

Moments of Affine Diffusions

Paul L. Fackler*

first draft: May 1, 2000

current revision: June 19, 2000

Abstract

Affine diffusion models are widely used in financial modeling; a general expression for the conditional covariance of an affine diffusion, however, has not appeared. This note provides such an expression, for both constant and time-varying parameter models. The conditional mean and covariance can be computed simultaneously by solving a system of linear first order differential equations. Computational strategies, including an eigenvalue/eigenvector decomposition, are discussed.

Keywords: Stochastic Processes, Affine Diffusions

*The author is an Associate Professor at North Carolina State University.

Mail: Department of Agricultural and Resource Economics

NCSU, Box 8109

Raleigh NC, 27695, USA

e-mail: paul_fackler@ncsu.edu

Web-site: <http://www4.ncsu.edu/~pfackler/>

© 2000, Paul L. Fackler

1 Introduction

Many parametric models for pricing financial assets are based on affine diffusions. Models for fixed income assets in this class include Vasicek and the one- and multi-factor models of Cox et al. (CIR) (e.g. Chen and Scott and Duan and Simonato) and general affine models of Duffie and Kan and Dai and Singleton. The models of futures price term structure in Schwartz are also included in the affine class, as are many of commonly used option pricing models, including the Gaussian (constant volatility) models of Black and Scholes and of Miltersen and Schwartz, and the stochastic volatility models of Hull and White and of Heston.

Despite their widespread use and the fact that it is widely recognized that both the conditional mean and variance are affine functions (with time-varying coefficients) of the current state, an explicit expression for the variance of future realizations has not appeared in the literature. Such an explicit expression is useful in a variety of ways. First, it makes it easier to characterize the effect of alternative parameter values on the behavior of the diffusion. Second, the variance is used in a variety of estimation strategies, including the maximum-likelihood and quasi-maximum likelihood methods described by Schwartz, Fisher and Gilles and Duan and Simonato. Third, it could be useful in Monte-Carlo simulations as a control variate (Judd).

This note provides expressions for the first two moments of affine diffusions. First, a general discussion of forward and backward methods for computing moments of dif-

fusions is presented. Then affine diffusions are defined and it is shown that $E[X_{t+h}|x_t]$ and $\text{Var}[X_{t+h}|x_t]$ solve an ordinary differential equation in h . This result, which has not previously appeared in the literature, provides a way of computing the first two moments of affine diffusions even when the parameters of the diffusion are time varying. This will facilitate modeling markets in which seasonality is important, such as energy, oil, and agricultural commodity markets. When the parameters of the diffusion are constant-in-time, explicit solutions are possible. The paper includes a discussion of computational considerations, including the use of generic differential equation solvers, eigenvalue/eigenvector decompositions and Padé approximations.

2 The Mean and Variance of a Diffusion

Suppose that an n -vector valued process, x is defined by

$$dx = \mu(x, t)dt + \sigma(x, t)dW.$$

There are, generically, two approaches that can be used to compute moments of stochastic differential equations. The forward approach fixes the initial time t and integrates forward with respect to the future time horizon, T . The backward approach fixes the future date, T , and integrates backwards with respect to t .

To find $E_t[x_T]$,¹ the forward approach uses the integral representation of the SDE

$$x_T = x_t + \int_t^T \mu(x_\tau, \tau)d\tau + \int_t^T \sigma(x_\tau, \tau)dW_\tau.$$

The diffusion term has expectation 0, so

$$\mathbb{E}_t[x_T] = x_t + \int_t^T \mathbb{E}_t[\mu(x_\tau, \tau)] d\tau$$

or, in differential form,

$$\frac{\partial \mathbb{E}_t[x_T]}{\partial T} = \mathbb{E}_t[\mu(x_T, T)]. \quad (2.1)$$

In contrast, the backward approach defines the function $y_t = \mathbb{E}_t[x_T] = Y(x_t, t)$.

By Ito's Lemma²

$$dy = \left[Y_t + Y_x \mu + \frac{1}{2} Y_{xx} \text{vec}(\sigma \sigma^\top) \right] dt + Y_x \sigma dW.$$

By the Law of Iterated Expectations, the drift on y must be 0; hence Y solves the partial differential equation

$$0 = Y_t + Y_x \mu + \frac{1}{2} Y_{xx} \text{vec}(\sigma \sigma^\top),$$

subject to the boundary condition that $Y(x_T, T) = x_T$. Unlike the forward approach, the backward approach provides a non-random functional equation to solve.

