	0.1677	0.1904
<i>α</i> ₀	0.0004 (0.0007)	0.0036 (0.0007)	0.0030 (0.0007)	0.0031 (0.0007)	0.0026 (0.0007)	0.0036 (0.0007)	0.0033 (0.0007)
<i>α</i> ₁	−0.0290 (0.0615)	−0.2736 (0.0615)	−0.2538 (0.0620)	−0.2999 (0.0611)	−0.2850 (0.0607)	−0.3081 (0.0612)	−0.3260 (0.0614)
<i>p</i> -value	0.3187	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

The models are tested through the in-sample regression

$$y^n_{t+\Delta t} - E_t[y^n_{t+\Delta t}] = \alpha_0 + \alpha_1 \times s_t + \varepsilon_{t+\Delta t},$$

where y^n_t is the yield of a zero-coupon bond maturing at time $t + n$. The conditional expectations, $E_t[y^n_{t+\Delta t}]$, are calculated with the parameters displayed in Table 1. The difference between the five-year yield and the three-month yield is represented by s_t . The forecasting period (Δt) is six months. The reported p -values are for a one-tailed test. If a model matches the time variability of the term premium, then α_1 should not be statistically different from zero. The variances of $\varepsilon_{t+\Delta t}$ are assumed constant in the above ordinary least squares (OLS) regression. The standard errors are displayed between parentheses. The standard errors are corrected for heteroscedasticity and autocorrelated residuals by the Newey and West estimator with nine lags. The null hypothesis that $\alpha_1 = 0$ is rejected in all models, but *EA*₀(3) at a 5% confidence level. There is a clear decreasing pattern in the p -values from the left to the right, indicating that models with time-varying yield volatilities have more difficulty in forecasting future yields. The SAS-R extension contributes to solving this tension because the SAS-R models capture more of the information on the slope of the curve.

yields and hence it does not match one of the stylized facts of the term structure literature.

To a certain extent we should expect models with stochastic volatility to perform worse than homoscedastic models in the Equation (11). Models with stochastic volatility are required to match not only the expected changes in yields but also the conditional variances of yields, while the homoscedastic model, *EA*₀(3), has to match only the expected change in yields. Under this point of view, the relevant question is whether the stochastic volatility models are using all the information in the term structure to produce forecasts.

To test if the stochastic volatility models are incorporating all the information available at time t to produce forecasts, I run a regression

similar to Equation (11):

$$y_{t+\Delta t}^n - E_t[y_{t+\Delta t}^n] = \alpha_0 + \alpha_1 \times s_t + \varepsilon_{t+\Delta t}. \quad (12)$$

However, as opposed to the Equation (11), the variances of $\varepsilon_{t+\Delta t}$ are not assumed constant. The variances of $\varepsilon_{t+\Delta t}$ are assumed to be equal to $\sigma^2 V_t$, where σ^2 is constant and V_t is the conditional variance of the $y_{t+\Delta t}^n$ calculated by each model. If a model incorporates all the information available at time t to forecast the changes in yields and their variances, then α_0 and α_1 should be equal to zero and σ^2 should be equal to one. Equation (12) is estimated by weighted least squares, the estimation results are displayed in Table 7. The null hypothesis that $\alpha_1 = 0$ in Equation (12) is not rejected as often as in the Equation (11). Thus there is evidence that part of the reported failure of affine models with stochastic volatilities in forecasting changes on yields is due to the fact that the test based on Equation (11) does not take into account all the information provided by the models.

Even though Equation (12) does not reject the affine models with stochastic volatilities as often as Equation (11), the results of Equation (12) indicate that the mean-volatility tension that has been described in the literature is still present. In the case of the $EA_1(3)$ model and yield $y^{6\text{-month}}$, the tension is present because the yield change forecast is worse than in the $EA_0(3)$ model. However, Equation (12) gives evidence that no information is missed by the model because, as indicated by the P -values, we cannot reject the assumption that $\alpha_1 = 0$ at usual confidence levels. Therefore, in this case, the mean-volatility tension happens because the information available at time t is shared to calculate the conditional mean and variances of yields. In the case of the yields with time to maturity longer than 6 months, the tension is present not only because the yield forecast is worse than in the $EA_0(3)$ model, but also because some information is missed by the models with stochastic volatilities, as indicated by the rejection of the assumption that $\alpha_1 = 0$ at the usual confidence levels.

The out-sample results of Equation (11) are displayed in Tables 8 and 9. Two out-sample periods are analyzed, one is from January 1994 to December 1998 and the other is from January 1995 to December 1998. The results displayed in Tables 5, 8, and 9 indicate that the out-sample results are less clear than the in-sample results in relation to the relative performance of the examined models. In some cases, the SAS-R model outperforms the corresponding essentially affine models, in other cases the SAS-R does not outperform the corresponding essentially affine models.

