

Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions*

Yacine Aït-Sahalia
Department of Economics
Princeton University and NBER[†]

Robert Kimmel
Department of Economics
Princeton University[‡]

October 11, 2002

Abstract

We develop and implement a technique for maximum likelihood estimation in closed-form of multivariate affine yield models of the term structure of interest rates. Affine yield models owe their popularity among both practitioners and academics to the fact that they allow for straightforward pricing of bonds and other interest rate derivatives. However, estimation still poses many challenging issues. Applying the method of Aït-Sahalia (2001), we derive closed-form approximations to the likelihood functions for all nine of the Dai and Singleton (2000) canonical affine models corresponding to dimensions 1, 2 and 3 of the state vector. Monte Carlo simulations reveal that our technique produces extremely accurate approximations of the exact likelihood function.

*Financial support from the NSF under grant SES-0111140 is gratefully acknowledged.

[†]Princeton, NJ 08544-1021. Phone: (609) 258-4015. E-mail: yacine@princeton.edu.

[‡]Princeton, NJ 08544-1021. Phone: (609) 258-0243. E-mail: rkimmel@princeton.edu.

1 Introduction

In this paper, we develop and implement a technique for closed-form maximum likelihood estimation of multivariate affine yield models of the term structure of interest rates. Affine yield models are very popular among both practitioners and academics, largely because they have very desirable analytical properties and allow for straightforward pricing of bonds and other interest rate derivatives. These prices are solutions to the Feynman-Kac 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, affine yield models owe their popularity to the fact that this partial differential equation decomposes into a system of ordinary differential equations, which can be solved quickly, even with a large number of underlying factors.

Despite their relatively desirable analytic properties, estimation of affine yield models still poses many challenges. The likelihood function of 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 now discuss in turn.

Moments of affine diffusions can be found in closed-form. Estimation of affine yield models through the generalized method of moments is therefore 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 (2001) 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 is not.

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. Because the method is computationally intensive, requiring heavy simulations, and is 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 it has been shown to perform poorly at least in the context of dynamic term structure models (see Duffee and Stanton (2001)). 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 boundaries 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 data observations could not have occurred.

One alternative to method of moments estimation is quasi-maximum likelihood (see e.g., Duffee (2002)).

In this approach, the density of the state vector, conditional on the previous observation, is assumed to have a multivariate Gaussian distribution, and the mean vector and covariance matrix of the state vector are assumed to be proportional to the length of time between observations. 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 for some affine yield models), it is necessary to assume that at least some of the yields are observed with error; Piazzesi (2002) discusses this issue. Quasi-maximum likelihood estimation has the advantages that it is feasible for all affine yield models, and never estimates models that imply the state vector took on unattainable values for some data observations. However, only some affine yield models have a Gaussian transition density, and even for those models, the assumptions of QML estimation regarding the means and variances of the transition density are not accurate.

Instead of quasi-maximum likelihood one can consider true maximum likelihood estimation, with the likelihood function calculated numerically or estimated through simulation techniques. The transition function can be found as the solution to the Kolmogorov forward equation; this partial differential equation must be solved numerically. However, the transition density must be calculated for each data observation, and for each value of the parameter vector considered during a likelihood search. Already computationally intensive for a scalar diffusion, estimation by this method becomes extremely difficult for multivariate diffusions; see Jensen and Poulsen (2002) for a comparison of different methods. Maximum likelihood estimation can be implemented via simulations instead. Pedersen (1995) developed a technique for estimating the likelihood function of discrete observations of a diffusion process by simulations, which Brandt and Santa-Clara (2002) extended to multivariate diffusions. When applied to term structure models, likelihood methods usually assume, as does Duffee (2002), that an arbitrary set of benchmark yields are observed without error, with all remaining yields observed with some error; however, 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 can still be considerable. Finally, Liu, Pan, and Pedersen (2001) propose to numerically Fourier-invert the known characteristic function of an affine diffusion to recover an approximation of its density.

As an alternative to the above techniques, we propose maximum likelihood estimation with the likelihood function approximated by a series of highly accurate expansions for the log-likelihood function (or equivalent the density) due to Aït-Sahalia (2001), which generalizes to arbitrary multivariate processes the univariate results developed in Aït-Sahalia (2002) (see also Aït-Sahalia (1999) for examples of application of the univariate method in finance). The key aspect of the method is that, unlike all the approaches described above, the resulting density expansion from this approach is in closed form. While the method can be applied to *any* multivariate diffusion, irrespectively of the particular specification adopted, we illustrate this technique here

in the context of affine term structure models.

This class of models has been studied extensively by Dai and Singleton (2000). They show that there are $N + 1$ non-nested families of affine models with N state variables, M of which entering the diffusion matrix. 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 when $N = 1$, where both single-factor affine models have closed-form likelihoods.) Duffie, Pedersen, and Singleton (2002) propose a decomposition of the likelihood function of an affine model which requires independence of the volatility variables and the simulation of the remaining part of the likelihood. The independence assumption is not satisfied as soon as $M > 1$. 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 (2002)), nor for that matter to affine specifications although this is our focus here. We show how maximum-likelihood estimation can be implemented using for instance bond yields as the observables.

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, before detailing in Section 4 the construction of the closed-form likelihood expansions. We then test in Section 5 the accuracy of our technique by imposing necessary parameter restrictions so that all nine models have closed-form likelihood functions, and compare estimates derived using the true likelihoods on simulated data to those derived using our approximations on the same data. We find uniformly that the maximum-likelihood estimates produced by our method are extremely close to the estimates produced by the exact likelihood function, and conclude in Section 6. The explicit formulae we obtain for the affine term structure models are contained in the Appendix. They are also available in computer form from the authors upon request.

2 Affine Term Structure Models

At its most general level, a multivariate term structure model specifies that the instantaneous riskless rate r_t is a deterministic function of an N -dimensional vector of state variables, X_t :

$$r_t = r(X_t; \theta). \tag{2.1}$$

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

$$dX_t = \mu^Q(X_t; \theta) dt + \sigma(X_t; \theta) dW_t^Q \tag{2.2}$$

where X_t and $\mu(X_t; \theta)$ are $N \times 1$ vectors, $\sigma(X_t; \theta)$ is an $N \times N$ matrix, θ is a p -dimensional parameter and W_t is an $N \times 1$ vector of independent Brownian motions. Let $v(x; \theta) \equiv \sigma(x; \theta) \sigma'(x; \theta)$ where $'$ denotes transposition.

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:

$$B(x, t, T; \theta) = E \left[\exp \left(- \int_t^T r_u du \right) \middle| X_t = x \right] \quad (2.3)$$

where the expectation is taken with respect to the risk-neutral dynamics of X specified in (2.2). It is also well-known that B satisfies the partial differential equation:

$$\frac{\partial B}{\partial t} + \mu(x; \theta)' \frac{\partial B}{\partial x} + \frac{1}{2} \text{Trace} \left[v(x; \theta) \frac{\partial^2 B}{\partial x \partial x'} \right] - r(x, t; \theta) B = 0 \quad (2.4)$$

with the final condition $B(x, T, T; \theta) = 1$ for all x and θ . Such a model is well-defined provided that (2.2) is well-defined, the expected value (2.3) is finite, or, equivalently, the PDE (2.4) has a well-defined solution.

Although there are several different ways to define affine yield term structure models, we shall use the following definition. An affine yield model is any model where the short rate (2.1) is an affine function of the state vector and the risk-neutral dynamics (2.2) are affine:

$$dX_t = \tilde{K} \left(\tilde{A} - X_t \right) dt + \sqrt{S(X_t; \beta)} dW_t^Q \quad (2.5)$$

where $S(X_t; \beta)$ is the diagonal matrix with elements $S_{ii} = 1 + X_t' \beta_i$. (Note that the requirement that $S(X_t; \beta)$ be diagonal does not result in a loss of generality.) 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 denotes the number of independent state variables entering the diffusion structure for the state variables. With N factors, there are therefore $N + 1$ non-nested families of affine models corresponding to $M = 0, 1, \dots, N$, and we will consider one, two and three factor models. Even holding the size of the state vector fixed, there are several distinct non-nested families of affine yield models, each with its own form of likelihood function. We characterize all possible affine yield models with three or fewer state variables in Appendix A.

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

$$B(x, t, T; \theta) = \exp \left(-\gamma_0(\tau; \theta) - \gamma(\tau; \theta)' x \right) \quad (2.6)$$

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

$$g(x, t, T; \theta) = \gamma_0(\tau; \theta) + \gamma(\tau; \theta)' x. \quad (2.7)$$

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 final condition for the bond price implies that $\gamma_0(0; \theta) = \gamma(0; \theta) = 0$, while

$$\begin{aligned} r_t &= \lim_{\tau \rightarrow 0} \frac{1}{\tau} \{ \gamma_0(\tau; \theta) + \gamma(\tau; \theta)' x \} \\ &= \frac{\partial \gamma_0}{\partial \tau}(0; \theta) + \frac{\partial \gamma}{\partial \tau}(0; \theta)' x \\ &\equiv \delta_0 + \delta' x. \end{aligned} \tag{2.8}$$

Affine yield models undoubtedly owe much of their popularity to the fact that bond prices can be calculated quickly as solutions to a system of ordinary differential equations. Under non-linear models, bond prices will normally be solutions to a partial differential equation that is far more difficult to solve. Plugging the solution (2.6) into the PDE (2.4), we see indeed that the functions γ_0 and γ jointly solve the two ordinary Riccati differential equations:

$$\frac{\partial \gamma_0(\tau; \theta)}{\partial \tau} = \delta_0 + \tilde{A}' \tilde{K}' \gamma(\tau; \theta) - \frac{1}{2} \sum_{i=M+1}^N [\gamma(\tau; \theta)]_i^2 \tag{2.9}$$

$$\frac{\partial \gamma(\tau; \theta)}{\partial \tau} = \delta + \tilde{K}' \gamma(\tau; \theta) - \frac{1}{2} \sum_{i=M+1}^N [\gamma(\tau; \theta)]_i^2 \beta_i. \tag{2.10}$$

The solution can be evaluated rapidly by standard numerical techniques for ordinary differential equations, yielding the functions γ_0 and γ .

3 Estimation Procedure

3.1 From Yields to State

We propose to estimate the parameter vector θ based on a panel data of observed bond prices or equivalently bond yields, recognizing that the state vector X_t is unobservable. The first task is therefore to infer the state vector X_t at date t from the cross-section of bond yields at date t with different maturities. We estimate the models using highly accurate approximations to the likelihood function of the state vector, applying the technique of Aït-Sahalia (2001). Affine yield models, as their name implies, make yields of zero coupon bonds linear functions of the state vector. Given this simple relationship between yields and the state vector, the likelihood function of bond yields is a simple transformation of the likelihood function of the state vector.

If the number of observed yields at that point in time is smaller than the number N of state variables in the model, then the state is not completely observed, and the vector of observed yields does not follow a Markov process, even if the (unobserved) state vector does, enormously complicating maximum likelihood estimation.

On the other hand, if the number of observed yields is larger than the number of state variables, then some of the yields can be expressed as deterministic functions of other observed yields, *without error*. Even tiny deviations from the predicted values have a likelihood of zero. A common practice (see, for example, Duffee (2002)) is to assume that certain benchmark yields are observed precisely, whereas other yields are observed with measurement error, where the error is generally held to be independent over time, of the error on other yields, and of the model.