¹The notation $\mathbb{E}_t[z]$ is a shorthand for expectation of z , conditioned on information at time t . By the Markov property of Ito diffusions this implies that $\mathbb{E}_t[z] = \mathbb{E}[z|x_t]$ when z is measurable with respect to x .

² Y_{xx} is defined as an $n \times n^2$ matrix the i th row of which equals vectorized Hessian of the i th element of Y : $[Y_{xx}]_i = \text{vec}(\partial[Y]_i / \partial x \partial x^\top)^\top$. The fact that $\text{trace}(AB) = \text{vec}(A^\top)^\top \text{vec}(B)$ implies that the i th row of $[Y_{xx} \text{vec}(\sigma \sigma^\top)]_i = \text{trace}(\partial[Y]_i / \partial x \partial x^\top \sigma \sigma^\top)$.

To compute the second moment, the forward approach utilizes the function $u_t = U(x_t) = \text{vec}(x_t x_t^\top)$, the dynamics of which can be expressed as

$$du_t = \left[U_x(x_t)\mu(x_t, t) + U_{xx}\text{vec}\left(\sigma(x_t, t)\sigma(x_t, t)\right) \right] dt + \frac{1}{2}U_x(x_t)\sigma(x_t, t)dW.$$

Thus

$$\mathbb{E}_t[u_T] = \text{vec}(x_t x_t^\top) + \int_t^T \mathbb{E}_t \left[U_x(x_\tau)\mu(x_\tau, \tau) + \frac{1}{2}U_{xx}(x_\tau, \tau)\text{vec}\left(\sigma(x_\tau, \tau)\sigma^\top(x_\tau, \tau)\right) \right] d\tau.$$

The vectorized variance is therefore

$$\text{vec}(\text{Var}_t(x_T)) = \mathbb{E}_t[u_T] - \text{vec}\left(\mathbb{E}_t[x_T]\mathbb{E}_t[x_T]^\top\right).$$

The backward approach to obtaining an expression for the conditional variance uses the facts that

$$dy = Y_x \sigma dW$$

and $y_T = x_T$, which together imply that x_T can be written as the stochastic integral

$$x_T = y_t + \int_t^T Y_x(x_\tau, \tau)\sigma(x_\tau, \tau)dW_\tau,$$

the variance of which is

$$\text{Var}_t(x_T) = \mathbb{E}_t[(x_T - y_t)(x_T - y_t)^\top] = \mathbb{E}_t \left[\left[\int_t^T Y_x \sigma dW_\tau \right] \left[\int_t^T Y_x \sigma dW_\tau \right]^\top \right],$$

where Y_x and σ are evaluated at (x_τ, τ) .

This result may be simplified using the following lemma is used.

Lemma: Given two functions, $f(x_t, t)$ and $g(x_t, t)$,

$$\begin{aligned} \mathbb{E}_t \left[f(x_T, T)g(x_T, T)^\top \right] &= \mathbb{E}_t \left[\left[\int_t^T f(x_\tau, \tau) dW_\tau \right] \left[\int_t^T g(x_\tau, \tau) dW_\tau \right]^\top \right] \\ &= \mathbb{E}_t \left[\int_t^T f(x_\tau, \tau)g(x_\tau, \tau)^\top d\tau \right]. \end{aligned}$$

The lemma is a simple result of the definition of an Ito integral and appears in scalar form in numerous places, e.g., Neftci, p.191; for completeness a proof is provided in an appendix. Using the lemma, the conditional covariance is given by

$$\text{Var}_t(x_T) = \mathbb{E}_t \left[\int_t^T Y_x \sigma \sigma^\top Y_x^\top d\tau \right]. \quad (2.2)$$

3 Affine Diffusions

The affine diffusion model is defined by the stochastic differential equation

$$dx = (a + Ax)dt + C \text{diag} \left(\sqrt{b + Bx} \right) dW.$$

or, in integral representation,

$$x_{t+h} = x_t + \int_t^{t+h} (a + Ax_\tau) d\tau + \int_t^{t+h} C \text{diag} \left(\sqrt{b + Bx} \right) dW.$$

The parameters a and b (both $n \times 1$) and A , B and C (all $n \times n$) can be constants or functions of time (the latter are particularly useful for processes displaying seasonality).

