Specification Issues for Multivariate Affine Diffusion Models

Paul L. Fackler*

first draft: September 15, 1999
current revision: June 19, 2000

Abstract

Affine diffusion models provide a tractable framework for pricing many financial assets, including bonds, futures and European options. This paper extends and clarifies recent work on invariant parameter transformations, admissibility and identification of 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/

Multivariate pricing models are increasingly recognized as essential for adequately describing the behavior of financial derivatives. A particularly convenient class of models for the underlying factors is the class of affine diffusions. Many standard models for fixed income assets fall into this class, including that of Vasicek and the one- and multi-factor models of Cox et al. (CIR). Recently Duffie and Kan have developed an integrated approach to the use of multi-factor affine diffusions for interest rate modeling. 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 Black and Scholes, and the stochastic volatility models of Hull and White and of Heston.

The usefulness of the affine diffusion class derives in part from the ease with which a number of important financial derivatives can be priced. In general, pricing financial assets that depend on multiple factors requires solving multi–dimensional partial differential equations (PDEs). If the underlying factors are described by affine diffusions, prices of simple bonds, futures and European options can be characterized in terms of systems of ordinary differential equations (ODEs), which are far easier to solve. Furthermore, with some of the simpler models, an explicit solution to the system of ODEs is possible. This feature of affine diffusions provides a practical method to overcome the so-called curse of dimensionality, which has limited the number of underlying factors that could practically be used.

In this paper, several issues relating to the specification of affine diffusions are examined. First, the notion of invariant transformations of affine models is discussed. Second, necessary and sufficient conditions for an affine diffusion to be admissible are presented, along with an explicit algorithm to check admissibility. Third, the characteristics defining invariance classes of affine diffusions are discussed. The paper ends with some concluding comments followed by proofs of theorems and lemmas.

The paper draws on and extends work by Duffie and Kan and Dai and Singleton. Although it covers some of the same ground as these papers, it provides a more intuitive and detailed treatment of the issues. For example, the discussion of invariant transformations is simpler than that of Dai and Singleton and the nature of the impact of these transformation is clarified. Perhaps more importantly, the paper provides necessary and sufficient conditions and an explicit operational method for checking admissibility. It also provides a complete characterization of the properties of affine diffusions that are invariant to affine transformations.

Affine Diffusions

An affine diffusion is a (possibly multi-factor) diffusion in which the instantaneous mean and variance are both affine in the levels of the factors. In stochastic differential equation form they can be represented by

$$dx = [a + Ax]dt + C^\top \text{diag}\left(\sqrt{b + Bx}\right) dW, \tag{1}$$

where W is a n-vector of independent standard Weiner processes. Any affine diffusion can be represented in terms of the set of parameters $\theta = \{a, A, b, B, C\}$.

Critical to the analysis of affine diffusions is the n-vector of volatility processes

$$v(x) = b + Bx.$$

The instantaneous covariance of x, $C^\top \text{diag}(v(x))C$, is well defined (positive semi-definite) for all states in which $v(x) \geq 0$.[1]

This class of models is sufficiently rich to capture such features as mean reversion and stochastic volatility. Examples of such models include one factor models such as Vasicek's

$$dx = [\kappa m - \kappa x]dt + \sigma dW$$

and Cox, Ingersoll and Ross's

$$dx = [\kappa m - \kappa x]dt + \sigma\sqrt{x}dW,$$

both of which are mean reverting processes. Heston's two factor stochastic volatility model also falls into this class:

$$dx = \left(\begin{bmatrix} \mu \\ \kappa m \end{bmatrix} + \begin{bmatrix} 0 & -\frac{1}{2} \\ 0 & -\kappa \end{bmatrix} x\right) dt + \begin{bmatrix} 1 & 0 \\ \rho\sigma & \sqrt{1-\rho^2}\sigma \end{bmatrix}^\top \text{diag}\left(\sqrt{\begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} x}\right) dW$$

($b = 0$). Schwartz's models of commodity prices with stochastic convenience yield also are affine diffusions; for example, his two-factor model can be expressed as

$$dx = \left(\begin{bmatrix} \mu \\ \kappa m \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & -\kappa \end{bmatrix} x\right) dt + \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sqrt{1-\rho^2}\sigma_2 \end{bmatrix}^\top dW$$

($b = [1\ 1]^\top$ and $B = 0$). Multivariate versions of the CIR model are also increasingly common (e.g. Chen and Scott); these take A and B to be diagonal, $b = 0$ and $C = \mathrm{I}_n$.

[1] Duffie and Kan show that any form that allowed more flexible dependence of the covariance terms on x would violate positive definiteness with positive probability. Additional restrictions are discussed below.

Invariant Transformations

Issues of admissibility and identification of affine models are facilitated by introducing a class of transformations that preserve essential qualitative aspects of the diffusion but make it easier to work with and/or interpret.

Define two affine diffusions as members of the same equivalence class if they can be written as affine functions of one another. Thus, any process y defined by the affine transformation

$$y = \Lambda x + \lambda$$

where Λ is non-singular, is in the same equivalence class as x. The process y can be written as

$$\begin{aligned} dy &= \Lambda\left[a + A\Lambda^{-1}(y-\lambda)\right]dt + \Lambda C^{\top}\text{diag}\left(\sqrt{b + B\Lambda^{-1}(y-\lambda)}\right)dW \\ &= \left[\tilde{a} + \tilde{A}y\right]dt + \tilde{C}^{\top}\text{diag}\left(\sqrt{\tilde{b} + \tilde{B}y}\right)dW, \end{aligned}$$

where, by matching coefficients,

$$\begin{aligned} \tilde{A} &= \Lambda A \Lambda^{-1} \\ \tilde{a} &= \Lambda(a - A\Lambda^{-1}\lambda) \\ \tilde{b} &= b - B\Lambda^{-1}\lambda \\ \tilde{B} &= B\Lambda^{-1} \\ \tilde{C} &= C\Lambda^{\top}. \end{aligned} \tag{2}$$

