Journal of Financial Economics 98 (2010) 113–144

Contents lists available at ScienceDirect

Journal of Financial Economics

journal homepage: www.elsevier.com/locate/jfec

Estimating affine multifactor term structure models using closed-form likelihood expansions ☆

Yacine Aït-Sahalia [a,*], Robert L. Kimmel [b]

[a] Department of Economics, Princeton University and NBER, Princeton, NJ 08544-1021, USA
[b] EDHEC Business School, 393, Promenade des Anglais BP3116 06202 Nice cedex 3, France

ARTICLE INFO

Article history:
Received 10 December 2008
Received in revised form
15 May 2009
Accepted 17 July 2009
Available online 31 May 2010

JEL classification:
C51
C52
G12

Keywords:
Term structure
Multifactor
Interest rates
Affine
Closed-form maximum-likelihood

ABSTRACT

We develop and implement a technique for closed-form maximum likelihood estimation (MLE) of multifactor affine yield models. We derive closed-form approximations to likelihoods for nine Dai and Singleton (2000) affine models. Simulations show our technique very accurately approximates true (but infeasible) MLE. Using US Treasury data, we estimate nine affine yield models with different market price of risk specifications. MLE allows non-nested model comparison using likelihood ratio tests; the preferred model depends on the market price of risk. Estimation with simulated and real data suggests our technique is much closer to true MLE than Euler and quasi-maximum likelihood (QML) methods.

1. Introduction

We develop and implement in this paper a technique for closed-form maximum likelihood estimation of multivariate affine yield models of the term structure of interest rates, and for statistical comparison of nested and non-nested families of these models. Affine yield models are popular among both practitioners and academics, largely because they have desirable analytical properties and allow for straightforward pricing of bonds and other interest rate derivatives. Bond prices, in general, are solutions to a second-order linear partial differential equation. For most non-affine term structure models, solutions to this differential equation must be found through numeric methods, which become increasingly impractical as the number of factors underlying the model increases. However, for affine yield models, this partial differential equation decomposes into a system of Ricatti-type ordinary differential equations (see, for example, Duffie and Kan, 1996) which can be solved quickly, even with a large number of underlying factors.

Despite their desirable analytic properties, the estimation of affine yield models still poses many challenges. The likelihood function for a vector of yields in an affine yield model is known in closed-form only for a few special cases. Most studies of estimation of affine yield models outside this relatively restricted subclass have therefore focused either on numeric techniques or method of moments estimators. Each of these methods has its advantages and disadvantages, which we will briefly discuss in turn.

☆ Financial support from the NSF under Grants SES-0111140 and SES-0350772 is gratefully acknowledged. We would like to thank Robert Bliss and Greg Duffee for the dataset used in this paper.

* Corresponding author. Tel.: +1 609 258 4015.
E-mail address: yacine@princeton.edu (Y. Aït-Sahalia).

0304-405X/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jfineco.2010.05.004

Moments of affine diffusions can be found in closed-form,[1] so estimation of affine yield models through the generalized method of moments (GMM) is feasible. As an early example, Gibbons and Ramaswamy (1993) use this method to estimate the model of Cox, Ingersoll, and Ross (1985). Chacko and Viceira (2003) and Singleton (2001) consider estimation methods based on moments derived from the characteristic function of the transition density, which is known in closed-form for affine diffusions, even though the density itself is, in general, not known in closed-form.

Dai and Singleton (2000) estimate several affine yield models using the simulation-based efficient method of moments. At least in theory, efficiency can be achieved if the number of moment conditions goes to infinity with the number of data observations. This method is computationally intensive, requiring many simulations, and is also highly flexible, requiring the a priori selection of an auxiliary model and resulting moment conditions. Little is known about its behavior in repeated simulation trials, although its performance has been questioned in the context of dynamic term structure models (see Duffee and Stanton, 2001; Tauchen, 2004). In addition, most affine yield models have lower bounds on one or more state variables. Most implementations of the method of moments techniques calculate moments of bond yields directly, and never explicitly calculate the values of the state variables implied by the observed bond yields. The implied values of some of the state variables may lie on the wrong side of the boundary for some observations, in which case the estimated model implies that the observed data could not have occurred. Duffee (2002) notes that the parameters estimated by Dai and Singleton (2000) imply that many of the observations in the data set could not have occurred if the estimated model were in fact the true data-generating model.

One alternative to method of moments estimation is the use of Gaussian likelihood approximations, in which the conditional density of the state vector is assumed to have a multivariate Gaussian distribution. The mean vector and covariance matrix of the state vector can be assumed to be proportional to the length of time between observations (i.e., the properties of the process at a finite time horizon are approximated by an Euler discretization). Alternately, the true means, variances, and covariances of the state vector can be used; this latter approach in the literature on affine processes is usually referred to as quasi-maximum likelihood (QML); see, for example, Fisher and Gilles (1996a) and Duffee (2002). With either approach, if the number of observed yields is greater than the number of state variables in the model being estimated (as is necessary for full identification of all parameters for some affine yield models), it must be assumed that at least some of the yields are observed with error; see, for example, Piazzesi (2009). Estimation using Euler or QML approximation of the likelihood function has the advantages that it is feasible for all affine yield models, and avoids the problem of estimated models that imply the state vector took on unattainable values for some observations. Use of the Euler approximate likelihood produces inconsistent estimates; QML is consistent but inefficient, except in the cases where the true likelihood is Gaussian.

Instead of quasi-maximum likelihood, one can consider approximate maximum likelihood estimation with the likelihood function calculated numerically or through simulation techniques. The transition function can be found as a solution to the Kolmogorov forward equation; in general, this partial differential equation must be solved numerically. Maximum likelihood estimation can be implemented via simulations. Pedersen (1995) develops a technique for estimating the likelihood function of discrete observations of a diffusion process by simulations, which Brandt and Santa-Clara (2002) extend to multivariate diffusions. When applied to term structure models, likelihood methods usually assume, as do Chen and Scott (1993) and Duffee (2002), that an arbitrary set of benchmark yields are observed without error, with all remaining yields observed with some error; Brandt and He (2002) perform simulated maximum likelihood estimation of a model when all yields are observed with some error. However, because new simulations are required for each parameter vector considered during the likelihood search, the computing time required is substantial, as is generally the case with simulation-based methods. On the other hand, their method does not require that an arbitrary set of bond yields be assumed to be observed without error. Finally, Liu, Pan, and Pedersen (2001) propose to Fourier-invert numerically the known characteristic function of an affine diffusion to recover an approximation of its density. The transition density must be calculated for each data observation, and for each value of the parameter vector considered during a likelihood search, which quickly becomes computationally very intensive; see Jensen and Poulsen (2002) for a comparison of different methods which demonstrates this in the univariate case. Already numerically intensive in the univariate case, estimation by numerical methods becomes exponentially more difficult for multivariate diffusions, which typically involve large numbers of parameters.

As an alternative to the above techniques, we propose maximum likelihood estimation with the likelihood function approximated by a series of accurate expansions for the log-likelihood function (or equivalently the density) due to Aït-Sahalia (2008), which generalizes to arbitrary multivariate processes the univariate results developed in Aït-Sahalia (2002).[2] What we do in this

[1] For some models with multiple factors, "closed-form" must be interpreted to include finding the spectral or Jordan decomposition of a non-symmetric matrix. See Fisher and Gilles (1996b) for a detailed discussion of the calculation of first and second moments of affine diffusions.

[2] Aït-Sahalia (1999) contains examples of application of the univariate method to models in finance. Jensen and Poulsen (2002), Stramer and Yan (2005), and Hurn, Jeisman, and Lindsay (2007) conducted extensive comparisons of different techniques for approximating the transition function and demonstrated that the method described is both the most accurate and the fastest to implement for the types of problems and sampling frequencies one encounters in finance. The method has been extended to time inhomogeneous processes by Egorov, Li and Xu (2003) and to jump-diffusions by Schaumburg (2001) and Yu (2007). DiPietro (2001) has extended the methodology to make it applicable in a Bayesian setting. Bakshi and Ju (2005) propose an alternative centering in the univariate case. Li (2005) considers the case of "damped diffusion" processes.

paper is a natural but not quite straightforward application of these results, because unlike the standard maximum-likelihood theory developed there, multifactor term structure models typically rely on a state vector that is at least partly unobservable, or latent. The key aspect of the method is that, unlike the approaches described above, the resulting density expansion from this approach is in closed-form. Furthermore, because it is maximum-likelihood, it is efficient, achieving the Cramér-Rao lower bound for the asymptotic variance of the estimators.

The affine class of models has been studied extensively by Dai and Singleton (2000). They show the existence of $N+1$ distinct non-nested families of affine models with N state variables, M of which enter the diffusion matrix; we use the notation $A_M(N)$ with $0 \leq M \leq N$ to identify each family of models. Without parameter restrictions, the likelihood function is known in closed-form for only one of these $N+1$ families, corresponding to $M=0$. (The only exception is $N=1$; in this case, both single-factor affine models have closed-form likelihoods.) Duffie, Pedersen, and Singleton (2003) propose a decomposition of the likelihood function of an affine model, an approach that requires independence of the volatility variables and the simulation of the remaining part of the likelihood. The independence assumption is, in general, not satisfied when $M > 1$.

By contrast, we derive closed-form approximations to the likelihood functions for all $N+1$ families for all $N \leq 3$ (a total of nine models, four of which have known likelihood functions). No simulations are required in our approach, and we are not limited to any particular affine model, such as those with independent volatility variables as in Duffie, Pedersen, and Singleton (2003), nor for that matter to affine specifications, although they are our focus here. We show how maximum-likelihood estimation can be implemented using, for instance, bond yields as the observables. The likelihood expansions computed in this paper for affine term structure models are also being used by Mosburger and Schneider (2005), Cheridito, Filipović, and Kimmel (2007), Thompson (2008), and Egorov, Li, and Ng (2008), the latter two of which apply this method in the four-factor case. Expansions using the same method but for stochastic volatility models are used by Bakshi, Ju, and Ou-Yang (2006) and Aït-Sahalia and Kimmel (2007).

Maximum likelihood estimation allows us to use likelihood ratio tests to evaluate the fit of non-nested models (see Vuong, 1989), which is difficult or impossible for other methods, such as the method of moments. Although other studies, such as Dai and Singleton (2000) and Cheridito, Filipović, and Kimmel (2007), have evaluated non-nested models, these studies have used ad hoc model selection criteria, rather than rigorous statistical tests. Based on synthetic US Treasury strips data, we implement a likelihood ratio test for non-nested models to compare the fit of different models with the same number of state variables. We find that, with the simple market price of risk specification for affine models proposed by Dai and Singleton (2000), models with fewer state variables entering the diffusion matrix (i.e., $A_M(N)$ models with smaller values of M) are usually preferred to models with more state variables in the diffusion matrix (although the preference is not always significant at the conventional 95% confidence level). However, when the more flexible market price of risk specification used by Cheridito, Filipović, and Kimmel (2007) is used, this preference sometimes reverses; models with very non-Gaussian state variable processes are preferred to those with more Gaussian state processes. Although these authors have commented on the apparent reversal of preference when the market price of risk specification is extended, they do not perform any rigorous statistical tests; the tests they do perform are confined to comparison of different market price of risk within the same class of affine model (that is, with the same values of M and N). Our procedure allows rigorous statistical testing of such hypotheses involving non-nested models. We also consider nested likelihood ratio tests of the Dai-Singleton market price of risk relative to the Cheridito-Filipović-Kimmel specifications, and also of parameter restrictions needed for existence of closed-form likelihoods. We find to some extent that these two sets of parameters are substitutes; although each set is still statistically significant (at the 95% level) in almost every model considered, the degree of statistical significance of either set is usually much reduced when the other set is introduced. We also perform the likelihood ratio tests using Euler and QML likelihood approximations, and find that they are sometimes much less accurate than our likelihood approximation method.

The paper is organized as follows. We start with a brief review of affine term structure models in Section 2. Next, we describe our estimation technique in Section 3, including the construction of the closed-form likelihood expansions. We then examine in Section 4, through Monte Carlo simulations, the accuracy of this method by imposing necessary parameter restrictions so that closed-form likelihood functions are available, and compare estimates derived using the true likelihoods on simulated data to those derived using our approximations on the same data. We also compare these estimates to those obtained using Euler and QML approximations to the likelihood function. We find uniformly that the maximum-likelihood estimates produced by our method are very close to the estimates produced by the exact likelihood function; the Euler and QML approximations are much less accurate (except in the case of Gaussian models, for which the QML likelihood approximation is exact). In Section 5, we apply the estimation technique to US Treasury security data and discuss the results, and also perform nested and non-nested model selection tests using the approximate likelihoods. We compare these results to those obtained using Euler and QML approximations to the likelihood function, and, as with the simulated data, find our method approximates true maximum likelihood much more closely than the other two methods. Section 6 concludes. The explicit formulae we derive for the affine term structure models are contained in Appendix B for the bivariate models. They are available in computer form from the authors in dimensions 1, 2, and 3 upon request.

2. Affine term structure models

A multivariate term structure model typically specifies that the instantaneous riskless rate r_t is a deterministic function of an $N \times 1$ vector of state variables, X_t

$$r_t = r(X_t;\theta). \tag{1}$$

Under an equivalent martingale measure Q, the state vector follows the dynamics:

$$dX_t = \mu^Q(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dW_t^Q, \tag{2}$$

where X_t and $\mu^Q(X_t;\theta)$ are $N \times 1$ vectors, $\sigma(X_t;\theta)$ is an $N \times N$ matrix, θ is a vector containing the model parameters, and W_t^Q is an $N \times 1$ vector of independent Brownian motions.

In order to avoid arbitrage opportunities, the price at t of a zero-coupon bond maturing at T is given by the Feynman-Kac representation

$$P(x,t,T;\theta) = E^Q\left[\exp\left(-\int_t^T r_u\,du\right)\middle| X_t = x\right], \tag{3}$$

where the expectation is taken with respect to the risk-neutral dynamics of X specified in (2). It is also well-known that P satisfies the partial differential equation (PDE)

$$\frac{\partial P}{\partial t} + \mu^Q(x;\theta)'\frac{\partial P}{\partial x} + \frac{1}{2}\mathrm{Trace}\left[\nu(x;\theta)\frac{\partial^2 P}{\partial x \partial x'}\right] - r(x;\theta)P = 0, \tag{4}$$

where $\nu(x;\theta) \equiv \sigma(x;\theta)\sigma'(x;\theta)$ (with $'$ denoting transposition). The bond price must also satisfy the final condition $P(x,T,T;\theta) = 1$ for all x, T, and θ. Such a model is well-defined provided that (2) is well-defined, and the expected value (3) is finite, or, equivalently, the PDE (4) has a well-defined solution (subject only to technical regularity conditions).

Although there are several different ways to define affine yield term structure models, we use the following definition. An affine yield model is any model where the short rate (1) is an affine function of the state vector and the risk-neutral dynamics (2) are affine. We write the risk-neutral dynamics of the state vector as

$$dX_t = (\tilde{A} + \tilde{B}X_t)\,dt + \Sigma\sqrt{S(X_t;\alpha,\beta)}\,dW_t^Q, \tag{5}$$

where $\tilde{A}$ is an N-element column vector, $\tilde{B}$ and Σ are $N \times N$ matrices, and $S(X_t;\alpha,\beta)$ is the diagonal matrix with elements $S_{ii} = \alpha_i + X_t'\beta_i$, with each α_i a scalar and each β_i an $N \times 1$ vector, $1 \le i \le N$. The parameters of $\tilde{A}$, $\tilde{B}$, Σ, α, and β cannot be chosen arbitrarily; as discussed in Dai and Singleton (2000), there are admissibility restrictions required for the existence of the process X_t. These authors demonstrate the existence, for each value of N, of $N+1$ disjoint admissible regions of the parameter space. Let β denote the $N \times N$ matrix whose i-th column is the vector β_i, and let M denote the rank of β. M is the number of independent linear combinations of state variables entering the diffusion structure for the state variables.

With N factors, there are $N+1$ non-nested families of affine models corresponding to $M = 0,1,\ldots,N$; in each of these families different restrictions are imposed on the parameters of $\tilde{A}$, $\tilde{B}$, Σ, α, and β. Dai and Singleton (2000) also note that affine models do not have a unique representation, that is, there exist different choices of the model parameters that generate identical behavior of the interest rate and yield processes. They proceed to describe a canonical representation for each family of affine yield models, in which the Σ matrix is equal to the identity matrix; this choice does not result in a loss of generality, since, for a Σ matrix not equal to identity, we can construct a new set of state variables

$$Y_t = \Sigma^{-1}X_t. \tag{6}$$

The diffusion matrix of Y_t is then diagonal. We consider $N = 1$, 2, and 3, in which case there are two, three, and four, respectively, non-nested families of models, each with its own form of likelihood function.[3] We characterize the affine yield models with three or fewer state variables, as per Dai and Singleton (2000) (with our alternate parameterization for the drift), in Appendix A.

It can then be shown that, in affine models, bond prices have the exponential affine form

$$P(x,t,T;\theta) = \exp(-\gamma_0(\tau;\theta) - \gamma(\tau;\theta)'x), \tag{7}$$

where $\tau = T - t$ is the bond's time to maturity. That is, bond yields (non-annualized, and denoted by $g(x,t,T;\theta) = -\ln[P(x,t,T;\theta)]$) are affine functions of the state vector

$$g(x,t,T;\theta) = \gamma_0(\tau;\theta) + \gamma(\tau;\theta)'x. \tag{8}$$

Alternatively, one can start with the requirement that the yields be affine, and show that the dynamics of the state vector must be affine; see Duffie and Kan (1996). The same authors also show that, when the coefficients are affine, (4) is equivalent to a system of Ricatti-type ordinary differential equations, for which numeric solution is typically much faster than a general PDE with non-linear coefficients. This analytic tractability of bond prices accounts for much of the popularity of affine models.

3. Estimation procedure

Maximum-likelihood estimation requires evaluation of the likelihood of the observed panel of yield data for each parameter vector considered during a search procedure. The procedure for evaluating the likelihood of the observed yields at a particular value of the parameter vector consists of four steps. First, we extract the value of the state vector X_t (which is not directly observed) from a set of benchmark yields, which are assumed to be observed without error. Second, we evaluate the joint likelihood of the series of implied observations of the state vector X_t, using accurate closed-form approximations to the likelihood function. Third, we multiply this joint likelihood by a Jacobian determinant, to find the likelihood of the panel of observations of the benchmark yields. Finally, for a set of additional yields assumed to be

[3] Gouriéroux and Sufana (2004) identify some affine diffusions that do not fit into the classification scheme of Dai and Singleton (2000). We do not consider this class of diffusions. Cheridito, Filipović, and Kimmel (2010) show that the assumption of a diffusion matrix of the form shown in (5) also excludes some affine diffusions with four or more factors. However, in this paper, we consider only models with three or fewer factors, for which (5) is not restrictive.

observed with error (the additional yields are needed to identify some model parameters; see below), we calculate the likelihood of the observation errors, and multiply this likelihood by the likelihood of the benchmark yields found in the previous step, to find the joint likelihood of the panel of all yields. This procedure is repeated for different values of the parameter vector, until the maximum likelihood estimate is found.

Fig. 1 summarizes our estimation method: for each parameter vector, we can evaluate the likelihood of the observed bond yields using a combination of the affine pricing model and our closed-form likelihood expansions. As the figure shows, the only role the affine structure plays in our estimation method consists in allowing the transformation from observed yields to state variables (i.e., the pricing model) to be easily solvable.

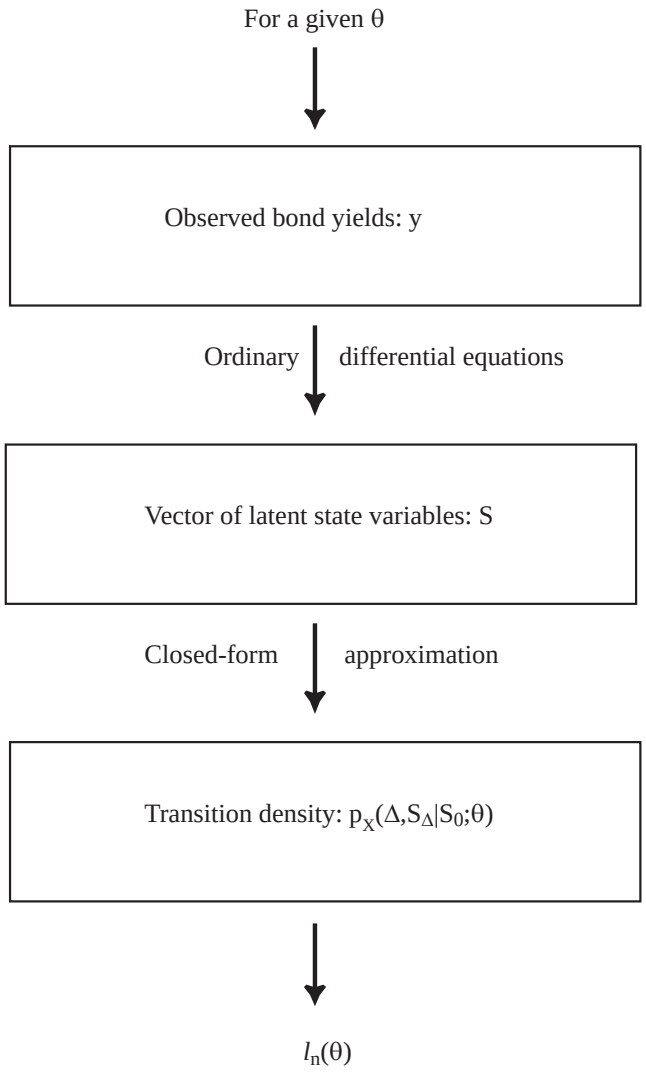

Fig. 1. The likelihood-based estimation method. This figure describes the steps involved in implementing the estimation method. The derivation of the closed-form expression for the log-likelihood is done once and for all for each model. Given that expression, maximum-likelihood proceeds as in the standard case with the additional step (requiring for affine models the solution of ordinary differential equations) necessary to transform the observed yields into the latent state variables of the model.

3.1. Extracting state from yields

We estimate the parameter vector θ based on a panel of observed bond prices or, equivalently, yields. The state vector X_t is not directly observable, and must be inferred from the cross-section of bond yields at date t. In affine yield models, zero-coupon yields are affine functions of the state vector, so the likelihood function of yields is readily determined from the likelihood function of the state vector.