Both the forward and backward approaches to computing the variance for the affine diffusion model have been discussed in the literature; specifically, Duan and Simonato used the forward and Fisher and Gilles used the backward approach. Neither of these papers, however, provided an explicit expression or an approach suitable for general computation.

The conditional expectation for affine diffusions is easily solved using the forward approach.

$$\frac{dE_t[x_{t+h}]}{dh} = a + AE_t[x_{t+h}] \quad (3.3)$$

is a first order linear differential equation in h with boundary condition x_t when $h = 0$.

It is well known that this is solved by an expression of the form

$$E_t[x_{t+h}] = m_0(t, h) + m_1(t, h)x_t, \quad (3.4)$$

where the $n \times 1$ and $n \times n$ coefficient matrices m_0 and m_1 satisfy the differential equations

$$\frac{\partial m_0(t, h)}{\partial h} = a(t + h) + A(t + h)m_0(t, h)$$

and

$$\frac{\partial m_1(t, h)}{\partial h} = A(t + h)m_1(t, h)$$

subject to the boundary conditions $m_0(t, 0) = 0$ and $m_1(t, 0) = I_n$. When the parameters are constants (in t) and A is non-singular, the expectation is

$$E_t[x_{t+h}] = m_0(h) + m_1(h)x_t = \left(e^{hA} - I_n\right)A^{-1}a + e^{hA}x_t.$$

The variance of an affine diffusion is more easily computed using the backward approach. Applying (2.2) with $Y_x(x_\tau) = m_1(\tau, t + h - \tau)$,

$$\begin{aligned} \text{Var}_t[x_{t+h}] &= \left[\int_t^{t+h} m_1(\tau, t+h-\tau) \sigma(x_\tau, \tau) \sigma(x_\tau, \tau)^\top m_1^\top(\tau, t+h-\tau) d\tau \right] \\ &= \int_t^{t+h} m_1(\tau, t+h-\tau) C \text{diag}(b + B E_t[x_\tau]) C^\top m_1^\top(\tau, t+h-\tau) d\tau, \end{aligned} \quad (3.5)$$

where C , b , and B are evaluated at $t + \tau$.

A few preliminary results are needed before computing this integral. First, the relationship between the vec and Kronecker product operators is³

$$\text{vec}(XYZ) = (Z^\top \otimes X) \text{vec}(Y).$$

The obvious product rule for derivatives of Kronecker products with respect to scalars applies, namely

$$\frac{dX(\tau) \otimes Y(\tau)}{d\tau} = \frac{dX(\tau)}{d\tau} \otimes Y(\tau) + X(\tau) \otimes \frac{dY(\tau)}{d\tau}.$$

The derivative of the matrix exponential is

$$\frac{de^{\tau A}}{d\tau} = A e^{\tau A} = e^{\tau A} A,$$

as is easily demonstrated using the series form $e^{\tau A} = \sum_{i=0}^{\infty} \frac{\tau^i}{i!} A^i$. Its integral, if A is non-singular, is

$$\int_0^h e^{\tau A} d\tau = A^{-1} (e^{hA} - I) = (e^{hA} - I) A^{-1}.$$

³A good source of background on the matrix operators used here is Dhrymes, chap 4.

It will be more convenient to work with the vectorized version of (3.5):

$$\text{vec}(\text{Var}_t[x_{t+h}]) = \int_t^{t+h} (m_1(\tau, t+h-\tau) \otimes m_1(\tau, t+h-\tau))(C \otimes C)D(b + BE_t[x_\tau])d\tau,$$

where D is an $n^2 \times n$ matrix such that $\text{vec}(\text{diag}(x)) = Dx$; specifically, $D_{ij} = 1$ if

$i = (j-1)n + j$ and 0 otherwise.⁴ This can be written in the form

$$\text{vec}(\text{Var}_t[x_{t+h}]) = V_0(t, h) + V_1(t, h)x_t,$$

where, using (3.4),

$$V_0(t, h) = \int_t^{t+h} (m_1(\tau, t+h-\tau) \otimes m_1(\tau, t+h-\tau))(C \otimes C)D(b + Bm_0(t, \tau-t))d\tau \quad (3.6)$$

and

$$V_1(t, h) = \int_t^{t+h} (m_1(\tau, t+h-\tau) \otimes m_1(\tau, t+h-\tau))(C \otimes C)DBm_1(t, \tau-t)d\tau. \quad (3.7)$$

This demonstrates the earlier mentioned results that the n^2 elements of the conditional covariance are affine in x_t .