A more subtle form of equivalence comes from that fact that parameters of the diffusion (covariance) term can be altered without altering the covariance itself. Such transformations, here termed covariance transformations, satisfy the relationship

$$\tilde{C}^{\top}\text{diag}\left(\tilde{b} + \tilde{B}x\right)\tilde{C} = C^{\top}\text{diag}\left(b + Bx\right)C.$$

In general, it can be assumed that C and $\tilde{C}$ are non-singular and so it is always possible to write $\tilde{C}$ as a linear transformation of C:

$$\tilde{C} = \Phi C.$$

For convenience, define the augmented matrix

$$B^a = [B \quad b].$$

A covariance transformation requires that[2]

$$C^{\top}\Phi^{\top}\text{diag}(\tilde{B}^a_{\cdot k})\Phi C = C^{\top}\text{diag}(B^a_{\cdot k})C,$$

[2]The notation $x_{\cdot i}$ indicates the ith column of x; $x_{i\cdot}$ indicates the ith row.

for all $k = 1, \ldots, n+1$. Because C and Φ are full rank, an equivalent condition is

$$\mathrm{diag}(\tilde{B}^a_{\cdot k}) = \Phi^{-\top}\mathrm{diag}(B^a_{\cdot k})\Phi^{-1}.$$

Breaking these conditions down,

$$[\Phi^{-1}]^\top_{\cdot i}\mathrm{diag}(B^a_{\cdot k})[\Phi^{-1}]_{\cdot j} = \begin{cases} \tilde{B}^a_{\cdot k} & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

The $i = j$ condition serves to define the transformed $\tilde{B}^a$ matrix:

$$\tilde{B}^a_{ij} = \sum_{k=1}^{n} [\Phi^{-1}]^2_{ki} B^a_{kj},$$

or, in matrix notation, $\tilde{B}^a = \Phi^* B^a$, where $\Phi^*_{ij} = [\Phi^{-1}]^2_{ji}$.

Thus the problem of identifying allowable covariance transformations is reduced to finding a nonsingular matrix, Φ, such that

$$\Phi^{-\top}\mathrm{diag}(B_{\cdot k})\Phi^{-1} \tag{3}$$

is diagonal for every k. Any Φ which consists of one and only one non-zero element per row (or column) satisfies this condition, regardless of the specific value of B^a. Such matrices are products of permutation matrices (the non-zero values all equal one) and diagonal (scaling) matrices.

The more difficult problem concerns whether any non-simple transformation exist in which one or more rows (columns) contain multiple non-zero elements. The following theorem discusses when such a transformation is possible.

Theorem 1 *If, for any k, i and j, $\Phi_{ki} \neq 0$ and $\Phi_{kj} \neq 0$, then $B_{i\cdot} \propto B_{j\cdot}$.*

Thus only if there are rows of B^a that are proportional to each other can there be more than one non-zero element in a row of Φ and, furthermore, any the nonzero elements must correspond to the rows of B^a that are mutually proportional. Mutually proportional rows of B^a correspond to situations in which some of the v_i are essentially equal to one another (a scaling transformation could be applied to ensure that the factors of proportionality all equal 1). Thus covariance transforms do not change the definitions of the v_i but merely scale and permute them.

Allowable forms of covariance transformation matrices can be specified more directly, as demonstrated in the following theorem.

Theorem 2 *Suppose be there are p sets of rows of B^a within each of which all there are p_i mutually proportional rows. Define the permutation matrix, $\hat{P}$, to group together proportional rows of B^a:*

$$\hat{P}B^a = \begin{bmatrix} \alpha_1\beta_1^\top \\ \alpha_2\beta_2^\top \\ \cdots \\ \alpha_p\beta_p^\top \end{bmatrix},$$

where the α_i are non-negative vectors of length p_i and the β_i are each n-vectors. Further let $\hat{D}^2$ be defined as the diagonal of the vertical stack of the α_i:

$$\hat{D}^2 = diag\left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \cdots \\ \alpha_p \end{bmatrix}\right)$$

Any covariance transformation can be expressed in the form:

$$\Phi = \hat{P}\hat{D}\begin{bmatrix} \Phi_1 & 0 & \dots & 0 \\ 0 & \Phi_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Phi_p \end{bmatrix} PD$$

where the Φ_i are arbitrary orthogonal matrices of dimensional $p_i \times p_i$ and where P and D are arbitrary permutation and diagonal (scaling) matrices.

An example of a covariance transform is a Φ equal to I_n except for elements ij and ji, where $B^a_{i.} = \alpha B^a_{j.}$. If element Φ_{ij} is selected arbitrarily to equal β then $\Phi_{ji} = -\alpha\beta$. A particularly useful transformation of this type is one that zeros elements of C. Suppose that $C_{ii} \neq 0$. Setting $\beta = \frac{C_{ji}}{\alpha C_{ii}}$ will cause $\tilde{C}_{ji}$ to equal 0. This transformation will be used later to normalize parameters.

Another useful transformation combines diagonal linear and covariance transforms to simultaneously scale C and B. Suppose elements C_{ii} and B_{ii} are both non-zero. They can be scaled to 1 by setting $\Phi_{ii}C_{ii}\Lambda_{ii} = 1$, and $\dfrac{B_{ii}}{\Phi_{ii}^2\Lambda_{ii}} = 1$. This is accomplished by setting $\Lambda_{ii} = \dfrac{1}{B_{ii}C_{ii}^2}$ and $\Phi_{ii} = \dfrac{1}{\sqrt{B_{ii}}C_{ii}}$.

In the Gaussian case ($B = 0$), all rows of B^a are, trivially, proportional to one another hence Φ can be defined by postmultiplying any orthogonal matrix by a diagonal matrix with its ith diagonal element equal to $\sqrt{b_i}$ if $b_i > 0$ and 1 if $b_i = 0$.