It is well-known that, in general, affine models are not identified under both true and risk-neutral probability measures if the number of observed yields is equal to the number of state variables. Following Chen and Scott (1993) and Duffee (2002), we use $N+H$ observed yields, $H>0$, in the postulated model $A_M(N)$, and include observation errors for H of those yields. At each date t, the state vector X_t is then exactly identified by the yields observed without error, and these N yields jointly follow a Markov process. Denoting the times to maturity of the yields observed without error as $\tau_1,\ldots,\tau_N$, the observed values of these yields, on the left-hand side, are equated with the predicted values from (8) given the model parameters and the current values of the state variables, $X_{1t},\ldots,X_{Nt}$:

$$\begin{bmatrix} g(X_t,t,t+\tau_1;\theta) \\ \vdots \\ g(X_t,t,t+\tau_N;\theta) \end{bmatrix} = \begin{bmatrix} \gamma_0(\tau_1;\theta) \\ \vdots \\ \gamma_0(\tau_N;\theta) \end{bmatrix} + \begin{bmatrix} \gamma(\tau_1;\theta)' \\ \vdots \\ \gamma(\tau_N;\theta)' \end{bmatrix} \begin{bmatrix} X_{1t} \\ \vdots \\ X_{Nt} \end{bmatrix}. \quad (9)$$

By introducing a vector $\Gamma_0(\theta)$ whose elements are $\gamma_0(\tau_1;\theta)$ through $\gamma_0(\tau_N;\theta)$, and a matrix $\Gamma(\theta)$ whose columns are $\gamma(\tau_1;\theta)$ through $\gamma(\tau_N;\theta)$, this equation can be expressed in matrix form,

$$g_t = \Gamma_0(\theta) + \Gamma(\theta)' X_t. \quad (10)$$

The current value of the state vector X_t can be found by inverting this equation:

$$X_t = [\Gamma(\theta)']^{-1}[g_t - \Gamma_0(\theta)]. \quad (11)$$

3.2. Likelihood of the state vector

While the only parameters entering the transformation from observed yields to the state variables are the parameters of the risk-neutral (or Q-measure) dynamics of the state variables, once we have constructed our time series of values of X_t sampled at dates t_0, $t_1,\ldots,t_n$, the dynamics of the state variables that can be inferred from this time series are the dynamics under the physical measure (denoted by P). In the first step of the estimation procedure, we rely on the tractability of the affine bond pricing model, but in this step, we do not. In particular, we can now specify freely (that is, without regard for considerations of analytical tractability) the market prices of risk of the different Brownian motions, or equivalently the Radon-Nikodym derivative dQ/dP. The P-measure dynamics are

$$\begin{aligned} dX_t &= \mu^P(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dW_t^P \\ &= \{\mu^Q(X_t;\theta) + \sigma(X_t;\theta)\Lambda(X_t;\theta)\}\,dt + \sigma(X_t;\theta)\,dW_t^P. \end{aligned} \quad (12)$$

For example, Dai and Singleton (2000) use the simple market price of risk specification

$$\Lambda(X_t;\theta)=\sigma(X_t;\theta)'\lambda, \tag{13}$$

with λ an $N\times 1$ vector of constant parameters, so that under P, the instantaneous drift of each state variable is its drift under the risk-neutral measure, plus a constant times its volatility squared. Under this specification, the drift of the state vector is then affine under both the physical and risk-neutral measures, since

$$\mu^P(X_t;\theta)=(\tilde{A}+\tilde{B}X_t)+\Sigma S(X_t;\beta)\Sigma'\lambda\equiv A+BX_t.$$

We consider a canonical representation of affine models where A, B, α, and β have the normalized form

$$A=\begin{bmatrix}A_{M\times 1}\\ 0_{(N-M)\times 1}\end{bmatrix},\quad B=\begin{bmatrix}B_{M\times M} & 0_{M\times(N-M)}\\ B_{(N-M)\times M} & B_{(N-M)\times(N-M)}\end{bmatrix},$$

$$\alpha=\begin{bmatrix}0_{M\times 1}\\ 1_{(N-M)\times 1}\end{bmatrix},\quad \beta=\begin{bmatrix}I_{M\times M} & \beta_{M\times(N-M)}\\ 0_{(N-M)\times M} & 0_{(N-M)\times(N-M)}\end{bmatrix}. \tag{14}$$

This normalization is similar, but not identical, to that of Dai and Singleton (2000), who, as discussed earlier, write the drift of the state variables as $B(A-X_t)$, and who also normalize the last $(N-M)$ elements of θ to be zero. In our normalization, the last $(N-M)$ state variables (when $M<N$) differ by a constant from their counterparts under the Dai and Singleton (2000) normalization.

Given the identification normalization in which Σ is taken to be the identity matrix, existence of the process and non-attainment of the boundaries require that the parameters governing the state variable dynamics be constrained as follows:

$$(A_{M\times 1})_i\geq\tfrac{1}{2},\quad 1\leq i\leq M, \tag{15}$$

$$(B_{M\times M})_{ij}\geq 0,\quad 1\leq i,\ j\leq M,\ i\neq j, \tag{16}$$

$$(\beta_{M\times(N-M)})_{ij}\geq 0,\quad 1\leq i\leq M,\ M+1\leq j\leq N. \tag{17}$$

We further discuss these restrictions, and additional restrictions for stationarity, in Appendix A. Note that, in an $A_M(N)$ model with $M>0$, there are restrictions on the values of the state variables, but (at least for the more complicated models), there is no straightforward way to restrict a search to those parameter values which imply that the panel of state variables extracted from the observed yields satisfies these restrictions. We deal with this problem simply by setting the likelihood equal to zero whenever a state variable violates the model restrictions.

As mentioned above, an affine μ^P is not required for our likelihood expansions. See Duarte (2004) for a model that is affine under the risk-neutral probability measure, but not under the true probability measure. However, we do rely on the affine character of the dynamics under Q, because those allow us to go from state to yields in the tractable manner given by (9).

3.3. Likelihood of yields observed without error

Since the relationship between the state vector and bond yields is affine, as given by (10), the transition function of the bond yields can be derived from the transition function of the state vector by a change of variables and multiplication by a Jacobian determinant, which is a constant (i.e., not dependent on the state vector) in this case. Specifically, consider the stochastic differential equation describing the dynamics of the state vector X_t under the measure P, as specified by (12). Let $p_X(\Delta,x|x_0;\theta)$ denote its transition function, that is the conditional density of $X_{t+\Delta}=x$ given $X_t=x_0$. Let $p_G(\Delta,g|g_0;\theta)$ similarly denote the transition function of the vector of yields observed Δ time units apart. Since $x=\Gamma'^{-1}(\theta)(g-\Gamma_0(\theta))$, we have

$$p_G(\Delta,g|g_0;\theta)=\det|\Gamma'^{-1}(\theta)|p_X(\Delta,\Gamma'^{-1}(\theta)(g-\Gamma_0(\theta))|\Gamma'^{-1}(\theta)\times(g_0-\Gamma_0(\theta));\theta). \tag{18}$$

Then, recognizing that the vector of yields is Markovian and applying Bayes' Rule, the log-likelihood function for discrete data on the yield vector g_t sampled at dates $t_0,t_1,\ldots,t_n$ has the simple form

$$\ell_n(\theta)\equiv n^{-1}\sum_{i=1}^{n} l_G(t_i-t_{i-1},g_{t_i}|g_{t_{i-1}};\theta), \tag{19}$$

where $l_G\equiv\ln p_G$. As usual in likelihood estimation, we discard the unconditional distribution of the first observation since it is asymptotically irrelevant.

We assume in this paper that the sampling process is deterministic (see Aït-Sahalia and Mykland, 2003 for a treatment of maximum likelihood estimation in the case of randomly spaced sampling times). In typical practical situations for term structure models, and hence in our Monte Carlo experiments below, these types of models are estimated on the basis of weekly or monthly data, so that $t_i-t_{i-1}=\Delta=\frac{7}{365}$ or $t_i-t_{i-1}=\Delta=\frac{1}{12}$ is a fixed number.

3.4. Likelihood of all yields

From the coefficients $\gamma_0(\tau;\theta)$ and $\gamma(\tau;\theta)$ and the value of the state vector X_t found in the first step, we can calculate the implied values of the yields which are assumed to be observed with error, whose maturities are denoted by $\tau_{N+1},\ldots,\tau_{N+H}$

$$\begin{bmatrix}g(X_t,t,t+\tau_{N+1};\theta)\\ \vdots\\ g(X_t,t,t+\tau_{N+H};\theta)\end{bmatrix}=\begin{bmatrix}\gamma_0(\tau_{N+1};\theta)\\ \vdots\\ \gamma_0(\tau_{N+H};\theta)\end{bmatrix}+\begin{bmatrix}\gamma(\tau_{N+1};\theta)'\\ \vdots\\ \gamma(\tau_{N+H};\theta)'\end{bmatrix}\begin{bmatrix}X_{1t}\\ \vdots\\ X_{Nt}\end{bmatrix}. \tag{20}$$

The observation errors, denoted by $\varepsilon(t,t+\tau_{N+i})$, are the difference between these implied yields and the yields from the data. By assumption, these errors are Gaussian with zero mean and constant variance, and are independent across time and maturity (and also independent of the state variable processes). The joint likelihood of the observation errors can be calculated from the Gaussian density function, where σ_i is the standard deviation of the observation error for the yield with maturity τ_{N+i}

$$\sigma_i=\sqrt{Var[\varepsilon(t,t+\tau_{N+i})]}. \tag{21}$$

Since the observation errors are independent of the state variable process, the joint likelihood of the panel of all

observed yields is simply the product of the likelihood of the yields observed without error, multiplied by the likelihood of the observation errors. Equivalently, the two log-likelihoods can simply be added to find the joint log-likelihood of the panel of all yields.

3.5. Closed-form likelihood expansions

We now describe how we obtain closed-form approximations to l_G, hence to the log-likelihood function of the discretely sampled vector of yields in light of (19). To construct an expansion for l_G, we first construct an expansion for $l_X \equiv \ln p_X$ and then take logs on both sides of (18) to recover the corresponding expansion for l_G. So we can reduce the problem to one of approximating l_X, and for that we use the closed-form method of Aït-Sahalia (2008), which extends to multivariate diffusions the univariate results of Aït-Sahalia (1999, 2002).

The expansion of the log-likelihood has the form of a Taylor series in Δ at order K

$$l_X^{(K)}(\Delta,x|x_0;\theta) = -\frac{m}{2}\ln(2\pi\Delta) - D_v(x;\theta) + \frac{C_X^{(-1)}(x|x_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(k)}(x|x_0;\theta)\frac{\Delta^k}{k!}, \tag{22}$$

with

$$D_v(x;\theta) \equiv \tfrac{1}{2}\ln(Det[v(x;\theta)]), \tag{23}$$

for $v(x;\theta) \equiv \sigma(x;\theta)\sigma^T(x;\theta)$.

As defined in Aït-Sahalia (2008), a diffusion X is *reducible* if and only if there exists a one-to-one transformation of the diffusion X into a diffusion Y whose diffusion matrix σ_Y is the identity matrix. That is, there exists an invertible function $\gamma(x;\theta)$ such that $Y_t \equiv \gamma(X_t;\theta)$ satisfies the stochastic differential equation

$$dY_t = \mu_Y(Y_t;\theta)\,dt + dW_t^P. \tag{24}$$

Every univariate diffusion is reducible. However, this is not the case for every multivariate diffusion, and Aït-Sahalia (2008) gives necessary and sufficient conditions for reducibility. In the reducible case, the coefficients $C_X^{(k)}$ for $k=-1,0,\ldots,K$ are obtained in closed-form.

In the irreducible case, the approach is to calculate in closed-form a Taylor series in $(x-x_0)$ of each coefficient $C_X^{(k)}$, at order j_k in $(x-x_0)$. Such an expansion is denoted by $C_X^{(j_k,k)}$ and it turns out that the order should be $j_k=2(K-k)$ for $k=-1,0,\ldots,K$. The resulting expansion will then be

$$\tilde{l}_X^{(K)}(\Delta,x|x_0;\theta) = -\frac{m}{2}\ln(2\pi\Delta) - D_v(x;\theta) + \frac{C_X^{(j_{-1},-1)}(x|x_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(j_k,k)}(x|x_0;\theta)\frac{\Delta^k}{k!}. \tag{25}$$

Such a further Taylor expansion in $x-x_0$ is unnecessary in the reducible case; however, even for an irreducible diffusion, it is still possible to compute the coefficients $C_X^{(j_k,k)}$ in closed-form. For details, we refer to Aït-Sahalia (2008). The specific expressions for the models under consideration in dimension 2 are reproduced in Appendix B; they are available in computer form from the authors for dimensions 1, 2, and 3 upon request.

3.6. Alternate methods

It is possible to substitute other approximations for the closed-form likelihood expressions described in Section 3.5. One such method is what we call the Euler method, in which the conditional likelihood of the vector of state variables is approximated by a Gaussian density, with mean and variance equal to its instantaneous mean and variance, multiplied by the time between observations. Although an accurate approximation when consecutive observations are very close together in time, the Euler approximate likelihood is inaccurate for longer time periods, because, for any $A_M(N)$ with $m>0$, the true transition density is not Gaussian, and the mean and variance of the state vector is calculated as if its drift and diffusion were constant over the entire time between observations, when in fact, they are state dependent. Nonetheless, the Euler approximate likelihood is very simple to calculate, and may be a reasonable alternative to our method, despite its potential inaccuracy. In Section 4, we therefore examine the performance of this method, alongside our method, to judge whether the inaccuracy from use of the Euler method is sizable or not.

Another alternative method, often used in the literature, is usually referred to as quasi-maximum likelihood (QML); see Fisher and Gilles (1996a). (These authors do not use the term quasi-maximum likelihood or its acronym.) With this method, the transition density is approximated (as with the Euler method) by a Gaussian density. However, the mean and variance of this Gaussian transition density are the true mean and variance, rather than Euler approximations. The only source of inaccuracy is therefore the assumption of a Gaussian density, although in the case of $A_0(N)$ models, the true transition density is Gaussian, and consequently coincides with the QML likelihood. We calculate the conditional means and variances numerically, as solutions to the Chapman–Kolmogorov backward equation; with affine coefficients, this partial differential equation decomposes into a system of ordinary differential equations, the solution to which can be found by fast numeric methods. It is also possible to find these moments in closed-form, provided the definition of "closed-form" includes calculation of the Jordan decomposition of the slope matrix in the drift of the state vector.[4] However, the expressions for the conditional means and variances are subject to numeric instability in some cases; they frequently contain expressions of the form $[f(b)-f(a)]/(b-a)$. If the parameters of the model are such that a and b are very close together, calculation of this expression is problematic, due to the finite numeric precision of computers.[5] One potential solution is to use a Taylor approximation to $f(b)$ (around a) when a and b are close together; we instead avoid the

[4] The spectral decomposition is a special case of the Jordan decomposition. However, since there is no requirement that the slope matrix in the drift of the state vector be symmetric, there is no guarantee that it has a spectral decomposition. The Jordan decomposition coincides with the spectral decomposition, when the latter exists. See Fisher and Gilles (1996b) for a derivation of the first and second conditional moments of affine processes.

[5] The same problem arises when the true likelihood function is used, for those cases in which it is known explicitly.

problem entirely by using numerically calculated moments, as discussed above. Since the QML method uses more accurate moments than Euler, but still retains the simplicity of the Gaussian approximation, we examine in Section 4 the performance of the QML method (with moments calculated numerically), alongside our method and the Euler method, to judge whether the inaccuracy from use of this method is sizable or not.

4. Monte Carlo results

To assess the accuracy of our technique, we now consider models for which the likelihood function is known in closed-form, and compare parameter estimates using our technique to those obtained using the true likelihood functions. We also compare estimates by alternate approximation methods, Euler and QML, to the true MLE. We calculate the conditional means and variances for QML numerically; estimates calculated with exact expressions for the conditional moments (not presented—see the discussion in Section 3.6) produce nearly identical results. In all models considered but one, we find our parameter estimates are very close to the true maximum likelihood estimates for simulated data at the weekly frequency; in the one case for which the approximate estimates are not close to the true MLE, the deviation can be traced to an identification issue, and we find that, when the model is rewritten in terms of observed yields, the results using our technique are extremely close to the true MLE results. The deviations between the Euler and QML methods and the true MLE are substantially larger (except for QML estimates of the $A_0(N)$ models, since QML then coincides with the true likelihood). Since our estimation approach is based on Taylor expansions in the sampling interval Δ, observations at the daily frequency would result in even greater accuracy.

4.1. Procedure

The full parameter vector θ consists of all the elements of $(A,B,\alpha,\beta,\delta_0,\delta,\lambda,\sigma)$; with our normalization, the α vector contains no free parameters. We consider all nine models corresponding to $N=1$, 2, and 3 and estimate each $A_M(N)$ model using n time series observations of $N+H$ zero-coupon bond yields.

The individual models themselves are described in Appendix A, and parameter restrictions needed for existence, boundary non-attainability, and stationarity are presented in Table 1. For the purpose of studying the accuracy of our likelihood expansion approach, we also consider further parameter restrictions whenever necessary to obtain a closed-form likelihood to which we can then compare our expansion. These further parameter restrictions are presented in Table 2. Note again that our method does *not* require these further restrictions, and remains closed-form in all cases. The only reason we impose them is to have an exact likelihood to which our expansion can be compared in these Monte Carlo simulations.

For each canonical model, we simulate 1,000 data series of 501 weekly observations ($\Delta = 7/365$) of the vector of N state variables, giving $n=500$ pairs of discrete transitions of that process. Each of the simulated sample paths is produced by an Euler discretization of the process, using 30 intervals per week. Twenty-nine out of every 30 observations are discarded, leaving only the observations at the weekly frequency. Each simulated data series is initialized based on the unconditional distribution of the yields.

From this time series of the canonical state variables, we calculate the instantaneous interest rate and yields of varying maturities. Including the instantaneous interest rate (which can be interpreted as a yield with maturity of zero), we generate twice as many yields as state variables in each model (i.e., $H=N$), with maturities spaced evenly every two years. For example, for the three-factor models, we calculate the instantaneous interest rate and yields with maturities of two, four, six, eight, and 10 years. As discussed, more yields than state variables are needed to ensure identification of all model parameters, including the market price of risk. For an N-factor model, the N longest maturity yields include observation errors, which are assumed to be Gaussian with mean zero and constant variance, and further assumed to be uncorrelated across maturities and over time. Since the parameter restrictions which ensure existence of closed-form likelihoods also ensure existence of closed-form yield expressions, we calculate yields exactly rather than through numeric solution of the pricing partial differential equation.

From the time series observations of $2 \times N$ yields for each N-factor model, we then proceed to estimate the model parameters, using the true likelihoods, the 2-term reducible approximate likelihoods, and also the Euler and QML approximate likelihood functions. For some models, numeric instability was encountered in evaluating the true MLE; this occurred for three of the 1,000 simulations for the $A_1(2)$ model, seven of the 1,000 simulations for the $A_2(3)$ model, and 82 of the 1,000 simulations for the $A_3(3)$ model. In these cases, the true MLE was calculated with considerable error, and was very far from estimates produced by all the approximation methods; these particular simulations therefore do not allow meaningful comparison of the error introduced by different approximation techniques, and were excluded in the presentation of the results.[6]

Because there are nine models, some of them containing many parameters, we present results only for the three-factor models; these results for simulations with weekly observations are presented in Tables 3–6, which also present the data-generating values of the parameters.[7] The results for the one-factor and two-factor

[6] The problem is with the implementation of the modified Bessel function of the first kind in Matlab, which has difficulty with certain combinations of inputs. Specifically, in some cases, the function returns an error code indicating a loss of accuracy at the estimated parameter vector. In other cases, it returns zero, instead of a very small positive number, for some parameter vectors, so the search algorithm avoids these values.

[7] The data-generating values are based on estimates from Cheridito, Filipović, and Kimmel (2007).

Table 1
Parameter restrictions for the models.

This table shows the parameter restrictions imposed on the different models under consideration, which are sufficient to ensure existence of the state variable process, non-attainment of the boundaries (if any), and stationarity. "Eigen" denotes the eigenvalues of the indicated matrix. In some cases (e.g., the $A_2(2)$ model), the eigenvalues must be real due to restrictions for existence. The restriction in the table refers to the real part anyway, since eigenvalues may be complex conjugate pairs for some other models. Note that, although the data-generating values used in simulations satisfy the stationarity constraints, these constraints were not imposed during the estimation procedure. Finally, to be able to apply the closed-form likelihood expansion to square-root variables, we need to strengthen the condition $a_i \geq \frac{1}{2}$ to $a_i \geq \frac{3}{4}$. After reduction to unit diffusion of a square-root variable $dX_t = \beta(\theta - X_t)\,dt + \sigma X_t^{1/2}\,dW_t$, the process becomes $dY_t = \mu_Y(Y_t)\,dt + dW_t$, where $\mu_Y(y) = \kappa/y - \beta y/2$. So $\mu_Y(y)$ diverges near 0 like κ/y, with $\kappa = 2\beta\theta/\sigma^2 - 1/2$. But in the limiting case represented by square-root variables, to make zero an entrance boundary requires $\kappa \geq \frac{1}{2}$; for technical reasons, namely the ability to apply Girsanov's Theorem, we require slightly more for the likelihood expansion for this model, specifically $\kappa \geq 1$: see Assumption 3.1 in Aït-Sahalia (2002, p. 228). That translates here into $a_i \geq \frac{3}{4}$. A divergence faster than κ/y (such as $\mu_Y(y) \sim \kappa/y^\gamma$, $\gamma > 1$) generates exponential convergence of the transition density, so this is only an issue in the limiting case represented by square-root processes.

