Unifying Gaussian Dynamic Term Structure Models from an HJM Perspective[*]

HAITAO LI[†]   XIAOXIA YE[‡]

Cheung Kong Graduate School of Business   University of Bradford

FAN YU[§]

Claremont McKenna College

Abstract

We show that the unified HJM-based approach of constructing Gaussian dynamic term structure models developed by Li, Ye, and Yu (2016) nests most existing GDTSMs as special cases. We also discuss issues of interest rate derivatives pricing under this approach and using integration to construct Markov representations of HJM models.

Keywords: Gaussian Dynamic Term Structure Models; HJM; Finite Dimensional Realizations; Interest Rate Derivatives

JEL classification: C61; E43; E44; G12

1 Introduction

Gaussian dynamic term structure models (GDTSM) have been extensively used in the literature to address a wide range of topics due to their tractability and robust performance. For example, Dai and Singleton (2002) examine the ability of GDTSMs in explaining the expectation puzzles; Joslin, Singleton, and Zhu (2011) study the implications of GDTSMs on no-arbitrage constraints and forecasts of yield factors; Ang and Piazzesi (2003) and Joslin, Le, and Singleton (2013) analyze Gaussian term structure models with macroeconomic factors. Most existing studies on GDTSMs, however, follow the affine term structure modeling framework of Duffie and Kan (1996) and Dai and Singleton (2000). This approach starts by assuming that the risk-free short rate is an affine function of a finite number of Markov state variables, which follow affine diffusions. The affine structure of the model leads to analytical solutions for the prices of a wide range of fixed-income securities.

Alternatively, one can construct GDTSMs as Markov representations or finite dimensional realizations (FDRs hereafter) of Heath, Jarrow, and Morton (1992) models (HJM hereafter).[1] By focusing on the instantaneous forward rate, HJM models are especially convenient for pricing interest rate derivatives because the risk neutral drift of the forward rate is completely determined by the volatility of the forward rate. Though the path-dependent nature of general HJM models makes their empirical implementations challenging, a large literature has been developed to construct Markov representations of HJM models.[2] For example, Bhar and Chiarella (1997) and Björk and Gombani (1999) impose restrictions on the volatility function of forward rates that permit Markov representations of Gaussian HJM models.

In this paper, we provide one of the first comprehensive studies of the connections between the two different approaches to constructing GDTSMs, which have been developed in parallel until recently. Specifically, we show that the systematic approach developed in Li, Ye, and Yu (2016) (LYY hereafter) nests most of the existing GDTSMs as special cases. This approach obtains Markov representations of Gaussian HJM models based on the linear

---

[1]Throughout the paper, we use Markov representation and finite-dimensional realization interchangeably.

[2]See e.g., Carverhill (1994), Ritchken and Sankarasubramanian (1995), Bliss et al. (1996), Bhar and Chiarella (1997), Inui and Kijima (1998), De Jong and Santa-Clara (1999), Björk and Gombani (1999), Björk and Svensson (2001), Björk and Landén (2002), and Chiarella and Kwon (2003), among others.

system approach of Björk and Gombani (1999). While the "direct integration" approach of Bhar and Chiarella (1997) leads to one Markov representation of a given HJM model,[3] the linear system approach leads to infinitely many equivalent Markov representations for the same HJM model based on the so-called "invariant transforms" that have been widely used in the term structure literature.[4]

The volatility specification in LYY guarantees Markov representation of Gaussian HJM models. We show that most existing GDTSMs developed under the affine framework can be derived as FDRs of Gaussian HJM models with the new volatility function based on the linear system approach. The "invariant transform" reveals close connections among the GDTSMs: though these models look very different from the affine modeling perspective, they can all be traced back to the HJM model with the same volatility specification.

The structure of the paper is as follows. In Section 2, we review the systematic HJM-based approach of GDTSMs developed by LYY. In Section 3, we show that most of the existing GDTSMs can be nested as special cases in our approach given the proposed volatility function. Section 4 concludes the paper. The appendices contain technical details and proofs.

2   Markov Representation of HJM Models

In this section, we review the systematic HJM-based approach of GDTSMs developed by LYY. To keep this paper self-contained, we reconstruct Sections III.A and III.B and Appendix A in LYY here by adding some concrete examples while limiting technical details. We first introduce the HJM framework for term structure modeling. In particular, we emphasize that HJM models generally do not have a Markov representation. Then we discuss a general approach for obtaining FDRs of HJM models based on the linear system approach. We also discuss how to modify the time-inhomogeneous forward curve under the HJM model to make it compatible with the time-homogeneous forward curve under traditional GDTSMs, as well as the implication of the HJM approach for interest rate derivatives pricing. Finally,

---

[3]Bhar and Chiarella (1997) obtain a Markov representation of an HJM model with a time-invariant and hump-shaped volatility function by directly integrating out the forward rate.

[4]Dai and Singleton (2000) use invariant transforms to classify all affine term structure models into a few subclasses of canonical representations.

we propose a general volatility function and derive a base realization for this volatility specification.

2.1 The HJM Framework

Let $f(t, T)$ be the instantaneous forward rate at time t for a future date $T > t$, which represents the rate that can be contracted at time t for instantaneous risk-free borrowing or lending at time T. Given $f(t, s)$ for all s between t and T, the price at time t of a zero-coupon bond with maturity T can be obtained as

$$P(t, T) = \exp\left\{-\int_t^T f(t, s)\, ds\right\}.$$

The spot interest rate at time t is simply $r_t = f(t, t)$.

HJM model term structure dynamics through the stochastic evolution of the forward rates,

$$df(t, T) = \mu(t, T)\, dt + \sigma(t, T)\, dW_t,$$

where $W_{m \times 1}$ is an m-dimensional independent Wiener process under the Q measure, $\mu(t, T)$ and $\sigma(t, T)_{1 \times m}$ are the instantaneous drift and volatility, respectively. The processes μ and σ are adapted processes in R and R^m, respectively, such that the forward rate processes are well defined. HJM establish the following no-arbitrage restriction on the drift of the forward rate process,

$$\mu(t, T) = \sigma(t, T)\left(\int_t^T \sigma(t, s)\, ds\right)^{\mathsf{T}}.$$

Therefore, the volatility function $\sigma(t, T)$ completely determines the drift of the forward rate under the Q measure.

The dynamics of the short rate is given by

$$dr_t = \left.\frac{\partial f(t, \tau)}{\partial \tau}\right|_{\tau = t} dt + \sigma(t, t)\, dW_t, \tag{1}$$

where

$$\left.\frac{\partial f(t,\tau)}{\partial \tau}\right|_{\tau=t} = \left.\frac{\partial f(0,\tau)}{\partial \tau}\right|_{\tau=t} + \int_0^t \left.\frac{\partial \sigma(v,\tau)}{\partial \tau}\right|_{\tau=t} \left(\int_v^t \sigma(v,s)\,ds\right)^{\mathsf{T}} dv \qquad (2)$$

$$+ \int_0^t \sigma(v,t)\,\sigma(v,t)^{\mathsf{T}}\,dv + \int_0^t \left.\frac{\partial \sigma(v,\tau)}{\partial \tau}\right|_{\tau=t}\,dW_v.$$

In general, the short rate is not Markovian, since the drift term in (1), $\left.\frac{\partial f(t,\tau)}{\partial \tau}\right|_{\tau=t}$, involves integrations over the entire history of the forward rate.[5]

For convenience, we consider the Musiela parameterization, which uses time to maturity x, rather than time of maturity T, to parameterize bonds and forward rates.

Definition 1 *For all $x \geqslant 0$, the forward rate in the Musiela parameterization, $r(t,x)$, is defined as*

$$r(t,x) = f(t,t+x) \ and \ P(t,T) = \exp\left\{-\int_0^{T-t} r(t,s)\,ds\right\}.$$

Following Brace and Musiela (1994), the standard HJM drift condition can be rewritten as:

$$dr(t,x) = \mu_r(t,x)\,dt + \sigma(t,x)\,dW_t,$$

$$\mu_r(t,x) = \frac{\partial}{\partial x}r(t,x) + \sigma(t,x)\int_0^x \sigma(t,s)^{\mathsf{T}}\,ds,$$

where $W_t = \left[[w_i]_{i=1}^m\right]^{\mathsf{T}}$, $\sigma(t,x) = \sigma^f(t,t+x) = [\sigma_i(t,x)]_{i=1}^m$, and $[\bullet_i]_{i=1}^m$ is a compact notation for a row vector $[\bullet_1, \bullet_2, \cdots, \bullet_m]$. We then have:

$$r(t,x) = r(0,t+x) + \Theta(t,x) + r_0(t,x), \qquad (3)$$

$$\Theta(t,x) = \int_0^t \sigma(s,x+t-s)\int_0^{x+t-s} \sigma(s,\tau)^{\mathsf{T}}\,d\tau ds, \qquad (4)$$

$$r_0(t,x) = \int_0^t \sigma(s,x+t-s)\,dW_s, \qquad (5)$$

$$dr_0(t,x) = \frac{\partial r_0(t,x)}{\partial x}dt + \sigma(t,x)\,dW_t, \quad r_0(0,x) = 0. \qquad (6)$$

---

[5]Note that the first term in (2) is the slope of the initial forward curve, the second and third terms depend on the history of the volatility process, and the last term depends on the histories of both the volatility process and the Brownian motion.