Summarizing, an affine diffusion can be transformed into another member of its equivalence class by defining an Λ, λ and Φ such that

$$\begin{aligned}
\tilde{A} &= \Lambda A \Lambda^{-1} \\
\tilde{a} &= \Lambda(a + A\Lambda^{-1}\lambda) \\
\tilde{C} &= \Phi C \Lambda^{\top} \\
\tilde{b} &= \Phi^*(b - B\Lambda^{-1}\lambda) \\
\tilde{B} &= \Phi^* B \Lambda^{-1}
\end{aligned} \tag{4}$$

where $[\Phi^*]_{ij} = [\Phi^{-1}]^2_{ji}$. Λ and λ are unrestricted but Φ must satisfy the restriction that

$$\Phi^{-\top}\text{diag}(B^a_{\cdot k})\Phi^{-1}$$

is diagonal for all $k = 1, \ldots, n+1$.[3]

It is useful to note that affine transforms applied in sequence can be expressed as a single transform:

$$y = \Lambda_2(\Lambda_1 x + \lambda_1) + \lambda_2 = \Lambda_2\Lambda_1 x + (\Lambda_2\lambda_1 + \lambda_2).$$

Similarly, sequential covariance transforms $\Phi = \Phi_2\Phi_1$ can be applied directly to alter C. It can also be shown that, for all k, $\Phi^* B^a_{\cdot k} = \Phi_2^*\Phi_1^* B^a_{\cdot k}$.[4]

Admissibility

The parameters of an affine diffusion (a, A, b, B, C) cannot be specified arbitrarily, but must satisfy certain restrictions to be admissible. Duffie and Kan show that the existence of a solution to the stochastic differential equation in (1) is ensured if the volatility process, $v(x) = b + Bx$, is non-negative with probability one. This, in turn,

[3] Dai and Singleton classify invariant transformations into four types. The presentation here clarifies the effect of allowable transformations on parameters and provides an operational approach to defining transformations. In particular, the form of allowable covariance transformations was not clear in their presentation.

[4] Let the operator $\text{Diag}(M)$ create a vector containing the diagonal elements of a square matrix M. Note that for any vector v, $\text{Diag}(\text{diag}(v)) = v$ and that for any diagonal matrix M, $\text{diag}(\text{Diag}(M)) = M$. The following steps outline the proof of the assertion in the text:

$$\begin{aligned}
\Phi^* B^a_{\cdot k} &= \text{Diag}\left(\Phi^{-\top}\text{diag}(B^a_{\cdot k})\Phi^{-1}\right) \\
&= \text{Diag}\left(\Phi_2^{-\top}\Phi_1^{-\top}\text{diag}(B^a_{\cdot k})\Phi_1^{-1}\Phi_2^{-1}\right) \\
&= \text{Diag}\left(\Phi_2^{-\top}\text{diag}\left(\text{Diag}\left(\Phi_1^{-\top}\text{Diag}(B^a_{\cdot k})\Phi_1^{-1}\right)\right)\Phi_2^{-1}\right) \\
&= \text{Diag}\left(\Phi_2^{-\top}\text{diag}(\Phi_1^* B^a_{\cdot k})\Phi_2^{-1}\right) \\
&= \Phi_2^*\Phi_1^* B^a_{\cdot k}.
\end{aligned}$$

is ensured if the parameters values are chosen so, for any x such that $v_i(x) = 0$, the drift associated with $v_i(x)$ is non-negative and its covariance is 0.

To facilitate discussion of the restrictions this places on parameters, define the feasible set to be

$$\mathcal{D}^x = \{x : b + Bx \geq 0\}.$$

Also define the set of points on the boundary of the feasible set for which $v_i(x) = 0$ by

$$\bar{\mathcal{D}}_i^x = \{x \in \mathcal{D}^x : b_i + B_{i\cdot}x = 0\},$$

(this set may by empty).

Noting that

$$dv = B(a + Ax)dt + BC^\top \text{diag}(\sqrt{b + Bx})dW,$$

admissibility is therefore demonstrated by showing that

$$\forall x \in \bar{\mathcal{D}}_i^x, \; B_{i\cdot}(a + Ax) \geq 0$$

and

$$\forall x \in \bar{\mathcal{D}}_i^x, \; B_{i\cdot}C^\top \text{diag}(b + Bx)CB_{i\cdot}^\top = B_{i\cdot}C^\top \text{diag}(CB_{i\cdot}^\top)(b + Bx) = 0.$$

The drift must maintain non-negativity, so certainly its minimum must be non-negative. Similarly, the covariance of each v_i is non-negative everywhere in the feasible set and so its maximum in $\bar{\mathcal{D}}_i^x$ must be zero. The sets $\bar{\mathcal{D}}_i^x$ are defined by linear constraints and both the drift and covariances are affine in x. Admissibility in affine models can, therefore, be checked using a series of linear programs:

$$\min_{x \in \bar{\mathcal{D}}_i^x} B_{i\cdot}(a + Ax) \geq 0$$

and

$$\max_{x \in \bar{\mathcal{D}}_i^x} B_{i\cdot}C^\top \text{diag}(CB_{i\cdot}^\top)(b + Bx) = 0,$$

for $i = 1, \ldots, n$.

Expressing the admissibility conditions in terms of a set of linear programs, however, does not provide any insight into how admissibility restricts the behavior of the process. An operational method of checking admissibility can be developed by exploiting the fact that the feasible domain for an admissible process can always be transformed to the domain $\mathbf{R}_+^m \otimes \mathbf{R}^{n-m}$ (shown below). Thus, there is a one-to-one mapping between feasible points for the original process and feasible points in this simple domain. Checking admissibility then amounts to checking a set of sign and zero restrictions on parameter values.