To avoid these issues, we propose to use a number of observed yields that is exactly equal to the number N of state variables in the postulated model $A_M(N)$. At each date t , the state vector X_t is then exactly identified, and the vector of observed yields follows a Markov process. However, the parameters of the model will not always be completely identified. Specifically, there are affine yield models that generate identical dynamics for the time series of observed yields, but different dynamics for additional yields. Equating N observed yields with time to maturity τ_1, \dots, τ_N on the left hand side with their values under the model, given in light of (2.7) by the affine expression on the right hand side, we have:

$$\begin{bmatrix} g(t, t + \tau_1) \\ \vdots \\ g(t, t + \tau_N) \end{bmatrix} = \begin{bmatrix} \gamma_0(\tau_1; \theta) \\ \vdots \\ \gamma_0(\tau_N; \theta) \end{bmatrix} + \begin{bmatrix} \gamma(\tau_1; \theta)' \\ \vdots \\ \gamma(\tau_N; \theta)' \end{bmatrix} \begin{bmatrix} X_{1t} \\ \vdots \\ X_{Nt} \end{bmatrix} \quad (3.1)$$

or, in matrix form,

$$g_t = \Gamma_{0t}(\theta) + \Gamma_t(\theta)' X_t. \quad (3.2)$$

3.2 The Observed State Dynamics

While the only parameters entering the transformation from observed yields to the state variables are the parameters of the dynamics of the state process under the risk-neutral measure Q , once we have constructed our time series of values of X_t sampled at dates t_0, t_1, \dots, t_n the dynamics of the state variable that we will be able to infer from this time series are the dynamics under the physical measure of the process, which we denote as P . The first step in the estimation procedure is the only place where we rely on the tractability of the affine bond pricing model. 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-Nykodym derivative dQ/dP :

$$\begin{aligned} dX_t &= \mu^P(X_t; \theta) dt + \sigma(X_t; \theta) dW_t^P \\ &= \{\mu^Q(X_t; \theta) + \sigma(X_t; \theta) \Lambda(X_t; \theta)\} dt + \sigma(X_t; \theta) dW_t^P. \end{aligned} \quad (3.3)$$

We adopt the simple specification for the market price of risk

$$\Lambda(X_t; \theta) = \sigma(X_t; \theta)' \lambda \quad (3.4)$$

with λ an $N \times 1$ vector of constant parameters, so that under P , the instantaneous drift of each state variables 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

$$\begin{aligned} \mu^P(X_t; \theta) &= \tilde{K}(\tilde{A} - X_t) + S(X_t; \beta)' \lambda \\ &\equiv K(A - X_t). \end{aligned} \quad (3.5)$$

Following Dai and Singleton (2000), we consider the canonical affine models where K , A and β have the normalized form

$$K = \begin{bmatrix} K_{M \times M} & 0_{M \times (N-M)} \\ K_{(N-M) \times M} & K_{(N-M) \times (N-M)} \end{bmatrix}, \quad A = \begin{bmatrix} A_{M \times 1} \\ 0_{(N-M) \times 1} \end{bmatrix}, \quad \beta = \begin{bmatrix} I_{M \times M} & \beta_{M \times (N-M)} \\ 0_{(N-M) \times M} & 0_{(N-M) \times (N-M)} \end{bmatrix}.$$

The parameters governing the state variable dynamics are constrained as follows:

$$A_{M \times 1} \geq 0 \quad (3.6)$$

$$\beta_{M \times (N-M)} \geq 0 \quad (3.7)$$

$$K_{M \times M} A_{M \times 1} \geq \frac{1}{2} \quad (3.8)$$

Furthermore, the off-diagonal elements of $\kappa_{M \times M}$ must be non-positive. We further discuss these restrictions in the Appendix.

As mentioned above, having an affine μ^P is not required for our likelihood expansions. Since we can derive likelihood expansions for arbitrary diffusions, we can allow μ^P to contain terms that are non-affine, such as square roots of linear functions of the state vector, as in Duarte (1999) for instance. Duffee (2002) also allows for a more general market price of risk specification than Dai and Singleton (2000), where μ^Q and μ^P (and the diffusion matrix) remain affine. 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 (3.1).

3.3 Maximum-Likelihood Estimation

Since the relationship between the state vector and bond yields is affine, as given by (3.2), 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, which is a constant matrix 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 (3.3). 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 units apart. Since $x = \Gamma^{-1}(\theta)(g - \Gamma_0(\theta))$, we have

$$p_G(\Delta, g|g_0; \theta) = \Gamma^{-1}(\theta) p_X(\Delta, \Gamma^{-1}(\theta)(g - \Gamma_0(\theta)) | \Gamma^{-1}(\theta)(g_0 - \Gamma_0(\theta)); \theta) \quad (3.9)$$

Then, recognizing that the yields vector 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, \dots, t_n has the simple form

$$\ell_n(\theta) \equiv n^{-1} \sum_{i=1}^n l_G(t_i - t_{i-1}, g_{t_i} | g_{t_{i-1}}; \theta) \quad (3.10)$$

where $l_G \equiv \ln p_G$. 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, and in our Monte Carlo experiments below, these types of models are estimated on the basis of weekly data, so that $t_i - t_{i-1} = \Delta = 7/365$ is a fixed number. Figure 1 describes 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.

4 Closed-Form Likelihood Expansions

In the next Section, we explain how to derive closed-form approximations to l_G , hence to the log-likelihood function of the discretely sampled vector of yields in light (3.10). Closed-form approximations to l_G are obtained by applying to the different term structure models (3.3) the general method described in Aït-Sahalia (2001), which extends to multivariate diffusions the univariate results of Aït-Sahalia (1999) and Aït-Sahalia (2002). 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 (3.9) to recover the corresponding expansion for l_G . So we can reduce the problem to one of approximating l_X , and we now turn to that question.

4.1 Reducibility

As defined in Aït-Sahalia (2001), 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

$$dY_t = \mu_Y(Y_t; \theta) dt + dW_t. \quad (4.1)$$

Every univariate diffusion is reducible. However, this is not the case for every multivariate diffusion. Whether or not a given multivariate diffusion is reducible depends on the specification of its σ matrix. Specifically, Proposition 1 of Aït-Sahalia (2001) provides a necessary and sufficient condition for reducibility: the diffusion X is reducible if and only if the inverse diffusion matrix $\sigma^{-1} = [\sigma_{i,j}^{-1}]_{i,j=1,\dots,m}$ satisfies on $\mathcal{S}_X \times \Theta$ the condition that

$$\frac{\partial \sigma_{ij}^{-1}(x; \theta)}{\partial x_k} = \frac{\partial \sigma_{ik}^{-1}(x; \theta)}{\partial x_j} \quad (4.2)$$

for each triplet $(i, j, k) = 1, \dots, m$ such that $k > j$, or equivalently

$$\sum_{l=1}^m \frac{\partial \sigma_{ik}(x; \theta)}{\partial x_l} \sigma_{lj}(x; \theta) = \sum_{l=1}^m \frac{\partial \sigma_{ij}(x; \theta)}{\partial x_l} \sigma_{lk}(x; \theta). \quad (4.3)$$

Whenever a diffusion is reducible, an expansion can be computed for the transition density p_X of X by first computing it for the density p_Y of Y and then transforming Y back into X (see Section 4.2). When a diffusion is not reducible, the situation is more involved (see Section 4.3).

Affine yield models of the class $A_M(N)$ with $M = 0$ or $M = N$ are reducible. $A_M(N)$ models with $0 < M < N$ are reducible only if the β_{ij} coefficients are constrained to be zero.

4.2 Determination of the Coefficients in the Reducible Case

The expansion for l_Y is of the form

$$l_Y^{(K)}(\Delta, y|y_0; \theta) = -\frac{m}{2} \ln(2\pi\Delta) + \frac{C_Y^{(-1)}(y|y_0; \theta)}{\Delta} + \sum_{k=0}^K C_Y^{(k)}(y|y_0; \theta) \frac{\Delta^k}{k!}. \quad (4.4)$$

As shown in Theorem 1 of Aït-Sahalia (2001), the coefficients of the expansion are given explicitly by:

$$C_Y^{(-1)}(y|y_0; \theta) = -\frac{1}{2} \sum_{i=1}^m (y_i - y_{0i})^2 \quad (4.5)$$

$$C_Y^{(0)}(y|y_0; \theta) = \sum_{i=1}^m (y_i - y_{0i}) \int_0^1 \mu_{Y_i}(y_0 + u(y - y_0); \theta) du \quad (4.6)$$

and, for $k \geq 1$,

$$C_Y^{(k)}(y|y_0; \theta) = k \int_0^1 G_Y^{(k)}(y_0 + u(y - y_0)|y_0; \theta) u^{k-1} du \quad (4.7)$$

where

$$\begin{aligned} G_Y^{(1)}(y|y_0; \theta) &= -\sum_{i=1}^m \frac{\partial \mu_{Y_i}(y; \theta)}{\partial y_i} - \sum_{i=1}^m \mu_{Y_i}(y; \theta) \frac{\partial C_Y^{(0)}(y|y_0; \theta)}{\partial y_i} \\ &\quad + \frac{1}{2} \sum_{i=1}^m \left\{ \frac{\partial^2 C_Y^{(0)}(y|y_0; \theta)}{\partial y_i^2} + \left[\frac{\partial C_Y^{(0)}(y|y_0; \theta)}{\partial y_i} \right]^2 \right\} \end{aligned} \quad (4.8)$$

and for $k \geq 2$

$$\begin{aligned}
G_Y^{(k)}(y|y_0; \theta) &= -\sum_{i=1}^m \mu_{Yi}(y; \theta) \frac{\partial C_Y^{(k-1)}(y|y_0; \theta)}{\partial y_i} + \frac{1}{2} \sum_{i=1}^m \frac{\partial^2 C_Y^{(k-1)}(y|y_0; \theta)}{\partial y_i^2} \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{h=0}^{k-1} \binom{k-1}{h} \frac{\partial C_Y^{(h)}(y|y_0; \theta)}{\partial y_i} \frac{\partial C_Y^{(k-1-h)}(y|y_0; \theta)}{\partial y_i}.
\end{aligned} \tag{4.9}$$

Given an expansion for the density p_Y of Y , an expansion for the density p_X of X can be obtained by a direct application of the Jacobian formula:

$$\begin{aligned}
l_X^{(K)}(\Delta, x|x_0; \theta) &= -\frac{m}{2} \ln(2\pi\Delta) - D_v(x; \theta) + \frac{C_Y^{(-1)}(\gamma(x; \theta) | \gamma(x_0; \theta); \theta)}{\Delta} \\
&\quad + \sum_{k=0}^K C_Y^{(k)}(\gamma(x; \theta) | \gamma(x_0; \theta); \theta) \frac{\Delta^k}{k!}
\end{aligned} \tag{4.10}$$

from $l_Y^{(K)}$ given in (4.4), using the coefficients $C_Y^{(k)}$, $k = -1, 0, \dots, K$ given above, and where

$$D_v(x; \theta) \equiv \frac{1}{2} \ln(\text{Det}[v(x; \theta)]). \tag{4.11}$$

4.3 Determination of the Coefficients in the Irreducible Case

In the irreducible case, we apply Theorem 2 of Ait-Sahalia (2001). The expansion of the log likelihood has the form

$$l_X^{(K)}(\Delta, x|x_0; \theta) = -\frac{m}{2} \ln(2\pi\Delta) - D_v(x; \theta) + \frac{C_X^{(-1)}(x|x_0; \theta)}{\Delta} + \sum_{k=0}^K C_X^{(k)}(x|x_0; \theta) \frac{\Delta^k}{k!}. \tag{4.12}$$

The approach is to calculate a Taylor series in $(x - x_0)$ of each coefficient $C_X^{(k)}$, at order j_k in $(x - x_0)$. Such an expansion will be denoted by $C_X^{(j_k, k)}$ at order $j_k = 2(K - k)$, for $k = -1, 0, \dots, K$.

The resulting expansion will then be

$$\tilde{l}_X^{(K)}(\Delta, x|x_0; \theta) = -\frac{m}{2} \ln(2\pi\Delta) - D_v(x; \theta) + \frac{C_X^{(j_{-1}, -1)}(x|x_0; \theta)}{\Delta} + \sum_{k=0}^K C_X^{(j_k, k)}(x|x_0; \theta) \frac{\Delta^k}{k!} \tag{4.13}$$

Such a Taylor expansion was unnecessary in the reducible case: the expressions given in Section 4.2 provide the explicit expressions of the coefficients $C_Y^{(k)}$ and then in (4.10) we have the corresponding ones for $C_X^{(k)}$. However, even for an irreducible diffusion, it is still possible to compute the coefficients $C_X^{(j_k, k)}$ explicitly.