2.2 A General Approach to Markov Representation of HJM Models

The non-Markov nature of HJM models makes their empirical implementations difficult. Consequently, a large literature has been developed to identify conditions that allow Markov representations or FDRs of HJM models. One widely recognized condition for HJM models to exhibit FDRs is a time-invariant volatility function that is a deterministic function of time to maturity and satisfies a multidimensional linear ODE with constant coefficients (see, e.g., Björk and Svensson, 2001, Corollary 5.1 or Chiarella and Kwon, 2003, Assumption 1).[6] Since the time and maturity dependent components of the volatility function are separable, we can obtain the Markov states by integrating over the historical Brownian shocks.

Specifically, for a one-factor HJM model with the following volatility function:

$$\sigma(x) = (\Omega_1 + \Omega_2 x) e^{-kx}, \tag{7}$$

$r_0(t, x)$ under the Q measure is

$$r_0(t, x) = \int_0^t \sigma(x + t - s) \, dW_s.$$

By Bhar and Chiarella (1997)'s "direct integration", we have:

$$Z_t = \left[\int_0^t e^{-k(t-s)} dW_s, \ \int_0^t (t - s) e^{-k(t-s)} dW_s \right]^\mathsf{T},$$

where Z_t satisfies:

$$dZ_t = \begin{bmatrix} -k & 0 \\ 1 & -k \end{bmatrix} Z_t dt + \begin{bmatrix} 1 \\ 0 \end{bmatrix} dW_t.$$

---

[6]Specifically, the ith component of σ, $\sigma_i(x)$ (which is n times differentiable with respect to x), satisfies an nth order ODE of the form

$$\frac{\partial^n}{\partial x^n} \sigma_i(x) - \sum_{j=0}^{n-1} \kappa_{ij}(x) \frac{\partial^j}{\partial x^j} \sigma_i(x) = 0,$$

where $\kappa_{ij}(x)$'s are continuous deterministic functions.

Then, $r_0(t, x)$ becomes

$$r_0(t, x) = \left[(\Omega_1 + \Omega_2 x) e^{-kx}, \quad \Omega_2 e^{-kx} \right] Z_t.$$

Therefore, the above volatility function allows one to obtain an FDR for the HJM model through "direct integration". It is important to point out that while the original HJM model is a one-factor model, the FDR has two state variables. The additional state variable is introduced to make the system Markov.

However, one limitation of the "direct integration" approach is that it leads to only one among many possible FDRs for the same HJM model, and when the volatility function is more complex, using it to obtain FDRs becomes intractable. An alternative way of obtaining FDRs of HJM models has been developed by Björk and Gombani (1999) based on the Linear System Theory, with the following definition of the FDR:

Definition 2 *A triplet* $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$, *where* $\mathbf{A}$ *is an* $n \times n$-*matrix*, $\mathbf{B}$ *is an* $n \times m$-*matrix and* $\mathbf{C}(x)$ *is an* n-*dimensional row-vector function, is called an* n-*dimensional realization of the system* $r_0(t, x)$ *if* $r_0(t, x)$ *has the representation:*

$$r_0(t, x) = \mathbf{C}(x) Z_t, \tag{8}$$

$$dZ_t = \mathbf{A} Z_t dt + \mathbf{B} dW_t, \quad Z_0 = \mathbf{0}. \tag{9}$$

Björk and Gombani (1999) has shown that for an HJM model to have an FDR as in (8)-(9), the volatility function must be written as:

$$\sigma(x) = \mathbf{C}(x)\mathbf{B} = \mathbf{C}_0 \exp(\mathbf{A}x)\mathbf{B},$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}(x)$ are given in Definition 2, and $\mathbf{C}_0 = \mathbf{C}(0)$.

One important advantage of the approach of Björk and Gombani (1999) is that given one FDR of an HJM model, one can apply the similarity transformation (see, e.g., Schutter, 2000) to the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ to construct a new realization

$$\left\{ M\mathbf{A}M^{-1}, M\mathbf{B}, \mathbf{C}(x) M^{-1} \right\}$$

given a nonsingular matrix M. Then

$$r_0(t, x) = \mathbf{C}(x) M^{-1}(MZ_t),$$

$$d(MZ_t) = M\mathbf{A}M^{-1}(MZ_t) dt + M\mathbf{B}dW_t,$$

is another FDR for the HJM model with a new state vector MZ. In the finance literature, these transformations are referred to as "invariant transforms" (see Dai and Singleton, 2000).

Therefore, while the "direct integration" approach can only construct one FDR, the linear system approach is much more flexible in constructing FDRs. The flexibility offered by the linear system approach turns out to be crucial for constructing traditional GDTSMs from Gaussian HJM models.

2.3   Time-homogeneous Forward Curves and Derivatives Pricing

Traditional GDTSMs are time-homogeneous. Though the FDR we obtain for $r_0(t, x)$ is time-homogeneous, the first two components of the forward curve in (3), $r(0, t + x)$ and $\Theta(t, x)$, tend to be time-inhomogeneous. In this section, we construct GDTSMs from HJM models by replacing $r(0, t + x) + \Theta(t, x)$ with $\lim_{t \to \infty} r(0, t + x) + \Theta(t, x)$, essentially assuming the model has evolved from the distant past. De Jong and Santa-Clara (1999) and Trolle and Schwartz (2009) adopt a similar treatment.

Denote $\lim_{t \to \infty} \Theta(t, x)$ as $\Theta^*(x)$, it is shown in Appendix B in LYY that for an invertible $\mathbf{A}$,

$$\Theta^*(x) = \mathbf{C}(x) \left(\mathbf{A}^{-1}\mathbf{B}\mathbf{B}^{\mathsf{T}}(\mathbf{A}^{\mathsf{T}})^{-1} \right) \mathbf{C}^{\mathsf{T}} - \frac{1}{2}\mathbf{C}(x) \left(\mathbf{A}^{-1}\mathbf{B}\mathbf{B}^{\mathsf{T}}(\mathbf{A}^{\mathsf{T}})^{-1} \right) \mathbf{C}(x)^{\mathsf{T}}. \tag{10}$$

Assuming a constant initial forward curve, we have

$$\lim_{t \to \infty} r(0, t + x) \equiv \varphi, \tag{11}$$

where φ is a non-negative real number.

Given (10) and (11), the forward rate can be rewritten as $r(t, x) = \varphi + \Theta^*(x) + \mathbf{C}_0 \exp(\mathbf{A}x) Z_t$,

and the zero coupon bond price is given by:

$$P(t, t+x) = \exp\left(-\int_0^x r(t,s)\, ds\right) = \exp\left(\mathbf{H}(x) - \mathbf{F}(x)^\mathsf{T} Z_t\right), \tag{12}$$

where

$$\mathbf{H}(x) = -\varphi x - \int_0^x \Theta^*(s)\, ds,$$

$$\mathbf{F}(x)^\mathsf{T} = \mathbf{C}_0 \int_0^x \exp(\mathbf{A}s)\, ds = (\mathbf{C}(x) - \mathbf{C}_0)\, \mathbf{A}^{-1}.$$

Therefore, the dynamics of $P(t, t+x)$ is:

$$\frac{dP(t, t+x)}{P(t, t+x)} = f(t,t)\, dt - \mathbf{F}(x)^\mathsf{T} \mathbf{B} dW_t. \tag{13}$$

These results can also be derived using the traditional GDTSM approach by starting from the short rate specification. That is, the short rate is:

$$r(t,0) = \varphi + \Theta^*(0) + \mathbf{C}_0 Z_t, \text{ where } dZ_t = \mathbf{A} Z_t dt + \mathbf{B} dW_t.$$

Then, $\mathbf{H}(x)$ and $\mathbf{F}(x)$ can be solved from the following Riccati equations:

$$\frac{d\mathbf{H}(x)}{dx} = \frac{1}{2}\mathbf{F}(x)^\mathsf{T} \mathbf{B}\mathbf{B}^\mathsf{T}\mathbf{F}(x) - (\varphi + \Theta^*(0)), \quad \frac{d\mathbf{F}(x)}{dx} = \mathbf{A}^\mathsf{T}\mathbf{F}(x) + \mathbf{C}_0^\mathsf{T},$$

with the boundary conditions $\mathbf{H}(0) = 0$, and $\mathbf{F}(0) = \mathbf{0}_{m \times 1}$. We see immediately that these are the same $\mathbf{H}(x)$ and $\mathbf{F}(x)$ presented earlier.