To see this, first transform the process in a way that isolates the set of variables involved in the volatility processes. Apply the QR decomposition (which is well defined for any matrix),

$$B = [R \;\; 0]Q^\top,$$

where R is $n \times m$ ($m = \text{rank}(B)$) and $Q = [Q_1 \;\; Q_2]$ is orthogonal. The linear transformation $\Lambda = Q^\top$ is then applied. The vector of volatility variables (v), under the transformed process $y = Q^\top x$ is

$$dv = RQ_1^\top(a + A[Q_1 \;\; Q_2]y)dt + RQ_1^\top C^\top \text{diag}(\sqrt{v})dW.$$

The volatility terms are influenced only by the first m variables in y. It is immediately clear that admissibility requires that the rest of the $n - m$ variables cannot affect the drift of the first m; otherwise it would not be possible for the drift in the volatilities to be correctly signed. Thus the first check for admissibility is that

$$Q_1^\top A Q_2 = 0.$$

It now suffices to work with the m-dimensional subspace defined by $z = Q_1^\top x$. The relevant feasible region is the set $D^z = \{z \in \mathbf{R}^m : b + Rz \geq 0\}$. Similarly, the set $\bar{\mathcal{D}}_i^z = \{z \in \mathcal{D}^z : b_i + R_{i\cdot}z = 0\}$ is the set of feasible points for which $v_i(z) = 0$.

Consider the corners of the feasible region, i.e., the feasible values of z such that m values of $b + Rz$ are equal to 0. Define ι to be an index of m integers from $\{1, \ldots, n\}$. Let R_ι be the $m \times m$ matrix and b_ι the m-vector formed from the rows indexed by ι. A corner is a point at which R_ι^{-1} exists and is defined by

$$z_\iota = -R_\iota^{-1} b_\iota.$$

A feasible corner is a corner in the feasible region:

$$b + Rz_\iota = b - RR_\iota^{-1} b_\iota \geq 0.$$

There are $\begin{pmatrix} n \\ m \end{pmatrix}$ possible choices for ι (the ordering of ι is irrelevant). There will be at least one ι for which R_ι is full rank (because $\text{rank}(R) = m$). For any such R_ι, z_ι is in the feasible set if $b + Rz_\iota \geq 0$. If this is never true, it must be the case that the feasible set is empty and hence the process is inadmissible.

The following theorem demonstrates that, for admissible processes, there is a unique feasible corner.

Theorem 3 *An admissible affine diffusion process has one and only one point, $\hat{z}$, such that m of the values of $b_i + R_{i\cdot}\hat{z} = 0$ and $b + R\hat{z} \geq 0$.*

The proof of this theorem relies on the condition that the variance of the $v_i(z)$ must go to 0 for all $z \in \bar{\mathcal{D}}_i^z$. If there are multiple feasible corners, the covariance of the $v_i(z)$ cannot be guaranteed to go to 0 for all z such that $v_i(z) = 0$. Hence, there is no way to ensure that the process remain in the feasible region.

Corollary 1 *The index, ι, associated with the unique feasible point satisfies $RR_\iota^{-1} \geq 0$.*

Corollary 2 *The feasible region, $\mathcal{D}^z$, is a convex polyhedral cone in R^m.*

The corollaries provide useful results that stem from the uniqueness of the feasible corner. The first provides a simple check for whether a feasible corner is the unique one and that degeneracy is not problematic.[5] One checks the set of possible combinations until a feasible corner with $RR_\iota^{-1} \geq 0$ is found. If no such feasible corner is found, the process is inadmissible.

The second corollary provides a way to visualize the feasible set geometrically. Furthermore, if a unique feasible corner exists, the feasible region may be transformed so the corner becomes the new origin and the zero curves indexed by ι become the new axes. Thus the feasible space for the transformed z is $\mathbf{R}_+^m$. This facilitates checking the remaining admissibility conditions.

For any feasible z, the drift and covariances associated with the rows indexed by ι must satisfy the requisite conditions:

$$R_{i\cdot}Q_1^\top(a + AQ_1 z_\iota) \geq 0$$

and that

$$R_{i\cdot}Q_1^\top C^\top \text{diag}(CQ_1 R_{i\cdot}^\top)(b + Rz_\iota) = 0,$$

$\forall i \in \iota$.

[5]It is possible that the feasible corner is degenerate, meaning it can be formed from multiple combinations. If so, the combination that results in the smallest feasible region should be selected. i.e., the one for which all of the values of RR_ι^{-1} are positive. This ensures that all non-negative points in a ball around the origin are feasible. Note that this use of the term degeneracy is consistent with the use of the term in the linear programming literature but is different from the meaning used by Duffie and Kan.

The existence of an appropriate permutation is guaranteed by admissibility; consequently, necessary and sufficient conditions for admissibility amount to determining the admissibility of the transformed process. Denoting this process as y,

$$dy = \begin{bmatrix} (P_1 R) Q_1^\top \\ Q_2 \end{bmatrix} \left(a + A \left[(P_1 R) Q_1^\top \;\; Q_2^\top \right] y \right) dt + \begin{bmatrix} (P_1 R) Q_1^\top \\ Q_2 \end{bmatrix} C^\top \mathrm{diag}\left(\sqrt{\tilde{b} + \tilde{B} y} \right) dz.$$

The covariance of the transformed process remains non-negative as long as the first m variables in y remain non-negative.

Two conditions are necessary and sufficient for each of the y_i $(i \le m)$ to remain non-negative with probability one. First, the drift on each y_i $(i \le m)$ must be non-negative whenever $y_i = 0$. Already discussed is the requirement that the $m \times n - m$ block $\tilde{A}_{12}$ be identically zero (i.e., that $Q_1 A Q_2 = 0$); the unboundedness of the y_i for $i > m$ requires this. Similarly, because the y_i, for $i \le m$, can achieve arbitrary values in $\mathbf{R}_+^m$, the sign restrictions $\tilde{a}_1 \ge 0$ and $\tilde{A}_{11} \ge 0$ (except on the diagonal) must hold. These restrictions are necessary and collectively they are sufficient for the drift on y_i $(i \le m)$ to be non-negative when $y_i = 0$.

