

A Dynamic Model for the Forward Curve

Choong Tze Chua
Singapore Management University

Dean Foster
University of Pennsylvania

Krishna Ramaswamy
University of Pennsylvania

Robert Stine
University of Pennsylvania

This article develops and estimates a dynamic arbitrage-free model of the current forward curve as the sum of (i) an unconditional component, (ii) a maturity-specific component and (iii) a date-specific component. The model combines features of the Preferred Habitat model, the Expectations Hypothesis (EH) and affine yield curve models; it permits a class of low-parameter, multiple state variable dynamic models for the forward curve. We show how to construct alternative parametric examples of the three components from a sum of exponential functions, verify that the resulting forward curves satisfy the Heath-Jarrow-Morton (HJM) conditions, and derive the risk-neutral dynamics for the purpose of pricing interest rate derivatives. We select a model from alternative affine examples that are fitted to the Fama-Bliss Treasury data over an initial training period and use it to generate out-of-sample forecasts for forward rates and yields. For forecast horizons of 6 months or longer, the forecasts of this model significantly outperform those from common benchmark models. (*JEL* C53, E43, E47)

The structure of forward rates for fixed maturity loans to begin at various dates in the future can be inferred from the prices of Treasury securities or directly observed from the extremely active Eurodollar Futures market. The constellation of these rates (or of the related yields) plays a central role in the allocation of capital. The random behavior of this “yield curve”—or the relationship between the yields and the term to maturity—is a subject of considerable theoretical and empirical study.

We would like to thank the editor, Yacine Aït-Sahalia, and two anonymous referees for their guidance. We are also grateful to Amir Yaron, Michael Brandt, Francis Diebold and seminar participants at The University of Pennsylvania, Singapore Management University, the 2005 Financial Management Association meeting in Siena and the 2005 Winter Research Conference at the Center for Analytical Finance, Indian School of Business in Hyderabad for their comments. Address correspondence to Choong Tze Chua, Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899, or e-mail: ctchua@smu.edu.sg.

© The Author 2007. Published by Oxford University Press on behalf of The Society for Financial Studies. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org.
doi:10.1093/rfs/hhm039

Yield curves have traditionally been modeled in one of two ways: equilibrium models and no-arbitrage models. Equilibrium models such as the Vasicek model (1977) and the Cox-Ingersoll-Ross model (1985) define stochastic processes driven by a small number of forcing factors. Once these processes are defined, the forward curve and its evolution can be derived either under various assumptions for the risk premia or from a more fundamental model that begins with preferences and imposes market clearing conditions. Empirically, however, these models do not fit the observed forward curve well on any given day. In fact, the fit is often so poor that the differences between the empirical and theoretical values can be construed as model mis-specification rather than random pricing errors.

In contrast, no-arbitrage models are calibrated to fit the observed forward curve perfectly on a given day. This approach to modeling the term-structure was pioneered by Ho and Lee (1986). Hull and White (1990) also built a no-arbitrage model that extended the Vasicek model to fit the initial term-structure. Other important contributors to the no-arbitrage model literature include Black, Derman, and Toy (1990) and Heath, Heath, Jarrow, and Morton (1992). HJM derived a framework for the arbitrage-free evolution of the entire forward curve, starting from the currently observed forward curve. The time-series dynamics of the forward curve evolution are constrained by the current shape of the curve. Also, these models are forced to fit measurement errors in the observed term-structure thereby generating erroneous implications for the time-series evolution. This has led some authors to argue that by forcibly calibrating the entire curve to the observed rates, arbitrage-free principles could be violated (See Backus, Foresi, and Zin (1998)).

There have been several new developments in term-structure modeling in the form of market models and stochastic string models. Market models seek to model observable quantities such as the London Interbank Offered Rates (LIBOR) directly within the framework of HJM. Brace, Gatarek, and Musiela (1997) and Miltersen, Sandmann, and Sondermann (1997) proposed the BGM model that falls into this category. Because these models can be calibrated using observed market rates instead of proxies, they can be estimated accurately. Stochastic string models were first developed by Kennedy (1994), who modeled the forward curve as a Gaussian random field. Goldstein (2000), Santa-Clara and Sornette (2001) and Collin-Dufresne and Goldstein (2003) later proposed similar models that allow for strings of shocks to the forward rate curve that are correlated with one another. In stochastic string models, instantaneous forward rates of differing maturities are driven by their own shocks that can be correlated with those of neighboring maturities. Stochastic string models can fit any day's cross-section of bond prices perfectly without any need for measurement error. Thus when measurement errors are in fact presents these models may over-fit the data.

In this article, we propose an easily interpretable and arbitrage-free general model of forward rates that appeals to economic intuition, and show that a specific form of the model that is (exponentially) affine in the state variables generates superior out-of-sample forecasts of forward rates and yield curves. Here is a brief summary of the general model's principal features:

1. The current term structure of forward rates is modeled as the sum of three components:
 - (a) an unconditional curve that represents the steady-state forward curve;
 - (b) a maturity-specific curve consisting of current deviations from the unconditional curve that can be driven by one or more state variables and embeds the influence of supply and demand from agents who have needs for loans of specific terms. To justify this curve we would appeal to investors' preferences, or to a preferred habitat model (see Modigliani and Sutch (1966)); and
 - (c) a date-specific curve that can be driven by one or more state variables and embeds the current influence of expectations about spot rates to prevail at specific future dates. The date-specific component is intended to summarize the influence of fundamental nominal and real factors on future expected interest rates.
2. The evolution of the maturity- and date-specific component curves is autoregressive in sensible ways described further in Section 1;
3. The dynamics of the sum of the three component curves (under certain conditions, for chosen parametric forms that are affine in the state variables) produce a model of the forward curve that is arbitrage free and meets the conditions imposed by Heath, Jarrow, and Morton (1992); and finally
4. A selected parametric model, when taken to the data, generates out-of-sample forecasts that are superior to those from available benchmark models.

An important feature of this model is that it recognizes, in reduced form, the influence of both maturity-specific effects from investors' choices and date-specific effects driven by economy-wide events. The framework outlined above permits one to build alternative dynamic models of the forward curve; indeed, the analysis permits nonlinear and nonaffine forms, and can be applied to model the dynamics of forward curves for market prices of commodities as well. We choose a specific version of the model from a menu of alternative models using Fama-Bliss US Treasury data over a preliminary training period, before testing its forecasting power over the remaining out-of-sample period.

There are several measures that can be used to compare competing interest rate models. These include comparing the implied parametric density to the nonparametric estimates (see Aït-Sahalia (1996b)), comparing the goodness-of-fit to the empirically observed data, and comparing the accuracy of out-of-sample forecasts. The latter two measures are more commonly used. However, we consider superior out-of-sample forecasting performance to be more important than superior in-sample fits. In-sample fits in many models (including ours) can always be improved by increasing the complexity of the model. However, as noted by Diebold and Li (2006), it is not obvious that such over-fitting leads to improved out-of-sample forecasting performance. Out-of-sample forecasting performance is therefore a more objective measure of model performance. Despite the sizable literature on the theory and estimation of term-structure models, few authors have produced forecasts that are significantly better than even the most elementary benchmark, the Random Walk (see Duffee (2002) for a survey of poor forecasts generated by the most common models). One exception is Diebold and Li (2006), who fit autoregressive models to parameter estimates of the Nelson–Siegel model (1987). They reported significantly better forecasts than the Random Walk model when forecasting yields of maturities less than 5 years at the 12-month ahead horizon.

The remainder of the article is organized as follows: Section 1 introduces the components of our model and provides the intuition and economic origins behind them. Section 2 shows how to construct, using examples, a class of models that can be developed from this economic intuition; it also shows that they conform to the HJM specification for arbitrage-free dynamics. The pricing of bonds and derivative securities under this framework are also explained here. Section 3 explains how the model can be implemented by using a Kalman filter and also chooses a particular model for implementation. Section 4 then applies the chosen model to Fama-Bliss Treasury data and shows that the forecasts generated by that model are significantly better than the forecasts generated by the benchmark models: the Random-Walk (RW) model, the Expectations Hypothesis (EH) model and the EH with Term-Premium model. We also compare our results to those in the recent literature, namely Diebold and Li (2006) and Duffee (2002). Section 5 concludes.

1. Qualitative Description and Economic Intuition of the Model

The current (date t) forward curve is written $f(\tau; t)$ and represents the curve of forward rates for instantaneous loans to begin at future dates $t + \tau$, $\tau > 0$. The proposed model of the forward curve is the sum of three

component curves:¹

$$f(\tau; t) = U(\tau) + M(\tau; t) + D(\tau; t) \quad (1)$$

where

1. $U(\tau)$ is the unconditional or steady-state forward curve;
2. $M(\tau; t)$ is the component curve of maturity-specific deviations; and
3. $D(\tau; t)$ is the component curve of date-specific deviations.

The first argument τ in parentheses refers to the time to maturity; where there is a second argument it refers to the calendar date for that component curve. Thus, $M(\tau; t)$ refers to the maturity-specific deviation embedded in the forward curve at date t for the future date $t + \tau$.

1.1 The unconditional forward curve

$U(\tau)$ represents the steady state or the unconditional forward curve; if we were to forecast the forward curve at a time in the distant future, all presently available information would be of little use. This unconditional curve can be written as

$$U(\tau) = \lim_{s \uparrow \infty} E_t[f(\tau; s)] \quad (2)$$

It is time invariant and may be estimated by taking an average of all available historical curves.

1.2 The curve of maturity-specific deviations

The curve of maturity-specific deviations recognizes that a part of the deviation of the current forward curve from the unconditional forward curve at some maturities has no implication for future spot rates. Rather, this abnormal component may be local to those particular maturities of the forward curve. The concept of a maturity-specific deviation originates from the Market Segmentation Hypothesis and the Preferred Habitat Theory (Modigliani and Sutch (1966)). These models postulate that some market participants are primarily concerned with their natural maturity habitat, with little regard for the implication of the forward rates on future spot rates. The actions of these participants affect only those maturities (and nearby maturities) of the forward curve at each date, instead of having effects that move progressively towards shorter maturities and eventually affect the spot rates.²

¹ We assume that the forward curve on any given date is observed with random measurement error. Thus, to recover the fitted forward curve on any given date, we do not fit the observed curve exactly. Instead, we use smooth functions to obtain the fitted curve, and assume that the residuals are random measurement errors.

² For instance, a decrease in medium-term liquidity in the loanable funds market may drive forward rates in the 5-year maturity higher, and such a change would be captured as maturity-specific deviation, all else equal.

Therefore, $M(\tau; t)$, the maturity-specific component curve at date t , captures abnormal activity that affects the forward curve at specific maturities τ . The entire maturity-specific deviation curve may be modeled as a point-wise mean-reverting process that reverts to zero, so that

$$E_t [M(\tau; T)] = e^{-K_m(T-t)} M(\tau; t) \quad \tau > 0, \quad (3)$$

where $K_m > 0$ is a parameter indicating the speed of reversion to zero. The overall maturity-specific deviation can be comprised of two (or more) different maturity-specific deviations; so that, for example,

$$M(\tau; t) = M_1(\tau; t) + M_2(\tau; t)$$

where $M_1(\tau; t)$ and $M_2(\tau; t)$ are both mean-reverting to zero at different rates. So for each component of the overall maturity-specific deviation we require that

$$E_t [M_j(\tau; T)] = e^{-K_{m_j}(T-t)} M_j(\tau; t) \quad \tau > 0, j = 1, 2, \dots \quad (4)$$

The arbitrage-free formulation of the overall curve of maturity-specific deviations (described further in Section 2) has the property that $M(\infty; t) = 0$ for all t . Note that instantaneous or spot rates are relevant to zero maturity loans, and we assume $M(0; t) = 0$ for all t to allow the date-specific deviations to capture the dynamics of present and future spot rates. Figure 1 illustrates the forecasted behavior of maturity-specific deviations: anchored at zero at extreme maturity values, the entire curve decays (in expectation) point-wise towards zero as time passes, satisfying relation (3).

1.3 The curve of date-specific deviations

A date-specific deviation is caused by information affecting expectations of the spot interest rate on a specific calendar date in the future. The concept of a date-specific deviation has its roots in the EH (Fisher (1896)). It is intuitive that forward rates contain information regarding future spot rates; therefore a high forward rate today should naturally point towards a higher spot rate at the corresponding date in the future. However, the EH fails in some basic ways, as shown in the literature. In the theoretical realm, it has been shown that most versions of the EH admit arbitrage (Cox, Ingersoll, and Ross (1981)).³ In empirical tests, forecasts of forward rates generated by the EH model are generally considered to be inferior to even the most basic benchmark, the Random Walk model. The model proposed

³ Some recent literature seems to vindicate theoretical aspects of the EH. McCulloch (1993) and Fisher and Gilles (1998) present examples to show that some forms of the EH are consistent with no-arbitrage. Longstaff (2000) shows that all traditional forms of the EH are consistent with no-arbitrage if markets are incomplete.

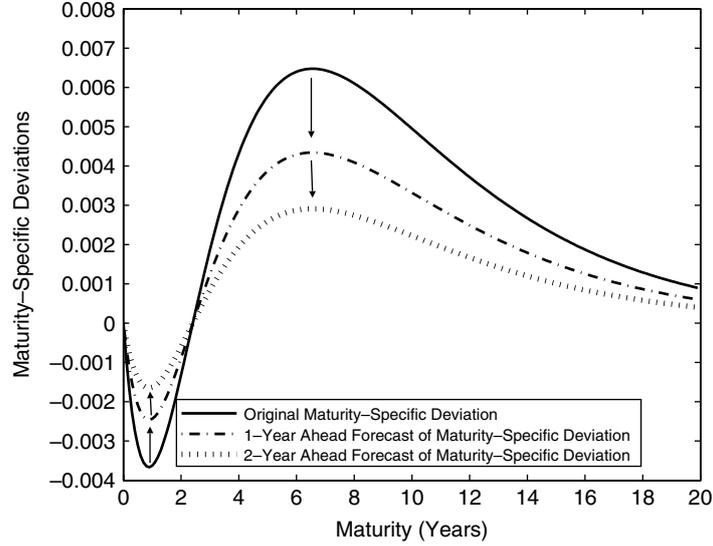


Figure 1
Illustration of maturity-specific deviation behavior
 Starting with any given maturity-specific deviation (for illustrative purposes, we set the original maturity-specific deviation to be $M(\tau; t) = 0.05e^{-0.2\tau} - 0.1e^{-0.4\tau} + 0.05e^{-0.8\tau}$). We expect the maturity-specific deviation to decay exponentially to zero at rate K_m (In this illustration, we set $K_m = 0.4$) as time passes (from relation (3)): $E_t[M(\tau; T)] = e^{-K_m(T-t)}M(\tau; t)$.

here attributes only a part of the current forward curve as containing information about future spot rates.