These integrals can be differentiated with respect to h to derive two systems of differential equations for V_0 and V_1 . Noting that $m_1(t, 0) = I_n$ and that $\partial m_1(t, h)/\partial h = A(t+h)$, we have the following results

$$\frac{\partial V_0(t, h)}{\partial h} = (C \otimes C)Db + (C \otimes C)DBm_0(t, h) + (A \otimes I_n + I_n \otimes A)V_0(t, h)$$

⁴The term $(C \otimes C)D$ is a matrix the i th column of which is the Kronecker product of the i th column of C with itself: $[(C \otimes C)D]_{\cdot i} = C_{\cdot i} \otimes C_{\cdot i}$, (columnwise Kronecker product).

and

$$\frac{\partial V_1(t, h)}{\partial h} = (C \otimes C)DBm_1(t, h) + (A \otimes I_n + I_n \otimes A)V_1(t, h)$$

(note that the parameters a , A , b , B , and C are evaluated at $t + h$). When the parameters are constants in time m_0 , m_1 , V_0 and V_1 are functions of the time horizon h alone.

Summarizing, the mean and covariance are both affine in the current state:

$$\begin{pmatrix} \mathbb{E}_t[x_{t+h}] \\ \text{vec}(\text{Var}_t[x_{t+h}]) \end{pmatrix} = \begin{pmatrix} m_0 \\ V_0 \end{pmatrix} + \begin{pmatrix} m_1 \\ V_1 \end{pmatrix} x_t.$$

The coefficient matrices satisfy two systems of ordinary differential equations in h ,

$$\begin{pmatrix} m_0 \\ V_0 \end{pmatrix}' = \tilde{a} + \tilde{A} \begin{pmatrix} m_0 \\ V_0 \end{pmatrix} \tag{3.8}$$

and

$$\begin{pmatrix} m_1 \\ V_1 \end{pmatrix}' = \tilde{A} \begin{pmatrix} m_1 \\ V_1 \end{pmatrix}. \tag{3.9}$$

where

$$\tilde{a} = \begin{bmatrix} a \\ (C \otimes C)Db \end{bmatrix}$$

and

$$\tilde{A} = \begin{bmatrix} A & 0 \\ (C \otimes C)DB & (A \otimes I_n + I_n \otimes A) \end{bmatrix}.$$

The first system is $(n + n^2) \times 1$, the second is $(n + n^2) \times n$. The initial ($h = 0$) conditions are $m_1(0) = 0$, $V_0(0) = 0$, $m_1(0) = I_n$ and $V_1(0) = 0$.

When parameters of the process are constant in time and A is nonsingular, the complete model is solved by

$$\begin{pmatrix} m_0 \\ V_0 \end{pmatrix} = \left(e^{h\tilde{A}} - I \right) \tilde{A}^{-1} \tilde{a} \quad (3.10)$$

and

$$\begin{pmatrix} m_1 \\ V_1 \end{pmatrix} = e^{h\tilde{A}} \begin{bmatrix} I_n \\ 0 \end{bmatrix}. \quad (3.11)$$

For stationary processes (ones for which the eigenvalues of A are strictly negative) the moments of the long-run (invariant) distribution are immediately available by noting that, as $h \rightarrow \infty$, $e^{\tilde{A}h} \rightarrow 0$. Solving explicitly for \tilde{A}^{-1} yields

$$E[x] = -A^{-1}a$$

and

$$\text{vec}(\text{Var}(x)) = -\left[A \otimes I + I \otimes A \right]^{-1} (C \otimes C) D \left(b - BA^{-1}a \right).$$

4 Computational Considerations

From a computational point of view, the issue is how best to solve linear first order ODEs. Although this may seem to be a trivial matter, efficient and accurate computation may make a large difference if these expectations must be calculated many

times, as is the case when they are used in parameter estimation and in Monte Carlo simulations.