One important advantage of our approach is that the volatility function completely determines the HJM model and is the most important factor for pricing interest rate derivatives. We illustrate this point by briefly discussing the pricing of interest rate caps under the FDR of the HJM model. Since a cap is a portfolio of caplets, we will focus our discussions on the pricing of a caplet. Following Björk (2009, Definition 22.2), the LIBOR forward rate for $[T, T+\delta]$ is defined as:

$$L(t; T, \delta) \equiv -\frac{P(t, T+\delta) - P(t, T)}{\delta P(t, T+\delta)}. \tag{14}$$

Given our Gaussian model, $Caplet\,(t; T, K)$, which is the price of a European call option on the LIBOR rate between $T - \delta$ and T with strike rate K, has a Black-Scholes type analytical expression summarized in the following proposition.[7] Since the underlying LIBOR rate is nonrandom after $T - \delta$, the caplet matures at $T - \delta$, while its payoff is realized at T.

Proposition 3 *$Caplet\,(t; T, K)$, representing the price of a European call option on the LIBOR rate from $T - \delta$ to T, with time-to-maturity $T - t - \delta$ and strike rate K, is given by:*

$$Caplet\,(t; T, K) = \delta P\,(t, T) \left[\left(L\,(t; T - \delta, \delta) + \frac{1}{\delta} \right) N\,(d_1) - \left(K + \frac{1}{\delta} \right) N\,(d_2) \right],$$

$$d_1 = \frac{\ln\left(L\,(t; T - \delta, \delta) + \frac{1}{\delta} \right) - \ln\left(K + \frac{1}{\delta} \right) + \frac{1}{2}\Sigma\,(t, T)}{\sqrt{\Sigma\,(t, T)}}, d_2 = d_1 - \sqrt{\Sigma\,(t, T)},$$

where $N\,(\cdot)$ is the standard normal CDF, $\Sigma\,(t, T) \equiv \int_t^{T - \delta} \left\| \sigma^P\,(t, s + \delta) - \sigma^P\,(t, s) \right\|^2 ds$, and $\sigma^P\,(t, T)$ is the volatility of $P\,(t, T)$.

Proof. See Appendix A. ∎

By (13), we have $\sigma^P\,(t, T) = \mathbf{F}\,(T - t)^{\mathsf{T}}\,\mathbf{B} = (\mathbf{C} - \mathbf{C}\,(T - t))\,\mathbf{A}^{-1}\mathbf{B}$. The caplet pricing formula is linked to the state variables via the LIBOR forward rate $L\,(t; T - \delta, \delta)$, which is a function of the prices of zero coupon bonds. Similarly, the LIBOR and swap rates are linked to the state variables via the prices of zero coupon bonds.

2.4 Volatility Specification and Base Realization

Our discussion so far has focused on the general approach. In this section, we introduce a volatility function specification that guarantees Markov representation of HJM models as well as a base realization that leads to FDRs of HJM models.

Björk and Svensson (2001) show that the most general deterministic volatility function that allows Markov representation of HJM models is the so-called "quasi-exponential" (or QE) function that has the following general form:

$$\sigma_{\text{QE}}\,(x) = \sum_i e^{\lambda_i x} + \sum_j e^{\alpha_j x} \left[p_j\,(x) \cos\,(\omega_j x) + q_j\,(x) \sin\,(\omega_j x) \right],$$

---

[7]In fact, the result given in Proposition 3 is the so-called "Shifted" or "Displaced" LIBOR Market Model. See Brace (2007, Chapter 3).

where λ_i, α_j, and ω_j are real numbers, and p_j and q_j are polynomials. Moreover, $\sigma_{QE}(x)$ can be written as $\mathbf{C}_0 \exp(\mathbf{A}x)\mathbf{B}$ (see Björk and Svensson, 2001, Remark 5.1 and Lemma 5.1).

The volatility function we consider is a special case of the QE function and consists of polynomials and exponentials. Specifically, for a m-factor Gaussian HJM model, our volatility function is

$$\sigma(x) \equiv \underbrace{\left[\begin{bmatrix} 1 & x & \cdots & x^{n_i-1} \end{bmatrix} e^{-k_i x}\right]^I_{i=1}}_{1 \times \sum_{i=1}^I n_i} \times \mathbf{\Omega}_{\left(\sum_{i=1}^I n_i\right) \times m}, \tag{15}$$

where n_i is a natural number, $\begin{bmatrix} 1 & x & \cdots & x^{n_i-1} \end{bmatrix}$ is a $1 \times n_i$ row vector, k_i is a positive real number with $k_i < k_j$ for $i < j$, and $\left[\begin{bmatrix} 1 & x & \cdots & x^{n_i-1} \end{bmatrix} e^{-k_i x}\right]^I_{i=1}$ is a $1 \times \sum_{i=1}^I n_i$ row vector. The matrix $\mathbf{\Omega}$ satisfies the following restrictions:[8]

1. $\mathbf{\Omega}$ is an $\sum_{i=1}^I n_i \times m$ full rank matrix with $\sum_{i=1}^I n_i \geq m$, i.e., Rank $(\mathbf{\Omega}) = m$.

2. In $\mathbf{\Omega}$, any $(\sum_{i=1}^j n_i)$th row is not a zero row, for $j = 1, 2, \ldots, I$.

3. $\mathbf{\Omega}(1,j) + \sum_{k=1}^{I-1} \mathbf{\Omega}\left(1 + \sum_{i=1}^k n_i, j\right) \geq 0$, for $j = 1, 2, \ldots, m$.

4. $\mathbf{\Omega}$ is set to a lower trapezoidal matrix (a generalization of the lower triangular form for a non-square matrix) for the purpose of identification.

For example, for a two-factor HJM model, our general volatility function simplifies to

$$\sigma(x)_{1 \times 2} = \underbrace{\begin{bmatrix} \begin{bmatrix} 1 & x & \cdots & x^{n_1-1} \end{bmatrix} e^{-k_1 x} & \begin{bmatrix} 1 & x & \cdots & x^{n_2-1} \end{bmatrix} e^{-k_2 x} \end{bmatrix}}_{1 \times (n_1+n_2)} \underbrace{\begin{bmatrix} \Omega_1 & \Omega_{(n_1+n_2)+1} \\ \Omega_2 & \Omega_{(n_1+n_2)+2} \\ \vdots & \vdots \\ \Omega_{n_1+n_2} & \Omega_{2(n_1+n_2)} \end{bmatrix}}_{(n_1+n_2) \times 2}. \tag{16}$$

Our volatility function generalizes a similar specification of Bhar and Chiarella (1997), which,

---

[8] For more details of these restrictions, see footnotes 12-15 in LYY.

for a two-factor HJM model, has the following form:

$$\sigma\left(x\right)_{1\times2} = \underbrace{\left[\begin{bmatrix}1 & x & \cdots & x^{n_1-1}\end{bmatrix}e^{-k_1x} \quad \begin{bmatrix}1 & x & \cdots & x^{n_2-1}\end{bmatrix}e^{-k_2x}\right]}_{1\times(n_1+n_2)} \underbrace{\begin{bmatrix} \Omega_1 & 0 \\ \vdots & \vdots \\ \Omega_{n_1} & 0 \\ 0 & \Omega_{2n_1+n_2+1} \\ \vdots & \vdots \\ 0 & \Omega_{2(n_1+n_2)} \end{bmatrix}}_{(n_1+n_2)\times2}, \quad (17)$$

and each component of $\sigma\left(x\right)$ can be written as

$$\sigma_1\left(x\right) = \left(\Omega_1 + \Omega_2 x + \cdots + \Omega_{n_1} x^{n_1-1}\right)e^{-k_1x},$$

$$\sigma_2\left(x\right) = \left(\Omega_{2n_1+n_2+1} + \Omega_{2n_1+n_2+2}x + \cdots + \Omega_{2\times(n_1+n_2)}x^{n_2-1}\right)e^{-k_2x}.$$

We see that (17) is a special case of (16) by restricting certain elements in the Ω matrix to be zero. This minor generalization provides additional flexibility that makes it possible for multi-factor Gaussian HJM models to nest many traditional GDTSMs.

We note that the volatility function presented in (16) is for illustrative purpose only, as out of the $2\left(n_1+n_2\right)$ parameters in Ω, only $2\left(n_1+n_2\right)-1$ are needed to guarantee that Ω is a full rank matrix. In general, only $\frac{\left(2\sum_{i=1}^{I}n_i+1-m\right)m}{2}$ free parameters are allowed in Ω when estimating a model. This restriction can be easily satisfied by setting Ω as a lower trapezoidal matrix.