The date-specific deviation curve $D(\tau; t)$ is influenced by abnormal events or information that affects the portions of the forward curve corresponding to specific maturity dates. In other words, this curve captures the deviations of expected future spot rates from the unconditional spot rate. For instance, suppose that on $t \equiv$ January 1, 2002 it is learned that the Treasury needs additional financing on (or around) $s \equiv$ January 2003; that will drive up interest rates during that period. On 1 January 2002, the 1-year forward rate would be elevated. As time passes, we expect the elevated portion of the forward curve to move closer to the origin, since in expectation the higher rates around 1 January 2003, would remain. Thus, the date-specific deviation has the property:

$$E_t [D(s - T; T)] = D(s - t; t) \quad t < T < s. \quad (5)$$

The date-specific deviation at zero maturity is simply the difference between the spot interest rate and the unconditional spot rate: $D(0; t) = f(0; t) - U(0)$. At infinite maturity, the date-specific deviation must be zero because it is not plausible that one can have any information about the spot rate in the infinite future other than that contained in the

unconditional spot rate, so $D(\infty; t) = 0$ for all t . Figure 2 illustrates the forecasted behavior of the date-specific deviation. Starting from a given date-specific curve that is anchored at zero at the long end, the entire curve shifts (in expectation) to the left as time passes, satisfying relation (5).

1.4 The dynamic behavior of the forward curve

The dynamic behavior of the forward curve in relation (1) depends only on the dynamic behavior of the date-specific and the maturity-specific deviations, as indicated in relations (3) and (5). Each of these, within a specific model that we specify in Section 2, is affected by one or more state variables representing the evolution of underlying economic factors.

The maturity-specific deviation is caused by abnormal pricing of forward rates specific to certain maturities, driven by habitat and preferences of individual and institutional investors. Changes in demand or supply at a given maturity habitat can affect a range of surrounding maturities—investors treat them as close substitutes—which allows us to treat the maturity-specific deviation as a smooth curve. Since these are deviations from the average, the average deviation should naturally be zero. Without additional information to guide us on how these deviations behave over time, a simple yet intuitive model for these deviations would be that they decay towards zero at some rate. In Section 2, where we

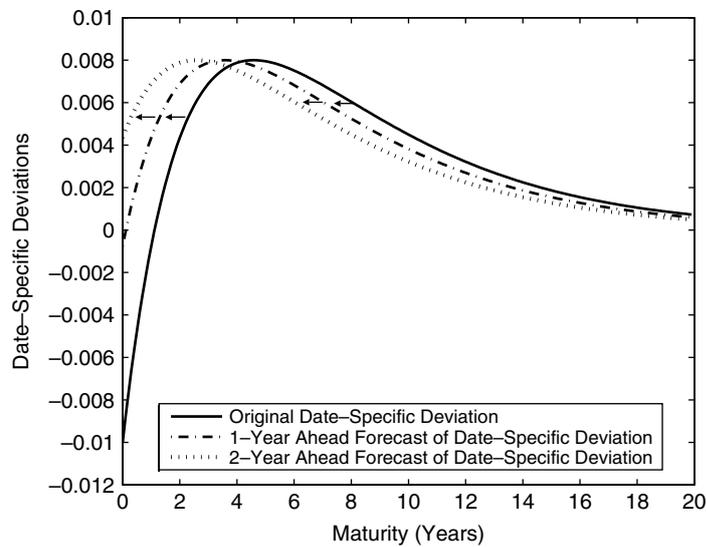


Figure 2
Illustration of Date-specific Deviation Behavior
 Starting with any given date-specific deviation (for illustrative purposes, we set the original date-specific deviation to be $D(\tau; t) = 0.04e^{-0.2\tau} - 0.05e^{-0.4\tau}$). We expect the date-specific deviation curve to shift to the left uniformly as time passes (from relation (5)): $E_t[D(\tau - (T - t); T)] = D(\tau; t)$.

develop an arbitrage-free framework for our model, we assume that the maturity-specific deviation decays at an exponential rate to satisfy the HJM requirement for the model to be arbitrage free.

Our model (in the general form under discussion so far) does not *a priori* preclude the possibility that there might be negative forward rates. Given an observed term structure of forward rates that is positive at all maturities, it is possible to find maturity- and date-specific deviations that fit the current term structure but produce forecasts of negative forward rates in the future. For example, an extremely large and positive maturity-specific deviation coupled with an extremely large and negative date-specific deviation can produce such negative forward rate forecasts. However, in the explicit parameterized forms of the model described in Section 2 we ensure that the model is arbitrage free by checking the HJM restrictions.

In the implementations of explicit forms of our general model (described in Section 2 and made clear in the estimation procedure in Section 3) we employ sums of exponential basis functions for $U(\tau)$ and similar basis functions (that are scaled by Brownian motions) to specify the functional forms for $M(\tau; t)$, and $D(\tau; t)$. The resulting function for forward rates $f(\tau; t)$ is affine in the state variables and has a structure that lends itself easily to estimation.

1.5 Closely related models

We now show that many well-known models of the term structure—the EH and the Vasicek and CIR models—are closely related⁴ to the model that we are proposing here, in the sense that these models share one or more of the three components described above.

1.5.1 Expectations hypotheses. As mentioned in Section 1 the date-specific deviation has its roots in the EH. Two forms of EH can be viewed as special cases of our class of models. The Pure EH—that forward rates are predictors of future spot rates—can be written as:

$$E_t[f(0; T)] = f(T - t; t),$$

where $f(0; T)$ is the spot rate on date T and $f(T - t; t)$ is the forward rate quoted at date t for an instantaneous loan to begin at date T . This relation is the same as relation (5): we can view the Pure EH as a special case of our model, influenced by date-specific events but without an unconditional curve or a maturity-specific deviation curve.

The EH with a Term Premium is different from Pure EH in that it assumes the existence of a (perhaps maturity dependent) time-invariant

⁴ We should emphasize here that the explicit forms of the model (described further in Section 2) that we work with are more complex and will not have these models as stylized special cases.

term premium (say $\lambda(\tau)$ for maturity τ), where the term premium is defined as the excess of the quoted forward rate over the expected spot rate. Denoting the term premium at maturity $T - t$ as $\lambda(T - t)$, the date t predictor of the future spot rate for date T is

$$E_t[f(0; T)] = f(T - t; t) - \lambda(T - t).$$

In the absence of a maturity-specific deviation, the date t predictor of the future spot rate for date T is

$$\begin{aligned} E_t[f(0; T)] &= E_t[U(0) + D(0; T)] \\ &= U(0) + D(T - t; t) \\ &= D(T - t; t) + U(T - t) - (U(T - t) - U(0)) \\ &= f(T - t; t) - [U(T - t) - U(0)]. \end{aligned} \quad (6)$$

The EH with Term Premium is therefore a special case of our model without any maturity-specific deviation, where the difference between the unconditional rate at zero maturity and its value for a given maturity τ corresponds to the time-invariant term premium $\lambda(\tau) \equiv U(\tau) - U(0)$.

1.5.2 Vasicek and Cox - Ingersoll - Ross models. Two well-known models are those of Vasicek (1977) and Cox-Ingersoll-Ross (CIR) (1985). The single factor versions of these two models specify a continuous-time autoregressive process for the spot interest rate whose long-run mean is θ , and derive a pricing formula for zero coupon bonds that is exponentially affine in that spot rate:

$$P_{zc}(\tau; t) = e^{A(\tau) - B(\tau)r_t},$$

where t is the current date, $P_{zc}(\tau; t)$ is the price of the zero-coupon bond of maturity τ , r_t is the spot interest rate at date t , and $A(\tau)$ and $B(\tau)$ are known expressions that involve other parameters including a market price of risk. These models imply that the forward rate quoted on date t for an instantaneous-maturity loan at date $t + \tau$ is

$$f(\tau; t) = -\frac{P'_{zc}(\tau; t)}{P_{zc}(\tau; t)} = -A'(\tau) + B'(\tau)r_t.$$

The forward rate is an affine function of the spot interest rate, and because affine functions of autoregressive processes are themselves autoregressive it follows that the forward rates of the Vasicek and CIR models are point-wise autoregressive processes. To draw the parallel between these two models and our model, first note that in both these models

$$E_t(r_T - \theta) = e^{-\kappa(T-t)}(r_t - \theta)$$

where κ is the speed-of-adjustment parameter common to both models. Then the conditional expectation of the τ -period forward rate at the future date T given information at date t is

$$\begin{aligned} E_t[f(\tau; T)] &= E_t[-A'(\tau) + B'(\tau)r_T] \\ &= E_t [(-A'(\tau) + B'(\tau)\theta) + B'(\tau)(r_T - \theta)] \\ &= \{(-A'(\tau) + B'(\tau)\theta)\} + [e^{-\kappa(T-t)}B'(\tau)(r_t - \theta)]. \quad (7) \end{aligned}$$

As a result, the forward curve in the Vasicek and CIR models can be decomposed into two components: $\{(-A'(\tau) + B'(\tau)\theta)\}$ which is time-invariant, and $[B'(\tau)(r_t - \theta)]$ which is autoregressive and decays towards zero. When comparing to our class of models, we simply omit the date-specific deviation from our model, and recognize the first component as analogous to the unconditional curve of our model;⁵ the second component is analogous to the maturity-specific deviation of our model, obeying the same dynamics as the maturity-specific deviation:

$$E_t [B'(\tau)(r_T - \theta)] = e^{-\kappa(T-t)} B'(\tau) (r_t - \theta).$$

In the multi-factor versions of the Vasicek and CIR models, there is one time-invariant component but there are multiple autoregressive components decaying toward zero at different rates. Thus, the multi-factor versions of the Vasicek and CIR models are analogous to our model with an unconditional curve and multiple maturity-specific deviation curves but without any date-specific deviation.

2. Explicit Forms of the Forward Rate Model

The general model we presented in Section 1 involves relations (1), (3) and (5). We rely on the guidance of other theoretical models and hypotheses while discussing our general model, but it remains a reduced form model whose three component curves need explicit forms before we can judge its effectiveness. However, for each explicit form that we propose, we would like to ensure that the dynamics for the resulting forward curve are arbitrage free.

In this section, we develop several explicit parametric forms of the forward rate dynamics in our model and also show that these models are arbitrage free under certain conditions. All the formulations of the forward curve that we propose (and test in subsequent sections) use representations of the maturity- and date-specific deviations that are continuous functions, and they share the property that they are (exponentially) affine in the

⁵ The first component in relation (7) corresponds exactly to the unconditional forward curve of the Vasicek and CIR models.

driving state variables that themselves follow Itô processes. Therefore, we need to establish the arbitrage-free property in this setting, and we do so by verifying the conditions imposed by HJM (1992). We should add that it is possible to develop explicit forms of our general model that are both arbitrage-free and not exponentially affine in the driving state variables. We include an example of a nonaffine parameterization of our model that is consistent with HJM in Appendix A.3.

The HJM paper shows that if forward rates are processes, then the drift $\mu(t, s)$ and diffusion $\sigma(t, s)$ of the Stochastic Differential Equation (SDE) for the forward rate $f_{HJM}(t, s)$ quoted at t for date $s > t$ (in our notation this would be $f(s - t; t)$), are related by:

$$\mu(t, s) = \sigma(t, s)^T \left(\int_t^s \sigma(t, v) dv - \bar{\kappa}_t \right) \quad (8)$$

for some vector $\bar{\kappa}_t$ that satisfies the equality:

$$E \left[\exp \left(\int_0^T \bar{\kappa}_t^T dB_t - \frac{1}{2} \int_0^T |\bar{\kappa}_t|^2 dt \right) \right] = 1 \quad (9)$$

Equivalently, under the risk-neutral measure, the drift $\mu^*(t, s)$ and diffusion $\sigma(t, s)$ terms of the forward rate SDE are necessarily related by:

$$\mu^*(t, s) = \sigma(t, s)^T \left(\int_t^s \sigma(t, v) dv \right) \quad (10)$$

Note that the diffusion, but not the drift, of the forward rate under the risk-neutral measure is identical to that under the real measure.

In developing the explicit forms of the model, it is useful to introduce the notion of an *Arbitrage-Free Unit* (AFU). An AFU is an elementary model of forward rates: each unit can be driven by one, two or more Brownian motions. While an AFU is theoretically possible under the HJM framework, it may offer too simple a structure to accurately represent real world data. In order to get a more flexible and realistic model of forward rates, these AFUs can be combined to form a composite arbitrage-free description of forward rates, and we now turn to these tasks.

2.1 1-Brownian motion arbitrage-free unit

Consider first a simple dynamic model for the forward rate that is driven by a single Brownian motion—we denote this $f_1(\tau; t)$ with a subscript indicating the number of Brownian motions,⁶ so that the innovations in both the maturity-specific and date-specific deviation are perfectly

⁶ The instantaneous forward rate on date t for maturity on date s is really a function of $\{t, s - t, m(t), \text{ and } d(t)\}$. For simplicity, we will continue to write the forward rate as a function of two variables: $\{\tau = s - t, t\}$, writing $f_1(\tau; t)$ in place of $f_1(t, s, m(t), d(t))$ and suppressing the dependence on the two state variables.

correlated. Later, we extend the AFU to embed two or more Brownian motions so as to get a richer set of models.

The explicit parameterization is chosen as a linear combination of exponential functions. The precise choice of exponential bases can affect the arbitrage-free status of the model. As an example, we now choose a particular basis set that we later show to conform to the HJM specification in matching the drift and the diffusion of the resulting forward rate process. The three components of the current forward curve $f_1(\tau; t)$ are as follows:

1. The time-invariant unconditional curve is now explicitly written as

$$U_1(\tau) = C_0 - C_1 e^{-2K_m \tau}, \quad (11)$$

where C_0, C_1 and K_m are positive constants. This form generates a smooth upward-sloping unconditional curve that starts at $U_1(0) = C_0 - C_1$ at the origin and asymptotes to C_0 at infinite maturity.

2. The maturity-specific deviation is explicitly written as

$$M_1(\tau; t) = m(t) \left[e^{-K_m \tau} - e^{-2K_m \tau} \right]. \quad (12)$$

By design, $M_1(0; t) = 0 \forall t$. Because $\lim_{\tau \rightarrow \infty} M(\tau; t) = 0$ the deviation has a humped shape. The $m(t)$ is an Itô process whose dynamics are induced by the Brownian motion, defined further below; $m(t)$ serves to scale the deviation which has a fixed shape with a peak value at maturity $\tau = \frac{\ln 2}{K_m}$.

3. The date-specific deviation is specified as

$$D_1(\tau; t) = d(t) \left[e^{-2K_m \tau} \right]. \quad (13)$$

Here $d(t)$ is an Itô process whose dynamics are related to the Brownian motion, also defined below; it serves to scale an exponential function which is either monotonically upward- or downward-sloping. Note that the overall date-specific deviation $D_1(0; t) = d(t)$ at zero maturity, and it asymptotes to zero at infinite maturity ($D_1(\infty; t) = 0$), reflecting the fact that there can be no expectation about the spot rate in the distant future other than the long-run mean.