Model	Existence	Boundary	Stationarity
$\mathbf{A}_0(1)$	–	–	$b_{11} < 0$
$\mathbf{A}_1(1)$	$a_1 \geq 0,\ \delta_1 \geq 0$	$a_1 \geq \frac{1}{2}$	$b_{11} < 0$
$\mathbf{A}_0(2)$	–	–	$b_{11} < 0,\ b_{22} < 0$
$\mathbf{A}_1(2)$	$a_1 \geq 0,\ \delta_1 \geq 0,\ \beta_{21} \geq 0$	$a_1 \geq \frac{1}{2}$	$b_{11} < 0,\ b_{22} < 0$
$\mathbf{A}_2(2)$	$a_1 \geq 0, a_2 \geq 0$ $\delta_1 \geq 0, \delta_2 \geq 0$ $b_{12} \geq 0, b_{21} \geq 0$	$a_1 \geq \frac{1}{2},\ a_2 \geq \frac{1}{2}$	$\mathrm{Re}\left[\mathrm{Eigen}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\right] < 0$
$\mathbf{A}_0(3)$	–	–	$b_{11} < 0,\ b_{22} < 0,\ b_{33} < 0$
$\mathbf{A}_1(3)$	$a_1 \geq 0, \delta_1 \geq 0$ $\beta_{21} \geq 0, \beta_{31} \geq 0$	$a_1 \geq \frac{1}{2}$	$b_{11} < 0$ $\mathrm{Re}\left[\mathrm{Eigen}\begin{bmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{bmatrix}\right] < 0$
$\mathbf{A}_2(3)$	$a_1 \geq 0, a_2 \geq 0$ $\delta_1 \geq 0, \delta_2 \geq 0$ $b_{12} \geq 0, b_{21} \geq 0$ $\beta_{31} \geq 0, \beta_{32} \geq 0$	$a_1 \geq \frac{1}{2},\ a_2 \geq \frac{1}{2}$	$\mathrm{Re}\left[\mathrm{Eigen}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}\right] < 0$ $b_{33} < 0$
$\mathbf{A}_3(3)$	$a_1 \geq 0, a_2 \geq 0, a_3 \geq 0$ $\delta_1 \geq 0, \delta_2 \geq 0, \delta_3 \geq 0$ $b_{12} \geq 0, b_{13} \geq 0, b_{21} \geq 0$ $b_{23} \geq 0, b_{31} \geq 0, b_{32} \geq 0$	$a_1 \geq \frac{1}{2},\ a_2 \geq \frac{1}{2},\ a_3 \geq \frac{1}{2}$	$\mathrm{Re}\left[\mathrm{Eigen}\begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix}\right] < 0$

Table 2
Parameter restrictions for Monte Carlo simulations of canonical affine processes.

This table shows the additional parameter restrictions (relative to those in Table 1) needed to ensure existence of a known exact likelihood function. The method does *not* require these restrictions, and remains closed-form in all cases. The only reason we impose them here is to have an exact likelihood so that the method's performance can be assessed in these Monte Carlo simulations. These restrictions are not imposed in the empirical work that follows. The purpose of imposing these restrictions is to allow us to test the accuracy of our expansion in Monte Carlo simulations by comparing it to the exact, closed-form, likelihood function. The restriction b_{23}=0 for the $A_1(3)$ model is not actually needed for existence of a closed-from likelihood function, but is inherent to that model: it is required for identification of that affine model if the other restrictions are imposed.

Model	Restrictions for exact density
$A_0(1)$	none
$A_1(1)$	none
$A_0(2)$	none
$A_1(2)$	$b_{21} = \beta_{21} = 0$
$A_2(2)$	$b_{12} = b_{21} = 0$
$A_0(3)$	none
$A_1(3)$	$b_{21} = b_{23} = b_{31} = \beta_{21} = \beta_{31} = 0$
$A_2(3)$	$b_{12} = b_{21} = b_{31} = b_{32} = \beta_{31} = \beta_{32} = 0$
$A_3(3)$	$b_{12} = b_{21} = b_{13} = b_{31} = b_{23} = b_{32} = 0$

models (not presented) are qualitatively similar. We present the results in a common format that allows for the comparison of the sampling noise error in the parameter estimates, $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$, to the error due to the approximation of the true likelihood by our approach, $\hat{\theta}^{(2)} - \hat{\theta}^{(MLE)}$, as well as the corresponding error for the approximation of the likelihood by the Euler and QML methods, $\hat{\theta}^{(Euler)} - \hat{\theta}^{(MLE)}$ and $\hat{\theta}^{(QML)} - \hat{\theta}^{(MLE)}$, respectively. The notation $\hat{\theta}^{(2)}$ indicates that we use an expansion at order 2 in Δ (i.e., K=2 in Section 3.5) to obtain the approximate likelihood estimator. The bias (i.e., the average deviation from $\theta^{(TRUE)}$) and the standard deviation for $\hat{\theta}^{(MLE)}$ are reported, as well as the mean and standard deviation of the differences between $\hat{\theta}^{(2)}$, $\hat{\theta}^{(Euler)}$, and $\hat{\theta}^{(QML)}$, respectively, from $\hat{\theta}^{(MLE)}$. The biases and standard deviations of the differences between the approximate MLE, $\hat{\theta}^{(2)}$, and the true MLE, $\hat{\theta}^{(MLE)}$, are reported as percentages of the corresponding biases and standard deviations of the difference between the true MLE $\hat{\theta}^{(MLE)}$ and the data-generating values $\theta^{(TRUE)}$. The biases and standard

Table 3
Monte Carlo simulations for the $A_0(3)$ model—structural parameters, weekly observations.

This table reports the results of 1,000 Monte Carlo simulations for the $A_0(3)$ model, with weekly observations, comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(MLE)}$ around the true value of the parameters $\theta^{(TRUE)}$, to the distribution of the differences between the reducible and Euler estimates ($\hat{\theta}^{(2)}$ and $\hat{\theta}^{(Euler)}$, respectively) and the exact MLE $\hat{\theta}^{(MLE)}$. The reducible likelihoods are based on the expansion with K=2 terms. QML estimates are not presented, since for this model, they coincide with the exact MLE. The means and standard deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$ are reported as percentages of the corresponding deviations of $\hat{\theta}^{(MLE)}$ from $\theta^{(TRUE)}$. The means and standard deviations of $\hat{\theta}^{(Euler)}$ from $\hat{\theta}^{(MLE)}$ are reported as multiples of the corresponding deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$. The results appear to show that the difference $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ is smaller, but still a significant fraction of, the difference $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$ due to the sampling noise, and that, for many parameters, the Euler likelihoods perform nearly as well as the approximate MLE. However, this conclusion is unwarranted, since there is an identification problem with this model, as discussed in Section 4.2. Table 7 shows that, when the model is written in terms of observed bond yields instead of latent state variables, the approximate MLE strongly outperforms the Euler likelihood.

Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$		$\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ (perc. of $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$)		$\hat{\theta}^{(Euler)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
b_{11}	−0.5000	−0.19	0.57	3.31%	25.83%	0.91	0.99
b_{21}	1.5000	−0.19	0.52	2.49%	20.29%	0.99	0.99
b_{22}	−1.5000	0.15	0.65	4.97%	22.26%	2.05	1.00
b_{31}	0.1000	0.10	0.35	−1.47%	20.99%	3.02	0.99
b_{32}	−0.2000	−0.25	0.52	−0.56%	18.56%	4.86	1.00
b_{33}	−0.0500	−0.11	0.34	0.80%	7.95%	0.51	1.01
d_0^M	0.0022	−3.4E−04	3.6E−03	0.03%	0.01%	23.80	198.17
d_1	0.0100	−4.3E−03	7.1E−03	2.54%	25.40%	−0.78	1.00
d_2	0.0200	−3.3E−03	8.2E−03	0.13%	15.45%	60.58	1.01
d_3	0.0100	1.9E−03	4.9E−03	1.27%	18.73%	−0.24	1.00
λ_1	−0.0500	0.053	0.35	7.10%	14.54%	1.26	1.01
λ_2	−0.5000	0.034	0.37	−0.59%	11.84%	29.43	1.00
λ_3	−0.1000	−0.090	0.37	1.10%	4.35%	1.85	1.23
σ_1	0.0100	−1.6E−05	3.3E−04	−0.01%	0.02%	−6.98	34.24
σ_2	0.0100	−6.7E−06	3.2E−04	0.04%	0.01%	−25.18	44.85
σ_3	0.0100	−3.9E−05	3.2E−04	0.00%	0.02%	2064.35	47.59

Table 4
Monte Carlo simulations for the $A_1(3)$ model—structural parameters, weekly observations.

This table reports the results of 1,000 Monte Carlo simulations for the $A_1(3)$ model, with weekly observations, comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(MLE)}$ around the true value of the parameters $\theta^{(TRUE)}$, to the distribution of the differences between the reducible, Euler, and QML estimates ($\hat{\theta}^{(2)}$, $\hat{\theta}^{(Euler)}$, and $\hat{\theta}^{(QML)}$, respectively) and the exact MLE $\hat{\theta}^{(MLE)}$. The reducible likelihoods are based on the expansion with K=2 terms. The means and standard deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$ are reported as percentages of the corresponding deviations of $\hat{\theta}^{(MLE)}$ from $\theta^{(TRUE)}$. The means and standard deviations of $\hat{\theta}^{(Euler)}$ and $\hat{\theta}^{(QML)}$ from $\hat{\theta}^{(MLE)}$ are reported as multiples of the corresponding deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$. The results show that the difference $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ is much smaller, and often several orders of magnitude smaller, than the difference $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$ due to the sampling noise. It also shows that the approximate MLE is much closer than the Euler or QML estimate to the exact MLE; the mean and standard deviation of the difference is much smaller for the two-term approximate MLE than for Euler or QML, for most parameters; for a few, the QML means are smaller than the approximate MLE means, but the mean-squared errors for the same parameters (which combine the mean and standard deviation) are much larger for QML. The QML estimates are generally better than the Euler estimates, although not dramatically so.

Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$		$\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ (perc. of $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$)		$\hat{\theta}^{(Euler)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)		$\hat{\theta}^{(QML)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
a_1^M	0.0189	6.9E−04	4.2E−03	−2.10%	4.00%	−12.84	5.76	−8.70	5.44
b_{11}	−0.5000	−0.14	0.33	0.10%	2.41%	−52.99	4.00	−43.86	4.07
b_{22}	−0.0400	4.6E−05	1.6E−03	−1.96%	1.76%	−26.61	15.08	−21.72	15.35
b_{32}	1.1000	−2.8E−03	0.11	−4.47%	1.33%	9.84	6.59	7.66	7.47
b_{33}	−2.0000	−6.0E−03	0.031	0.19%	2.11%	3.79	12.84	13.88	12.89
d_0^M	0.0000	0.059	0.021	−0.11%	3.59%	2.81	3.69	1.06	3.98
d_1	0.0200	4.7E−04	2.3E−03	5.72%	4.52%	−22.53	7.32	−9.93	7.21
d_2	0.0400	−6.8E−05	3.0E−03	−4.85%	1.77%	178.88	5.62	0.22	5.42
d_3	0.0600	−2.5E−04	2.1E−03	0.01%	0.84%	43,405.07	19.01	256.44	13.24
λ_1	−0.1000	−0.11	0.30	0.21%	2.47%	−25.17	3.26	−21.42	3.29
λ_2	−0.2500	0.88	0.32	−0.13%	3.45%	2.06	3.79	0.34	3.99
λ_3	−0.3500	0.92	0.33	−0.02%	0.72%	−25.50	12.84	−3.15	11.22
σ_1	0.0100	4.8E−06	3.2E−04	−0.02%	0.07%	−2852.16	88.08	−2491.88	90.13
σ_2	0.0100	−1.5E−05	3.3E−04	−0.06%	0.09%	304.16	76.89	247.41	52.77
σ_3	0.0100	−8.8E−06	3.2E−04	−0.13%	0.09%	191.37	58.84	230.47	70.59

Table 5
Monte Carlo simulations for the $A_2(3)$ model—structural parameters, weekly observations.

This table reports the results of 993 Monte Carlo simulations for the $A_2(3)$ model, with weekly observations, comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(MLE)}$ around the true value of the parameters $\theta^{(TRUE)}$, to the distribution of the differences between the reducible, Euler, and QML estimates ($\hat{\theta}^{(2)}$, $\hat{\theta}^{(Euler)}$, and $\hat{\theta}^{(QML)}$, respectively) and the exact MLE $\hat{\theta}^{(MLE)}$. Seven of the originally generated 1,000 simulations have been deleted from the sample due to numeric stability problems in calculating the exact MLE, $\hat{\theta}^{(MLE)}$. The reducible likelihoods are based on the expansion with K=2 terms. The means and standard deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$ are reported as percentages of the corresponding deviations of $\hat{\theta}^{(MLE)}$ from $\theta^{(TRUE)}$. The means and standard deviations of $\hat{\theta}^{(Euler)}$ and $\hat{\theta}^{(QML)}$ from $\hat{\theta}^{(MLE)}$ are reported as multiples of the corresponding deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$. The results show that the difference $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ is much smaller, and often several orders of magnitude smaller, than the difference $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$ due to the sampling noise. It also shows that the approximate MLE is much closer than the Euler or QML estimate to the exact MLE; the mean and standard deviation of the difference is much smaller for the two-term approximate MLE than for Euler or QML, for most parameters; for a few, the QML means are smaller than the approximate MLE means, but the mean-squared errors for the same parameters (which combine the mean and standard deviation) are much larger for QML. The QML estimates are generally better than the Euler estimates, although not dramatically so.

Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$		$\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ (perc. of $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$)		$\hat{\theta}^{(Euler)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)		$\hat{\theta}^{(QML)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
a_1^M	0.0025	0.018	0.050	−0.01%	0.04%	−1119.50	943.40	−1390.04	884.09
a_2^M	0.0300	0.035	0.081	−0.06%	0.16%	−112.48	214.35	−122.17	230.79
b_{11}	−0.1000	0.067	0.038	0.01%	0.09%	6.23	156.01	−64.26	139.22
b_{22}	−0.7500	0.060	0.17	0.27%	0.16%	23.62	47.13	−2.93	44.12
b_{33}	−5.0000	−0.20	0.52	−0.93%	0.61%	24.71	19.88	−0.69	12.68
d_0^M	−0.0017	−9.2E−03	0.039	−0.00%	0.04%	−6864.24	965.82	−8190.57	917.43
d_1	0.0025	−4.9E−04	8.4E−04	0.07%	0.10%	73.28	146.54	55.89	140.87
d_2	0.0100	1.6E−04	2.9E−03	−0.21%	0.08%	980.70	263.29	376.34	265.30
d_3	0.0500	−4.0E−04	1.9E−03	−1.11%	0.20%	−509.33	48.38	−0.47	22.95
λ_1	0.0050	0.074	0.036	−0.02%	0.09%	50.53	151.78	44.51	130.36
λ_2	−0.0500	0.11	0.12	−0.03%	0.05%	33.27	121.83	7.16	117.19
λ_3	−0.1000	6.25	0.70	−0.04%	0.50%	−113.09	18.09	−0.69	13.64
σ_1	0.0100	−1.0E−05	3.2E−04	0.01%	0.00%	11.90	43.79	20.59	37.47
σ_2	0.0100	−1.4E−05	3.2E−04	0.02%	0.01%	26.85	53.36	18.29	44.20
σ_3	0.0100	−9.8E−06	3.2E−04	0.04%	0.01%	16.52	59.90	7.93	51.01

deviations of the differences between the alternative approximations, $\hat{\theta}^{(Euler)}$ and $\hat{\theta}^{(QML)}$, and the true MLE, $\hat{\theta}^{(MLE)}$, are reported as multiples of the biases and standard deviations of the differences between $\hat{\theta}^{(2)}$ and $\hat{\theta}^{(MLE)}$.

4.2. Identification

Affine models do not have unique representations; given any affine yield model, an equivalent model can be constructed by replacing the state vector with a non-singular affine transformation of itself. To ensure that the estimated models have a unique representation, additional restrictions are needed. We use the parameterization of the Dai-Singleton canonical models, with the modification to the drift of the state variables discussed earlier. However, we make two additional modifications to the parameterization when reporting the results. The Dai-Singleton normalizations, although useful in characterizing the varying families of affine yield models, are unstable at certain values of the parameter vector, and result in poor identification of some parameters near these values. For example, consider the single-factor affine yield model

$$dr_t = (c_1 + c_2 r_t)\,dt + c_3 dW_t. \tag{26}$$

This model is in the $A_0(1)$ family, and is equivalent to the model of Vasicek (1977). In the Dai-Singleton canonical form (with our alternate drift parameterization), this model is described as

$$dX_{1t} = (b_{11}X_t)\,dt + dW_t, \tag{27}$$

$$r_t = \delta_0 + \delta X_{1t}, \tag{28}$$

where b_{11}=c_2, $\delta_0 = -c_1/c_2$, and $\delta = c_3$. If c_1=c_2=0, then the δ_0 parameter is unidentified. When c_2 is very small and c_1 is not, then the δ_0 parameter is very poorly identified; even small sampling variation in the estimated c_2 parameter (i.e., the constant term in the drift of the observed quantity, r_t) results in huge variation in the estimated δ_0 parameter. When a large number of simulations are performed, it is likely that some of the estimated values of b_{11} (and therefore c_2, in the above representation) will be close to zero, and the δ_0 parameter estimates then take on very extreme values. In this case, the distribution of the estimated δ_0 parameter provides little useful information about the accuracy of the estimation; although the δ_0 estimate may be very far from the true value, this deviation has little effect on the implied process followed by the interest rate and bond yields. This problem is exacerbated in models with multiple state variables, where δ_0 is poorly identified when any of the eigenvalues of the B matrix are close to zero. We therefore report instead of δ_0 a modified parameter δ_0^M, which is the δ_0 parameter multiplied by the harmonic average of the eigenvalues of B, that is

$$\delta_0^M = \delta_0 \Big/ \sum_{i=0}^{N} \frac{1}{\nu_i}, \tag{29}$$

Table 6
Monte Carlo simulations for the $A_3(3)$ model—structural parameters, weekly observations.

This table reports the results of 918 Monte Carlo simulations for the $A_3(3)$ model, with weekly observations, comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(MLE)}$ around the true value of the parameters $\theta^{(TRUE)}$, to the distribution of the differences between the reducible, Euler, and QML estimates ($\hat{\theta}^{(2)}$, $\hat{\theta}^{(Euler)}$, and $\hat{\theta}^{(QML)}$, respectively) and the exact MLE $\hat{\theta}^{(MLE)}$. Of the originally generated 1,000 simulations, 82 have been deleted from the sample due to numeric stability problems in calculating the exact MLE, $\hat{\theta}^{(MLE)}$. The reducible likelihoods are based on the expansion with K=2 terms. The means and standard deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$ are reported as percentages of the corresponding deviations of $\hat{\theta}^{(MLE)}$ from $\theta^{(TRUE)}$. The means and standard deviations of $\hat{\theta}^{(Euler)}$ and $\hat{\theta}^{(QML)}$ from $\hat{\theta}^{(MLE)}$ are reported as multiples of the corresponding deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$. The results show that the difference $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ is much smaller, and often several orders of magnitude smaller, than the difference $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$ due to the sampling noise. It also shows that the approximate MLE is much closer than the Euler or QML estimate to the exact MLE; the mean and standard deviation of the difference is much smaller for the two-term approximate MLE than for Euler or QML, for most parameters; for a few, the QML means are smaller than the approximate MLE means, but the mean-squared errors for the same parameters (which combine the mean and standard deviation) are much larger for QML. The QML estimates are generally better than the Euler estimates, although not dramatically so.

Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$		$\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ (perc. of $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$)		$\hat{\theta}^{(Euler)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)		$\hat{\theta}^{(QML)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
a_1^M	0.0025	0.102	0.097	0.04%	0.91%	−23.04	19.32	19.79	8.28
a_2^M	0.1767	0.52	0.92	−0.29%	3.09%	−16.89	16.42	−36.16	15.42
a_3^M	0.0300	2.7	1.3	0.13%	2.32%	−4.58	10.35	12.77	6.11
b_{11}	−0.1000	0.091	0.019	−0.01%	0.60%	−19.72	22.89	13.26	10.43
b_{22}	−5.0000	−0.44	1.3	−2.76%	0.76%	23.53	13.63	0.04	3.80
b_{33}	−1.0000	0.38	0.18	−0.04%	0.77%	−2.76	14.39	14.24	8.69
d_0^M	−0.0009	−0.023	0.058	0.15%	0.56%	−27.73	51.51	15.18	14.11
d_1	0.0025	−6.2E−04	3.3E−04	−0.01%	0.79%	68.08	15.10	−41.46	7.18
d_2	0.0035	8.9E−04	3.0E−03	−0.02%	1.14%	2565.82	33.53	−514.41	36.21
d_3	0.0100	−6.7E−03	2.1E−03	0.03%	0.83%	15.28	13.52	13.88	8.20
λ_1	0.0200	0.093	0.018	−0.01%	0.59%	−16.27	22.81	12.88	10.33
λ_2	−0.2000	0.47	0.76	0.07%	0.74%	−5.64	37.37	29.62	39.48
λ_3	−0.1000	0.38	0.18	−0.05%	0.77%	0.40	14.40	13.53	8.67
σ_1	0.0100	−1.3E−05	3.2E−04	0.17%	0.15%	1.81	2.38	1.21	1.10
σ_2	0.0100	−1.7E−05	3.1E−04	−0.05%	0.09%	3.66	3.70	−0.68	1.87
σ_3	0.0100	−4.5E−06	3.2E−04	−0.10%	0.07%	3.57	2.87	−4.20	2.95

where $\nu_1,\ldots,\nu_N$ are the eigenvalues of B. For consistency, this normalization is used in all models, even those for which non-identification of the δ_0 parameter is not an issue.

A similar problem occurs in some models when elements of the δ vector are very small. Consider the single factor affine yield in which the interest rate process is

$$dr_t = (c_1 + c_2 r_t)\,dt + \sqrt{c_3 + c_4 r_t}\,dW_t. \tag{30}$$

(These parameters must be restricted to ensure existence of this process.) This model is in the $A_1(1)$ family (which is a slight generalization of the model of Cox, Ingersoll, and Ross, 1985), and the Dai-Singleton canonical representation (with our modified parameterization of the drift) is

$$dX_{1t} = (a_1 + b_{11}X_t)\,dt + \sqrt{X_{1t}}\,dW_t, \tag{31}$$

$$r_t = \delta_0 + \delta X_{1t}, \tag{32}$$

where a_1=$c_1/c_4 - c_2c_3/c_4^2$, b_{11}=c_2, $\delta_0 = -c_3/c_4$, and $\delta = c_4$.

Note that, when c_4 is very small, two of the parameters may take on very large values; the a_1 parameter is more severely affected, since the square of c_4 appears in the denominator. The distribution of the estimated values of a_1 therefore provides little useful information by which we can judge the quality of the estimation. In multiple-factor models, this problem can occur with the a_i parameter for any $1 \le i \le M$. We therefore report instead a scaled version of a_i parameter[8]

$$a_i^M = a_i \delta_i. \tag{33}$$

With some models, even with the normalizations discussed, another identification issue can arise, which is the reordering of state variables (see Babbs and Nowman, 1999). For example, in the $A_3(3)$ model, if the off-diagonal elements of the drift are zero, then it is possible simply to reorder the three state variables. The parameter vector is then different, even though the implied interest rate and bond yield processes are exactly the same as without the reordering. To guard against this problem, we sort by the eigenvalues of the drift matrix. With the parameter restrictions imposed on this model in the Monte Carlo simulations, the eigenvalues of the drift matrix are simply the diagonal elements, and each one is clearly associated with a state variable. When reporting the results, if the size ordering of the eigenvalues of the estimated drift matrix is different than the size ordering of the eigenvalues from the

[8] Collin-Dufresne, Goldstein, and Jones (2003) propose an alternative to the parameterization scheme in which the choice of state variables is based on observed quantities such as level and slope of the term structure. Although our parameterization is based on Dai and Singleton (2000), the modifications to this scheme we propose are similar in spirit to those of Collin-Dufresne, Goldstein, and Jones (2003).

data-generating drift matrix, we reorder the state variables. This procedure eliminates this final identification problem for the $A_2(2)$, $A_2(3)$, and $A_3(3)$ models.

