Markov Chain Approximations For Term Structure Models*

David Backus,† Liuren Wu,‡ and Stanley Zin§

First draft: March 5, 1999
This version: August 18, 1999

Abstract

We derive discrete markov chain approximations for continuous state equilibrium term structure models. The states and transition probabilities of the markov chain are chosen efficiently according to a quadrature rule as in Tauchen and Hussey (1991). Quadrature provides a simple yet flexible method which can easily incorporates complication like non-normality, heteroskedasticity, and multiple factors. We use the extended Vasicek model of the term structure as an example for this procedure and compare its *pricing efficiency and accuracy* to the popular trinomial tree approximation of Hull and White (1990). We further illustrate, with numerical examples, the *flexibility* and efficiency of this procedure in pricing interest rate options when the underlying interest rate has conditional non-normality and multiple factors.

JEL Classification Codes: E43, C63, C61, G12, G13

Keywords: markov chain, term structure, interest rates, mean-reversion, quadrature, option pricing.

---

*We welcome comments, especially references to related papers we inadvertently overlooked. We thank seminar participants at New York University and the Federal Reserve Bank of New York for helpful comments. The latest version of the paper can be downloaded from http://www.bnet.fordham.edu/public/finance/lwu/index.html.

†Stern School of Business, New York University and NBER, dbackus@stern.nyu.edu.

‡Graduate School of Business, Fordham University; liwu@mary.fordham.edu.

§Stern School of Business, New York University, Graduate School of Industrial Administration, Carnegie Mellon University, and NBER, szin@bobbyorr.gsia.cmu.edu.

1 Introduction

Binomial trees have become a workhorse of applied finance. Prominent examples in the term-structure literature include Ho and Lee (1986), Black, Derman and Toy (1990), and Black and Karasinski (1991). In addition to providing a simple computational algorithm for approximating arbitrage-free prices implied by an equilibrium asset-pricing model, they are used extensively by practitioners to price complex, path-dependent derivative securities for which analytical forms are unavailable. Most binomial trees used in the finance literature are *recombining* due to its appealing properties. That is, *event 1* followed in time by *event 2* leads to the same point in the statespace as *event 2* followed by *event 1.* Therefore, the statespace grows linearly in time (rather than the exponential growth of the statespace implied by non-recombining trees), greatly reducing the computational burden of solving the model.

One significant drawback of this approach becomes evident when the state variable exhibits mean reversion. The recombining property of the state space is not as natural in this case, and *ad hoc* methods of capturing mean reversion while restricting states to recombine must be developed. For example, the trinomial tree approach of Hull and White (1990) has become a standard approximation for this situation.

In this paper, we develop a time-homogeneous markov chain approach for problems of discrete approximation of term-structure models. This approach provides a somewhat more systematic treatment of approximation and is easily generalized to more complex models. In addition, there is a well-developed literature on convergence of these approximations to the true process. The drawback of this approach is the large number of parameters that must be calibrated. An N-state markov chain has $N + N(N-1)$ free parameters: N states and $N(N-1)$ transition probabilities. Since in many applications, accuracy demands a large value for N, the number of free parameters to fit seems daunting. To address this problem, we rely heavily on the quadrature-based methods of Tauchen and Hussey (1991). As we will see, this provides a very general method of efficiently choosing states and transition probabilities given the underlying continuous-state process.

For term structure models of interest rates, where interest rates and bond yields exhibit significant mean-reversion, a markov chain representation may be a more natural way to go. Traditionally, markov chains have been used to model stationary processes. Interest rates and bond yields fit perfectly into this category. In contrast, trees are more appropriate for random walk processes. Hull and White (1990) use trinomial trees to model the term structure of interest rates. Due to the stationarity of interest rates, the tree cannot grow infinitely over time with positive probability. To avoid negative probability, the tree has to be truncated, rather arbitrarily, up to a certain stage. The calibration of such a truncated tree can be very messy, especially at the edges: Depending on whether a state is at or near the edge or not and when the tree is truncated, that state can be linked to 1, 2, 3, 4, or 5 branches from the previous period. When computing state prices, therefore, extreme care must be taken on where the state lies on the tree. Yet with the markov chain representation, multiperiod state prices are merely simple powers of the transition matrix. The time-homogeneity of the representation releases us from the burden of step-by-step calibration.