If the diffusion model has time varying coefficients, then the explicit solutions given in (3.10) and (3.11) are not applicable. It may be possible to find an explicit expression if the time-variation is nicely behaved, otherwise the ODEs will need to be solved using one of a variety of general techniques such as Runge-Kutta (Press et al., chap.13) or collocation (Judd).

When the coefficients are constant, however, other computational approaches are typically superior in both speed and accuracy. Moler and Loan discuss numerous strategies and conclude that no one method is best in all situations. For the problem at hand, the methods using the eigenvalue/eigenvector decomposition of A and Padé approximations appear to yield excellent results.

The first of these uses the eigenvalue/eigenvector decomposition $A = Z\Lambda Z^{-1}$, where Λ is a diagonal matrix with i th diagonal element, λ_i , equal to the i th eigenvalue of A . Problems can arise when A is singular (one or more eigenvalues equal to 0) and, more importantly, when A is defective (its eigenvector matrix is singular), implying that the decomposition does not exist. A necessary, but not sufficient condition is for A to have repeated eigenvalues. Even if A is not exactly defective, its eigenvector matrix may be poorly conditioned and computations based on the decomposition may be very inaccurate.

The eigenvalues of A determine the time scales of the speeds of mean reversion for the diffusion. If these are known to all be distinct then A can be assumed to be non-defective, and the eigenvalue/eigenvector decomposition will be efficient and accurate. The expressions for the first moment are straightforward:

$$m_0(h) = Z(e^{h\Lambda} - I_n)\Lambda^{-1}Z^{-1}a$$

and

$$m_1(h) = Ze^{h\Lambda}Z^{-1}.$$

The approach requires special treatment when any of the λ_i are near 0. Eigenvalues equal to 0 imply that the diffusion is non-stationary (but non-explosive) and are common in many models, including Heston's stochastic volatility model for stock prices and Schwartz's Gaussian model of futures price term structure. Any elements in the diagonal matrix $(e^{h\Lambda} - I_n)\Lambda^{-1}$ associated with 0 eigenvalues can be replaced by the limiting value of h .⁵

The second moment is a bit more cumbersome to deal with. Rather than attempting to use \tilde{A} ,⁶ the integral expression for V_1 (3.7) can be written as

$$(Z \otimes Z) \int_0^h e^{\tau(\Lambda \otimes I + I \otimes \Lambda)} (Z^{-1}C \otimes Z^{-1}C) DB e^{(h-\tau)\Lambda} d\tau Z^{-1} = (Z \otimes Z)q(h)Z^{-1}.$$

⁵The expression $(e^{\lambda h} - 1)/\lambda$ can generally only be computed accurately using floating point arithmetic when λ is larger than the square root of machine precision; approximately $1.5 * 10^{-8}$ using 32 bit (double precision) floating point numbers. Values below this should be set to h .

⁶ \tilde{A} is defective when A has a zero eigenvalue and hence cannot be used directly.

The vectorized value of $q(h)$ is

$$\text{vec}(q(h)) = \left[e^{h\Lambda \otimes I_{n^2}} \int_0^h e^{\tau(I \otimes (\Lambda \otimes I + I \otimes \Lambda) - \Lambda \otimes I_{n^2})} d\tau \right] \text{vec}((Z^{-1}C \otimes Z^{-1}C)DB).$$

This isolates the time varying part of the computation to the $n^3 \times n^3$ diagonal matrix in [], which can be computed with a triple loop; the generic diagonal element has the form

$$q_{ijk}^1 = e^{h\lambda_k} \int_0^h e^{\tau(\lambda_i + \lambda_j - \lambda_k)} d\tau = \begin{cases} e^{h\lambda_k} h & \text{if } \lambda_i + \lambda_j - \lambda_k = 0 \\ \frac{1}{\lambda_i + \lambda_j - \lambda_k} (e^{h(\lambda_i + \lambda_j)} - e^{h\lambda_k}) & \text{otherwise} \end{cases}$$