A similar identification issue potentially arises in the $A_0(2)$, $A_0(3)$, and $A_1(3)$ models, although in our simulations, it has manifested itself only in the $A_0(3)$ case. For this latter model, the matrix of slope coefficients in the drift, in the Dai-Singleton canonical form, is

$$B = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix}. \tag{34}$$

For over 400 of the 1,000 simulations for this model, the ordering of the diagonal elements is different, for at least one of the estimation methods, from the data-generating values. In this case, the state variables cannot simply be reordered while still preserving the Dai-Singleton canonical form, because a reordering would cause some of the elements of B above the diagonal to be non-zero. Investigation of these estimates shows, for these simulations, that the true likelihood function is extremely flat in some directions in the parameter space. In one extreme case, the true likelihoods of the MLE estimated with and without an ordering constraint on the diagonal elements of B differ beginning only in the twelfth significant decimal place. The reversal of the diagonal elements is therefore indicative of poor identification, and, absent any convenient alternate parameterization that is better identified, we present the results for this model in a second form as well, expressing the state variable dynamics in terms of observed quantities.

Affine yield models are usually expressed (as we express them) in terms of latent variables, rather than directly observable quantities such as yields. As noted by Duffie and Kan (1996), these models can be expressed in terms of bond yields instead of latent state variables; however, expressing the parameter restrictions needed to satisfy existence and no-arbitrage conditions in a straightforward way, and imposing them in estimation, has proved challenging; see Collin-Dufresne, Goldstein, and Jones (2003) for one approach. However, given an estimated model (expressed in a latent variable form), it is relatively easy to construct an observationally equivalent model in terms of bond yields. Furthermore, since bond yields are well-defined and observable quantities, there are no identification problems. Therefore, even though the models are estimated using their latent variable representations (with the Dai-Singleton normalizations, and the modifications discussed above), we can also report the results in yield autoregressive form

$$y_m = \varphi_m + \sum_{n=1}^{N} \psi_{mn} y_n + \eta_m, \tag{35}$$

where $y_1,\ldots,y_N$ are the first N yields generated (i.e., those observed without error); unlike the yields referenced in Section 3, these yields are annualized. The residuals $\eta_1,\ldots,\eta_N$ have mean zero and covariance structure

$$Cov[\eta_i,\eta_j] = \omega_{ij0} + \sum_{k=1}^{M} \omega_{ijk} y_k, \tag{36}$$

with the stipulation that $\omega_{ijk} = \omega_{jik}$ for any $1 \le i,j \le N$ and $0 \le k \le M$.

Presenting the results for the $A_0(3)$ model in yield autoregressive form resolves the final identification issue for this model; these results are presented in Table 7, and, as discussed in the next section, confirm the accuracy of our method. The same issue could also have affected other models, but an examination of the parameter estimates shows that it did not; for these models, not a single simulation experienced a reversal of the diagonal elements of the drift matrix, for either the true MLE, or for the estimates using any of the approximation methods.

4.3. Results

Tables 3–6 present the Monte Carlo results for the three-factor models with weekly observations, in terms of the model parameters. Deferring discussion of the $A_0(3)$ model for now, our technique produces parameter estimates for the other models that are extremely close to the MLE based on the true likelihood, both in absolute terms and relative to the sampling distribution of the latter relative to the true parameter. The mean difference between the two estimates is typically very small compared to the mean difference between the true maximum likelihood estimator and the true parameter value; the standard deviation of the difference between the two estimators is also typically very small compared to the standard deviation of the MLE itself; the ratios are no more than a few percent, and for many parameters, much less than 1%. These results suggest that the approximation error introduced by our likelihood approximation is swamped by the sampling error of the MLE estimator, i.e., the noise resulting from the fact that the parameters are estimated from random data. Consequently, the exact MLE can be replaced by our estimator at almost no cost (and of course, our estimator can always be calculated, unlike the exact MLE, which is only available for models which have a known closed-form likelihood).

Tables 3–6 also present the difference between estimates obtained using the Euler approximate likelihood and the true MLE, and between estimates using the QML approximate likelihood and the true MLE (for those models for which these two are different). For QML, the conditional means and variances were calculated numerically; estimates using explicit expressions for the conditional moments for the one- and two-factor models (not presented) are very close to the estimates using numeric moments. QML estimates are only presented when they are different from the true MLE, i.e., for all models except $A_0(3)$. Again deferring discussion of the $A_0(3)$ model, both of these methods produce for the other models estimates with substantially larger deviations from the true MLEs than our method. For nearly all parameters in all models, the mean of the difference between the estimates using these two methods and the true MLE is much larger (in absolute magnitude) than the mean difference between estimates from the 2-term approximate likelihood and the true MLE; in all cases, the Euler and QML methods have

Table 7
Monte Carlo simulations for the $A_0(3)$ model—yield autoregressions, weekly observations.

This table reports the results of 1,000 Monte Carlo simulations for the $A_0(3)$ model in yield autoregressive form, with weekly observations, comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(MLE)}$ around the true value of the parameters $\theta^{(TRUE)}$, to the distribution of the differences between the reducible and Euler estimates ($\hat{\theta}^{(2)}$ and $\hat{\theta}^{(Euler)}$, respectively) and the exact MLE $\hat{\theta}^{(MLE)}$. The reducible likelihoods are based on the expansion with $K=2$ terms. QML estimates are not presented, since for this model, they coincide with the exact MLE. The means and standard deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$ are reported as percentages of the corresponding deviations of $\hat{\theta}^{(MLE)}$ from $\theta^{(TRUE)}$. The means and standard deviations of $\hat{\theta}^{(Euler)}$ from $\hat{\theta}^{(MLE)}$ are reported as multiples of the corresponding deviations of $\hat{\theta}^{(2)}$ from $\hat{\theta}^{(MLE)}$. The results show that the difference $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ is several orders of magnitude smaller than the difference $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$ due to the sampling noise. It also shows that the approximate MLE is much closer than the Euler estimate to the exact MLE; the mean and standard deviation of the difference is much smaller for the two-term approximate MLE than for the Euler likelihood, for all parameters. The analogous information expressed in terms of the parameters of the canonical model is presented in Table 3.

Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$		$\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$ (perc. of $\hat{\theta}^{(MLE)}-\theta^{(TRUE)}$)		$\hat{\theta}^{(Euler)}-\hat{\theta}^{(MLE)}$ (mult. of $\hat{\theta}^{(2)}-\hat{\theta}^{(MLE)}$)	
		Mean	Std. dev.	Mean	Std. dev.	Mean	Std. dev.
φ_1	−0.0115	−5.1E−04	8.1E−03	−0.09%	0.07%	43.11	33.32
φ_2	−0.0041	−2.0E−04	6.6E−03	0.01%	0.00%	178.09	359.59
φ_3	−0.0022	−1.7E−04	4.9E−03	−0.00%	0.00%	−175.75	539.41
ψ_{11}	−2.7824	−0.14	0.38	−0.19%	0.17%	32.75	25.76
ψ_{12}	5.4942	0.40	1.2	−0.19%	0.18%	32.10	25.20
ψ_{13}	−2.7540	−0.27	0.89	−0.19%	0.18%	31.61	25.46
ψ_{21}	−0.5818	4.3E−03	0.016	−0.17%	0.08%	35.90	32.98
ψ_{22}	0.0613	−0.012	0.049	−0.17%	0.08%	35.00	28.78
ψ_{23}	0.5251	7.7E−03	0.039	−0.18%	0.08%	34.10	27.34
ψ_{31}	−0.2150	−1.8E−03	8.6E−03	−0.19%	0.06%	36.55	35.71
ψ_{32}	−0.4608	4.4E−03	0.029	−0.20%	0.06%	35.27	28.64
ψ_{33}	0.6710	−2.2E−03	0.026	−0.26%	0.05%	33.79	26.40
ω_{110}	0.0006	−3.6E−06	3.7E−05	−0.06%	0.01%	−3634.91	653.28
ω_{120}	0.0004	−2.4E−06	2.6E−05	−0.03%	0.01%	643.76	565.89
ω_{130}	0.0003	−1.8E−06	2.0E−05	−0.03%	0.01%	2307.07	526.67
ω_{210}	0.0004	−2.4E−06	2.6E−05	−0.03%	0.01%	643.76	565.89
ω_{220}	0.0004	−1.4E−06	2.3E−05	0.02%	0.01%	3259.58	190.58
ω_{230}	0.0003	−1.0E−06	1.8E−05	0.02%	0.01%	2836.08	167.21
ω_{310}	0.0003	−1.8E−06	2.0E−05	−0.03%	0.01%	2307.07	526.67
ω_{320}	0.0003	−1.0E−06	1.8E−05	0.02%	0.01%	2836.08	167.21
ω_{330}	0.0002	−7.6E−07	1.4E−05	0.03%	0.01%	3037.14	153.93

much larger standard deviation from the true MLE than the 2-term approximate likelihood.[9] Even in the few cases where the other two methods have smaller mean differences, the standard deviation of the difference is much larger, so that the 2-term likelihood method produces much smaller mean-squared error. Perhaps surprisingly, the QML method usually produces only a small improvement over the Euler method (except for the Gaussian models, for which the QML estimates coincide with the true MLE); most of the differences from the true MLE seem to be driven by the use of a Gaussian density when the true density is non-Gaussian, rather than the Euler method's use of approximate rather than true conditional moments.

The results for the unmodified a_i and δ_0 parameters (not presented) are characterized by a few very extreme outliers, in both the true and approximate MLE results. The estimates, as well as the conformance between the true and approximate estimates, of the modified versions of these parameters suggest that the extreme values of the a_i parameters occur only when the corresponding δ_i parameters are very close to zero, and the extreme values of the δ_0 parameter occur only when the B matrix contains a very small eigenvalue. As argued above, these extreme values are then indicative of the poor identification of the parameter which is intrinsic to the parameterization method, rather than poor performance of the estimation technique. The results in yield autoregressive form are presented only for the $A_0(3)$ model in Table 7, to save space, but were calculated for all models, and confirm this interpretation. The dynamics of observed yields implied by the approximate MLE closely conform to those implied by the true MLE.

From Table 3, our method appears to perform relatively poorly for the $A_0(3)$ model[10]; however, closer examination

[9] Note that the biases for the Euler and QML estimates are reported as multiples of the corresponding biases for the reducible estimates. In a few cases, the sample bias for the reducible estimates was very close to zero, with the result that the biases for the other two methods are many times larger. In these cases, it does not necessarily follow that the bias for the other two methods is extreme, merely that it is very small for the reducible method. This phenomenon did not occur for any parameter estimates in reverse, that is, none of the Euler or QML estimates had extremely small bias, so that this multiple was extremely close to zero. Also, this phenomenon does not occur with the standard deviations. In any event, the tables contain all the information necessary to back out the absolute instead of relative estimation errors, if so desired.

[10] It is unlikely that anyone would use our method to estimate an $A_0(3)$ model, since the true likelihood function is known explicitly, but we include results anyway for completeness.

shows that this is not the case. The errors in estimation of the parameter vector introduced by our method as reported in the table are still substantially smaller than the sampling variability inherent in maximum likelihood estimation, but are relatively larger than in the other models. However, much of the deviation of our approximate MLE from the true MLE is symptomatic of an identification issue, as discussed in the previous section, not a problem with our method. The results for the $A_0(3)$ model are presented in yield autoregressive form as well in Table 7. As shown, the parameters of the yield autoregressive representation of the model estimated using our method are extremely close to those parameters estimated using the exact likelihood function; the mean difference and standard deviation of difference are, in all cases, much smaller than 1% of the bias and standard deviation of the true MLE itself. The Euler estimates are much less accurate, with biases and standard deviations always dozens, and often hundreds of times larger than our technique.[11] For this model, the QML likelihood coincides with true likelihood, so results are not presented.

In summary, we find that any additional bias and variance introduced by the use of an approximate likelihood are insignificant in magnitude relative to the bias and variance of the MLE estimator itself, so that use of our approximations does not result in a degradation of the quality of the MLE estimates. We further find that estimates using the Euler and QML methods deviate much more from the MLE estimate, with the additional error sometimes nearly as large as the intrinsic variability of MLE itself. Although the $A_0(3)$ model (for which the QML likelihood approximation is exact) appears to be an exception, in which Euler performs as well as our method for at least some of the parameters, this is the side-effect of an identification problem; when the results for this model are viewed in yield autoregressive form, our method strongly outperforms the Euler method, as it does in the other models. This outperformance relative to Euler occurs with the $A_0(3)$ model, despite the fact that the Gaussian likelihood of the state vector for this model would seem to give Gaussian approximation methods an advantage.

5. Empirical application

We now estimate the nine affine models with one, two, and three factors on a data set of synthetic US Treasury strips, constructed by the method of McCulloch (1975). This method for construction of strips prices from prices of coupon securities was used to construct the data set of McCulloch and Kwon (1993); the method was evaluated by Bliss (1997), who also produces periodic extensions of the data set. We use synthetic strips for a period of 31 years, from January of 1972 until December of 2002; although the data set includes yields from January 1970, yields for the longer maturities used in this study are not available for the first two years. The selection of maturities, parameterization of the models, etc., follows that of Section 4 closely. In particular, we use the modified a_i^M parameters; however, since poor identification of the δ_0 parameter is not evident in the estimates for any of the models, we do not use the modified version of this parameter. We use a number of observed yields equal to twice the number of factors in the model, with maturities spaced every two years, except that the shortest maturity is one month rather than zero.[12] For an N-factor model, the N shortest maturities are considered observed without error, and the remaining yields are considered to have observation error. For example, in the $A_M(3)$ models, maturities of one month, two years, and four years are assumed observed without error, and maturities of six, eight, and 10 years are observed with error. All models are estimated four ways: with and without the parameter restrictions used in Section 4 to ensure existence of a closed-form likelihood, and with the Dai-Singleton and Cheridito-Filipović-Kimmel market price of risk specifications. For the $A_0(N)$ models, there are no parameter restrictions needed for existence of a closed-form likelihood, so there are only two distinct estimations to perform; for the other models, all four possible combinations are estimated. In all cases, the estimates are found using the 2-term approximate likelihoods as well as the Euler and QML approximations; the true MLE is also estimated when the exact likelihood function is known in closed-form.

The parameter estimates, with standard errors, for the three-factor models with the Dai-Singleton market price of risk specification, are presented in Tables 8, 10, and 12 using the 2-term approximate likelihoods, Euler approximations, and QML approximations, respectively. Analogous estimates using the Cheridito-Filipović-Kimmel market price of risk specification are presented in Tables 9, 11, and 13. Standard errors have been calculated by evaluating the sample means of the cross products of the score functions, which in turn have been calculated by numeric differentiation of the log likelihood. All six tables present results for all three-factor models, both with and without the parameter restrictions for existence of a closed-form likelihood, from Table 2 (Tables 10–13). Not all parameters appear in all model specifications, so the tables contain some blank entries. Estimates using the true likelihood function and for the one- and two-factor models are not presented, although those estimates are used in the likelihood ratio tests discussed later. An examination of the true estimates (which are available only for the restricted versions of the models, and, as per the above, are not presented) suggests that they are generally quite close to the estimates using reducible likelihoods. Estimates for the $A_3(3)$ restricted model, with the Dai-Singleton market price of risk, using the true likelihood function, were not calculated, due to a

[11] Further confirmation that the results presented in Table 3 are affected by an identification problem comes from the d_0^M, σ_1, σ_2, and σ_3 results. These parameters are invariant to linear transformation of the state vector, and therefore not subject to an identification problem. As shown, our method strongly outperforms Euler for these parameters.

[12] Although the original data sets of McCulloch (1975) and McCulloch and Kwon (1993) included zero-maturity yields, the subsequent extensions do not always do so.

Table 8
Parameter estimates for three-factor models—Dai-Singleton market price of risk, reducible likelihood.

This table reports the parameter estimates for the $A_0(3)$, $A_1(3)$, $A_2(3)$, and $A_3(3)$ models with the reducible likelihood approximation with K=2 terms. For the latter three models, results for both the restricted (i.e., with the parameter restrictions from Table 2 imposed) and unrestricted versions of the model are reported. Standard errors are reported in parentheses. The data used are synthetic US Treasury strips, constructed by the method of McCulloch (1975), from January of 1972 until December of 2002. Yields with maturities of one month, two years, and four years are considered observed without error; additional yields with maturities of six, eight, and 10 years are considered observed with error. See Section 5 for additional detail.

	$A_0(3)$		$A_1(3)$—Restricted		$A_1(3)$—Unrestricted		$A_2(3)$—Restricted		$A_2(3)$—Unrestricted		$A_3(3)$—Restricted		$A_3(3)$—Unrestricted	
Parameter	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.
a_1^M	–	–	0.00	(0.21)	0.00	(0.21)	0.04	(0.20)	0.01	(0.22)	0.00	(0.23)	0.0	(7.2)
a_2^M	–	–	–	–	–	–	0.00	(0.24)	0.00	(0.40)	0.0E+06	(8.9E+06)	0.0E+05	(4.4E+05)
a_3^M	–	–	–	–	–	–	–	–	–	–	0.05	(0.43)	0.0	(127.4)
b_{11}	−0.0168	(0.0015)	0.000	(0.039)	0.000	(0.037)	0.00	(0.16)	0.00	(0.14)	−0.037	(0.038)	−0.041	(0.039)
b_{12}	–	–	–	–	–	–	–	–	0.0135	(0.0079)	–	–	0.00000	(0.00023)
b_{13}	–	–	–	–	–	–	–	–	–	–	–	–	0.03	(0.18)
b_{21}	−0.40	(0.16)	–	–	−0.0	(3.3)	–	–	0.00	(0.14)	–	–	22.5	(53.9)
b_{22}	−2.96	(0.26)	−0.880	(0.026)	−2.0	(76.9)	−0.038	(0.038)	−0.039	(0.038)	−5.71	(0.39)	−3.28	(0.37)
b_{23}	–	–	–	–	0.4	(7.5)	–	–	–	–	–	–	191.6	(414.3)
b_{31}	0.64	(0.13)	–	–	0.14	(0.95)	–	–	1.67	(0.17)	–	–	0.066	(0.060)
b_{32}	2.56	(0.14)	2.32	(0.12)	2.9	(7.5)	–	–	0.180	(0.051)	–	–	0.0000	(0.0039)
b_{33}	−0.841	(0.022)	−2.69	(0.24)	−1.7	(77.0)	−5.65	(0.38)	−3.15	(0.26)	−0.81	(0.12)	−0.88	(0.26)
d_0	0.00	(0.18)	0.000	(0.015)	0.000	(0.014)	0.000	(0.015)	0.000	(0.012)	−0.3	(160.4)	−0.2	(10.2)
d_1	0.0048	(0.0012)	0.00275	(0.00051)	0.00093	(0.00028)	0.0194	(0.0016)	0.00390	(0.00063)	0.00278	(0.00053)	0.00069	(0.00032)
d_2	−0.0130	(0.0011)	0.0052	(0.0014)	−0.01	(0.63)	0.00282	(0.00055)	0.00094	(0.00029)	0.0002	(0.0043)	0.00015	(0.00030)
d_3	0.02413	(0.00045)	0.02811	(0.00070)	0.03	(0.13)	0.0391	(0.0013)	0.02699	(0.00062)	0.0145	(0.0014)	0.00231	(0.00074)
λ_1	−0.19	(0.19)	0.001	(0.039)	0.003	(0.037)	0.68	(0.16)	0.77	(0.14)	−0.034	(0.039)	−0.034	(0.035)
λ_2	0.61	(0.24)	−0.12	(0.20)	0.3	(25.8)	−0.035	(0.038)	−0.036	(0.038)	−0.01	(0.12)	−0.007	(0.013)
λ_3	−0.97	(0.23)	−0.97	(0.23)	−1.1	(6.7)	−1.10	(0.28)	−1.17	(0.26)	−0.10	(0.12)	−0.080	(0.098)
σ_1	0.001002	(6.1E−05)	0.000998	(6.1E−05)	0.000997	(6.1E−05)	0.000998	(6.1E−05)	0.000999	(6.2E−05)	0.000994	(6.2E−05)	0.000997	(6.2E−05)
σ_2	0.00172	(0.00021)	0.00172	(0.00022)	0.00172	(0.00022)	0.00173	(0.00021)	0.00172	(0.00022)	0.00172	(0.00021)	0.00172	(0.00022)
σ_3	0.00226	(0.00020)	0.00226	(0.00021)	0.00226	(0.00021)	0.00226	(0.00020)	0.00225	(0.00020)	0.00225	(0.00020)	0.00225	(0.00021)

Table 9
Parameter estimates for three-factor models—Cheridito-Filipović-Kimmel market price of risk, reducible likelihood.

This table reports the parameter estimates for the $A_0(3)$, $A_1(3)$, $A_2(3)$, and $A_3(3)$ models with the reducible likelihood approximation with K=2 terms. For the latter three models, results for both the restricted (i.e., with the parameter restrictions from Table 2 imposed) and unrestricted versions of the model are reported. Standard errors are reported in parentheses. The data used are synthetic US Treasury strips, constructed by the method of McCulloch (1975), from January of 1972 until December of 2002. Yields with maturities of one month, two years, and four years are considered observed without error; additional yields with maturities of six, eight, and 10 years are considered observed with error. See Section 5 for additional detail.