It is interesting to notice how the inclusion of the year 1994 changes the out-sample results. There were a series of rate increases in 1994 [see Campbell (1995)]. These rate increases make the statistical relationship between future changes in rates and the slope of the term structure

Table 7**Regression of the forecasting residuals on the slope of the term structure considering the heteroscedasticity of the residuals forecasted by each model**

	$EA_0(3)$	$EA_1(3)$	$SAS-R_1(3)$	$CA_2(3)$	$SAS-R_2(3)$	CIR	$SAS-R_3(3)$
$y^{6-month}$							
Adj. R^2	-0.0018	0.0281	0.0249	0.0535	0.0523	0.0816	0.1495
α_0	-0.0001 (0.0017)	0.0028 (0.0017)	0.0021 (0.0017)	0.0025 (0.0018)	0.0027 (0.0018)	0.0053 (0.0018)	0.0057 (0.0017)
α_1	0.0157 (0.1223)	-0.1555 (0.1251)	-0.1413 (0.1230)	-0.2502 (0.1280)	-0.2466 (0.1282)	-0.4217 (0.1301)	-0.5534 (0.1182)
p -value	0.4490	0.1072	0.1256	0.0256	0.0275	0.0006	0.0000
y^{5-year}							
Adj. R^2	-0.0010	0.1023	0.0890	0.1260	0.1211	0.1226	0.1678
α_0	0.0004 (0.009)	0.0036 (0.0009)	0.0026 (0.0010)	0.0030 (0.0009)	0.0025 (0.0009)	0.0038 (0.0009)	0.0034 (0.0009)
α_1	-0.0248 (0.0744)	-0.2159 (0.0787)	-0.1976 (0.0793)	-0.2603 (0.0787)	-0.2496 (0.0782)	-0.2945 (0.0753)	-0.3418 (0.0744)
p -value	0.3695	0.0031	0.0065	0.0005	0.0008	0.0001	0.0000
$y^{10-year}$							
Adj. R^2	-0.0002	0.1341	0.1156	0.1534	0.1423	0.1645	0.1869
α_0	0.0004 (0.0007)	0.0031 (0.0007)	0.0021 (0.0008)	0.0028 (0.0007)	0.0020 (0.0008)	0.0035 (0.0007)	0.0027 (0.0008)
α_1	-0.0290 (0.0615)	-0.2168 (0.0653)	-0.1976 (0.0669)	-0.2448 (0.0651)	-0.2297 (0.0650)	-0.2708 (0.0637)	-0.2880 (0.0646)
p -value	0.3187	0.0005	0.0016	0.0001	0.0002	0.0000	0.0000

Similar to the Table 6 regression, the models are tested through the in-sample regression

$$y_{t+\Delta t}^n - E_t[y_{t+\Delta t}^n] = \alpha_0 + \alpha_1 \times s_t + \varepsilon_{t+\Delta t},$$

where y_t^n is the yield of a zero-coupon bond maturing at time $t + n$. The conditional expectations, $E_t[y_{t+\Delta t}^n]$, are calculated with the parameters displayed in Table 1. The difference between the five-year yield and the three-month yield is represented by s_t . The forecasting period (Δt) is six months. As opposed to the regression in Table 6, the variances of $\varepsilon_{t+\Delta t}$ are not assumed constant, but $\varepsilon_{t+\Delta t}$ is assumed to have variance equal to $\sigma^2 V_t$, where σ^2 is constant and V_t is the conditional variance of the $y_{t+\Delta t}^n$ calculated by each model. This regression is estimated by weighted least squares. The standard errors are displayed between parentheses. The standard errors are corrected for autocorrelated residuals by Newey and West estimators with nine lags and Bartlett kernel. The results do not change qualitatively if Newey and West estimator with five lags and a rectangular kernel is used. The reported p -values are for a one-tailed test. If a model matches the time variability of the term premium, then α_1 should be statistically not different from zero. The null hypothesis that $\alpha_1 = 0$ is not rejected as often as in the regression displayed in Table 6 and hence there is evidence that part of the reported failure of affine models with stochastic volatilities is due to the fact that the test on Table 6 does not take into account all the information provided by the models.

between 1994 and 1998 different from the usual relationship, and thus the estimated α_1 's in the 1994–1998 period are substantially different from those estimated in the 1952–1993 period. If the year 1994 is not included in the out-sample analysis, then the usual statistical relationship between future changes in rates and the slope of the term structure holds, and the out-sample results are qualitatively similar with the in-sample results.

2.5 Intuition for improvement caused by the semiaffine models

The SAS-R model produces higher variability of the term premium than the corresponding affine models because the price of risk in all the elements of the price of risk vector in the SAS-R model can change sign. The

Table 8
Out-of-sample regression of the forecasting residuals on the slope of the term structure out-of-sample period starting in 1994

	$EA_0(3)$	$EA_1(3)$	$SAS-R_1(3)$	$CA(3)$	$SAS-R(3)$	CIR	$SAS-R_3(3)$
$y^{6-month}$							
Adj. R^2	0.3816	0.1313	0.1525	0.0101	0.0525	-0.0154	-0.0192
α_0	-0.0061 (0.0017)	-0.0033 (0.0018)	-0.0039 (0.0018)	-0.0037 (0.0023)	-0.0039 (0.0022)	-0.0016 (0.0024)	-0.0022 (0.0019)
α_1	0.5263 (0.1698)	0.2764 (0.1772)	0.2973 (0.1764)	0.1373 (0.2297)	0.2296 (0.2304)	0.0511 (0.2396)	-0.0015 (0.1913)
p -value	0.0016	0.0624	0.0489	0.2763	0.1616	0.4160	0.4969
y^{5-year}							
Adj. R^2	0.1144	0.0050	0.0146	-0.0147	0.0137	-0.0131	-0.0136
α_0	-0.0063 (0.0032)	-0.0032 (0.0031)	-0.0041 (0.0031)	-0.0039 (0.0035)	-0.0050 (0.0034)	-0.0033 (0.0035)	-0.0042 (0.0033)
α_1	0.3780 (0.2673)	0.1483 (0.2590)	0.1752 (0.2583)	0.0690 (0.2937)	0.1979 (0.3050)	0.0809 (0.2942)	0.0723 (0.2709)
p -value	0.0816	0.2847	0.2503	0.4076	0.2595	0.3922	0.3953
$y^{10-year}$							
Adj. R^2	0.0584	-0.0161	-0.0113	-0.0192	0.0002	-0.0192	-0.0192
α_0	-0.0050 (0.0033)	-0.0023 (0.0031)	-0.0033 (0.0032)	-0.0027 (0.0033)	-0.0040 (0.0033)	-0.0022 (0.0033)	-0.0031 (0.0032)
α_1	0.2646 (0.2658)	0.0494 (0.2522)	0.0778 (0.2513)	0.0045 (0.2711)	0.1364 (0.2863)	-0.0019 (0.2695)	0.0071 (0.2561)
p -value	0.1620	0.4227	0.3790	0.4934	0.3178	0.4972	0.4890