One can also interpret this parameterization of the unconditional curve, the maturity-specific deviation curve and the date-specific deviation curve as polynomials in the log-maturity scale. For instance, if we let $p(x) = e^{-K_m x}$, then $U_1(x) = C_0 - C_1 p(x)^2$, $M_1(x; t) = m(t)(p(x) - p(x)^2)$ and $D_1(x; t) = d(t)p(x)^2$.

Given this parameterization, we can use Itô's lemma to derive the following SDE for the evolution of the forward rate:

$$df_1(\tau; t) = \frac{\partial f_1}{\partial t} dt + \frac{\partial f_1}{\partial m(t)} dm(t) + \frac{\partial f_1}{\partial d(t)} dd(t), \quad (14)$$

indicating dependence on the driving Itô processes $m(t)$ and $d(t)$; all second-order terms are zero.

Recall that the model requires (see relation (3)) the maturity-specific deviation to decay exponentially towards zero at rate K_m . Therefore we require the SDE for the state variable $m(t)$ to have the drift $-m(t)K_m$, and specify its diffusion coefficient γ_t later, when we impose the arbitrage-free condition:

$$dm(t) = -m(t)K_m dt + \gamma_t dB(t), \quad (15)$$

where $B(t)$ is a Brownian motion.

In the SDE for the Itô process $d(t)$ we make its drift rate equal to $-2d(t)K_m$ so that we satisfy the relation (5) above; and we specify the diffusion of the process for $d(t)$ to be identical to that of $m(t)$, which is necessary to ensure that the drift and diffusion of the forward rate conform to the HJM condition in relation (8)

$$dd(t) = -2d(t)K_m dt + \gamma_t dB(t). \quad (16)$$

Note that the maturity-specific and the date-specific deviations are driven by the same Brownian motion, so their innovations in the 1-Brownian motion AFU are necessarily perfectly correlated.⁷

Relation (14), the SDE for the forward rate in this explicit 1-Brownian motion setup, can now be rewritten as

$$df_1(\tau; t) = \left\{ -K_m(2C_1 + m(t))e^{-2K_m(\tau)} \right\} dt + \left\{ e^{-K_m(\tau)} \gamma_t \right\} dB(t). \quad (17)$$

2.2 Checking the HJM restriction

We must now verify that the proposed dynamics in relation (17) is arbitrage free. Denoting the diffusion of the forward rate SDE as

$$\sigma(t, s) = e^{-K_m(s-t)} \gamma_t, \quad \tau \equiv s - t \quad (18)$$

we have

$$\int_t^s \sigma(t, v) dv = -\frac{1}{K_m} e^{-K_m(s-t)} \gamma_t + \frac{1}{K_m} \gamma_t.$$

⁷ By choosing the overall forward curve as the sum of several AFUs driven by one or more Brownian motions we avoid this extreme implication.

For this version of the 1-Brownian motion AFU, we choose the market price of risk κ_t as⁸

$$\kappa_t = \frac{\gamma_t}{K_m} \quad (19)$$

Notice that the market price of risk is proportional to the diffusion term of the state variable, just as in the CIR model. Then the HJM condition says

$$\sigma(t, s) \left(\int_t^s \sigma(t, v) dv - \kappa_t \right) = -\frac{1}{K_m} \gamma_t^2 e^{-2K_m(s-t)}. \quad (20)$$

By specifying γ_t^2 as

$$\gamma_t^2 = (m(t) + 2C_1)K_m^2, \quad (21)$$

relation (20) becomes

$$\sigma(t, s) \left(\int_t^s \sigma(t, v) dv - \kappa_t \right) = -K_m(2C_1 + m(t))e^{-2K_m(s-t)}$$

which is exactly the drift of $df_1(s-t; t)$ (see relation (17)), thus satisfying the HJM condition.

Within the Dai and Singleton (2000) classification scheme the 1-Brownian motion AFU would be a special case of an $A_1(2)$ model since there are 2 state variables, and the correlation structure of the diffusion process is driven by a single state variable. Although the 1-Brownian motion AFU can theoretically be viewed as a specific “model” of forward rates, it is not designed to be a complete model. Rather, we take it to be a basic building block that can be combined with other similar AFUs to form a more comprehensive and complete model.

2.3 2-Brownian motion arbitrage-free unit

The 1-Brownian motion arbitrage-free system can be extended to a system with two Brownian motions driving the forward rate, where the first Brownian motion drives the maturity-specific deviation and a portion of the date-specific deviation, and the second Brownian motion drives the remaining date-specific deviation. This 2-Brownian motion AFU can therefore permit less-than-perfect correlations between the two types of deviations. The economic interpretation behind this system is that there are two independent sets of shocks, the first set of shocks coming from changes in market participants’ supply and demand for funds. This generates some

⁸ It is not necessary that $\kappa_t = \frac{\gamma_t}{K_m}$. If the market price of risk takes on another form, the model requires a different specification for γ_t or $U(s-t)$ or both so that the system remains arbitrage-free.

repercussions in terms of both the expected spot rate in the future (date-specific deviation) and also portions of the forward curve that have no effect on the expected spot rate (maturity-specific deviation). The second set of shocks comes purely from changes in market-wide information about expected spot rates in the future. For instance, if a specific event (for example, a change in future budget deficits) is anticipated to affect the spot rate at some future date, then the date-specific deviation curve shifts to accommodate the change in expectation, while the maturity-specific deviation is unaffected.

We again parameterize the three components of the forward rate curve, $f_2(\tau; t)$, now recognizing the subscript to refer to the 2 Brownian motions. However, each of these components now combines additional exponential functions, thereby allowing for flexible responses to the two driving state variables.

1. We parameterize

$$U_2(\tau) = C_0 - C_1 e^{-2K_m\tau} - C_2 e^{-K_2\tau}, \quad (22)$$

where K_2 and K_m are positive constants. $U_2(\tau)$ is again time-invariant but it is more flexible than in the 1 Brownian motion case; it starts at $C_0 - C_1 - C_2$ at zero maturity, it can be humped or monotonic in maturity, but it asymptotes to C_0 at infinite maturity. Thus, whereas C_0 has to be positive, C_1 and C_2 can be positive or negative. As long as $C_1 + C_2 > 0$, the unconditional forward rate curve is eventually upward-sloping.

2. The maturity-specific deviation is now parameterized as

$$M_2(\tau; t) = m(t) \left\{ 2e^{-K_m\tau} - e^{-2K_m\tau} - e^{-K_2\tau} \right\}. \quad (23)$$

The maturity-specific deviation is zero at zero maturity, and zero at infinite maturity; it is driven by one state variable $m(t)$ that serves to scale the exponential function in braces. Note that it is also more flexible than the maturity-specific deviation in the 1 Brownian motion case as it can now have either 1 or 2 humps, thus effectively emphasizing the influence at two maturities.

3. The date-specific deviation is now parameterized as

$$D_2(\tau; t) = d_1(t)e^{-2K_m\tau} + d_2(t)e^{-K_2\tau} + d_3(t)e^{\frac{-K_2}{2}\tau}. \quad (24)$$

It is now a sum of three exponential functions driven by 3 state variables: $d_j(t)$, $j = 1, 2, 3$, which are all Itô processes whose dynamics are defined below. It can take on a variety of shapes in regions around any given maturity, but asymptotes to zero at infinite maturity.

Given the above parameterization, we can use Itô's lemma to derive the following SDE for the evolution of the forward rate:

$$df_2(\tau; t) = \frac{\partial f_2}{\partial t} dt + \frac{\partial f_2}{\partial m(t)} dm(t) + \sum_{j=1}^3 \frac{\partial f_2}{\partial d_j(t)} dd_j(t), \quad (25)$$

where all second-order terms are zero.

Note that the maturity-specific deviation again decays exponentially towards zero at rate K_m . We specify the SDE for the state variable $m(t)$ (driven by the first Brownian motion $B_1(t)$) with a drift $-m(t)K_m$:

$$dm(t) = -m(t)K_m dt + \gamma_{1,t} dB_1(t)$$

but with a diffusion term $\gamma_{1,t}$ that is chosen to satisfy the HJM condition

$$\gamma_{1,t}^2 = (m(t) + 2C_1) \frac{K_m^2}{4}.$$

For the date-specific deviation to satisfy relation (5), the drift rates for its state variables $d_1(t)$, $d_2(t)$ and $d_3(t)$ must be $-2d_1(t)K_m$, $-d_2(t)K_2$ and $d_3(t)\frac{-K_2}{2}$ respectively. To keep the system arbitrage free we need to specify the diffusions of these three SDEs to be $\gamma_{1,t}$, $\gamma_{1,t}$ and $\gamma_{2,t}$ respectively:

$$dd_1(t) = -2d_1(t)K_m dt + \gamma_{1,t} dB_1(t)$$

$$dd_2(t) = -d_2(t)K_2 dt + \gamma_{1,t} dB_1(t)$$

$$dd_3(t) = -d_3(t)\frac{K_2}{2} dt + \gamma_{2,t} dB_2(t)$$

where it should be noted that the third state variable $d_3(t)$ is driven by the second Brownian motion, with its diffusion term defined as

$$\gamma_{2,t}^2 = \frac{(C_2 K_2 + m(t)(K_2 - K_m))(K_2)}{2},$$

and $B_1(t)$ and $B_2(t)$ are the two independent Brownian motions. The maturity- and date-specific innovations can now exhibit a richer correlation structure.

Finally, we specify the market price of risk to be

$$\vec{\kappa}_t = \begin{bmatrix} \kappa_{1,t} \\ \kappa_{2,t} \end{bmatrix} = \begin{bmatrix} \frac{2}{K_m} \gamma_{1,t} \\ \frac{2}{K_2} \gamma_{2,t} \end{bmatrix} \quad (26)$$

Note that this specification of the market price of risk makes it proportional to the diffusion terms of the respective state variables, as in the case of the 1-Brownian motion arbitrage-free unit. Appendix A.1 contains the

proof that the 2-Brownian motion model shown here is indeed arbitrage-free. Within the Dai and Singleton (2000) classification scheme, the 2-BM AFU would be a special case of an $A_1(4)$ model since there are four state variables, and the correlation structure of the diffusion process is driven by a single state variable. Once again, although the 2-BM AFU can theoretically be viewed as a specific “model” of forward rates, it is not designed to be a complete model and should not be studied in isolation.⁹

2.4 Combining multiple arbitrage-free units

The parameterization of a single AFU (whether it is driven by one or more Brownian motions) is subject to the restriction that although its date-specific deviations can have arbitrary economic influences that decay separately at various rates, its maturity-specific deviation is permitted only one shape that decays at a fixed rate. By combining multiple AFUs it is possible to permit a wider set of effects for agents with different habitats, and hence additional maturity-specific influences on the forward curve. We show in this section that sums of independent AFUs are also arbitrage-free, thus adding flexibility in this way.

Assume that the i -th AFU follows the SDE:

$$df_i(s-t; t) = \theta_i(t, s)dt + \sigma_i(t, s)^T dB_i(t)$$

where

$$\sigma_i(t, s)^T \left(\int_t^s \sigma_i(v, t)dv - \kappa_{i,t} \right) = \theta_i(t, s).$$

All the units are independent of each other. Denote the forward curve as their sum:

$$f(s-t; t) = \sum_i f_i(s-t; t)$$

and

$$\theta(t, s) = \sum_i \theta_i(t, s).$$

Further denote $\vec{\sigma}(t, s)$, $\vec{\kappa}_t$ and \vec{B}_t as column vectors where $\sigma_i(t, s)$, $\kappa_{i,t}$ and $B_{i,t}$ respectively are stacked on one another in the same order. Then, it follows that

$$df(s-t; t) = \theta(t, s)dt + \vec{\sigma}(t, s)^T d\vec{B}_t$$

⁹ The extension to n Brownian motions is available from the authors.

and

$$\vec{\sigma}(t, s)^T \left(\int_t^s \vec{\sigma}(t, v) dv - \vec{\kappa}_t \right) = \theta(t, s)$$

We need to check that relation (9) holds in the combined system: $E[\exp(\int_0^T \vec{\kappa}_t^T d\vec{B}_t - \frac{1}{2} \int_0^T |\vec{\kappa}|^2 dt)] = 1$. This condition is easy to establish because each individual $\kappa_{i,t}$ satisfies that equality and the $\kappa_{i,t}$ are independent of each other:

$$\begin{aligned} & E \left[\exp \left(\int_0^T \vec{\kappa}_t^T d\vec{B}_t - \frac{1}{2} \int_0^T |\vec{\kappa}|^2 dt \right) \right] \\ &= E \left[\exp \left(\int_0^T \sum_i \kappa_{i,t} dB_{i,t} - \frac{1}{2} \int_0^T \sum_i |\kappa_{i,t}|^2 dt \right) \right] \\ &= E \left[\exp \left(\sum_i \left(\int_0^T \kappa_{i,t} dB_{i,t} - \frac{1}{2} \int_0^T |\kappa_{i,t}|^2 dt \right) \right) \right] \\ &= E \left[\prod_i \exp \left(\int_0^T \kappa_{i,t} dB_{i,t} - \frac{1}{2} \int_0^T |\kappa_{i,t}|^2 dt \right) \right] \\ &= \prod_i E \left[\exp \left(\int_0^T \kappa_{i,t} dB_{i,t} - \frac{1}{2} \int_0^T |\kappa_{i,t}|^2 dt \right) \right] \\ &= 1 \end{aligned}$$

Note also that the different AFUs need not have the same decay rate for their maturity-specific deviations. In this way, a selected forward curve can have several maturity-specific deviations that decay at different rates. This allows us to produce a maturity-specific deviation curve that can take on various shapes and follow a wider range of time-series dynamics. We interpret an individual AFU as an economic variable that drives the forward curve: these economic variables have effects on the forward curve that last for varying amounts of time. For instance, a temporary supply shock of 5-year loanable funds may be very short-lived, thereby corresponding to an AFU with a high decay rate for its maturity-specific deviation. On the other hand, a structural shift in the economy may produce a longer lasting effect on the forward curve, corresponding to an AFU with a low decay rate for its maturity-specific deviation.