	$A_0(3)$		$A_1(3)$—Restricted		$A_1(3)$—Unrestricted		$A_2(3)$—Restricted		$A_2(3)$—Unrestricted		$A_3(3)$—Restricted		$A_3(3)$—Unrestricted	
Parameter	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.
a_1^M	–	–	0.0	(3.6)	0.0	(3.5)	−0.03	(0.82)	−0.0	(1.8)	0.0	(5.8)	0.0	(6.0)
a_2^M	–	–	–	–	–	–	−0.0	(4.0)	−0.0	(4.6)	0.00	(0.81)	0.0	(2.5)
a_3^M	–	–	–	–	–	–	–	–	–	–	0.0	(2.0)	0.0	(5.6)
b_{11}	−0.15	(0.12)	−0.16	(0.12)	−0.16	(0.11)	0.00	(0.12)	−0.00	(0.18)	−0.24	(0.13)	−0.50	(0.19)
b_{12}	–	–	–	–	–	–	–	–	0.000	(0.083)	–	–	0.16	(0.84)
b_{13}	–	–	–	–	–	–	–	–	–	–	–	–	1.0	(1.1)
b_{21}	−0.00	(0.66)	–	–	−0.09	(0.21)	–	–	0.82	(0.49)	–	–	0.00	(0.10)
b_{22}	−3.72	(0.42)	−3.48	(0.41)	−3.55	(0.42)	−0.18	(0.12)	−0.37	(0.16)	−0.16	(0.12)	−0.61	(0.46)
b_{23}	–	–	–	–	−0.3	(13.8)	–	–	–	–	–	–	0.54	(0.52)
b_{31}	−0.39	(0.24)	–	–	0.04	(0.44)	–	–	0.36	(0.13)	–	–	0.00	(0.21)
b_{32}	0.95	(0.57)	0.53	(0.56)	0.3	(13.7)	–	–	0.073	(0.046)	–	–	1.91	(0.74)
b_{33}	−0.85	(0.24)	0.0000	(0.0041)	−0.63	(0.24)	−2.65	(0.35)	−3.33	(0.37)	−0.00	(0.18)	−2.71	(0.41)
d_0	0.006	(0.031)	−0.0	(10.4)	0.000	(0.026)	−0.0	(12.3)	0.0	(6.5)	−0.000	(0.016)	−0.002	(0.016)
d_1	0.0171	(0.0054)	0.00087	(0.00037)	0.00039	(0.00053)	−0.0	(1.7)	−0.1	(2.1)	0.00037	(0.00043)	0.00027	(0.00050)
d_2	−0.029	(0.012)	−0.0202	(0.0037)	−0.027	(0.079)	−0.00	(0.11)	−0.01	(0.31)	0.00006	(0.00077)	0.00045	(0.00089)
d_3	0.0080	(0.0048)	0.0202	(0.0014)	0.02	(0.13)	0.3	(10.5)	0.3	(11.0)	0.0076	(0.0014)	0.0089	(0.0018)
a_1^Q	1.5	(1.8)	0.76	(0.41)	0.76	(0.23)	1.95	(0.89)	1.73	(0.99)	0.5	(1.2)	0.5	(1.1)
a_2^Q	−2.7	(4.8)	−60.8	(589.3)	4.8	(8.4)	0.50	(0.49)	0.50	(0.53)	0.7	(2.1)	0.6	(1.9)
a_3^Q	0.44	(0.62)	−10.6	(73.7)	−1.3	(22.5)	−0.2E+03	(7.5E+03)	−0.1E+03	(4.0E+03)	0.5	(7.5)	0.5	(8.0)
b_{11}^Q	0.9	(1.2)	−0.0054	(0.0044)	−0.0055	(0.0032)	−0.728	(0.031)	−0.720	(0.038)	−0.124	(0.051)	−0.129	(0.054)
b_{12}^Q	7.0	(2.5)	–	–	–	–	0.145	(0.055)	0.158	(0.065)	0.5	(1.1)	0.4	(1.1)
b_{13}^Q	0.06	(0.88)	–	–	–	–	–	–	–	–	0.00	(0.58)	0.00	(0.93)
b_{21}^Q	−2.6	(3.9)	−0.164	(0.097)	−0.49	(0.99)	0.12	(0.11)	0.14	(0.11)	0.187	(0.088)	0.216	(0.084)
b_{22}^Q	−15.8	(8.8)	−7.9	(3.3)	−12.1	(9.1)	−0.031	(0.029)	−0.039	(0.035)	−1.70	(0.58)	−1.18	(0.67)
b_{23}^Q	−1.3	(1.8)	−1.13	(0.68)	−3.2	(58.5)	–	–	–	–	0.52	(0.21)	0.42	(0.42)
b_{31}^Q	0.06	(0.51)	0.131	(0.059)	0.2	(2.3)	0.0E+03	(1.0E+03)	0.0E+03	(1.3E+03)	0.00	(0.28)	0.00	(0.33)
b_{32}^Q	1.8	(2.5)	4.0	(1.5)	3.9	(58.7)	1.8	(66.5)	5.9	(192.6)	5.5	(1.6)	4.7	(1.7)
b_{33}^Q	−0.30	(0.10)	−0.15	(0.26)	0.3	(3.1)	−0.2E+03	(6.4E+03)	−0.2E+03	(6.8E+03)	−3.46	(0.81)	−4.5	(1.2)
σ_1	0.000991	(6.3E−05)	0.000988	(6.2E−05)	0.000987	(6.3E−05)	0.000987	(6.1E−05)	0.000987	(6.2E−05)	0.000992	(6.2E−05)	0.000991	(6.3E−05)
σ_2	0.00171	(0.00023)	0.00171	(0.00023)	0.00171	(0.00023)	0.00171	(0.00022)	0.00171	(0.00022)	0.00172	(0.00023)	0.00171	(0.00023)
σ_3	0.00226	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)

Table 10
Parameter estimates for three-factor models—Dai-Singleton market price of risk, Euler likelihood.

This table reports the parameter estimates for the $A_0(3)$, $A_1(3)$, $A_2(3)$, and $A_3(3)$ models with the Euler likelihood approximation. For the latter three models, results for both the restricted (i.e., with the parameter restrictions from Table 2 imposed) and unrestricted versions of the model are reported. Standard errors are reported in parentheses. The data used are synthetic US Treasury strips, constructed by the method of McCulloch (1975), from January of 1972 until December of 2002. Yields with maturities of one month, two years, and four years are considered observed without error; additional yields with maturities of six, eight, and 10 years are considered observed with error. See Section 5 for additional detail.

	$A_0(3)$		$A_1(3)$—Restricted		$A_1(3)$—Unrestricted		$A_2(3)$—Restricted		$A_2(3)$—Unrestricted		$A_3(3)$—Restricted		$A_3(3)$—Unrestricted	
Parameter	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.
a_1^M	–	–	0.00	(0.23)	0.00	(0.23)	0.04	(0.23)	0.01	(0.29)	0.00	(0.41)	0.0E+03	(1.2E+03)
a_2^M	–	–	–	–	–	–	0.00	(0.26)	0.00	(0.48)	0.0E+06	(1.9E+06)	0.0E+05	(2.3E+05)
a_3^M	–	–	–	–	–	–	–	–	–	–	0.05	(0.55)	0.0	(223.9)
b_{11}	−0.0164	(0.0015)	0.000	(0.039)	0.000	(0.037)	0.00	(0.14)	0.00	(0.13)	−0.037	(0.038)	−0.041	(0.081)
b_{12}	–	–	–	–	–	–	–	–	0.020	(0.012)	–	–	0.0000	(0.0095)
b_{13}	–	–	–	–	–	–	–	–	–	–	–	–	0.1	(1.1)
b_{21}	−0.33	(0.14)	–	–	0.0	(7.4)	–	–	0.00	(0.17)	–	–	31.7	(30.0)
b_{22}	−2.67	(0.21)	−0.898	(0.028)	−0.8	(137.0)	−0.037	(0.038)	−0.038	(0.040)	−4.83	(0.26)	−2.95	(0.39)
b_{23}	–	–	–	–	−0.0	(106.1)	–	–	–	–	–	–	347.7	(91.0)
b_{31}	0.59	(0.12)	–	–	0.13	(0.90)	–	–	1.48	(0.15)	–	–	0.068	(0.056)
b_{32}	2.42	(0.12)	2.23	(0.10)	2.4	(106.0)	–	–	0.143	(0.045)	–	–	0.0000	(0.0017)
b_{33}	−0.860	(0.024)	−2.46	(0.20)	−2.7	(136.9)	−4.78	(0.25)	−2.83	(0.22)	−0.80	(0.11)	−0.88	(0.24)
d_0	0.00	(0.18)	0.000	(0.016)	0.000	(0.015)	0.000	(0.016)	0.000	(0.014)	−0.4	(16.5)	−0.3	(2.1)
d_1	0.0059	(0.0011)	0.00273	(0.00055)	0.00113	(0.00031)	0.0180	(0.0016)	0.00485	(0.00058)	0.00277	(0.00083)	0.00084	(0.00064)
d_2	−0.00889	(0.00093)	0.0066	(0.0014)	0.0	(1.4)	0.00280	(0.00059)	0.00112	(0.00031)	0.00008	(0.00011)	6.7E−05	(2.0E−05)
d_3	0.02354	(0.00044)	0.02545	(0.00063)	0.02	(0.41)	0.0319	(0.0011)	0.02413	(0.00056)	0.0132	(0.0013)	0.00308	(0.00055)
λ_1	−0.19	(0.20)	0.001	(0.039)	0.003	(0.037)	0.68	(0.14)	0.77	(0.13)	−0.035	(0.039)	−0.031	(0.047)
λ_2	0.60	(0.24)	−0.10	(0.20)	−0.1	(68.0)	−0.035	(0.039)	−0.036	(0.040)	−0.0029	(0.0040)	−0.0034	(0.0012)
λ_3	−1.03	(0.23)	−1.02	(0.23)	−1.2	(8.7)	−1.16	(0.28)	−1.22	(0.26)	−0.09	(0.10)	−0.076	(0.096)
σ_1	0.001005	(6.1E−05)	0.001000	(6.1E−05)	0.001000	(6.1E−05)	0.001001	(6.1E−05)	0.001003	(6.2E−05)	0.000996	(6.1E−05)	0.001000	(6.2E−05)
σ_2	0.00172	(0.00021)	0.00172	(0.00022)	0.00172	(0.00022)	0.00173	(0.00021)	0.00173	(0.00022)	0.00172	(0.00021)	0.00172	(0.00022)
σ_3	0.00226	(0.00020)	0.00227	(0.00020)	0.00226	(0.00021)	0.00226	(0.00020)	0.00226	(0.00020)	0.00226	(0.00020)	0.00225	(0.00021)

Table 11
Parameter estimates for three-factor models—Cheridito-Filipović-Kimmel market price of risk, Euler likelihood.

This table reports the parameter estimates for the $A_0(3)$, $A_1(3)$, $A_2(3)$, and $A_3(3)$ models with the Euler likelihood approximation. For the latter three models, results for both the restricted (i.e., with the parameter restrictions from Table 2 imposed) and unrestricted versions of the model are reported. Standard errors are reported in parentheses. The data used are synthetic US Treasury strips, constructed by the method of McCulloch (1975), from January of 1972 until December of 2002. Yields with maturities of one month, two years, and four years are considered observed without error; additional yields with maturities of six, eight, and 10 years are considered observed with error. See Section 5 for additional detail.

	$A_0(3)$		$A_1(3)$—Restricted		$A_1(3)$—Unrestricted		$A_2(3)$—Restricted		$A_2(3)$—Unrestricted		$A_3(3)$—Restricted		$A_3(3)$—Unrestricted	
Parameter	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.
a_1^M	–	–	0.0	(3.6)	0.0	(3.3)	−0.03	(0.79)	−0.0	(1.8)	0.0	(5.5)	0.0	(5.6)
a_2^M	–	–	–	–	–	–	−0.0	(4.1)	−0.0	(4.7)	−0.00	(0.65)	0.0	(2.2)
a_3^M	–	–	–	–	–	–	–	–	–	–	0.0	(1.8)	0.0	(4.8)
b_{11}	−0.14	(0.12)	−0.15	(0.11)	−0.15	(0.11)	0.00	(0.11)	−0.00	(0.16)	−0.21	(0.13)	−0.45	(0.18)
b_{12}	–	–	–	–	–	–	–	–	0.000	(0.080)	–	–	0.00	(0.86)
b_{13}	–	–	–	–	–	–	–	–	–	–	–	–	1.18	(0.99)
b_{21}	0.02	(0.53)	–	–	−0.08	(0.23)	–	–	0.79	(0.45)	–	–	0.000	(0.086)
b_{22}	−3.25	(0.31)	−3.07	(0.30)	−3.13	(0.40)	−0.19	(0.12)	−0.37	(0.16)	−0.25	(0.12)	−0.56	(0.35)
b_{23}	–	–	–	–	−0.3	(14.1)	–	–	–	–	–	–	0.41	(0.39)
b_{31}	−0.38	(0.22)	–	–	0.04	(0.45)	–	–	0.31	(0.11)	–	–	0.00	(0.17)
b_{32}	0.79	(0.47)	0.46	(0.49)	0.3	(14.1)	–	–	0.058	(0.039)	–	–	1.88	(0.57)
b_{33}	−0.82	(0.23)	0.0003	(0.0032)	−0.62	(0.37)	−2.41	(0.28)	−2.94	(0.28)	−0.00	(0.17)	−2.29	(0.32)
d_0	0.006	(0.031)	−0.0	(5.1)	0.000	(0.026)	−0.0	(5.1)	0.00	(0.43)	−0.000	(0.017)	−0.002	(0.015)
d_1	0.0169	(0.0050)	0.00088	(0.00038)	0.00055	(0.00046)	−0.03	(0.71)	−0.03	(0.14)	0.00057	(0.00048)	0.00046	(0.00047)
d_2	−0.026	(0.012)	−0.0170	(0.0031)	−0.022	(0.096)	−0.001	(0.034)	−0.004	(0.017)	−0.00027	(0.00063)	0.00079	(0.00064)
d_3	0.0076	(0.0046)	0.0202	(0.0014)	0.02	(0.13)	0.3	(4.2)	0.18	(0.68)	0.0076	(0.0011)	0.0080	(0.0013)
a_1^Q	1.6	(2.0)	0.75	(0.41)	0.76	(0.25)	1.6	(1.0)	1.3	(1.1)	0.5	(1.1)	0.5	(1.1)
a_2^Q	−3.3	(5.9)	−92.7	(338.4)	5.1	(9.8)	0.50	(0.54)	0.50	(0.62)	0.7	(2.0)	0.7	(1.8)
a_3^Q	0.62	(0.83)	−12.4	(33.6)	−1.4	(28.8)	−0.2E+03	(3.5E+03)	−80.3	(305.8)	0.5	(7.4)	0.5	(8.1)
b_{11}^Q	0.9	(1.2)	−0.0054	(0.0044)	−0.0055	(0.0033)	−0.715	(0.033)	−0.703	(0.037)	−0.105	(0.047)	−0.110	(0.049)
b_{12}^Q	6.5	(2.3)	–	–	–	–	0.160	(0.056)	0.167	(0.063)	0.5	(1.3)	0.4	(1.3)
b_{13}^Q	0.13	(0.88)	–	–	–	–	–	–	–	–	0.00	(0.68)	0.00	(0.95)
b_{21}^Q	−3.0	(4.4)	−0.18	(0.11)	−0.5	(1.1)	0.165	(0.100)	0.204	(0.099)	0.163	(0.077)	0.184	(0.077)
b_{22}^Q	−16.8	(10.2)	−7.7	(3.1)	−11.5	(7.9)	−0.045	(0.031)	−0.057	(0.035)	−1.64	(0.47)	−1.31	(0.58)
b_{23}^Q	−1.6	(2.1)	−1.27	(0.74)	−3.2	(66.1)	–	–	–	–	0.43	(0.18)	0.43	(0.32)
b_{31}^Q	0.19	(0.62)	0.130	(0.058)	0.2	(2.7)	28.8	(488.0)	25.7	(96.4)	0.00	(0.27)	0.00	(0.33)
b_{32}^Q	2.3	(2.5)	3.5	(1.3)	3.6	(66.3)	1.4	(23.2)	3.2	(11.9)	5.3	(1.3)	5.3	(1.6)
b_{33}^Q	−0.23	(0.17)	−0.13	(0.25)	0.3	(2.2)	−0.2E+03	(2.9E+03)	−126.7	(478.6)	−3.15	(0.57)	−4.3	(1.1)
σ_1	0.000991	(6.3E−05)	0.000988	(6.2E−05)	0.000987	(6.3E−05)	0.000988	(6.1E−05)	0.000988	(6.1E−05)	0.000993	(6.1E−05)	0.000991	(6.3E−05)
σ_2	0.00171	(0.00023)	0.00171	(0.00023)	0.00171	(0.00023)	0.00171	(0.00022)	0.00171	(0.00022)	0.00172	(0.00022)	0.00171	(0.00022)
σ_3	0.00226	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)

Table 12
Parameter estimates for three-factor models—Dai-Singleton market price of risk, QML likelihood.

This table reports the parameter estimates for the $A_0(3)$, $A_1(3)$, $A_2(3)$, and $A_3(3)$ models with the QML likelihood approximation. For the latter three models, results for both the restricted (i.e., with the parameter restrictions from Table 2 imposed) and unrestricted versions of the model are reported. Standard errors are reported in parentheses. The data used are synthetic US Treasury strips, constructed by the method of McCulloch (1975), from January of 1972 until December of 2002. Yields with maturities of one month, two years, and four years are considered observed without error; additional yields with maturities of six, eight, and 10 years are considered observed with error. See Section 5 for additional detail.

	$A_0(3)$		$A_1(3)$—Restricted		$A_1(3)$—Unrestricted		$A_2(3)$—Restricted		$A_2(3)$—Unrestricted		$A_3(3)$—Restricted		$A_3(3)$—Unrestricted	
Parameter	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.
a_1^M	–	–	0.00	(0.23)	0.00	(0.23)	0.04	(0.21)	0.01	(0.29)	0.00	(0.26)	0.0E+03	(1.6E+03)
a_2^M	–	–	–	–	–	–	0.00	(0.26)	0.00	(0.44)	0.0E+07	(6.0E+07)	0.0E+05	(9.2E+05)
a_3^M	–	–	–	–	–	–	–	–	–	–	0.04	(0.47)	0.0	(305.0)
b_{11}	−0.0168	(0.0015)	0.000	(0.039)	0.000	(0.037)	0.00	(0.14)	0.00	(0.13)	−0.038	(0.039)	−0.046	(0.092)
b_{12}	–	–	–	–	–	–	–	–	0.024	(0.013)	–	–	0.000	(0.011)
b_{13}	–	–	–	–	–	–	–	–	–	–	–	–	0.1	(1.3)
b_{21}	−0.41	(0.16)	–	–	0.0	(7.3)	–	–	0.00	(0.15)	–	–	34.5	(44.5)
b_{22}	−2.99	(0.26)	−0.878	(0.026)	−0.8	(125.9)	−0.039	(0.039)	−0.040	(0.040)	−5.86	(0.40)	−3.42	(0.47)
b_{23}	–	–	–	–	−0.0	(107.4)	–	–	–	–	–	–	410.9	(333.7)
b_{31}	0.65	(0.13)	–	–	0.15	(0.87)	–	–	1.50	(0.15)	–	–	0.075	(0.057)
b_{32}	2.58	(0.15)	2.33	(0.12)	2.5	(107.4)	–	–	0.167	(0.052)	–	–	0.0000	(0.0021)
b_{33}	−0.839	(0.022)	−2.70	(0.24)	−3.0	(125.8)	−5.78	(0.40)	−3.21	(0.28)	−0.81	(0.11)	−0.87	(0.30)
d_0	0.00	(0.18)	0.000	(0.016)	0.000	(0.015)	0.000	(0.016)	0.000	(0.013)	−0.5	(562.7)	−0.4	(9.4)
d_1	0.0048	(0.0012)	0.00274	(0.00055)	0.00092	(0.00029)	0.0173	(0.0014)	0.00360	(0.00054)	0.00279	(0.00059)	0.00059	(0.00069)
d_2	−0.0131	(0.0011)	0.0052	(0.0014)	0.0	(1.3)	0.00282	(0.00060)	0.00090	(0.00030)	0.0001	(0.0043)	7.2E−05	(6.7E−05)
d_3	0.02411	(0.00045)	0.02810	(0.00070)	0.03	(0.28)	0.0392	(0.0014)	0.02702	(0.00063)	0.0141	(0.0014)	0.00252	(0.00069)
λ_1	−0.19	(0.19)	0.001	(0.039)	0.003	(0.037)	0.68	(0.14)	0.76	(0.13)	−0.036	(0.039)	−0.029	(0.045)
λ_2	0.61	(0.24)	−0.12	(0.20)	−0.2	(56.4)	−0.036	(0.039)	−0.037	(0.040)	−0.00	(0.12)	−0.0032	(0.0032)
λ_3	−0.98	(0.23)	−0.97	(0.23)	−1.1	(8.3)	−1.13	(0.28)	−1.19	(0.27)	−0.11	(0.11)	−0.092	(0.097)
σ_1	0.001002	(6.1E−05)	0.000997	(6.1E−05)	0.000997	(6.1E−05)	0.000998	(6.1E−05)	0.000999	(6.2E−05)	0.000994	(6.1E−05)	0.000997	(6.2E−05)
σ_2	0.00172	(0.00021)	0.00172	(0.00022)	0.00172	(0.00022)	0.00173	(0.00021)	0.00172	(0.00022)	0.00172	(0.00021)	0.00172	(0.00022)
σ_3	0.00226	(0.00020)	0.00226	(0.00020)	0.00226	(0.00021)	0.00226	(0.00020)	0.00225	(0.00020)	0.00225	(0.00020)	0.00225	(0.00021)

Table 13
Parameter estimates for three-factor models—Cheridito-Filipović-Kimmel market price of risk, QML likelihood.

This table reports the parameter estimates for the $A_0(3)$, $A_1(3)$, $A_2(3)$, and $A_3(3)$ models with the QML likelihood approximation. For the latter three models, results for both the restricted (i.e., with the parameter restrictions from Table 2 imposed) and unrestricted versions of the model are reported. Standard errors are reported in parentheses. The data used are synthetic US Treasury strips, constructed by the method of McCulloch (1975), from January of 1972 until December of 2002. Yields with maturities of one month, two years, and four years are considered observed without error; additional yields with maturities of six, eight, and 10 years are considered observed with error. See Section 5 for additional detail.