The models are tested through the out-sample regression

$$y_{t+\Delta t}^n - E_t[y_{t+\Delta t}^n] = \alpha_0 + \alpha_1 \times s_t + e_{t+\Delta t},$$

where y_t^n is the yield of a zero-coupon bond maturing at time $t+n$. The conditional expectations, $E_t[y_{t+\Delta t}^n]$, are calculated with the parameters displayed in Table 1. The difference between the five-year yield and the three-month yield is represented by s_t . The forecasting period (Δt) is six months. The reported p -values are for a one-tailed test. If a model matches the time variability of the term premium, then α_1 should not be statistically different from zero. The standard errors are displayed in parentheses. Standard errors are corrected for heteroscedasticity and autocorrelated residuals by the Newey and West estimator with nine lags. This table displays the results of this regression using data from January 1994 to December 1998.

instantaneous term premium in the SAS-R model is

$$\mu_P - r = -B(\tau)' \times \Sigma \sqrt{S_t} \times (\Sigma^{-1} \lambda_0 + \sqrt{S_t} \lambda_1 + \sqrt{S_t} \lambda_2 X_t). \quad (13)$$

While the instantaneous term premium in the essentially affine models is

$$\mu_P - r = -B(\tau)' \times \Sigma \sqrt{S_t} \times (\sqrt{S_t} \lambda_1 + \sqrt{S_t} \lambda_2 X_t). \quad (14)$$

The $EA_1(3)$, $CA_2(3)$ and CIR , models do not produce a high time variability of the term premium because not all the individual elements of the price of risk vector, that is, the individual elements of the vector $\sqrt{S_t} \lambda_1 + \sqrt{S_t} \lambda_2 X_t$ in Equation (14), can change sign and they are very close to zero. Consequently changes in the state variables $X_{i,t}$ cause small changes in the instantaneous term premium $\mu_P - r$. The estimated SAS-R models produce higher time variability of the term premium because all the individual elements in the price of risk vector, that is, the individual elements of the vector $\Sigma^{-1} \lambda_0 + \sqrt{S_t} \lambda_1 + \sqrt{S_t} \lambda_2 X_t$ in Equation (13) can change sign. Consequently changes in the state variables $X_{i,t}$ cause

Table 9

Out-of-sample regression of the forecasting residuals on the slope of the term structure out-of-sample period starting in 1995

	$EA_0(3)$	$EA_1(3)$	$SAS-R_1(3)$	$CA_2(3)$	$SAS-R_2(3)$	CIR	$SAS-R_3(3)$
$y^{6\text{-month}}$							
Adj. R^2	-0.0250	0.0845	0.0687	0.2531	0.2501	0.3299	0.3443
α_0	-0.0022 (0.0016)	0.0008 (0.0017)	0.0001 (0.0017)	0.0014 (0.0022)	0.0016 (0.0022)	0.0037 (0.0022)	0.0023 (0.0018)
α_1	-0.0003 (0.1369)	-0.2816 (0.1377)	-0.2578 (0.1393)	-0.5539 (0.1819)	-0.5496 (0.1815)	-0.6710 (0.1826)	-0.6121 (0.1437)
p -value	0.4991	0.0237	0.0357	0.0020	0.0021	0.0003	0.0001
$y^{5\text{-year}}$							
Adj. R^2	-0.0016	0.0701	0.0583	0.1296	0.1251	0.1245	0.1132
α_0	-0.0019 (0.0037)	0.0011 (0.0035)	0.0002 (0.0036)	0.0011 (0.0039)	0.0007 (0.0039)	0.0017 (0.0039)	0.0004 (0.0037)
α_1	-0.2140 (0.2982)	-0.4382 (0.2867)	-0.4074 (0.2903)	-0.6042 (0.3144)	-0.5946 (0.3166)	-0.5931 (0.3146)	-0.5474 (0.2993)
p -value	0.2385	0.0670	0.0840	0.0308	0.0337	0.0332	0.0373
$y^{10\text{-year}}$							
Adj. R^2	0.0347	0.1225	0.1068	0.1598	0.1507	0.1621	0.1477
α_0	-0.0006 (0.0038)	0.0019 (0.0036)	0.0010 (0.0036)	0.0019 (0.0038)	0.0012 (0.0038)	0.0024 (0.0038)	0.0012 (0.0037)
α_1	-0.3283 (0.3003)	-0.5253 (0.2838)	-0.4921 (0.2879)	-0.6175 (0.2988)	-0.6002 (0.3021)	-0.6206 (0.2971)	-0.5789 (0.2910)
p -value	0.1403	0.0357	0.0475	0.0226	0.0268	0.0215	0.0267