2.5 Pricing zero-coupon bonds and interest rate derivatives

The price of a zero-coupon bond at date t maturing at a future date T , $P_{zc}(t, T)$, can be derived from the instantaneous forward rates via the formula:

$$P_{zc}(t, T) = e^{-\int_t^T f(s-t;t) ds}$$

The zero-coupon yield is then

$$\begin{aligned} y_{zc}(t, T) &= \frac{-\ln(P_{zc}(t, T))}{T-t} \\ &= \frac{\int_t^T f(s-t; t) ds}{T-t} \end{aligned} \quad (27)$$

In the case of a 1-Brownian motion AFU, $P_{zc}(t, T)$ is

$$\begin{aligned} P_{zc}(t, T) &= e^{-\int_t^T C_0 - C_1 e^{-2K_m(s-t)} + m(t)(e^{-K_m(s-t)} - e^{-2K_m(s-t)}) + d(t)e^{-2K_m(s-t)} ds} \\ &= e^{-\left[C_0(T-t) - C_1 \frac{e^{-2K_m(T-t)} - 1}{-2K_m} + m(t) \frac{e^{-K_m(T-t)} - e^{-2K_m(T-t)}}{-K_m} + d(t) \frac{e^{-2K_m(T-t)} - 1}{-2K_m} \right]} \\ &= e^{-\left[-C_0(T-t) - C_1 \frac{(e^{-2K_m(T-t)} - 1)}{2K_m} + m(t) \frac{2e^{-K_m(T-t)} - e^{-2K_m(T-t)} - 1}{2K_m} + d(t) \frac{(e^{-2K_m(T-t)} - 1)}{2K_m} \right]} \end{aligned}$$

Similarly, the zero-coupon yield can be expressed as

$$\begin{aligned} y_{zc}(t, T) &= \frac{-\ln(P_{zc}(t, T))}{T-t} \\ &= \frac{1}{T-t} \left[C_0(T-t) + C_1 \frac{(e^{-2K_m(T-t)} - 1)}{2K_m} \right] \\ &\quad - \frac{1}{T-t} \left[m(t) \frac{2e^{-K_m(T-t)} - e^{-2K_m(T-t)} - 1}{2K_m} \right] \\ &\quad + d(t) \frac{(e^{-2K_m(T-t)} - 1)}{2K_m} \end{aligned}$$

It is clear from this expression for $y_{zc}(t, T)$ that the zero-coupon yields are affine functions of the state variables. The prices of zero-coupon bonds and zero-coupon yields for 2-Brownian motion units, n -Brownian motion units and for multiple AFUs can be worked out in a similar fashion. Our class of models is a special case of the Affine Term Structure models studied by Duffie and Kan (1996) and the results for the general affine case in their paper are also applicable to our model.

Pricing any interest rate derivative in the framework of this model is also relatively simple. Given the diffusion term $\sigma(t, s)$, relation (10) gives us the drift under the risk-neutral measure, thereby specifying the risk-neutral SDE completely. The distribution of forward rates under the risk-neutral measure then follows from its risk-neutral SDE:

$$df(s-t; t) = \sigma(t, s)^T \left(\int_t^s \sigma(t, v) dv \right) dt + \sigma(t, s)^T dB_t^*$$

where B_t^* is an n -dimensional Brownian motion under the risk-neutral measure. The price of any derivative product is then obtained by taking

the expectations of the payoff, given that the forward rates follow the risk-neutral process specified above. The expectation of the payoff under the risk-neutral process can either be solved for in closed form from the PDE for the derivative security, or by performing Monte–Carlo simulations of the risk-neutral process. Aït-Sahalia (1996a) and Hull and White (1990) show how derivative prices can be calculated using these approaches.

3. Empirical Implementation

3.1 Data: Fama-Bliss treasury

For the period July 1964 to December 2004, we obtain monthly prices of 16 zero-coupon bonds of various maturities ranging from 8 days to approximately 5 years from the Center for Research in Security Prices (CRSP). For maturities of less than 1 year, we use the *Fama Treasury Bill Term Structure File*. For maturities of 1 year or more, we use the *Fama-Bliss Discount Bonds File*. Both these files are in the Monthly CRSP US Treasury Database.

The implied continuously-compounded forward rate for maturities between any two adjacent bonds is computed and taken as the instantaneous forward rate associated with a maturity that is at the mid-point between the two bonds' maturities.¹⁰ This procedure converts each adjacent pair of zero-coupon bonds into an instantaneous forward rate with an associated maturity. Thus, at each date we have 16 point estimates of instantaneous forward rates (we introduced a new bond with maturity zero and price \$1 at each date to get a total of 17 bonds and 16 adjacent pairs). Although the 16 point estimates of instantaneous forward rates do not have identical maturities across different dates, the maturities are nevertheless stable. Summary statistics of the constructed forward rates are displayed in Table 1.

¹⁰ For instance, if on date t we have 2 zero-coupon bonds with prices P_1 and P_2 maturing on dates T_1 and T_2 respectively, we set

$$f\left(\left(\frac{T_1 + T_2}{2} - t\right); t\right) = \frac{\ln(P_1) - \ln(P_2)}{T_2 - T_1}$$

Table 1
Summary statistics of constructed monthly Fama-Bliss forward rates, July 1964 to December 2004

Mean maturity (years)	SD Maturity (years)	Mean forward rates	SD Forward rates
0.03590	0.01018	0.05708	0.02723
0.11351	0.01841	0.06011	0.02837
0.19689	0.01855	0.06263	0.02936
0.28021	0.01840	0.06281	0.02897
0.36354	0.01834	0.06451	0.02922
0.44688	0.01816	0.06489	0.02935
0.53022	0.01801	0.06468	0.02821
0.61357	0.01813	0.06642	0.02875
0.69693	0.01816	0.06703	0.02937
0.78028	0.01815	0.06574	0.02892
0.86360	0.01842	0.06630	0.02804
0.95262	0.01027	0.06827	0.02708
1.5	0	0.06893	0.02645
2.5	0	0.07204	0.02449
3.5	0	0.07409	0.02427
4.5	0	0.07402	0.02358

Fama-Bliss zero-coupon prices each month are converted into implied continuously compounded forward rates by assuming a flat term-structure between any two adjacent bonds.

3.2 The statistical model

We explicitly model the three components of the forward curve as sums of exponentials:¹¹

$$M(\tau; t) = \sum_{i=1}^n \sum_{j=1}^{n_m} e^{k_i \tau} \mathbf{M}_{ij} m_j(t) \quad (28)$$

$$D(\tau; t) = \sum_{i=1}^n \sum_{j=1}^{n_d} e^{k_i \tau} \mathbf{D}_{ij} d_j(t) \quad (29)$$

$$U(\tau) = \sum_{i=1}^n u_i e^{k_i \tau} \quad (30)$$

In Section 2 we show that restrictions (for the 1- and 2-Brownian motion AFU examples given there) on k_i , \mathbf{M}_{ij} , \mathbf{D}_{ij} and u_i ensure that the model is arbitrage free. The models we estimate in this article have the property that

$$k_i = i \times K_m, i = 0, 1, \dots, n - 1 \quad (31)$$

for some free parameter $K_m > 0$ that is to be estimated from the data. For a given model the coefficients of the matrices \mathbf{M} and \mathbf{D} are fully determined by the arbitrage constraints. The coefficients u_i of the unconditional curve will be fitted to an average forward curve across all the data.

¹¹ For notational simplicity, relations (28) to (30) assume that there are n_m maturity-specific state variables and n_d date-specific state variables fitted to n bases.

Define the vector $\vec{k} = [k_1, \dots, k_n]'$, and the vector of exponentials $e^{\vec{x}} \equiv [e^{x_1}, \dots, e^{x_n}]'$. Then we can simplify the specifications for the three components to

$$M(\tau; t) = (e^{\vec{k}\tau})' \mathbf{M} \vec{m}(t) \quad (32)$$

$$D(\tau; t) = (e^{\vec{k}\tau})' \mathbf{D} \vec{d}(t) \quad (33)$$

$$U(\tau) = (e^{\vec{k}\tau})' \vec{u} \quad (34)$$

Putting these together we get

$$f(\tau; t) = (e^{\vec{k}\tau})' \left(\vec{u} + \mathbf{M} \vec{m}(t) + \mathbf{D} \vec{d}(t) \right)$$

The vector-valued stochastic processes $\vec{m}(t)$ and $\vec{d}(t)$ are modeled by the following SDEs:

$$d\vec{m}(t) = \mathbf{V}_m \vec{m} dt + \Sigma_m(\vec{m}) d\vec{B}(t) \quad (35)$$

$$d\vec{d}(t) = \mathbf{V}_d \vec{d} dt + \Sigma_d(\vec{m}) d\vec{B}(t) \quad (36)$$

The matrices \mathbf{V}_m and \mathbf{V}_d and the matrix valued functions $\Sigma_m(\vec{m})$ and $\Sigma_d(\vec{m})$ are determined by the model's no-arbitrage conditions. Note that the matrix valued function \mathbf{V}_d depends on \vec{m} and not on \vec{d} to satisfy the no-arbitrage condition.

Here is a simplified outline of our estimation procedure:

1. On date T , use all the data (across dates and maturities) available up to date T to fit the unconditional forward curve with exponentials (which are functions of K_m ; see relations (34) and (31)) to obtain $\hat{U}(\cdot)$.
2. Use a Kalman filter to estimate the stochastic process $\vec{z}_t \equiv f(\cdot; t) - \hat{U}(\cdot)$.
3. Maximize a quasi-likelihood to estimate K_m and σ^{*2} , where σ^{*2} is the variance of the measurement errors, which we define later in Section 3.4.

Each of these steps is more completely described in the remainder of this section.

3.3 Fitting the unconditional curve

Because the basis functions for the unconditional curve are parameterized as exponential functions of K_m we begin with an initial value for K_m . We then create an ‘‘average’’ forward curve by taking the mean maturity and mean forward rates for each of the 16 daily rates over the relevant sample period. The unconditional curve is fitted to this ‘‘average’’ forward curve using the basis functions. Given the fitted unconditional curve, we subtract

the unconditional rates at corresponding maturities from each observed forward rate in the sample, leaving a “deviations-only” data vector, which we define below as z_t and feed into the Kalman filter as observations.

3.4 Fitting the Kalman filter

A standard Kalman filter can be used to estimate a system of unobserved state variables in which the observed variables are linked to the unobserved state variables via a measurement equation, and the transition equation for the unobserved state variables is specified as a system of linear equations with Gaussian innovations (see Hamilton (1994) Chapter 13 for a discussion of the Kalman filter’s implementation and estimation). If the innovations in the unobserved state variables are not Gaussian (which is the case for our model), estimates from the standard Kalman filter are, in general, not conditionally unbiased estimators of the true state variables (Chen and Scott (2002)). However, it is still possible to proceed with the implementation of the Kalman filter by assuming that the innovations are indeed Gaussian in order to obtain a quasi-log-likelihood from the Kalman filter, and then optimize over that quasi-log-likelihood to obtain quasi-maximum likelihood (QML) estimates for parameters of the model.¹² The parameters in the model that we need to optimize over the quasi-log-likelihood are K_m , the decay rate of the maturity-specific deviation, and σ^{*2} , the variance of the measurement errors. Duffee and Stanton (2004) show that the use of QML via the Kalman filter in term-structure model estimation yields favorable results as compared to those from another common estimator, the Efficient Method of Moments of Gallant and Tauchen (1996). Duffee (2002) also highlights several other advantages of using QML including the fact that there is a positive probability that the fitted model could generate the empirically observed data, unlike method of moments-type estimators.

By viewing $\vec{m}(t)$ and $\vec{d}(t)$ as latent state variables we are able to fit our model directly into a Kalman filter framework. Stack a sequence of maturities into a vector $\vec{\tau} = [\tau_1, \dots, \tau_\ell]'$. Next place $\vec{m}(t)$ and $\vec{d}(t)$ into a vector \vec{x}_t :

$$\vec{x}_t = \begin{bmatrix} \vec{m}(t) \\ \vec{d}(t) \end{bmatrix}. \quad (37)$$

At each date t , we can relate these to the observed data with the measurement equation:

$$\vec{z}_t \equiv f(\vec{\tau}; t) - \hat{U}(\vec{\tau}) = \mathbf{A}\vec{x}_t + \vec{\epsilon}_t \quad (38)$$

¹² Several authors (including Geyer and Pichler (1999), Chen and Scott (2002) and DeJong and Santa-Clara (1999)) have also estimated term-structure models with nonGaussian innovations and made use of such a QML estimator.

where \mathbf{A} is the measurement matrix for the state variables, and $\vec{\epsilon}_t$ is the vector of measurement errors. The j -th row of the matrix \mathbf{A} is defined as

$$\mathbf{A}_j \equiv \left[(e^{\vec{k}\tau_j})' \mathbf{M}, (e^{\vec{k}\tau_j})' \mathbf{D} \right]. \quad (39)$$

To allow for statistical estimation, we now simplify the model by adding the assumption that the measurement errors are homoscedastic and both cross-sectionally and serially uncorrelated:

$$\Sigma_\epsilon \equiv \text{Var}(\vec{\epsilon}_t) = \sigma^{*2} I. \quad (40)$$

We estimate the noise variance σ^{*2} from the data when maximizing the quasi-likelihood.

We can now derive the transition equation of the Kalman filter as the discretized version of the stochastic process for \vec{x}_t . First let

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_d \end{bmatrix} \text{ and } \Sigma(\vec{m}) = \begin{bmatrix} \Sigma_m(\vec{m}) \\ \Sigma_d(\vec{m}) \end{bmatrix}.$$

The transition equation is therefore

$$\vec{x}_t = W \vec{x}_{t-1} + \xi_t, \quad (41)$$

where W is a diagonal matrix with

$$W_{ii} = e^{\delta V_{ii}}, \quad (42)$$

where δ is the step size, and we approximate $Q_t \equiv \text{Var}_{t-1}(\xi_t)$ by

$$Q_t \approx \delta \Sigma(\vec{m}) \Sigma(\vec{m})'. \quad (43)$$

Given this specification for the Kalman filter, we set the initial estimates of the state vector at its unconditional mean, which is zero ($\hat{x}_0 = 0$), and set the initial covariance matrix at the unconditional variance $\text{Var}(\vec{x}_t)$. We can then run the Kalman filter to estimate the state variables by iterating between the prediction equations and the updating equations as in DeJong and Santa-Clara (1999), Geyer and Pichler (1999) and Babbs and Nowman (1999).¹³ We provide a copy of the standard equations of the Kalman filter in Appendix A.4.

Standard code for the Kalman filter generates a log-likelihood function which we maximize to fit the parameters. The next section shows how to use this likelihood to select a model.

¹³ The framework of the model places boundaries on the values of some of the state variables. The diffusion terms, which are functions of the maturity-specific state variables, must be constrained to be nonnegative. This in turn places constraints on those state variables. In the empirical implementation, a simple and common way of enforcing this restriction is to replace the values of the state variables that do breach the constraints with ones that just satisfy it. See Chen and Scott (2002) and Geyer and Pichler (1999) for further examples of such restrictions in a Kalman filter.

3.5 Picking a particular model

Sections 2.1 through 2.4 showed how AFUs can be constructed and combined so that the resulting forward curve $f(\tau; t)$ has arbitrage-free dynamics. We now select a specific model combining AFUs that will later be used in Section 4 for estimation, forecasting, and comparison to benchmark models.