By similar means (3.6) can be used to compute V_0 :

$$\begin{aligned} V_0 &= (Z \otimes Z) \left[\int_0^h e^{\tau(\Lambda \otimes I + I \otimes \Lambda)} d\tau \right] (Z^{-1}C \otimes Z^{-1}C)Db \\ &\quad + (Z \otimes Z) \left[\int_0^h e^{\tau(\Lambda \otimes I + I \otimes \Lambda)} (Z^{-1}C \otimes Z^{-1}C)DBZ (e^{(h-\tau)\Lambda} - I) \Lambda^{-1} d\tau \right] Z^{-1}a. \end{aligned}$$

The first bracketed term is an $n^2 \times n^2$ diagonal matrix with generic element

$$q_{ij}^{00} = \int_0^h e^{\tau(\lambda_i + \lambda_j)} d\tau = \begin{cases} h & \text{if } \lambda_i + \lambda_j = 0 \\ \frac{1}{\lambda_i + \lambda_j} (e^{h(\lambda_i + \lambda_j)} - 1) & \text{otherwise} \end{cases}$$

The second bracketed term can be written in vectorized form as

$$\left[\int_0^h (e^{(h-\tau)\Lambda} - I) \Lambda^{-1} \otimes e^{\tau(\Lambda \otimes I + I \otimes \Lambda)} d\tau \right] \text{vec}((Z^{-1}C \otimes Z^{-1}C)DBZ).$$

This is an $n^3 \times n^3$ diagonal matrix, the typical element of which depends on whether $\lambda_k = 0$. For $\lambda_k = 0$ the value is

$$q_{ijk}^{01} = \int_0^h (h - \tau) e^{\tau(\lambda_i + \lambda_j)} d\tau = \begin{cases} \frac{h^2}{2} & \text{if } \lambda_i + \lambda_j = 0 \\ \frac{1}{\lambda_i + \lambda_j} \left(\frac{e^{h(\lambda_i + \lambda_j)} - 1}{\lambda_i + \lambda_j} - h \right) & \text{otherwise} \end{cases}$$

For $\lambda_k \neq 0$ the value is

$$q_{ijk}^{01} = \int_0^h \frac{e^{(h-\tau)\lambda_k} - 1}{\lambda_k} e^{\tau(\lambda_i + \lambda_j)} d\tau = \frac{1}{\lambda_k} (q_{ijk}^1 - q_{ij}^{00})$$

Coding details are available from the author. Note that only the eigenvectors and eigenvalues of A need be computed and the only inverse needed is Z^{-1} . Thus, although some complications arise in dealing with the possibility of zero eigenvalues, the method is computational efficient and accurate. Furthermore, it can be used to compute the mean and variance for multiple time horizons with little increase in computational time.

Another approach uses Padé (rational) approximations

$$f(x) \approx \frac{\sum_{i=0}^n a_i x^i}{\sum_{i=0}^n b_i x^i}.$$

to compute the matrix exponential and its integral.⁷ Padé approximations with rescaling and squaring are recommended by Moler and Van Loan as a useful general approach to computing matrix exponentials. A similar Padé approximation is easily developed for the integral of a matrix exponential as well and produces accurate results with no special coding required to handle defective and singular A . The Padé approximation uses the fact that

$$\int_0^t e^{sA} ds = \int_0^t \sum_{i=0}^{\infty} \frac{s^i X^i}{i!} ds = \sum_{i=0}^{\infty} \int_0^t \frac{s^i X^i}{i!} ds = t \sum_{i=0}^{\infty} \frac{t^i X^i}{(i+1)!}.$$

⁷Padé approximations and methods for computing their coefficients are discussed in Press et al., section 5.12.

The coefficients for order 6 approximations are given in Table 1.

It is a good practice to rescale the problem so tA has a norm less than 1 before computing the Padé approximations (see Moler and Van Loan for discussion). By dividing by the minimum power of 2 that results in a norm of less than 1, the rescaling can be undone by doubling operations. For e^{tA} the doubling uses $e^{2tA} = (e^{tA})^2$. For $\int_0^t e^{sA} ds$ the relationship

$$\int_0^t e^{2sA} ds = \int_0^t e^{sA} ds + e^{tA} \int_0^t e^{sA} ds$$

is used.

These relationships can also be used to compute moments for multiple time horizons by defining all required time horizons to be integer multiples of a time step Δ .