This table is analogous to Table 8. It displays the results of the out-of-sample analysis using data from January 1995 to December 1998. For a discussion of the reasons behind the effect of excluding the year of 1994, see Section 2.4.

significant changes in the term premium $\mu_P - r$. The observed range of the elements of the price of risk vectors are displayed in Table 3. Notice that the $EA_1(3)$, $CA_2(3)$, and CIR models do not produce the same wide range of values for the elements of the price of risk vector as their corresponding SAS-R models.

It is possible to argue that it is not necessary to build affine models that allow the change in sign of the price of risk in order to explain the time variability of the term premium. A close examination of Equation (13) reveals that the instantaneous term premium could change sign in a multifactor completely affine model where the elements of the vector λ_1 have opposite signs. Therefore a completely affine model could potentially explain the time variability of the term premium. However, this argument does not take into consideration that term structure models must explain not only the term premium variation, but also the relative movements of bonds with different times to maturity.

To see why the argument above fails, take for instance a two-factor CIR model where the yield of a zero-coupon bond with time to maturity τ is given by $y_t^\tau = \frac{(-A(\tau) + \sum_{i=1}^2 B_i(\tau) \times X_{i,t})}{\tau}$ and the instantaneous term premium is

$$\mu_P - r = -B(\tau)' \times \begin{pmatrix} \beta_1 X_{1,t} & 0 \\ 0 & \beta_2 X_{2,t} \end{pmatrix} \times \lambda_1 \quad (15)$$

In the model above, assume that $X_{1,t}$ explains the movements of the level of the term structure and $X_{2,t}$ explains the movement of the slope of the term structure. This is equivalent to saying that for a long time to maturity, $B_2(\tau)$ is small in relation to $B_1(\tau)$, and hence the term premium of a long-term bond is mostly driven by X_1 [see Equation (15)]. Because the first element of the vector $B_1(\tau)' \times \beta_1 X_{1,t} \times \lambda_1(1)$ does not change sign, the instantaneous term premium of a long-term bond will rarely have changes in sign. Therefore this model cannot explain at the same time the term premium variation, that is, the changes in the sign of $\mu_P - r$, and the relative movements of bonds with different time to maturity, that is, $B_2(\tau) \ll B_1(\tau)$ for large τ .

2.6 The mean reversion in the state variables

The estimation of the SAS-R model indicates that some of the state variables are mean reverted, while the estimation of the essentially and completely affine models do not indicate the same degree of mean reversion. For instance, the estimated half-life of the level state variable in the $EA_1(3)$ model is 220 years, while the approximated half-life of this state variable in the $SAS-R_1(3)$ model is 8.88 years.

This limitation of the essentially affine models is clarified in Figure 2. The estimated drift functions of the level state variable of the estimated $EA_1(3)$ and $SAS-R_1(3)$ models are represented in Figure 2. The drift functions under the equivalent martingale measure are also displayed.

The drift functions under the equivalent martingale measure have a positive slope because the state variable driving the level of the term structure is nonstationary under the equivalent martingale measure. The nonstationary state variable under the equivalent martingale measure

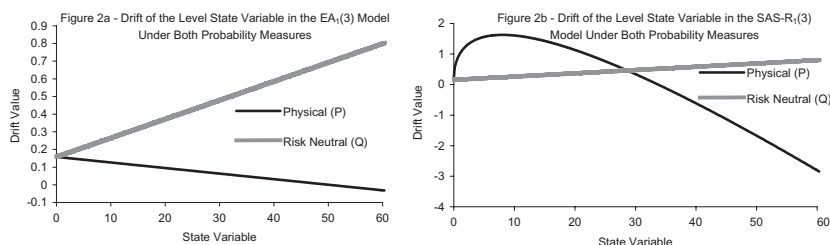


Figure 2

Figure 2a and 2b plot the drift (under the physical and the equivalent martingale measures) of level state variable in the $EA_1(3)$ and in the $SAS-R_1(3)$ models. In Figures 2a and 2b, the drift of the level state variable has a positive slope under the equivalent martingale measure. Figure 2a shows the limitation of the $EA_1(3)$ model in terms of probability measures, it indicates that, in the $EA_1(3)$ model, the drift functions under the different probability measures can intercept only at the origin. Figure 2a also points that the non-stationarity of the level state variable under the equivalent martingale measure coupled with the limitation that the $EA_1(3)$ price of risk does not allow the $EA_1(3)$ model to match the degree of mean reversion present in the data. Figure 2b indicates that the $SAS-R_1(3)$ model solves the limitation of the $EA_1(3)$ model by allowing the drift functions under different probability measures to intercept at more than one point.

attached to the limitation of the price of risk in the essentially affine model results in a very slow mean reversion. The $EA_1(3)$ model restricts the first element of the price of risk vector to be zero only at the origin, and hence, in the $EA_1(3)$ model, the drift functions under both measures can intercept only at the origin (see Figure 2a). The $EA_1(3)$ model would match the mean reversion in the data if the drift of the level state variable were positive for small values of the state variable and negative for large values of the state variable, and hence, to match the mean reversion in the data the drift functions would have to intercept at some intermediate value of the state variable. However, the $EA_1(3)$ restrictions do not allow the existence of such an intermediate intercept point, and hence the $EA_1(3)$ model cannot match the mean reversion present in the data.