	$A_0(3)$		$A_1(3)$—Restricted		$A_1(3)$—Unrestricted		$A_2(3)$—Restricted		$A_2(3)$—Unrestricted		$A_3(3)$—Restricted		$A_3(3)$—Unrestricted	
Parameter	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.	Estimate	Std. dev.
a_1^M	–	–	0.0	(3.5)	0.0	(3.2)	−0.03	(0.78)	−0.0	(1.9)	0.0	(5.5)	0.0	(4.9)
a_2^M	–	–	–	–	–	–	−0.0	(4.1)	−0.0	(4.6)	−0.00	(0.62)	0.0	(2.3)
a_3^M	–	–	–	–	–	–	–	–	–	–	0.0	(1.8)	0.0	(4.4)
b_{11}	−0.15	(0.12)	−0.16	(0.12)	−0.16	(0.11)	0.00	(0.11)	−0.00	(0.16)	−0.22	(0.13)	−0.48	(0.19)
b_{12}	–	–	–	–	–	–	–	–	0.000	(0.086)	–	–	0.14	(0.80)
b_{13}	–	–	–	–	–	–	–	–	–	–	–	–	1.2	(1.3)
b_{21}	0.00	(0.66)	–	–	−0.10	(0.24)	–	–	0.80	(0.45)	–	–	0.00	(0.10)
b_{22}	−3.79	(0.42)	−3.55	(0.41)	−3.62	(0.42)	−0.19	(0.12)	−0.39	(0.16)	−0.31	(0.12)	−0.49	(0.43)
b_{23}	–	–	–	–	−0.3	(16.4)	–	–	–	–	–	–	0.51	(0.61)
b_{31}	−0.39	(0.24)	–	–	0.04	(0.53)	–	–	0.36	(0.13)	–	–	0.00	(0.20)
b_{32}	0.97	(0.57)	0.54	(0.57)	0.3	(16.4)	–	–	0.069	(0.047)	–	–	1.90	(0.65)
b_{33}	−0.85	(0.24)	0.0003	(0.0041)	−0.63	(0.21)	−2.69	(0.35)	−3.38	(0.37)	−0.00	(0.17)	−3.08	(0.48)
d_0	0.006	(0.031)	−0.0	(7.4)	0.000	(0.026)	−0.0	(37.1)	0.0	(1.6)	−0.000	(0.016)	−0.001	(0.014)
d_1	0.0170	(0.0057)	0.00090	(0.00039)	0.00044	(0.00050)	−0.0	(5.1)	−0.04	(0.56)	0.00061	(0.00043)	0.00041	(0.00055)
d_2	−0.030	(0.013)	−0.0200	(0.0034)	−0.026	(0.092)	−0.00	(0.25)	−0.005	(0.075)	−0.00048	(0.00068)	0.00083	(0.00089)
d_3	0.0080	(0.0048)	0.0202	(0.0013)	0.02	(0.14)	0.3	(34.1)	0.2	(3.0)	0.0075	(0.0011)	0.0102	(0.0019)
a_1^Q	1.6	(1.9)	0.73	(0.40)	0.74	(0.24)	1.5	(1.0)	1.4	(1.1)	0.5	(1.1)	0.50	(0.93)
a_2^Q	−2.8	(5.1)	−68.0	(411.6)	4.4	(8.3)	0.50	(0.53)	0.50	(0.58)	0.7	(1.9)	0.6	(1.4)
a_3^Q	0.46	(0.65)	−12.5	(56.2)	−1.2	(24.2)	−0.0E+04	(2.3E+04)	−0.1E+03	(1.0E+03)	0.5	(6.6)	0.5	(6.7)
b_{11}^Q	0.9	(1.2)	−0.0051	(0.0045)	−0.0052	(0.0034)	−0.715	(0.033)	−0.708	(0.038)	−0.103	(0.047)	−0.116	(0.048)
b_{12}^Q	7.4	(2.4)	–	–	–	–	0.162	(0.057)	0.174	(0.067)	0.6	(1.1)	0.4	(1.1)
b_{13}^Q	0.10	(0.91)	–	–	–	–	–	–	–	–	0.00	(0.52)	0.0	(1.2)
b_{21}^Q	−2.7	(4.2)	−0.159	(0.093)	−0.5	(1.1)	0.162	(0.099)	0.178	(0.099)	0.155	(0.074)	0.208	(0.068)
b_{22}^Q	−16.6	(9.3)	−7.7	(3.0)	−11.4	(8.1)	−0.045	(0.031)	−0.052	(0.036)	−1.72	(0.47)	−0.77	(0.66)
b_{23}^Q	−1.3	(1.9)	−1.10	(0.64)	−3.1	(64.3)	–	–	–	–	0.41	(0.16)	0.10	(0.63)
b_{31}^Q	0.07	(0.53)	0.129	(0.057)	0.2	(2.6)	0.0E+03	(3.2E+03)	28.4	(346.7)	0.00	(0.24)	0.00	(0.32)
b_{32}^Q	1.9	(2.6)	3.9	(1.4)	3.7	(64.5)	1.3	(151.5)	3.9	(46.3)	5.6	(1.4)	4.3	(1.5)
b_{33}^Q	−0.29	(0.11)	−0.16	(0.25)	0.3	(3.3)	−0.0E+04	(2.1E+04)	−0.1E+03	(1.8E+03)	−2.98	(0.53)	−5.0	(1.3)
σ_1	0.000991	(6.3E−05)	0.000988	(6.2E−05)	0.000987	(6.3E−05)	0.000987	(6.1E−05)	0.000987	(6.1E−05)	0.000993	(6.2E−05)	0.000990	(6.3E−05)
σ_2	0.00171	(0.00023)	0.00171	(0.00023)	0.00171	(0.00023)	0.00171	(0.00022)	0.00171	(0.00022)	0.00172	(0.00022)	0.00171	(0.00023)
σ_3	0.00226	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)	0.00225	(0.00021)

numeric stability problem in evaluating the likelihood function.[13]

A few common features are evident. The estimates for the standard deviations of the observation errors across different models, market price of risk specifications, and likelihood method, are quite close, suggesting that neither the choice of M for an $A_M(3)$ model, nor choice of the market price of risk specification, has a strong influence on the implied cross-sectional shape of the yield curve. However, the implied time series behavior of the yield curve may be very different across models and market price of risk specification. Note, for example, that some combinations of model and market price of risk specification have a very slowly mean-reverting (or non-mean-reverting) component, with an eigenvalue of the B matrix very close to zero; other combinations do not. This particular feature of the estimates, as well as some others, does vary somewhat across likelihood methods, as well as across models and market price of risk specifications.

However, rather than focusing on the many individual parameter estimates across a wide variety of model specifications and estimation methods, we examine likelihood ratio tests for groups of parameters. Specifically, we test the significance of the Cheridito-Filipović-Kimmel market price of risk specification, relative to the Dai-Singleton specification, both with and without the parameter restrictions of Table 2. We also test for the significance of the unrestricted models, relative to those with the restrictions of Table 2 imposed, under both market price of risk specifications. These likelihood ratio tests are presented in Table 14, and are performed using the 2-term approximate likelihood, the Euler and QML approximate likelihoods, and the true likelihood (when known), allowing comparison of the likelihood methods, in addition to the empirical results. As previously discussed, numeric instability occurs when calculating the true likelihood of the restricted $A_3(3)$ model with the Dai-Singleton market price of risk at the estimated parameter vector; likelihood ratio tests involving this model are therefore not presented.

As shown, the Cheridito-Filipović-Kimmel specification is significant at the 95% level for every model, both restricted and unrestricted, except the $A_0(1)$ model, and often at much higher levels of significance. Our method, using the 2-term reducible likelihood approximation, closely reproduces the true likelihood ratios for all cases where the true likelihood is known, providing further confirmation of the accuracy of our approach. The likelihood ratios using the Euler and QML approximations are typically not as close to the true likelihood ratios as our methods.

The parameter restrictions in Table 2 are statistically significant in every case. No comparison of the various likelihood approximation methods to the true likelihood is possible for these tests, although we note that the 2-term approximate likelihood ratio and QML ratio are usually closer to each other than to the Euler likelihood ratio. In all cases, the parameter restrictions are much less significant under the Cheridito-Filipović-Kimmel market price of risk than under the Dai-Singleton market price of risk. Similarly, the more general market price of risk is much less significant, in all cases, when the parameter restrictions of Table 2 are not imposed. Together, these results suggest that the additional parameters of the more general market price of risk and the additional parameters included when the restrictions of Table 2 are relaxed are, to some extent, substitutes; when one group of parameters is included, the other is much less statistically significant.

Tests of nested models (for example, an $A_M(N)$ model with the Cheridito-Filipović-Kimmel market price of risk specification, relative to the same model with the Dai-Singleton model) are possible, if perhaps more difficult, to perform when the estimation is by method of moments or other techniques. However, the use of maximum likelihood estimation allows us to perform likelihood ratio tests of non-nested models, as described in Vuong (1989). Such tests would be difficult or impossible with estimation methods other than maximum likelihood. We perform tests of $A_M(N)$ models with the same N but different M, using both the Dai-Singleton and Cheridito-Filipović-Kimmel market price of risk specifications, and with and without the parameter restrictions of Table 2, in Table 15. As shown, our method (the reducible MLE) closely approximates the results for the true MLE, for all cases in which the true likelihood ratio statistic can be calculated. (As discussed above, there is numeric instability in calculating the true likelihood of the restricted $A_3(3)$ model with the Dai-Singleton market price of risk.) With the Dai-Singleton market price of risk and the parameter restrictions of Table 2 imposed, smaller values of M are usually preferred to larger values of M; the only exceptions are that the $A_1(1)$ model is preferred to the $A_0(1)$ model, and the $A_3(3)$ model is preferred to the $A_2(3)$. The Euler and QML likelihood ratios are sometimes quite different from the true likelihood ratio, although not enough to reverse the model preference. Dai and Singleton (2000) discussed choice of non-nested three-factor models, and argued that the $A_1(3)$ model provided the best fit, but this conclusion was based on ad hoc model comparison criteria. Here, we reach a rather different conclusion, using a rigorous statistical test, made possible by the use of maximum likelihood estimation (although it should be noted that those authors use a different data set and a different estimation method). Some of the model preferences are statistically significant at the 95% level, and some are not.

From the analogous results for the Cheridito-Filipović-Kimmel market price of risk specification, the preferred model is always the model with the higher M, although the preference is not always statistically significant. As with the previous set of results, the likelihood ratio statistic using our method closely approximates the true likelihood ratio statistic. The Euler and QML likelihood

[13] The search algorithm attempts to explore a region of the parameter space in which the Matlab function for the modified Bessel function of the first kind reports an overflow. The same type of overflow occurs at the parameter vectors estimated using the approximation methods, suggesting that the true MLE is in a region in which the true likelihood cannot be evaluated accurately.

Table 14
Likelihood ratio tests for market price of risk specifications and parameter restrictions.

This table reports likelihood ratio statistics for nested versions of each of the nine basic models. The third through sixth columns show likelihood ratio statistics for the Cheridito, Filipović, and Kimmel (2007) (CFK) market price of risk specification, relative to the Dai and Singleton (2000) (DS) specification, at the estimated parameter vector for both the restricted (i.e., with the parameter restrictions of Table 2 imposed) and unrestricted versions of each model. The seventh through 10th columns show likelihood ratio statistics for the restricted version of each model, relative to the unrestricted version. For comparisons involving two restricted models, the likelihood ratio is calculated using the true likelihood function, and the two-term reducible, Euler, and QML approximate likelihoods. For comparisons in which at least one of the models is unrestricted, the true likelihood is not known, and therefore not used. Comparisons involving the restricted $A_3(3)$ model with the DS market price of risk are not made using the true likelihood, due to numeric instability in calculating the modified Bessel function at the estimated parameter vector. The results using QML for the $A_0(1)$, $A_0(2)$, and $A_0(3)$ models are not presented, since for these models, QML coincides with true MLE. For these models and the $A_1(1)$ model, the restricted and unrestricted models coincide, so no comparison is made. As shown, the CFK market price of risk specification is statistically significant, for each feasible likelihood method for both restricted and unrestricted models, in all cases except the $A_0(1)$ model. The difference between the restricted and unrestricted models is also statistically significant, using every feasible likelihood method under both market price of risk specifications. As shown, the two-term reducible likelihood approximations reproduce the results of the true likelihoods (when known) more closely, and often much more closely, than either the Euler or the QML likelihoods.

		Likelihood ratio tests Market price of risk specifications				Likelihood ratio tests Parameter restrictions			
Model	Likelihood	Restricted model	Unrestricted model	Degrees of freedom	Cutoff value	DS	CFK	Degrees of freedom	Cutoff value
$A_0(1)$	MLE	2.45	–	1	3.84	–	–	–	–
	Reducible	2.45	2.45						
	Euler	2.44	2.44						
$A_1(1)$	MLE	7.82	–	1	3.84	–	–	–	–
	Reducible	7.83	7.83						
	Euler	10.69	10.69						
	QML	10.55	10.55						
$A_0(2)$	MLE	20.65	–	4	9.49	–	–	–	–
	Reducible	20.57	20.57						
	Euler	21.52	21.52						
$A_1(2)$	MLE	182.68	–	3	7.82	–	–	1	3.84
	Reducible	182.72	10.75			183.41	11.43		
	Euler	181.02	11.34			181.10	11.42		
	QML	181.37	10.77			182.09	11.49		
$A_2(2)$	MLE	636.55	–	4	9.49	–	–	2	5.99
	Reducible	636.95	278.03			373.92	15.00		
	Euler	700.79	293.38			420.42	13.01		
	QML	685.17	297.53			402.47	14.83		
$A_0(3)$	MLE	54.84	–	9	16.92	–	–	–	–
	Reducible	55.01	55.01						
	Euler	64.42	64.42						
$A_1(3)$	MLE	74.89	–	7	14.07	–	–	2	5.99
	Reducible	74.97	45.59			42.44	13.06		
	Euler	81.19	54.75			39.44	13.00		
	QML	74.77	45.14			42.64	13.01		
$A_2(3)$	MLE	326.02	–	7	14.07	–	–	4	9.49
	Reducible	328.25	158.56			196.63	26.95		
	Euler	369.65	182.46			214.14	26.94		
	QML	338.50	166.57			199.18	27.24		
$A_3(3)$	MLE	N/A	–	9	16.92	–	–	6	12.59
	Reducible	299.87	132.94			221.27	54.34		
	Euler	341.17	160.22			234.74	53.79		
	QML	311.84	153.27			219.69	61.12		

ratios are generally less accurate. Considering the three-factor models, the $A_1(3)$ model is preferred to the $A_0(3)$ model, and the $A_3(3)$ model is preferred to the $A_2(3)$ model, neither preference is statistically significant. By contrast, comparisons between these two groups of models are all statistically significant at the 95% level.

For the unrestricted models, under the Dai-Singleton market price of risk, whichever $A_M(N)$ model has the higher M value is preferred when comparing single- or three-factor models; for two-factor models, the preference is usually for smaller M. Some results are statistically significant, and some are not; no comparison with the true likelihood ratio is possible. When the Cheridito-Filipović-Kimmel market price of risk is used, the preference is always for higher values of M, and is usually statistically significant.

Perhaps surprisingly, the QML likelihood ratios are sometimes further from the true ratios (when known), or from ratios using our approximate likelihood, than the Euler likelihood ratios are. A possible explanation for this

Table 15
Non-nested likelihood ratio tests.

This table reports likelihood ratio statistics comparing non-nested models (with the same number of factors). For each pair of models, both the Dai and Singleton (2000) (DS) and Cheridito, Filipović, and Kimmel (2007) (CFK) market price of risk specification are used. For the restricted models (i.e., those satisfying the restrictions of Table 2), the likelihood ratio is calculated using the exact likelihood function, and the two-term reducible, Euler, and QML approximations. For the unrestricted models, the true likelihood is not known, and therefore not used. The comparison is not made for the restricted $A_3(3)$ model with the DS market price of risk using the true likelihood function, due to numeric instability in calculating the modified Bessel function at the estimated parameter value. As shown, the approximate likelihood closely reproduces the results obtained using the true likelihood (when known), sometimes much more closely than the Euler and QML likelihoods.

Model		Restricted model				Unrestricted model			
		DS		CFK		DS		CFK	
comparison	Likelihood	Statistic	Preferred model	Statistic	Preferred model	Statistic	Preferred model	Statistic	Preferred model
$A_0(1)$ vs. $A_1(1)$	MLE	−6.78	$A_1(1)$	−7.68	$A_1(1)$	–	–	–	–
	Reducible	−6.78	$A_1(1)$	−7.68	$A_1(1)$	−6.78	$A_1(1)$	−7.68	$A_1(1)$
	Euler	−6.62	$A_1(1)$	−7.91	$A_1(1)$	−6.62	$A_1(1)$	−7.91	$A_1(1)$
	QML	−6.62	$A_1(1)$	−7.91	$A_1(1)$	−6.62	$A_1(1)$	−7.91	$A_1(1)$
$A_0(2)$ vs. $A_1(2)$	MLE	2.84	$A_0(2)$	−1.33	$A_1(2)$	–	–	–	–
	Reducible	2.85	$A_0(2)$	−1.34	$A_1(2)$	−2.76	$A_1(2)$	−2.11	$A_1(2)$
	Euler	2.81	$A_0(2)$	−1.15	$A_1(2)$	−2.65	$A_1(2)$	−1.88	$A_1(2)$
	QML	2.82	$A_0(2)$	−1.17	$A_1(2)$	−2.67	$A_1(2)$	−1.92	$A_1(2)$
$A_0(2)$ vs. $A_2(2)$	MLE	8.14	$A_0(2)$	−4.97	$A_2(2)$	–	–	–	–
	Reducible	8.16	$A_0(2)$	−4.98	$A_2(2)$	1.69	$A_0(2)$	−5.99	$A_2(2)$
	Euler	8.16	$A_0(2)$	−5.14	$A_2(2)$	1.83	$A_0(2)$	−6.07	$A_2(2)$
	QML	7.90	$A_0(2)$	−5.18	$A_2(2)$	1.86	$A_0(2)$	−6.09	$A_2(2)$
$A_1(2)$ vs. $A_2(2)$	MLE	9.09	$A_1(2)$	−6.68	$A_2(2)$	–	–	–	–
	Reducible	9.12	$A_1(2)$	−6.68	$A_2(2)$	3.17	$A_1(2)$	−6.56	$A_2(2)$
	Euler	9.16	$A_1(2)$	−7.21	$A_2(2)$	3.16	$A_1(2)$	−7.11	$A_2(2)$
	QML	8.86	$A_1(2)$	−7.20	$A_2(2)$	3.18	$A_1(2)$	−7.08	$A_2(2)$
$A_0(3)$ vs. $A_1(3)$	MLE	0.66	$A_0(3)$	−0.31	$A_1(3)$	–	–	–	–
	Reducible	0.65	$A_0(3)$	−0.32	$A_1(3)$	−2.64	$A_1(3)$	−1.59	$A_1(3)$
	Euler	0.59	$A_0(3)$	−0.24	$A_1(3)$	−2.60	$A_1(3)$	−1.48	$A_1(3)$
	QML	0.72	$A_0(3)$	−0.24	$A_1(3)$	−2.59	$A_1(3)$	−1.47	$A_1(3)$
$A_0(3)$ vs. $A_2(3)$	MLE	2.85	$A_0(3)$	−3.35	$A_2(3)$	–	–	–	–
	Reducible	2.88	$A_0(3)$	−3.37	$A_2(3)$	−0.85	$A_2(3)$	−4.73	$A_2(3)$
	Euler	3.02	$A_0(3)$	−3.29	$A_2(3)$	−0.77	$A_2(3)$	−4.47	$A_2(3)$
	QML	2.75	$A_0(3)$	−3.29	$A_2(3)$	−0.91	$A_2(3)$	−4.49	$A_2(3)$
$A_0(3)$ vs. $A_3(3)$	MLE	N/A	–	-4.18	$A_3(3)$	–	–	–	–
	Reducible	1.83	$A_0(3)$	-4.22	$A_3(3)$	−4.09	$A_3(3)$	−6.54	$A_3(3)$
	Euler	1.90	$A_0(3)$	-4.55	$A_3(3)$	−3.71	$A_3(3)$	−6.52	$A_3(3)$
	QML	1.61	$A_0(3)$	-4.55	$A_3(3)$	−3.71	$A_3(3)$	−6.53	$A_3(3)$
$A_1(3)$ vs. $A_2(3)$	MLE	2.96	$A_1(3)$	−2.78	$A_2(3)$	–	–	–	–
	Reducible	3.01	$A_1(3)$	−2.79	$A_2(3)$	−0.16	$A_2(3)$	−4.31	$A_2(3)$
	Euler	3.16	$A_1(3)$	−2.72	$A_2(3)$	−0.10	$A_2(3)$	−4.13	$A_2(3)$
	QML	2.83	$A_1(3)$	−2.71	$A_2(3)$	−0.26	$A_2(3)$	−4.14	$A_2(3)$
$A_1(3)$ vs. $A_3(3)$	MLE	N/A	–	-3.98	$A_3(3)$	–	–	–	–
	Reducible	1.81	$A_1(3)$	-4.01	$A_3(3)$	−3.28	$A_3(3)$	−6.19	$A_3(3)$
	Euler	1.91	$A_1(3)$	-4.12	$A_3(3)$	−2.93	$A_3(3)$	−6.29	$A_3(3)$
	QML	1.54	$A_1(3)$	-4.12	$A_3(3)$	−2.95	$A_3(3)$	−6.32	$A_3(3)$
$A_2(3)$ vs. $A_3(3)$	MLE	N/A	–	-1.18	$A_3(3)$	–	–	–	–
	Reducible	−4.35	$A_3(3)$	-1.19	$A_3(3)$	−4.62	$A_3(3)$	−3.44	$A_3(3)$
	Euler	−4.22	$A_3(3)$	-1.48	$A_3(3)$	−4.46	$A_3(3)$	−3.69	$A_3(3)$
	QML	−4.22	$A_3(3)$	-1.51	$A_3(3)$	−4.45	$A_3(3)$	−4.17	$A_3(3)$

phenomenon is that there are two sources of error in the Euler likelihood; a Gaussian likelihood is used in place of the true density, and approximate moments are used in place of the true first and second moments. With the QML likelihood, only the first source of error is present. Although the two types of errors in the Euler likelihood may sometimes reinforce each other, it is possible that sometimes they will pull in opposite directions, with the effect that the combined error of the two sources may be smaller in magnitude than the first source alone. To take an example from Table 15, in the non-nested likelihood ratio tests comparing the unrestricted versions of the $A_2(3)$ and $A_3(3)$ models, using the Cheridito-Filipović-Kimmel market price of risk, the Euler results reproduce the reducible likelihood results more closely than the QML results do. Although

comparison to the true likelihood results in this case is not possible, the reducible likelihoods have been found to be quite accurate in all cases where a comparison is possible. The actual QML log-likelihood (not reported—only differences are presented in the table) for the $A_2(3)$ model in this case is quite close to the reducible likelihood for the same model; most of the discrepancy in the likelihood ratio test comes from the difference between the reducible and QML likelihoods for the $A_3(3)$ model. Tables 9 and 13 present that the QML parameter estimates for this model are often further from the reducible estimates than the Euler estimates are, suggesting that the discrepancy (across methods) in the likelihood ratio has led to discrepancy in the parameter estimates as well.

In summary, for all results in which comparison is possible, our method closely reproduces results obtained using the true likelihood function. The Euler and QML likelihood approximations are almost always less accurate; sometimes the reduction in accuracy is small, but in other cases, the loss of accuracy is quite noticeable.

