Arrow–Debreu Prices for Affine Models

Régis J. Van Steenkiste† and Silverio Foresi‡

April 26, 1999

Abstract

We put forward a general methodology to price arbitrary payoffs linked to the realization of interest rates, asset prices, or other variables driven by the multivariate Affine Jump–Diffusion process of Duffie and Kan (1996). We attack and solve the basic problem of computing the Arrow–Debreu state prices or, equivalently, Green’s functions associated with the process. Given the Arrow–Debreu state prices, one can price derivative instruments with payoffs of arbitrary complexity. Within this framework, we also develop a scheme to price derivatives with early exercise at intermediate dates. To derive Arrow–Debreu state prices we exploit the basic observation that the integral of the overnight interest rate is itself affine. We augment the state space to add the integral of the overnight rate and we use transform methods to compute the density of the augmented affine process to calculate Arrow–Debreu prices. The main goal of the paper is to provide a viable numerical implementation of the proposed methodology, and we illustrate with applications the concepts introduced below.

JEL Classification Codes: G12, G13, C63

Keywords: Pricing kernel; Green’s function; Fourier Transform; FFT; DFT; term structure; derivatives; numerical methods; early exercise.

---

*We welcome comments, including references to related papers we inadvertently overlooked.
†Salomon Smith Barney, Inc.; regis.vansteenkiste@ssmb.com
‡Goldman Sachs Asset Management; silverio.foresi@gx.com.
1 Introduction

We put forward a general methodology to price arbitrary payoffs linked to the realization of interest rates, asset prices, or other variables driven by the multivariate Affine Jump-Diffusion process (AJD) of Duffie and Kan (1996). We attack and solve the basic problem of computing the Arrow-Debreu (A-D) state prices or, equivalently, Green's functions associated with the process. Given the A-D state prices, one can price derivative instruments with payoffs of arbitrary complexity. Within this framework, we also develop a scheme to price derivatives with early exercise at intermediate dates.

The motivation for the exercise is essentially a practical one. One often needs to price contingent claims whose payoffs are not affine, and for which "closed form" or other analytical methods such as those used in Chacko and Das (1999) and Duffie, Pan, and Singleton (1998) cannot be applied. While simulation offers a viable alternative to handle non-affine payoffs, it cannot handle early exercise in a natural way. A typical example of a contingent claim which is difficult to price by either methods is an interest rate swap which is cancelable on specific dates.

To derive A-D state prices we start from the basic observation that the pricing kernel is itself exponential affine. We proceed by augmenting the state space with a state variable equal to the integral of the overnight rate, which is equal to the natural logarithm of the pricing kernel, after an appropriate change of measure. All the ingredients used in this paper are well known. The idea of augmenting the state space to include the integral of the interest rate is used by Chacko and Das (1999) to price Asian-style options, but it is implicit already in Jamshidian (1989). Given that the augmented process is affine, we exploit transform methods to calculate the density of the augmented affine process, a technique used in many applications, including Bates (1995), Das and Foresi (1996), Heston (1993), and Stein and Stein (1991). Given the distribution of the joint process, we calculate A-D prices using the same logic that Jamshidian (1989) used to compute A-D state prices for the Vasicek model.

Our primary interest lies in exploring the viability of the numerical implementation, and we will measure advantages and disadvantages of our approach in the associated metric. The method is well suited to price payoffs for which transform methods as, e.g., in Chacko and Das (1999) and Duffie, Pan, and Singleton (1998), cannot be applied. This is typically the case when payoffs are non-linear or non-loglinear in the underlying factors. Non-linear payoffs can be handled quite nat-
urally by simulation methods. It is however not easy to simulate exactly a general $A JD$ process, one typical problem being how to handle effectively time varying volatility. More importantly, our approach deals in a natural way with early exercise and overcomes the difficulties of simulation methods in this respect.

Most practitioners will probably argue that, for the class of problems we attack, the real contenders are neither simulation-based nor transform methods, but implementations based on a discrete-time lattice (or “tree”). Jamshidian (1989), for example, proposes an approximated procedure to compute $A-D$ prices on a tree. The approximation is usually very good for small time steps. However, for particular parameterization or values of the state variables, the approximation is acceptable only at the expense of prohibitive computational costs. In contrast, our method provides $A-D$ prices which can be computed in a continuous time setting (within the precision of the method) regardless of the horizon or time step.

The main limitation of the approach we develop here is that it is computationally intensive. It should not be considered as a substitute for pricing “vanilla” products for which closed form solutions or direct transform methods are readily available. While the present technique is computationally intensive, it is amenable to massive parallel computing.

While the techniques we exploit rely in essence on transform methods, this paper should be of interest also to researchers who prefer simulation or tree-based implementations. A scheme for improving the accuracy of tree-based methods is presented. In a similar vein, we suggest a simulation procedure for the general $AJD$, which recovers arbitrage-free prices regardless of the time step. Both improvements take advantage of a normalized Green’s function.

The plan of this paper goes as follows:

In Section 2, we review well-known principles of asset pricing and set forth the key insight of augmenting the state space to derive the Green’s function in terms of the density of the augmented process. We also introduce the concept of Convexity-Adjusted Density (CAD).

In Section 3, we derive for the class of $AJD$ processes the characteristic function of the augmented process and use transform methods to obtain its PDF, which is then integrated to compute the $A-D$ prices. We illustrate the technique for the case of the generalized multi-variate Vasicek model, for which a closed form solution is available (Beaglehole and Tenney, 1991).
In general, a closed form expression for the Green's function of an \( AJD \) process is not available. We solve the general problem by numerical methods in Section 4. We discuss in detail how to improve the speed of the algorithm by Fast Fourier Transforms. To improve the accuracy, we propose a rotation of the state space based on the conditional variance–covariance matrix of the process. We provide a derivation for the moments of the joint conditional distribution of the process for moments of any order, which is of independent interest.

In Section 5, the general methodology, known as backward induction, for pricing early exercise is reviewed. Through the use of Green's functions, a simplification in the procedure for calibrating the model to the initial term structure of interest rates, known as forward induction, is presented as a generalization of the results in Janshidian (1991). Finally, the present method is compared to existing methods. In particular, the use of the \( CAD \) in lieu of the \( PDF \) of the process is suggested to potentially improve the accuracy of both Monte-Carlo and tree-based methods.

We conclude in Section 6 with a series of applications to gauge the accuracy of the numerical method and its ease of implementation.

2 Arrow–Debreu State Prices

In this section, we review the basic principles of asset pricing by emphasizing the relationship between the \( A-D \) state prices or Green's functions and risk–neutral pricing. Once defined, the Green's function is normalized, leading to the concept of \( CAD \).

2.1 Green's Function

It is assumed that the uncertainty in the economy is captured by the \( n \times 1 \) state vector \( X_t \in \mathbb{R}^n \)

\[
X_t = \begin{pmatrix}
   x_t^{(1)} \\
   x_t^{(2)} \\
   \vdots \\
   x_t^{(n)}
\end{pmatrix}.
\]  

(1)

It is assumed that \( X_t \) affects asset prices and in particular the overnight rate \( r_t \).
For valuation purposes, we use a *pricing kernel* $m(X_t)$ which regulates prices of contingent claims. The existence of such a pricing kernel is equivalent to the absence of pure arbitrage opportunities (see, e.g., Duffie, 1996). The price $V(X_s, s)$ of an asset with maturity $T$ and terminal payoff $V(X_T, T)$ conditional on the state at time $s$ and paying a stream of cashflows $h(X_t, t), t \in (s, T]$ is (see, e.g., Duffie, 1996)

$$V(X_s, s) = E^q_s \left[ \frac{m(X_T)}{m(X_s)} V(X_T, T) \right] + \int_{t=s}^{t=T} \frac{m(X_t)}{m(X_s)} h(X_t, t) dt,$$  

(2)

where the expectation $E^q_s [ ]$ is taken with respect to the objective probability density function of the original process. The instantaneous expected rate of change of the pricing kernel is equal to minus the instantaneous interest rate, $r_t$ (see, e.g., Cox, Ingersoll, and Ross, 1985a, Theorem 1, in a diffusion setting), $E^q_t [d m_t/m_t] = -r_t dt$. Under the risk-neutral probability measure, changes in the pricing kernel are locally riskless, and $m$ is given by

$$\frac{m(X_{t+\tau})}{m(X_t)} = e^{-\int_{u=t}^{u=t+\tau} r_u du}. \tag{3}$$

We now introduce the state variable $x_t^{(0)}$, the integral of the short rate (Jamshidian, 1991)

$$x_t^{(0)} = \int_{u=s}^{u=t} r_u du, \tag{4}$$

and introduce the augmented state vector $\tilde{X}_t \in \mathbb{R}^{n+1}$

$$\tilde{X}_t = \left\{ \begin{array}{c} x_t^{(0)} \\ x_t^{(1)} \\ \vdots \\ x_t^{(n)} \end{array} \right\}. \tag{5}$$

Figure 1 illustrates the original state space and extended state space.

Now, we combine Eqs. (3)–(4) to rewrite the valuation equation as (Eq. 2)

$$V(X_s, s) = E^q_s \left[ e^{-x_t^{(0)}(s)} V(X_T, T) + \int_{t=s}^{t=T} e^{-x_t^{(0)}(t)} h(X_t, t) dt \right].$$  

(6)

where the expectation $E^q_s [ ]$ is taken with respect to the probability density function of the augmented process $\tilde{X}(\tilde{X}_t, t|X_s)$

$$V(X_s, s) = \int_{\tilde{X}_T \in D_T} e^{-x_t^{(0)}(s)} V(X_T, T) \tilde{\Psi}(\tilde{X}_T, T|\tilde{X}_s) d\tilde{X}_T$$

$$+ \int_{t=s}^{t=T} \int_{\tilde{X}_t \in D_t} e^{-x_t^{(0)}(t)} h(X_t, t) \tilde{\Psi}(\tilde{X}_t, t|\tilde{X}_s) d\tilde{X}_t dt,$$  

(7)
where $\tilde{D}_t$ is the allowable domain of the augmented process. If the terminal payoff $V(X_T, T)$ and the cashflow $h(X_t, t)$ do not depend on the integral of the overnight rate $x^{(0)}$, we can integrate along the $x^{(0)}$ axis, arriving at

$$
V(X_s, s) = \int_{X_T \in D_t} G(X_T, T | X_s) V(X_T, T) dX_T + \int_{t=s}^{t=T} \int_{X_t \in D_t} G(X_t, t | X_s) h(X_t, t) dX_t dt,
$$

(8)

where $G(X_t, t | X_s)$ is the Green's function defined as

$$
G(X_t, t | X_s) = \int_{x^{(0)}_t \in D_t^{(0)}} \Psi(\tilde{X}_t, t | \tilde{X}_s) e^{-x^{(0)}_t} dx^{(0)}_t.
$$

(9)

### 2.2 CAD of a Process

From the definition of the Green's function, we see that it is always positive and integrates over the $X_t$ space to a non-zero constant. We define $\Psi^c(X_t, t | X_s)$ as

$$
\Psi^c(X_t, t | X_s) = \frac{G(X_t, t | X_s)}{\int_{X_t \in D_t} G(X_t, t | X_s) dX_t},
$$

(10)
Table 1: Equivalence between the Green's function pricing and the expectation pricing formulae

**Expectations**

\[ V(X_s, s) = E_s \left[ e^{-\int_{s}^{t} r_s \, dX_s} V(X_T, T) + \int_{t}^{T} e^{-\int_{t}^{u} r_s \, dX_s} h(X_t, t) \, dt \right] \]

**Green's Function**

\[ V(X_s, s) = \int_{X_T \in \Omega_T} G(X_T, T | X_s) V(X_T, T) \, dX_T + \int_{t}^{T} \int_{X_s \in \Omega_s} G(X_t, t | X_s) h(X_t, t) \, dX_t \, dt \]

\[ G(X_t, t | X_s) = \int_{s^{(e)} \in \Omega^{(e)}} \Psi(X_t, t | X_s) e^{-s_t^{(e)}} \, dz_t^{(e)} \]

**Convexity-Adjusted Density (CAD)**

\[ \Psi_c(X_t, t | X_s) = \frac{G(X_t, t | X_s)}{\int_{X_t \in \Omega_t} G(X_t, t | X_s) \, dX_t} \]

\( V(X_s, s) \) is the value of an asset conditional on the state at time \( s \) and \( V(X_T, T) \) is the terminal value of the asset. \( h(X_t, t) \) is the state-contingent cashflow stream of the asset. \( \Psi(\tilde{X}_s, t | \tilde{X}_s) \) is the PDF of the augmented process.

which amounts to dividing the Green's function by the price of a Pure Discount Bond (PDB) maturing at time \( t \) conditional on the state at time \( s \). The function \( \Psi_c(X_t, t | X_s) \) is always positive and its integral over the allowable space for \( X_t \) is equal to one. Therefore, \( \Psi_c(X_t, t | X_s) \) can be interpreted as a Probability Density Function (PDF) and is referred to as the convexity-adjusted PDF. In Section 3.6, we provide an example of this concept.