The mean reversion of the level state variable in the $SAS-R_1(3)$ model indicates that the very strong persistence of the level state variable in the $EA_1(3)$ model is partially caused by the limitation of the price of risk parametrization. The $EA_1(3)$ model cannot detect the change in the mean reversion of the level because the level state variable in all models with stochastic volatility is related to the volatility of yields, and thus the level state variable in the $EA_1(3)$ model, is equal to the first element of the vector X_t . In the $EA_1(3)$ model, this state variable cannot have a general specification of the price of risk, while in the $SAS-R_1(3)$ it can. Since in the $SAS-R_1(3)$ model the price of risk can change sign, the drift of the state variables under the equivalent martingale measure and under the physical probability measure can intercept at more than one point (Figure 2b). This additional flexibility of the $SAS-R_1(3)$ model allows it to let the level state variable be nonstationary under the equivalent martingale measure and at the same time match the degree of mean reversion present in the data.

The difference in mean reversion is present not only between the $EA_1(3)$ and $SAS-R_1(3)$ models, but also between other estimated models. The estimated half-lives are displayed in Table 4 and they indicate that interest rates are persistent in the essentially affine and in the semiaffine models. However part of the persistence of the state variables in the essentially affine models is caused by its restricted parametrization for the price of risk. It is also interesting to notice that all semiaffine models have half-lives similar to the half-lives of the $EA_0(3)$ model, which is the model with the best forecasting performance.

3. Conclusion

The results in this article indicate that the parametrization of the price of risk has important implications for affine term structure models. First, richer parametrizations for the price of risk, such as the one in the SAS-R model, help affine models produce better term premium forecasts. The improvement in explaining the time variation of the term premium to the

corresponding essentially affine models is caused by the lack of flexibility of the essentially affine models in generating enough variation of all elements of the price of risk vector.

Second, the result common to previous empirical studies, indicating that some of the state variables underlying the term structure of interest rates have a large persistence, may have been partially caused by restrictions in the price of risk. Consequently some of the relationships between the term structure of interest rates and macrovariables derived in these previous studies may have been skewed by the price of risk restrictions.

The results also indicate that the tension between matching the variability of the term premium and matching the variability of the volatilities of yields cannot be completely solved by the examined parametrization of the price of risk. All the estimated models with stochastic volatility performed worse than a model with homoscedastic interest rates in forecasting changes in yields. Some of this trade-off is attributed to the fact that, by construction, stochastic volatility models are required to match the first and second moments of the data, while homoscedastic models are not. However, even after taking into account the time-varying conditional volatility of the yields calculated by stochastic volatility models, stochastic volatility models do not seem to incorporate all the information in the slope of the term structure.

Appendix

A.1 The SAS-R model is arbitrage free

The purpose of this section is to provide a guideline for proof that the SAS-R model satisfies sufficient conditions for an arbitrage-free economy. For this purpose, details are omitted but they can be found in the references cited herein.

Let the state variable X be an n -dimensional $Itô$ process on the filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ with time set $[0, T]$. Let X be defined by some starting point x_0 and by the stochastic differential equation (SDE),

$$dX_t = (\kappa\theta + \Sigma\sqrt{S_t}\Sigma^{-1}\lambda_0 - \kappa X_t) dt + \Sigma\sqrt{S_t}dW_t^P; \quad 0 \leq t \leq T, \quad (16)$$

where W_t^P are standard independent Brownian motions under the physical measure P . The parameters κ , θ , Σ , λ_0 , and S_t are defined in Equations (2–4).

Proposition 1. Equation (16) has a weak solution with $X_0 = x_0$, if for all i :

1. For all x such that $\alpha_i + \beta'_i x = 0$, $\beta'_i(\theta - \kappa \cdot x) > \beta'_i \Sigma' \Sigma \beta_i / 2$.
2. For all j , if $(\beta'_i \Sigma)_j \neq 0$, then $\alpha_i + \beta'_i x = \alpha_j + \beta'_j x$.

Proof of Proposition. Under Equations (1) and (2), Duffie and Kan (1996) show that the following SDE has a unique strong solution:

$$dX_t = (\kappa\theta - \kappa X_t) dt + \Sigma\sqrt{S_t}dW_t^{PAUX}; \quad 0 \leq t \leq T, \quad (17)$$

where $W_t^{P_{AUX}}$ are standard independent Brownian motions under an auxiliary probability measure P_{AUX} . Since $E_t^{P_{AUX}}[\exp(\frac{1}{2}\int_0^T \Sigma^{-1} \lambda_0 \cdot \Sigma^{-1} \lambda_0 ds)] < \infty$, the process

$$\xi_t = \exp \left[\int_0^t \Sigma^{-1} \lambda_0 dW_s^{P_{AUX}} - \frac{1}{2} \int_0^t \Sigma^{-1} \lambda_0 \cdot \Sigma^{-1} \lambda_0 ds \right] \quad (18)$$

is a martingale. Consequently Girsanov's theorem implies that, under the probability measure P given by $(dP/dP_{AUX}) = \xi_t$, the process $W_t^P = W_t^{P_{AUX}} - \int_0^t \Sigma^{-1} \lambda_0 ds$; $0 \leq t \leq T$ is a Brownian motion with $W_0^P = 0$. Simply rewriting Equation (17) as

$$dX_t = (\kappa\theta + \Sigma\sqrt{S_t}\Sigma^{-1}\lambda_0 - \kappa X_t) dt + \Sigma\sqrt{S_t}dW_t^P; \quad 0 \leq t \leq T, \quad (19)$$

we see that (X, W^P) , $(\Omega, \mathcal{F}, \mathbf{F}, P)$ is a weak solution to Equation (16).