Using the interest rate cap as an example, we find that, with the same number of final states, the markov chain representation yields the cap price much *faster* than the trinomial tree approach. The speed gain comes from several sources. First, particular in languages such as matlab and GAUSS which handles matrix efficiently, power computation are much faster than loops. Since the state price at each period can be represented as simple powers of the adjusted transition matrix, we can compute the state price efficiently. In the trinomial tree approach, however, several conditional arguments needed in the program to take care of the edges. Secondly, since the number of states in recombining tree increases linearly with number of periods, to obtain a finer grid for the final period, we need use finer time intervals even if all we care is the final period. The number of states is therefore linked closely to the number of periods of the tree. In case of the trinomial tree, due to the truncation at the edge, we generally need even finer intervals to obtain the desired number of states. This implies a waste of computing time for these middle steps. With the markov chain representation, however, the specification of the number of states and the number of periods are independent. We can freely choose the number of states to satisfy the precision requirement and choose the number of periods based on the specification of the option. For example, to price a single European option with maturity τ, we only need to specify a one-step markov chain which is calibrated to the τ-period conditional distribution. The pricing accuracy does not depend on the number of steps in this case, but just depend on the approximation of the conditional distribution, which is determined by the separately-chosen number of states. To price a cap with tenor τ, on the other hand, we only need to calibrate the markov chain to the conditional density with intervals no finer than the tenor τ. Pricing with trees generally need much finer intervals to obtain the required number of states needed for accurate pricing.

The markov chain representation not only leads to faster pricing but also yields *more accurate* results. First, trees have to be equally spaced to be able to recombine. For mean-reverting processes, equal-spaced trees will eventually come up with negative probabilities at a certain stage and that is where arbitrary truncation is needed. It is obviously more efficient to put finer grids at higher probability areas and coarser ones at low probability areas. The quadrature-based method of markov chain calibration just does that. As a result, while tree calibration is fitting merely the mean and variance of the innovation, a N-state markov chain calibrated using Gauss quadrature fits up to $2N-1$ moments. Since a tree has less states at the beginning periods and more in later periods, it can be regarded as a markov chain with many zero constraints.

Yet another advantage of the markov chain representation is its extreme *flexibility* in modeling complex dynamics such as conditional non-normality, heteroskedasticity, and multiple factors, all of which are salient features of the interest rates. As mentioned above, a N-state markov chain calibrated using Gauss quadratures can fit up to $2N-1$ moments and can therefore readily capture most types of non-normality specifications, as long as the underlying distribution has finite moments and converges in the sense that the distribution can be relatively well captured by its first $2N-1$ moments. Trees, on the other hand, are very limited. Binomial trees only have two free parameters and can therefore only capture conditional mean and variance. Trinomial trees have one more degree of freedom and can in principle fit another moment, say, the fourth cumulant.[1] Yet, even this is stretched. Fatter tail (leptokurtosis) implies a wider grid and a smaller branching probability, both of which lead to early truncations of the tree. As shown in the paper, trinomial trees begin to exhibit unstable behavior as we try to incorporate very moderate kurtosis. It has been well known that skewness and excess kurtosis induce pricing biases for options compared with Black's log-normal assumption. The markov chain representation can capture these biases without any extra difficulty. In this paper, we derive an analytical pricing formula for caps when the underlying interest rate follows an *extended* Vasicek (1977) model, where the innovation exhibits non-normality, captured by a Gram-Charlier A-type distribution. We then illustrate how we can readily calibrate the markov chain on such a Gram-Charlier density and price caps on it.

Empirical evidence has also indicated that generally two or more factors are needed to model the term structure of interest rates. While multi-factor modeling on a tree can be extremely complicated, it can relatively easily be done on a markov

[1] Due to the symmetric feature of the tree, it cannot be used to fit a non-zero third moment.

chain. Tauchen and Hussey (1991) provide examples on how to calibrate a VAR process and GARCH process on a markov chain. We calibrate a two-factor affine model with conditional heteroskedasticity and price a series of interest rate caplets to illustrate the wide applicability and flexibility of this procedure in pricing derivatives with complex dynamics.

The paper is structured as follows. The next section formally introduce the markov chain representation including the quadrature-based method of calibration and the markov chain representation of the state prices in the term structure of interest rates. In Section 3, we review the Hull and White (1990) trinomial tree representation for comparison. In Section 4, we derive the analytical pricing formula for caplets when the underlying interest rate follows an extended Vasicek model and the innovation is captured by a Gram-Charlier A-type distribution. Section 5 compares, through numerical examples, the pricing accuracy and efficiency of the markov chain representation and the trinomial tree approach. Section 6 demonstrates the flexibility of the markov chain approach in pricing derivatives when the underlying interest rate has (i) conditional non-normality and (ii) multiple factors. Section 7 concludes.