To select a model for empirical implementation, we evaluated seven candidate models over a period of training data (July 1964 to June 1984) that is prior to the period in which we examine out-of-sample forecasting (July 1984 onwards). This is to ensure that the forecasts generated in later sections are truly out-of-sample—both model selection and parameter estimation are dependent only on the training data available up to the date the forecast is made. We evaluate the model using both log-likelihood and the Akaike Information Criterion (AIC, Akaike (1973)). The log-likelihood function for each model is directly obtainable from the Kalman filter that we implement and describe in Section 3.4. The AIC adjusts the log-likelihood of a model by penalizing additional degrees of freedom. Our latent state space approach does not lend itself to the usual application of AIC for model selection; however, a correction for degrees of freedom can still be implemented by adjusting the log-likelihood appropriately. We present the log-likelihood, the number of state variables, the number of free parameters, and the AIC for each of the seven models that we test in Table 2. As that table shows, model 6 has the highest log-likelihood as well as the lowest AIC value (the log-likelihood for model 6 is much larger than the closest competitor; for reasonable penalty functions used to adjust for the number of free parameters, the relative rank among the competing models will not change). Model 6 is thus the model of our choice (henceforth the Chua, Foster, Ramaswamy, Stine (CFRS) model).

4. The CFRS Model—Structure and Forecasts

4.1 Description of the CFRS model

The CFRS model (model 6) combines two 2-Brownian motion AFUs and fits the forward curve to three exponential functions $\{e^{-K_m}, e^{-2K_m}, e^{-4K_m}\}$. The parameterization of the forward curve (now written $F(\tau; t)$ to distinguish it from the explicit versions of AFUs in Section 2) is as follows:

$$F(\tau; t) = (e^{\vec{k}\tau})' (\vec{u} + \mathbf{M}\vec{m}(t) + \mathbf{D}\vec{d}(t)) + \epsilon(\tau; t), \quad (44)$$

where

$$(e^{\vec{k}\tau}) = \begin{bmatrix} 1 \\ e^{-K_m\tau} \\ e^{-2K_m\tau} \\ e^{-4K_m\tau} \end{bmatrix}, \mathbf{M} = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ -2 & -1 \\ 0 & -1 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix},$$

Table 2
Log-likelihood and Akaike Information Criterion values (AIC) in model selection using monthly Fama-Bliss forward rates from July 1964 to June 1984

Model	Composition	State dimensions	Free parameters	Log-likelihood	AIC
Model 1	AFU1	2	3 $\{C_0, C_1, K_m\}$	12417	-24827
Model 2	AFU2	3	3 $\{C_0, C_1, K_m\}$	12716	-25427
Model 3	AFU3	4	4 $\{C_0, C_1, C_2, K_m\}$	12904	-25800
Model 4	AFU1 + AFU2	5	3 $\{C_0, C_1, K_m\}$	12600	-25194
Model 5	AFU1 + AFU3	6	4 $\{C_0, C_1, C_2, K_m\}$	12872	-25736
Model 6	AFU2 + AFU3	7	4 $\{C_0, C_1, C_2, K_m\}$	13014	-26020
Model 7	AFU1 + AFU2 + AFU3	9	4 $\{C_0, C_1, C_2, K_m\}$	12999	-25990

We use the Fama-Bliss training data to derive the log-likelihood and AIC values for seven competing models. The model with the highest log-likelihood and lowest AIC value is our model of choice (CFRS model). AIC is calculated via the formula $-2 \ln L + 2K$, where L is the likelihood, and K is the number of free parameters in the model. The models that we consider are made up of combinations of the following three AFUs (where $m_i(t)$ refers to the maturity-specific state variable of the i -th AFU, and $d_{j,k}(t)$ refers to the k -th date-specific state variable of the j -th AFU):

1. AFU1: $U(s-t) + m_1(t)(e^{-K_m(s-t)} - e^{-2K_m(s-t)}) + d_{1,1}(t)e^{-2K_m(s-t)}$
2. AFU2: $U(s-t) + m_2(t)(2e^{-K_m(s-t)} - 2e^{-2K_m(s-t)}) + 2d_{2,1}(t)e^{-2K_m(s-t)} + d_{2,2}(t)e^{-K_m(s-t)}$
3. AFU3: $U(s-t) + m_3(t)(2e^{-K_m(s-t)} - e^{-2K_m(s-t)} - e^{-4K_m(s-t)}) + d_{3,1}(t)e^{-2K_m(s-t)} + d_{3,2}(t)e^{-4K_m(s-t)} + d_{3,3}(t)e^{-2K_m(s-t)}$

$$\vec{m}(t) = \begin{bmatrix} m_1(t) \\ m_2(t) \end{bmatrix}, \vec{d}(t) = \begin{bmatrix} d_1(t) \\ d_2(t) \\ d_3(t) \\ d_4(t) \\ d_5(t) \end{bmatrix}, \vec{u} = \begin{bmatrix} C_0 \\ 0 \\ -C_1 \\ -C_2 \end{bmatrix}.$$

In this parameterization of the forward curve $m_1(t)$, $d_1(t)$ and $d_2(t)$ correspond to the maturity and date-specific deviations of the first 2-Brownian motion AFU and $m_2(t)$, $d_3(t)$, $d_4(t)$ and $d_5(t)$ correspond to the maturity and date-specific deviations of the second 2-Brownian motion AFU, and these AFUs are independent. Therefore there are seven state variables in this system: $m_1(t)$, $m_2(t)$, $d_1(t)$, $d_2(t)$, $d_3(t)$, $d_4(t)$ and $d_5(t)$.

The stochastic processes for vectors $\vec{m}(t)$ and $\vec{d}(t)$ are

$$d\vec{m}(t) = \mathbf{V}_m \vec{m} dt + \Sigma_m(\vec{m}) d\vec{B}(t) \quad (45)$$

$$d\vec{d}(t) = \mathbf{V}_d \vec{d} dt + \Sigma_d(\vec{m}) d\vec{B}(t) \quad (46)$$

where

$$\mathbf{V}_m = \begin{bmatrix} -K_m & 0 \\ 0 & -K_m \end{bmatrix},$$

$$\mathbf{V}_d = \begin{bmatrix} -2K_m & 0 & 0 & 0 & 0 \\ 0 & -K_m & 0 & 0 & 0 \\ 0 & 0 & -2K_m & 0 & 0 \\ 0 & 0 & 0 & -4K_m & 0 \\ 0 & 0 & 0 & 0 & -2K_m \end{bmatrix}$$

$$\Sigma_m(\vec{m}) = \begin{bmatrix} \gamma_{1,t} & 0 & 0 & 0 \\ 0 & 0 & \gamma_{3,t} & 0 \end{bmatrix}, \Sigma_d(\vec{m}) = \begin{bmatrix} \gamma_{1,t} & 0 & 0 & 0 \\ 0 & \gamma_{2,t} & 0 & 0 \\ 0 & 0 & \gamma_{3,t} & 0 \\ 0 & 0 & \gamma_{3,t} & 0 \\ 0 & 0 & 0 & \gamma_{4,t} \end{bmatrix},$$

$$dB_t = \begin{bmatrix} dB_{1,t} \\ dB_{2,t} \\ dB_{3,t} \\ dB_{4,t} \end{bmatrix},$$

using four independent Brownian motions, and

$$\begin{aligned} \gamma_{1,t}^2 &= \left(m_1(t) + \frac{2}{3}C_1 \right) \frac{K_m^2}{4} \\ \gamma_{2,t}^2 &= \left(m_1(t) + \frac{2}{3}C_1 \right) K_m^2 \\ \gamma_{3,t}^2 &= \left(m_2(t) + \frac{2}{3}C_1 \right) \frac{K_m^2}{4} \\ \gamma_{4,t}^2 &= (3m_2(t) + 4C_2) 2K_m^2 \end{aligned}$$

In Appendix A.2 we show that this model satisfies the HJM conditions. As before, this requires us to specify a market price of risk, which is

$$\hat{\kappa}_t = \begin{bmatrix} \kappa_{1,t} \\ \kappa_{2,t} \\ \kappa_{3,t} \\ \kappa_{4,t} \end{bmatrix} = \begin{bmatrix} \frac{2}{K_m} \gamma_{1,t} \\ \frac{1}{K_m} \gamma_{2,t} \\ \frac{2}{K_m} \gamma_{3,t} \\ \frac{1}{2K_m} \gamma_{4,t} \end{bmatrix} \quad (47)$$

Within the Dai and Singleton (2000) classification scheme, the CFRS model is a special case of an $A_2(7)$ model, because there are seven state variables and the correlation structure of the diffusion process is driven by two state variables: $m_1(t)$ and $m_2(t)$. While estimation of a seven-state variable model is typically intractable due to the large number of parameters that need to be estimated, this is not the case with the CFRS model. In the seven-state variable framework, the CFRS model already has most of the parameters fixed relative to the maximally flexible member of that class and only a few parameters need to be estimated. This makes the seven-state variable CFRS model both simpler and easier to estimate than even the maximal three-state variable model.

To implement the CFRS model within the Kalman filter framework we set $\delta = \frac{1}{12}$ and follow the procedure described in Section 3.2.

4.2 Fitting the CFRS model to the full sample

In this section, we fit the CFRS model to the full sample to demonstrate the estimation procedure and illustrate its properties. It should be emphasized that this is not an out-of-sample forecasting exercise; that is left to Sections 4.3 to 4.5.

We first estimate the unconditional curve. Fitting the unconditional curve on the original bases, $\{1, e^{-2K_m(s-t)}, e^{-4K_m(s-t)}\}$, results in unstable estimates of C_1 and C_2 due to the significant collinearity between $e^{-2K_m(s-t)}$ and $e^{-4K_m(s-t)}$. So, we fit the unconditional curve to two bases, $\{1, e^{-2K_m(s-t)}\}$. In other words, instead of using $\{1, e^{-2K_m(s-t)}, e^{-4K_m(s-t)}\}$ as the bases for fitting C_0 , C_1 and C_2 , we use $\{1, e^{-2K_m(s-t)}\}$ as bases for fitting C_0 and \tilde{C}_1 . We then make the assumption that $C_1 = \frac{2}{3}\tilde{C}_1$ and $C_2 = \frac{1}{3}\tilde{C}_1$.¹⁴ Summary statistics for the unconditional curve of the CFRS model are displayed in Table 3. The coefficients for C_0 and \tilde{C}_1 are 0.08045 and 0.01911 respectively, implying an unconditional curve that starts at 6.134% and rises monotonically to 8.045% at infinite maturity.

Next, we construct the Kalman filter “observations” by subtracting the estimated unconditional rates from the forward rates. Using this data and the Kalman filter equations specified in the previous section, we can calculate the log-likelihood for any given set of parameters. We then optimize over the parameter space to find the parameter values (K_m, σ^{*2}) that maximize this quasi-log-likelihood.

Summary statistics for the fitted curves of the CFRS model are displayed in Table 4 while Figure 3 displays the fitted unconditional curve, the “average” curve and the average fitted curve of the CFRS model. The half-life of the maturity-specific deviation is 4.30 years, corresponding to the estimate of K_m of 0.1612. It is important to note that this number is not comparable to the half-lives estimated in the CIR and Vasicek models, because our model also simultaneously estimates a date-specific deviation that, at zero maturity, dictates the level of the short term or spot rate. The fitted spot rate on any date t is $F(0; t) = U(0) + D(0; t)$. Therefore, we have

$$F(0; t) = U(0) + 2d_1(t) + d_2(t) + d_3(t) + d_4(t) + d_5(t) \quad (48)$$

By looking at the slope of the estimated curve of date-specific deviations at zero maturity we can then estimate the instantaneous drift of the spot

¹⁴ This particular split of \tilde{C}_1 into C_1 and C_2 permits maximal flexibility for the evolutionary dynamics of $m_2(t)$. Notice that the diffusion terms of the state variables, $\gamma_{j,t}$ must be nonnegative. In particular, the model restricts $\gamma_{3,t}^2 \geq 0$ and $\gamma_{4,t}^2 \geq 0$. This translates to the following constraints: $m_2(t) \geq -\frac{2}{3}C_1$ and $m_2(t) \geq -\frac{4}{3}C_2$. To allow for maximal flexibility, it is ideal for both constraints to be binding simultaneously. We therefore have $C_1 = 2C_2$, which then leads to the proposed split of $C_1 = \frac{2}{3}\tilde{C}_1$ and $C_2 = \frac{1}{3}\tilde{C}_1$.

Table 3
Unconditional curve of CFRS model, Fama-Bliss treasury data, June 1964 to December 2004

Coefficient	Fit	Lower 95% CI	Upper 95% CI
C_0	0.08045	0.07726	0.08364
\tilde{C}_1	0.01911	0.01500	0.02322

The unconditional curve is of the form $U(s-t) = C_0 - \tilde{C}_1 e^{-2K_m(s-t)}$, where $K_m = 0.1612$. An “average” forward curve is created by taking the mean maturity and mean forward rates for each of the 16 daily rates across time. The values C_0 and \tilde{C}_1 are then obtained through the least squares criterion. We report the fitted values and their 95% confidence intervals.

Table 4
Estimated state variables of CFRS model, Fama-Bliss treasury data, June 1964 to December 2004

State Variable	Mean	SD
$m_1(t)$	-0.00087	0.01145
$d_1(t)$	0.00285	0.00736
$d_2(t)$	-0.03314	0.03069
$m_2(t)$	0.01212	0.01314
$d_3(t)$	0.00082	0.01038
$d_4(t)$	-0.00209	0.00824
$d_5(t)$	0.02828	0.04378

We choose a set of parameters that maximizes the quasi-log-likelihood of the Kalman filter. This table provides the summary statistics of the state variables generated by the Kalman filter using that set of optimal parameters.

rate at time t :

$$\left. \frac{\partial D_t(\tau)}{\partial t} \right|_{\tau=0} = -4K_m d_1(t) - K_m d_2(t) - 2K_m d_3(t) - 4K_m d_4(t) - 2K_m d_5(t), \quad (49)$$

showing that the current values of the state variables in the CFRS specification that drives the date-specific component also dictate the local drift of the spot rate.

From Table 4, we also observe that the first AFU is dominated by one state variable, $d_2(t)$ which has a strong negative mean and a relatively large standard deviation. The second AFU is dominated by $d_5(t)$, which has a strong positive mean and an even larger standard deviation. The average deviation is approximately zero, resulting in close agreement between the average fitted curve and the unconditional curve, as observed in Figure 3.

Figure 4 displays the term-structure of the time-series standard deviations of the maturity-specific deviation, the date-specific deviation and the total deviation (sum of maturity- and date-specific deviations). This figure shows that the standard deviation of the total deviation decreases as maturity increases. At short maturities, the date-specific deviation accounts for most of the variability; beyond maturities of 4 years, the variability of the maturity-specific deviation exceeds that of the date-specific deviation.