Proposition 2. *There exists a probability measure Q on $(\Omega, \mathcal{F})$ that is an equivalent martingale measure to P and under which the latent variables X_t follow the Itô process represented by the SDE:*

$$dX_t = ((\kappa\theta)^Q - \kappa^Q X_t) dt + \Sigma\sqrt{S_t}dW_t^Q, \quad (20)$$

where W_t^Q are standard independent Brownian motions under the measure Q , and κ^Q and θ^Q are defined in Equation (2).

Proof of Proposition. Since $E_t^P[\exp(\frac{1}{2}\int_0^T \Sigma^{-1} \lambda_0 \cdot \Sigma^{-1} \lambda_0 ds)] < \infty$, the process

$$\xi_t = \exp \left[- \int_0^t \Sigma^{-1} \lambda_0 dW_s^P - \frac{1}{2} \int_0^t \Sigma^{-1} \lambda_0 \cdot \Sigma^{-1} \lambda_0 ds \right] \quad (21)$$

is a martingale, and hence, Girsanov's theorem implies that, under the auxiliary probability measure P_{AUX} given by $(dP_{AUX}/dP) = \xi_t$, the process $W_t^{P_{AUX}} = W_t^P + \int_0^t \Sigma^{-1} \lambda_0 ds$; $0 \leq t \leq T$ is a Brownian motion with $W_0^{P_{AUX}} = 0$. The state variables X follow the process under the probability measure P_{AUX} : $dX_t = ((\kappa\theta)^P - \kappa^P \cdot X_t) dt + \Sigma\sqrt{S_t}dW_t^{P_{AUX}}$; $0 \leq t \leq T$.

From Duffee (2002), it is known that there exists a probability measure Q equivalent to P_{AUX} , under which the state variables follow the process: $dX_t = ((\kappa\theta)^Q - \kappa^Q X_t) dt + \Sigma\sqrt{S_t}dW_t^Q$; $0 \leq t \leq T$ and hence, there exists a probability measure Q on $(\Omega, \mathcal{F})$ that is an equivalent martingale measure to P and under which the latent variables follow the diffusion represented by Equation (20).

Under technical conditions [Duffee (1996)], the existence of an equivalent martingale measure is equivalent to nonarbitrage. Therefore the SAS-R model is arbitrage free. It is interesting to note that under conditions 1 and 2 of Proposition 1, Equation (20) has a unique strong solution. Condition 1 of Proposition 1 can be relaxed further as noted by Dai and Singleton (2000). The proof presented herein can be easily expanded to incorporate the price of risk given by $\lambda(X_t) = \Sigma^{-1} \lambda_0(X_t) + \sqrt{S_t} \lambda_1 + \sqrt{S_t} \lambda_2 X_t$, where $\lambda_0(X_t)$ is a strictly bounded function of X .

A.2 The quasi-maximum-likelihood procedure and the calculation of conditional expectations

Quasi-maximum-likelihood (QML) is used to estimate all models. In the case of the completely and essentially affine models, the QML method is the one proposed by Fisher and Gilles (1996) and Duffee (2002). In this method, the density for $X_{t+\Delta t}$ conditional on X_t is assumed normal with mean $E[X_{t+\Delta t} | X_t]$ and covariance $\text{var}[X_{t+\Delta t} | X_t]$.

Assume that the matrix κ has eigenvalue decomposition $\kappa = N \times \Lambda \times N^{-1}$, where N is the eigenvector matrix of κ and Λ is a diagonal matrix with elements given by the eigenvalues of

$\kappa, (d_1, d_2, \dots, d_n)$. The conditional expectation $E[X_{t+\Delta t} | X_t]$ is given by

$$E[X_{t+\Delta t} | X_t] = (I - e^{-\kappa\Delta t})\theta + e^{-\kappa\Delta t}X_t, \quad (22)$$

where $e^{-\kappa\Delta t} = N \times e^{-\Lambda\Delta t} \times N^{-1}$ and $e^{-\Lambda\Delta t}$ is the diagonal matrix with the i th diagonal element given by $e^{-d_i\Delta t}$. The covariance matrix $\text{var}[X_{t+\Delta t} | X_t]$ calculation involves three steps:

1. Define the state variable X_t^* as $N^{-1} \times X_t$. The dynamics of the state variable X_t^* is $dX_t^* = \Lambda(\theta^* - X_t^*)dt + \Sigma^* S_t^* dW_t^P$, where $S_{t(i,i)} = \sqrt{\alpha_i + \beta_i^* X_t^*}$, $\theta^* = N^{-1}\theta$, $\Sigma^* = N^{-1}\Sigma$ and $\beta^* = \beta N$.
2. Calculate the covariance matrix of $X_{t+\Delta t}^*$ conditional on X_t^* by using the expression

$$\begin{aligned} \text{var}[X_t^* | X_t^*] = & \{(d_j + d_k)^{-1} [G_0]_{j,k} (1 - e^{-(T-t)(d_j + d_k)})\} \\ & \times \sum_{i=1}^n [\theta_i^* \{(d_j + d_k)^{-1} [G_i]_{j,k} (1 - e^{-(T-t)(d_j + d_k)})\}] \\ & + \sum_{i=1}^n [(X_{t,i}^* - \theta_i^*) \{(d_j + d_k - d_i)^{-1} [G_i]_{j,k} \times (e^{-d_i(T-t)} - e^{-(d_j + d_k)(T-t)})\}], \end{aligned} \quad (23)$$

where $\{f(j, k)\}$ denotes the matrix with the element (j, k) given by $f(j, k)$, G_0 is the matrix given by $\sum^* \text{diag}(\alpha^*) \sum^{*t}$, and G_i is the matrix given by $\sum^* \text{diag}(\beta_i^*) \sum^{*t}$.

3. Calculate the covariance matrix of $X_{t+\Delta t}$ conditional on X_t with the expression

$$\text{var}[X_{t+\Delta t} | X_t] = N \times \text{var}[X_{t+\Delta t}^* | X_t^*] \times N'. \quad (24)$$

In the case of the SAS-R models, the density for $X_{t+\Delta t}$ conditional on X_t is assumed multivariate normal. Unfortunately there is no known closed-form solution to the expected values and covariance matrix of $X_{t+\Delta t}$ conditional on X_t in the SAS-R model and hence, in the SAS-R case, I adopt an approximation similar in nature to the one proposed by Duffee and Stanton (2001).³ The approximation is based on a Taylor expansion of the drift of the state variables around X_t

$$\mu(X) = \kappa\theta + \Sigma\sqrt{S}\Sigma^{-1}\lambda_0 - \kappa \times X, \quad (25)$$

where $\sqrt{S}$ is a diagonal matrix with the i th diagonal given by $\sqrt{\alpha_i + \beta_i' X}$.

By means of a Taylor expansion around X_t , $\sqrt{\alpha_i + \beta_i' X} \sim \sqrt{\alpha_i + \beta_i' X_t} + \frac{\beta_i'(X - X_t)}{2\sqrt{\alpha_i + \beta_i' X_t}}$. Substituting this Taylor approximation in Equation (25) and rearranging terms results in

$$\mu(X) \sim \kappa\theta + \Sigma A(X_t)\Sigma^{-1}\lambda_0 - (\kappa \times X - \Sigma B(X_t)\Sigma^{-1}\lambda_0) \quad (26)$$

where $A(X_t)$ and $B(X_t)$ are diagonal matrices with the i th diagonal term given by $A_{i,i}(X_t) = \sqrt{\alpha_i + \beta_i' X_t} + \frac{-\beta_i' X_t}{2\sqrt{\alpha_i + \beta_i' X_t}}$ and $B_{i,i}(X_t) = \frac{\beta_i' X_t}{2\sqrt{\alpha_i + \beta_i' X_t}}$.

Observe that Equation (26) is a linear approximation for the drift of the state variables in the semiaffine model. Consequently the procedure used to calculate the expectation and the covariance matrix in essentially affine models can be used to compute an approximation for the expectation and for the covariance matrix of $X_{t+\Delta t}$ conditional on X_t in the semiaffine model.

³ Duffee and Stanton (2001) propose a Kalman filter estimator for the SAS-R model that is based on three approximations: The first is the use of the instantaneous dynamics of the state variables as proxy for the discrete-time dynamics. The second is the linearization of the drift of the SAS-R model. The third is the evaluation of these dynamics at filtered values instead of evaluating them at exactly identified values. Herein the only approximation is the linearization of the drift of the SAS-R model.

The approximations above are tested by Monte Carlo simulations and they seem to work very well for Δt equal to one month. I have compared the calculated expectations and variances of $X_{t+1 \text{ month}}$ conditional on X_t with those computed by the Monte Carlo simulations. The comparison is made for all the estimated semiaffine models with the parameters displayed in Table 1. A total of 10,000 Monte Carlo simulation runs with antithetic paths, control variate, Euler discretization of SAS-R diffusion and discretization interval equal to 1/251 year are performed. The control variate is with the estimated essentially/completely affine model corresponding to the estimated semiaffine model. For instance, for the estimated SAS- $R_2(3)$ model the estimated $CA_2(3)$ is used.

The approximations are analyzed through the mean square percentage error. In all semiaffine models the square root of the mean square percentage error of the conditional expectations are less than 0.35% and the square root of the mean square percentage error of the conditional variances are less than 0.2%. An examination of Figure 2 reveals that it is not surprising that these approximations work so well for small periods Δt (1 month). Figure 2 reveals that the nonlinear drift of the SAS- $R_1(3)$ is very close to a linear function in the range where data are observed. In the case of the SAS- $R_2(3)$ and the SAS- $R_3(3)$ models, plots similar to the ones in Figure 2 reveal that the nonlinear drift of the state variables is very close to a linear function in the observed state variables range. This suggests that the approximations based on the linearization of the drift should work well for the estimated semiaffine models.