Given the general volatility function in (15), we immediately have one straightforward Markov representation of the HJM model:

Theorem 4 *For the HJM volatility function defined in (15), one realization triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is:*

$$
\mathbf{C}(x) = \left[\begin{bmatrix} 1 & x & \cdots & x^{n_i-1} \end{bmatrix} e^{-k_i x} \right]_{i=1}^{I}, \mathbf{B} = \mathbf{\Omega},
$$

$$
\mathbf{A} = \underbrace{\begin{bmatrix} \mathbf{A}_1 & & & \\ & \mathbf{A}_2 & & \\ & & \ddots & \\ & & & \mathbf{A}_I \end{bmatrix}}_{n \times n}, \mathbf{A}_i = \underbrace{\begin{bmatrix} -k_i & 1 & & & \\ & -k_i & 2 & & \\ & & -k_i & \ddots & \\ & & & \ddots & n_i - 1 \\ & & & & -k_i \end{bmatrix}}_{n_i \times n_i},
$$

where $n = \sum_{i=1}^{I} n_i$, and $\mathbf{A}$ is in block diagonal form with each block given by $\mathbf{A}_i$, whose non-zero elements are indicated above.

We refer to this realization as the Ascending Jordan-form Realization (AJR) as the system matrix $\mathbf{A}$ is similar to the Jordan Canonical Form, except its off-diagonal elements are a series of ascending numbers between 1 and $n_i - 1$ instead of 1's. We will call this realization the base realization from now on.

We emphasize that given the volatility function (15), the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is a minimal realization in the sense that it has the minimum number of state variables to make the HJM model Markov. The minimal number of states of a linear system is given by the degree of its characteristic polynomial,[9] which is

$$
\dim \left[span \left\{ \left[e^{-k_i x}, x e^{-k_i x}, \cdots, x^{n_i-1} e^{-k_i x} \right]_{i=1}^{I} \right\} \right] = \sum_{i=1}^{I} n_i = n,
$$

and n is also the number of states of the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ as $\mathbf{A}$ is an $n \times n$ matrix. There-fore, all the realizations (based on the volatility function (15)) considered in this paper are minimal, since the dimension of $\mathbf{A}$ remains the same under different invariant transforms.

Although the original HJM model has only m stochastic factors, $n - m$ additional state variables are needed to make the model Markov. Though the new system may look much more complicated than the original one, in most applications a small m and n (e.g., $m = 2$, $n = 5$) are more than sufficient to generate models with excellent performance. The extra

---

[9] It is also called the "McMillan degree". See, e.g., Chen (1999).

state variables provide additional flexibility to capture term structure dynamics, especially the non-Markov property of bond yields (see LYY). Such flexibility is not available to traditional GDTSMs, which are special cases under our framework with $m = n$.

3 Traditional GDTSMs as FDRs of HJM Models

In this section, we show that most traditional GDTSMs in the existing literature are special cases of our HJM model with the general volatility specification in (15). We also show how to obtain each GDTSM as an FDR of our Gaussian HJM model using the base realization, the AJR. For convenience, we list all the models in terms of their original notations in Table 1.

The most popular GDTSMs in the literature are the $\mathbb{A}_0(m)$ models of Dai and Singleton (2000). One common feature of $\mathbb{A}_0(m)$ is $n = m$. In Section 3.1 and 3.2, we present different variations of $\mathbb{A}_0(m)$ as FDRs of our HJM model. In Section 3.3, we present two GDTSMs that are not nested by $\mathbb{A}_0(m)$, with the notable feature of $n > m$.

3.1 Maximal $\mathbb{A}_0(m)$

We use $AM_0(m)$ to denote the maximal model of $\mathbb{A}_0(m)$, in the sense that the model has the maximal number of identifiable free parameters. The corresponding HJM volatility specification for $AM_0(m)$ is:

$$
\sigma(x) = \begin{bmatrix} e^{-k_1 x} & e^{-k_2 x} & \cdots & e^{-k_m x} \end{bmatrix} \Omega_{m \times m}, \tag{18}
$$

where Ω is an $m \times m$ square matrix.

The base realization triplet of (18) is:

$$
\mathbf{A}^{\text{Base}} = \begin{bmatrix} -k_1 & 0 & 0 & 0 \\ 0 & -k_2 & 0 & \vdots \\ 0 & 0 & \ddots & 0 \\ 0 & \cdots & 0 & -k_m \end{bmatrix}, \mathbf{B}^{\text{Base}} = \Omega_{m \times m}, \mathbf{C}^{\text{Base}}(x) = \begin{bmatrix} e^{-k_1 x} & e^{-k_2 x} & \cdots & e^{-k_m x} \end{bmatrix}.
$$

3.1.1 Dai and Singleton (2000)

For the $AM_0(3)$ model considered in Dai and Singleton (2000), the corresponding HJM volatility matrix Ω is a lower-triangular square matrix, i.e.,

$$
\Omega = \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 & \Omega_4 & 0 \\ \Omega_3 & \Omega_5 & \Omega_6 \end{bmatrix},
$$

and the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is given by

$$
\mathbf{A} = \begin{bmatrix} -k_1 & 0 & 0 \\ \frac{\Omega_2}{\Omega_4}(k_1 - k_2) & -k_2 & 0 \\ \frac{\Omega_3\Omega_4(k_1-k_3)-\Omega_2\Omega_5(k_1-k_2)}{\Omega_4\Omega_6} & \frac{\Omega_5}{\Omega_6}(k_2 - k_3) & -k_3 \end{bmatrix},
$$

$$
\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},
$$

$$
\mathbf{C}(x) = \begin{bmatrix} \Omega_1 e^{-k_1 x} + \Omega_2 e^{-k_2 x} + \Omega_3 e^{-k_3 x} & \Omega_4 e^{-k_2 x} + \Omega_5 e^{-k_3 x} & \Omega_6 e^{-k_3 x} \end{bmatrix}.
$$

We refer to this realization as the Lower Triangular Realization (LTR). It is only available when the roots of the minimal polynomial of the linear system are all distinct (cf. Chen, 1999). The transform matrix M that connects the base realization to the LTR is $\left(\mathbf{B}^{\text{Base}}\right)^{-1}$. All of the transform matrices defined in the paper are applied to the base realization by default.

3.1.2 Collin-Dufresne, Goldstein, and Jones (2008)

The Gaussian model considered in Collin-Dufresne et al. (2008) is also maximal. The realization form they employ is the so-called Companion Form Realization (CR). For a three-factor Gaussian model, Ω is also a lower-triangular square matrix, and their triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is

given by:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -k_1 k_2 k_3 & -(k_1 k_2 + k_1 k_3 + k_3 k_2) & -(k_1 + k_2 + k_3) \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & -k_3 \\ k_1^2 & k_2^2 & k_3^2 \end{bmatrix} \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 & \Omega_4 & 0 \\ \Omega_3 & \Omega_5 & \Omega_6 \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^3 \Omega_i & \sum_{i=4}^5 \Omega_i & \Omega_6 \\ -\sum_{i=1}^3 k_i \Omega_i & -\sum_{i=4}^5 k_{i-3} \Omega_i & -k_3 \Omega_6 \\ \sum_{i=1}^3 k_i^2 \Omega_i & \sum_{i=4}^5 k_{i-3}^2 \Omega_i & k_3^2 \Omega_6 \end{bmatrix},$$

$$\mathbf{C}(x) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \exp(\mathbf{A}x).$$

The transform matrix is:

$$M = \begin{bmatrix} 1 & 1 & 1 \\ -k_1 & -k_2 & -k_3 \\ k_1^2 & k_2^2 & k_3^2 \end{bmatrix}.$$

The detailed algorithm for deriving the CR from the volatility function is presented in Appendix B.

One advantage of the CR formulated in Collin-Dufresne et al. (2008) is that the realized state variables are directly observable, which is very convenient for model diagnostics.