With $v(x; \theta) \equiv \sigma(x; \theta) \sigma^T(x; \theta)$, define the following functions of the coefficients and their derivatives:

$$\begin{aligned}
G_X^{(0)}(x|x_0; \theta) &= \frac{m}{2} - \sum_{i=1}^m \mu_i(x; \theta) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial v_{ij}(x; \theta)}{\partial x_i} \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_j} \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \frac{\partial^2 C_X^{(-1)}(x|x_0; \theta)}{\partial x_i \partial x_j} \\
&\quad - \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} \frac{\partial D_v(x; \theta)}{\partial x_j},
\end{aligned} \tag{4.14}$$

$$\begin{aligned}
G_X^{(1)}(x|x_0; \theta) &= -\sum_{i=1}^m \frac{\partial \mu_i(x; \theta)}{\partial x_i} + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2 v_{ij}(x; \theta)}{\partial x_i \partial x_j} \\
&\quad - \sum_{i=1}^m \mu_i(x; \theta) \left(\frac{\partial C_X^{(0)}(x|x_0; \theta)}{\partial x_i} - \frac{\partial D_v(x; \theta)}{\partial x_i} \right) \\
&\quad + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial v_{ij}(x; \theta)}{\partial x_i} \left(\frac{\partial C_X^{(0)}(x|x_0; \theta)}{\partial x_j} - \frac{\partial D_v(x; \theta)}{\partial x_j} \right) \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \left\{ \frac{\partial^2 C_X^{(0)}(x|x_0; \theta)}{\partial x_i \partial x_j} - \frac{\partial^2 D_v(x; \theta)}{\partial x_i \partial x_j} \right. \\
&\quad \left. + \left(\frac{\partial C_X^{(0)}(x|x_0; \theta)}{\partial x_i} - \frac{\partial D_v(x; \theta)}{\partial x_i} \right) \left(\frac{\partial C_X^{(0)}(x|x_0; \theta)}{\partial x_j} - \frac{\partial D_v(x; \theta)}{\partial x_j} \right) \right\}
\end{aligned} \tag{4.15}$$

and for $k \geq 2$:

$$\begin{aligned}
G_X^{(k)}(x|x_0; \theta) &= -\sum_{i=1}^m \mu_i(x; \theta) \frac{\partial C_X^{(k-1)}(x|x_0; \theta)}{\partial x_i} + \sum_{i=1}^m \sum_{j=1}^m \frac{\partial v_{ij}(x; \theta)}{\partial x_i} \frac{\partial C_X^{(k-1)}(x|x_0; \theta)}{\partial x_j} \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \frac{\partial^2 C_X^{(k-1)}(x|x_0; \theta)}{\partial x_i \partial x_j} \\
&\quad + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \left\{ 2 \left(\frac{\partial C_X^{(0)}(x|x_0; \theta)}{\partial x_i} - \frac{\partial D_v(x; \theta)}{\partial x_i} \right) \frac{\partial C_X^{(k-1)}(x|x_0; \theta)}{\partial x_j} \right. \\
&\quad \left. + \sum_{h=1}^{k-2} \binom{k-2}{h} \frac{\partial C_X^{(h)}(x|x_0; \theta)}{\partial x_i} \frac{\partial C_X^{(k-1-h)}(x|x_0; \theta)}{\partial x_j} \right\}.
\end{aligned} \tag{4.16}$$

For each $k = -1, 0, \dots, K$, the coefficient $C_X^{(k)}(x|x_0; \theta)$ in (4.12) solves the equation

$$f_X^{(k-1)}(x|x_0; \theta) = 0 \tag{4.17}$$

where

$$\begin{aligned}
f_X^{(-2)}(x|x_0; \theta) &= -2C_X^{(-1)}(x|x_0; \theta) - \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_j} \\
f_X^{(-1)}(x|x_0; \theta) &= -\sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} \frac{\partial C_X^{(0)}(x|x_0; \theta)}{\partial x_j} - G_X^{(0)}(x|x_0; \theta)
\end{aligned}$$

and for $k \geq 1$

$$f_X^{(k-1)}(x|x_0; \theta) = C_X^{(k)}(x|x_0; \theta) - \sum_{i=1}^m \sum_{j=1}^m v_{ij}(x; \theta) \frac{\partial C_X^{(-1)}(x|x_0; \theta)}{\partial x_i} \frac{\partial C_X^{(k)}(x|x_0; \theta)}{\partial x_j} - G_X^{(k)}(x|x_0; \theta)$$

where the functions $G_X^{(k)}$, $k = 0, 1, \dots, K$ are given above. $G_X^{(k)}$ involves only the coefficients $C_X^{(h)}$ for $h = -1, \dots, k-1$, so this system of equation can be utilized to solve recursively for each coefficient at a time. Specifically, the equation $f_X^{(-2)} = 0$ determines $C_X^{(-1)}$; given $C_X^{(-1)}$, $G_X^{(0)}$ becomes known and the equation $f_X^{(-1)} = 0$ determines $C_X^{(0)}$; given $C_X^{(-1)}$ and $C_X^{(0)}$, $G_X^{(1)}$ becomes known and the equation $f_X^{(0)} = 0$ then determines $C_X^{(1)}$, etc. It turns out that this results in a system of linear equations in the coefficients of the

polynomials $C_X^{(j_k, k)}$, so each one of these equations can be solved explicitly in the form of the Taylor expansion $C_X^{(j_k, k)}$ of the coefficient $C_X^{(k)}$, at order j_k in $(x - x_0)$.

5 Monte Carlo Results

In order to determine 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. In all examples considered, we find our parameter estimates are very close to the true maximum likelihood estimates for simulated data at the weekly frequency. 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.

As shown in (3.9), the likelihood function of a yield vector is simply the likelihood of the canonical state variables times a Jacobian factor. The full parameter vector θ consists of all the elements of (K, A, β) . As mentioned, we consider all nine models corresponding to $N = 1, 2$ and 3 and we estimate each $A_M(N)$ model using n time series observations of N zero-coupon bond yields.

The individual models themselves are shown in Appendix A, the parameter restrictions are shown in Table 1, while the actual parameter values used in the simulations are contained in Table 2. The parameter values satisfy all existence, boundary non-attainability, and stationarity conditions. 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 shown in Table 2. Note again that our expansion does *not* require these further restrictions. The only reason we impose them is to have an exact likelihood to compare our expansion to.

For each canonical model, we simulate 5,000 data series of 501 weekly observations each ($\Delta = 1/52$) of the vector of N state variables, giving $n = 500$ pairs of discrete transitions of that process. The parameter values we use to generate the simulated data are specified in Table 2. Each of the simulated sample path samples is produced by a Milshtein discretization of the process, using thirty intervals per week. Twenty nine out of every thirty 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.

We then proceed to estimate the model parameters, using both our approximate likelihood expressions and the true likelihood functions. We go through all nine models, in Tables 3 through 6, reporting 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, $\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$. The notation $\hat{\theta}^{(2)}$ indicates that we use an expansion at order 2 in Δ (i.e., $K = 2$ in Section 4) to obtain the approximate likelihood estimator. The bias and standard deviation of $\hat{\theta}^{(2)}$ around $\hat{\theta}^{(MLE)}$ are reported as a

percentage of the corresponding quantities for $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$.

These tables show that our technique produces parameter estimates 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 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 very small compared to the standard deviation of the MLE itself. This means 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.)

Finally, we report in Figures 2, 3 and 4 the empirical distribution of the estimation of the exact MLE around the true value (left column) and the approximation error from replacing the exact MLE with our estimator (left column), for the parameters estimated under the two-dimensional models $A_M(2)$, $M = 0, 1, 2$ based on the same 5,000 simulations as above. Results are similar for the one and three dimensional models and are not reported to save space. As expected, the small sample distribution of the MLE estimates of the mean reversion parameters κ tends to be slightly skewed to the right, as is typical in a near unit root case (our values of κ are all between 0.5 and 2.0). The right columns also show that the estimation error is largely uneventful. The difference in the scale of the x axis between the left and right columns is another way of showing that the approximation error induced by replacing the exact MLE estimator with our approximation is negligible.

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.

6 Conclusions

We have developed and implemented a technique for maximum likelihood estimation of affine yield models, and implemented this technique for several families of such 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 extremely close to the true maximum likelihood estimates. 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. And 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 trivial computational cost given the closed form nature of our formulae.

Much remains to be done. Our technique can be applied to models that are non-affine. One special case of this are models that are affine under Q but non-affine under P , since being affine 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. Our technique allows estimation of a much broader class of such models, and does so in closed form.

References

- Aït-Sahalia, Y. (1999). Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.
- Aït-Sahalia, Y. (2001). Closed-form likelihood expansions for multivariate diffusions. Technical report, Princeton University.
- Aït-Sahalia, Y. (2002). Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach. *Econometrica* 70, 223–262.
- Aït-Sahalia, Y. and P. A. Mykland (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. *Econometrica*, forthcoming.
- Brandt, M. and P. He (2002). Simulated likelihood estimation of affine term structure models from panel data. Working paper, University of Pennsylvania.
- Brandt, M. and P. Santa-Clara (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. and L. Viceira (2001). Spectral GMM estimation of continuous-time processes. Working paper, Harvard University.
- Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica* 53, 385–408.
- Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *Journal of Finance* 55, 1943–1978.
- Duarte, J. (1999). The relevance of the price of risk in affine term structure models. Technical report, The University of Chicago.
- Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance* 57, 405–443.
- Duffee, G. R. and R. H. Stanton (2001). Estimation of dynamic term structure models. Technical report, Haas School of Business, University of California-Berkeley.
- Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance* 6, 379–406.
- Duffie, D., L. H. Pedersen, and K. J. Singleton (2002). Modeling sovereign yield spreads: A case study of russian debt. *Journal of Finance*, forthcoming.
- Gibbons, M. R. and K. Ramaswamy (1993). A test of the Cox, Ingersoll, and Ross model of the term structure. *Review of Financial Studies* 6, 619–658.
- Jensen, B. and R. Poulsen (2002). Transition densities of diffusion processes: Numerical comparison of approximation techniques. *Journal of Derivatives* 9, 1–15.
- Liu, J., J. Pan, and L. H. Pedersen (2001). Density-based inference in affine jump-diffusions. Working paper, UCLA.
- 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. (2002, June). Affine term structure models. Forthcoming in the Handbook of Financial Econometrics.
- Singleton, K. (2001). Estimation of affine asset pricing models using the empirical characteristic function. *Journal of Econometrics* 102, 111–141.
- Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

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 and diffusion coefficients. Furthermore, there will typically be infinitely many model specifications that produce exactly the same 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.

Using their notation, 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 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 will discuss each model in turn. Although 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), we nonetheless find it useful to explore all nine in full detail. 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 a single state variable:

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

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:

$$dX_{1t} = -\kappa_{11}X_{1t}dt + dW_{1t}^P$$

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:

$$dX_{1t} = \kappa_{11}[\alpha_1 - X_{1t}]dt + \sqrt{X_{1t}}dW_{1t}^P$$

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 Q , the dynamics of the $A_0(1)$ and $A_1(1)$ model are respectively.

$$\begin{aligned} dX_{1t} &= [-\lambda_1 - \kappa_{11}X_{1t}]dt + dW_{1t}^Q \\ dX_{1t} &= [\kappa_{11}\alpha_1 - (\kappa_{11} + \lambda_1)X_{1t}]dt + \sqrt{X_{1t}}dW_{1t}^Q. \end{aligned}$$

A.2 Two Factor Models

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

$$r_t = \delta_0 + \delta_1 X_{1t} + \delta_2 X_{2t}$$

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

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} -X_{1t} \\ -X_{2t} \end{bmatrix} dt + d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \end{bmatrix} \quad (\text{A.1})$$

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 :

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} \alpha_1 - X_{1t} \\ -X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{1 + \beta_{21} X_{1t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \end{bmatrix}. \quad (\text{A.2})$$

In general, the likelihood function for this type of diffusion is not known in closed-form; however, if we impose the constraints $\kappa_{21} = 0$ 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.

The $A_2(2)$ model has the representation:

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{2t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \end{bmatrix} \quad (\text{A.3})$$

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

The three canonical specifications are as presented in Dai and Singleton (2000); however, there are 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:

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} -X_{1t} \\ -X_{2t} \end{bmatrix} dt + d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \end{bmatrix}$$

with the constraint $(\kappa_{11} - \kappa_{22})^2 < 4\kappa_{12}\kappa_{21}$. 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 κ_{11} and κ_{22} to be complex conjugate pairs. An example of the second type of non-conforming diffusion is:

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} \alpha_1 - X_{1t} \\ -X_{2t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{1t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \end{bmatrix}$$

This diffusion most closely resembles the $A_1(2)$ canonical form, but no change of variables can generate the 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 κ matrix.