For convenience, the expressions for the prices of an asset by expectations and by Green's functions are presented in Table 1 along with the definition of the convexity-adjusted density.
3  Affine Jump–Diffusion Models

We now particularize the general principles set forth in Section 2 to the class of AJD models. First, we recall the definition of the class of affine models considered in this paper. Next, we compute the characteristic function and, by inversion, the PDF of the process by taking advantage of its affine nature. The augmented process obtained by adding the integral of the overnight rate to the initial process is shown to be affine and its associated PDF is integrated to yield the Green’s function. As an example, this procedure is applied to the case of the multivariate Vasicek model, a subset of the AJD class of models. We conclude this section with suggestions for improving tree-based and Monte-Carlo implementations.

3.1 Affine Specification

We follow Duffie, Pan, and Singleton (1998) and Chacko and Das (1999)’s notation to describe the following general admissible Markov AJD model to represent the risk-neutral processes in an economy with trading horizon $H$

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t + J(t)^Tdn_t, \quad \mathbb{E}\left[dW_t dW_t^T\right] = I_m. \quad (11)$$

$W_t \in R^m$ is an $m \times 1$ vector of orthogonal Wiener processes and $I_m$ is the $m \times m$ identity matrix. $N_t \in R^l$ is an $l \times 1$ vector of orthogonal Poisson processes characterized by an $l \times 1$ frequency vector $\lambda(X_t, t) \in R^l$. $X_t, \mu(X_t, t) \in R^m$ are vectors, $\sigma(X_t, t) \in R^{m \times m}$ is a matrix. $J(t) \in R^{l \times n}$ is a matrix governing the amplitude of the jumps and is assumed to be a matrix of correlated random variables which are independent of the state vector $X_t$.

The following functional forms characterize the affine specification of the model

$$\begin{align*}
\mu(X, t) &= K_0(t) + K_1(t)X \\
\sigma(X, t)\sigma(X, t)^T &= H_0(t) + H_1(t)X \\
\lambda(X, t) &= L_0(t) + L_1(t)X,
\end{align*} \quad (12)$$

where $K_0(t) \in R^m$ is an $n \times 1$ vector, $K_1(t), H_0(t) \in R^{m \times n}$ are matrices. $L_0(t) \in R^l$ is an $l \times 1$ vector and $L_1(t) \in R^{l \times n}$ is a matrix. $H_1(t) \in R^{n \times n \times n}$ is a third order tensor such that

$$[H_1(t)X]_{ij} = [H_1(t)]_{ijk} X_k, \quad (13)$$

where a repeated index implies summation. These results are summarized in Table 3.
It is often useful to allow state-contingent discrete time changes in the affine specification of the model. The resulting local affine approximation represents a way to extend the current methodology to non-affine models (see Section 7 below). In the remainder of this section, we drop the state dependence of the affine specification to lighten up the notation.

3.2 Characteristic Function and PDF of the General AJD Process

The characteristic function of the process in Eq. (11) is given by

\[ \phi(X_s, t; \Omega) = \mathbb{E} \left[ e^{-\Omega^T X_t} \mid X_s \right], \quad (14) \]

where \(\mathbb{E} \left[ \cdot \mid X_s \right] \) is the expectation given the state at time \(s\). \(\Omega \in \mathbb{R}^n\) is an \(n \times 1\) vector defined in the dual frequency space

\[ \Omega = \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}. \quad (15) \]

Equation (14) can be rewritten as

\[ \phi(X_s, t; \Omega) = \int_{X_t \in \mathcal{D}_t} \Psi(X_t, t \mid X_s) e^{-\Omega^T X_t} dX_t, \quad (16) \]

where \(\mathcal{D}_t\) is the allowable state space. \(\Psi(X_t, t \mid X_s)\) is the conditional PDF at time \(t\) of the process conditional on the state at time \(s\). Thus, the characteristic function \(\phi(X_s, t; \Omega)\) is the Fourier Transform of the PDF of the process.

The characteristic function satisfies the following backward Kolmogorov PDE

\[ \mu_i(X_t, t) \frac{\partial \phi(X_t, t; \Omega)}{\partial x(i)} + \frac{1}{2} \left[ \sigma(X_t, t) \sigma^T(X_t, t) \right]_{ij} \frac{\partial^2 \phi(X_t, t; \Omega)}{\partial x(i) \partial x(j)} - \frac{\partial \phi(X_t, t; \Omega)}{\partial t} + \lambda_k(X_t, t) E_J \left[ \phi(X + J_k(t)^T, t; \Omega) - \phi(X, t; \Omega) \right] = 0 \quad i, j = 1, \ldots, n, \quad k = 1, \ldots, l, \quad (17) \]

where \(J_k(t)\) is the \(k^{th}\) line of the \(J(t)\) matrix, and a repeated index implies summation. \(E_J[\cdot]\) denotes the expectation taken with respect to the matrix \(J\). The solution \(\phi(X_t, t; \Omega)\) of this PDE is

\[ \log \phi(X_t, t; \Omega) = A^T(t; \Omega) X + B(t; \Omega), \quad (18) \]
where \( A(t; \Omega) \in C^n \) is an \( n \times 1 \) complex valued vector and \( B(t; \Omega) \in C \) is a complex function of time. Direct substitution of this expression into Eq. (17) yields

\[
\mu^T(X, t)A(t; \Omega) + \frac{1}{2}A^T(t; \Omega)\sigma(X, t)\sigma^T(X, t)A(t; \Omega) - X^TA'(t; \Omega) - B'(t; \Omega) + \lambda(X, t)TE \left[ e^{J(t)A(t; \Omega)} - 1 \right] = 0.
\]

(19)

In the jump related term, the exponential of the vector \( J(t)A(t; \Omega) \) is a vector whose components are the exponential of each individual component of \( J(t)A(t; \Omega) \). The expectation of this vector is taken with respect to the joint probability distribution of the mixture rule implied by \( J(t) \).

Due to the affine nature of the model, the above expression is identically zero whenever

\[
B'(t; \Omega) = K^T_0(t)A(t; \Omega) + \frac{1}{2}A^T(t; \Omega)H_0(t)A(t; \Omega) + L_0(t)E_J \left[ e^{J(t)A(t; \Omega)} - 1 \right]
\]

\[
A'(t; \Omega) = K^T_1(t)A(t; \Omega) + \frac{1}{2}A^T(t; \Omega)H_1(t)A(t; \Omega) + L_1(t)E_J \left[ e^{J(t)A(t; \Omega)} - 1 \right].
\]

(20)

In the above expression, \( A^T(t; \Omega)H_1(t)A(t; \Omega) \) is an \( n \times 1 \) vector with \( k \)th component equal to \( A_k(t; \Omega) [H_1(t)]_{kk} A_j(t; \Omega) \). This system of Riccati complex valued ODE's is subjected to the initial conditions

\[
B(s; \Omega) = 0 \quad A(s; \Omega) = -\Omega.
\]

(21)

It is assumed here that the system of differential equations in Eq. (20) has a solution for all times in the trading window \([0, T] \).

The PDF of the process is obtained by inversion of Eq. (16)

\[
\Psi(X_t, t|X_s) = \frac{1}{(2\pi)^{\frac{n}{2}}} \int_{\Omega} \phi(X_s, t; \Omega) e^{\sigma^T X_s t} d\Omega.
\]

(22)

A numerical implementation of this inversion procedure is presented in Section 4.1 and relies on the use of multi-variate DFT's.

The expressions for the characteristic function and the PDF of the process are reported in Table 2.
Table 2: Characteristic function and PDF of an affine process

Characteristic Function

\[
\log \phi(X, t, \Omega) = A^T(t, \Omega)X + B(t, \Omega)
\]

where

\[
\begin{align*}
B(t, \Omega) &= K_0(t)A(t, \Omega) + \frac{1}{2}A^T(t, \Omega)H_0(t)A(t, \Omega) + L_0(t)E_J \left[ e^{\gamma(t)A(t, \Omega)} - 1 \right] \\
A(t, \Omega) &= K_1(t)A(t, \Omega) + \frac{1}{2}A^T(t, \Omega)H_1(t)A(t, \Omega) + L_1(t)E_J \left[ e^{\gamma(t)A(t, \Omega)} - 1 \right]
\end{align*}
\]

Initial condition

\[
\begin{align*}
B(s, \Omega) &= 0 \\
A(s, \Omega) &= -\gamma \Omega
\end{align*}
\]

Probability Density Function (PDF)

\[
\Psi(X_t, t | X_s) = \frac{1}{(2\pi)^T} \int_{\Omega} \phi(X_s, t, \Omega)e^{\gamma(T-s)}d\Omega
\]

\(K_0(t), K_1(t), H_0(t), H_1(t), L_0(t), \) and \(L_1(t)\) are defined in Eq. (12). \(\Omega\) is defined in Eq. (15). The above expressions with a tilde are valid for the augmented process \(\sim\).
3.3 Augmented AJD Model

In the multi-factor models typically encountered in practice, the short rate is assumed to be one of the state variables or a linear combination of the state variables

\[ r_t = \alpha_s^T X_t + \beta_s \quad t > s, \]  

(23)

where \( \alpha_s \in \mathbb{R}^n \) is an \( n \times 1 \) vector and \( \beta_s \in \mathbb{R} \) is a scalar. Note that both \( \alpha_s \) and \( \beta_s \) may be functions on the state \( X_s \). In Section 7, we suggest a methodology for selecting the functions \( \alpha_s \) and \( \beta_s \) to approximate more general non-linear dependencies. In the following derivation, to lighten the notation, we do not write explicitly the state-contingent dependence of the affine specification of the overnight rate.

We extend the model in Eq. (11) by adding a variable \( x_t^{(0)} \) which represents the integral of the short rate \( r_t \) defined as

\[ dx_t^{(0)} = \left[ \alpha^T X_t + \beta \right] dt. \]

(24)

The addition of this factor also appears in Jamshidian (1989) and Chacko and Das (1999). The key observation here is that the integral of the short term rate is itself affine in the factors of the original process. We define the augmented state vector \( \tilde{X}_t \in \mathbb{R}^{\tilde{n}=n+1} \)

\[ \tilde{X}_t = \left\{ \begin{array}{c} x_t^{(0)} \\ x_t^{(1)} \\ \vdots \\ x_t^{(n)} \end{array} \right\}. \]

(25)

The process for the new state vector is

\[ d\tilde{X}_t = \tilde{\mu}(\tilde{X}_t,t)dt + \tilde{\sigma}(\tilde{X}_t,t)dW_t + \tilde{J}(t)^T dN_t. \]

(26)

\( \tilde{X}_t, \tilde{\mu}(\tilde{X}_t,t) \in \mathbb{R}^{\tilde{n}} \) are vectors defined as

\[ \tilde{X}_t = \left\{ \begin{array}{c} x_t^{(0)} \\ X_t \end{array} \right\} \quad \tilde{\mu}(\tilde{X}_t,t) = \left\{ \begin{array}{c} \alpha^T X_t + \beta \\ \mu(X_t,t) \end{array} \right\}. \]

(27)

\( \tilde{\sigma}(\tilde{X}_t,t) \in \mathbb{R}^{\tilde{n} \times \tilde{n}} \) and \( \tilde{J}(t) \in \mathbb{R}^{l \times \tilde{n}} \) are matrices defined as follows

\[ \tilde{\sigma}(\tilde{X}_t,t) = \begin{bmatrix} \mathcal{O}^T \\ \sigma(X_t,t) \end{bmatrix} \quad \tilde{J}(t) = \left[ \begin{array}{c} \mathcal{O} \\ J(t) \end{array} \right]. \]

(28)

where \( \mathcal{O} \) is an \( n \times 1 \) vector of zeros. Moreover, the intensity vector \( \tilde{\lambda}(\tilde{X}_t,t) \in \mathbb{R}^{\tilde{n}} \) of the jumps in the augmented process is

\[ \tilde{\lambda}(\tilde{X}_t,t) = \left\{ \begin{array}{c} 0 \\ \lambda(X_t,t) \end{array} \right\}. \]

(29)
Figure 1 illustrates the original state space and extended state space.

By using the affine specification of the original process (Eq. 12), we can write the affine specification of the augmented process in terms of the one of the original process according to

\[
\tilde{K}_0(t) = \begin{cases} \beta \\ K_0(t) \end{cases} \quad \tilde{K}_1(t) = \begin{bmatrix} 0 & \alpha^T \\ \mathcal{O} & K_1(t) \end{bmatrix} \quad \tilde{L}_0(t) = L_0(t) \quad \tilde{L}_1(t) = \begin{bmatrix} \mathcal{O} & L_1(t) \end{bmatrix}.
\] (30)

Similarly, the augmented matrix \( \tilde{H}_0(t) \) and augmented tensor \( \tilde{H}_1(t) \) are defined as

\[
\tilde{H}_0(t) = \begin{bmatrix} 0 & \mathcal{O}^T \\ \mathcal{O} & H_0(t) \end{bmatrix}
\]

\[
\left[ \tilde{H}_1(t) \right]_{i,j,k} = \begin{cases} 0 \\ \mathcal{O}^T [H_1(t)]_{i,j,k} \\ \mathcal{O}^T [H_1(t)]_{i,j,k} \end{cases} \quad k = 1, \ldots, n.
\] (31)

These results are shown in Table 3 where they are compared with the parameters of the original process.