3.2 Non-maximal $\mathbb{A}_0(m)$

3.2.1 Fisher (1998)

With restrictions on the parameters of $\mathbb{A}_0(m)$, the model is no longer maximal. For example, Fisher (1998) considers a $\mathbb{A}_0(2)$ model with the constraint that $\Omega_2 = -\Omega_1$. The corresponding volatility function is:

$$\sigma(x) = \begin{bmatrix} e^{-k_1 x} & e^{-k_2 x} \end{bmatrix} \begin{bmatrix} \Omega_1 & 0 \\ -\Omega_1 & \Omega_3 \end{bmatrix}.$$

The triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ that Fisher (1998) employs, which is a variant of the LTR, is given by:

$$\mathbf{A} = \begin{bmatrix} -k_1 & 0 \\ k_2 & -k_2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \Omega_1 \frac{k_2 - k_1}{k_2} & 0 \\ 0 & \Omega_3 \end{bmatrix}, \mathbf{C}(x) = \begin{bmatrix} \frac{k_2}{k_2 - k_1} \left(e^{-k_1 x} - e^{-k_2 x} \right) & e^{-x k_2} \end{bmatrix}.$$

The transform matrix is:

$$M = \begin{bmatrix} \frac{k_2 - k_1}{k_2} & 0 \\ 1 & 1 \end{bmatrix}.$$

3.2.2 Joslin, Singleton, and Zhu (2011)

Joslin et al. (2011) "develop a novel canonical GDTSM in which the pricing factors are observable portfolios of yields." We can obtain the "JPC" model of Joslin et al. (2011) using the Jordan Form Realization (JR) with the following volatility specification,[10]

$$\sigma(x) = \begin{bmatrix} e^{-k_1 x} & e^{-k_2 x} & x e^{-k_2 x} \end{bmatrix} \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 + \Omega_3 & \Omega_5 + \Omega_4 & \Omega_6 \\ \Omega_3 & \Omega_5 & \Omega_6 \end{bmatrix}.$$

The associated triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is:

$$\mathbf{A} = \begin{bmatrix} -k_1 & 0 & 0 \\ 0 & -k_2 & 1 \\ 0 & 0 & -k_2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 & \Omega_4 & 0 \\ \Omega_3 & \Omega_5 & \Omega_6 \end{bmatrix},$$

$$\mathbf{C}(x) = \begin{bmatrix} e^{-x k_1} & e^{-x k_2} & e^{-x k_2} (x + 1) \end{bmatrix},$$

---

[10]Jordan Form Realization (JR) is one of the most popular realizations in Linear System Theory. It is less sensitive to parameter variations, and is therefore a good candidate for practical implementation (see, e.g., Chen, 1999).

and the transform matrix is:

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

We note that in Joslin et al. (2011), specifying the eigenvalues and volatility matrix is analogous to specifying the k_i's and $\mathbf{B}$, respectively, in our framework. However, the volatility matrix in Joslin et al. (2011) can only be square, whereas the matrix $\mathbf{B}$ in our setting can be non-square by specifying the n_i's, which are the orders of the polynomials corresponding to the k_i's. Therefore, we can obtain models that are not covered by Joslin et al. (2011)'s approach.

3.2.3 Christensen, Diebold, and Rudebusch (2011)

The model considered in Christensen et al. (2011) is a non-maximal $\mathbb{A}_0(3)$ with more constraints on the parameters. The corresponding volatility function for their model is:

$$\sigma(x) = \begin{bmatrix} 1 & e^{-k_2 x} & x e^{-k_2 x} \end{bmatrix} \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 & \Omega_4 & 0 \\ \Omega_3 & \Omega_5 & \Omega_6 \end{bmatrix},$$

which is similar to that of "JPC" in Joslin et al. (2011), but with one of the eigenvalues set to zero. However, Christensen et al. (2011) employ a different form of realization, which we refer to as the Quasi-Jordan Form Realization (QJR). Specifically, in Christensen et al. (2011), the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is:

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -k_2 & k_2 \\ 0 & 0 & -k_2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \Omega_1 & 0 & 0 \\ \Omega_2 & \Omega_4 & 0 \\ \frac{\Omega_3}{k_2} & \frac{\Omega_5}{k_2} & \frac{\Omega_6}{k_2} \end{bmatrix}, \mathbf{C}(x) = \begin{bmatrix} 1 & e^{-k_2 x} & k_2 x e^{-k_2 x} \end{bmatrix}.$$

The transform matrix M takes the form:

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{k_2} \end{bmatrix}.$$

Under this realization, the state variables have clear economic interpretations as level, slope, and curvature factors, since their factor loadings coincide exactly with those of Nelson and Siegel (1987).[11]

3.3 Others

In this section, we show that some of the non-Markov Gaussian models with $n > m$ can also fit into our framework. Interested readers can refer to LYY for further implications and empirical analyses of non-Markov models constructed using the systematic approach.

3.3.1 Ang and Piazzesi (2003)

Ang and Piazzesi (2003) use a discrete-time AR(p) process to describe macro variables in their term structure models. However, the corresponding GDTSMs with AR(p) state variables no longer belong to $\mathbb{A}_0(m)$, as the lagged state variables are free of innovation terms. Therefore, there are more states than factors in these models, i.e., $n > m$.

We modify Ang and Piazzesi (2003)'s model by replacing the discrete-time AR(p) process with a continuous-time AR(p) (CAR) process.[12] The corresponding volatility specification

---

[11] By setting the (demeaned) state variables to QJR, the zero-coupon bond yield of maturity T is:

$$y(t, T) = \text{constant} + X_t^1 + \frac{e^{k_2(T-t)} - 1}{k_2(T - t)} X_t^2 + \left[\frac{e^{k_2(T-t)} - 1}{k_2(T - t)} - e^{k_2(T-t)} \right] X_t^3,$$

where X_t^1 to X_t^3 are the state variables in QJR. This representation coincides with the dynamic version of the original Nelson-Siegel model. See Christensen et al. (2011, Proposition 1).

[12] We emphasize that the CAR process is not an equivalent analogy to the AR process in the continuous-time sense. As mentioned in He and Wang (1989), it is extremely difficult, if not impossible, to find the condition for an AR(p) to be embedded in a CAR(p) when $p \geq 3$. However, they do share the important feature that current states are functions of their own lagged values.

for the model with CAR(2) state variables is:

$$\sigma(x) = \begin{bmatrix} e^{-k_1 x} & e^{-k_2 x} \end{bmatrix} \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix}.$$

We obtain an FDR for the model using the CR, which has the following triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -k_1 k_2 & -(k_1 + k_2) \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ \frac{\Omega_2 \Omega_1 (k_1 - k_2)^2}{\Omega_1 + \Omega_2} \end{bmatrix},$$

$$\mathbf{C}(x) = \frac{\Omega_1 + \Omega_2}{\Omega_1 \Omega_2 (k_1 - k_2)^2} \begin{bmatrix} (\Omega_1 k_2 e^{-xk_1} + \Omega_2 k_1 e^{-xk_2}) & (\Omega_1 e^{-xk_1} + \Omega_2 e^{-xk_2}) \end{bmatrix}.$$

The transform matrix M is:

$$M = \frac{k_1 - k_2}{\Omega_1 + \Omega_2} \begin{bmatrix} -\Omega_2 & \Omega_1 \\ \Omega_2 k_1 & -\Omega_1 k_2 \end{bmatrix}.$$

3.3.2 Bhar and Chiarella (1997)

Bhar and Chiarella (1997) present a one-factor Gaussian model directly under the HJM setup, with the following volatility specification:

$$\sigma(x) = e^{-kx} \begin{bmatrix} 1 & \cdots & x^{n-1} \end{bmatrix} \begin{bmatrix} \Omega_1 \\ \vdots \\ \Omega_n \end{bmatrix}.$$

For $n = 3$, the triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is given by:

$$\mathbf{A} = \begin{bmatrix} -k & 0 & 0 \\ 1 & -k & 0 \\ 0 & 2 & -k \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{C}(x) = \begin{bmatrix} \Omega_1 + \Omega_2 x + \Omega_3 x^2 & \Omega_2 + 2\Omega_3 x & \Omega_3 \end{bmatrix} e^{-kx}.$$

The transform matrix M is:

$$M = \begin{bmatrix} \Omega_1 & \Omega_2 & \Omega_3 \\ \Omega_2 & 2\Omega_3 & 0 \\ \Omega_3 & 0 & 0 \end{bmatrix}^{-1}.$$

We refer to this realization as the Lower Ascending Jordan Form Realization (LAJR). The LAJR is obtained using Bhar and Chiarella (1997)'s "direct integration" (their Proposition 1), which only works with one factor. We extend their result to handle the multi-factor case. The result is presented in Appendix C.