2 Markov Chain Representation

Markov chains provide a natural way to approximate stationary processes. In a markov chain representation, the number and the values of states are fixed, as are the transition probabilities. As an example, consider the state variable z that follows the process

$$dz = -\kappa z dt + \sigma dW,$$

where dW is a standard Brownian motion. Its discrete-time equivalent can be written as

$$\begin{aligned} z_{t+\Delta t} &= e^{-\kappa \Delta t} z_t + \sigma \sqrt{\frac{1-e^{-2\kappa \Delta t}}{2\kappa}} \varepsilon_{t+\Delta t} \\ &\cong \varphi z_t + \sigma \sqrt{\Delta t} \varepsilon_{t+\Delta t}, \end{aligned} \tag{1}$$

where $\varphi = 1 - \kappa \Delta t$ and $\varepsilon \sim N(0,1)$. The second line represents a first order approximation. An N-state markov chain approximation to this process is given by a set of values for the state $\{z_1, z_2, \ldots, z_N\}$ and a matrix of transition probabilities π, with

$$\pi_{ij} = \mathbb{P}\left(z_{t+\Delta t} = z_j | z_t = z_i\right).$$

The role of quadrature is to calibrate the states and probabilities from the conditional distribution of $z_{t+\Delta t}$.

2.1 Transition matrix: The quadrature-based methods

Tauchen and Hussey (1991) proposed a quadrature-based methods for the calibration of a markov chain. It greatly eases the burden of the calibration of the transition matrix. The grid points and transition probabilities in the markov chain are determined by the conditional density function $f(y|x)$ together with a numerical quadrature rule. The method can handle large problems and complex dynamics.

A N-point quadrature rule for integration of function $g(u)$, $u \in R^M$, against a density $\omega(u)$ is a set of N abscissa $u_k \in R^M$ and weights $w_k \in R^M$ such that

$$\int g(u)\,\omega(u)\,du \doteq \sum_{k=1}^{N} g(u_k)\,w_k \tag{2}$$

with convergence for each function g (under regularity conditions) of the approximating sum to the integral as $N \to \infty$. The abscissa u_k and the weights w_k depends only on the density ω and not directly on the function g. For most good rules, the weights are nonnegative and the rules integrate the constant function exactly; i.e. the weights sum to unity. Thus a quadrature rule can be viewed as a discrete probability model that approximates the density ω.

For a classical N-point Gaussian rule along the real line, $u \in R$, the abscissa u_k and the weights w_k are determined by forcing the rule to be exact for all polynomials of degrees less than or equal to $2N-1$. See Davis and Rabinowitz (1984) for details. Gaussian rules are close to minimum norm rules and possess several optimum properties. They are the best that can be done with N points using moments as a criterion, because if two probability distributions have the same moments up through $2N$, and if one of the distribution is a discrete distribution concentrated on N points, then the two distribution must coincide (Norton and Arnold, 1985).

Specifically, Tauchen and Hussey (1991) approximates the conditional density function $f(y|x)$ by a N-state markov chain $\left[\pi_{ij} = \pi^N(y_j|y_i)\right]$ with

$$\pi_{ij} = \frac{f(y_j|y_i)}{s(y_i)\,\omega(y_j)} w_j,\ \ i,j = 1,\cdots,N$$

where

$$s(y_i) = \sum_{j=1}^{N} \frac{f(y_j|y_i)}{\omega(y_j)} w_j$$

provides a normalization. The transition probability is obtained by replacing integration against $\omega(y_j)$ with summation using the quadrature rule, and then normalizing so that the weights add to unity. The Nystrom extension of the solution to the entire domain of x is

$$\pi^N(y_j|x) = \frac{f(y_j|x)}{s(x)\,\omega(y_j)} w_j, \quad j = 1, ..., N$$

with

$$s(x) = \sum_{j=1}^{N} \frac{f(y_j|x)}{\omega(y_j)} w_j.$$

$\omega(y)$ is some strictly positive weighting function. There is great latitude in the selection of the weighting function, though clearly one would want to select a weighting function such that the resulting sums are very close to exact integration against $f(y|x)$. One reasonable choice for the weighting function is $\omega(y) = f(y|0)$, i.e., the conditional density given that the process is at the unconditional mean. Another reasonable choice is $\omega(y) = f(y)$ where $f(y)$ is the unconditional or stationary density of the process. The density $f(y|0)$ places relatively more weight in the central part of the distribution and less weight in the tails than does $f(y)$. Tauchen and Hussey (1991) find that $\omega(y) = f(y|0)$ gives much better approximations except when the maximum magnitude of the characteristic roots of the autoregression is close to zero, in which case there is little difference between $f(y|0)$ and $f(y)$ and the approximations are essentially equivalent.

2.2 Gaussian Quadrature rules

Gaussian quadrature rules build on the orthogonal polynomial approach to functional approximation. For any fixed nonnegative weighting function $w(x)$, Gaussian creates approximations of the form

$$\int_a^b f(x)\, w(x)\, dx \doteq \sum_{i=1}^{N} w_i f(x_i) \tag{3}$$

for some nodes $x_i \in [a, b]$ and positive weights w_i and the approximation (3) is exact whenever f is a degree $2N-1$ polynomial. Specifically, given a nonnegative function

$w(x)$ and the $2N$-dimensional family $\mathcal{F}_{2N-1}$ of degree $2N-1$ polynomials, we can find N points $\{x_i\}_{i=1}^N \subset [a,b]$, and N nonnegative weights $\{w_i\}_{i=1}^N$ such that

$$\int_a^b f(x)\, w(x)\, dx = \sum_{i=1}^N w_i f(x_i) \tag{4}$$

for all $f \in \mathcal{F}_{2N-1}$. The following theorem from Judd (1998) summarizes the discussion in Davis and Rabinowitz (1984).