Thus, the affine structure of the model is preserved in the extended model. Therefore, it lends itself to the same range of techniques that are applicable to the original process. In particular, we can compute the characteristic function of the augmented process and, by inversion, the PDF of the augmented process. This is the object of the next Section.

3.4 Characteristic Function and PDF of the Augmented Process

A procedure similar to the one described to compute the characteristic function of the original process can be applied here to compute the characteristic function of the augmented process. We simply quote the final result

\[
\log \tilde{\phi}(\tilde{X}, t; \tilde{\Omega}) = \tilde{A}^T(t; \tilde{\Omega}) \tilde{X} + \tilde{B}(t; \tilde{\Omega}),
\] (32)

where \( \tilde{\Omega} \in \mathbb{R}^\tilde{n} \) is the augmented \( \tilde{n} \times 1 \) frequency vector defined as

\[
\tilde{\Omega} = \begin{bmatrix} \omega_0 \\ \omega_1 \\ \vdots \\ \omega_\tilde{n} \end{bmatrix} = \begin{bmatrix} \omega_\tilde{n} \\ \Omega \end{bmatrix}.
\] (33)
Table 3: Affine specification of the original model and augmented model

**Original Model**

d\(X_i = \mu(X_i, t)dt + \sigma(X_i, t)dW_i + J(t)^T dN_i\)

\[
\begin{align*}
\mu(X, t) &= K_0(t) + K_1(t)X \\
\sigma(X, t)\sigma(X, t)^T &= H_0(t) + H_1(t)X \quad \lambda(X, t) = L_0(t) + L_1(t)X
\end{align*}
\]

**Augmented Model**

\[d\tilde{X}_i = \tilde{\mu}(\tilde{X}_i, t)dt + \tilde{\sigma}(\tilde{X}_i, t)dW_i + \tilde{J}(t)^T dN_i\]

\[
\tilde{X}_i \begin{cases}
X_i^{(0)} = \int_{\gamma_{\omega, -2}}^{\gamma_{\omega, 1}} [\alpha^T X_u + \beta] \, du \\
X_i
\end{cases}
\]

\[
\begin{align*}
\tilde{\mu}(\tilde{X}, t) &= \tilde{K}_0(t) + \tilde{K}_1(t)\tilde{X} \\
\tilde{\sigma}(\tilde{X}, t)\tilde{\sigma}(\tilde{X}, t)^T &= \tilde{H}_0(t) + \tilde{H}_1(t)\tilde{X} \quad \Rightarrow \\
\tilde{\lambda}(\tilde{X}, t) &= \tilde{L}_0(t) + \tilde{L}_1(t)\tilde{X}
\end{align*}
\]

\[
\begin{align*}
\tilde{K}_0(t) &= \begin{bmatrix} \beta \\ K_0(t) \end{bmatrix} \\
\tilde{K}_1(t) &= \begin{bmatrix} 0 \\ \alpha^T \\ K_1(t) \end{bmatrix} \\
\tilde{H}_0(t) &= \begin{bmatrix} 0 \\ \alpha^T \end{bmatrix} \\
\tilde{H}_1(t) &= \begin{bmatrix} 0 \\ [H_1(t)]_{ij} \end{bmatrix} \quad k = 1, \ldots, n \\
\tilde{L}_0(t) &= L_0(t) \\
\tilde{L}_1(t) &= \begin{bmatrix} \alpha \\ L_1(t) \end{bmatrix}
\end{align*}
\]

\(\mathcal{O}\) is an \(n \times 1\) vector of zeros.
The vector $\bar{A}(t)$ and scalar function $\bar{B}(t)$ satisfy the following system of differential equations

$$
\begin{align*}
\bar{B}'(t; \tilde{\Omega}) & = \bar{K}_0^T(t)\bar{A}(t; \tilde{\Omega}) + \frac{1}{2} \bar{A}^T(t; \tilde{\Omega})\bar{H}_0(t)\bar{A}(t; \tilde{\Omega}) + \tilde{\mathcal{L}}_0(t) T_{EJ} \left[ e^{\tilde{J}(t; \tilde{\Omega})} - 1 \right] \\
\bar{A}'(t; \tilde{\Omega}) & = \bar{K}_1^T(t)\bar{A}(t; \tilde{\Omega}) + \frac{1}{2} \bar{A}^T(t; \tilde{\Omega})\bar{H}_1(t)\bar{A}(t; \tilde{\Omega}) + \tilde{\mathcal{L}}_1(t) T_{EJ} \left[ e^{\tilde{A}'(t; \tilde{\Omega})} \tilde{J}(t) - 1 \right],
\end{align*}
$$

subjected to the initial conditions

$$
\begin{align*}
\bar{B}(s; \Omega) & = 0 \\
\bar{A}(s; \Omega) & = -\omega_0.
\end{align*}
$$

This constitutes a system of differential equations, which usually requires numerical integration. We now substitute Eqs. (30) and (31) into Eq. (34) and write the differential equation that must be satisfied by $A_0(t; \Theta)$

$$
A_0'(t; \tilde{\Omega}) = 0. \quad (36)
$$

The solution is

$$
A_0(t; \tilde{\Omega}) = -\omega_0. \quad (37)
$$

Note that it is assumed that the solution of the system of differential equations in Eq. (34) exists over the economy trading horizon $[0, H]$.

Similarly to Eq. (22), we take the inverse Fourier Transform of the characteristic function to obtain the PDF of the augmented process

$$
\hat{\Psi}(X_t, t | \bar{X}_s) = \frac{1}{(2\pi)^n} \int_{\Omega} \hat{\phi}(\bar{X}_s, t; \Omega) e^{i\bar{X}_s \bar{X}_t} d\Omega. \quad (38)
$$

### 3.5 Green’s Function and CAD of the Process

After the density of the augmented process is computed, the Green’s function of the original process is obtained by integrating this density along the additional state variable $x_t^{(0)}$. The CAD of the process is then computed from Eq. (10). Both results are shown in Table 1.

In the next section, we examine a class of affine models for which the characteristic function, the PDF, and the Green’s function can be derived analytically.
3.6 General Vasicek Model

The general Vasicek process is a subset of the \( AJD \) class for which \( \tilde{H}_1(t) = 0 \) and for which there is no jump \( J(t) \equiv 0 \). We seek a solution for \( \tilde{A}(t; \tilde{\Omega}) \) of the form

\[
\tilde{A}(t; \tilde{\Omega}) = -j \tilde{M}(t) \tilde{\Omega},
\]

where \( \tilde{M}(t) \) is an \( \tilde{n} \times \tilde{n} \) time varying real-valued matrix. Substitution of the above equation into Eq. (34) yields

\[
\tilde{M}'(t) \tilde{\Omega} = \tilde{K}_0^T(t) \tilde{M}(t) \tilde{\Omega}.
\]

For this equation to hold for all \( \tilde{\Omega} \), we must have

\[
\tilde{M}'(t) = \tilde{K}_0^T(t) \tilde{M}(t).
\]

The boundary condition on \( \tilde{A}(s; \tilde{\Omega}) \) (Eq. 35) imposes

\[
\tilde{M}(s) = I_{\tilde{n}},
\]

where \( I_{\tilde{n}} \) is the \( \tilde{n} \times \tilde{n} \) identity matrix. By substituting Eq. (39) into Eq. (34) and integrating the resulting expression, we arrive at

\[
\tilde{B}(t; \tilde{\Omega}) = -j \left[ \int_{u=s}^{u=t} \tilde{K}_0^T(u) M(u) du \right] \tilde{\Omega} - \frac{1}{2} \tilde{\Omega}^T \left[ \int_{u=s}^{u=t} M^T(u) \tilde{H}_0(u) M(u) du \right] \tilde{\Omega}.
\]

Therefore, the conditional characteristic function of the augmented process is (Eq. 32)

\[
\log \tilde{\phi} (\tilde{X}_s; t; \tilde{\Omega}) = -j \tilde{\Omega}^T \left[ \tilde{M}^T(t) \tilde{X}_s + \int_{u=s}^{u=t} \tilde{M}^T(u) \tilde{K}_0(u) du \right]
- \frac{1}{2} \tilde{\Omega}^T \left[ \int_{u=s}^{u=t} M^T(u) \tilde{H}_0(u) M(u) du \right] \tilde{\Omega}.
\]

By comparing the above expression with Eq. (106) in Appendix A.1, we conclude that the augmented process is distributed normally

\[
\tilde{X}_t \overset{\text{dist}}{\sim} N(\tilde{\mu}, \tilde{\Sigma}).
\]

The mean \( \tilde{\mu} \) and variance–covariance matrix \( \tilde{\Sigma} \) are given by

\[
\tilde{\mu} = \tilde{M}^T(t) \tilde{X}_s + \int_{u=s}^{u=t} \tilde{M}^T(u) \tilde{K}_0(u) du
\]

\[
\tilde{\Sigma} = \int_{u=s}^{u=t} M^T(u) \tilde{H}_0(u) M(u) du.
\]
The Green's function of the original process is obtained by integration of the PDF of the augmented process according to Eq. (9). We refer to the result derived in Appendix A.3 to show that the conditional Green's function is

\[
G(X_t, t|X_s) = e^{-m_0 + \frac{\tilde{E}_0}{2}} e^{-\frac{1}{2} (x - m + \tilde{\Sigma}_0)^T \Sigma^{-1} (x - m + \tilde{\Sigma}_0)} \frac{1}{(2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}},
\]

(47)

where \(m_0, \tilde{\Sigma}_0,\) and \(\tilde{\Sigma}_0\) are defined in Table 4. This result is in agreement with Beaglehole and Tenney (1991). Note that the Green's function is proportional to the PDF of a multi-variate normal distribution with mean \(m - \Sigma_0\) and variance-covariance matrix \(\Sigma\). According to Eq. (10), this distribution is referred to as the convexity-adjusted PDF

\[
G(X_t, t|X_s) = \Psi^c(X_t, t|X_s) e^{-m_0 + \frac{\tilde{E}_0}{2}}.
\]

(48)

The second factor in the above expression is the price contingent on the state at time \(s\) of a zero-coupon bond maturing at time \(t\).

This example illustrates the difference between the convexity-adjusted density and the actual density of the original process. The magnitude of the difference is not only a function of the time step \(t - s\) but also of the probability distribution of the joint process, which depends on the parametrization of the model. This convexity adjustment has implications when implementing numerically a model of the AJD class as explained in the next section.

Note that, as expected, due to the Gaussian nature of the process, the variance-covariance matrix of the convexity-adjusted distribution is identical to the variance-covariance matrix of the original distribution. However, this is not true in general. We have summarized the results for the general Vasicek model in Table 4.

3.7 Applications of the CAD: Monte-Carlo and Tree Based Implementations

Unlike tree-based implementations, the current method provides a pricing scheme that preserves the no-arbitrage condition regardless of the trading horizon. Tree-based methods are arbitrage free for the given trading horizon but cannot ensure arbitrage free pricing if the trading horizon is smaller than the steps chosen when discretizing the time domain. Also it is often expedient to price different instruments on lattices of different time steps. We suggest to build trees that match the moments
Table 4: Green’s function and convexity–adjusted density of the general Vasicek model

\[ G(X_t, t|X_s) = \Psi^c(X_t, t|X_s)e^{-m_0 + \frac{\hat{\Sigma}}{2}} \]

where

\[
\begin{align*}
\Psi^c(X_t, t|X_s) &= \frac{e^{-\frac{1}{2}(x - m + \hat{\Sigma}u)^T\Sigma^{-1}(x - m + \hat{\Sigma}u)}}{(2\pi)^{\frac{m}{2}} \sqrt{\det \Sigma}} \\
\hat{m} &= \begin{cases} m_0 \\ m \end{cases} = M^T(t)\hat{X}_s + \int_{u=s}^{u=t} M^T(u)\hat{K}_0(u)du \\
\hat{\Sigma} &= \begin{bmatrix} \hat{\Sigma}_{00} & \hat{\Sigma}_{0}\Sigma' \\ \hat{\Sigma}_{0} & \Sigma' \end{bmatrix} = \int_{u=s}^{u=t} M^T(u)\hat{H}_0(u)M(u)du \\
M(t) &\text{ is solution of} \\
\begin{cases} \\
M'(t) = \hat{K}_0^T(t)M(t) \\
\bar{M}(s) = I_{\bar{n}}
\end{cases}
\end{align*}
\]

\(\Psi^c(X_t, t)\) is the convexity–adjusted PDF of the model and is defined in Table 1. \(m\) and \(\Sigma\) are the mean and variance–covariance matrices of the process, respectively. \(\hat{K}_0(t), \hat{K}_1(t),\) and \(\hat{H}_0(t)\) are defined in Table 3. \(I_{\bar{n}}\) is the \(\bar{n} \times \bar{n}\) identity matrix.
of the \( CAD \) instead of the \( PDF \) of the process. Using only one parametrization to produce the \( CAD \) of the process ensures consistency in pricing across instruments regardless of the different time discretizations used to price each individual instrument.

In general, Monte–Carlo methods are also prone to similar discretization problems. The \( CAD \), however, provides the appropriate distribution that guarantees arbitrage free pricing regardless of the discretization horizon. Based on the \( CAD \), one can envision an integrated framework which combines Monte–Carlo paths and a backward induction scheme for pricing early exercise.