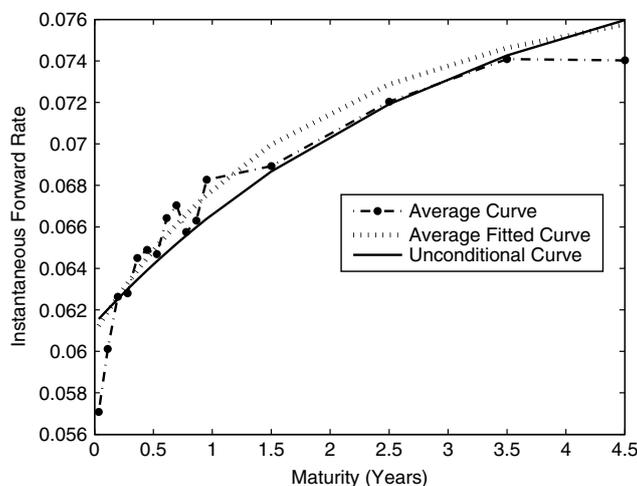


Figure 3
Unconditional curve, “Average” curve and average fitted curve of CFRS model, Fama-Bliss treasury data, June 1964 to December 2004

An “average” forward curve is created by taking the mean maturity and mean forward rates for each of the 16 daily rates across time. The data used to derive the “average” forward curve are reported in columns 1 and 3 of Table 1. The unconditional curve is of the form $U(s-t) = C_0 - \tilde{C}_1 e^{-2K_m(s-t)}$, where $K_m = 0.1612$. The values C_0 and \tilde{C}_1 are then obtained through the least squares criterion, and are reported in Table 3. The average fitted curve is obtained by creating an average deviations curve based on the average state variables reported in Table 4, and adding that average deviations curve to the unconditional curve.

4.3 Out-of-sample forward rate forecasts from the CFRS model

We evaluate the predictive power of the CFRS model by comparing its out-of-sample forecasting accuracy to that of standard benchmarks such as the RW model, the EH model and the Expectations Hypothesis with Term-Premium model (EHTP). To ensure that our forecasts are truly out-of-sample, all fits and parameter values are calibrated using only data available on or before the date that any forecast is made. In making out-of-sample forecasts, we use a rolling training interval with a fixed length of 20 years. The fitted curves and parameters based on the 20-year training data are then used to forecast future forward curves 3-months, 6-months, 12-months and 24-months from the last date in the training data. The training data is then rolled ahead by one month to estimate a new set of parameters to be used in making forecasts from the last date in the new training set.¹⁵

¹⁵ For instance, data from July 1964 to June 1984 are used to calibrate the Kalman filter parameters via QML and to get the estimated state variables for June 1984. These parameters, along with the estimated state variables for June 1984, are then used to forecast the forward curve for September 1984, December 1984, June 1985 and June 1986. Then, data from August 1964 to July 1984 are used to produce forecasts for forward curves 3-months, 6-months, 12-months and 24-months from July 1984. This process is repeated until the final forecast is for December 2004, which is the end of our sample.

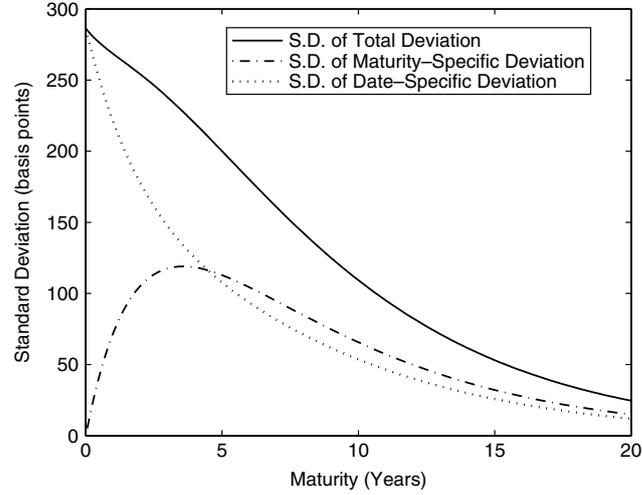


Figure 4
Standard deviation of Maturity-Specific deviation, Date-Specific deviation and total deviation of the CFRS model, Fama-Bliss treasury data, June 1964 to December 2004

Using the full sample period from June 1964 to December 2004, we compute the monthly maturity-specific, date-specific and total deviation based upon the monthly fitted state variables obtained from the Kalman filter. We then report the time-series standard deviation of these quantities at each maturity up to 20 years.

Generating forecasts of future forward curves given the fitted parameters and the state variables on the last day of the training data is straightforward. We only need to generate forecasts of the state variables at the future date, and then convert those forecasted state variables into the implied forward curve. Assuming that the last date in the training period is t , we generate the forecast for a future date T in the CFRS model by using

$$\begin{aligned}\hat{m}_1(T) &= E[m_1(T) | \hat{m}_1(t)] = \hat{m}_1(t) e^{-\hat{K}_m(T-t)} \\ \hat{d}_1(T) &= E[d_1(T) | \hat{d}_1(t)] = \hat{d}_1(t) e^{-2\hat{K}_m(T-t)} \\ \hat{d}_2(T) &= E[d_2(T) | \hat{d}_2(t)] = \hat{d}_2(t) e^{-\hat{K}_m(T-t)} \\ \hat{m}_2(T) &= E[m_2(T) | \hat{m}_2(t)] = \hat{m}_2(t) e^{-\hat{K}_m(T-t)} \\ \hat{d}_3(T) &= E[d_3(T) | \hat{d}_3(t)] = \hat{d}_3(t) e^{-2\hat{K}_m(T-t)} \\ \hat{d}_4(T) &= E[d_4(T) | \hat{d}_4(t)] = \hat{d}_4(t) e^{-4\hat{K}_m(T-t)} \\ \hat{d}_5(T) &= E[d_5(T) | \hat{d}_5(t)] = \hat{d}_5(t) e^{-2\hat{K}_m(T-t)}\end{aligned}$$

With the forecasted future forward curves, we can calculate the forecast errors as the differences between the forecasted forward rates and the observed forward rates on the forecasted date. In reporting the forecast

performance, we put our forecast errors into several maturity buckets: 0 to 1 year, 1 to 5 years and 0 to 5 years.

4.4 Out-of-sample forward rate forecasts from other benchmark models

We use several benchmark models to act as comparisons to the forecasts of our model: the RW model, the EH model and the EHTP model.

To generate forecasts from the RW model, we fit each date's observed Fama-Bliss forward rates to the following bases: $\{\beta_{\tau_i}\} \equiv \{1, e^{-0.5\tau_i}, e^{-1\tau_i}, e^{-1.5\tau_i}\}$,¹⁶ where τ_i is the maturity of the i -th forward rate for that date, thereby generating a smooth fit for the date's forward rates. The forecast of the RW model for any future date is the same fitted curve.¹⁷

To generate forecasts from the EH model, we fit each date's forward rates to the bases $\{\beta_{\tau_i}\}$. On date t , let the estimated coefficients corresponding to the basis functions $\{\beta_{\tau_i}\}$ be $\{\widehat{EH}_1(t), \widehat{EH}_2(t), \widehat{EH}_3(t), \widehat{EH}_4(t)\}$ respectively. The forecast of the EH model for date T in the future, conditional on the fit on date t , would be based on the following coefficients: $\{\widehat{EH}_1(t), \widehat{EH}_2(t)e^{-0.5(T-t)}, \widehat{EH}_3(t)e^{-1(T-t)}, \widehat{EH}_4(t)e^{-1.5(T-t)}\}$. This implies that if the fitted forward rates on date t are

$$f(s-t; t) = \widehat{EH}_1(t) + \widehat{EH}_2(t)e^{-0.5(s-t)} \\ + \widehat{EH}_3(t)e^{-1(s-t)} + \widehat{EH}_4(t)e^{-1.5(s-t)},$$

then the forecasted forward rate on some future date T for maturity on date s would be the same as the forward rate on date t for maturity on date s :

$$\hat{f}(s-T; T) = E[f(s-T; T) | f(s-t; t)] = f(s-t; t) \\ = \widehat{EH}_1(t) + \widehat{EH}_2(t)e^{-0.5(s-t)} \\ + \widehat{EH}_3(t)e^{-1(s-t)} + \widehat{EH}_4(t)e^{-1.5(s-t)}.$$

To generate forecasts from the EHTP model, we first estimate an average term-premium by calibrating the implied "steady state" curve to match as closely as possible the "average" curve described in the previous section via the least-squares criterion. We parameterize the "steady state" curve as

$$f_{SS}(s-t) = SS_0 + SS_1e^{-0.5(s-t)}$$

¹⁶ We investigate alternative specifications of the bases for the benchmark models, including $\{1, e^{-K_m\tau_i}, e^{-2K_m\tau_i}, e^{-4K_m\tau_i}\}$, but these have much poorer out-of-sample forecasting accuracy than the ones that we adopt.

¹⁷ An alternative method of generating forecasts for the RW model is simply to use the actual rates, as opposed to fitted rates, as forecasts of future rates. We implemented this method and found that the out-of-sample forecasting accuracy of this method is virtually identical to the one that we adopt.

where a negative value for SS_1 indicates an upward-sloping “steady state” curve, and therefore positive term-premiums. We assume that this term-premium is time-invariant. Next, we subtract the steady state rates from all the forward rates in the data, leaving us with the residuals. These residuals are then assumed to conform to the EH. Again, we fit the residuals to the bases: $\{\beta_{\tau_i}\}$. On date t , let the estimated coefficients corresponding to the basis functions $\{\beta_{\tau_i}\}$ be $\{\widehat{TP}_1(t), \widehat{TP}_2(t), \widehat{TP}_3(t), \widehat{TP}_4(t)\}$ respectively. The forecast of these coefficients for date T in the future conditional on the coefficients on date t would be $\{\widehat{TP}_1(t), \widehat{TP}_2(t)e^{-0.5(T-t)}, \widehat{TP}_3(t)e^{-1(T-t)}, \widehat{TP}_4(t)e^{-1.5(T-t)}\}$. To generate the forecasts of the forward rates, we need to add back the term premiums:

$$\begin{aligned} \hat{f}(s - T; T) = & \widehat{SS}_0 + \widehat{SS}_1 e^{-0.5(s-t)} + \widehat{TP}_1(t) + (\widehat{TP}_2(t)e^{-0.5(T-t)}) \\ & \times e^{-0.5(s-T)} + (\widehat{TP}_3(t)e^{-1(T-t)})e^{-1(s-t)} + (\widehat{TP}_4(t)e^{-1.5(T-t)})e^{-1.5(s-t)} \end{aligned}$$

We compare the accuracy of the forecasts generated from our model, the RW model, the EH model, and the EHTP model by looking at the forecast errors generated by each model. We first compute the difference in RMSE between two competing models. We then use the Newey–West estimator (1987) to compute the variance estimate of the RMSE-difference series, correcting for auto-correlation and heteroscedasticity¹⁸ in the series. The z -score (NW-stat) for the significance of differences between two competing forecasts can then be directly derived from the differences in means and the computed variance.

The results are shown in Table 5. At the 3-month horizon, the forecasts of the CFRS model mildly underperform relative to the RW model and mildly outperform relative to the EH and the EHTP models (a negative value of the NW-stat indicates that the CFRS model performs better than the competing models). The CFRS model’s performance at the 3-month horizon can be attributed to its less than perfect cross-sectional fits: the CFRS model’s fits are more constrained due to stronger restrictions in its parameterization. Short horizon forecasts are necessarily very similar to the cross-sectional fit. Therefore, poor cross-sectional fits naturally result in poor short horizon forecasts. Another explanation for the observed forecast performance at short horizons is that the signal-to-noise ratio here is extremely low so most of the innovations in the forward curve at short horizons are noise, and are likely left unexplained by any model. At horizons of 12 months or longer, a good predictive model should be able to capture more signal relative to the random movements of the forward curve. This explanation is supported by the fact that the relative forecasting performance of the CFRS model improves dramatically when we move from the 3-month to the 6-, 12- or 24-month horizons.

¹⁸ See Diebold and Mariano (1995) for another possible test of significance for auto-correlated series.

Table 5
RMSE and NW-stat for out-of-sample forward rate forecasts, June 1964 to December 2004

Maturity Bucket (years) Number of Bonds	3-month-ahead forecast			6-month-ahead forecast			12-month-ahead forecast			24-month-ahead forecast		
	0 to 1 12	1 to 5 4	0 to 5 16	0 to 1 12	1 to 5 4	0 to 5 16	0 to 1 12	1 to 5 4	0 to 5 16	0 to 1 12	1 to 5 4	0 to 5 16
RMSE (basis points)												
CFRS	67.09	62.93	67.95	89.68	79.36	89.63	122.61	102.03	121.03	152.33	101.42	143.77
RW	66.83	60.62	66.97	92.00	82.79	92.21	134.99	113.46	133.56	194.14	133.76	183.90
EH	73.04	62.19	72.00	109.83	87.04	106.39	169.27	121.23	160.93	238.49	139.84	219.78
EHTP	69.43	61.34	69.02	100.94	84.69	98.94	150.01	115.96	144.58	204.20	129.04	190.00
NW-stat												
CFRS vs RW	0.185	1.521	0.901	-0.768	-1.664	-1.067	-1.822	-3.794	-2.268	-3.572	-5.499	-3.928
CFRS vs EH	-2.456	0.427	-2.035	-3.106	-2.776	-3.091	-3.253	-3.692	-3.314	-3.381	-4.358	-3.486
CFRS vs EHTP	-1.131	0.944	-0.630	-2.183	-2.165	-2.159	-2.462	-3.391	-2.523	-2.772	-4.337	-2.920

This table shows the 3-, 6-, 12- and 24-month-ahead forward rate forecasts produced by the CFRS model, the Random Walk (RW) model, the EH model and the EHTP model. We use the first 20 years of training data to generate the first forecast. The training window is then moved forward one month at a time to generate successive forecasts. We first calculate the cross-sectional RMSE of forecast error within each maturity bucket in any given month. We then calculate the time-series mean of the cross-sectional RMSE and report that mean in this table. The NW-stat (quoted as a z-score) is used to test the significance of the differences in RMSE between any two models. A negative value of the NW-stat indicates that the first model (the model mentioned before “vs”) is performing better than the second model.

At the 6-month horizon, the CFRS model significantly outperforms the EH model. While the CFRS model also outperforms the RW and EHTP models, the difference in RMSE is not statistically significant. At the 12-month and 24-month horizons, the CFRS model performs significantly better than all benchmark models, at all maturities. The only exception is the 12-month ahead forecast of short maturity forward rates where the CFRS does not outperform the RW model significantly.