Under Q , the dynamics of the state vector in the three models are respectively

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(- \begin{bmatrix} \lambda_{10} \\ \lambda_{20} \end{bmatrix} - \begin{bmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + d \begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \end{bmatrix}$$

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} \alpha_1 \kappa_{11} \\ -\lambda_2 \end{bmatrix} - \begin{bmatrix} \kappa_{11} + \lambda_1 & 0 \\ \kappa_{21} + \lambda_2 \beta_{21} & \kappa_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{1 + \beta_{21} X_{1t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \end{bmatrix}$$

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \left(\begin{bmatrix} \kappa_{11} \alpha_1 + \kappa_{12} \alpha_2 \\ \kappa_{21} \alpha_1 + \kappa_{22} \alpha_2 \end{bmatrix} - \begin{bmatrix} \kappa_{11} + \lambda_{11} & \kappa_{11} \\ \kappa_{21} & \kappa_{22} + \lambda_{22} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \right) dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 \\ 0 & \sqrt{X_{1t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \end{bmatrix}$$

A.3 Three Factor Models

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

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

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

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} -X_{1t} \\ -X_{2t} \\ -X_{3t} \end{bmatrix} dt + d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}$$

The transition density of the state vector is trivariate Gaussian.

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

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} \alpha_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{1 + \beta_{21} X_{1t}} & 0 \\ 0 & 0 & \sqrt{1 + \beta_{31} X_{1t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}$$

The transition density function is known in closed-form only if the first state variable is independent of the other two, i.e., if $\kappa_{21} = 0$, $\kappa_{31} = 0$, $\beta_{21} = 0$, 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) and a bivariate Gaussian (the distribution of the other two).

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

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \\ -X_{3t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{1 + \beta_{31} X_{1t} + \beta_{32} X_{2t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}$$

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

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

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} \alpha_1 - X_{1t} \\ \alpha_2 - X_{2t} \\ \alpha_3 - X_{3t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{X_{3t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}$$

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

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

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

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 κ matrix are complex conjugate pairs. Similarly, the model

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} \alpha_1 - X_{1t} \\ -X_{2t} \\ -X_{3t} \end{bmatrix} dt + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{1t}} & 0 \\ 0 & 0 & \sqrt{1 + \beta_{31} X_{1t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^P \\ W_{2t}^P \\ W_{3t}^P \end{bmatrix}$$

is similar to the $A_1(3)$ model, but the absence of a constant coefficient in the diffusion of the second state variable make 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 Q , the dynamics of the state vector in the four models are respectively

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

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \left(\begin{bmatrix} \kappa_{11}\alpha_1 \\ -\lambda_2 \\ -\lambda_3 \end{bmatrix} - \begin{bmatrix} \kappa_{11} + \lambda_1 & 0 & 0 \\ \kappa_{21} + \lambda_2\beta_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} + \lambda_3\beta_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} \right) dt \\ + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{1 + \beta_{21}X_{1t}} & 0 \\ 0 & 0 & \sqrt{1 + \beta_{31}X_{1t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \\ W_{3t}^Q \end{bmatrix}$$

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \left(\begin{bmatrix} \kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 \\ \kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 \\ -\lambda_3 \end{bmatrix} - \begin{bmatrix} \kappa_{11} + \lambda_1 & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} + \lambda_2 & 0 \\ \kappa_{31} + \lambda_3\beta_{31} & \kappa_{32} + \lambda_3\beta_{32} & \kappa_{33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} \right) dt \\ + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{1 + \beta_{31}X_{1t} + \beta_{32}X_{2t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \\ W_{3t}^Q \end{bmatrix}$$

$$d \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} = \begin{bmatrix} \kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 + \kappa_{13}\alpha_3 \\ \kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 + \kappa_{23}\alpha_3 \\ \kappa_{31}\alpha_1 + \kappa_{32}\alpha_2 + \kappa_{33}\alpha_3 \end{bmatrix} - \begin{bmatrix} \kappa_{11} + \lambda_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} + \lambda_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} + \lambda_{33} \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ X_{3t} \end{bmatrix} dt \\ + \begin{bmatrix} \sqrt{X_{1t}} & 0 & 0 \\ 0 & \sqrt{X_{2t}} & 0 \\ 0 & 0 & \sqrt{X_{3t}} \end{bmatrix} d \begin{bmatrix} W_{1t}^Q \\ W_{2t}^Q \\ W_{3t}^Q \end{bmatrix}.$$

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 and CIR 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 from the authors upon request.

B.1 The $A_0(2)$ Model

The coefficients below correspond to the SDE (A.1).

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

$$C_X^{(0)}(x|x_0; \theta) = -\frac{1}{2}(x_1 - x_{10})^2 \kappa_{11} - (x_1 - x_{10})x_{10}\kappa_{11} - \frac{1}{2}(x_1 - x_{10})(x_2 - x_{20})\kappa_{21} \\ - \frac{1}{2}(x_2 - x_{20})^2 \kappa_{22} + (x_2 - x_{20})(-x_{10}\kappa_{21} - x_{20}\kappa_{22})$$

$$C_X^{(1)}(x|x_0; \theta) = \frac{1}{24}(x_1 - x_{10})^2(-4\kappa_{11}^2 - 3\kappa_{21}^2) - \frac{1}{3}(x_1 - x_{10})(x_2 - x_{20})\kappa_{21}\kappa_{22} \\ - \frac{1}{2}(x_2 - x_{20})\kappa_{22}(x_{10}\kappa_{21} + x_{20}\kappa_{22}) + \frac{1}{2}(x_1 - x_{10})(-x_{10}(\kappa_{11}^2 + \kappa_{21}^2) - x_{20}\kappa_{21}\kappa_{22}) \\ + \frac{1}{24}(x_2 - x_{20})^2(\kappa_{21}^2 - 4\kappa_{22}^2) + \frac{1}{2}(\kappa_{11} - x_{10}^2\kappa_{11}^2 + \kappa_{22} - (x_{10}\kappa_{21} + x_{20}\kappa_{22})^2)$$

$$C_X^{(2)}(x|x_0; \theta) = -\frac{1}{12}((x_1 - x_{10})^2 \kappa_{21}^2 \kappa_{22}) + \frac{1}{12}(x_2 - x_{20})^2 \kappa_{21}^2 \kappa_{22} \\ - \frac{1}{6}(x_1 - x_{10})\kappa_{21}\kappa_{22}(x_{10}\kappa_{21} + x_{20}\kappa_{22}) + \frac{1}{6}(x_2 - x_{20})\kappa_{21}(x_{10}(\kappa_{11}^2 + \kappa_{21}^2) + x_{20}\kappa_{21}\kappa_{22}) \\ + \frac{1}{12}(-2\kappa_{11}^2 - \kappa_{21}^2 - 2\kappa_{22}^2) + \frac{1}{12}(x_1 - x_{10})(x_2 - x_{20})\kappa_{21}(\kappa_{11}^2 + \kappa_{21}^2 - \kappa_{22}^2)$$

B.2 The $A_1(2)$ Model

The coefficients below correspond to the SDE (A.2).

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

$$\begin{aligned}
C_X^{(0)}(x|x_0; \theta) &= \frac{(x_1-x_{10})(1+4x_{10}\kappa_{11}+4x_{10}^2\beta_{21}\kappa_{11}-4\alpha_1\kappa_{11}-4x_{10}\beta_{21}\alpha_1\kappa_{11})}{-4x_{10}-4x_{10}^2\beta_{21}} - \frac{x_{10}(x_2-x_{20})^3\beta_{21}(\kappa_{21}+\beta_{21}\alpha_1\kappa_{21}-x_{20}\beta_{21}\kappa_{22})}{24(1+x_{10}\beta_{21})^4} \\
&+ \frac{(x_1-x_{10})^2(6-24\alpha_1\kappa_{11}+x_{10}\beta_{21}(13-48\alpha_1\kappa_{11})-2x_{10}^2\beta_{21}^2(1+12\alpha_1\kappa_{11}))}{48x_{10}^2(1+x_{10}\beta_{21})^2} \\
&+ \frac{(x_1-x_{10})^3(-8+32\alpha_1\kappa_{11}+4x_{10}^3\beta_{21}^3(1+8\alpha_1\kappa_{11}))+3x_{10}^2\beta_{21}^2(-9+32\alpha_1\kappa_{11})+x_{10}\beta_{21}(-25+96\alpha_1\kappa_{11})}{96x_{10}^3(1+x_{10}\beta_{21})^3} \\
&+ \frac{(x_1-x_{10})^4(x_{10}^3\beta_{21}^3(659-2304\alpha_1\kappa_{11})+2x_{10}^2\beta_{21}^2(461-1728\alpha_1\kappa_{11}))}{2304x_{10}^4(1+x_{10}\beta_{21})^4} + \frac{(x_2-x_{20})(-(x_{10}\kappa_{21})+\alpha_1\kappa_{21}-x_{20}\kappa_{22})}{1+x_{10}\beta_{21}} \\
&+ \frac{(x_1-x_{10})^4(3x_{10}\beta_{21}(197-768\alpha_1\kappa_{11})+144(1-4\alpha_1\kappa_{11})-2x_{10}^4\beta_{21}^4(43+288\alpha_1\kappa_{11}))}{2304x_{10}^4(1+x_{10}\beta_{21})^4} \\
&- \frac{(x_1-x_{10})(x_2-x_{20})(\kappa_{21}+\beta_{21}\alpha_1\kappa_{21}-x_{20}\beta_{21}\kappa_{22})}{2(1+x_{10}\beta_{21})^2}
\end{aligned}$$

$$\begin{aligned}
C_X^{(1)}(x|x_0; \theta) &= \frac{(x_2-x_{20})(x_{20}\kappa_{22}(\beta_{21}-12\kappa_{22}-24x_{10}\beta_{21}\kappa_{22}-2x_{10}\beta_{21}^2(1+6x_{10}\kappa_{22})))}{24(1+x_{10}\beta_{21})^3} + \frac{(x_2-x_{20})(\kappa_{21}(-12x_{10}^3\beta_{21}^2\kappa_{22}+12\alpha_1\kappa_{22}))}{24(1+x_{10}\beta_{21})^3} \\
&+ \frac{(x_2-x_{20})(\kappa_{21}(12x_{10}^2\beta_{21}(-2+\beta_{21}\alpha_1)\kappa_{22}+2x_{10}(\beta_{21}+\beta_{21}^2\alpha_1-6\kappa_{22}+12\beta_{21}\alpha_1\kappa_{22})))}{24(1+x_{10}\beta_{21})^3} + \frac{(x_2-x_{20})(\kappa_{21}(-1-\beta_{21}\alpha_1))}{24(1+x_{10}\beta_{21})^3} \\
&+ \frac{(x_1-x_{10})(x_2-x_{20})(-16\kappa_{21}\kappa_{22}+\beta_{21}(20x_{20}\kappa_{22}^2+\kappa_{21}(5-28x_{10}\kappa_{22}-20\alpha_1\kappa_{22})))}{48(1+x_{10}\beta_{21})^4} \\
&+ \frac{(x_1-x_{10})(x_2-x_{20})(4x_{10}\beta_{21}^3(\kappa_{22}(x_{20}+x_{10}^2\kappa_{21}+5x_{10}x_{20}\kappa_{22})-\alpha_1(\kappa_{21}+5x_{10}\kappa_{21}\kappa_{22})))}{48(1+x_{10}\beta_{21})^4} \\
&+ \frac{(x_1-x_{10})(x_2-x_{20})(-\beta_{21}^2(-5\alpha_1\kappa_{21}+5x_{20}\kappa_{22}+8x_{10}^2\kappa_{21}\kappa_{22}+4x_{10}(\kappa_{21}+10\alpha_1\kappa_{21}\kappa_{22}-10x_{20}\kappa_{22}^2)))}{48(1+x_{10}\beta_{21})^4} \\
&- \frac{(x_2-x_{20})^2(8x_{10}^3\beta_{21}^4(-1+30\alpha_1\kappa_{11}-30\alpha_1^2\kappa_{11}^2+60x_{10}\kappa_{22}+30x_{10}^2(\kappa_{11}^2+8\kappa_{22}^2)))}{11520x_{10}(1+x_{10}\beta_{21})^5} \\
&- \frac{(x_2-x_{20})^2(+4x_{10}^2\beta_{21}^3(-79+60x_{10}^3\kappa_{21}^2-180\alpha_1^2(\kappa_{11}^2+x_{10}\kappa_{21}^2)))}{11520x_{10}(1+x_{10}\beta_{21})^5} - \frac{(x_2-x_{20})^2(-480x_{10}(x_{10}\kappa_{21}^2-4\kappa_{22}^2))}{11520x_{10}(1+x_{10}\beta_{21})^5} \\
&- \frac{(x_2-x_{20})^2(+4x_{10}^2\beta_{21}^3(-180x_{10}\kappa_{22}(-1+x_{20}^2\kappa_{22})+180\alpha_1(\kappa_{11}+2x_{10}x_{20}\kappa_{21}\kappa_{22})+60x_{10}^2(3\kappa_{11}^2+32\kappa_{22}^2)))}{11520x_{10}(1+x_{10}\beta_{21})^5} \\
&- \frac{(x_2-x_{20})^2(+15\beta_{21}(-3-16\alpha_1^2\kappa_{11}^2+16\alpha_1(\kappa_{11}-6x_{10}^2\kappa_{21}^2)-16x_{10}\kappa_{22}+16x_{10}^2(\kappa_{11}^2+6x_{20}\kappa_{21}\kappa_{22}+32\kappa_{22}^2)))}{11520x_{10}(1+x_{10}\beta_{21})^5} \\
&- \frac{(x_2-x_{20})^2(x_{10}\beta_{21}^2(-113+720x_{10}^3\kappa_{21}^2-720\alpha_1^2(\kappa_{11}^2+x_{10}\kappa_{21}^2)))}{11520x_{10}(1+x_{10}\beta_{21})^5} \\
&- \frac{(x_2-x_{20})^2(x_{10}\beta_{21}^2(-720x_{10}x_{20}^2\kappa_{22}^2+720x_{10}^2(\kappa_{11}^2+2x_{20}\kappa_{21}\kappa_{22}+16\kappa_{22}^2)))}{11520x_{10}(1+x_{10}\beta_{21})^5} \\
&- \frac{(x_2-x_{20})^2(x_{10}\beta_{21}^2(720\alpha_1(\kappa_{11}+2x_{10}\kappa_{21})(-x_{10}\kappa_{21})+x_{20}\kappa_{22})))}{11520x_{10}(1+x_{10}\beta_{21})^5} \\
&+ \frac{(x_1-x_{10})(9-48\alpha_1\kappa_{11}+48\alpha_1^2\kappa_{11}^2+9x_{10}\beta_{21}(3-16\alpha_1\kappa_{11}+16\alpha_1^2\kappa_{11}^2))}{192x_{10}^2(1+x_{10}\beta_{21})^3} \\
&+ \frac{(x_1-x_{10})(-48x_{10}^5\beta_{21}^2(\beta_{21}\kappa_{11}^2+\kappa_{21}^2)-16x_{10}^4\beta_{21}(9\beta_{21}\kappa_{11}^2+9\kappa_{21}^2+\beta_{21}^2\kappa_{22}))}{192x_{10}^2(1+x_{10}\beta_{21})^3} \\
&+ \frac{(x_1-x_{10})(8x_{10}^3(\beta_{21}^3(1-6\alpha_1\kappa_{11}+6\alpha_1^2\kappa_{11}^2)-12\kappa_{21}^2))}{192x_{10}^2(1+x_{10}\beta_{21})^3}
\end{aligned}$$