Second, the instantaneous variance on y_i $(i \le m)$ must go to zero when y_i goes to zero. To examine the implications of this, write out the variance of y_i:

$$\begin{aligned} \mathrm{Var}(y_i) &= \sum_{k=1}^{m} \tilde{C}_{ki}^2 y_k + \sum_{j=m+1}^{n} \tilde{C}_{ji}^2 \left(\tilde{b}_j + \sum_{k=1}^{m} \tilde{B}_{jk} y_k \right) \\ &= \sum_{k=1}^{m} \left(\tilde{C}_{ki}^2 + \sum_{j=m+1}^{n} \tilde{C}_{ji}^2 \tilde{B}_{jk} \right) y_k + \sum_{j=m+1}^{n} \tilde{C}_{ji}^2 \tilde{b}_j . \end{aligned}$$

The constant term and the coefficients on the y_k for $k \ne i$ must all be zero. Given the already stated non-negativity of $\tilde{b}$ and $\tilde{B}$, these conditions are, for all $i, k \le m$ $(i \ne k)$ and $j > m$, $\tilde{C}_{ki} = 0$, $\tilde{C}_{ji} \tilde{B}_{jk} = 0$ and $\tilde{C}_{ji} \tilde{b}_j = 0$.

Summarizing, the following transformations are possible for all admissible affine processes:

$$\begin{aligned} \tilde{a} &= \begin{bmatrix} R_\iota Q_1^\top \\ Q_2 \end{bmatrix} (a + A Q_1 R_\iota^{-1} b_\iota) \\ \tilde{A} &= \begin{bmatrix} R_\iota Q_1^\top \\ Q_2 \end{bmatrix} A [Q_1 R_\iota^{-1} \;\; Q_2] \\ \tilde{b} &= P_\iota (b - R R_\iota^{-1} b_\iota) \\ \tilde{B} &= P_\iota R R_\iota^{-1} \\ \tilde{C} &= P_\iota C [Q_1 R_\iota \;\; Q_2] \end{aligned} \quad (5)$$

Necessary and sufficient conditions for the admissibility of an affine diffusion can be expressed in terms of these parameters:

Theorem 4 *An affine diffusion with $B = RQ_1$, where R is $n \times m$ and $Q = [Q_1 \;\; Q_2]$ is orthogonal, is admissible if and only if:*

1. *$Q_1 A Q_2 = 0$.*
2. *There exists a unique m-element index, ι, of row indices of R such that R_ι^{-1} exists, $b - RR_\iota^{-1} b_\iota \geq 0$, and $RR_\iota^{-1} \geq 0$.*
3. *Using ι to determine the appropriate transform the following parameter restrictions must hold:*

$$\begin{aligned} \tilde{a}_i &\geq 0 \text{ for } i \leq m \\ \tilde{A}_{ik} &\geq 0 \text{ for } i, k \leq m, \;\; i \neq k \\ \tilde{C}_{ki} &= 0 \text{ for } i, k \leq m, \;\; i \neq k \\ \tilde{C}_{ji}\tilde{B}_{jk} &= 0 \text{ for } i, k \leq m, \;\; i \neq k, \;\; j > m \\ \tilde{C}_{ji}\tilde{b}_j &= 0 \text{ for } i \leq m, \;\; j > m. \end{aligned}$$

An admissibility check is thus easily implemented by performing a QR decomposition and checking condition (1), checking (2) by searching over the $n!/(m!(n-m)!)$ possible ι, checking the sign conditions on the drift and the zero restrictions on the diffusion parameters given in (3). MATLAB code implementing the approach is available on the author's website.[6]

In the Gaussian case ($m = 0$), admissibility is always ensured because the state space is $\mathbf{R}^n$. In the $m = n$ case, there is only one possible index vector ι and hence the initial transformation is effected by $\Lambda = B$ and $\lambda = b$. Admissibility requires that $B(a - AB^{-1}b) \geq 0$, that the off–diagonal elements of BAB^{-1} are non-negative and that $CB^\top$ must be diagonal.

Here it has only been assumed that $\tilde{a}_i \geq 0$ for $i \leq m$. This ensures that the process is admissible but does not rule out absorbing boundaries. Duffie and Kan and Dai and Singleton demonstrate that

$$0 < \tilde{a}_i \leq \tfrac{1}{2}\tilde{C}_{\cdot i}^\top \tilde{C}_{\cdot i}$$

ensures that $y_i = 0$ ($i \leq m$) is a reflecting boundary and that

$$\tilde{a}_i > \tfrac{1}{2}\tilde{C}_{\cdot i}^\top \tilde{C}_{\cdot i}$$

[6] These conditions are similar to those given in appendix A.2 of Dai and Singleton. The main substantive difference is their assumption that $C_{ji} = 0$ for $i \leq m, \; j > m$. Furthermore, they only claimed sufficiency for their conditions; the conditions given here are necessary and sufficient.

ensures that it is an entrance boundary (i.e., that, almost surely, it cannot be reached from the interior of the feasible region).

Invariant Properties of Affine Diffusions

A useful feature of affine diffusions is the flexibility one has in interpreting the factors. For example, in bond pricing examples, some models take the factors to be yields on bonds of different maturities, whereas others interpret the factors as the short rate process and time-varying components if it, such as its mean and volatility. Schwartz and Smith pointed out the equivalence of two models of futures prices, one with the factors interpreted as the (unobserved) spot price and its convenience yield, the other with factors interpreted as the short and long run mean price.

It is therefore useful to ask what properties are invariant to alternative interpretations of the factors. Such alternative interpretations arise via performing invariant transformations and hence an equivalent question is what are the properties of invariance classes.

One way to address this question is to identify a canonical representation for each equivalence class and study its properties. Dai and Singleton suggest such a canonical model but it is not general. A more complete treatment is provided in this section.

The attempt to determine a minimal parameterized model is especially useful in determining what parameters are identified econometrically. In particular, any parameter restrictions that are not violated by covariance transformations are over-parameterized. In addition, a canonical model that is guaranteed to be admissible for all parameter values that meet simple sign restrictions facilitates both model specification and estimation.

Some of the properties of an invariance class can be obtained immediately by examining the forms of the transformations given in (4). One that has been utilized extensively already is the rank of B. Given the uniqueness of the feasible corner, the decomposition of the model into m non-negative pure volatility factors and its $n-m$ unbounded other factors is an invariant feature.