From (3.10) and (3.11), it can be readily verified that:

$$\begin{pmatrix} m_0(i\Delta) \\ V_0(i\Delta) \end{pmatrix} = \begin{pmatrix} m_0(\Delta) \\ V_0(\Delta) \end{pmatrix} + \begin{bmatrix} m_1(\Delta) & 0 \\ V_1(\Delta) & m_1(\Delta) \otimes m_1(\Delta) \end{bmatrix} \begin{pmatrix} m_0((i-1)\Delta) \\ V_0((i-1)\Delta) \end{pmatrix} \quad (4.12)$$

and

$$\begin{pmatrix} m_1(i\Delta) \\ V_1(i\Delta) \end{pmatrix} = \begin{bmatrix} m_1(\Delta) & 0 \\ V_1(\Delta) & m_1(\Delta) \otimes m_1(\Delta) \end{bmatrix} \begin{pmatrix} m_1((i-1)\Delta) \\ V_1((i-1)\Delta) \end{pmatrix}. \quad (4.13)$$

Thus once the solution is obtained for the time step Δ , the solution for integer multiples of Δ can be computed using a simple linear recursion.

It should also be pointed out that the presentation above, in terms of $\text{vec}(\text{Var}_t(x_{t+h}))$, is easily modified to compute $\text{vech}(\text{Var}_t(x_{t+h}))$, avoiding computation of symmetric

elements. This reduces the size of the ODEs in (3.8) and (3.9) considerably. For example, for $n = 3$, the ODE systems for z_0 and z_1 contain 12 and 36 elements, respectively, if vec is used and 9 and 18 elements if vech is used. To accomplish this, modify \tilde{a} and \tilde{A} as follows:

$$\tilde{a} = \begin{bmatrix} a \\ S(C \otimes C)Db \end{bmatrix}$$

and

$$\tilde{A} = \begin{bmatrix} A & 0 \\ S(C \otimes C)DB & S(A \otimes I_n + I_n \otimes A)S^i \end{bmatrix},$$

where $\text{vech}(x) = \text{Svec}(x)$ and, for symmetric x , $\text{vec}(x) = S^i \text{vech}(x)$. This will yield the values of $\text{vech}(\text{Var}_t[x_{t+h}])$. Similar reductions in computations are possible for the direct solution methods as well.

5 Concluding Comments

This note provides explicit expressions for the first two moments of an affine diffusion. In particular, it shows that these moments are affine in the current level of the process, with the coefficients of this affine function satisfying simple first order linear ODEs in the time horizon. Explicit solutions can be obtained by several means, which depend on the problem and goals. If the parameters of the process are constant-in-time a solution based on the eigenvalue/eigenvector decomposition is probably the

most efficient and accurate computational approach, especially when multiple time horizons are needed. This may fail in certain situations (defective A), in which case the solution can be computed using Padé approximations for matrix exponentials and their integrals. When the process has time-varying parameters, as would be the case in modeling prices displaying seasonality, the solution can be obtained using generic ODE solvers (e.g. Runge-Kutta or collocation).

Appendix

Lemma: Given two functions, $f(x_t, t)$ and $g(x_t, t)$,

$$\mathbb{E}_t \left[\left[\int_t^T f(x_\tau, \tau) dW_\tau \right] \left[\int_t^T g(x_\tau, \tau) dW_\tau \right]^\top \right] = \mathbb{E}_t \left[\int_t^T f(x_\tau, \tau) g(x_\tau, \tau)^\top d\tau \right].$$

Proof: The Ito integral can be written as the infinite sum:

$$\int_t^T f(x_\tau, \tau) dW_\tau = \lim_{n \rightarrow \infty} \sqrt{\frac{T}{n}} \sum_{i=0}^{n-1} f(x_{t+iT/n}, t + iT/n) e_i,$$

where the e_i are independent standard normal random vectors.