$$\begin{aligned}
& + \frac{(x_1-x_{10})(8x_{10}^3(-6\beta_{21}(3\kappa_{11}^2+2\kappa_{21}(-\alpha_1\kappa_{21})+x_{20}\kappa_{22}))+2\beta_{21}^2(3\alpha_1^2\kappa_{21}^2-6x_{20}\alpha_1\kappa_{21}\kappa_{22}+\kappa_{22}(-2+3x_{20}^2\kappa_{22})))}{192x_{10}^2(1+x_{10}\beta_{21})^3} \\
& + \frac{(x_1-x_{10})(16x_{10}^2(\beta_{21}^2(2-9\alpha_1\kappa_{11}+9\alpha_1^2\kappa_{11}^2)-3(\kappa_{11}^2+2\kappa_{21}(-\alpha_1\kappa_{21})+x_{20}\kappa_{22})))}{192x_{10}^2(1+x_{10}\beta_{21})^3} \\
& + \frac{(x_1-x_{10})(16x_{10}^2(\beta_{21}(3\alpha_1^2\kappa_{21}^2-6x_{20}\alpha_1\kappa_{21}\kappa_{22}+\kappa_{22}(-1+3x_{20}^2\kappa_{22})))}{192x_{10}^2(1+x_{10}\beta_{21})^3} \\
& + \frac{(x_1-x_{10})^2(-135(3-16\alpha_1\kappa_{11}+16\alpha_1^2\kappa_{11}^2)-540x_{10}\beta_{21}(3-16\alpha_1\kappa_{11}+16\alpha_1^2\kappa_{11}^2))}{11520x_{10}^3(1+x_{10}\beta_{21})^4} \\
& + \frac{(x_1-x_{10})^2(240x_{10}^6\beta_{21}^3(\beta_{21}\kappa_{11}^2+\kappa_{21}^2)+480x_{10}^5\beta_{21}^2(2\beta_{21}\kappa_{11}^2+2\kappa_{21}^2+\beta_{21}^2\kappa_{22}))}{11520x_{10}^3(1+x_{10}\beta_{21})^4} \\
& + \frac{(x_1-x_{10})^2(\beta_{21}^2(-2461+12960\alpha_1\kappa_{11}-12960\alpha_1^2\kappa_{11}^2)-240\beta_{21}(\alpha_1\kappa_{21}-x_{20}\kappa_{22})^2+240(\kappa_{11}^2-2\alpha_1\kappa_{21}^2+2x_{20}\kappa_{21}\kappa_{22}))}{11520x_{10}^3(1+x_{10}\beta_{21})^4} \\
& + \frac{(x_1-x_{10})^2(-4x_{10}^3(\beta_{21}^3(473-2160\alpha_1\kappa_{11}+2160\alpha_1^2\kappa_{11}^2)+240\kappa_{21}^2-240\beta_{21}(\kappa_{11}^2+4\kappa_{21}(-\alpha_1\kappa_{21})+x_{20}\kappa_{22})))}{11520x_{10}^3(1+x_{10}\beta_{21})^4} \\
& + \frac{(x_1-x_{10})^2(-4x_{10}^3(120\beta_{21}^2(4\alpha_1^2\kappa_{21}^2-8x_{20}\alpha_1\kappa_{21}\kappa_{22}+\kappa_{22}(-1+4x_{20}^2\kappa_{22})))}{11520x_{10}^3(1+x_{10}\beta_{21})^4} \\
& + \frac{(x_1-x_{10})^2(-8x_{10}^4\beta_{21}(\beta_{21}^3(47-270\alpha_1\kappa_{11}+270\alpha_1^2\kappa_{11}^2)+30\kappa_{21}^2-60\beta_{21}(3\kappa_{11}^2+7\kappa_{21}(-\alpha_1\kappa_{21})+x_{20}\kappa_{22})))}{11520x_{10}^3(1+x_{10}\beta_{21})^4} \\
& + \frac{(x_1-x_{10})^2(-8x_{10}^4\beta_{21}(30\beta_{21}^2(7\alpha_1^2\kappa_{21}^2-14x_{20}\alpha_1\kappa_{21}\kappa_{22}+\kappa_{22}(-4+7x_{20}^2\kappa_{22})))}{11520x_{10}^3(1+x_{10}\beta_{21})^4} \\
& - \frac{9-48\alpha_1\kappa_{11}+48\alpha_1^2\kappa_{11}^2+48x_{10}^4\beta_{21}(\beta_{21}\kappa_{11}^2+\kappa_{21}^2)-48x_{10}^3(-\kappa_{21}^2+\beta_{21}^2(2\alpha_1\kappa_{11}^2+\kappa_{22}))-2\beta_{21}(\kappa_{11}^2-\alpha_1\kappa_{21}^2+x_{20}\kappa_{21}\kappa_{22})}{96x_{10}(1+x_{10}\beta_{21})^2} \\
& - \frac{4x_{10}(\beta_{21}(5-24\alpha_1\kappa_{11}+24\alpha_1^2\kappa_{11}^2)+12(\alpha_1^2\kappa_{21}^2+\kappa_{22}(-1+x_{20}^2\kappa_{22}))-2\alpha_1(\kappa_{11}^2+x_{20}\kappa_{21}\kappa_{22}))}{96x_{10}(1+x_{10}\beta_{21})^2} \\
& - \frac{8x_{10}^2(\beta_{21}^2(1-6\alpha_1\kappa_{11}+6\alpha_1^2\kappa_{11}^2)+6(\kappa_{11}^2+2\kappa_{21}(-\alpha_1\kappa_{21})+x_{20}\kappa_{22})))}{96x_{10}(1+x_{10}\beta_{21})^2} \\
& - \frac{8x_{10}^2(6\beta_{21}(\alpha_1^2\kappa_{21}^2+\kappa_{22}(-2+x_{20}^2\kappa_{22}))-2\alpha_1(2\kappa_{11}^2+x_{20}\kappa_{21}\kappa_{22})))}{96x_{10}(1+x_{10}\beta_{21})^2} \\
C_Y^{(2)}(x|x_0; \theta) & = \frac{-(45(3-16\alpha_1\kappa_{11}+16\alpha_1^2\kappa_{11}^2)+195x_{10}\beta_{21}(3-16\alpha_1\kappa_{11}+16\alpha_1^2\kappa_{11}^2))}{5760x_{10}^2(1+x_{10}\beta_{21})^4} - \frac{2x_{10}^2(-120\beta_{21}(\alpha_1\kappa_{21}-x_{20}\kappa_{22})^2)}{5760x_{10}^2(1+x_{10}\beta_{21})^4} \\
& - \frac{2x_{10}^2(\beta_{21}^2(467-2520\alpha_1\kappa_{11}+2520\alpha_1^2\kappa_{11}^2)+120(\kappa_{11}^2-2\alpha_1\kappa_{21}^2+2x_{20}\kappa_{21}\kappa_{22}+4\kappa_{22}^2))}{5760x_{10}^2(1+x_{10}\beta_{21})^4} - \frac{960x_{10}^6\beta_{21}^4\kappa_{22}^2}{5760x_{10}^2(1+x_{10}\beta_{21})^4} \\
& - \frac{240x_{10}^5\beta_{21}^2(\kappa_{21}^2+\beta_{21}(\kappa_{11}^2+16\kappa_{22}^2))}{5760x_{10}^2(1+x_{10}\beta_{21})^4} - \frac{8x_{10}^3(\beta_{21}^3(91-450\alpha_1\kappa_{11}+450\alpha_1^2\kappa_{11}^2)+120\kappa_{21}^2+30\beta_{21}^2(\alpha_1\kappa_{21}-x_{20}\kappa_{22})^2)}{5760x_{10}^2(1+x_{10}\beta_{21})^4} \\
& - \frac{8x_{10}^3(30\beta_{21}(3\kappa_{11}^2+2\alpha_1\kappa_{21}^2-2x_{20}\kappa_{21}\kappa_{22}+16\kappa_{22}^2))}{5760x_{10}^2(1+x_{10}\beta_{21})^4} - \frac{8x_{10}^4\beta_{21}(\beta_{21}^3(23-120\alpha_1\kappa_{11}+120\alpha_1^2\kappa_{11}^2))}{5760x_{10}^2(1+x_{10}\beta_{21})^4} \\
& - \frac{8x_{10}^4\beta_{21}(150\kappa_{21}^2+60\beta_{21}^2(\alpha_1\kappa_{21}-x_{20}\kappa_{22})^2+30\beta_{21}(3\kappa_{11}^2+4\alpha_1\kappa_{21}^2-4x_{20}\kappa_{21}\kappa_{22}+24\kappa_{22}^2))}{5760x_{10}^2(1+x_{10}\beta_{21})^4}
\end{aligned}$$

B.3 The $A_2(2)$ Model

The coefficients below correspond to the SDE (A.3).