Theorem 1 *Suppose that $\{\phi_k(x)\}_{k=0}^\infty$ is an orthogonal family of polynomials with respect to $w(x)$ on $[a,b]$. Furthermore, define q_k so that $\phi_k(x) = q_k x^k + \cdots$. Let x_i, $i = 1, ..., N$ be the N zeros of $\phi_N(x)$. Then $a < x_1 < x_2 < \cdots < x_N < b$, and if $f \in C^{(2N)}[a,b]$, then*

$$\int_a^b f(x)\, w(x)\, dx = \sum_{i=1}^N w_i f(x_i) + \frac{f^{(2N)(\xi)}}{q_N^2 (2N)!}$$

for some $\xi \in [a,b]$, where[2]

$$w_i = -\frac{q_{N+1}/q_N}{\phi_N'(x_i)\, \phi_{N+1}(x_i)} > 0. \tag{5}$$

Furthermore, the formula $\sum_{i=1}^N w_i f(x_i)$ is the unique Gaussian integration formula on n nodes which exactly integrates $\int_a^b f(x)\, w(x)\, dx$ for all polynomials in $\mathcal{F}_{2N-1}$.

We can develop a Gaussian quadrature scheme over any interval $[a,b]$ using any weighting function. A key result in Theorem 1 is that Gaussian quadrature uses the zeros of the orthogonal polynomials, and that they lie in the interval $[a,b]$. Furthermore, the weights w_i are always positive, avoiding the precision problems of higher-order Newton-Cotes formulas. The formulas in Theorem 1 tell us how to compute the necessary nodes and weights. Typically one does not do the computation indicated in Theorem 1. Instead, there are some Gaussian quadrature formulas which are particularly useful, and the values of the nodes and weights are kept in tables.

[2] The weight w_i generally needs to be normalized by

$$w_i = \frac{\mu_0 w_i}{\sum w_i},$$

where μ_0 is the zero-th moment of the weight function. For example, for the Hermite weighting function $\omega(x) = e^{-x^2}$, we have $\mu_0 = \sqrt{\pi}$.

2.2.1 Gauss-Hermite Quadrature

Gauss-Hermite quadrature arises naturally because normal random variables are used often in economic problems. To evaluate $\int_{-\infty}^{\infty} f(x) e^{-x^2} dx$ using N points, the Gauss-Hermite quadrature rule uses the weights w_i and the nodes x_i, $i = 1, \cdots, N$ that are defined by

$$\int_{-\infty}^{\infty} f(x) e^{-x^2} dx = \sum_{i=1}^{N} w_i f(x_i) + \frac{N!\sqrt{\pi}}{2^N} \frac{f^{(2N)}(\xi)}{(2N)!}$$

for some $\xi \in (-\infty, \infty)$. The hermite polynomials are given by

$$\phi_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n} \left[e^{-x^2} \right]$$

and the orthogonality relation is

$$\int_{-\infty}^{\infty} e^{-x^2} \phi_m(x) \phi_n(x) dx = \begin{cases} 0, & m \neq n \\ \sqrt{\pi} 2^n n!, & m = n. \end{cases}$$

$\phi_n(x)$ can be generated by the following recursive relation,

$$\phi_{n+1}(x) = 2x\phi_n(x) - 2n\phi_{n-1}(x),$$

starting from $\phi_0(x) = 1$ and $\phi_1(x) = 2x$. As is clear from this recursive equation, $q_{n+1}/q_n = 2$ for all n in equation (5). Therefore, the nodes x_i can be solved from a polynomial equation and then the weights w_i is given by equation (5). w_i and x_i are often listed in tables. However, in the numerical exercise later in this paper, we did compute the Gauss-Hermite quadrature rules using matlab programs adapted from Elhay and Kautsky (1982)'s IQPACK FORTRAN program, available from Netlib.