$$\begin{aligned} & \mathbb{E}_t \left[\left[\int_t^T f(x_\tau, \tau) dW_\tau \right] \left[\int_t^T g(x_\tau, \tau) dW_\tau \right]^\top \right] \\ &= \mathbb{E}_t \left[\left[\lim_{n \rightarrow \infty} \sqrt{\frac{T}{n}} \sum_{i=0}^{n-1} f(x_{t+iT/n}, t + iT/n) e_i \right] \left[\lim_{n \rightarrow \infty} \sqrt{\frac{T}{n}} \sum_{j=0}^{n-1} g(x_{t+jT/n}, t + jT/n) e_j \right]^\top \right] \\ &= \mathbb{E}_t \left[\lim_{n \rightarrow \infty} \frac{T}{n} \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} f(x_{t+iT/n}, t + iT/n) e_i e_j^\top g(x_{t+jT/n}, t + jT/n)^\top \right] \\ &= \mathbb{E}_t \left[\lim_{n \rightarrow \infty} \frac{T}{n} \sum_{i=0}^{n-1} f(x_{t+iT/n}, t + iT/n) g(x_{t+iT/n}, t + iT/n)^\top \right] \\ &= \mathbb{E}_t \left[\int_t^T f(x_\tau, \tau) g(x_\tau, \tau)^\top d\tau \right]. \end{aligned}$$

The proof uses the fact that $\mathbb{E}_t[e_i e_j] = 0$ for $i \neq j$ and equals I_n for $i = j$.

References

- Black, Fischer and Myron Scholes (1973): “The Pricing of Options and Corporate Liabilities,” *Journal of Political Economy*, 81, 637-654.
- Chen, Ren-Raw and Louis Scott (1995): “Interest Rate Options in Multi-Factor Cox-Ross-Rubinstein Models of the Term Structure,” *Journal of Derivatives*, 3, 53-72.
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross (1985): “A Theory of the Term Structure of Interest Rates,” *Econometrica*, 53, 385-407.
- Dai, Qiang and Kenneth J. Singleton (1999): “Specification Analysis for Affine Term Structure Models,” *Journal of Finance*, forthcoming.
- Dhrymes, Phoebus J (1978): *Mathematics for Econometrics*. New York, Springer-Verlag.
- Duan, Jin-Chuan and Jean-Guy Simonato (1999): “Estimating and Testing Exponential–Affine Term Structure Models by Kalman Filter,” *Review of Quantitative Finance and Accounting*, 13, 111-135.
- Duffie, Darrell and Rui Kan (1996): “A Yield-Factor Model of Interest Rates,” *Mathematical Finance*, 6, 379-406.
- Fisher, Mark and Christian Gilles (1996): “Estimating Exponential Affine Models of the Term Structure,” Federal Reserve Board, Washington, DC.
- Heston, Steven L (1993): “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options,” *Review of Financial Studies*, 6, 327-343.
- Hull, John and Alan White (1987): “The Pricing of Options on Assets with Stochastic Volatilities,” *Journal of Finance*, 42, 281-300.
- Judd, Kenneth L (1998): *Numerical Methods in Economics*. Cambridge, MA, MIT Press.
- Miltersen, Kristian R. and Eduardo S. Schwartz (1998): “Pricing of Options on Commodity Futures with Stochastic Term Structures of Convenience Yields and Interest Rates,” *Journal of Financial and Quantitative Analysis*, 33, 33-59.
- Moler, Cleve and Charles Van Loan (1978): “Nineteen Dubious Ways to Compute the Exponential of a Matrix,” *SIAM Review*, 20, 801-836.

- Neftci, Salih N (1996): *An Introduction to the Mathematics of Financial Derivatives*. San Diego, CA, Academic Press.
- Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery (1992): *Numerical Recipes*. Cambridge, Cambridge University Press.
- Schwartz, Eduardo S (1997): "The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging," *Journal of Finance*, 52, 923-973.
- Vasicek, Oldrich (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

Table 1. Padé Approximation Coefficients

i	0	1	2	3	4	5	6
e^{tA}							
a_i	1	1/2	5/44	1/66	1/792	1/15840	1/665280
b_i	1	-1/2	5/44	-1/66	1/792	-1/15840	1/665280
$\frac{1}{t} \int_0^t e^{sA} ds$							
a_i	1	1/26	5/156	1/858	1/5720	1/205920	1/8648640
b_i	1	-6/13	5/52	-5/429	1/1144	-1/25740	1/1235520

$$f(x) \approx \frac{\sum_{i=0}^n a_i x^i}{\sum_{i=0}^n b_i x^i}.$$