4 Numerical Implementation

In general, a closed form expression for the Green’s function of an \( AJD \) process is not available. In this section, we present a numerical scheme for computing the density of an \( AJD \) process, and by extension, the associated Green’s function.

4.1 DFT Approximation

We present the formalism for computing numerically Fourier Transforms and their inverse. This section generalizes to the multi–variate case the paper by Carr and Madan (1998). This derivation is valid for the general class of \( AJD \) processes. The method applies to both the original process and the augmented process and relies on the Laurent–Schwartz notation described in Appendix A.4 to represent the following symbols

\[
X_t^{\text{min}}, X_t^{\text{max}}, \Delta X_t \in \mathbb{R}^n \\
N, \beta \in \mathbb{N}^n \\
\gamma, \gamma_{\text{min}}, \pm 1 \in \mathbb{Z}^n.
\]  

The characteristic function of the augmented process is the Fourier Transform of the density of the distribution over the allowable state space (Eq. 16)

\[
\phi(X_\gamma, t; \Omega) = \int_{X_t \in \mathcal{D}_t} \Psi(X_t, t|X_\gamma) e^{-i \Omega^T X_t} dX_t.
\]  

In this case, the characteristic function is assumed to be known and we wish to determine the density by inversion of the above equation.
The procedure starts from the observation that most of the probability of the process is concentrated in a limited hypervolume $\mathcal{V}$ of the state space. We select an $n$-dimensional hypercube $\mathcal{H}$ that contains $\mathcal{V}$ and denote by $X_t^{\text{min}}$ and $X_t^{\text{max}}$ the opposite corners of the hypercube. The choice for the dimension of the hypercube can for instance be guided by examining the moments of the distribution of the augmented affine process. In Section 4.3, we show that these moments can be obtained by solving a finite number of linear ODEs.

The Fourier Transform of the density of the augmented process is approximated by the integral over the hypercube $\mathcal{H} \subset \mathcal{D}_t$

$$
\phi(X_s, t; \Omega) \approx \int_{X_t \in \mathcal{H}} \Psi(X_t, t|X_s)e^{-\mathcal{A}^T X_t} dX_t.
$$

We now select $N$ sample points in the state space equally distributed within the hypercube and approximate the integral by the following sum

$$
\phi(X_s, t; \Omega) \approx \sum_{0 \leq \beta < N} \Psi(X_t^{\text{min}} + \beta \Delta X_t, t|X_s)e^{-\mathcal{A}^T (X_t^{\text{min}} + \beta \Delta X_t)} \Delta X_t^1, \tag{52}
$$

where $\Delta X_t$ is defined as

$$
N \Delta X_t = X_t^{\text{max}} - X_t^{\text{min}}. \tag{53}
$$

The next step is to sample the dual frequency space

$$
\Omega = \frac{2\pi\gamma}{N \Delta X_t} \quad \gamma_{\text{min}} \leq \gamma < \gamma_{\text{min}} + N. \tag{54}
$$

Intuitively, the presence of $\Delta X_t$ in the denominator is dictated by the fact that a tightly packed PDF $\Psi$ in the state space contains a lot of high “frequency” components and is therefore very widely distributed in the frequency domain.

Substitution of Eq. (54) into Eq. (52) yields

$$
\phi \left( X_s, t; \frac{2\pi\gamma}{N \Delta X_t} \right) \approx \sum_{0 \leq \beta < N} \Psi(X_t^{\text{min}} + \beta \Delta X_t, t|X_s)e^{-2\pi j \frac{\gamma X_t^{\text{min}} + \beta \Delta X_t}{N \Delta X_t}}. \tag{55}
$$

We recognize the above expression for $\phi$ as the $n$-dimensional DFT of the sampled PDF (Bracewell, 1986). By inversion of Eq. (55), we obtain

$$
\Psi(X_t^{\text{min}} + \beta \Delta X_t, t|X_s) \approx (N \Delta X_t)^{-1} \times \sum_{\gamma_{\text{min}} \leq \gamma < \gamma_{\text{min}} + N} \phi \left( X_s, t; \frac{2\pi\gamma}{N \Delta X_t} \right) e^{2\pi j \frac{\gamma X_t^{\text{min}} + \beta \Delta X_t}{N \Delta X_t}}. \tag{56}
$$
Table 5: Discretization of the Fourier Transform by DFT or FFT

<table>
<thead>
<tr>
<th>Fourier Transform</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \phi(X_s, t; \Omega) = \int_{X_t \in D_t} \Psi(X_t, t</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Discrete Fourier Transform (DFT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ \Psi(X_t^{\text{min}} + \beta \Delta X_t, t</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fast Fourier Transform (FFT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[ N = 2^q \iff \text{Acceleration} \approx \left( \frac{N}{q} \right)^1 ]</td>
</tr>
</tbody>
</table>

The characteristic function \( \phi(X_s, t; \Omega) \) is defined in Eq. (14). The same expressions with a tilde are valid for the augmented process (~). The above equations use the Laurent–Schwartz formalism defined in Appendix A.4.

Cooley and Tukey (1965) developed a method for computing the DFT if the number of sample points is a power of 2. This algorithm is known as the FFT and allows to drastically reduce the number of computations in the ratio

\[ \left( \frac{N}{\log_2 N} \right)^1, \quad (57) \]

where \( \log_2 \) denotes the logarithm in base two. The computational savings becomes huge even for relatively small values of \( N \). The FFT algorithm is based on a factorization of the transform tensor that maps the samples in the state space to the ones in the frequency space and vice-versa. We implemented the FFT through a time designation “butterfly” algorithm. For more details, we refer the reader to Bracewell (1986). The main results of this section are summarized in Table 5.
4.2 Rotation of the State Space

To improve the accuracy of the FFT method, we do not actually compute the Fourier Transform in the original state space. Instead, we rotate the state space to greatly reduce the volume of the hypercube $\mathcal{H}$ whose size, we recall, is determined such as to enclose the allowable states of the process almost surely. The motivation for this procedure lies in the fact that, in general, the state variables are correlated in level and a hypercube in the original state space spans regions of the state space with low associated probability as illustrated in Figure 2. Even if the factors are uncorrelated in level, the overnight rate and the integral of the short rate show some degree of level correlation, thereby justifying the pre-rotation of the state space.

If we denote by $\mathcal{V}_s(t)$ the variance–covariance matrix of the process at time $t$ conditional on the state at time $s$, the state vector $X_t$ is expressed in terms of the state vector $X_t^R$ in the rotated state space

$$
X_t = E_s[X_t] + \mathcal{R}X_t^R.
$$

(58)
\( \mathcal{R} \) is a matrix whose columns are the eigenvectors of the variance-covariance matrix \( \mathcal{V}_x(t) \)
\[
\mathcal{V}_x(t) = \mathcal{R} \Lambda_x(t) \mathcal{R}^{-1},
\]
where \( \Lambda_x(t) \) is the diagonalized variance-covariance matrix which can be used to choose the size of the hypercube \( \mathcal{H}^\mathcal{R} \).

Conversely, the state variable in the rotated space is
\[
X_t^\mathcal{R} = \mathcal{R}^{-1} (X_t - \mathbb{E}_x[X_i]) .
\]

By inspection of Eq. (60), we note that the state variables in the rotated state space are linear combinations of the initial state variables. Therefore, the process expressed in the rotated space is also in the affine class and the techniques developed so far do apply.

Thus, in order to improve the accuracy, it will be useful to compute the conditional variance-covariance matrix of the process. This is the object of the next section where we show that all the conditional moments of the joint distribution of the process can be obtained by solving a finite number of linear ODEs.

### 4.3 Moments of the Joint Distribution

We derive a general recursive system of linear differential equations that relate the moments of the distribution for the general class of AJD. The class of AJD models for which this system of differential equations can be solved analytically is wider than the class of models for which the Riccati ODEs (Eq. 20) can be solved exactly.

Analogous expressions hold for the moments of the subjective process in the economy. These expressions for the moments of the joint distribution can be used within the scope of a Generalized Method of Moments (GMM) to estimate, from market data, the parameters in the objective process.

The derivation relies on the notation introduced by Laurent–Schwartz to greatly reduce the number of indices used in multi-variate derivations (Section A.4). In the remainder of this section, we use the following symbols within the confines of the Laurent–Schwartz notation
\[
\beta, \gamma, \delta, \theta \in \mathbb{N}^n \\
\epsilon \in \mathbb{N}^l.
\]
With this notation, we proceed by writing Itô's lemma for $X_t^\alpha$

$$d(X_t^\alpha) = \sum_{|\beta|-1} D^\beta [X_t^\alpha] \mu(X_t, t)^\beta dt + \sum_{|\beta|-1} D^\beta [X_t^\alpha] (\sigma(X_t, t) dW_t)^\beta$$

$$+ \frac{1}{2} \sum_{|\beta|-|\gamma|-1} D^{\beta+\gamma} [X_t^\alpha] (\sigma(X_t, t) dW_t)^\beta (\sigma(X_t, t) dW_t)^\gamma$$

$$+ \sum_{|\epsilon|=1} \lambda^\epsilon (X_t, t) E_J [\{X_t + J(t)\}^\alpha - X_t^\alpha] dt + dJ_t^F. \quad (62)$$

$dJ_t^F$ is the unpredictable jump innovation

$$dJ_t^F = (X_t^+) - X_t^\alpha - \sum_{|\epsilon|=1} \lambda^\epsilon (X_t, t) E_J [\{X_t + J(t)\}^\alpha - X_t^\alpha] dt, \quad (63)$$

where

$$X_t^+ = \lim_{s \to t} X_s, \quad s > t. \quad (64)$$

Given the affine structure of the model, we now expand Eq. (62) by combining Eqs. (11) and (12)

$$d(X_t^\alpha) = \sum_{|\beta|-1} 1_{\alpha \geq \beta} \frac{\alpha!}{(\alpha - \beta)!} X_t^{\alpha-\beta} \left\{ K_0(t)^\beta + [K_1(t) X_t]^{\beta} \right\}$$

$$+ \sum_{|\beta|-|\gamma|-1} \frac{1_{\alpha \geq \beta+\gamma}}{2} \frac{\alpha!}{(\alpha - \beta - \gamma)!} X_t^{\alpha-\beta-\gamma} \left\{ H_0(t)^{\beta+\gamma} + [H_1(t) X_t]^{\beta+\gamma} \right\} dt$$

$$+ \sum_{|\epsilon|=1} \lambda^\epsilon (X_t, t) E_J [\{X_t + J(t)\}^\alpha - X_t^\alpha] dt$$

$$+ \sum_{|\beta|-1} D^\beta [X_t^\alpha] (\sigma(X_t, t) dW_t)^\beta + dJ_t^F. \quad (65)$$

Next, we integrate the above expression from time $s$ to time $t$ and take the expectation conditional on the state at time $s$. From Fubini's theorem and the martingale property of $dJ_t^F$ and $dW_t$, we have after taking the derivative with respect to time $t$

$$\frac{d}{dt} M_s^\alpha(t) - M_s^\alpha(t) \sum_{|\beta|-1} 1_{\alpha \geq \beta} \frac{\alpha!}{(\alpha - \beta)!} K_1(t)^{\beta\gamma}$$

$$- M_s^\alpha(t) \sum_{\beta+\gamma=\alpha} \frac{\alpha!}{\beta\gamma!} L_1(t)^{\beta\gamma} E_J [J(t)\eta^\epsilon] =$$

$$L_1(t)^{\beta\gamma} \eta^\epsilon$$

$$\beta \neq \alpha$$

$$|\epsilon| = 1$$

23
\[
\sum_{|\beta|-1} 1_{\alpha \geq \beta} \frac{\alpha!}{(\alpha - \beta)!} K_0(t)^\beta \mathcal{M}_s^{\alpha-\beta}(t) + \sum_{|\beta|=1} 1_{\alpha \geq \beta} \frac{\alpha!}{(\alpha - \beta)!} K_1(t)^\beta \mathcal{M}_s^{\alpha-\beta+\gamma}(t) \\
+ \sum_{|\beta|-|\gamma|-1} 1_{\alpha \geq \beta+\gamma} \frac{\alpha!}{(\alpha - \beta - \gamma)!} H_0(t)^\beta \mathcal{M}_s^{\alpha-\beta-\gamma}(t) \\
+ \sum_{|\beta|-|\gamma|-|\epsilon|-1} 1_{\alpha \geq \beta+\gamma} \frac{\alpha!}{(\alpha - \beta - \gamma)!} H_1(t)^\beta \mathcal{M}_s^{\alpha-\beta-\gamma+\delta}(t) + \\
+ \sum_{|\beta|=\alpha, |\epsilon|=1} L_0(t)^\beta \frac{\alpha!}{|\beta\gamma!} E_J [J(t)^\epsilon] \mathcal{M}_s^{\beta}(t) \\
+ \sum_{|\beta|=\alpha, |\epsilon|=|\theta|=1} L_1(t)^\beta \frac{\alpha!}{|\beta\gamma\theta|} E_J [J(t)^\epsilon] \mathcal{M}_s^{\beta+\theta}(t).
\]

(66)

where \( \mathcal{M}_s^{\alpha}(t) \) is defined as

\[
\mathcal{M}_s^{\alpha}(t) = E [X_t^\alpha | X_s].
\]

(67)

These results are reported in Table 6.