Gauss-Hermite quadratures are used in connection with normal random variables. In particular, if Y is distributed $N(\mu, \sigma^2)$ then

$$E(f(Y)) = \left(2\pi\sigma^2\right)^{-1/2} \int_{-\infty}^{\infty} f(y) e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy$$

However, one must remember that to use Gauss-Hermite quadrature to compute such expectations, it is necessary to use the linear change of variables, $x = (y - \mu) / \sqrt{2}\sigma$, and use the identity

$$\int_{-\infty}^{\infty} f(y) e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy = \int_{-\infty}^{\infty} f\left(\sqrt{2}\sigma x + \mu\right) e^{-x^2} \sqrt{2}\sigma dx.$$

Hence, the general Gauss-Hermite quadrature for expectations of functions of Normal variables is

$$\begin{aligned} E(f(Y)) &= \left(2\pi\sigma^2\right)^{-1/2} \int_{-\infty}^{\infty} f(y)\, e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy \\ &= \pi^{-\frac{1}{2}} \sum_{i=1}^{N} w_i f\left(\sqrt{2}\sigma x_i + \mu\right), \end{aligned}$$

where the w_i and x_i are the Gauss-Hermite quadrature weights and nodes over $[-\infty, \infty]$.

In the applications below, we set $\omega(y) = f(y|0)$, unless otherwise noted, and use a Gauss-Hermite quadrature rule.

2.3 State price construction

We use the quadrature $\{x_i, w_i\}_{i=1}^{N}$ to provide an approximation of the stationary process z in (1). Recall the discrete-time approximation of the process,

$$z_{t+\Delta t} = \varphi z_t + \sigma\sqrt{\Delta t}\varepsilon_{t+\Delta t},$$

where $\varepsilon_{t+\Delta t}$ is a standard normal variable. Later we incorporate conditional non-normality by replacing the standard normal assumption with a Gram-Charlier type-A distribution.

Since the Gauss-Hermite quadrature nodes $\{x_i\}_{i=1}^{N}$ are based on the distribution $\omega(x) = e^{-x^2}$, the corresponding nodes for the stationary process z based on the conditional density $\omega(z_{t+\Delta t}|0) = \frac{1}{\sqrt{2\pi}\sigma\sqrt{\Delta t}} e^{-\frac{z^2}{2\sigma^2\Delta t}}$ is thus

$$z_i = \sqrt{2}\sigma\sqrt{\Delta t} x_i.$$

The transition matrix is given by

$$\pi_{ij} = \frac{f(z_j|z_i)}{s(z_i)\, f(z_j|0)} w_j, \quad j = 1, ..., N$$

where

$$\begin{aligned} f(z_j|z_i) &= \frac{1}{\sqrt{2\pi}\sigma\sqrt{\Delta t}} \exp\left(-\frac{(z_j - \varphi z_i)^2}{2\sigma^2\Delta t}\right); \\ f(z_j|0) &= \frac{1}{\sqrt{2\pi}\sigma\sqrt{\Delta t}} \exp\left(-\frac{z_j^2}{2\sigma^2\Delta t}\right); \\ s(z_i) &= \sum_{j=1}^{N} \frac{f(z_j|z_i)}{f(z_j|0)} w_j. \end{aligned}$$

At each state, we set the interest rate $r(t) = z(t) + \alpha(t)$, where $\alpha(t)$ is calibrated to match the current term structure of the yield curve. This yields the so-called *extended* Vasicek (1977) model due to the time-dependent specification of $\alpha(t)$. [3]

For notational convenience, we assume that the current state is the first state z_1, then from the one-period bond price,

$$P(1) = e^{-r(1)\Delta t} = e^{-(\alpha(1)+z_1)\Delta t} \sum_{j=1}^{N} \pi_{1j} = \sum_{j=1}^{N} Q_{1j},$$

we obtain the drift parameter,

$$\alpha(1) = -\frac{1}{\Delta t} \ln P(1) - z_1,$$

and one-period ahead state price density

$$Q_{1j} = \exp(-(\alpha(1) + z_1)\Delta t)\, \pi_{1,j} = P(1)\, \pi_{1j},$$

where π_{1j} denotes the transition probability from state z_1 (the current state) to state j. The one-period ahead state price Q_{1j} is today's price of an instrument which pays $1 next period if state j occurs and nothing if any other states occur. Any state-contingent claim with a maturity of one period can therefore be priced by the state price density.

For two-period ahead state prices, we consider a two-period bond,

$$P(2) = \sum_{k=1}^{N} Q_{1k} \exp(-(\alpha(2) + z_k)\Delta t) \sum_{j=1}^{N} \pi_{kj} = \sum_{j=1}^{N} Q_{2j},$$

from which we can solve for $\alpha(2)$,

$$\alpha(2) = -\frac{1}{\Delta t} \ln P(2) + \frac{1}{\Delta t} \ln \sum_{k=1}^{N} Q_{1,k} \exp(-z_k \Delta t).$$

[3] To avoid negative interest rate, Hull and White (1994) also proposed to do the following transformation,

$$\log r = z(t) + \alpha(t),$$

which results in the Black-Karasinski (1991) lognormal model with constant volatility,

$$d \log r = \kappa(\theta(t) - a \log r) + \sigma dW.$$

This specification of interest rates, however, may lead to infinite prices for Eurodollar futures (Hogan and Weintraub, 1993).