The rank of A is also invariant to transformation. Furthermore, the transformation $\Lambda A \Lambda^{-1}$ is a similarity transform and hence the eigenvalues of A are invariant. One can go further. Given the uniqueness of the decomposition into pure volatility factors and all other factors, the invariant eigenvalues of A can be partitioned into those associated with the pure volatility factors and those associated with the other factors (the eigenvalues of $Q_1^\top A Q_1$ and $Q_2^\top A Q_2$). This implies that the qualitative behavior of future expected time paths of the two sets of factors is invariant to transformation.

The sign of the eigenvalues determine the stationarity properties of the process. Perhaps obviously, the invariance of the eigenvalues implies that a non-stationary model cannot be transformed into a stationary one and vice versa. In addition, the eigenvalues determine the time scales over which shocks influence the behavior of the system.

One final invariant property concerns the topological features of the feasible region. Invariant transformations cannot alter the number of non-proportional rows of B^a or the number of rows in each class. Suppose there are p sets of rows that are mutually proportional and, for convenience, the rows of B^a are permuted so the first n_1 are mutually proportional, the second n_2 are mutually proportional, etc. (with the n_i summing to n). Admissible covariance transforms must be block diagonal, with the size of the blocks corresponding to the n_i. Thus, the p values of the n_i are invariant features of the diffusion.

One implication is that the degeneracy of the unique feasible corner is an invariant feature. Intuitively, this determines the number of the v_i that are bounded away from 0. As will be shown below, Dai and Singleton's canonical model makes the implicit assumption that the unique corner is non-degenerate.

From the model used to check admissibility it is easy to transform the model so the C matrix is an identity matrix. This is accomplished by first zeroing the elements of C_{21}, then the elements of C_{12}, and then transforming C_{22} to an identity matrix. Finally, the diagonal elements of C_{11} are normalized to 1.

If element ji in the lower left block of C (C_{21}) is non-zero it must be the case that $B^a_{j\cdot} = \alpha B^a_{i\cdot}$ for some positive α. The covariance transform used to zero elements of C (discussed on page 6) can be applied to each of these non-zero elements. This has no effect on b but might change the diagonal elements of C_{11} and B_{11}. These will be renormalized to identity matrices below.

C_{12} and C_{22} can be normalized with a linear transformation:

$$\hat{y} = \begin{bmatrix} I_m & 0 \\ -\left[\tilde{C}_{22} - \tilde{C}_{21}\tilde{C}_{11}^{-1}\tilde{C}_{12}\right]^{-\top} \tilde{C}_{12}^{\top}\tilde{C}_{11}^{-1} & \left[\tilde{C}_{22} - \tilde{C}_{21}\tilde{C}_{11}^{-1}\tilde{C}_{12}\right]^{-\top} \end{bmatrix} y.$$

This is well defined given the nonsingularity of both C and C_{11}. It serves to make $\tilde{C}_{22} = I_{n-m}$ and $\tilde{C}_{12} = 0$. This transformation alters $\tilde{a}_2$, $\tilde{A}_{21}$ and $\tilde{A}_{22}$, but leaves unchanged $\tilde{a}_1$ $\tilde{A}_{11}$, $\tilde{C}_{11}$, $\tilde{C}_{21}$, $\tilde{b}_2$ and $\tilde{B}_{21}$. It thus normalizes the last $n-m$ variables (the ones that do not enter into the volatility terms) and makes the entire set of variables instantaneously uncorrelated.

Finally, the scale the model can be normalized by making $C_{11} = B_{11} = I_m$ and any non-zero values of b_2 equal to 1, using diagonal matrices Λ and Φ such that $\Phi_{ii}C_{ii}\Lambda_{ii} = 1$, for all i, $\dfrac{B_{ii}}{\Phi_{ii}^2\Lambda_{ii}} = 1$, for $i \leq m$ and $\dfrac{b_i}{\Phi_{ii}^2} = 1$ for any $i > m$ such that $b_i \neq 0$. This is accomplished by setting $\Lambda_{ii} = \dfrac{1}{B_{ii}C_{ii}^2}$ and $\Phi_{ii} = \dfrac{1}{\sqrt{B_{ii}}C_{ii}}$ for $i \leq m$,

$\Lambda_{ii} = \dfrac{1}{\sqrt{b_i}C_{ii}}$ and $\Phi_{ii} = \sqrt{b_i}$ for $i > m$ and $b_i \neq 0$ and $\Lambda_{ii} = \Phi_{ii} = 1$ for $i > m$ and $b_i = 0$.

Thus any admissible model can be obtained as an affine transformation of a model of the form

$$dy = \left(\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} y \right) dt + \text{diag}\left(\sqrt{\begin{bmatrix} 0 \\ b_2 \end{bmatrix} + \begin{bmatrix} I_m & 0 \\ B_{21} & 0 \end{bmatrix} y} \right) dz, \qquad (6)$$

where b_2 is composed of 0s and 1s.

If A_{22} is full rank, a_2 can be normalized to 0 by setting $v_2 = A_{22}^{-1}a_2$. More generally, for some $n-m \times n-m$ permutation matrix, $P = [P_1 \;\; P_2]$, we can write

$$P^\top A_{22} P = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} q_1^\top,$$

where r_2 is a full rank $p \times p$ matrix and q_1 is $n-m \times p$ and orthogonal. The affine transformation with

$$\Lambda = \begin{bmatrix} I_m & 0 \\ 0 & P^\top \end{bmatrix} \text{ and } \lambda = \begin{bmatrix} 0 \\ q_1 r_2^{-1} P_2^\top a_2 \end{bmatrix}$$

causes the last p values of a to equal 0. Further affine transformations are possible that set $\lambda_2 = q_2\gamma$, where $q_1^\top q_2 = 0$ and γ is an arbitrary $n-m-p$ vector. Such transformations have no affect on the parameters of the stochastic differential equation that defines the process, although they would affect the levels of the variables (or, equivalently, their initial values). Models with $\text{rank}(A_{22}) = p < n-m$ exhibit $n-m-p$ sources of non-stationarity in their drifts; this is roughly analogous to saying that the last $n-m$ variables (the non-volatility variables) exhibit p cointegrating relationships.