3.3.3 Cox, Ingersoll, and Ross (1981)

In Section VIII of their paper (Eqn. (71)), Cox et al. (1981) develop a one-factor, three-state term structure model, in which the state variables for the term structure consist of the short rate and two smoothed averages of the short rate. The Gaussian version of this model is a special case of our general framework. The corresponding volatility specification for this model is:

$$\sigma(x) = \begin{bmatrix} e^{-k_1 x} & e^{-k_2 x} & e^{-k_3 x} \end{bmatrix} \begin{bmatrix} \Omega_1 \\ \Omega_2 \\ \Omega_3 \end{bmatrix}.$$

The triplet $\{\mathbf{A}, \mathbf{B}, \mathbf{C}(x)\}$ is given by:

$$\mathbf{A} = \begin{bmatrix} -\frac{k_1 \Omega_1 + k_2 \Omega_2 + k_3 \Omega_3}{\Omega_1 + \Omega_2 + \Omega_3} & a_1(\Psi) & a_2(\Psi) \\ b_1(\Psi) & -b_1(\Psi) & 0 \\ b_2(\Psi) & 0 & -b_2(\Psi) \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \Omega_1 + \Omega_2 + \Omega_3 \\ 0 \\ 0 \end{bmatrix},$$

$$\mathbf{C}(x) = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} e^{-\mathbf{A}x}.$$

The transform matrix M is:

$$
M = \begin{bmatrix}
\frac{\Omega_1}{\Omega_1+\Omega_2+\Omega_3} & \kappa\left(\Psi\right) & \varpi - \kappa\left(\Psi\right) \\[2mm]
\frac{\Omega_2}{\Omega_1+\Omega_2+\Omega_3} & \vartheta\left(\Psi\right) & \varkappa - \vartheta\left(\Psi\right) \\[2mm]
\frac{\Omega_3}{\Omega_1+\Omega_2+\Omega_3} & -\kappa\left(\Psi\right) - \vartheta\left(\Psi\right) & \kappa\left(\Psi\right) + \vartheta\left(\Psi\right) - \varpi - \varkappa
\end{bmatrix}^{-1},
$$

where $\Psi = \{\Omega_1, \Omega_2, \Omega_3, k_1, k_2, k_3\}$,

$$
\varpi = -\frac{\Omega_1\left(k_1 k_2 \Omega_3 + k_1 k_3 \Omega_2 - k_2 k_3 \Omega_2 - k_2 k_3 \Omega_3\right)}{\left(k_1 k_2 \Omega_3 + k_1 k_3 \Omega_2 + k_2 k_3 \Omega_1\right)\left(\Omega_1 + \Omega_2 + \Omega_3\right)},
$$

$$
\varkappa = \frac{\Omega_2\left(k_1 k_3 \Omega_1 - k_1 k_3 \Omega_3 - k_2 k_3 \Omega_1 + k_1 k_3 \Omega_3\right)}{\left(k_1 k_2 \Omega_3 + k_1 k_3 \Omega_2 + k_2 k_3 \Omega_1\right)\left(\Omega_1 + \Omega_2 + \Omega_3\right)},
$$

and $a_i\left(\Psi\right)$, $b_i\left(\Psi\right)$, $i = 1, 2$, $\kappa\left(\Psi\right)$, and $\vartheta\left(\Psi\right)$ are too lengthy to be presented in the paper.[13] If $a_1\left(\Psi\right) > 0 > a_2\left(\Psi\right)$, then the short rate would promote regression to the first moving average and extrapolation to the second. For this reason, we refer to this realization as the Regression and Extrapolation Realization (RER). RER is very useful in extracting information about time-varying targets of the short rate.

4   Conclusion

In this paper, we review and enrich the general HJM-based framework for modeling GDTSMs developed in LYY. Based on a general volatility specification, we obtain infinitely many Markov realizations of the same HJM model, which nests most existing GDTSMs as special cases. We also discuss the implication of our approach for interest rate derivatives pricing, and demonstrate the advantage of using Linear System Theory to construct FDRs over the integration approach. In LYY, we present further empirical evidence that some of the non-Markov GDTSMs constructed using this general approach can outperform those from the extant literature. Taken together, our approach seems to hold promise for term structure modeling.

---

[13] These functions (in terms of MATLAB code) are available upon request from the authors.

Appendices

A   Proof of Proposition 3

For a Gaussian model, the Q-dynamics of the zero-coupon bond maturing at T (T is constant) is given by:

$$dP\left(t,T\right) = r_t P\left(t,T\right) dt - P\left(t,T\right) \sigma^P\left(t,T\right) dW_t,$$

where $\sigma^P\left(t,T\right) = \int_t^T \sigma\left(t,s\right) ds$. By definition, the LIBOR forward rate for $[T - \delta, T]$ is given by:

$$L\left(t;T-\delta,\delta\right) = -\frac{P\left(t,T\right) - P\left(t,T-\delta\right)}{\delta P\left(t,T\right)}.$$

Then, we define a new variable $L^\delta\left(t\right)$ as $L\left(t;T-\delta,\delta\right) + \frac{1}{\delta}$, i.e.,

$$L^\delta\left(t\right) = \frac{P\left(t,T-\delta\right)}{\delta P\left(t,T\right)}.$$

Next, we will show that $L^\delta\left(t\right)$ is a martingale under the forward measure Q^T for $t <$

$T - \delta$. By Itô's lemma,

$$
\begin{aligned}
dL^\delta(t) &= \frac{\partial L^\delta(t)}{\partial P(t,T)} dP(t,T) + \frac{\partial L^\delta(t)}{\partial P(t,T-\delta)} dP(t,T-\delta) \\
&\quad + \frac{1}{2} \frac{\partial^2 L^\delta(t)}{\partial P(t,T)^2} [dP(t,T)]^2 + \frac{\partial^2 L^\delta(t)}{\partial P(t,T-\delta)\partial P(t,T)} [dP(t,T) dP(t,T-\delta)] \\
&= \frac{1}{\delta P(t,T)} \left(r_t P(t,T-\delta) dt - P(t,T-\delta) \sigma^P(t,T-\delta) dW_t \right) \\
&\quad - \frac{P(t,T-\delta)}{\delta P(t,T)^2} \left(r_t P(t,T) dt - P(t,T) \sigma^P(t,T) dW_t \right) \\
&\quad + \frac{P(t,T-\delta)}{\delta P(t,T)^3} \left(P(t,T)^2 \left\| \sigma^P(t,T) \right\|^2 dt \right) \\
&\quad - \frac{1}{\delta P(t,T)^2} \left(P(t,T-\delta) \sigma^P(t,T-\delta) \sigma^P(t,T)^\mathsf{T} P(t,T) dt \right) \\
&= L^\delta(t) \left(\sigma^P(t,T) dW_t - \sigma^P(t,T-\delta) dW_t \right. \\
&\quad \left. + \left\| \sigma^P(t,T) \right\|^2 dt - \sigma^P(t,T-\delta) \sigma^P(t,T)^\mathsf{T} dt \right).
\end{aligned}
$$

We know that the Girsanov kernel for the transition from Q to Q^T is given by the T-maturity bond volatility $\sigma^P(t,T)$ (see Björk, 2009, Proposition 26.7), i.e., $W_t^T = W_t + \sigma^P(t,T)^\mathsf{T} t$ is a Wiener process under the forward measure Q^T. Therefore:

$$
dL^\delta(t) = L^\delta(t) \left[\sigma^P(t,T) - \sigma^P(t,T-\delta) \right] dW_t^T \tag{A1}
$$

$$
\equiv L^\delta(t) \sigma^{L^\delta}(t,T) dW_t^T. \tag{A2}
$$

This shows that $L^\delta(t)$ is a Q^T-martingale.

The Caplet price is given by:

$$
\begin{aligned}
Caplet(t;T,K) &= \delta \mathbb{E}^Q \left[e^{-\int_t^T r_s ds} (L(T-\delta;T-\delta,\delta) - K)^+ \middle| \mathcal{F}_t \right] \\
&= \delta P(t,T) \mathbb{E}^{Q^T} \left[(L(T-\delta;T-\delta,\delta) - K)^+ \middle| \mathcal{F}_t \right] \\
&= \delta P(t,T) \mathbb{E}^{Q^T} \left[\left(L^\delta(T-\delta) - K - \frac{1}{\delta} \right)^+ \middle| \mathcal{F}_t \right].
\end{aligned}
$$

By (A1), we have:

$$
L^\delta(T-\delta) = L^\delta(t) \exp \left\{ -\frac{1}{2} \int_t^{T-\delta} \left\| \sigma^{L^\delta}(t,s+\delta) \right\|^2 ds + \int_t^{T-\delta} \sigma^{L^\delta}(t,s+\delta) dW_s^T \right\}.
$$

Therefore, computing the expectation with respect to the forward measure leads to the result of Proposition 3.