4.5 Out-of-sample yield forecasts

Forward rates are seldom directly forecasted: most authors including Duffee (2002) and Diebold and Li (2006) use their models to forecast bond yields. Thus, to compare the accuracy of forecasts of our model with forecasts made by other models, we translate the forecasts of forward rates from our model into forecasts of yields. Because the forward curve in our model is parameterized as sums of exponential functions, yields are analytically obtainable via relation (27). We also convert the forecasts of forward rates from the RW model into the implied forecasts of yields. The differences between these forecasts and the actual realized yields from the Fama-Bliss zero-coupon bond data are taken to be the forecast errors. As additional yield-based benchmarks, we replicate Diebold and Li's (DL) procedure as well as the various classes of completely affine and essentially affine models studied in Duffee (2002).

Diebold and Li forecast the yield curve using US Treasury bonds of fixed maturities from January 1985 to December 2000 by applying an autoregressive model to the fitted coefficients from the Nelson–Siegel model (1987). They use a 9-year training window from January 1985 to January 1994. Thus, their out-of-sample test statistic is generated from approximately 6 remaining years of out-of-sample forecasts.

Duffee forecasts government bond yields using data from January 1952 to December 1998. He studies several classes of completely affine and essentially affine models. For each class of models, he employs QML to estimate the parameters of the models and to generate out-of-sample forecasts; he uses a 43-year training window from January 1952 to December 1994 to generate approximately 4 years of out-of-sample forecasts (from January 1995 to December 1998).

For the purpose of this study, we repeat the forecasting techniques of Diebold and Li, and Duffee. However, we apply their procedures on our full data sample: June 1964 to December 2004. Similar to the forward rate forecasts, we use a 20-year training window, generating approximately 20 years of out-of-sample forecasts.¹⁹ We compare the relative performance

¹⁹ Some authors have argued that the years of the “Fed Experiment” from 1979 to 1982 are an unusual period for interest rates. To ensure the robustness of our results, we repeat the entire forecasting exercise using only data from January 1983 through December 2004, with a 10-year training window. The results are the same, albeit with lower statistical significance due to the shorter sample period.

of the CFRS model, the RW model, the DL procedure (DL model) as well as multiple classes of models from Duffee (2002). Specifically, in replicating Duffee's study, we compute out-of-sample forecasts for both completely affine and essentially affine versions of maximal $A_0(3)$, $A_1(3)$ and $A_2(3)$ models using QML. Following Duffee, we use a set of 1000 random admissible starting parameters, and all parameters that have t -statistics less than one are set to zero.²⁰ The forecasts from the maximal $A_1(3)$ models are, in general, better than those from the $A_0(3)$ and $A_2(3)$ models. Therefore, we only report the forecasting performance of the maximal completely affine $A_1(3)$ model (written as CA_A13) and the maximal essentially affine $A_1(3)$ model (written as EA_A13).

We report the RMSE and NW-stats of the out-of-sample forecasts from the various methods in Table 6. Here is a summary of the results:

1. At the 3-month ahead horizon, the CFRS model significantly outperforms all competing forecasts for yields between 0 and 1 year to maturity. However, it underperforms the RW model for yields between 1 and 5 years to maturity, although not significantly. The CFRS model also outperforms the DL and EA_A13 models significantly for maturities between 1 to 5 years.
2. At the 6-month ahead forecast horizon, the CFRS model outperforms all competitors across all maturities. These differences in performance are also statistically significant for maturities from 0 to 1 year (DL model, CA_A13 model and EA_A13 model) and from 1 to 5 years (for DL model and EA_A13 model).
3. At the 12-month and 24-month ahead horizon, the forecasts from the CFRS model are significantly better than all competing forecasts across all maturities.

The remarks in the summary above indicate that the CFRS model has more forecast power than all comparable methods, especially at longer horizons and maturities.

5. Conclusion

We have introduced a class of models for forward rates that is arbitrage free while retaining coherent and economically sensible dynamics. The three components of this class of models, namely the unconditional curve, the date-specific deviation and the maturity-specific deviation are economically easily interpretable and are consistent with other models and hypotheses relating to the term structure of interest rates.

²⁰ However, instead of using simplex followed by NPSOL to execute the optimization, we use the *fminunc* function in Matlab to find the set of optimal parameters in any given iteration.

Table 6
RMSE and NW-stat for Out-of-sample yield forecasts, June 1964 to December 2004

Maturity Bucket (years) Number of Bonds	3-month-ahead forecast			6-month-ahead forecast			12-month-ahead forecast			24-month-ahead forecast		
	0 to 1 12	1 to 5 4	0 to 5 16	0 to 1 12	1 to 5 4	0 to 5 16	0 to 1 12	1 to 5 4	0 to 5 16	0 to 1 12	1 to 5 4	0 to 5 16
RMSE (basis points)												
CFRS	44.06	53.15	48.32	69.83	72.82	73.08	105.50	99.94	106.90	142.80	112.95	137.65
RW	46.97	51.88	50.18	74.52	77.19	78.12	120.09	113.11	121.80	185.72	149.51	179.70
DL	74.52	79.34	77.10	109.30	109.37	110.99	148.48	139.26	148.20	210.49	187.40	206.61
CA_A13	57.88	58.47	59.48	97.58	91.67	97.86	160.66	146.92	159.52	264.35	236.98	259.90
EA_A13	95.46	74.30	92.44	131.59	105.65	127.57	170.97	144.63	167.12	208.23	173.96	201.96
NW-stat												
CFRS vs RW	-2.361	0.958	-1.830	-1.633	-1.969	-2.084	-2.242	-3.014	-2.626	-3.624	-4.522	-3.914
CFRS vs DL	-7.304	-5.056	-6.687	-4.627	-3.688	-4.343	-2.917	-2.619	-2.850	-2.452	-3.708	-2.719
CFRS vs CA_A13	-3.273	-1.347	-2.800	-2.488	-1.883	-2.339	-2.595	-2.658	-2.681	-2.340	-2.613	-2.453
CFRS vs EA_A13	-9.217	-6.024	-9.096	-5.534	-4.592	-5.555	-3.640	-3.795	-3.873	-3.194	-4.279	-3.507

This table shows the 3-, 6-, 12- and 24-month-ahead yield forecasts produced by the CFRS model, the Random Walk (RW) model, the DL procedure, the completely affine $A_1(3)$ (CA_A13) model and the essentially affine $A_1(3)$ (EA_A13) model. We use the first 20 years of training data to generate the first forecast. The training window is then moved forward one month at a time to generate successive forecasts. We first calculate the cross-sectional RMSE of forecast error within each maturity bucket in any given month. We then calculate the time-series mean of the cross-sectional RMSE and report that mean in this table. The NW-stat (quoted as a z-score) is used to test the significance of the differences in RMSE between any 2 models. A negative value of the NW-stat indicates that the first model (the model mentioned before "vs") is performing better than the second model.

Our parameterization of these quantities also simplifies the conversion of forward rates into bond prices and yields. The stochastic dynamics can be conveniently expressed under the risk-neutral measure. This leads to straightforward pricing of interest rate derivatives.

This class of models is empirically feasible to implement and can be used to generate forecasts of future forward rate curves. The forecasts at 6-, 12- and 24-month-ahead horizons generated by our particular specification are significantly better than the benchmark RW model, EH model and Expectations Hypothesis model with Term-Premium. Forecasts of future yields also perform better than those in the current literature.

This particular arbitrage-free formulation of our model is by no means the only possible formulation of the concept of maturity- and date-specific deviations. Even though our simple class of models can generate good forecasts, other more sophisticated formulations may produce even better fits and forecasts.

The models we present for the partitioning of the forward curve into the three components is immediately and easily extended to forward curves for commodity prices, such as crude oil: there the maturity-specific components have similar and intuitive interpretations, while the date-specific deviations are affected by weather forecasts and output predictions affecting convenience yields. It is also possible to specify a nonlinear version of the model but that poses formidable problems in testing because the model is no longer affine. These topics are left for future research.

Appendix A:

A.1 Proof that the 2-Brownian motion AFU conforms to the HJM specification

Reiterating relation (25), we have

$$df(s-t; t) = f_t dt + f_{m(t)} dm(t) + f_{d_1(t)} dd_1(t) + f_{d_2(t)} dd_2(t) + f_{d_3(t)} dd_3(t) \quad (A1)$$

where all second-order terms are zero.

Since

$$dm(t) = -m(t)K_m dt + \gamma_{1,t} dB_{1,t}$$

$$dd_1(t) = -2d_1(t)K_m dt + \gamma_{1,t} dB_{1,t}$$

$$dd_2(t) = -d_2(t)K_2 dt + \gamma_{1,t} dB_{1,t}$$

$$dd_3(t) = -d_3(t) \frac{K_2}{2} dt + \gamma_{2,t} dB_{2,t}$$

Relation (A1), the SDE for the forward rate, can be rewritten as

$$df(s-t; t) = \left[-K_m(2C_1 + m(t))e^{-2K_m(s-t)} - (C_2K_2 + m(t)(K_2 - K_m))e^{-K_2(s-t)} \right] dt \\ + \left[2e^{-K_m(s-t)}\gamma_{1,t} \right] dB_{1,t} + \left[e^{-\frac{K_2}{2}(s-t)}\gamma_{2,t} \right] dB_{2,t}.$$

Denote the diffusion of the forward rate SDE as

$$\sigma(t, s) = \begin{bmatrix} 2e^{-K_m(s-t)}\gamma_{1,t} \\ e^{-\frac{K_2}{2}(s-t)}\gamma_{2,t} \end{bmatrix} \quad (\text{A2})$$

$$\int_t^s \sigma(t, v)dv = \begin{bmatrix} -\frac{2}{K_m}e^{-K_m(s-t)}\gamma_{1,t} + \frac{2}{K_m}\gamma_{1,t} \\ -\frac{2}{K_2}e^{-\frac{K_2}{2}(s-t)}\gamma_{2,t} + \frac{2}{K_2}\gamma_{2,t} \end{bmatrix} \quad (\text{A3})$$

From our earlier assumption that

$$\vec{\kappa}_t = \begin{bmatrix} \kappa_{1,t} \\ \kappa_{2,t} \end{bmatrix} = \begin{bmatrix} \frac{2}{K_m}\gamma_{1,t} \\ \frac{2}{K_2}\gamma_{2,t} \end{bmatrix} \quad (\text{A4})$$

we have

$$\sigma(t, s)^T \left(\int_t^s \sigma(t, v)dv - \vec{\kappa}_t \right) = -\frac{4}{K_m}\gamma_{1,t}^2 e^{-2K_m(s-t)} - \frac{2}{K_2}\gamma_{2,t}^2 e^{-K_2(s-t)}. \quad (\text{A5})$$

Since

$$\gamma_{1,t}^2 = (m(t) + 2C_1)\frac{K_m^2}{4}$$

and

$$\gamma_{2,t}^2 = \frac{(C_2K_2 + m(t)(K_2 - K_m))(K_2)}{2},$$

relation (A5) becomes

$$\sigma(t, s)^T \left(\int_t^s \sigma(t, v)dv - \vec{\kappa}_t \right) = -K_m(2C_1 + m(t))e^{-2K_m(s-t)} \\ - (C_2K_2 + m(t)(K_2 - K_m))e^{-K_2(s-t)}.$$

The last expression exactly equals the drift of $df(s-t; t)$, and thus satisfies the HJM condition specified in relation (8).

A.2 Proof that the CFRS model conforms to the HJM specification

From Itô's lemma, we have the following SDE for the CFRS model:

$$df(s-t; t) = f_t dt + f_{m_1(t)} dm_1(t) + f_{d_1(t)} dd_1(t) + f_{d_2(t)} dd_2(t) \\ + f_{m_2(t)} dm_2(t) + f_{d_3(t)} dd_3(t) + f_{d_4(t)} dd_4(t) + f_{d_5(t)} dd_5(t)$$

where all second-order terms are zero.

Since

$$\begin{aligned}
 dm_1(t) &= -m_1(t)K_m dt + \gamma_{1,t} dB_{1,t} \\
 dm_2(t) &= -m_2(t)K_m dt + \gamma_{3,t} dB_{3,t} \\
 dd_1(t) &= -2d_1(t)K_m dt + \gamma_{1,t} dB_{1,t} \\
 dd_2(t) &= -d_2(t)K_m dt + \gamma_{2,t} dB_{2,t} \\
 dd_3(t) &= -2d_3(t)K_m dt + \gamma_{3,t} dB_{3,t} \\
 dd_4(t) &= -4d_4(t)K_m dt + \gamma_{3,t} dB_{3,t} \\
 dd_5(t) &= -2d_5(t)K_m dt + \gamma_{4,t} dB_{4,t}
 \end{aligned}$$

the SDE of the forward rate can thus be rewritten as

$$\begin{aligned}
 df(s-t; t) &= -K_m(2C_1 + 2m_1(t) + m_2(t))e^{-2K_m(s-t)} dt \\
 &\quad - (4C_2 K_m + 3m_2(t)K_m)e^{-4K_m(s-t)} dt \\
 &\quad + [2e^{-K_m(s-t)}\gamma_{1,t}]dB_{1,t} + [e^{-K_m(s-t)}\gamma_{2,t}]dB_{2,t} \\
 &\quad + [2e^{-K_m(s-t)}\gamma_{3,t}]dB_{3,t} + [e^{-2K_m(s-t)}\gamma_{4,t}]dB_{4,t}.
 \end{aligned}$$

Denote the diffusion of the forward rate SDE as

$$\begin{aligned}
 \sigma(t, s) &= \begin{bmatrix} 2e^{-K_m(s-t)}\gamma_{1,t} \\ e^{-K_m(s-t)}\gamma_{2,t} \\ 2e^{-K_m(s-t)}\gamma_{3,t} \\ e^{-2K_m(s-t)}\gamma_{4,t} \end{bmatrix} \\
 \int_t^s \sigma(t, v)dv &= \begin{bmatrix} -\frac{2}{K_m}e^{-K_m(s-t)}\gamma_{1,t} + \frac{2}{K_m}\gamma_{1,t} \\ -\frac{1}{K_m}e^{-K_m(s-t)}\gamma_{2,t} + \frac{1}{K_m}\gamma_{2,t} \\ -\frac{2}{K_m}e^{-K_m(s-t)}\gamma_{3,t} + \frac{2}{K_m}\gamma_{3,t} \\ -\frac{1}{2K_m}e^{-2K_m(s-t)}\gamma_{4,t} + \frac{1}{2K_m}\gamma_{4,t} \end{bmatrix}
 \end{aligned}$$

From our earlier assumption that

$$\hat{\kappa}_t = \begin{bmatrix} \kappa_{1,t} \\ \kappa_{2,t} \\ \kappa_{3,t} \\ \kappa_{4,t} \end{bmatrix} = \begin{bmatrix} \frac{2}{K_m}\gamma_{1,t} \\ \frac{1}{K_m}\gamma_{2,t} \\ \frac{2}{K_m}\gamma_{3,t} \\ \frac{1}{2K_m}\gamma_{4,t} \end{bmatrix}$$

we have,

$$\begin{aligned}
 \sigma(t, s)^T \left(\int_t^s \sigma(t, v)dv - \hat{\kappa}_t \right) &= \left(-\frac{4}{K_m}\gamma_{1,t}^2 - \frac{1}{K_m}\gamma_{2,t}^2 - \frac{4}{K_m}\gamma_{3,t}^2 \right) e^{-2K_m(s-t)} \\
 &\quad + \left(-\frac{1}{2K_m}\gamma_{4,t}^2 \right) e^{-4K_m(s-t)}. \tag{A6}
 \end{aligned}$$

Since

$$\begin{aligned}\gamma_{1,t}^2 &= \left(m_1(t) + \frac{2}{3}C_1\right) \frac{K_m^2}{4} \\ \gamma_{2,t}^2 &= \left(m_1(t) + \frac{2}{3}C_1\right) K_m^2 \\ \gamma_{3,t}^2 &= \left(m_2(t) + \frac{2}{3}C_1\right) \frac{K_m^2}{4} \\ \gamma_{4,t}^2 &= (3m_2(t) + 4C_2) 2K_m^2\end{aligned}$$

relation (A6) becomes

$$\begin{aligned}\sigma(t, s)^T \left(\int_t^s \sigma(t, v) dv - \hat{\kappa}_t \right) &= -K_m(2C_1 + 2m_1(t) + m_2(t))e^{-2K_m(s-t)} \\ &\quad - (4C_2K_m + 3m_2(t)K_m)e^{-4K_m(s-t)}.\end{aligned}$$

Because the last expression exactly equals the drift term of $df(s-t; t)$, this model satisfies the HJM condition specified in relation (8).