$$\begin{aligned}
C_X^{(-1)}(x|x_0; \theta) & = \frac{-21(x_1-x_{10})^6}{256x_{10}^5} + \frac{7(x_1-x_{10})^5}{64x_{10}^4} - \frac{5(x_1-x_{10})^4}{32x_{10}^3} + \frac{(x_1-x_{10})^3}{4x_{10}^2} - \frac{(x_1-x_{10})^2}{2x_{10}} \\
& - \frac{(x_2-x_{20})^2}{2x_{20}} - \frac{21(x_2-x_{20})^6}{256x_{20}^5} + \frac{7(x_2-x_{20})^5}{64x_{20}^4} - \frac{5(x_2-x_{20})^4}{32x_{20}^3} + \frac{(x_2-x_{20})^3}{4x_{20}^2} \\
C_X^{(0)}(x|x_0; \theta) & = \frac{(x_1-x_{10})^4(1-4\alpha_1\kappa_{11}+4x_{20}\kappa_{12}-4\alpha_2\kappa_{12})}{16x_{10}^4} + \frac{(x_1-x_{10})^2(1-4\alpha_1\kappa_{11}+4x_{20}\kappa_{12}-4\alpha_2\kappa_{12})}{8x_{10}^2} \\
& - \frac{(x_1-x_{10})(1+4x_{10}\kappa_{11}-4\alpha_1\kappa_{11}+4x_{20}\kappa_{12}-4\alpha_2\kappa_{12})}{4x_{10}} + \frac{(x_1-x_{10})^3(-1+4\alpha_1\kappa_{11}-4x_{20}\kappa_{12}+4\alpha_2\kappa_{12})}{12x_{10}^3} \\
& - \frac{(x_1-x_{10})^2(x_2-x_{20})^2(x_{20}\kappa_{12}+x_{10}\kappa_{21})}{48x_{10}^2x_{20}^2} - \frac{(x_1-x_{10})(x_2-x_{20})(x_{20}\kappa_{12}+x_{10}\kappa_{21})}{2x_{10}x_{20}} \\
& + \frac{(x_1-x_{10})^2(x_2-x_{20})(7x_{20}\kappa_{12}+x_{10}\kappa_{21})}{24x_{10}^2x_{20}} - \frac{(x_1-x_{10})^3(x_2-x_{20})(10x_{20}\kappa_{12}+x_{10}\kappa_{21})}{48x_{10}^3x_{20}} \\
& + \frac{(x_1-x_{10})(x_2-x_{20})^2(x_{20}\kappa_{12}+7x_{10}\kappa_{21})}{24x_{10}x_{20}^2} - \frac{(x_1-x_{10})(x_2-x_{20})^3(x_{20}\kappa_{12}+10x_{10}\kappa_{21})}{48x_{10}x_{20}^3} \\
& + \frac{(x_2-x_{20})^4(1+4x_{10}\kappa_{21}-4\alpha_1\kappa_{21}-4\alpha_2\kappa_{22})}{16x_{20}^4} + \frac{(x_2-x_{20})^2(1+4x_{10}\kappa_{21}-4\alpha_1\kappa_{21}-4\alpha_2\kappa_{22})}{8x_{20}^2} \\
& - \frac{(x_2-x_{20})(1+4x_{10}\kappa_{21}-4\alpha_1\kappa_{21}+4x_{20}\kappa_{22}-4\alpha_2\kappa_{22})}{4x_{20}} + \frac{(x_2-x_{20})^3(-1-4x_{10}\kappa_{21}+4\alpha_1\kappa_{21}+4\alpha_2\kappa_{22})}{12x_{20}^3}
\end{aligned}$$

$$\begin{aligned}
C_X^{(1)}(x|x_0; \theta) = & \frac{(x_1-x_{10})(x_2-x_{20})(16x_{20}^3\kappa_{12}^2+x_{20}^2\kappa_{12}(9-16\alpha_1\kappa_{11}-16\alpha_2\kappa_{12})+x_{10}^2\kappa_{21}(9+16x_{10}\kappa_{21}-16\alpha_1\kappa_{21}-16\alpha_2\kappa_{22}))}{48x_{10}^2x_{20}^2} \\
& + \frac{(x_1-x_{10})(48x_{20}^3\kappa_{12}^2-48x_{20}^2\kappa_{12}(-1+2\alpha_1\kappa_{11}+2\alpha_2\kappa_{12}))}{192x_{10}^2x_{20}} + \frac{(x_1-x_{10})(8x_{10}^2\kappa_{21}(-7-12x_{10}\kappa_{21}+12\alpha_1\kappa_{21}+12\alpha_2\kappa_{22}))}{192x_{10}^2x_{20}} \\
& + \frac{(x_1-x_{10})(x_{20}(9+48\alpha_1^2\kappa_{11}^2-8x_{10}\kappa_{12}-48\alpha_2\kappa_{12}+48\alpha_2^2\kappa_{12}^2+48\alpha_1\kappa_{11}(-1+2\alpha_2\kappa_{12})-48x_{10}^2(\kappa_{11}^2+2\kappa_{21}\kappa_{22})))}{192x_{10}^2x_{20}} \\
& + \frac{(x_1-x_{10})^2(-144x_{20}^3\kappa_{12}^2+16x_{20}^2\kappa_{12}(-9+18\alpha_1\kappa_{11}+2x_{10}\kappa_{12}+18\alpha_2\kappa_{12}))}{768x_{10}^3x_{20}} \\
& + \frac{(x_2-x_{20})(48x_{10}^3\kappa_{21}^2-48x_{10}^2\kappa_{21}(-1+2\alpha_1\kappa_{21}+2\alpha_2\kappa_{22}))}{192x_{10}x_{20}^2} + \frac{(x_1-x_{10})^2(-16x_{10}^2\kappa_{21}(-1+4x_{10}\kappa_{21}+2\alpha_1\kappa_{21}+2\alpha_2\kappa_{22}))}{768x_{10}^3x_{20}} \\
& + \frac{(x_1-x_{10})^2(x_{20}(-144\alpha_1^2\kappa_{11}^2+16x_{10}\kappa_{12}-144\alpha_1\kappa_{11}(-1+2\alpha_2\kappa_{12})-9(3-16\alpha_2\kappa_{12}+16\alpha_2^2\kappa_{12}^2)))}{768x_{10}^3x_{20}} \\
& - \frac{(x_2-x_{20})^2(x_{20}^2\kappa_{12}(-1+2\alpha_1\kappa_{11}+4x_{20}\kappa_{12}+2\alpha_2\kappa_{12}))}{48x_{10}x_{20}^3} + \frac{(x_2-x_{20})^2(16x_{10}^2\kappa_{21}(-9+2x_{20}\kappa_{21}+18\alpha_1\kappa_{21}+18\alpha_2\kappa_{22}))}{768x_{10}x_{20}^3} \\
& + \frac{(x_2-x_{20})^2(x_{10}(16x_{20}\kappa_{21}+16x_{20}^2(2\kappa_{11}\kappa_{12}-4\kappa_{12}\kappa_{21}+\kappa_{22}^2)))}{768x_{10}x_{20}^3} - \frac{(x_2-x_{20})^2(144x_{10}^3\kappa_{21}^2)}{768x_{10}x_{20}^3} \\
& + \frac{(x_2-x_{20})^2(x_{10}(-9(3+16\alpha_1^2\kappa_{21}^2-16\alpha_2\kappa_{22}+16\alpha_2^2\kappa_{22}^2+16\alpha_1\kappa_{21}(-1+2\alpha_2\kappa_{22}))))}{768x_{10}x_{20}^3} \\
& + \frac{(x_2-x_{20})(8x_{20}^2\kappa_{12}(-7+12\alpha_1\kappa_{11}-12x_{20}\kappa_{12}+12\alpha_2\kappa_{12}))}{192x_{10}x_{20}^2} + \frac{(x_2-x_{20})(48x_{10}^3\kappa_{21}^2-48x_{10}^2\kappa_{21}(-1+2\alpha_1\kappa_{21}+2\alpha_2\kappa_{22}))}{192x_{10}x_{20}^2} \\
& + \frac{(x_2-x_{20})(x_{10}(-8x_{20}\kappa_{21}-48x_{20}^2(2\kappa_{11}\kappa_{12}+\kappa_{22}^2)))}{192x_{10}x_{20}^2} + \frac{(x_1-x_{10})^2(x_{20}(16x_{10}^2(\kappa_{11}^2-4\kappa_{12}\kappa_{21}+2\kappa_{21}\kappa_{22})))}{768x_{10}^3x_{20}} \\
& + \frac{(x_2-x_{20})(x_{10}(3(3+16\alpha_1^2\kappa_{21}^2-16\alpha_2\kappa_{22}+16\alpha_2^2\kappa_{22}^2+16\alpha_1\kappa_{21}(-1+2\alpha_2\kappa_{22}))))}{192x_{10}x_{20}^2} \\
& - \frac{x_{20}(3+16\alpha_1^2\kappa_{11}^2-16\alpha_2\kappa_{12}+16x_{20}^2\kappa_{12}^2+16\alpha_2^2\kappa_{12}^2+16x_{20}\kappa_{12}(1-2\alpha_2\kappa_{12})-16\alpha_1\kappa_{11}(1+2x_{20}\kappa_{12}-2\alpha_2\kappa_{12}))}{32x_{10}x_{20}} \\
& - \frac{16x_{10}^3\kappa_{21}^2+16x_{10}^2(\kappa_{21}(1-2\alpha_1\kappa_{21}-2\alpha_2\kappa_{22})+x_{20}(\kappa_{11}^2+2\kappa_{21}\kappa_{22}))}{32x_{10}x_{20}} - \frac{x_{10}(3+16\alpha_1^2\kappa_{21}^2-16\alpha_2\kappa_{22}+16\alpha_2^2\kappa_{22}^2)}{32x_{10}x_{20}} \\
& - \frac{x_{10}(16\alpha_1\kappa_{21}(-1+2\alpha_2\kappa_{22})+16x_{20}^2(2\kappa_{11}\kappa_{12}+\kappa_{22}^2))}{32x_{10}x_{20}} - \frac{x_{10}(-32x_{20}(\alpha_1(\kappa_{11}^2+\kappa_{21}\kappa_{22})+\alpha_2(\kappa_{11}\kappa_{12}+\kappa_{22}^2)))}{32x_{10}x_{20}}
\end{aligned}$$

$$\begin{aligned}
C_X^{(2)}(x|x_0; \theta) = & -\frac{16x_{10}x_{20}^2\kappa_{12}(1-2\alpha_1\kappa_{11}+4x_{20}\kappa_{12}-2\alpha_2\kappa_{12})}{384x_{10}^2x_{20}^2} \\
& - \frac{(3x_{20}^2(3+16\alpha_1^2\kappa_{11}^2-16\alpha_2\kappa_{12}+16x_{20}^2\kappa_{12}^2+16\alpha_2^2\kappa_{12}^2+16x_{20}\kappa_{12}(1-2\alpha_2\kappa_{12}))-16\alpha_1\kappa_{11}(1+2x_{20}\kappa_{12}-2\alpha_2\kappa_{12}))}{384x_{10}^2x_{20}^2} \\
& - \frac{(48x_{10}^4\kappa_{21}^2+16x_{10}^3\kappa_{21}(3+4x_{20}\kappa_{21}-6\alpha_1\kappa_{21}-6\alpha_2\kappa_{22}))}{384x_{10}^2x_{20}^2} + \frac{(x_{10}^2(-16x_{20}\kappa_{21}(-1+2\alpha_1\kappa_{21}+2\alpha_2\kappa_{22})))}{384x_{10}^2x_{20}^2} \\
& - \frac{(x_{10}^2(16x_{20}^2(\kappa_{11}^2+2\kappa_{11}\kappa_{12}+4\kappa_{12}\kappa_{21}+2\kappa_{21}\kappa_{22}+\kappa_{22}^2)))}{384x_{10}^2x_{20}^2} \\
& - \frac{(x_{10}^2(3(3+16\alpha_1^2\kappa_{21}^2-16\alpha_2\kappa_{22}+16\alpha_2^2\kappa_{22}^2+16\alpha_1\kappa_{21}(-1+2\alpha_2\kappa_{22}))))}{384x_{10}^2x_{20}^2}
\end{aligned}$$