6. Conclusions

We have developed and implemented a technique for maximum likelihood estimation of affine yield models, implemented this technique for several families of such models, and performed likelihood ratio tests, both nested and non-nested, on various pairs of models. In those cases where the likelihood function for a model is known in closed-form, we find through simulations that estimates obtained through our technique are very close to the true maximum-likelihood estimates, and substantially more accurate than alternative methods, such as Euler or QML approximation. Our technique, which applies to all affine yield models (including those for which the likelihood function is not known in closed-form), therefore promises to be an accurate and computationally efficient estimation method. The bias and variance introduced by using an approximation to the likelihood function, rather than the true likelihood function, are trivial compared to the bias and variance of the true maximum likelihood estimator itself; alternate approximation methods, such as Euler or QML, generate much larger approximation errors. Not only do we produce maximum-likelihood estimates (as opposed to second-best solutions such as GMM or other estimators), but we do so at a small computational cost given the closed-form nature of our formulae. Furthermore, we use the likelihood statistics generated by the estimation process to perform rigorous statistical tests of both nested and non-nested pairs of models, comparing models with different market price of risk specifications, models with and without parameter restrictions needed to ensure existence of a closed-form likelihood, and models with different numbers of bounded state variables. For the cases where the true likelihood is available, we find likelihood ratio tests performed using our method reproduce the true results more accurately, and often much more accurately, than Euler or QML likelihood approximations.

Although we have focused exclusively on affine models, much broader application of our method is possible. For example, our technique can be applied to models that are affine under the risk-neutral probability measure, but non-affine under true probabilities. The affine property is useful only for pricing; it is irrelevant as far as deriving closed-form likelihood expansions, which are available for unconstrained multivariate diffusions. Such models have been proposed by several papers, but have been estimated only in restricted special cases. For other implementations of our method in various contexts and with various data sets, see Bakshi, Ju, and Ou-Yang (2006), Cheridito, Filipović, and Kimmel (2007), Mosburger and Schneider (2005), Thompson (2008), and Egorov, Li, and Ng (2008).

Appendix A. Families of admissible affine diffusions

Several practical issues arise when we study affine yield models. First, as discussed in Duffie and Kan (1996), existence considerations impose constraints on the coefficients of both the drift vector and diffusion matrix. Furthermore, there will typically be infinitely many model specifications that produce identical interest rate dynamics. Dai and Singleton (2000) consider these issues, and, for affine yield models with N state variables, specify $N+1$ non-nested canonical models that very nearly achieve three goals: (1) each canonical model satisfies all existence and uniqueness requirements, (2) each affine yield model is observationally equivalent to a canonical model, and (3) each canonical model is observationally different from all others. As we show in this appendix, neither of the last two goals is completely achieved, although Dai and Singleton (2000) come very close. We detail in Table 1 the parameter restrictions corresponding to the various models.

Each affine diffusion can be assigned to a family $A_M(N)$, in which N is the number of state variables and M is the number of independent linear combinations of those state variables that appear in the diffusion matrix. The vector of state variables is premultiplied by a non-singular matrix of constants; the result is taken to be a new state vector. If the diffusion is affine in the old state vector, the diffusion followed by the alternate state vector is also affine, and by judicious choice of the matrix of constants, also corresponds to one of the canonical models.

Considering affine yield models with one, two, or three state variables, there are a total of nine observationally distinct canonical models, not counting the trivial zero-factor model with a constant interest rate. The likelihood function for each of the nine models is different, so we discuss each model in turn. The likelihood function is known in closed-form for four of the nine canonical models (as well as for special cases of the other five). Those models for which a closed-form likelihood function is known provide useful test cases for evaluating our estimation technique.

A.1. One-factor models

In single-factor affine yield models, the interest rate is a linear function of a single state variable

$$r_t = \delta_0 + \delta X_{1t}. \tag{A.1}$$

The dynamics of the state variable (under the physical measure P) may take one of two distinct forms. In the $A_0(1)$ model, we have

$$dX_{1t} = b_{11} X_{1t}\, dt + dW^P_{1t}. \tag{A.2}$$

This model is an Ornstein-Uhlenbeck process, corresponding to the model of Vasicek (1977), and has a Gaussian transition function.

The $A_1(1)$ model has the dynamics

$$dX_{1t} = (a_1 + b_{11} X_{1t})\, dt + \sqrt{X_{1t}}\, dW^P_{1t}. \tag{A.3}$$

When $\delta_0 = 0$, the $A_1(1)$ model reduces to Feller's square-root model, corresponding to the model of Cox, Ingersoll, and Ross (1985), and the transition density of the state variable is non-central chi-squared. When $\delta_0 \neq 0$, the transition function readily follows by a simple change of variable. The likelihood function is therefore known for all single-factor affine yield models.

Under the risk-neutral probability measure Q, the dynamics of the state variables of the $A_0(1)$ and $A_1(1)$ model are, respectively,

$$dX_{1t} = [-\lambda_1 + b_{11} X_{1t}]\, dt + dW^Q_{1t}, \tag{A.4}$$

$$dX_{1t} = [a_1 + (b_{11} - \lambda_1) X_{1t}]\, dt + \sqrt{X_{1t}}\, dW^Q_{1t}. \tag{A.5}$$

A.2. Two-factor models

There are three families of two-factor affine yield models. In all three the interest rate is specified as

$$r_t = \delta_0 + \delta_1 X_{1t} + \delta_2 X_{2t}. \tag{A.6}$$

In the $A_0(2)$ family, the dynamics of the state variables are (under the physical measure P)

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + d\begin{bmatrix} W^P_{1t} \\ W^P_{2t} \end{bmatrix}. \tag{A.7}$$

The transition function for this type of diffusion is known in closed-form, and is bivariate Gaussian.

The $A_1(2)$ model has dynamics under P

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ 0 \end{bmatrix} + \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{1 + \beta_{21} X_{1t}} \end{bmatrix} d\begin{bmatrix} W^P_{1t} \\ W^P_{2t} \end{bmatrix}. \tag{A.8}$$

In general, the likelihood function for this type of diffusion is not known in closed-form; however, if we impose the constraints $b_{21}=0$ and $\beta_{21} = 0$, then the two state variables are independent, and their joint transition density is the product of the two marginal transition densities, which are Gaussian and non-central chi-squared, respectively (after translation of the first state variable by a constant).

The $A_2(2)$ model has the representation

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{2t}} \end{bmatrix} d\begin{bmatrix} W^P_{1t} \\ W^P_{2t} \end{bmatrix}. \tag{A.9}$$

The transition density of this type of diffusion is known only if $b_{12}=0$ and $b_{21}=0$, in which case the two state variables are (after translation) independent non-central chi-squared random variables.

The three canonical specifications are as presented in Dai and Singleton (2000), apart from our modified parameterization of the drift; however, there are at least two types of two-variable affine diffusions that are not observationally equivalent to any of the three canonical models. An example of the first type is

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} dt + d\begin{bmatrix} W^P_{1t} \\ W^P_{2t} \end{bmatrix}, \tag{A.10}$$

with the constraint $(b_{11} - b_{22})^2 < 4 b_{12} b_{21}$. This diffusion shares many properties of the $A_0(2)$ model (the transition density is bivariate Gaussian, both state variables are unbounded, etc.), but cannot be expressed in the $A_0(2)$ canonical form unless

we allow b_{11} and b_{22} to be complex conjugate pairs. An example of the second type of non-conforming diffusion is

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ 0 \end{bmatrix} + \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{1t}} \end{bmatrix} d\begin{bmatrix} W_{1t}^P \\ W_{2t}^P \end{bmatrix}. \tag{A.11}$$

This diffusion most closely resembles the $A_1(2)$ canonical form, but no change of variables can generate the non-zero constant coefficient in the diffusion term of the second state variable in the $A_1(2)$ model.

In all but a few special cases, each canonical model is observationally unique. In the $A_2(2)$ model, the two state variables can switch places; in the $A_0(2)$, model, there are infinitely many representations of observationally equivalent models for some restricted values of the B matrix.

Under the risk-neutral probability measure Q, the dynamics of the state vector in the three models are, respectively,

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(-\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} + \begin{bmatrix} b_{11} & 0 \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + d\begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \end{bmatrix}, \tag{A.12}$$

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ -\lambda_2 \end{bmatrix} + \begin{bmatrix} b_{11}-\lambda_1 & 0 \\ b_{21}-\lambda_2\beta_{21} & b_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{1+\beta_{21}X_{1t}} \end{bmatrix} d\begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \end{bmatrix}, \tag{A.13}$$

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} b_{11}-\lambda_1 & b_{12} \\ b_{21} & b_{22}-\lambda_2 \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{1t}} \end{bmatrix} d\begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \end{bmatrix}. \tag{A.14}$$

A.3. Three-factor models

In three-factor affine yield models, the instantaneous interest rate is defined as

$$r_t = \delta_0 + \delta_1 X_{1t} + \delta_2 X_{2t} + \delta_3 X_{3t}. \tag{A.15}$$

The first of the four canonical three-factor models is the $A_0(3)$ family, in which the state variables have the following dynamics:

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} dt + d\begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}. \tag{A.16}$$

The transition density of the state vector is trivariate Gaussian.

The $A_1(3)$ model has the following dynamics:

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} b_{11} & 0 & 0 \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{1+\beta_{21}X_{1t}} & 0 \\ 0 & 0 & \sqrt{1+\beta_{31}X_{1t}} \end{bmatrix} d\begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}. \tag{A.17}$$

The transition density function is known in closed-form only if the first state variable is independent of the other two, i.e., if $b_{21}=0$, $b_{31}=0$, $\beta_{21}=0$, and $\beta_{31}=0$. In this case, the joint transition density is the product of a non-central chi-squared (the distribution of the first state variable, after translation) and a bivariate Gaussian (the distribution of the other two).

In the $A_2(3)$ model, the state vector has the following dynamics:

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ a_2 \\ 0 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{1+\beta_{31}X_{1t}+\beta_{32}X_{2t}} \end{bmatrix} d\begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}. \tag{A.18}$$

The transition density is known in closed-form only if the three state variables are independent of each other, i.e., if $b_{12}=b_{21}=b_{31}=b_{32}=\beta_{31}=\beta_{32}=0$. In this case, the density is the product of two non-central chi-squared densities and a Gaussian density (after translation of the first two state variables).

Finally, in the $A_3(3)$ model, the dynamics are:

$$d\begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \left(\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{X_{3t}} \end{bmatrix} d\begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}. \tag{A.19}$$

The transition density is known in closed-form only if the three state variables are independent of each other, i.e., if $b_{12}=b_{13}=b_{21}=b_{23}=b_{31}=b_{32}=0$. In this case, the density is the product of three independent non-central chi-squared densities (after translation of the state variables).

As in the two-factor case, there are two types of affine diffusions with three state variables that are not observationally equivalent to any of the canonical models. The model

$$d\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}=\begin{bmatrix}b_{11}&b_{12}&b_{13}\\b_{21}&b_{22}&b_{23}\\b_{31}&b_{32}&b_{33}\end{bmatrix}\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}dt+d\begin{bmatrix}W_{1t}^P\\W_{2t}^P\\W_{3t}^P\end{bmatrix} \tag{A.20}$$

is similar to the $A_0(3)$ canonical model, but there is no change of variables that results in the $A_0(3)$ model if any two eigenvalues of the B matrix are complex conjugate pairs. Similarly, the model

$$d\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}=\left(\begin{bmatrix}a_1\\0\\0\end{bmatrix}+\begin{bmatrix}b_{11}&0&0\\b_{21}&b_{22}&b_{23}\\b_{31}&b_{32}&b_{33}\end{bmatrix}\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}\right)dt+\begin{bmatrix}\sqrt{X_{1t}}&0&0\\0&\sqrt{X_{1t}}&0\\0&0&\sqrt{1+\beta_{31}X_{1t}}\end{bmatrix}d\begin{bmatrix}W_{1t}^P\\W_{2t}^P\\W_{3t}^P\end{bmatrix} \tag{A.21}$$

resembles the $A_1(3)$ model, but the absence of a constant coefficient in the diffusion of the second state variable makes it impossible to convert this model into the $A_1(3)$ model by a change of variables. Similar variants of the $A_2(3)$ model exist.

Under the risk-neutral probability measure Q, the dynamics of the state vector in the four models are, respectively,

$$d\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}=\left(\begin{bmatrix}-\lambda_1\\-\lambda_2\\-\lambda_3\end{bmatrix}+\begin{bmatrix}b_{11}&0&0\\b_{21}&b_{22}&0\\b_{31}&b_{32}&b_{33}\end{bmatrix}\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}\right)dt+d\begin{bmatrix}W_{1t}^Q\\W_{2t}^Q\\W_{3t}^Q\end{bmatrix}, \tag{A.22}$$

$$d\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}=\left(\begin{bmatrix}a_1\\-\lambda_2\\-\lambda_3\end{bmatrix}+\begin{bmatrix}b_{11}-\lambda_1&0&0\\b_{21}-\lambda_2\beta_{21}&b_{22}&b_{23}\\b_{31}-\lambda_3\beta_{31}&b_{32}&b_{33}\end{bmatrix}\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}\right)dt+\begin{bmatrix}\sqrt{X_{1t}}&0&0\\0&\sqrt{1+\beta_{21}X_{1t}}&0\\0&0&\sqrt{1+\beta_{31}X_{1t}}\end{bmatrix}d\begin{bmatrix}W_{1t}^Q\\W_{2t}^Q\\W_{3t}^Q\end{bmatrix}, \tag{A.23}$$

$$d\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}=\left(\begin{bmatrix}a_1\\a_2\\-\lambda_3\end{bmatrix}+\begin{bmatrix}b_{11}-\lambda_1&b_{12}&0\\b_{21}&b_{22}-\lambda_2&0\\b_{31}-\lambda_3\beta_{31}&b_{32}-\lambda_3\beta_{32}&b_{33}\end{bmatrix}\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}\right)dt+\begin{bmatrix}\sqrt{X_{1t}}&0&0\\0&\sqrt{X_{2t}}&0\\0&0&\sqrt{1+\beta_{31}X_{1t}+\beta_{32}X_{2t}}\end{bmatrix}d\begin{bmatrix}W_{1t}^Q\\W_{2t}^Q\\W_{3t}^Q\end{bmatrix}, \tag{A.24}$$

$$d\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}=\begin{bmatrix}a_1\\a_2\\a_3\end{bmatrix}+\begin{bmatrix}b_{11}-\lambda_1&b_{12}&b_{13}\\b_{21}&b_{22}-\lambda_2&b_{23}\\b_{31}&b_{32}&b_{33}-\lambda_3\end{bmatrix}\begin{bmatrix}X_{1t}\\X_{2t}\\X_{3t}\end{bmatrix}dt+\begin{bmatrix}\sqrt{X_{1t}}&0&0\\0&\sqrt{X_{2t}}&0\\0&0&\sqrt{X_{3t}}\end{bmatrix}d\begin{bmatrix}W_{1t}^Q\\W_{2t}^Q\\W_{3t}^Q\end{bmatrix}. \tag{A.25}$$

Appendix B. Formulae for the log-transition functions

In this section, we give the coefficients of the closed-form expansions for the log-transition functions corresponding to the three two-dimensional models. Expansions for the two univariate models (Vasicek, 1977; Cox, Ingersoll, and Ross, 1985, respectively) can be found in Aït-Sahalia (1999), while the expressions for the four three-dimensional models are not reported here to save space. They are available in computer form from the authors upon request.

B.1. The $A_0(2)$ model

The coefficients below correspond to the stochastic differential equation (SDE) (A.7).

$$C_X^{(-1)}(x|x_0;\theta)=-\frac{(x_1-x_{10})^2}{2}-\frac{(x_2-x_{20})^2}{2},$$

$$C_X^{(0)}(x|x_0;\theta)=\frac{x_1-x_{10}}{2}b_{11}(x_1+x_{10})+\frac{x_2-x_{20}}{2}(b_{21}(x_1+x_{10})+b_{22}(x_2+x_{20})),$$

$$C_X^{(1)}(x|x_0;\theta)=-\frac{b_{11}+b_{22}}{2}-\frac{b_{11}^2}{6}(x_1^2+x_1x_{10}+x_{10}^2)-\frac{b_{22}^2}{6}(x_2^2+x_2x_{20}+x_{20}^2)-\frac{b_{21}^2}{8}(x_1+x_{10})^2+\frac{b_{21}^2}{24}(x_2-x_{20})^2$$
$$-\frac{b_{21}b_{22}}{6}(2x_1x_2+x_1x_{20}+x_{10}x_2+2x_{10}x_{20}),$$

$$C_X^{(2)}(x|x_0;\theta) = -\frac{2b_{11}^2+b_{21}^2+2b_{22}^2}{12} - \frac{b_{11}b_{21}}{12}(x_2-x_{20})b_{11}(x_1+x_{10}) + \left(\frac{b_{21}b_{22}}{12}(x_1-x_{10}), -\frac{b_{21}^2}{12}, (x_2,-,x_{20},)\right)(b_{21}(x_1+x_{10}) + b_{22}(x_2+x_{20})).$$

B.2. The $A_1(2)$ model

The coefficients below correspond to the SDE (A.8).

$$\begin{aligned}
C_X^{(-1)}(x|x_0;\theta) = &-\frac{21(x_1-x_{10})^6}{256x_{10}^5} + \frac{7(x_1-x_{10})^5}{64x_{10}^4} - \frac{5(x_1-x_{10})^4}{32x_{10}^3} + \frac{(x_1-x_{10})^3}{4x_{10}^2} + \frac{x_{10}(x_2-x_{20})^6\beta_{21}^3(3-7x_{10}\beta_{21})}{11520(1+x_{10}\beta_{21})^7} \\
&+ \frac{x_{10}(x_2-x_{20})^4\beta_{21}^2}{96(1+x_{10}\beta_{21})^4} + \frac{(x_1-x_{10})(x_2-x_{20})^2\beta_{21}}{4(1+x_{10}\beta_{21})^2} - \frac{(x_1-x_{10})^2(x_2-x_{20})^2\beta_{21}(1+7x_{10}\beta_{21})}{48x_{10}(1+x_{10}\beta_{21})^3} \\
&- \frac{x_{10}(x_2-x_{20})^2+(x_1-x_{10})^2(1+x_{10}\beta_{21})}{2(x_{10}+x_{10}^2\beta_{21})} + \frac{(x_1-x_{10})^3(x_2-x_{20})^2\beta_{21}(1+4x_{10}\beta_{21}+9x_{10}^2\beta_{21}^2)}{96x_{10}^2(1+x_{10}\beta_{21})^4} \\
&+ \frac{(x_1-x_{10})^2(x_2-x_{20})^4\beta_{21}^2(-7-114x_{10}\beta_{21}+193x_{10}^2\beta_{21}^2)}{11520x_{10}(1+x_{10}\beta_{21})^6} + \frac{(x_1-x_{10})(x_2-x_{20})^4(\beta_{21}^2-3x_{10}\beta_{21}^3)}{192(1+x_{10}\beta_{21})^5} \\
&- \frac{(x_1-x_{10})^4(x_2-x_{20})^2\beta_{21}(15+71x_{10}\beta_{21}+133x_{10}^2\beta_{21}^2+149x_{10}^3\beta_{21}^3)}{2304x_{10}^3(1+x_{10}\beta_{21})^5},
\end{aligned}$$

$$\begin{aligned}
C_X^{(0)}(x|x_0;\theta) = &\frac{(x_2-x_{20})(b_{21}x_{10}+b_{22}x_{20})}{1+x_{10}\beta_{21}} + \frac{(x_1-x_{10})(x_2-x_{20})(b_{21}-b_{22}x_{20}\beta_{21})}{2(1+x_{10}\beta_{21})^2} - \frac{x_{10}(x_2-x_{20})^3\beta_{21}(-b_{21}+b_{22}x_{20}\beta_{21})}{24(1+x_{10}\beta_{21})^4} \\
&+ \frac{(x_1-x_{10})(x_2-x_{20})^3\beta_{21}(-1+3x_{10}\beta_{21})(-b_{21}+b_{22}x_{20}\beta_{21})}{48(1+x_{10}\beta_{21})^5} \\
&+ \frac{(x_1-x_{10})^2(x_2-x_{20})(1+7x_{10}\beta_{21})(-b_{21}+b_{22}x_{20}\beta_{21})}{24x_{10}(1+x_{10}\beta_{21})^3} \\
&+ \frac{(x_1-x_{10})^3(x_2-x_{20})(b_{21}-b_{22}x_{20}\beta_{21})(1+4x_{10}\beta_{21}+9x_{10}^2\beta_{21}^2)}{48x_{10}^2(1+x_{10}\beta_{21})^4} \\
&+ \frac{(x_1-x_{10})(-1+4a_1(1+x_{10}\beta_{21})+4b_{11}x_{10}(1+x_{10}\beta_{21}))}{4x_{10}(1+x_{10}\beta_{21})} \\
&+ \frac{(x_1-x_{10})^2(6+13x_{10}\beta_{21}-2x_{10}^2\beta_{21}^2-24a_1(1+x_{10}\beta_{21})^2)}{48x_{10}^2(1+x_{10}\beta_{21})^2} + \frac{(x_2-x_{20})^2(\beta_{21}-2x_{10}\beta_{21}^2+24b_{22}(1+x_{10}\beta_{21})^2)}{48(1+x_{10}\beta_{21})^3} \\
&- \frac{(x_1-x_{10})(x_2-x_{20})^2\beta_{21}(\beta_{21}(5-4x_{10}\beta_{21})+32b_{22}(1+x_{10}\beta_{21})^2)}{96(1+x_{10}\beta_{21})^4} \\
&- \frac{(x_2-x_{20})^4\beta_{21}^2(-2+61x_{10}\beta_{21}-42x_{10}^2\beta_{21}^2+240b_{22}x_{10}(1+x_{10}\beta_{21})^2)}{11520(1+x_{10}\beta_{21})^6} \\
&+ \frac{(x_1-x_{10})^3(-8-25x_{10}\beta_{21}-27x_{10}^2\beta_{21}^2+4x_{10}^3\beta_{21}^3+32a_1(1+x_{10}\beta_{21})^3)}{96x_{10}^3(1+x_{10}\beta_{21})^3} \\
&+ \frac{(x_1-x_{10})^4(144+591x_{10}\beta_{21}+922x_{10}^2\beta_{21}^2+659x_{10}^3\beta_{21}^3-86x_{10}^4\beta_{21}^4-576a_1(1+x_{10}\beta_{21})^4)}{2304x_{10}^4(1+x_{10}\beta_{21})^4} \\
&+ \frac{(x_1-x_{10})^2(x_2-x_{20})^2\beta_{21}(120b_{22}(1+x_{10}\beta_{21})^2(1+11x_{10}\beta_{21})+\beta_{21}(31+377x_{10}\beta_{21}-194x_{10}^2\beta_{21}^2))}{5760x_{10}(1+x_{10}\beta_{21})^5}
\end{aligned}$$