Under the assumption that the limits exist, the unconditional moments are obtained by letting time go to \( \infty \) in the above expression. Accordingly, the unconditional moments are obtained by setting in Eq. (66) the derivative with respect to time to zero and computing the limit as \( t \) approaches \( \infty \) of the different functions of time appearing in the expression for the conditional moments.

The recursive property of this differential system of equations holds for centered moments because centered moments are linear combinations of moments of equal or smaller order.

5 Forward and Backward Induction

In this section, we examine a possible simplification, known as forward induction, if the model has to be parameterized to match the initial term structure of interest rates. Next, we review the use of Green's function to price early exercise.
Table 6: Moments of the distribution of a general AJD model

\[
\frac{d}{dt} \mathcal{M}_0^0(t) = \left \{ \mathcal{M}_0^0(t) \sum_{|\beta|=1} 1_{\beta \geq \beta} \frac{\alpha!}{(\alpha-\beta)!} K_0(t)_{\beta \beta} + \mathcal{M}_0^0(t) \sum_{\beta \neq \alpha} L_1(t)_{\beta \gamma} \frac{\alpha!}{\beta! \gamma!} E_J [J(t)_{\beta \gamma}^\alpha] \right \} = \\
\sum_{|\beta|=1} 1_{\beta \geq \beta} \frac{\alpha!}{(\alpha-\beta)!} K_0(t)_{\beta \beta} \mathcal{M}_0^{\alpha-\beta}(t) + \sum_{|\beta|=1} \frac{\alpha!}{(\alpha-\beta)!} K_1(t)_{\beta \gamma} \mathcal{M}_0^{\alpha-\beta+\gamma}(t) + \sum_{|\beta|=1} 1_{\beta \geq \beta+\gamma} \frac{\alpha!}{(\alpha-\beta-\gamma)!} H_0(t)_{\beta \gamma} \mathcal{M}_0^{\alpha-\beta-\gamma}(t) \\
+ \sum_{|\beta|=1} 1_{\beta \geq \beta+\gamma} \frac{\alpha!}{(\alpha-\beta-\gamma)!} H_1(t)_{\beta \gamma} \mathcal{M}_0^{\alpha-\beta-\gamma+\delta}(t) + \sum_{\beta \neq \alpha} L_2(t)_{\beta \gamma} \frac{\alpha!}{\beta! \gamma!} E_J [J(t)_{\beta \gamma}^\alpha] \mathcal{M}_0^{\beta+\theta}(t) \\
+ \sum_{\beta \neq \alpha} \frac{\alpha!}{\beta! \gamma!} E_J [J(t)_{\beta \gamma}^\alpha] \mathcal{M}_0^{\beta+\theta}(t)
\]

The vectors $K_0(t)$, $L_0(t)$, the matrices $K_1(t)$, $H_0(t)$, and the tensor $H_1(t)$ are defined in Eq. (12). $\mathcal{M}_0^0(t)$ is defined in Eq. (67).
5.1 Term Structure Calibration

A model is often required to be calibrated to the initial term structure of interest rates. In practice, if the term structure is composed of $p$ PDB's with maturities $t_1, \ldots, t_p$ and priced at $P_0^{t_1}, \ldots, P_0^{t_p}$, the model will contain a vector $\Theta$ of $p$ parameters. This vector $\Theta$ is often interpreted as a time-dated risk-premium which is readjusted on a daily base. Backus, Foresi, and Zin (1998) argue that such model allow for self-inflicted arbitrage opportunities since the risk-premia behave like stochastic factors even though they are treated as deterministic functions.

As a result of the martingale property of discounted asset prices, the Green's function conditional on today's state satisfies the following recursive relationship

$$G(X_{t_{i+1}}, t_{i+1}; \Theta_{1 \rightarrow i+1} | X_0) = \int_{X_{t_i}} G(X_{t_{i+1}}, t_{i+1}; \Theta_{i+1} | X_{t_i}) G(X_{t_i}, t_i; \Theta_{1 \rightarrow i} | X_0) dX_{t_i},$$  \hspace{1cm} (68)

where $\Theta_{1 \rightarrow i}$ is a subvector of $\Theta$ built by taking the first $i$ components of $\Theta$. In the above expression, it is assumed that the Green's function at time $t_{i+1}$ conditional on the state at time $t_i$ depends only on one component of $\Theta$.

By integrating the above expression over all states $X_{t_{i+1}}$, we have the following expression for $P_0^{t_{i+1}}$

$$P_0^{t_{i+1}} = \int_{X_{t_{i+1}}} \int_{X_{t_i}} G(X_{t_{i+1}}, t_{i+1}; \Theta_{i+1} | X_{t_i}) G(X_{t_i}, t_i; \Theta_{1 \rightarrow i} | X_0) dX_{t_i} dX_{t_{i+1}}.$$  \hspace{1cm} (69)

This equation suggests the following usual recursive procedure for calibrating the model to the initial term structure of interest rates:

1. Start $i = 0$
2. Adjust the value for the parameter $\Theta_{i+1}$.
3. Compute the value of the conditional Green's function $G(X_{t_{i+1}}, t_{i+1}; \Theta_{i+1} | X_t)$ and integrate over the state $X_{t_{i+1}}$ resulting in a function of $X_t$.
4. Integrate the product of the function of $X_t$ obtained in Step 3 with the known conditional Green's function $G(X_{t_i}, t_i; \Theta_{1 \rightarrow i} | X_0)$ to obtain the model price for the bond (Eq. 69).
5. If the model price error is too large, readjust $\Theta_{i+1}$ and go back to Step 3. If the error is small enough, compute the Green's function $G\left(X_{t_{i+1}}, t_{i+1}; \Theta_{1 \rightarrow t_{i+1}}; X_0\right)$ (Eq. 68).

6. $i = i + 1$. Go back to Step 2

In implementing this procedure, practitioners have often approximated the Green's function to be the discounted density of the process. This idea was originally suggested by Jamshidian (1991) and is further discussed in, e.g., Rebonato (1996)

$$G(X_{t_{i+1}}, t_{i+1}|X_{t_i}) \approx \Psi(X_{t_{i+1}}, t_{i+1}|X_{t_i})e^{-f(X_{t_i}; \Theta_{i+1})(t_{i+1} - t_i)},$$  \hspace{1cm} (70)

where $f(X_{t_i}; \Theta_{i+1})$ is a function, usually chosen to be a short term rate contingent on the state at time $t_i$. The key observation here is that the Green's function is approximated by the product of the density of the process which integrates to one and a function which does not depend on $X_{t_{i+1}}$. The error between the actual Green's function and this approximation is a function of the length of the interval $t - s$ and can be quite substantial as illustrated in Section 6.3. Substitution of the above expression into Eq. (69) yields

$$P^t_{0+1} \approx \int_{X_{t_i}} G(X_{t_i}, t_i; \Theta_{1 \rightarrow t_i}; X_0) e^{-f(X_{t_i}; \Theta_{i+1})(t_{i+1} - t_i)} dX_{t_i}. \hspace{1cm} (71)$$

In the above equation, we have integrated out the density $\Psi(X_{t_{i+1}}, t_{i+1}|X_{t_i})$ of $X_{t_{i+1}}$ conditional on the state $X_{t_i}$. This relation, generalized here, was noted by Jamshidian (1991) and is valid only for infinitesimal $t_{i+1} - t_i$.

The main advantage of Eq. (71) over Eq. (69) is the speed at which the calibration takes place since we have one less integral to compute. The disadvantage is that it destroys the no-arbitrage condition and the time steps chosen need to stay small. In contrast, the method we propose satisfies the local no-arbitrage condition regardless of the size of the step chosen when discretizing in the time domain. The approximation in Eq. (71) can be used to obtain, at a relatively lower computational cost, a fairly accurate estimate of the parameter vector $\Theta$. If required, a more precise value can then be searched by making use of the correct expression for the price of the bond in Eq. (69).

5.2 Early Exercise

The use of the Green's function methodology for pricing options with early exercise relies on the fact that a Bermuda-style derivative can be converted into a compound European-style derivative. Standard limiting arguments allow to extend the
methodology to American-style options. This is the key observation that allows for a recursive application of the Green's function methodology known as backward induction, yet another form of the law of iterated expectation or the martingale property of discounted asset prices.

Accordingly, we define a series of events at time $t_1, \ldots, t_m = T$ where $T$ is the maturity of the security we need to price. An event can be an option exercise date, a cashflow date, or a change in the affine specification of the model due to changes in economic conditions (Section 7). We denote by $V_t$ the value of the security at time $t$ and by $E_i(V_t)$ the function that characterizes the event taking place at time $t$. The function $E_m(V_T)$ is the terminal payoff of the security. We want to compute the value of the security conditional on the state at time $t_0$. As mentioned, the terminal payoff of the security to be priced is

$$V_{T^-} = E_m(V_T).$$

(72)

The value right after an event $i - 1$ is (Eq. 8)

$$V_{t_{i-1}^+} = \int_{X_{t_i} \in \mathcal{D}} G(X_{t_i}, t_i|X_{t_{i-1}}) V_{t_{i-1}}^+ dX_{t_i}.$$  

(73)

The event $i - 1$ changes the value of the asset according to

$$V_{t_{i-1}} = E_{i-1}(V_{t_{i-1}^+}),$$

(74)

and the above two equations are applied recursively until the price is calculated for $i = 0$, which gives the value of the security. Note that in the case of a change in the affine specifications of the model, the value of the asset is not changed per se but the Green's function needs to be reevaluated based on the new specification. The recursive feature of Eq. (73) allows the pricing of any Bermuda-style options. The pricing of American-style options cannot be done except as a limiting case of the present methodology as the space between two events tends to zero. The backward induction pricing methodology is illustrated in Figure 3.

Because of the equivalence between conditional expectations used in Duffie, Pan, and Singleton (1998) and Chacko and Das (1999) and the present Green's function methodology, applying the present technique to the case where $m = 1$ to the valuation of European-style and Asian-style options yields the same results as in these two papers. In particular, in Appendix A.5, we generalize the concept of pricing probabilities introduced by Chacko and Das (1999) to the case of multiple exercises.
6 Applications

In this section, we consider a series of applications to gauge the accuracy and usefulness of the numerical techniques developed in Section 4 above.

First, in Section 6.1, we consider a one-factor Vasicek model for which we have closed form solutions for bond prices and for the Green’s function (Section 3.6). We compare the analytical results with those obtained by the algorithm proposed in Section 4.1.

Next, in Section 6.2, we assess the accuracy of the numerical algorithm in dealing with non-linear payoffs. To this end, we price power options and use the analytical result of Section 4.3 to verify a put-call parity relationship between power options.

In Section 6.3, we consider a two-factor stochastic volatility interest rate model. For this model, the Green’s function is computed numerically and normalized to yield the CAD of the process. The CAD of the process is compared to the PDF of the process to show the effect of the convexity adjustment.

Finally, in Section 6.4, we examine the case of a jump–diffusion process. In particular, we compute the Green’s function and show how a rotation of the state
space can be used to improve accuracy. The term structure of interest rates obtained by numerical integration of the Green's function is compared with the term structure obtained directly by integration of a system of Riccati ODE's.

6.1 One–Factor Vasicek

We examine a particular example of a Vasicek process for the short rate $y_t$

$$dy_t = -\kappa(y_t - \theta)\, dt + \xi dW_t,$$  \hspace{1cm} (75)

with the following parameterization

$$\kappa = 0.20 \quad \xi = 0.012 \quad \theta = 0.065.$$  \hspace{1cm} (76)

The model is extended by adding an additional factor, which is the integral of the overnight rate (Eq. 24)

$$dx_t = y_t dt.$$  \hspace{1cm} (77)

A direct application of the method described in Section 3.6 yields the Green's function for this process (Eq. 48)

$$G(y_t, t| y_s) = \psi^c(y_t, t| y_s)e^{-m_0 + \frac{S_0}{2}},$$  \hspace{1cm} (78)

where

$$m_0 = x_s + y_s \varphi(\kappa, t - s) + \theta \left[(t - s) - \varphi(\kappa, t - s)\right]$$

$$\Sigma_0 = \xi^2 \left[t - s - \varphi(\kappa, t - s) - \frac{\kappa}{2} \varphi(\kappa, t - s)^2\right].$$  \hspace{1cm} (79)

The first term in Eq. (78) is the convexity–adjusted density, which is normally distributed with mean $m^c$ and variance $\Sigma^c$

$$m^c = x_s + y_s \varphi(\kappa, t - s) + \theta \left[(t - s) - \varphi(\kappa, t - s)\right] - \frac{\xi^2 \varphi(\kappa, t - s)^2}{2}$$

$$\Sigma^c = \xi^2 \varphi(2\kappa, t - s).$$  \hspace{1cm} (80)

The second term in Eq. (78), we recall (Section 3.6), is the term structure of discount factors. From these discount factors, we can build the term structure of spot rates. The resulting term structure is illustrated in Figure 4 where it is compared to the one obtained by numerical integration of the Green's function obtained by the transform method described in Section 4. The match is essentially perfect, and the result for the term structure is identical to the one in Jamshidian (1989).