A.3 A nonaffine parameterization that conforms to the HJM specification

In this section, we present a simple 1-Brownian motion nonaffine model for the forward rate that is consistent with HJM specifications. Extensions of this nonaffine parameterization to 2 Brownian motions as well as n Brownian motions can also be specified and are available from the corresponding author.

Parameterize the forward curve $f(s-t; t)$ as

$$\begin{aligned}f_t(s-t) &= U(s-t) + M(s-t; t) + D(s-t; t) \\ &= C_0 - C_1e^{-K_m(s-t)} - C_2e^{-2K_m(s-t)} \\ &\quad + m(t)(e^{-K_m(s-t)} - e^{-2K_m(s-t)}) \\ &\quad + qm(t)^p(e^{-K_m(s-t)} - e^{-2K_m(s-t)}) \\ &\quad + d(t)e^{-2K_m(s-t)}\end{aligned}$$

where $m(t)$ is the maturity-specific state variable and $d(t)$ is the date-specific state variable. The parameter p is positive and generates the nonlinear relationship between forward rates and the maturity-specific state variable, while q is a scaling parameter on the nonlinear component. An intuitive interpretation for such a model is that whenever a shock occurs to the maturity-specific deviations, different types of agents cause the dissipation of such a shock at different rates. For instance, suppose that there is a sudden surge in demand for loanable funds at the 5-year maturity. The forward rate at around that maturity should increase (reflected by an increase in the maturity-specific deviation around the 5-year forward rate). There may be a small number of agents who are able to adjust their borrowing/lending habits very quickly in response to that shock (by either borrowing at different maturities or by shifting their lending to that maturity). However, the majority of agents may take a longer time to respond to such shocks. The small set of fast-moving agents can be represented by the nonlinear term $qm(t)^p$ where q is smaller than 1 and p is larger than 1; the large set of slower-moving agents is represented by the usual $m(t)$ term.

The following mathematical derivation explicitly defines the restrictions on the dynamics of the model to ensure that the model conforms to the HJM specification.

Let the stochastic processes for the maturity- and date-specific state variables be

$$dm(t) = -m(t)K_m dt + \gamma_t dB_t$$

and

$$dd(t) = -d(t)(2K_m)dt + (1 + pqm(t)^{p-1})\gamma_t dB_t,$$

where B_t is a standard Brownian motion.

We then have

$$\begin{aligned} df(s-t; t) &= f_t dt + f_{m(t)} dm(t) + f_{d(t)} dd(t) + \frac{1}{2} f_{m(t)m(t)} dm(t) \cdot dm(t) \\ &= (-C_1 K_m e^{-K_m(s-t)} - C_2(2K_m) e^{-2K_m(s-t)}) dt \\ &\quad + (m(t)(K_m e^{-K_m(s-t)} - 2K_m e^{-2K_m(s-t)})) dt \\ &\quad + (qm(t)^p (K_m e^{-K_m(s-t)} - 2K_m e^{-2K_m(s-t)})) dt \\ &\quad + d(t)(2K_m) e^{-2K_m(s-t)} dt \\ &\quad + (e^{-K_m(s-t)} - e^{-2K_m(s-t)}) (-m(t)K_m dt + \gamma_t dB_t) \\ &\quad + pqm(t)^{p-1} (e^{-K_m(s-t)} - e^{-2K_m(s-t)}) (-m(t)K_m dt + \gamma_t dB_t) \\ &\quad + e^{-2K_m(s-t)} (-d(t)(2K_m) dt + (1 + pqm(t)^{p-1}) \gamma_t dB_t) \\ &\quad + \frac{p(p-1)}{2} qm(t)^{p-2} (e^{-K_m(s-t)} - e^{-2K_m(s-t)}) \gamma_t^2 dt \\ &= \left(-C_1 K_m + (1-p)qm(t)^p K_m + \frac{p(p-1)}{2} qm(t)^{p-2} \gamma_t^2 \right) e^{-K_m(s-t)} dt \\ &\quad \left(-2C_2 K_m - m(t)K_m + (p-2)qm(t)^p K_m - \frac{p(p-1)}{2} qm(t)^{p-2} \gamma_t^2 \right) \\ &\quad \times e^{-2K_m(s-t)} dt + \gamma_t (pqm(t)^{p-1} + 1) e^{-K_m(s-t)} dB_t. \end{aligned}$$

Denote

$$\sigma(t, s) = \gamma_t (pqm(t)^{p-1} + 1) e^{-K_m(s-t)}.$$

Then

$$\begin{aligned} \int_t^s \sigma(t, v) dv - \kappa_t &= -\frac{\gamma_t (pqm(t)^{p-1} + 1)}{K_m} e^{-K_m(s-t)} + \frac{\gamma_t (pqm(t)^{p-1} + 1)}{K_m} - \kappa_t \\ \sigma(t, s) \left(\int_t^s \sigma(t, v) dv - \kappa_t \right) &= -\frac{\gamma_t^2 (pqm(t)^{p-1} + 1)^2}{K_m} e^{-2K_m(s-t)} \\ &\quad + \gamma_t (pqm(t)^{p-1} + 1) \left(\frac{\gamma_t (pqm(t)^{p-1} + 1)}{K_m} - \kappa_t \right) e^{-K_m(s-t)}. \end{aligned}$$

To ensure that $\sigma(t, s) \left(\int_t^s \sigma(t, v) dv - \kappa_t \right)$ equals the drift of $df_t(s - t)$, we match coefficients. Matching coefficients for $e^{-2K_m(s-t)}$, we have

$$-\frac{\gamma_t^2 (pqm(t)^{p-1} + 1)^2}{K_m} = -2C_2 K_m - m(t) K_m + (p-2)qm(t)^p K_m - \frac{p(p-1)}{2} qm(t)^{p-2} \gamma_t^2.$$

Matching coefficients for $e^{-K_m(s-t)}$, we have

$$\begin{aligned} & \gamma_t (pqm(t)^{p-1} + 1) \left(\frac{\gamma_t (pqm(t)^{p-1} + 1)}{K_m} - \kappa_t \right) \\ &= -C_1 K_m + (1-p)qm(t)^p K_m + \frac{p(p-1)}{2} qm(t)^{p-2} \gamma_t^2. \end{aligned}$$

Solving these equations for the values of γ_t^2 and κ_t , we find

$$\gamma_t^2 = \frac{-2C_2 K_m - m(t) K_m + (p-2)qm(t)^p K_m}{\frac{p(p-1)}{2} qm(t)^{p-2} - \frac{(pqm(t)^{p-1} + 1)^2}{K_m}}$$

and

$$\kappa_t = \frac{\gamma_t (pqm(t)^{p-1} + 1)}{K_m} - \frac{-C_1 K_m + (1-p)qm(t)^p K_m + \frac{p(p-1)}{2} qm(t)^{p-2} \gamma_t^2}{\gamma_t (pqm(t)^{p-1} + 1)}.$$

With these specifications for γ_t and κ_t , the model conforms to HJM and is arbitrage free.

A.4 Standard equations of the Kalman filter

First we need the prediction equations:

$$x_{t|t-1} = W \hat{x}_{t-1|t-1} \tag{A7}$$

where $x_{t|t-1}$ is the time $t-1$ prediction of x_t and $\hat{x}_{t-1|t-1}$ is the time $t-1$ estimate of x_{t-1} and W is as defined in relation (42) in Section 3.4.

$$P_{t|t-1} = W \hat{P}_{t-1|t-1} W' + Q_t \tag{A8}$$

where $P_{t|t-1}$ is the time $t-1$ prediction of P_t and $\hat{P}_{t-1|t-1}$ is the time $t-1$ estimate of P_{t-1} (P is the covariance matrix of the state vector x).

Updating equations:

$$\hat{x}_{t|t} = x_{t|t-1} + P_{t|t-1} A' F_t^{-1} v_t \tag{A9}$$

$$\hat{P}_{t|t} = P_{t|t-1} - P_{t|t-1} A' F_t^{-1} A P_{t|t-1} \tag{A10}$$

where

$$v_t = z_t - A x_{t|t-1} \tag{A11}$$

are the prediction errors, A is as defined in Section 3.4 and

$$F_t = A P_{t|t-1} A' + \sigma^{*2} I \tag{A12}$$

is the conditional variance of the prediction errors

Reference

- Aït-Sahalia, Y. 1996a. Nonparametric Pricing of Interest Rate Derivative Securities. *Econometrica* 64:527–60.
- Aït-Sahalia, Y. 1996b. Testing Continuous-Time Models of the Spot Interest Rate. *Review of Financial Studies* 9:385–426.
- Akaike, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. *Second International Symposium on Information Theory*, Budapest 267–81.
- Babbs, S., and K. Nowman. 1999. Kalman Filtering of Generalized Vasicek Term Structure Models. *Journal of Financial and Quantitative Analysis* 34:115–30.
- Backus, D., S. Foresi, and S. Zin. 1998. Arbitrage Opportunities in Arbitrage-Free Models of Bond Pricing. *Journal of Business and Economic Statistics* 16:13–26.
- Black, F., E. Derman, and W. Toy. 1990. A One-Factor Model of Interest Rates and its Application to Treasury Bond Options. *Financial Analysts Journal* 46:33–9.
- Brace, A., D. Gatarek, and M. Musiela. 1997. The Market Model of Interest Rate Dynamics. *Mathematical Finance* 7:127–55.
- Chen, R., and L. Scott. 2002. Multi-Factor Cox-Ingersoll-Ross Models of the Term Structure: Estimates and Tests from a Kalman Filter Model. Working Paper, Rutgers University.
- Collin-Dufresne, P., and R. Goldstein. 2003. Generalizing the Affine Framework to HJM and Random Field Models. Working Paper, University of California, Berkeley.
- Cox, J., J. Ingersoll, and S. Ross. 1981. A Re-Examination of Traditional Hypotheses About the Term Structure of Interest Rates. *Journal of Finance* 36:769–99.
- Cox, J., J. Ingersoll, and S. Ross. 1985. A Theory of the Term Structure of Interest Rates. *Econometrica* 53:385–407.
- Dai, Q., and K. Singleton. 2000. Specification Analysis of Affine Term Structure Models. *Journal of Finance* 55:1943–78.
- DeJong, F., and P. Santa-Clara. 1999. The Dynamics of the Forward Interest Rate Curve: a Formulation with State Variables. *Journal of Financial and Quantitative Analysis* 31:131–57.
- Diebold, F., and C. Li. 2006. Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics* 130:337–64.
- Diebold, F., and R. Mariano. 1995. Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13:253–63.
- Duffee, G. 2002. Term Premia and Interest Rate Forecasts in Affine Models. *Journal of Finance* 57:405–43.
- Duffee, G., and R. Stanton. 2004. Estimation of Dynamic Term Structure Models. Working Paper, U.C. Berkeley.
- Duffie, D., and R. Kan. 1996. A Yield-Factor Model of Interest Rates. *Mathematical Finance* 6:179–406.
- Fisher, I. 1896. *Appreciation and Interest*, vol. 11: Publications of the American Economic Association, New York 21–9.
- Fisher, M., and C. Gilles. 1998. Around and Around: The Expectations Hypothesis. *Journal of Finance* 53:365–83.
- Gallant, A., and G. Tauchen. 1996. Which Moments to Match. *Econometric Theory* 12:657–81.
- Geyer, A., and S. Pichler. 1999. A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure. *Journal of Financial Research* 22:107–30.
- Goldstein, R. 2000. The Term Structure of Interest Rates as a Random Field. *Review of Financial Studies* 13:365–84.

- Hamilton, J. 1994. *Time Series Analysis*, Princeton: Princeton University Press.
- Heath, D., R. Jarrow, and A. Morton. 1992. Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica* 60:77–105.
- Ho, T., and S. Lee. 1986. Term Structure Movements and Pricing Interest Rate Contingent Claims. *Journal of Finance* 41:1011–29.
- Hull, J., and A. White. 1990. Pricing Interest-Rate Derivative Securities. *Review of Financial Studies* 3:573–92.
- Kennedy, D. 1994. The Term Structure of Interest Rates as a Gaussian Random Field. *Mathematical Finance* 4:247–58.
- Longstaff, F. 2000. Arbitrage and the Expectations Hypothesis. *Journal of Finance* 55:989–94.
- McCulloch, J. 1993. A Reexamination of Traditional Hypotheses About the Term Structure: A Comment. *Journal of Finance* 48:779–89.
- Miltersen, K., K. Sandmann, and D. Sondermann. 1997. Closed Form Solutions for Term Structure Derivatives with Log Normal Interest Rates. *Journal of Finance* 52:409–30.
- Modigliani, F., and R. Sutch. 1966. Innovations in Interest Rate Policy. *The American Economic Review* 56:178–97.
- Nelson, C., and A. Siegel. 1987. Parsimonious Modeling of Yield Curves. *Journal of Business* 60:473–89.
- Newey, W., and K. West. 1987. A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica* 55:703–8.
- Santa-Clara, P., and D. Sornette. 2001. The Dynamics of the Forward Interest Rate Curve with Stochastic String Shocks. *The Review of Financial Studies* 14:149–85.
- Vasicek, O. 1977. An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics* 5:177–88.