$$\begin{aligned}
C_X^{(1)}(x|x_0;\theta) = &-(a_1b_{11}) - \frac{5b_{22}}{8} - \frac{7b_{11}^2x_1}{24} - \frac{47x_1^2}{1440x_{10}^3} - \frac{3(-1+a_1)a_1x_1^2}{16x_{10}^3} + \frac{77x_1}{720x_{10}^2} + \frac{5(-1+a_1)a_1x_1}{8x_{10}^2} - \frac{b_{22}x_1^2}{24x_{10}^2} - \frac{227}{1440x_{10}} \\
&- \frac{15(-1+a_1)a_1}{16x_{10}} + \frac{b_{22}x_1}{6x_{10}} + \frac{b_{11}^2x_1^2}{48x_{10}} - \frac{11b_{11}^2x_{10}}{48} - \frac{(x_2-x_{20})^2}{48x_{10}^2(1+x_{10}\beta_{21})^5} + \frac{3x_1^2}{128x_{10}^3(1+x_{10}\beta_{21})^4} \\
&- \frac{3x_1}{64x_{10}^2(1+x_{10}\beta_{21})^4} + \frac{3b_{21}x_1(x_2-x_{20})}{16x_{10}(1+x_{10}\beta_{21})^4} + \frac{49(x_2-x_{20})^2}{768x_{10}^2(1+x_{10}\beta_{21})^4} + \frac{b_{21}^2x_{10}(x_2-x_{20})^2}{16(1+x_{10}\beta_{21})^4} + \frac{3b_{22}x_1(x_2-x_{20})x_{20}}{16x_{10}^2(1+x_{10}\beta_{21})^4} \\
&+ \frac{b_{21}(x_2-x_{20})(-3+2b_{22}(x_2-x_{20})x_{20})}{16(1+x_{10}\beta_{21})^4} + \frac{3+8b_{22}(x_2-x_{20})x_{20}(-3+b_{22}(x_2-x_{20})x_{20})}{128x_{10}(1+x_{10}\beta_{21})^4} - \frac{b_{21}^2x_1^2}{8(1+x_{10}\beta_{21})^3} \\
&- \frac{29x_1^2}{384x_{10}^3(1+x_{10}\beta_{21})^3} + \frac{35x_1}{192x_{10}^2(1+x_{10}\beta_{21})^3} + \frac{b_{21}^2x_1x_{10}}{4(1+x_{10}\beta_{21})^3} - \frac{b_{21}^2x_{10}^2}{8(1+x_{10}\beta_{21})^3} - \frac{13b_{21}x_1x_2}{48x_{10}(1+x_{10}\beta_{21})^3}
\end{aligned}$$

$$
\begin{aligned}
&-\frac{787(x_2-x_{20})^2}{11520x_{10}^2(1+x_{10}\beta_{21})^3}+\frac{b_{21}b_{22}x_1x_{20}}{2(1+x_{10}\beta_{21})^3}+\frac{13b_{21}x_1x_{20}}{48x_{10}(1+x_{10}\beta_{21})^3}-\frac{b_{21}b_{22}x_1^2x_{20}}{4x_{10}(1+x_{10}\beta_{21})^3}-\frac{b_{21}b_{22}x_{10}x_{20}}{4(1+x_{10}\beta_{21})^3}\\
&-\frac{11b_{22}x_1x_2x_{20}}{24x_{10}^2(1+x_{10}\beta_{21})^3}+\frac{11b_{22}x_1x_{20}^2}{24x_{10}^2(1+x_{10}\beta_{21})^3}-\frac{b_{22}^2x_1^2x_{20}^2}{8x_{10}^2(1+x_{10}\beta_{21})^3}+\frac{b_{22}^2x_1x_{20}^2}{4x_{10}(1+x_{10}\beta_{21})^3}\\
&-\frac{6b_{22}^2x_{20}^2+b_{21}(x_2-x_{20})(-19+6b_{22}(x_2-x_{20})x_{20})}{48(1+x_{10}\beta_{21})^3}-\frac{41+8b_{22}(x_2-x_{20})(-3x_2-25x_{20}+6b_{22}x_2x_{20}^2-6b_{22}x_{20}^3)}{384x_{10}(1+x_{10}\beta_{21})^3}\\
&+\frac{b_{21}^2x_1^2}{48(1+x_{10}\beta_{21})^2}+\frac{959x_1^2}{11520x_{10}^3(1+x_{10}\beta_{21})^2}-\frac{1349x_1}{5760x_{10}^2(1+x_{10}\beta_{21})^2}-\frac{b_{22}x_1^2}{24x_{10}^2(1+x_{10}\beta_{21})^2}+\frac{b_{22}x_1}{12x_{10}(1+x_{10}\beta_{21})^2}\\
&-\frac{7b_{21}^2x_1x_{10}}{24(1+x_{10}\beta_{21})^2}+\frac{13b_{21}^2x_{10}^2}{48(1+x_{10}\beta_{21})^2}+\frac{b_{21}x_1x_2}{12x_{10}(1+x_{10}\beta_{21})^2}-\frac{(-71+60(-1+a_1)a_1)(x_2-x_{20})^2}{2880x_{10}^2(1+x_{10}\beta_{21})^2}\\
&-\frac{b_{21}x_1x_{20}}{12x_{10}(1+x_{10}\beta_{21})^2}+\frac{7b_{21}b_{22}x_1^2x_{20}}{24x_{10}(1+x_{10}\beta_{21})^2}+\frac{13b_{22}^2x_1^2x_{20}^2}{48x_{10}^2(1+x_{10}\beta_{21})^2}-\frac{17b_{22}x_1x_{20}(-x_2+x_{20})}{48x_{10}^2(1+x_{10}\beta_{21})^2}\\
&-\frac{b_{21}b_{22}x_1(5x_2+8x_{20})}{12(1+x_{10}\beta_{21})^2}-\frac{b_{22}^2x_1x_{20}(10x_2+9x_{20})}{24x_{10}(1+x_{10}\beta_{21})^2}+\frac{-2b_{22}+(-8b_{21}+b_{11}^2(x_2-x_{20}))(x_2-x_{20})+5b_{22}^2x_{20}(4x_2+x_{20})}{48(1+x_{10}\beta_{21})^2}\\
&-\frac{b_{21}x_{10}(b_{21}(x_2-x_{20})^2-2b_{22}(10x_2+9x_{20}))}{48(1+x_{10}\beta_{21})^2}+\frac{2099+240b_{22}(x_2-x_{20})(-5x_2-22x_{20}+3b_{22}x_2x_{20}^2-3b_{22}x_{20}^3)}{11520x_{10}(1+x_{10}\beta_{21})^2}\\
&+\frac{b_{21}^2x_1^2}{48(1+x_{10}\beta_{21})}-\frac{97x_1^2}{2880x_{10}^3(1+x_{10}\beta_{21})}+\frac{157x_1}{1440x_{10}^2(1+x_{10}\beta_{21})}+\frac{b_{22}x_1^2}{12x_{10}^2(1+x_{10}\beta_{21})}-\frac{b_{22}x_1}{4x_{10}(1+x_{10}\beta_{21})}\\
&-\frac{7b_{21}^2x_1x_{10}}{24(1+x_{10}\beta_{21})}-\frac{11b_{21}^2x_{10}^2}{48(1+x_{10}\beta_{21})}+\frac{b_{21}b_{22}x_1(x_2-x_{20})}{12(1+x_{10}\beta_{21})}+\frac{(1-30a_1+30a_1^2)(x_2-x_{20})^2}{1440x_{10}^2(1+x_{10}\beta_{21})}-\frac{7b_{22}^2x_1^2x_{20}^2}{48x_{10}^2(1+x_{10}\beta_{21})}\\
&+\frac{b_{22}x_1x_{20}(-x_2+x_{20})}{12x_{10}^2(1+x_{10}\beta_{21})}+\frac{b_{22}^2x_1x_{20}(10x_2+3x_{20})}{24x_{10}(1+x_{10}\beta_{21})}-\frac{b_{21}b_{22}x_{10}(7x_2+5x_{20})}{12(1+x_{10}\beta_{21})}-\frac{337-120b_{22}(x_2^2+2x_2x_{20}-3x_{20}^2)}{2880x_{10}(1+x_{10}\beta_{21})}\\
&-\frac{-8b_{22}+b_{11}^2(x_2-x_{20})^2+b_{22}^2(8x_2^2+28x_2x_{20}+7x_{20}^2)}{48(1+x_{10}\beta_{21})},
\end{aligned}
$$

$$
\begin{aligned}
C_X^{(2)}(x|x_0;\theta)=&\frac{-(23-120a_1+120a_1^2+120b_{22}^2x_{10}^2)}{720x_{10}^2}+\frac{1}{96x_{10}^2(1+x_{10}\beta_{21})^4}-\frac{11+48b_{21}^2x_{10}^3+96b_{21}b_{22}x_{10}^2x_{20}+48b_{22}^2x_{10}x_{20}^2}{384x_{10}^2(1+x_{10}\beta_{21})^3}\\
&-\frac{-73+120b_{21}^2x_{10}^3-480b_{21}b_{22}x_{10}^2x_{20}-600b_{22}^2x_{10}x_{20}^2}{2880x_{10}^2(1+x_{10}\beta_{21})^2}+\frac{1-30a_1+30a_1^2-30b_{11}^2x_{10}^2-60b_{22}^2x_{10}x_{20}^2}{720x_{10}^2(1+x_{10}\beta_{21})}.
\end{aligned}
$$

B.3. The $A_2(2)$ model

The coefficients below correspond to the SDE (A.9).

$$
\begin{aligned}
C_X^{(-1)}(x|x_0;\theta)=&\frac{-21(x_1-x_{10})^6}{256x_{10}^5}+\frac{7(x_1-x_{10})^5}{64x_{10}^4}-\frac{5(x_1-x_{10})^4}{32x_{10}^3}+\frac{(x_1-x_{10})^3}{4x_{10}^2}-\frac{(x_1-x_{10})^2}{2x_{10}}-\frac{(x_2-x_{20})^2}{2x_{20}}-\frac{21(x_2-x_{20})^6}{256x_{20}^5}\\
&+\frac{7(x_2-x_{20})^5}{64x_{20}^4}-\frac{5(x_2-x_{20})^4}{32x_{20}^3}+\frac{(x_2-x_{20})^3}{4x_{20}^2}\\
C_X^{(0)}(x|x_0;\theta)=&\frac{(1-4a_2-4b_{21}x_{10})(x_2-x_{20})^4}{16x_{20}^4}+\frac{(-1+4a_2+4b_{21}x_{10})(x_2-x_{20})^3}{12x_{20}^3}+\frac{(1-4a_2-4b_{21}x_{10})(x_2-x_{20})^2}{8x_{20}^2}\\
&+\frac{(x_1-x_{10})^4(1-4a_1-4b_{12}x_{20})}{16x_{10}^4}+\frac{(x_1-x_{10})^2(1-4a_1-4b_{12}x_{20})}{8x_{10}^2}+\frac{(x_1-x_{10})^2(x_2-x_{20})^2(b_{21}x_{10}+b_{12}x_{20})}{48x_{10}^2x_{20}^2}\\
&+\frac{(x_1-x_{10})(x_2-x_{20})(b_{21}x_{10}+b_{12}x_{20})}{2x_{10}x_{20}}-\frac{(x_1-x_{10})(x_2-x_{20})^2(7b_{21}x_{10}+b_{12}x_{20})}{24x_{10}x_{20}^2}\\
&+\frac{(x_1-x_{10})(x_2-x_{20})^3(10b_{21}x_{10}+b_{12}x_{20})}{48x_{10}x_{20}^3}+\frac{(x_1-x_{10})^3(-1+4a_1+4b_{12}x_{20})}{12x_{10}^3}\\
&+\frac{(x_1-x_{10})(-1+4a_1+4b_{11}x_{10}+4b_{12}x_{20})}{4x_{10}}-\frac{(x_1-x_{10})^2(x_2-x_{20})(b_{21}x_{10}+7b_{12}x_{20})}{24x_{10}^2x_{20}}\\
&+\frac{(x_1-x_{10})^3(x_2-x_{20})(b_{21}x_{10}+10b_{12}x_{20})}{48x_{10}^3x_{20}}+\frac{(x_2-x_{20})(-1+4a_2+4b_{21}x_{10}+4b_{22}x_{20})}{4x_{20}},
\end{aligned}
$$

$$
C_X^{(1)}(x|x_0;\theta)=\frac{(x_1-x_{10})(x_2-x_{20})((-9+16a_2)b_{21}x_{10}^2+16b_{21}^2x_{10}^3+b_{12}x_{20}^2(-9+16a_1+16b_{12}x_{20}))}{48x_{10}^2x_{20}^2}+\frac{(x_2-x_{20})(-1+2a_2)b_{21}x_{10}}{4x_{20}^2}
$$

$$+\frac{(x_2-x_{20})(48b_{21}^2x_{10}^3-8b_{12}x_{20}^2(-7+12a_1+12b_{12}x_{20})+x_{10}(9-48a_2+48a_2^2+8b_{21}x_{20}-96b_{11}b_{12}x_{20}^2-48b_{22}^2x_{20}^2))}{192x_{10}x_{20}^2}$$
$$+\frac{(x_1-x_{10})^2(-4b_{21}^2x_{10}+b_{21}(-1+2a_2-4b_{12}x_{20}+2b_{22}x_{20}))}{48x_{10}x_{20}}+\frac{(x_1-x_{10})(-12b_{21}^2x_{10}-b_{21}(-7+12a_2+12b_{22}x_{20}))}{24x_{20}}$$
$$+\frac{(x_1-x_{10})^2(x_{20}(-27-144a_1^2+16b_{11}^2x_{10}^2-16b_{12}(x_{10}-9x_{20})+16b_{12}^2(2x_{10}-9x_{20})x_{20}-144a_1(-1+2b_{12}x_{20})))}{768x_{10}^3x_{20}}$$
$$+\frac{(x_1-x_{10})(9+48a_1^2-48b_{11}^2x_{10}^2+8b_{12}(x_{10}-6x_{20})+48b_{12}^2x_{20}^2+48a_1(-1+2b_{12}x_{20}))}{192x_{10}^2}$$
$$+\frac{(x_2-x_{20})^2(-9b_{21}^2x_{10}^3-b_{12}x_{20}^2(1-2a_1+4b_{12}x_{20})+b_{21}x_{10}^2(9-18a_2+2b_{21}x_{20}))}{48x_{10}x_{20}^3}$$
$$+\frac{(x_2-x_{20})^2(-27+144a_2-144a_2^2+32b_{11}b_{12}x_{20}^2+16b_{22}^2x_{20}^2-16b_{21}x_{20}(1+4b_{12}x_{20}))}{768x_{20}^3}$$
$$-\frac{(3+16a_1^2-16b_{12}x_{20}+16b_{12}^2x_{20}^2+16a_1(-1+2b_{12}x_{20}))}{32x_{10}}$$
$$-\frac{16b_{21}^2x_{10}^2+(3+16a_2^2+32a_1b_{11}x_{20}+32b_{11}b_{12}x_{20}^2+16b_{22}^2x_{20}^2+16a_2(-1+2b_{22}x_{20}))+16x_{10}(b_{11}^2x_{20}+b_{21}(-1+2a_2+2b_{22}x_{20}))}{32x_{20}},$$

$$C_X^{(2)}(x|x_0;\theta)=-\frac{b_{21}^2x_{10}^2}{8x_{20}^2}-\frac{b_{12}(-1+2a_1+4b_{12}x_{20})}{24x_{10}}-\frac{b_{21}x_{10}(-3+6a_2+4b_{21}x_{20})}{24x_{20}^2}$$
$$-\frac{3+16a_1^2-16b_{12}x_{20}+16b_{12}^2x_{20}^2+16a_1(-1+2b_{12}x_{20})}{128x_{10}^2}$$
$$-\frac{9+48a_2^2+16b_{11}^2x_{20}^2+32b_{11}b_{12}x_{20}^2+16b_{22}^2x_{20}^2+16a_2(-3+2b_{21}x_{20})+16b_{21}x_{20}(-1+4b_{12}x_{20}+2b_{22}x_{20})}{384x_{20}^2}.$$

References

Aït-Sahalia, Y., 1999. Transition densities for interest rate and other nonlinear diffusions. Journal of Finance 54, 1361–1395.

Aït-Sahalia, Y., 2002. Maximum-likelihood estimation of discretely-sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262.

Aït-Sahalia, Y., 2008. Closed-form likelihood expansions for multivariate diffusions. Annals of Statistics 36, 906–937.

Aït-Sahalia, Y., Kimmel, R., 2007. Maximum likelihood estimation of stochastic volatility models. Journal of Financial Economics 83, 413–452.

Aït-Sahalia, Y., Mykland, P.A., 2003. The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica 71, 483–549.

Babbs, S.H., Nowman, K.B., 1999. Kalman filtering of generalized Vasicek term structure models. Journal of Financial and Quantitative Analysis 34, 115–130.

Bakshi, G., Ju, N., Ou-Yang, H., 2006. Estimation of continuous-time models with an application to equity volatility dynamics. Journal of Financial Economics 82, 227–249.

Bakshi, G.S., Ju, N., 2005. A refinement to Aït-Sahalia's (2002) "Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach". Journal of Business 78, 2037–2052.

Bliss, R.R., 1997. Testing term structure estimation methods. In: Boyle, P., Pennachi, G., Ritchken, P. (Eds.), Advances in Futures and Options Research, vol. 9. JAI Press, Greenwich, CT, pp. 197–231.

Brandt, M., He, P., 2002. Simulated likelihood estimation of affine term structure models from panel data. Unpublished Working Paper, University of Pennsylvania.

Brandt, M., Santa-Clara, P., 2002. Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. Journal of Financial Economics 63, 161–210.

Chacko, G., Viceira, L.M., 2003. Spectral GMM estimation of continuous-time processes. Journal of Econometrics 116, 259–292.

Chen, R.-R., Scott, L.O., 1993. Maximum likelihood estimation of a multi-factor equilibrium model of the term structure of interest rates. Journal of Fixed Income December, 14–31.

Cheridito, P., Filipović, D., Kimmel, R.L., 2007. Market price of risk specifications for affine models: theory and evidence. Journal of Financial Economics 83, 123–170.

Cheridito, P., Filipović, D., Kimmel, R.L., 2010. A note on the Dai-Singleton canonical representation of affine term structure models. Mathematical Finance 20, 509–519.

Collin-Dufresne, P., Goldstein, R., Jones, C., 2003. Can interest rate volatility be extracted from the yield curve? A test of unspanned stochastic volatility. Unpublished Working Paper, UC Berkeley and University of Minnesota.

Cox, J.C., Ingersoll, J.E., Ross, S.A., 1985. A theory of the term structure of interest rates. Econometrica 53, 385–408.

Dai, Q., Singleton, K.J., 2000. Specification analysis of affine term structure models. Journal of Finance 55, 1943–1978.

DiPietro, M., 2001. Bayesian inference for discretely sampled diffusion processes with financial applications. Ph.D. Thesis, Department of Statistics, Carnegie-Mellon University.

Duarte, J., 2004. Evaluating an alternative risk preference in affine term structure models. Review of Financial Studies 17, 379–404.

Duffee, G.R., 2002. Term premia and interest rate forecasts in affine models. Journal of Finance 57, 405–443.

Duffee, G.R., Stanton, R.H., 2001. Estimation of dynamic term structure models. Technical Report, Haas School of Business, University of California-Berkeley.

Duffie, D., Kan, R., 1996. A yield-factor model of interest rates. Mathematical Finance 6, 379–406.

Duffie, D., Pedersen, L.H., Singleton, K.J., 2003. Modeling sovereign yield spreads: a case study of Russian debt. Journal of Finance 58, 119–159.

Egorov, A.V., Li, H., Ng, D., 2008. A tale of two yield curves: modeling the joint term structure of dollar and euro interest rates. Journal of Econometrics, forthcoming.

Egorov, A.V., Li, H., Xu, Y., 2003. Maximum likelihood estimation of time inhomogeneous diffusions. Journal of Econometrics 114, 107–139.

Fisher, M., Gilles, C., 1996a. Estimating exponential-affine models of the term structure. Technical Report, Federal Reserve Bank of Atlanta.

Fisher, M., Gilles, C., 1996b. Term premia in exponential-affine models of the term structure. Technical Report, Federal Reserve Bank of Atlanta.

Gibbons, M.R., Ramaswamy, K., 1993. A test of the Cox, Ingersoll, and Ross model of the term structure. Review of Financial Studies 6, 619–658.

Gouriéroux, C., Sufana, R., 2004. A classification of two factor affine diffusion term structure models. Technical Report, University of Toronto.

Hurn, A., Jeisman, J., Lindsay, K., 2007. Seeing the wood for the trees: a critical evaluation of methods to estimate the parameters of stochastic differential equations. Journal of Financial Econometrics 5, 390–455.

Jensen, B., Poulsen, R., 2002. Transition densities of diffusion processes: numerical comparison of approximation techniques. Journal of Derivatives 9, 1–15.

Li, M., 2005. A damped diffusion framework for financial modeling and closed-form maximum likelihood estimation. Technical Report, Georgia Institute of Technology.

Liu, J., Pan, J., Pedersen, L.H., 2001. Density-based inference in affine jump-diffusions. Technical Report, UCLA.

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

McCulloch, J.H., Kwon, H., 1993. US term structure data, 1947–1991. Technical Report, Working Paper 93-6, Ohio State University.

Mosburger, G., Schneider, P., 2005. Modelling international bond markets with affine term structure models. Technical Report, Vienna University of Economics and Business Administration.

Pedersen, A.R., 1995. A new approach to maximum-likelihood estimation for stochastic differential equations based on discrete observations. Scandinavian Journal of Statistics 22, 55–71.

Piazzesi, M., 2009. Affine term structure models. In: Aït-Sahalia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics, vol. 1. North-Holland, Amsterdam, The Netherlands, pp. 691–766.

Schaumburg, E., 2001. Maximum likelihood estimation of jump processes with applications to finance. Ph.D. Thesis, Princeton University.

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

Stramer, O., Yan, J., 2005. On simulated likelihood of discretely observed diffusion processes and comparison to closed-form approximation. Technical Report, University of Iowa.

Tauchen, G.E., 2004. The EMM reply to the Duffee-Stanton paper: "Estimation of dynamic term structure models". Unpublished working paper, Duke University.

Thompson, S., 2008. Identifying term structure volatility from the LIBOR-swap curve. Review of Financial Studies 21, 819–854.

Vasicek, O., 1977. An equilibrium characterization of the term structure. Journal of Financial Economics 5, 177–188.

Vuong, Q.H., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Yu, J., 2007. Closed-form likelihood estimation of jump-diffusions with an application to the realignment risk premium of the Chinese yuan. Journal of Econometrics 141, 1245–1280.