Model	Existence	Boundary	Stationarity
$\mathbf{A}_0(1)$	–	–	$\kappa_{11} > 0$
$\mathbf{A}_1(1)$	$\kappa_{11}\alpha_1 \geq 0$ $\alpha_1 \geq 0$ $\delta_1 \geq 0$	$\kappa_{11}\alpha_1 \geq \frac{1}{2}$	$\kappa_{11} > 0$
$\mathbf{A}_0(2)$	–	–	$\kappa_{11} > 0$ $\kappa_{22} > 0$
$\mathbf{A}_1(2)$	$\kappa_{11}\alpha_1 \geq 0$ $\alpha_1 \geq 0$ $\delta_1 \geq 0$ $\beta_{21} \geq 0$	$\kappa_{11}\alpha_1 \geq \frac{1}{2}$	$\kappa_{11} > 0$ $\kappa_{22} > 0$
$\mathbf{A}_2(2)$	$\kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 \geq 0$ $\kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 \geq 0$ $\alpha_1 \geq 0, \alpha_2 \geq 0$ $\delta_1 \geq 0, \delta_2 \geq 0$ $\kappa_{12} \leq 0, \kappa_{21} \leq 0$	$\kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 \geq \frac{1}{2}$ $\kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 \geq \frac{1}{2}$	Eigen $\begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} > 0$
$\mathbf{A}_0(3)$	–	–	$\kappa_{11} > 0$ $\kappa_{22} > 0$ $\kappa_{33} > 0$
$\mathbf{A}_1(3)$	$\kappa_{11}\alpha_1 \geq 0$ $\alpha_1 \geq 0$ $\delta_1 \geq 0$ $\beta_{21} \geq 0, \beta_{31} \geq 0$	$\kappa_{11}\alpha_1 \geq \frac{1}{2}$	Re $\left[\text{Eigen} \begin{bmatrix} \kappa_{11} & \kappa_{23} \\ \kappa_{32} & \kappa_{33} \end{bmatrix} \right] > 0$
$\mathbf{A}_2(3)$	$\kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 \geq 0$ $\kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 \geq 0$ $\alpha_1 \geq 0, \alpha_2 \geq 0$ $\delta_1 \geq 0, \delta_2 \geq 0$ $\kappa_{12} \leq 0, \kappa_{21} \leq 0$ $\beta_{31} \geq 0, \beta_{32} \geq 0$	$\kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 \geq \frac{1}{2}$ $\kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 \geq \frac{1}{2}$	Re $\left[\text{Eigen} \begin{bmatrix} \kappa_{11} & \kappa_{12} \\ \kappa_{21} & \kappa_{22} \end{bmatrix} \right] > 0$ $\kappa_{33} > 0$
$\mathbf{A}_3(3)$	$\kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 + \kappa_{13}\alpha_3 \geq 0$ $\kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 + \kappa_{23}\alpha_3 \geq 0$ $\kappa_{31}\alpha_1 + \kappa_{32}\alpha_2 + \kappa_{33}\alpha_3 \geq 0$ $\alpha_1 \geq 0, \alpha_2 \geq 0, \alpha_3 \geq 0$ $\delta_1 \geq 0, \delta_2 \geq 0, \delta_3 \geq 0$ $\kappa_{12} \leq 0, \kappa_{13} \leq 0$ $\kappa_{21} \leq 0, \kappa_{23} \leq 0$ $\kappa_{31} \leq 0, \kappa_{32} \leq 0$	$\kappa_{11}\alpha_1 + \kappa_{12}\alpha_2 + \kappa_{13}\alpha_3 \geq \frac{1}{2}$ $\kappa_{21}\alpha_1 + \kappa_{22}\alpha_2 + \kappa_{23}\alpha_3 \geq \frac{1}{2}$ $\kappa_{31}\alpha_1 + \kappa_{32}\alpha_2 + \kappa_{33}\alpha_3 \geq \frac{1}{2}$	Eigen $\begin{bmatrix} \kappa_{11} & \kappa_{12} & \kappa_{13} \\ \kappa_{21} & \kappa_{22} & \kappa_{23} \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} > 0$

Table 1: Parameter Restrictions

This table shows the parameter restrictions imposed on the different models under consideration. “Eigen” denotes the eigenvalues of the matrix.

Model	Restrictions for Exact Density	Parameter Values
$A_0(1)$	none	$\kappa_{11} = 1.0$
$A_1(1)$	none	$\kappa_{11} = 0.50, \alpha_1 = 2.67$
$A_0(2)$	none	$\kappa_{11} = 1.00, \kappa_{21} = -0.50, \kappa_{22} = 2.00$
$A_1(2)$	$\kappa_{21} = 0, \beta_{21} = 0$	$\kappa_{11} = 0.50, \kappa_{22} = 2.00, \alpha_1 = 2.00$
$A_2(2)$	$\kappa_{12} = 0, \kappa_{21} = 0$	$\kappa_{11} = 1.00, \kappa_{22} = 2.00$ $\alpha_1 = 2.00, \alpha_2 = 3.00$
$A_0(3)$	none	$\kappa_{11} = 0.50, \kappa_{21} = -0.20, \kappa_{22} = 1.00$ $\kappa_{31} = 0.10, \kappa_{32} = 0.20, \kappa_{33} = 2.00$
$A_1(3)$	$\kappa_{21} = 0, \kappa_{23} = 0, \kappa_{31} = 0,$ $\beta_{21} = 0, \beta_{31} = 0$	$\kappa_{11} = 0.50, \kappa_{22} = 2.00, \kappa_{32} = -0.10$ $\kappa_{33} = 5.00, \alpha_1 = 2.00$
$A_2(3)$	$\kappa_{12} = 0, \kappa_{21} = 0, \kappa_{31} = 0, \kappa_{32} = 0,$ $\beta_{31} = 0, \beta_{32} = 0$	$\kappa_{11} = 0.50, \kappa_{22} = 2.00, \kappa_{33} = 5.0$ $\alpha_1 = 2.00, \alpha_2 = 1.00$
$A_3(3)$	$\kappa_{12} = \kappa_{21} = \kappa_{13} = \kappa_{31} = \kappa_{23} = \kappa_{32} = 0$	$\kappa_{11} = 0.50, \kappa_{22} = 2.00, \kappa_{33} = 1.00$ $\alpha_1 = 2.00, \alpha_2 = 1.00, \alpha_3 = 1.00$

Table 2: Parameter Values for Monte-Carlo Simulations of Canonical Affine Processes

This table shows the parameter values used in the Monte Carlo simulations. For each model, the additional parameter restrictions (relative to those in Table 1) needed to ensure existence of a known exact likelihood function are shown in the second column. The sole 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 third column shows the values of the remaining parameter values used in the simulations.

Model	Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$		$\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$	
			Mean	Std. Dev.	Mean	Std. Dev.
$A_0(1)$	κ_{11}	1.00	0.1927	0.5419	0.09%	0.06%
$A_1(1)$	κ_{11}	0.5000	0.4843	0.5494	0.02%	0.06%
	α_1	2.6667	0.0792	1.3341	-0.06%	0.11%

Table 3: Monte Carlo Simulations for the One-Dimensional Models

This table reports the results of 5,000 Monte Carlo simulations comparing the distribution of the maximum-likelihood estimator $\hat{\theta}^{(MLE)}$, based on the exact transition density for this model, around the true value of the parameters $\theta^{(TRUE)}$, to the distribution of the difference between the exact MLE $\hat{\theta}^{(MLE)}$ and the approximate MLE $\hat{\theta}^{(2)}$, based on the expansion with $K = 2$ terms, for the $A_0(1)$ and $A_1(1)$ models. The results in the table show that the difference $\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$ is several orders of magnitude smaller than the difference $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$ due to the sampling noise.

Model	Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$		$\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$	
			Mean	Std. Dev.	Mean	Std. Dev.
$A_0(2)$	κ_{11}	1.00	0.2131	0.5442	0.09%	0.06%
	κ_{21}	-0.50	-0.0386	0.5872	0.82%	0.09%
	κ_{22}	2.00	0.3301	0.7428	0.34%	0.16%
$A_1(2)$	κ_{11}	0.50	0.4931	0.5560	0.05%	3.34%
	κ_{22}	2.00	0.1976	0.6916	0.54%	1.64%
	α_1	2.00	0.1265	2.7746	0.90%	2.32%
$A_2(2)$	κ_{11}	1.00	0.4462	0.6277	0.09%	0.95%
	κ_{22}	2.00	0.4477	0.7615	0.36%	3.89%
	α_1	2.00	0.0190	0.4699	0.91%	2.78%
	α_2	3.00	0.0001	0.2802	-5.93%	0.32%

Table 4: Monte Carlo Simulations for the Two-Dimensional Models

This table reports the results of 5,000 Monte Carlo simulations for the $A_0(2)$, $A_1(2)$ and $A_2(2)$ models 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 difference between the exact MLE $\hat{\theta}^{(MLE)}$ and the approximate MLE $\hat{\theta}^{(2)}$, based on the expansion with $K = 2$ terms. The results in the table show that the difference $\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$ is much smaller, and often several orders of magnitude smaller, than the difference $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$ due to the sampling noise.

Model	Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$		$\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$	
			Mean	Std. Dev.	Mean	Std. Dev.
$A_0(3)$	κ_{11}	0.50	0.1922	0.4229	0.03%	0.03%
	κ_{21}	-0.20	-0.0489	0.4730	0.47%	2.31%
	κ_{22}	1.00	0.3568	0.6090	0.17%	3.11%
	κ_{31}	0.10	0.0240	0.4893	0.47%	0.71%
	κ_{32}	0.20	0.0277	0.6268	0.59%	1.05%
	κ_{33}	2.00	0.5343	0.8227	0.29%	0.87%
$A_1(3)$	κ_{11}	0.50	0.4949	0.5565	-0.02%	0.74%
	κ_{22}	2.00	0.1963	0.6979	0.45%	0.45%
	κ_{32}	-0.10	0.0007	0.7284	-6.33%	3.08%
	κ_{33}	5.00	0.3781	1.1021	2.84%	0.94%
	α_1	2.00	0.0846	1.3796	-0.87%	7.16%

Table 5: Monte Carlo Simulations for the $A_0(3)$ and $A_1(3)$ Models

This table reports the results of 5,000 Monte Carlo simulations for the $A_0(3)$ and $A_1(3)$ models 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 difference between the exact MLE $\hat{\theta}^{(MLE)}$ and the approximate MLE $\hat{\theta}^{(2)}$, based on the expansion with $K = 2$ terms. The results in the table show that the difference $\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$ is much smaller, and often several orders of magnitude smaller, than the difference $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$ due to the sampling noise.

Model	Parameter	$\theta^{(TRUE)}$	$\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$		$\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$	
			Mean	Std. Dev.	Mean	Std. Dev.
$A_2(3)$	κ_{11}	0.50	0.4923	0.5556	-0.04%	0.40%
	κ_{22}	2.00	0.4379	0.7703	0.30%	0.21%
	κ_{33}	5.00	0.2025	1.0681	4.64%	0.57%
	α_1	2.00	0.1141	2.6185	1.27%	2.98%
	α_2	1.00	-0.0011	0.1600	1.81%	0.01%
$A_3(3)$	κ_{11}	0.50	0.4918	0.5466	-0.04%	0.43%
	κ_{22}	2.00	0.4411	0.7658	0.29%	0.21%
	κ_{33}	1.00	0.4533	0.6231	-0.22%	0.76%
	α_1	2.00	0.0909	1.8325	1.03%	1.81%
	α_2	1.00	0.0016	0.1607	-1.30%	0.04%
	α_3	1.00	0.0179	0.3435	0.46%	0.59%

Table 6: Monte Carlo Simulations for the $A_2(3)$ and $A_3(3)$ Models

This table reports the results of 5,000 Monte Carlo simulations for the $A_2(3)$ and $A_3(3)$ models 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 difference between the exact MLE $\hat{\theta}^{(MLE)}$ and the approximate MLE $\hat{\theta}^{(2)}$, based on the expansion with $K = 2$ terms. The results in the table show that the difference $\hat{\theta}^{(MLE)} - \hat{\theta}^{(2)}$ is much smaller, and often several orders of magnitude smaller, than the difference $\hat{\theta}^{(MLE)} - \theta^{(TRUE)}$ due to the sampling noise.