Define $\beta(i) = \sum_{j=1}^{i} \alpha(j) + z_1$, we have

$$\begin{aligned}\beta(2) &= -\frac{1}{\Delta t}\ln P(2) + \frac{1}{\Delta t}\ln\sum_{k=1}^{N}\pi_{1k}\exp(-z_k\Delta t) \\ &= -\frac{1}{\Delta t}\ln P(2) + \frac{1}{\Delta t}\ln \pi_1 \exp(-z\Delta t),\end{aligned}$$

where $\exp(-z\Delta t)$ is a vector with ith element given by $\exp(-z_i\Delta t)$ and π_1 the first row of the transition matrix. The two-step ahead state prices are then given by

$$\begin{aligned}Q_{2j} &= \sum_{k=1}^{N} Q_{1k}\exp(-(\alpha(2)+z_k)\Delta t)\,\pi_{kj} \\ &= \exp(-\beta(2)\Delta t)\sum_{k=1}^{N}\pi_{1,k}\left(e^{-z_k\Delta t}\right)\pi_{kj},\end{aligned}$$

or in matrix notation,

$$\begin{aligned}Q_2 &= \exp(-\beta(2)\Delta t)\,\pi_1 D\left(e^{-z\Delta t}\right)\pi \\ &= P(2)\,\frac{\pi_1 D\left(e^{-z\Delta t}\right)\pi}{\pi_1\left(e^{-z\Delta t}\right)},\end{aligned}$$

where $D\left(e^{-z\Delta t}\right)$ denotes a diagonal matrix with the ith diagonal element given by $\exp(-z_i\Delta t)$. In general, we have

$$\beta(i) = -\frac{1}{\Delta t}\ln P(i) + \frac{1}{\Delta t}\ln \pi_1^{i-1}\exp(-z\Delta t).$$

$\alpha(i)$ can be obtained from β by sheer difference. The state price density (a $1\times N$ row) at period i is

$$Q_{ij} = \frac{P(i)}{\pi_1^{(i-1)}\exp(-z\Delta t)}\pi_{1j}^{(i)},$$

where $\pi^{(i)} = \left[\pi^{(i-1)}D\left(e^{-z\Delta t}\right)\pi\right] = \left[\pi D\left(e^{-z\Delta t}\right)\right]^{i-1}\pi$. The interest rate at period i is given by

$$r(i) = \alpha(i) + z.$$

Again, Q_{ij} is today's price of a claim to one dollar in case state j happens in i periods and zero otherwise. Therefore, once the transition matrix is calibrated, the state price Q_{ij} can be computed easily from simple matrix manipulation.

3 Hull and White Trinomial Trees

For comparison, we present in this section a review of the trinomial tree approach of Hull and White (1990). As discussed above, once the state price densities are computed, contingent claims can be priced easily as functions of the state prices. In what follows, we illustrate how the trinomial tree is calibrated and how the state prices are constructed on the tree.

Let $z(i,j)$ denote the state variable at period i and state j. Conditional on $z(i,j)$, the state variable at next period can only happen in its three neighboring states: $\{z(i+1,j-1), z(i+1,j), z(i+1,j+1)\}$, with probabilities $\{\pi_u(j), \pi_m(j), \pi_d(j)\}$. This trinomial tree structure essentially assumes that from state (i,j) to time $i+1$, the probability of reaching any states other than the three neighboring ones is zero. For the tree to recombine, the tree has to be equal spaced. Let δ denote the distance between two adjacent states, then starting from $z(0,0) = 0$, we have $z(i,j) = j\delta$, which does not depend on the time period i, but just depend on the state index j. To avoid negative probability, the tree needs to be truncated at a certain point. At the upper bound of the truncation, the branching configuration changes to $\{z(i+1,j-2), z(i+1,j-1), z(i+1,j)\}$, and at the lower $\{z(i+1,j), z(i+1,j+1), z(i+1,j+1)\}$. In what follows, we will detail on how the tree is calibrated and how the state prices are constructed on the tree.

3.1 Probabilities

Recall the discrete-time approximation of the state variable process,

$$z_{t+\Delta t} = \varphi z_t + \sigma\sqrt{\Delta t}\varepsilon_{t+\Delta t}.$$

The probabilities are picked to match the mean and variance of the increment $\Delta z = z_{t+\Delta t} - z_t$:

$$\begin{aligned} E\left[\Delta z(i,j) \,|z(i,j)\right] &= -(1-\varphi)\, z(i,j) = M(j\delta); \\ Var\left[\Delta z(i,j) \,|z(i,j)\right] &= \sigma^2 \Delta t = V. \end{aligned}$$