It remains to normalize the scale of the v_i terms, for $i > m$. There does not appear to be a generally satisfactory way to do this. One normalization that works for any model is to scale these terms so $\max_j B^a_{ij} = 1$. This approach can cause difficulties for econometric estimation of model parameters, as it implies a complicated constraint on parameter values that would impose difficult to impose on parameters estimates. An alternative approach is to specify which of the b_2 are non-zero and normalize these to 1. This still leaves the possibility of further scaling transformations, however, when the feasible corner is degenerate.

The problem here is that there is not a simple way to characterize invariance classes for affine models. Recall that an invariance class is defined as a set of affine diffusions that are affine transforms of one another. Invariant properties of affine models take several forms. The rank of B is the most important, but the ranks of $Q_1^\top AQ_1$ and $Q_2^\top AQ_2$ are also invariant. More specifically, the transform applied to A is a similarity transform ($\tilde{A} = \Lambda A \Lambda^{-1}$) and hence the eigenvalues of A are an invariant property of an affine diffusion. This is sensible, as the eigenvalues determine the qualitative properties of the expected path of the process. In addition, the number of unique v_i and how many of each there are is an invariant property. A different minimally parameterized model is required for each. The definitions of the v_i define the nature of the feasible region; the topological characteristics of this region are not altered by affine transformations.

The canonical model of Dai and Singleton as a special case of (6) with b_2 equal to a vector of ones and $[A_{21}\ A_{22}]a = 0$.[7] They implicitly assumed non-degeneracy, thus ensuring that b_2 is bounded away from 0. They also assumed stationarity (rank$(A) = n$). This allows them to normalize a_2 to any desired values. Their normalization ensures that the long-run means of the last $n - m$ variables are equal to 0. It should be noted, however, that the imposition of stationarity rules out certain models in the literature, notably the models of futures price term structure in Schwartz.

One use of minimally parameterized models is that it allows a count of the number of free parameters needed to generate any model in a invariance class. For an n variable model with rank$(B) = m$, the number of free parameters in the full rank A, $b_2 > 0$ case is equal to $n^2 + m(m + 1)$. This provides a necessary condition for a minimally parameterized model, and any model with more free parameters is over-parameterized.

Concluding Comments

With the increased use of multi-factor asset pricing models based on affine diffusions, it is important to specify models that are well defined and can be estimated from data. This paper develops an operational method for checking admissibility of any affine diffusion. It also discusses the problem of model over–parameterization, especially with respect to equivalent representations of the covariance parameters. The operationalization of both the admissibility check and the form of allowable parameter transformations provides analysts employing affine diffusion models an important set of tools to facilitate model specification.

[7] The canonical model of Dai and Singleton is written in terms of $\theta = Aa$, with θ_2 normalized to equal 0.

References

Black, Fischer and Myron Scholes. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81(1973):637–654.

Chen, Ren-Raw and Louis Scott. "Pricing Interest Rate Options in a Two-Factor Cox-Ingersoll-Ross Model of the Term Structure." *Review of Financial Studies* 5(1992):613–636.

Chen, Ren-Raw and Louis Scott. "Maximum-Likelihood Estimation of a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates." *Journal of Fixed Income* 3(1993):14–31.

Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross. "A Theory of the Term Structure of Interest Rates." *Econometrica* 53(1985):385–407.

Dai, Qiang and Kenneth J. Singleton. "Specification Analysis for Affine Term Structure Models." *Journal of Finance* 55(2000):forthcoming.

Duffie, Darrell and Rui Kan. "A Yield-Factor Model of Interest Rates." *Mathematical Finance* 6(1996):379–406.

Heston, Steven L. "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options." *Review of Financial Studies* 6(1993):327–343.

Hull, John and Alan White. "The Pricing of Options on Assets with Stochastic Volatilities." *Journal of Finance* 42(1987):281–300.

Schwartz, E.S. and J.E. Smith. "Short-Term Variations and Long-Term Dynamics in Commodity Prices." The John E. Anderson Graduate School of Management at UCLA, Working Paper No.14-97, 1997.

Schwartz, Eduardo S. "The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging." *Journal of Finance* 52(1997):923–973.

Vasicek, Oldrich. "An Equilibrium Characterization of the Term Structure." *Journal of Financial Economics* 5(1977):177–188.

Appendix: Proofs

Theorem 1 *If, for any k, i and j, $\Phi_{ki} \neq 0$ and $\Phi_{kj} \neq 0$, then $B_{i\cdot} \propto B_{j\cdot}$.*

Assume (without loss of generality) that the first column of Φ^{-1} contains non-zero elements in it first p rows (one can always assure this by performing permutation transformations). The diagonality of (3) requires that

$$\left[[\Phi^{-1}]_{\cdot 2} \ \dots \ [\Phi^{-1}]_{\cdot n}\right]^\top \mathrm{diag}(B^a_{\cdot k})[\Phi^{-1}]_{\cdot 1} = 0$$

for $k = 1, \dots, n+1$. In other words, the $n+1$ vectors $\mathrm{diag}(B^a_{\cdot k})[\Phi^{-1}]_{\cdot 1}$ all lie in the column null space of

$$\left[[\Phi^{-1}]_{\cdot 2} \ \dots \ [\Phi^{-1}]_{\cdot n}\right].$$

Because Φ is non-singular, this null space is one-dimensional and hence there must exist a set of scalar constants, c_k, such that

$$\mathrm{diag}(b)[\Phi^{-1}]_{\cdot 1} = c_1 \mathrm{diag}(B_{\cdot 1})[\Phi^{-1}]_{\cdot 1} = \dots = c_m \mathrm{diag}(B_{\cdot m})[\Phi^{-1}]_{\cdot 1}.$$

Of course, the conditions hold generically for any row j with $[\Phi^{-1}]_{j1} = 0$. We need to focus, therefore, only on the first p rows of $[\Phi^{-1}]_{\cdot 1}$. The covariance restrictions require that there exist c_k such that

$$b_j = c_1 B_{j1} = \dots = c_n B_{jn}$$

for $j \leq p$. Thus the first p rows of B^a must be proportional.