The values of the different vectors and matrices characterizing the process are shown in Table 7 and not reproduced in the text.
Table 7: Characterization of the augmented affine process for a one-factor Vasicek interest rate model

\[
\begin{align*}
\frac{dx_t}{dt} &= y_t dt \\
\frac{dy_t}{dt} &= -\kappa (y_t - \theta) dt + \xi dW_t
\end{align*}
\]

\[X_t = \{y_t\}, \quad \tilde{X}_t = \{x_t\}\]

\[\bar{\mu} = \left\{ \begin{array}{c} y_t \\ \theta - \kappa y_t \end{array} \right\}, \quad \bar{K}_0 = \left\{ \begin{array}{c} 0 \\ \theta \end{array} \right\}, \quad \bar{K}_1 = \left[ \begin{array}{cc} 0 & 1 \\ 0 & -\kappa \end{array} \right] \]

\[
\tilde{\sigma} \tilde{\sigma}^T = \begin{bmatrix} 0 & 0 \\ 0 & \xi^2 \end{bmatrix}, \quad \tilde{H}_0 = \begin{bmatrix} 0 & 0 \\ 0 & \xi^2 \end{bmatrix}, \quad \tilde{H}_1^{(0)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \tilde{H}_1^{(1)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}
\]

\[\tilde{A}(t; \tilde{\Omega}) = \begin{bmatrix} A_x(t; \tilde{\Omega}) \\ A_y(t; \tilde{\Omega}) \end{bmatrix}, \quad \tilde{\Omega} = \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} \]

\[\begin{cases} A'_x(t; \tilde{\Omega}) = 0 \\ A'_y(t; \tilde{\Omega}) = A_x(t; \tilde{\Omega}) - \kappa A_y(t; \tilde{\Omega}) \end{cases}, \quad \tilde{B}(t; \tilde{\Omega}) = \kappa \theta A_y(t; \tilde{\Omega}) + \frac{\xi^2}{2} A_y(t; \tilde{\Omega})^2 \]

\[
\begin{align*}
A_x(t; \tilde{\Omega}) &= -j \omega_x \\
A_y(t; \tilde{\Omega}) &= -j \frac{\omega_x^2}{\kappa} + j e^{-\kappa (t-s)} \left[ \frac{\omega_x^2}{\kappa} - \omega_y \right] \\
\tilde{B}(t; \tilde{\Omega}) &= -j \kappa \theta \omega_x \varphi(k, t - s) - \frac{\xi^2 \omega_x^2}{2} \xi^2 (2k, t - s) + j \omega_x \theta [\varphi(k, t - s) - (t - s)] - \frac{\omega_x \omega_y}{2} \xi^2 \varphi(k, t - s)^2 + \frac{\omega_x^2 \xi^2}{2 \kappa^2} \left[ \varphi(k, t - s) + \frac{\kappa}{2} \varphi(k, t - s)^2 - (t - s) \right]
\end{align*}
\]

\[
\begin{aligned}
\log \tilde{\phi}(x_s, y_s, t; \tilde{\Omega}) &= A_x(t; \tilde{\Omega}) x_s + A_y(t; \tilde{\Omega}) y_s + \tilde{B}(t) \\
\tilde{m} &= \left\{ \begin{array}{c} x_s + y_s \varphi(k, t - s) + \theta [(t - s) - \varphi(k, t - s)] \\ y_s e^{-\kappa (t-s)} + \kappa \theta \varphi(k, t - s) \end{array} \right\}, \quad \tilde{\Sigma} = \xi^2 \begin{bmatrix} t - s - \varphi(k, t - s) - \frac{\xi^2}{2} \varphi(k, t - s)^2 & \varphi(k, t - s)^2 \\ \frac{\xi^2}{2} & \varphi(2k, t - s) \end{bmatrix}
\end{aligned}
\]

The function \(\varphi(k, t)\) is defined as \(1 - e^{-kt} / k\). The affine specification of the model characterized by \(K_0(t), K_1(t), H_0(t), H_1(t), L_0(t),\) and \(L_1(t)\) is given in Table 3 for both the original and augmented models.
Figure 4: Comparison of (i) the term structure of interest rates obtained by numerical integration of the Green's function and (ii) the exact solution for the term structure given in Eq. (78).

6.2 Power Option Price

In this section, we price non-linear payoffs that cannot be solved by direct transform methods. We consider the case of power options. While there is probably limited interest in pricing these options, the exercise provides a natural test of the accuracy of the proposed numerical method. When the interest rate process is independent of the state variables appearing in the payoff of the security, the problem of pricing becomes separable, and the PDF of the augmented process can be written as the product of (i) a PDF for the interest rate process and (ii) a PDF for the other factors in the economy

$$
\Psi(\tilde{X}_t, t|\tilde{X}_s) = \Psi^I(\tilde{X}_t, t|\tilde{X}_s) \Psi^F(\tilde{X}_t^F|\tilde{X}_t^F),
$$

where $\tilde{X}_t^F$ are the factors appearing in the model for the interest rate and $\tilde{X}_t^F$ are the others factors in the economy. Accordingly, if the payoff of the security is only a polynomial function of $\tilde{X}_t^f$, we have for the value of a power security with payoff $f(\tilde{X}_T) = (\tilde{X}_T^f)^α$ (Eq. 8)

$$
V_s^α = \int_{\tilde{X}_T} \Psi(\tilde{X}_T, T|\tilde{X}_s)f(\tilde{X}_T^f)e^{-\tilde{X}_T^f(0)} d\tilde{X}_T
= \int_{\tilde{X}_T} e^{-\tilde{X}_T^f(0)} \Psi^F(\tilde{X}_T^F, T|\tilde{X}_s) \int_{\tilde{X}_T^f} \Psi^I(\tilde{X}_T^f, T|\tilde{X}_s)f(\tilde{X}_T^f)d\tilde{X}_T^f d\tilde{X}_T^f
= P_s \int_{\tilde{X}_T} \Psi^F(\tilde{X}_T^F, T|\tilde{X}_s)f(\tilde{X}_T^f)d\tilde{X}_T^f,
$$

32
Figure 5: Probability Density Function of the $S_t$ process obtained by using the Discrete Fourier Transform with $N = 32$ samples (left). Illustration of the payoff of square call and put options (right).

where $P_s^T$ is the price as of time $s$ of a zero maturing at time $T$.

Consider, for instance, the following square-root process for the stock price

$$dS_t = -\eta(S_t - S_\infty)dt + \xi\sqrt{S_t}dW_t. \quad (83)$$

The current stock price is $1.00$ and the instantaneous percentage volatility is assumed to be $35\%$, leading to the following possible parameterization of the process

$$S_0 = 1.00 \quad \xi = 0.35$$
$$S_\infty = 20.00 \quad \eta = 0.02. \quad (84)$$

The 1-year spot rate is assumed to be $5.00\%$.

We proceed to price square call and square put options struck at $1.50$ expiring in $T = 1$ year. The payoffs of these two options are illustrated in Figure 5. Also shown in this figure is the $PDF$ $\psi_I(S_T, T| S_0)$ of the stock price at expiration.

The prices of these options are obtained by numerically integrating the payoffs multiplied by the $PDF$ of the stock price and discounting at the 1-year riskless rate. The resulting values for the call and put options are

$$p = 0.103521$$
$$c = 0.046979$$

$$p + c = 0.150500 \approx e^{-0.05} m_2 = 0.150501 \quad (85)$$

33
In the above equation, the sum of the prices of the two options is compared to the present value of the second order moment centered around $1.50, denoted here by $m_2$, of the distribution of the stock price at expiry. This is the equivalent of a put–call parity relationship for square options. An expression for the second moment $m_2$ can be derived by solving a finite number of linear ODE’s as demonstrated in Section 4.3.

6.3 Stochastic Volatility Model

We now consider a stochastic volatility model characterized by the following risk–neutral process

\[
d y_t = [\alpha - \beta (y_t - y_\infty) - \gamma \left( \sigma_t^2 - \sigma_\infty^2 \right)] dt + \sigma_t dW_t, \\
d (\sigma_t^2) = -\eta \left( \sigma_t^2 - \sigma_\infty^2 \right) dt + 2\xi \sigma_t dW_t, \tag{86}
\]

where $y_t$ is the short term rate and $\sigma_t$ is the volatility of the short term rate. We select the following parameterization

\[
\begin{align*}
\beta &= 0.045 \\
\gamma &= 60 \\
\eta &= 1.3 \\
\xi &= 0.06 \\
\sigma_0 &= \sigma_\infty = 0.16.
\end{align*}
\tag{87}
\]

Note that in choosing this parameterization, we have chosen a high value of the volatility of the short rate to clearly show the effect of the convexity adjustment on the density.

Figure 6 shows the two-dimensional DFT approximation of the density of the process at time $t = s + \Delta t$ where $\Delta t$ is chosen to be one year. Figure 6 also shows the contour plot of the same PDF. In this figure, the density is centered around the conditional expectation of $y_t$ given the information at time $s$. The centered variable is denoted by $p_t$ and defined as

\[
p_t = y_t - \mathbb{E}_s [y_t]. \tag{88}
\]

If we integrate the density along the $y_t$ dimension, we obtain the distribution of the process in $\sigma_t^2$. The process in $\sigma_t^2$ is a square-root process (Eq. 86). The distribution of $\sigma_t^2$ conditional on the value of $\sigma_s^2$ at time $t = s$ is a non-central \chi-squared (Cox, Ingersoll, and Ross, 1985b) with $F$ degrees of freedom and parameter of non-centrality $G$

\[
2Q\sigma_t^2 \overset{\text{dist}}{\longrightarrow} \chi^2(F, G), \tag{89}
\]

34
Figure 6: Probability Density Function of the $p_k, \sigma_k^2$ process obtained by using the Discrete Fourier Transform with $N_y = N_x = 32$ samples (left). Contour plot of the PDF (right)
where

\[ Q = \frac{\eta}{2\xi^2 \left(1 - e^{-\eta(t-s)}\right)} \]

\[ F = \frac{\eta\sigma_0^2}{\xi^2} \]

(90)

\[ G = 2Q\sigma_0^2 e^{-\eta(t-s)}. \]

In Figure 7, the exact solution for the distribution in \( \sigma_t^2 \) is compared to the one obtained by numerical integration of the PDF in Figure 6 along the \( y_t \) axis. Also shown in this figure, is the result of the numerical integration of the PDF along the \( \sigma_t^2 \) axis.

Since the number of nodes chosen to discretize the hypercube is equal to a power of 2, an FFT algorithm can be used in lieu of the DFT algorithm. In that case, the acceleration factor is (Eq. 57)

\[ \left( \frac{32}{\log_2 32} \right)^2 \approx 41. \]

(91)

Thus, the FFT algorithm in this particular case is about 41 times faster than the DFT algorithm.

For the Green's function, we create the augmented process by adding the following factor (Eq. 24)

\[ dx_t = y_t dt. \]

(92)

The Green's function can be obtained by numerical integration of the PDF of the augmented process in \( x_t, y_t, \sigma_t^2 \), which is computed by FFT inversion of a 3-D characteristic function. The resulting Green's function is a function of the state variables \( y_t \) and \( \sigma_t^2 \). The CAD is obtained by dividing the Green's function by its integral over the entire allowable space for the state variables \( y_t, \sigma_t^2 \) according to Eq. (10). Figure 7 illustrates the result of a numerical integration of the CAD along the \( y_t \) and \( \sigma_t^2 \) axes. As shown in this figure, the difference between the CAD of the process and the PDF of the initial process is not negligible.

Next, we derive expressions for the conditional moments and unconditional moments of the distribution in \( y_t, \sigma_t^2 \) by using the methodology developed in Section 4.3. From the expression for the conditional moments for the general AJD in Table 6,
Figure 7: Probability Density Function of the $\sigma_t^2$ (top left) and of the $y_t$ (top right) processes obtained by integrating numerically the Discrete Fourier Transform approximation of the density of $p_t$, $\sigma_t^2$ along the $p_t$ axis (●). The continuous line in the graph on the left is the analytical solution for the density of $\sigma_t^2$ (Eq. 89). Comparison of the CAD of the process (●) with the density of the process. The CAD of the process is obtained by numerical integration of the Green’s function along the $y_t$ axis (bottom left) and the $\sigma_t^2$ axis (bottom right).
we can write the ODE's that must be satisfied by the conditional centered moments

\[
\frac{d}{dt} C_s^{m,n}(t) + (m\beta + mn) C_s^{m,n}(t) = -m\gamma C_s^{m-1,n+1}(t) + \\
\frac{m(m - 1)}{2} \left[ C_s^{m-2,n+1}(t) + E_s \left[ \sigma^2_t \right] C_s^{m-2,n}(t) \right] + \\
2\xi^2 n(n - 1) \left[ C_s^{m,n-1}(t) + E_s \left[ \sigma^2_t \right] C_s^{m-2,n}(t) \right],
\]

where the centered moments $C_s^{m,n}(t)$ are defined as

\[
C_s^{m,n}(t) = E \left[ (y_t - E_s[y_t])^m \right] \left( \sigma^2 - E_s \left[ \sigma^2_t \right] \right)^n.
\]

The fact that the $\sigma_t$ process does not depend on $y_t$ means that the system of differential equations in Eq. (93) is fully recursive. This implies that solving one ODE yields one new moment. The order in which these ODE's for $C_s^{m,n}(t)$ have to be solved for is given by

\[
C_s^{0,2}(t) \quad C_s^{1,1}(t) \quad C_s^{2,0}(t) \\
C_s^{0,3}(t) \quad C_s^{1,2}(t) \quad C_s^{2,1}(t) \quad C_s^{3,0}(t) \\
C_s^{0,4}(t) \quad C_s^{1,3}(t) \quad C_s^{2,2}(t) \quad C_s^{3,1}(t) \quad C_s^{4,0}(t) \\
\ldots
\]

Analytical expressions for the centered conditional moments up to the second order and centered unconditional moments up to the third order are presented in Table 8. In this particular stochastic volatility process, moments of any order can be derived analytically in terms of exponential functions by solving a finite number of linear ordinary differential equations.