However, since we have three parameters (two probabilities and δ), we need one extra moments to nail them down. Due to the symmetric nature of the tree, the third cumulant κ_3 of the increment has to be zero by construction. We can therefore use the condition on the fourth cumulant κ_4 of the increment. The three parameters (π_u, π_d, δ) can be solved from the following moment conditions:

$$\begin{aligned}
\pi_u(j)\,\delta + \pi_d(j)\,(-\delta) &= M(j\delta)\,; \\
\pi_u(j)\,\delta^2 + \pi_d(j)\,\delta^2 &= V + M^2(j\delta)^2\,; \\
\pi_u(j)\,\delta^4 + \pi_d(j)\,(-\delta)^4 &= \kappa_4 + 3V^2 + 6VM(j\delta) + M^4(j\delta)^4\,,
\end{aligned}$$

With the Brownian motion assumption in (1), we have $\kappa_4 = 0$. Then at $j = 0$, we have

$$\pi_0 = \pi_u(0) = \pi_d(0) = V/2\delta^2 = 3V^2/2\delta^4,$$

from which we have the grid

$$\delta^2 = 3V.$$

The start-up probability π_0 is hence given by

$$\pi_0 = V/2\delta^2 = 1/6.$$

For arbitrary j, the probabilities are

$$\begin{aligned}
\pi_u(j) &= \pi_0 + \frac{1}{2}\left[jM + (jM)^2\right]; \\
\pi_m(j) &= (1 - 2\pi_0) - (jM)^2\,; \\
\pi_d(j) &= \pi_0 + \frac{1}{2}\left[-jM + (jM)^2\right].
\end{aligned}$$

In discrete time, the Gaussian variable ε can be replaced by any other distributions that best fit the data. Using monthly data on U.S. Treasuries, we run a simple AR(1) type regression on the one-month yield (the short rate) and find that residuals departure profoundly from normality. The skewness ($\gamma_1 = \kappa_3/\kappa_2^{3/2}$) of the residual is -0.48 and the kurtosis ($\gamma_2 = \kappa_4/\kappa_2^2$) 9.10. Both measures are zero for normal distribution.

The current trinomial tree structure cannot accommodate non-zero skewness but, at least in theory, it can capture the observed positive kurtosis by using a larger value for the grid δ and a smaller value for the branching probability π_0. With $k_4 = \gamma_2 V^2$, we have

$$\pi_0 = V/2\delta^2 = (\gamma_2 + 3)\,V^2/2\delta^4,$$

from which we have the grid

$$\delta = \sqrt{(\gamma_2 + 3)\,V},$$

and the start-up probability π_0

$$\pi_0 = V/2\delta^2 = \frac{1}{2(\gamma_2+3)}.$$

A kurtosis of $\gamma_2 = 9$ greatly reduces the branching probability from 1/6 to about 4%. Once π_0 is determined, $\pi(j)$ can be determined as before.

One problem with this tree specification is that after a few nodes, we start to get negative probabilities. The intuition underlying this phenomenon is that since the process for z is stationary, the state can not expand forever. In the case of $r = z$, the stationarity of the interest rate implies that interest rates cannot be too large with positive probabilities. Specifically, let $j_{\max}$ and $j_{\min}$ denote the upper and lower bound of states. They can be proved to satisfy the following condition:

- If $\pi_0 \geq 1/8$, $\pi_m(j)$ becomes negative first when either $j_{\max} \geq -\sqrt{1-2\pi_0}/M$ or $j_{\min} \leq \sqrt{1-2\pi_0}/M$.
- If $\pi_0 < 1/8$, $\pi_u(j)$ hits zero first when $j_{\max} \geq \left(-1+\sqrt{1-8\pi_0}\right)/2M$, and $\pi_d(j)$ hits zero first when $j_{\min} \leq \left(1-\sqrt{1-8\pi_0}\right)/2M$.

As can be seen from these conditions, the smaller the initial probability π_0 is, the narrower the bounds are, that is, truncations have to be made at an earlier stage. Incorporating positive kurtosis reduces π_0 and hence causes earlier truncation. Also, not surprisingly, stronger mean-reversion, as captured by a larger absolute value for M, leads to early truncation.

The tree configuration is modified at the two bounds. At the upper bound, the tree can stay put, go down one step, or go down two steps. Similarly, at the lower bound, the tree can stay put, go up one step, or go up two steps. If we call these by the same names (up, middle, and down), the probabilities at the upper bound solve the following equations:

$$\begin{aligned} E[\Delta z(i,j)|z(i,j)] &= \pi_m^u(j)(-\delta) + \pi_d^u(j)(-2\delta) = M(j\delta); \\ Var[\Delta z(i,j)|z(i,j)] &= \pi_m^u(j)(-\delta)^2 + \pi_d^u(-2\delta)^2 = V + (j\delta)^2 M^2. \end{aligned}$$

The solution is

$$\begin{aligned} \pi_u^u(j) &= 1 + \pi_0 + \frac{1}{2}\left[3jM + (jM)^2\right]; \\ \pi_m^u(j) &= -2\pi_0 - 2jM - (jM)^2; \\ \pi_d^u(j) &= \pi_0 + \frac{1}{2}\left[jM + (jM)^2\right]. \end{aligned}$$

At the lower bound, we have $\pi_u^d(j) = \pi_d^u(j)$, $\pi_m^d(j) = \pi_m^u(j)$, and $\pi_d^d(j) = \pi_u^u(j)$.