B CR Algorithm

The details presented here can also be found in Brockett (1970, page 107). Specifically, $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}(x)$ can be solved from the equation of the transfer functions (see, Lemma 3.2 and 3.3 in Björk and Gombani, 1999) using the CR algorithm:

$$\mathcal{L}\left[\sigma\left(t\right)\right]\left(s\right) = \mathbf{C}\left(s\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{B} \text{ and } \mathbf{C}\left(x\right) = \mathbf{C}\exp\left(\mathbf{A}x\right),$$

where $\mathcal{L}\left[\cdot\right]$ is the Laplace transform of a function and is defined as:

$$\mathcal{L}\left[g\left(t\right)\right]\left(s\right) = \int_0^\infty e^{-st}g\left(t\right)dt.$$

Therefore, given the volatility function (15),

$$\mathcal{L}\left[\sigma\left(t\right)\right]\left(s\right) = \underbrace{\left[\frac{0!}{\left(k_i + s\right)}, \frac{1!}{\left(k_i + s\right)^2}, \cdots, \frac{\left(n_i - 1\right)!}{\left(k_i + s\right)^{n_i}}\right]_{i=1}^{I}}_{1 \times \sum_{i=1}^{I} n_i} \times \Omega. \tag{A3}$$

First, expand the first part of (A3) into an infinite power series to obtain:

$$\left[\frac{0!}{\left(k_i + s\right)}, \frac{1!}{\left(k_i + s\right)^2}, \cdots, \frac{\left(n_i - 1\right)!}{\left(k_i + s\right)^{n_i}}\right]_{i=1}^{I} = \sum_{\alpha=0}^{\infty} \left[\mathbf{L}_{\alpha 1}, \mathbf{L}_{\alpha 2}, \cdots, \mathbf{L}_{\alpha n}\right] s^{-(\alpha+1)},$$

where $\mathbf{L}_{\alpha j}$, $j = 1, 2, \ldots, n$, $\alpha = 0, 1, \ldots$, are called Markov parameters. They can be easily computed for any specific example by applying the Taylor expansion, e.g., expanding $\frac{1}{k_i + s}$ into an infinite power series by regarding $\frac{1}{k_i + s}$ as a function of k_i and expanding it about $k_i = 0$,

$$\frac{1}{k_i + s} = s^{-1} - k_i s^{-2} + k_i^2 s^{-3} - k_i^3 s^{-4} + \ldots.$$

Secondly, denote by $p\left(s\right)$ the monic least common multiple of all $\left(k_i + s\right)^{n_i+1}$ for $i =$

$1, 2, \ldots, m$, it follows that:

$$p(s) = \prod_{i=1}^{m} (k_i + s)^{n_i + 1} = s^n + p_{n-1} s^{n-1} + \cdots + p_1 s + p_0,$$

where $n = m + \sum_{i=1}^{m} n_i$. Like $\mathbf{L}_{\alpha j}$ above, p_k, $k = 0, 1, \ldots, n-1$, can also be easily computed for any specific example by expanding $\prod_{i=1}^{m} (k_i + s)^{n_i + 1}$.

Then,

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -p_0 & -p_1 & -p_2 & \cdots & -p_{n-1} \end{bmatrix}_{n \times n}, \quad \bar{\mathbf{B}} = \begin{bmatrix} \mathbf{L}_{01} & \mathbf{L}_{02} & \cdots & \mathbf{L}_{0m} \\ \mathbf{L}_{11} & \mathbf{L}_{12} & \cdots & \mathbf{L}_{1m} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{L}_{n-2,1} & \mathbf{L}_{n-2,2} & \cdots & \mathbf{L}_{n-2,m} \\ \mathbf{L}_{n-1,1} & \mathbf{L}_{n-1,2} & \cdots & \mathbf{L}_{n-1,m} \end{bmatrix}_{n \times n},$$

$\mathbf{B} = \bar{\mathbf{B}}_{n \times n} \times \mathbf{\Omega}_{n \times m}$, and $\mathbf{C}(x) = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}_{1 \times n} \times \exp(\mathbf{A}x)$.

C "Direct Integration" for the Multi-factor Case

The result is summarized in the following proposition:

Proposition 5 *If the forward rate volatility function of the ith Brownian Motion, $i = 1, 2, \ldots, m$,*
assumes the form:

$$\sigma_i(t, T) = \left(a_{i,0} + a_{i,1}(T-t) + a_{i,2}(T-t)^2 + \cdots + a_{i,np_i - 1}(T-t)^{np_i - 1} \right) e^{-k_i(T-t)},$$

then the forward rate $f(t, T)$ can be expressed as an affine function of Markovian state variables, i.e.,

$$f(t, T) = f(0, T) + \Theta(t, T-t) + \mathbf{C}^{LAJR}(T-t) Z_t^{LAJR}, \tag{A4}$$

$$dZ_t^{LAJR} = \mathbf{A}^{LAJR} Z_t^{LAJR} + \mathbf{B}^{LAJR} dW_t, \quad Z_0^{LAJR} = \mathbf{0}, \tag{A5}$$

where $\Theta(t, T - t)$ is defined in (4), and

$$\left[Z_t^{LAJR}\right]^\mathsf{T} = \underbrace{\left[\left[\int_0^t (t-s)^j e^{-k_i(t-s)} dw_i(s)\right]_{j=0}^{np_i-1}\right]_{i=1}^m}_{1 \times n},$$

$$\mathbf{C}^{LAJR}(x) = \underbrace{\left[\left[\sum_{l=j}^{np_i-1} a_{il} \binom{l}{j} x^{l-j} e^{-k_i x}\right]_{j=0}^{np_i-1}\right]_{i=1}^m}_{1 \times n},$$

$$\mathbf{A}^{LAJR} = \underbrace{\begin{bmatrix} \breve{\mathbf{A}}_1 & 0 & \cdots & 0 \\ 0 & \breve{\mathbf{A}}_2 & \cdots & 0 \\ & & \ddots & \\ 0 & \cdots & \cdots & \breve{\mathbf{A}}_m \end{bmatrix}}_{n \times n}, \breve{\mathbf{A}}_i = \underbrace{\begin{bmatrix} -k_i & 0 & \cdots & 0 \\ 1 & -k_i & \cdots & 0 \\ & & \ddots & \\ 0 & \cdots & n_i & -k_i \end{bmatrix}}_{(n_i+1) \times (n_i+1)},$$

$$\mathbf{B}^{LAJR} = \underbrace{\begin{bmatrix} \breve{\mathbf{B}}_1 \\ \breve{\mathbf{B}}_2 \\ \vdots \\ \breve{\mathbf{B}}_m \end{bmatrix}}_{n \times m},$$

$$\breve{\mathbf{B}}_1 = \underbrace{\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ & & \cdots & \\ 0 & \cdots & \cdots & 0 \end{bmatrix}}_{(n_1+1) \times m}, \breve{\mathbf{B}}_2 = \underbrace{\begin{bmatrix} 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ & & \cdots & \\ 0 & \cdots & \cdots & 0 \end{bmatrix}}_{(n_2+1) \times m}, \cdots, \breve{\mathbf{B}}_m = \underbrace{\begin{bmatrix} 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ & & \cdots & \\ 0 & \cdots & \cdots & 0 \end{bmatrix}}_{(n_m+1) \times m},$$

and $W_t = \left[[w_i]_{i=1}^m\right]^\mathsf{T}$.

Proof. In (A4), only the last term on the RHS needs to be verified. For each i,

$$\int_0^t \sigma_i(s, T)\, dw_i(s)$$

$$= \sum_{j=0}^{n_i} \int_0^t a_{i,j}(T-s)^j e^{-k_i(T-s)} dw_i(s)$$

$$= e^{-k_i(T-t)} \sum_{j=0}^{n_i} \int_0^t \sum_{l=0}^{j} \binom{j}{l}(T-t)^{j-l} a_{i,j}(t-s)^l e^{-k_i(t-s)} dw_i(s)$$

$$= e^{-k_i(T-t)} \sum_{j=0}^{n_i} \int_0^t \sum_{l=j}^{n_i} a_{i,l}\binom{l}{j}(T-t)^{l-j}(t-s)^j e^{-k_i(t-s)} dw_i(s)$$

$$= \sum_{j=0}^{n_i} \left[e^{-k_i(T-t)} \sum_{l=j}^{n_i} a_{i,l}\binom{l}{j}(T-t)^{l-j} \right] Z_t^{\text{LAJR}}(\mathbf{q}(i,j)).$$

where $Z_t^{\text{LAJR}}(\bullet)$ denotes the $\bullet$th element of Z_t^{LAJR}, and $\mathbf{q}(i,j)$, for $i = 1, 2, \ldots, m$ and $j = 0, 1, \ldots, n_i$, is defined as:

$$\mathbf{q}(i,j) = \begin{cases} j+1, & i = 1, \\ i + j + \sum_{l=1}^{i-1} n_l, & i > 1. \end{cases}$$

Next, we show that Z_t^{LAJR} follows SDE (A5). First, for $j = 0$, $Z_t^{\text{LAJR}}(\mathbf{q}(i,0)) = \int_0^t e^{-k_i(t-s)} dw_i(s)$ follows the SDE:

$$dZ_t^{\text{LAJR}}(\mathbf{q}(i,0)) = -k_i Z_t^{\text{LAJR}}(\mathbf{q}(i,0))\, dt + dw_i(t).$$

Then, for $j \geqslant 1$, $Z_t^{\text{LAJR}}(\mathbf{q}(i,j)) = \int_0^t (t-s)^j e^{-k_i(t-s)} dw_i(s)$ follows the SDE:

$$dZ_t^{\text{LAJR}}(\mathbf{q}(i,j)) = j\int_0^t (t-s)^{j-1} e^{-k_i(t-s)} dw_i(s) - k_i \int_0^t (t-s)^j e^{-k_i(t-s)} dw_i(s)$$

$$= \left[j Z_t^{\text{LAJR}}(\mathbf{q}(i,j-1)) - k_i Z_t^{\text{LAJR}}(\mathbf{q}(i,j)) \right] dt.$$

This completes the proof. ∎

References

Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary economics 50*(4), 745–787.