### 6.4 Jump-Diffusion Process

We consider the following jump-diffusion process for describing the dynamics of the overnight rate $y_t$

\[
dy_t = -\kappa(y_t - \theta)dt + \sigma dW_t + JdN_t,
\]

where $dN_t$ is a jump process with intensity $\lambda$, and $J$ is a random variable with two equiprobable states at $-\xi$ and $\xi$. We select the following parameterization of the model

\[
\begin{align*}
\kappa &= 0.22 & \xi &= 0.0050 \\
\sigma &= 0.0055 & \lambda &= 10.0 \\
\theta &= 0.06.
\end{align*}
\]
Table 8: Conditional and unconditional centered moments of the distribution in $y_t$ and $\sigma_t^2$

$$
C_{s,t}^{0,2}(t) = 4\xi^2 \left[ \sigma_{\infty}^2 \varphi(2\eta, \Delta t) + (\sigma_s^2 - \sigma_{\infty}^2) \varphi(\eta, \Delta t) e^{-\eta \Delta t} \right] \\
C_{s,t}^{1,1}(t) = -4\gamma \xi^2 \left[ \frac{\sigma_{\infty}^2}{2\eta} \varphi(\beta + \eta, \Delta t) + \left( \sigma_s^2 - \sigma_{\infty}^2 \right) \frac{e^{-\eta \Delta t}}{\eta} \varphi(\beta, \Delta t) + \left( \frac{\sigma_s^2}{2} - \sigma_s^2 \right) \frac{e^{-2\eta \Delta t}}{\eta} \varphi(\beta - \eta, \Delta t) \right] \\
C_{s,t}^{2,0}(t) = \sigma_{\infty}^2 \varphi(2\beta, \Delta t) + \left( \sigma_s^2 - \sigma_{\infty}^2 \right) e^{-\eta \Delta t} \varphi(2\beta - \eta, \Delta t) + 8\gamma^2 \xi^2 \left[ \frac{\sigma_{\infty}^2}{2\eta(\beta + \eta)} \varphi(2\beta, \Delta t) + \left( \frac{\sigma_s^2}{2} - \sigma_s^2 \right) \frac{e^{-2\eta \Delta t}}{\eta(\beta - \eta)} \varphi(2(\beta - \eta), \Delta t) + \left( \sigma_s^2 - \sigma_{\infty}^2 \right) \frac{e^{-\eta \Delta t}}{\eta(\beta - \eta)} \varphi(2\beta - \eta, \Delta t) + \left( \sigma_s^2 - \sigma_{\infty}^2 \right) \frac{e^{-\eta \Delta t}}{\beta(\beta - \eta)} \varphi(\beta - \eta, \Delta t) \right] \\
C_{s,t}^{0,3} = \frac{8\xi^4 \sigma_{\infty}^2}{\eta^2 \beta \eta} \\
C_{s,t}^{1,2} = -\frac{8\gamma \xi^4 \sigma_{\infty}^2}{\eta^2 (\beta + \eta)} \\
C_{s,t}^{2,1} = \frac{2\xi^2 \sigma_{\infty}^2}{\eta(2\beta + \eta)} \left[ 1 + \frac{\gamma^2 \xi^4}{\eta(\beta + \eta)} \right] \\
C_{s,t}^{3,0} = -\frac{2\gamma^2 \xi^4 \sigma_{\infty}^2}{\eta^2 (\beta + \eta)(2\beta + \eta)} \left[ 3\beta + 2\eta + 2\gamma^2 \xi^4 \right]
$$

The parameters $\eta$, $\beta$, $\xi$, $\gamma$, $\sigma_{\infty}$, and $\sigma_0$ refer to the stochastic volatility model in Eq. (86).
The addition of a jump term is justified in that it allows to model discrete changes in interest rate levels of 50 bp (parameterized here by $\xi$) during a period of active monetary policy. The initial value of the overnight rate is assumed to be 5.00%.

We select a one month horizon. The resulting probability density function of the state variable $y_t$ is illustrated in Figure 8. Note the presence of the two peaks in the PDF for the overnight rate $y_t$ at 5.00% ± 50 bp.

The Green's function is obtained by adding a state variable $x_t$, equal to the integral of the short rate (Eq. 24). The PDF of the joint process of $x_t, y_t$ is shown in Figure 9 (left). By inspection, we note that the overnight rate $y_t$ and the integral of the overnight rate $x_t$ are strongly correlated in level. This results in a loss of accuracy of the method since a lot of sample points are wasted in the numerical integration scheme used to obtain the PDF of the augmented process, and illustrates the importance of the rotation procedure proposed in Section 4.2. Figure 9 shows the same PDF for the joint process $x_t, y_t$ after a rotation of the state space. The procedure to select the rotation matrix is described in Section 4.2.

We now integrate numerically the PDF of the joint process in the rotated state space to obtain the Green's function of the process. From the Green's function, one can compute the one month rate implied by Eq. (95). By repeating this operation for different horizons, one can compute the term structure of interest rates. This is illustrated in Figure 10 where the term structure in the presence of a jump is
Figure 9: PDF of the joint process $x_t, y_t$ (Eqs. 95) in the initial state space (left) and in the rotated state space (right).
compared to the term structure of the same model without the jump term. Note that the jump adds to the second and higher order moments, and depresses the term structure for the usual convexity reason.

To gain confidence in the accuracy of the method, we also compute the term structure of interest rates directly from the solution for discount factors \( P(X, t; \tau) = \exp \left( A^T(t; \tau) X + B(t; \tau) \right) \) (see details in Appendix A.2). For the process described by Eq. (95), Eq. (110) becomes

\[
\begin{align*}
B'(t; \tau) + \kappa \theta A(t; \tau) + \frac{\sigma^2}{2} A(t; \tau)^2 + \lambda \mathbb{E}_J \left[ e^{J(t)A(t; \tau)} - 1 \right] &= 0 \\
A'(t; \tau) - \kappa A(t; \tau) &= 1,
\end{align*}
\]

subjected to the final condition

\[
A(\tau; \tau) = 0 \quad B(\tau; \tau) = 0.
\]

In Eq. (97), the last term in the expression for \( B(t; \tau) \) can be computed based on the fact that \( J(t) \) takes the values \( \pm \xi \) with probability 0.5

\[
\mathbb{E}_J \left[ e^{J(t)A(t; \tau)} - 1 \right] = \frac{1}{2} \left[ e^{\xi A(t; \tau)} - 1 \right] + \frac{1}{2} \left[ e^{-\xi A(t; \tau)} - 1 \right] = \cosh \left[ \xi A(t; \tau) \right] - 1.
\]

The term structure is obtained from Eq. (109) after numerical integration of the system of ODE's in Eq. (97)

\[
P(y_0, 0; \tau) = e^{A(0; \tau)y_0 + B(0; \tau)},
\]

where \( P(y_0, 0; \tau) \) is the price as of time 0 of a zero coupon bond maturing at time \( \tau \) and \( y_0 \) is the initial value of the state variable \( y_t \). The resulting term structure of interest rates is shown in Figure 10.

7 Affine Approximation of Non–Affine Models

This section presents a procedure by which one can still apply the methods developed above to the case where the model is not affine.

One way to handle non–affine models is to allow state–contingent discrete time changes of the affine specification of the model defined in Eq. (12) and illustrated in Figure 11. The dates corresponding to changes in the affine specification of the model are denoted by \( s_1, s_2, \ldots \) and between any such two dates \( s_n \) and \( s_{n+1} \), the state contingent affine specification of the model is characterized by the parameters
Figure 10: Term structure of interest rates implied by the Jump-Diffusion process for the overnight rate in Eq. (95) by numerical integration of the Green’s function (●) and analytically from Eq. (100). Comparison with the term structure of interest rates implied by the same process without the jump.

\[ \mu_{s_n}(X_t, t), \sigma_{s_n}(X_t, t), J_{s_n}(t), \text{and } \lambda_{s_n}(X_t, t) \] which are contingent on the state of the process at time \( s_n \). In other words, these parameters are a function of \( X_{s_n} \) between times \( s_n \) and \( s_{n+1} \). The idea here is to approximate a non-affine specification for a model by an affine model in order to apply the techniques developed for affine models.

In particular, this procedure applies when the overnight rate is not affine in the underlying factors of the economy. Suppose that the overnight spot rate is a function \( f \) of the state vector \( X_s \)

\[ r_s = f(X_s). \]  

(101)

In the neighborhood of \( X_s \), the function \( f \) of \( X_t \) is assumed to be approximated by an affine function of the state vector

\[ r_t = f(X_t) \approx \alpha_s^T X_t + \beta_s \quad t > s, \]  

(102)

where \( \alpha_s \) is an \( n \times 1 \) vector and \( \beta_s \) is a scalar. We refer to Figure 11 for the definition of the time window between \( s \) and \( t \). The superscript \( T \) denotes the transpose. We consider the following two scenarios for the selection of the vector \( \alpha_s \) and the scalar \( \beta_s \).
Figure 11: Illustration of the discrete time state-contingent changes of the affine specification of the model. $\alpha$ and $\beta$ are defined in Eq. (23)
1. The constants are chosen to be as follows

\[
\alpha_s = \left. \frac{df(X_t)}{dX_t} \right|_{t=s} \quad \beta_s = f(X_s) - \alpha_s^T X_s,
\]

(103)

where the differentiation symbol \( d \) denotes the derivative taken with respect to each component of the vector \( X_t \).

2. We select the vector \( \alpha_s \) and the scalar \( \beta_s \) such as to minimize the following objective function

\[
\begin{align*}
\{ \alpha_s \} &= \text{arg } \min_{\{ \alpha_s \}} E_a \left\{ e^{-\int_{t}^{u=1} f(X_u) du} - e^{-\int_{t}^{u=1} [\alpha_s^T X_u + \beta_s] du} \right\}^2 \\
\{ \beta_s \}
\end{align*}
\]

(104)

In the above, the expected values can be obtained by running Monte-Carlo simulations. Note that we have written the vector \( \alpha_s \) and scalar \( \beta_s \) as a function of state \( X_t \) at time \( s \). This means that the affine characterization of the overnight rate can be state dependent when looking at the conditional Green’s function given the state of the process at time \( s \).

8 Conclusions

In this paper, we have proposed a general methodology to price contingent claims when the interest rates and other asset prices are driven by a multivariate Affine Jump–Diffusion process. We have implemented a scheme based on numerical transform methods, and gauged the accuracy of the method by using a number of examples for which analytical results are available. Given the richness of the current examples considered, we are comfortable with the accuracy of the method. We also believe that it allows to highlight particular features of the distribution implied by an Affine Jump–Diffusion model, and that it can serve as a tool to detect problems in alternative implementations. Consider the case of a jump for instance; our method suggests that the resulting distribution can be multimodal. It is difficult to envision that a tree-based implementation would easily recover the correct state prices without some form of tinkering with the implementation.
A Appendix

A.1 Characteristic Function of a Gaussian

The probability density function of an \( n \)-dimensional Gaussian is

\[
N(m, \Sigma) \propto e^{-\frac{1}{2}(x-m)^T \Sigma^{-1} (x-m)}
\]

where \( m \in \mathbb{R}^n \) is an \( n \times 1 \) vector and \( \Sigma \in \mathbb{R}^{n\times n} \) is the variance-covariance matrix. The Fourier Transform of the PDF is

\[
N(m, \Sigma) \xrightarrow{F} e^{-\frac{1}{2} m^T \Sigma^{-1} m}.
\]

A.2 Term Structure of Interest Rates

For the AJD process described in Eq. (11), the price at time \( t \) of a zero-coupon bond maturing at time \( \tau \), denoted by \( P(X, t; \tau) \) satisfies the following backward Kolmogorov PDE

\[
\begin{align*}
\mu_i(X, t) \frac{\partial P(X, t; \tau)}{\partial x_i} &+ \frac{1}{2} \left[ \sigma_i(X, t) \sigma_j(X, t) \right] \frac{\partial^2 P(X, t; \tau)}{\partial x_i \partial x_j} \\
&+ \lambda_k(X, t) \mathbb{E}_J \left[ P(X + J_k(t; \tau), t; \tau) - P(X, t; \tau) \right] = r_t P(X, t; \tau) \\
&\quad \text{for } i, j = 1, \ldots, n, k = 1, \ldots, l,
\end{align*}
\]

where \( r_t \), the short rate, is defined in Eq. (23). The above PDE is subjected to the final condition

\[
P(X, \tau; \tau) = 1.
\]