The likelihood function is maximized, imposing the constraints displayed in Section 1.2 and in Dai and Singleton (2000). In addition, all the state variables that affect the matrix $\sqrt{S}$ are constrained to be positive in the estimation period. The likelihood maximization of the models with stochastic volatility involves a large number of constraints and thus the possibility of local optimum is a major concern. To avoid this possibility, I tried a large number of different starting values.

The starting values of the likelihood maximization for the models $EA_0(3)$, $EA_1(3)$, and $CA_2(3)$ are based on the results of the maximum-likelihood displayed in Duffee (2002). For the CIR model, many different starting values were tried and usually the maximization procedure converged to the same result. The starting values for the models SAS- $R_1(3)$, SAS- $R_2(3)$, and SAS- $R_3(3)$ are based on the results of the likelihood maximization of the $EA_1(3)$, $CA_2(3)$, and CIR models, respectively. The likelihood function is reasonably flat in relation to the parameters λ_0 , consequently I make a grid search by changing the parameters that control the mean reversion of the state variables (κ and λ_0), while keeping the diffusion parameters under the equivalent martingale measure fixed by changing the parameters in the λ_1 vector and in the λ_2 matrix.

The expectations and variances used in Equations (11) and (12) are calculated by Equations (22) and (24) for the $EA_0(3)$, $EA_1(3)$, $CA_2(3)$, and CIR models and they are calculated by Monte Carlo simulation for the semiaffine models. A total of 10,000 Monte Carlo simulation runs with antithetic paths, Euler discretization of SAS-R diffusion, and discretization interval equal to 1/36 year are performed. Monte Carlo simulations are used in this case instead of using the approximations given by Equation (26) because the approximations do not work so well for a six-month forecasting period.

When calculating the expectations used in Equations (11) and (12) for the out-of-sample period, it is possible to find negative values for the state variables affecting the volatilities of rates. These negative values are not admissible in the stochastic volatility models. When they occur, the negative state variable X_i is made equal to a very small positive number and the other two state variables are found by assuming that the yields of the 6-month and 10-year yields are observed without errors. These negative values are not found in the estimation period because in the likelihood maximization, all the state variables that affect the matrix $\sqrt{S}$ are constrained to be positive in the estimation period.

References

- Ahn, D.-H., Dittmar, R. F., and Gallant, A. R., 2002, "Quadratic Term Structure Models: Theory and Evidence", *Review of Financial Studies*, 15, 243–288.
- Bliss, R. R., 1997, "Testing Term Structure Estimation Methods," *Advances in Futures and Options Research*, 9, 197–231.
- Campbell, J. Y., 1995, "Some Lessons from the Yield Curve," *Journal of Economic Perspectives* 9, 129–152.
- Campbell, J., and Viceira, L., 1997, "Who Should Buy Long-Term Bonds?," *American Economic Review*, 91, 99–127.
- Chen, R. R., and Scott, L., 1993, "Maximum Likelihood Estimation for a Multifactor Equilibrium Model of the Term Structure of Interest Rates," *Journal of Fixed Income*, December, 14–30.
- Cox, J. C., Ingersoll, J. E., Ross, S. A., 1985, "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 386–406.
- Dai, Q., and Singleton, K. J., 2000, "Specification Analysis of Affine Term Structure Models." *Journal of Finance* 55, 1943–1978.
- Dai, Q. and Singleton, K. J., 2002, "Expectation Puzzles, Time-Varying Risk Premia, and Dynamic Models of the Term Structure." *Journal of Financial Economics* 63, 415–441.
- Dai, Q., and Singleton, K. J., 2003, "Term Structure Dynamics in Theory and Reality," forthcoming in *Review of Financial Studies*.
- Duffee, G. R., 2002, "Term Premia and Interest Rate Forecasts in Affine Models," *Journal of Finance*, 57, 405–443.
- Duffee, G. R., and Stanton, R. H., 2001, "Estimation of Dynamic Term Structure Models," working paper, Haas School of Business.
- Duffie, D., 1996, *Dynamic Asset Pricing Theory*, 2nd ed., Princeton University Press, Princeton, NJ.
- Duffie, D. and Kan, R., 1996, "A Yield Factor Model of Interest Rates." *Mathematical Finance*, 6, 379–406.
- Duffie, D., and Singleton, K., 1997, "An Econometric Model of the Term Structure of Interest-Rate Swap Yield," *Journal of Finance*, LII, 1287–1321.
- Fisher, M., and Gilles, C., 1996, "Estimating Exponential-Affine Models of the Term Structure." working paper, Federal Reserve Board.
- Ingersoll, J. E., 1987, *Theory of Financial Decision Making*, Rowman & Littlefield, Savage, MD.
- Jagannathan, R., and Sun, G., 1998, "An Evaluation of Multi-Factor CIR Models Using LIBOR, Swap Rates, and Cap Swaption Price," working paper, Northwestern University.
- Litterman, R. and Scheinkman, J., 1988, "Common Factors Affecting Bond Returns," Research paper, Goldman Sachs Financial Strategies Group.
- McCulloch, J. H., and Kwon, H. C. 1993, "U.S. Term Structure Data, 1947–1991," Working Paper 93-6, Ohio State University.
- Pearson, N. D. and Sun, T., 1994 "Exploiting the Conditional Density in Estimating the Term Structure: An Application to the Cox, Ingersoll, and Ross Model," *Journal of Finance*, XLIX, 1279–1304.
- Vasicek, O., 1977, "An equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177–188.

Copyright of Review of Financial Studies is the property of Society for Financial Studies and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.