Bhar, R. and C. Chiarella (1997). Transformation of Heath-Jarrow-Morton Models to Markovian Systems. *The European Journal of Finance 3*(1), 1–26.

Björk, T. (2009). *Arbitrage theory in continuous time*. Oxford Univ Pr.

Björk, T. and A. Gombani (1999). Minimal realizations of interest rate models. *Finance and Stochastics 3*(4), 413–432.

Björk, T. and C. Landén (2002). On the construction of finite dimensional realizations for nonlinear forward rate models. *Finance and Stochastics 6*(3), 303–331.

Björk, T. and L. Svensson (2001). On the Existence of Finite-Dimensional Realizations for Nonlinear Forward Rate Models. *Mathematical Finance 11*(2), 205–243.

Bliss, R., P. Ritchken, and E. Ronn (1996). Empirical Tests of Two State-Variable Heath-Jarrow-Morton Models. *Journal of Money, Credit & Banking 28*(3), 452–476.

Brace, A. (2007). *Engineering BGM*. Chapman & Hall.

Brace, A. and M. Musiela (1994). A MULTIFACTOR GAUSS MARKOV IMPLEMENTATION OF HEATH, JARROW, AND MORTON. *Mathematical Finance 4*(3), 259–283.

Brockett, R. (1970). *Finite dimensional linear systems*. Wiley.

Carverhill, A. (1994). When is the Short Rate Markovian? *Mathematical Finance 4*(4), 305–312.

Chen, C.-T. (1999). *Linear System Theory and Design. 3rd*. New York, NY: Oxford University Press.

Chiarella, C. and O. Kwon (2003). Finite Dimensional Affine Realisations of HJM Models in Terms of Forward Rates and Yields. *Review of Derivatives Research 6*(2), 129–155.

Christensen, J., F. Diebold, and G. Rudebusch (2011). The affine arbitrage-free class of nelson-siegel term structure models. *Journal of Econometrics*.

Collin-Dufresne, P., R. Goldstein, and C. Jones (2008). Identification of maximal affine term structure models. *The Journal of Finance 63*(2), 743–795.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1981). A re-examination of traditional hypotheses about the term structure of interest rates. *The Journal of Finance 36*(4), 769–799.

Dai, Q. and K. Singleton (2000). Specification analysis of affine term structure models. *Journal of Finance 55*(5), 1943–1978.

Dai, Q. and K. Singleton (2002). Expectation puzzles, time-varying risk premia, and affine models of the term structure. *Journal of Financial Economics 63*(3), 415–441.

De Jong, F. and P. Santa-Clara (1999). The Dynamics of the Forward Interest Rate Curve: A Formulation with State Variables. *Journal of Financial and Quantitative Analysis 34*(1), 131–157.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance 6*(4), 379–406.

Fisher, M. (1998). A simple model of the failure of the expectations hypothesis. *Unpublished working paper, Federal Reserve Bank of Atlanta*.

He, S. and J. Wang (1989). On embedding a discrete-parameter arma model in a continuous-parameter arma model. *Journal of Time Series Analysis 10*(4), 315–323.

Heath, D., R. Jarrow, and A. Morton (1992). Bond Pricing and the Term Structure of Interest Rates: A New Methodology for Contingent Claims Valuation. *Econometrica 60*(1), 77–105.

Inui, K. and M. Kijima (1998). A Markovian framework in multi-factor Heath-Jarrow-Morton models. *Journal of Financial and Quantitative Analysis*, 423–440.

Joslin, S., A. Le, and K. J. Singleton (2013). Why gaussian macro-finance term structure models are (nearly) unconstrained factor-vars. *Journal of Financial Economics 109*(3), 604–622.

Joslin, S., K. Singleton, and H. Zhu (2011). A new perspective on gaussian dynamic term structure models. *Review of Financial Studies*.

Li, H., X. Ye, and F. Yu (2016). A unified hjm approach to non-markov gaussian dynamic term structure models: International evidence. Working paper, Available at SSRN 2155538.

Nelson, C. and A. Siegel (1987). Parsimonious Modeling of Yield Curves. *Journal of Business 60*(4), 473.

Ritchken, P. and L. Sankarasubramanian (1995). Volatility Structures of Forward Rates and the Dynamics of the Term Structure. *Mathematical Finance 5*(1), 55–72.

Schutter, B. (2000). Minimal state-space realization in linear system theory: an overview. *Journal of Computational and Applied Mathematics 121*(1-2), 331–354.

Trolle, A. and E. Schwartz (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *Review of Financial Studies 22*(5), 2007.

Table 1: Summary of Specifications of GDTSMs in the Literature

Paper	Realization (Model Notation)	Specification	
Dai and Singleton (2000)	LTR ($AM_0(3)$)	$r(t) = \delta_0 + \delta_y' Y(t)$	$dY(t) = \begin{bmatrix} \kappa_{11} & 0 & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ \kappa_{31} & \kappa_{32} & \kappa_{33} \end{bmatrix} Y(t)\,dt + dW(t)$
Collin-Dufresne et al. (2008)	CR	$r(t) = r(t)$	$\begin{bmatrix} dr \\ d\mu_1 \\ d\mu_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \\ \gamma + \kappa_0 r + \kappa_1 \mu_1 + \kappa_2 \mu_2 \end{bmatrix} dt, \begin{bmatrix} V_{00} & V_{01} & V_{02} \\ V_{01} & V_{11} & V_{12} \\ V_{02} & V_{12} & V_{22} \end{bmatrix} \right)$
Fisher (1998)	LTR	$r(t)$ is directly modeled as one of the state variables	$dr(t) = \hat{\kappa}_r (z(t) - r(t))\,dt + \sigma_r d\hat{W}_1(t)$ $dz(t) = \hat{\kappa}_z \left(\hat{\theta}_z - z(t) \right) dt + \sigma_z d\hat{W}_2(t)$
Joslin et al. (2011)	JR (JPC)	$r_t = \iota \cdot X_t$, where ι is a vector of ones	$\Delta X_t = K_{0X}^Q + K_{1X}^Q X_{t-1} + \Sigma_X \epsilon_t^Q$, where K_{1X}^Q is of the Jordan form
Christensen et al. (2011)	QJR	$r_t = X_t^1 + X_t^2$	$\begin{pmatrix} dX_t^1 \\ dX_t^2 \\ dX_t^3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & \lambda \end{pmatrix} \left[\begin{pmatrix} \theta_1^Q \\ \theta_2^Q \\ \theta_3^Q \end{pmatrix} - \begin{pmatrix} X_1 \\ X_2 \\ X_3 \end{pmatrix} \right] dt + \Sigma \begin{pmatrix} dW_t^{1,Q} \\ dW_t^{2,Q} \\ dW_t^{3,Q} \end{pmatrix}$
Ang and Piazzesi (2003)	CR	$r_t = b_0 + b_1' X_t^o + v_t$, where $X_1^o = \left(f_t^{o'}, \cdots, f_{t-p-1}^{o'} \right)'$ and v_t is a monetary policy shock	$f_t^o = \rho_1 f_{t-1}^o + \cdots + \rho_p f_{t-p-1}^o + \Omega u_t^o$
Bhar and Chiarella (1997)	LAJR	$f(t,T) = f(0,T) + \int_0^t \sigma_f(u,T) \int_u^t \sigma_f(u,s)\,ds\,du + \int_0^t \sigma_f(u,T)\,d\tilde{W}(u)$	$\sigma_f(t,T) = \left(a_0 + a_1(T-t) + \cdots + a_n(T-t)^n \right) e^{-\lambda(T-t)}$
Cox, Ingersoll, and Ross (1981)	RER	$r_t = [1,0,0]\,[r_t, x_1, x_2]^\top$	$dr_t = [b_0(\mu - r_t) + b_1(x_1 - r_t) + b_2(x_2 - r_t)]\,dt + \sigma dz$ $dx_i = \beta_i(r_t - x_i)dt, i = 1,2$

This table provides a partial list of the GDTSMs in the literature that can be nested by our Gaussian HJM model with the general volatility function. For convenience, we adopt the notations in the original papers.