We guess a solution of the form

\[
P(X, t; \tau) = e^{A^T (r, \tau) X + B(t, \tau)}.
\]

The PDE is satisfied if

\[
\begin{align*}
B'(t; \tau) + K_0^T(t) A(t; \tau) &+ \frac{1}{2} A^T(t; \tau) H_0(t) A(t; \tau) + L_0(t)^T \mathbb{E}_J \left[ e^{J(t) A(t; \tau)} - 1 \right] = \beta(t) \\
A'(t; \tau) + K_1^T(t) A(t; \tau) &+ \frac{1}{2} A^T(t; \tau) H_1(t) A(t; \tau) + L_1(t)^T \mathbb{E}_J \left[ e^{J(t) A(t; \tau)} - 1 \right] = \alpha(t),
\end{align*}
\]

46
subjected to the final condition

$$A^T (\tau; \tau) = \left\{ \begin{array}{c} 0 \\
0 \end{array} \right\} \quad B(\tau; \tau) = 0. \quad (111)$$

A.3 Convexity Adjustment to the PDF in the Vasicek Model

In this section, we derive the Green's function from the PDF of the augmented process, which is assumed to be normal with mean \(\tilde{m}\) and variance-covariance matrix \(\tilde{\Sigma}\)

$$X_t \xrightarrow{\text{dist}} N(\tilde{m}, \tilde{\Sigma}). \quad (112)$$

\(\tilde{m}\) is an \(\tilde{n} \times 1\) vector and \(\tilde{\Sigma}\) is an \(\tilde{n} \times \tilde{n}\) matrix. We suppose that \(x_0\) is the additional factor. Therefore, the mean and variance-covariance matrix of the augmented process can be written in terms of the mean and variance of the original process according to

$$\tilde{m} = \left\{ \begin{array}{c} m_0 \\
m \end{array} \right\} \quad \tilde{\Sigma} = \left[ \begin{array}{cc} \tilde{\Sigma}_{00} & \tilde{\Sigma}_{0}^T \\
\tilde{\Sigma}_{0} & \Sigma \end{array} \right]. \quad (113)$$

We denote by \(\tilde{Q}\) the inverse of the variance-covariance matrix

$$\tilde{Q} = \tilde{\Sigma}^{-1} = \left[ \begin{array}{cc} \tilde{Q}_{00} & \tilde{Q}_{0}^T \\
\tilde{Q}_{0} & Q \end{array} \right]. \quad (114)$$

Since \(\tilde{Q}\) is the inverse of \(\tilde{\Sigma}\), the following must hold

$$\tilde{Q}_{00} \tilde{\Sigma}_{00} + \tilde{Q}_{0}^T \tilde{\Sigma}_{0} = 1$$
$$\tilde{Q}_{00} \tilde{\Sigma}_{0}^T + \tilde{Q}_{0}^T \Sigma = 0$$
$$\tilde{Q}_{0} \Sigma_{00} + Q \Sigma_{0} = 0$$
$$\tilde{Q}_{0} \Sigma_{0}^T + Q \Sigma = I_n, \quad (115)$$

where \(I_n\), we recall, is the \(n \times n\) identity matrix. The general form for the PDF of an \(n\)-variate Gaussian is given in Eq. (105). With this definition, we now compute

$$\frac{1}{(2\pi)^{\frac{n+1}{2}}} \sqrt{\det \Sigma} \int_{x_0 = -\infty}^{x_0 = \infty} e^{-\frac{1}{2} (x - \tilde{m})^T \tilde{Q} (x - \tilde{m})} e^{-\delta x_0} dx_0, \quad (116)$$

47
where $\delta$ is a constant chosen to be one if we wish to compute the Green’s function of the original process and to be zero if we wish to compute the PDF of the original process. $\bar{x}$ is a realization of the random multi-variate

$$\bar{x} = \begin{Bmatrix} x^{(0)} \\ x^{(1)} \\ \vdots \\ x^{(n)} \end{Bmatrix}.$$  

(117)

Even though the PDF of the Gaussian can be computed directly, we keep the constant $\delta$ in the integral above to clearly show the effect of the convexity adjustment.

The computation proceeds by rearranging the argument of the exponential based on the definition of $\bar{x}$, $\bar{m}$, and $\Sigma$

$$a(\bar{x}) = -\frac{1}{2} (\bar{x} - \bar{m})^T \bar{Q} (\bar{x} - \bar{m}) - \delta x_0.$$  

(118)

Substitution of Eqs. (113)–(115) into the above expression yields

$$a(\bar{x}) = -\frac{1}{2} (x - m)^T Q(x - m) + \frac{\delta^2}{2Q_{00}} + \frac{\delta}{Q_{00}}(x - m)^T \bar{Q}_0$$

$$+ \frac{(x - m)^T \bar{Q}_0 \bar{Q}_0^T (x - m)}{2Q_{00}} - \delta m_0$$

$$- \frac{1}{2} \bar{Q}_{00} \left[ x_0 - m_0 + \frac{\delta + (x - m)^T \bar{Q}_0}{Q_{00}} \right]^2.$$  

(119)

From Eq. (115), we observe that

$$\left( Q - \frac{\bar{Q}_0 \bar{Q}_0^T}{Q_{00}} \right) \Sigma = I_n.$$  

(120)

With this observation, $a(\bar{x})$ can be further rearranged

$$a(\bar{x}) = -\frac{1}{2} \left( x - m - \frac{\delta}{Q_{00}} \Sigma \bar{Q}_0 \right)^T \Sigma^{-1} \left( x - m - \frac{\delta}{Q_{00}} \Sigma \bar{Q}_0 \right) - \delta m_0$$

$$+ \frac{\delta^2}{2Q_{00}} \left[ 1 + \frac{\bar{Q}_0^T \Sigma \bar{Q}_0}{Q_{00}} \right] - \frac{\bar{Q}_{00}}{2} \left[ x_0 - m_0 + \frac{\delta + (x - m)^T \bar{Q}_0}{Q_{00}} \right]^2.$$  

(121)

Again, from Eq. (115), we note the following

$$\frac{\Sigma \bar{Q}_0}{Q_{00}} = -\bar{\Sigma}_0 \quad \frac{\bar{Q}_0^T \Sigma \bar{Q}_0}{Q_{00}} = \bar{Q}_{00} \bar{\Sigma}_{00} - 1.$$  

(122)
The argument of the exponential in Eq. (116) is now a separated quadratic form in $x_0$ and $x$

$$a(x) = -\frac{1}{2} \left( x - m + \delta \Sigma_0 \right)^T \Sigma^{-1} \left( x - m + \delta \Sigma_0 \right) - \delta m_0$$

$$+ \frac{\delta^2}{2} \Sigma_{00} - \frac{\tilde{Q}_{00}}{2} \left[ x_0 - m_0 + \frac{\delta + (x - m)^T \tilde{Q}_0}{Q_{00}} \right]^2.$$  \hspace{1cm} (123)

Since $\tilde{Q}$ is the inverse of $\tilde{\Sigma}$, the determinants of $\Sigma$ and $\Sigma$ are related by

$$\tilde{Q}_{00} \det \tilde{\Sigma} = \det \Sigma.$$  \hspace{1cm} (124)

We substitute the above expression into Eq. (116) and integrate along the $x^{(0)}$ axis to arrive at

$$e^{-\delta m_0 + \frac{\delta^2}{2} \Sigma_{00}} e^{-\frac{1}{2} (x-m+\delta \Sigma_{00})^T \Sigma^{-1} (x-m+\delta \Sigma_{00})} \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}}.$$  \hspace{1cm} (125)

The Green's function is obtained by setting $\delta = 1$ in the above expression and the density is obtained by setting $\delta = 0$.

### A.4 Laurent–Schwartz Formalism

The multi-index notation introduced by Laurent–Schwartz (see Friedman, 1969 or Hörmander, 1963) is very useful when dealing with multi-variate PDEs or stochastic processes in order to avoid numerous indices. For the reader's convenience, we redefine the notation here.

The zero vector and unit vector are

$$0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \quad 1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$  \hspace{1cm} (126)

The vectors $x, y \in \mathbb{R}^n$ and $\alpha \in \mathbb{Z}^n$ are defined

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \quad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \quad \alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}.$$  \hspace{1cm} (127)
The usual operations on a vector correspond to the following vector operations

\[(x + y)_i = x_i + y_i \quad (x y)_i = x_i y_i\]

\[(x/y)_i = x_i/y_i \quad (f(x))_i = f(x_i)\].

(128)

Next, we have the following monomials

\[|\alpha| = \sum_{i=1}^{n} \alpha_i \quad x^\alpha = \prod_{i=1}^{n} x_i^{\alpha_i} \quad \alpha! = \prod_{i=1}^{n} \alpha_i!\].

(129)

The differential operator is

\[D^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}\].

(130)

An ordering relation is defined

\[\alpha \geq \beta \text{ when } \alpha_i \geq \beta_i \quad \forall i = 1, \ldots, n\]

(131)

\[\alpha \geq \beta \iff \beta \leq \alpha\].

For a second order tensor \(H\), \(H_{\beta\gamma}\) is the tensor obtained by taking the lines and columns of \(H\) corresponding to the non-zero elements in \(\beta\) and \(\gamma\).

Finally, the Newton binomial is

\[(x + y)^\alpha = \sum_{\beta + \gamma = \alpha} \frac{\alpha!}{\beta! \gamma!} x^\beta y^\gamma\].

(132)

### A.5 Pricing Probability

In this section, we generalize to multiple exercise events the concept of risk–neutral probabilities of an option expiring in the money derived by Chacko and Das (1999) in the case of a single event. We also relate the pricing probabilities in Chacko and Das (1999) to our concept of convexity–adjusted density. In other words, we generalize to Bermuda–style options the concept introduced for European–style options in Chacko and Das (1999).

The event \(i\) is assumed to be the possible exercise of an option struck at \(K\)

\[V_{i_{-1}} = \max(V_{i_{-1}}^+, K)\].

(133)
Substitution into Eq. (73) yields

\[ V_{t_{i-1}}^{+} = \int_{X_{t_{i}} \in \mathcal{D}} G(X_{t_{i}}, t_{i} | X_{t_{i-1}}) \max(V_{t_{i}}^{+}, K) dX_{t_{i}}, \quad (134) \]

which can be rewritten as

\[ V_{t_{i-1}}^{+} = \int_{X_{t_{i}} \in \mathcal{D}} G(X_{t_{i}}, t_{i} | X_{t_{i-1}}) \mathbb{1}\{V_{t_{i}}^{+} \geq K\} V_{t_{i}}^{+} dX_{t_{i}} - K \int_{X_{t_{i}} \in \mathcal{D}} G(X_{t_{i}}, t_{i} | X_{t_{i-1}}) \mathbb{1}\{V_{t_{i}}^{+} \geq K\} V_{t_{i}}^{+} dX_{t_{i}}, \quad (135) \]

where \(1_{\{S\}}\) is an indicator function equal to unity if \(S\) is true and zero otherwise. In the second integral, we replace the Green’s function by the discounted convexity-adjusted density

\[ V_{t_{i-1}}^{+} = \int_{X_{t_{i}} \in \mathcal{D}} G(X_{t_{i}}, t_{i} | X_{t_{i-1}}) \mathbb{1}\{V_{t_{i}}^{+} \geq K\} V_{t_{i}}^{+} dX_{t_{i}} - K P_{t_{i}}^{D} \int_{X_{t_{i}} \in \mathcal{D}} \Psi^c(X_{t_{i}}, t_{i} | X_{t_{i-1}}) \mathbb{1}\{V_{t_{i}}^{+} \geq K\} dX_{t_{i}}. \quad (136) \]

The second integral can be interpreted as a probability for the option to be in-the-money. If the payoff \(V_{t_{i}}^{+}\) is positive, the first integral is always positive and if normalized properly can be interpreted as a probability. Accordingly, we define

\[ \Pi_{0, t_{i-1}} = \int_{X_{t_{i}} \in \mathcal{D}} G(X_{t_{i}}, t_{i} | X_{t_{i-1}}) V_{t_{i}}^{+} dX_{t_{i}} \]

\[ \Pi_{1, t_{i-1}} = \int_{X_{t_{i}} \in \mathcal{D}} G(X_{t_{i}}, t_{i} | X_{t_{i-1}}) \mathbb{1}\{V_{t_{i}}^{+} \geq K\} V_{t_{i}}^{+} dX_{t_{i}} \]

\[ \Pi_{2, t_{i-1}} = \int_{X_{t_{i}} \in \mathcal{D}} \Psi^c(X_{t_{i}}, t_{i} | X_{t_{i-1}}) \mathbb{1}\{V_{t_{i}}^{+} \geq K\} V_{t_{i}}^{+} dX_{t_{i}}. \quad (137) \]

With this notation, Eq. (136) becomes

\[ V_{t_{i-1}}^{+} = \Pi_{0, t_{i-1}} \Pi_{1, t_{i-1}} - K P_{t_{i-1}}^{D} \Pi_{2, t_{i-1}}. \quad (138) \]

This constitutes a recursive generalization of the pricing equation in Chacko and Das (1999). The probabilities keep the same interpretation for Bermuda-style options as they do for European-style options, namely that they measure in the risk-neutral world the probability for the option value to exceed its intrinsic value.
References