Theorem 2 *Suppose be there are p sets of rows of B^a within each of which all there are p_i mutually proportional rows. Define the permutation matrix, $\hat{P}$, to group together proportional rows of B^a:*

$$\hat{P} B^a = \begin{bmatrix} \alpha_1 \beta_1^\top \\ \alpha_2 \beta_2^\top \\ \dots \\ \alpha_p \beta_p^\top \end{bmatrix},$$

where the α_i are non-negative vectors of length p_i and the β_i are each n-vectors. Further let $\hat{D}^2$ be defined as the diagonal of the vertical stack of the α_i:

$$\hat{D}^2 = diag\left(\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \dots \\ \alpha_p \end{bmatrix}\right)$$

Any covariance transformation can be expressed in the form:

$$\Phi = \hat{P}\hat{D} \begin{bmatrix} \Phi_1 & 0 & \dots & 0 \\ 0 & \Phi_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \Phi_p \end{bmatrix} PD$$

where the Φ_i are arbitrary orthogonal matrices of dimensional $p_i \times p_i$ and where P and D are arbitrary permutation and diagonal (scaling) matrices.

[TO BE COMPLETED]

Theorem 3 *An admissible affine diffusion process has one and only one point, $\hat{z}$, such that m of the values of $b_i + R_{i\cdot}\hat{z} = 0$ and $b + R\hat{z} \geq 0$.*

It has been shown that admissible processes have at least one feasible corner. Let ι be the index associated with this corner and apply to z the linear transformation $\Lambda = R_\iota$ and $\lambda = b_\iota$. If the corner is degenerate pick ι such that $R_{i\cdot}R_\iota^{-1}$ for each i with $b_i - R_{i\cdot}R_\iota^{-1}b_\iota = 0$. This ensures that all nonnegative values in a ball around the origin in the transformed space are feasible.[8] For convenience, perform a permutation transformation so the first m rows of $\tilde{R}$ equal I_m and the first m elements of $\tilde{b}$ equal 0.

This demonstrates that any admissible affine diffusion process can be transformed to yield an m-dimensional vector of pure volatility processes with dynamics described by

$$dz = (a + Az)dt + \begin{bmatrix} C_{11}^\top & C_{21}^\top \end{bmatrix} \operatorname{diag}\left(\sqrt{\begin{bmatrix} 0 \\ b_2 \end{bmatrix} + \begin{bmatrix} I_m \\ B_{21} \end{bmatrix} z} \right) dW$$

(tildes have been dropped to avoid clutter). The volatility variables associated with this process are

$$v = \begin{bmatrix} 0 \\ b_2 \end{bmatrix} + \begin{bmatrix} \mathrm{I}_m \\ B_{21} \end{bmatrix} z.$$

Admissibility requires that, for every i and for all $z \in \bar{\mathcal{D}}_i^z$, the variance of $v_i = 0$:

$$B_{i\cdot} \begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix}^\top \operatorname{diag}(b + Bz) \begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix} B_{i\cdot}^\top = 0.$$

Consider the implications of these conditions for the first m elements of v, which equals z itself. Because the origin and non-negative points close to it are feasible C_{11}

[8] $\exists \delta > 0 : \forall \epsilon \geq 0$ with $\|\epsilon\| < \delta$, $z = z_\iota + \mathbf{R}_\iota^{-1}\epsilon$ is feasible.

must be diagonal and C_{21} must only contain non-zero values in elements ij such that $B^a_{i\cdot}$ is proportional to $B^a_{j\cdot}$.[9] Given the previous discussion of covariance transformations, it is possible to apply further transformations to ensure that C_{21}=0, without changing C_{11}, B_{11} or b_1. Without loss of generality, therefore, assume that these transformations have been applied and that $C_{21} = 0$.

Suppose now that there are feasible points such that $b_i + B_{i\cdot}z = 0$ for $i > m$ such that $z \gg 0$, i.e., on the interior of $\mathbf{R}^m_+$. This means that

$$B_{i\cdot}C_{11}^\top \mathrm{diag}(z)C_{11}B_{i\cdot}^\top = [B_{i\cdot}C_{11}]^2 z = 0,$$

where the square is taken element by element. This is the sum of products of non-negative terms and strictly positive terms. Hence, it must be the case that all of the terms in $[B_{i\cdot}C_{11}]$ are zero. Because C_{11} is full rank, this implies that $B_{i\cdot} = 0$. This contradicts the assumption that $b_i + B_{i\cdot}z = 0$ with $z \gg 0$.

Corollary 1 *At the unique feasible corner* $RR_\iota^{-1} \geq 0$.

Suppose that, for some i and j, $[RR_\iota^{-1}]_{ij} < 0$. Consider the point $\tilde{z}$ for which $\tilde{z}_k = 0$ for $k \neq j$ and $\tilde{z}_j = -(b_i - R_{i\cdot}R_\iota^{-1}b_\iota)/[RR_\iota^{-1}]_{ij} > 0$. $\tilde{z}$ is a feasible corner not equal to z_ι, contradicting the assumption that there is a unique feasible corner.

Corollary 2 *The feasible region,* $\mathcal{D}^z$*, is a convex polyhedral cone in* R^m.

The m-faceted convex polyhedral cone defined by the $n \times m$ matrix W and vertex $x_0 \in \mathbf{R}^n$ is the set of points $\{x : \exists \alpha \geq 0 \text{ with } x - x_0 = W\alpha\}$. Let W be defined by the matrix R_ι^{-1} and the vertex z_ι. Define $z = -R_\iota^{-1}b_\iota + R_\iota^{-1}\alpha$, for some $\alpha \geq 0$. The volatility process at this z, $v(z) = b - RR_\iota^{-1}b_\iota + RR_\iota^{-1}\alpha$, is feasible because $b - RR_\iota^{-1}b_\iota \geq 0$ and $RR_\iota^{-1} \geq 0$.

[9]This is the second part of Duffie and Kan's Condition A.