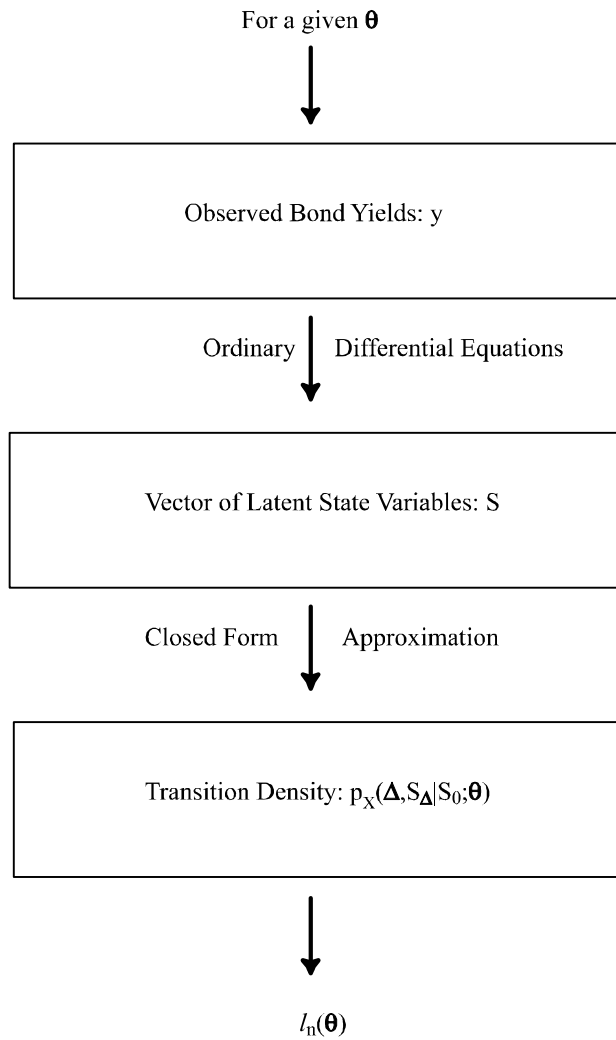


Figure 1: The Likelihood-Based Estimation Method

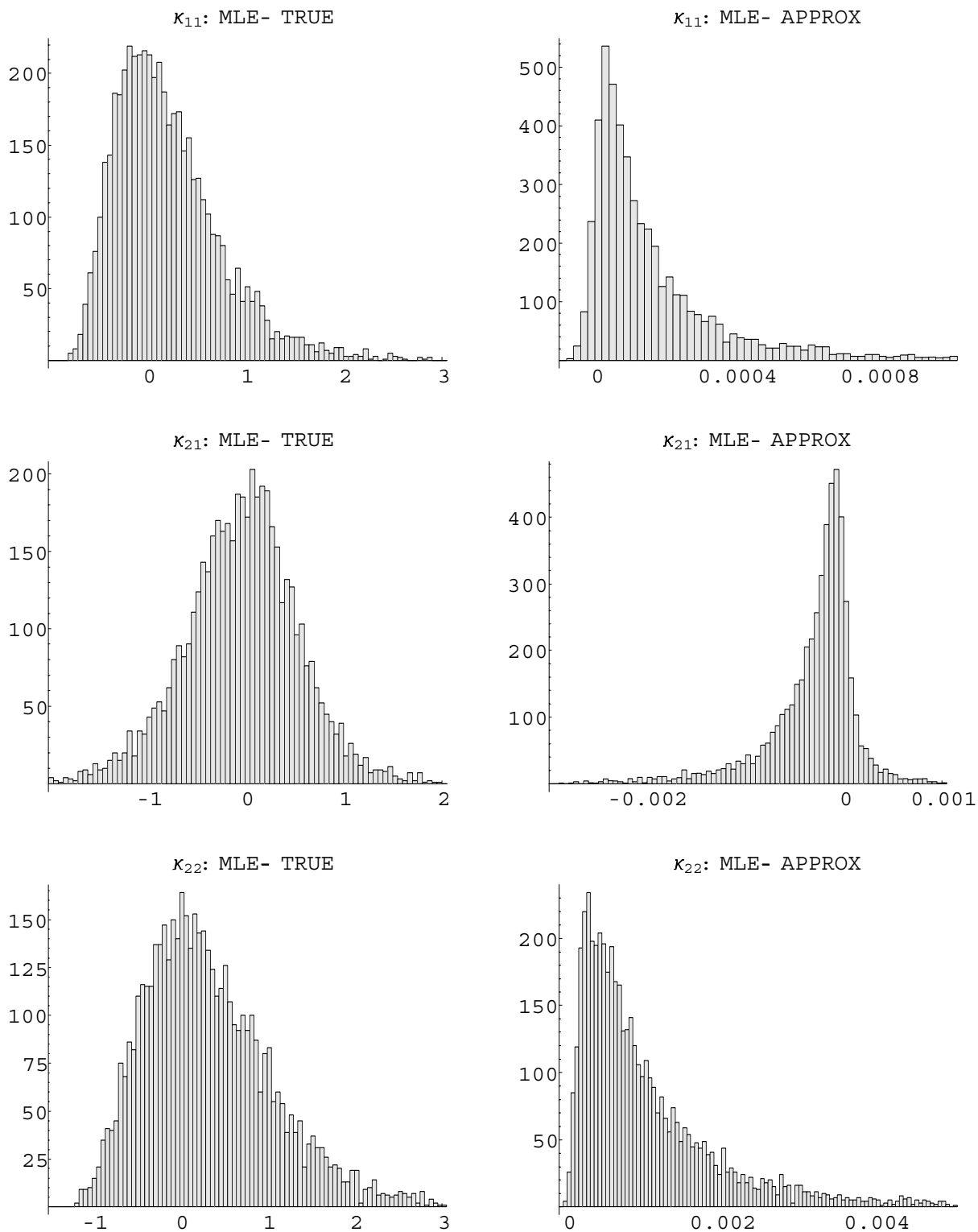


Figure 2: Empirical Distributions of the MLE Estimator (left column) and the Approximation Error (right column) in the $A_0(2)$ Model

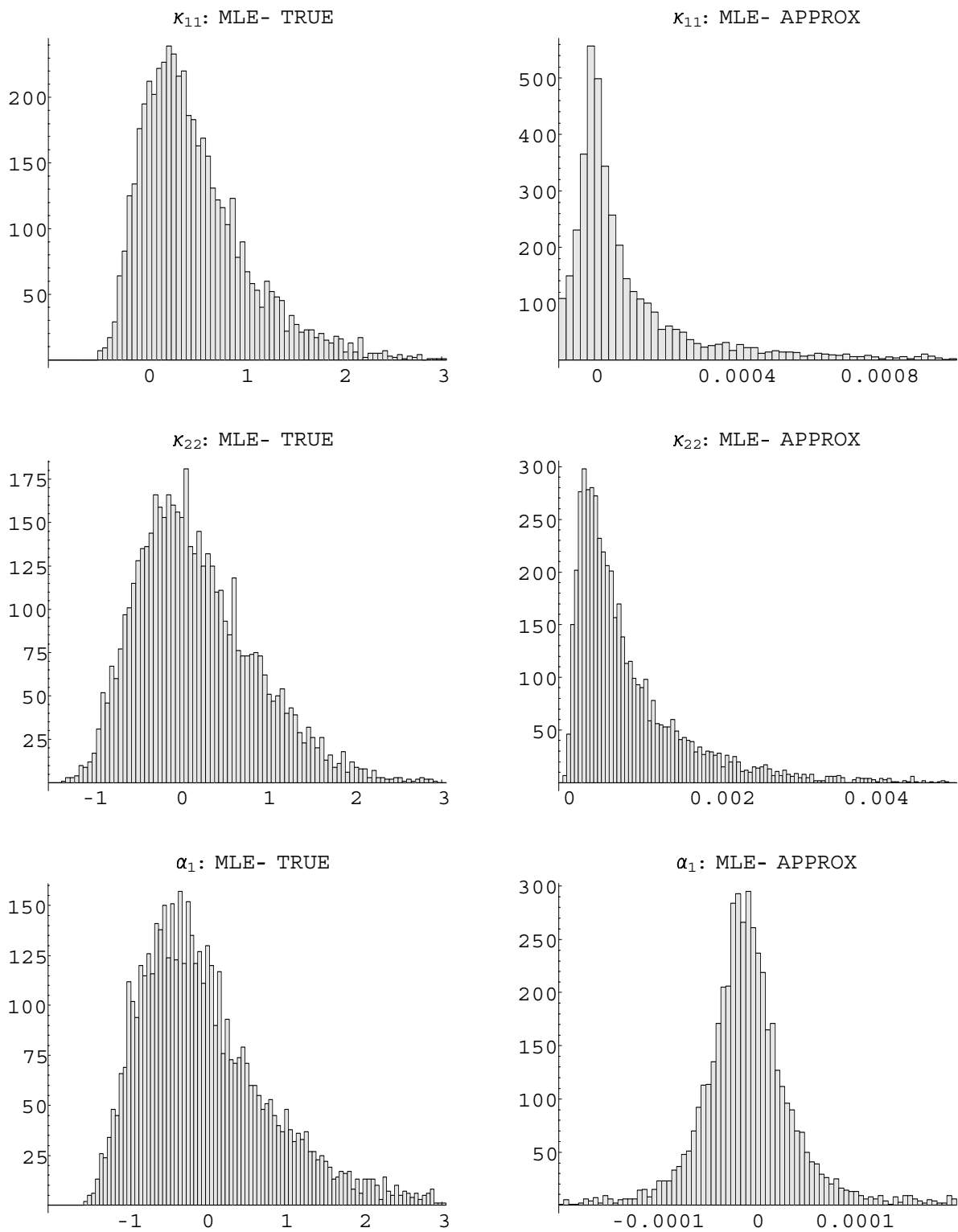


Figure 3: Empirical Distributions of the MLE Estimator (left column) and the Approximation Error (right column) in the $A_1(2)$ Model

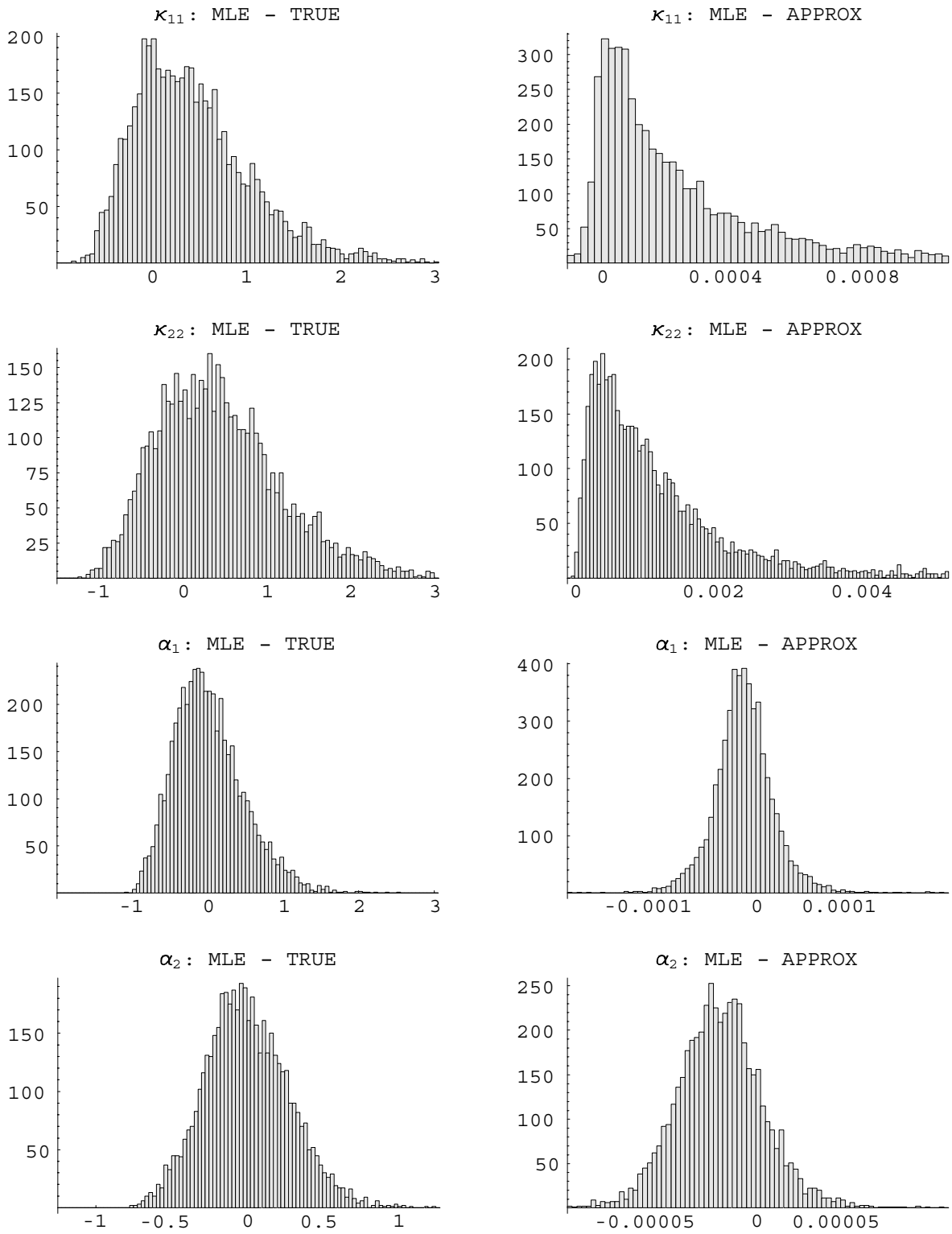


Figure 4: Empirical Distributions of the MLE Estimator (left column) and the Approximation Error (right column) in the $A_2(2)$ Model