3.2 State-price density

Once the probabilities are set, we can derive the state-price density based on the current term structure. The process is similar to that in Section 2.3. The only difference is that, due to the restricted structure of the trinomial tree and its boundaries, the notation is messier and several conditional arguments have to incorporated in the process.

Similarly, given the current price for the one-period bond, we have

$$P(1) = \exp(-\alpha(1)\Delta t) \sum_{k=-1}^{1} \pi_k(0) = \sum_{k=-1}^{1} Q_{1k}, \tag{6}$$

where $\pi_k(0) = \{\pi_u(0), \pi_m(0), \pi_d(0)\}$, and Q_{ij} is the price of the state j at period i. The shift parameter $\alpha(1)$ can be solved from (6),

$$\alpha(1) = -\frac{1}{\Delta t} \ln P(1),$$

and the one-period state price can be written as

$$Q_{1j} = \exp(-\alpha(1)\Delta t)\, \pi_j(0).$$

Similarly, from the two-period bond price,

$$P(2) = \sum_{k=-1}^{1} Q_{1k} \exp(-(\alpha(2) + k\delta)\Delta t) \sum_{j=-1}^{1} \pi_j(k) = \sum_{j=-2}^{2} Q_{2j},$$

we can derive the two-period drift $\alpha(2)$ and the state price Q_{2j}:

$$\begin{aligned} \alpha(2) &= -\frac{1}{\Delta t} \ln P(2) + \frac{1}{\Delta t} \ln \left[\sum_{k=-1}^{1} Q_{1k} \exp(-k\delta\Delta t) \right]; \\ Q_{2j} &= \sum_{k} Q_{1k} \exp(-(\alpha(2) + k\delta)\Delta t)\, \pi_j(k), \end{aligned}$$

where the summation over k is over all intermediate states that leads to the final branch j. Special care needs to be applied at the edges where irregularities arise. In general terms, we have

$$\begin{aligned} \alpha(i) &= -\frac{1}{\Delta t} \ln P(i) + \frac{1}{\Delta t} \ln \left[\sum_{k=\max[-i+1, j_{\min}]}^{\min[i-1, j_{\max}]} Q_{i-1,k} \exp(-k\delta\Delta t) \right]; \\ Q_{ij} &= \sum_{k} Q_{i-1,k} \exp(-(\alpha(i) + k\delta)\Delta t)\, \pi_j(k). \end{aligned}$$

4 Cap Pricing

In this section, we derive the analytical pricing formula for caps when the underlying interest rate process is represented by the *extended* Vasicek model:

$$dr = -\kappa \left(r - \theta(t)\right) dt + \sigma dW$$

The extension here comes from the time-dependent specification of $\theta(t)$. Such a model can be calibrated to match the current term structure perfectly, as we did in the state price construction. For ease of comparison, we rewrite the process in the discrete-time version:

$$r_{t+\Delta} = -\left(1-\varphi\right)\theta(t) + \varphi r_t + \sigma\sqrt{\Delta t}\varepsilon_{t+\Delta} \tag{7}$$

where $\varphi \approx 1-\kappa\Delta t$ captures the autocorrelation of the interest rate. $\varepsilon_{t+\Delta t}$ is assumed to be i.i.d. standard normal in the Vasicek model. We further extend the model by incorporating the well-documented non-normality in the innovation. Specifically, we replace the normal ε with a Gram-Charlier density:

$$g\left(\varepsilon\right) = f\left(\varepsilon\right) - \gamma_1 \frac{1}{3!} D^3 f\left(\varepsilon\right) + \gamma_2 \frac{1}{4!} D^4 f\left(\varepsilon\right),$$

where $f\left(\varepsilon\right) = \left(2\pi\right)^{-1/2} \exp\left(-\varepsilon^2/2\right)$ is the standard normal density and D^j denotes the jth derivative of what follows. As a standardized variable, γ_1 and γ_2 capture, respectively, the skewness and kurtosis of ε.

An interest rate cap is composed of a series of caplets. The payoff function of a typical (ith) caplet can be written as

$$\tau L \left(R_{t+i\tau} - K\right)^+,$$

where τ is the *tenor* of the caplet, which dictates the payment interval, L is the principal, K is the strike price, and R is the simply compounded interest rate given by

$$R_t = \frac{1}{\tau}\left(e^{\tau r_t} - 1\right).$$

To price the caplet, we need to know the conditional density of interest rates multiple periods ahead. From the interest rate process in (7), we have the n-period ahead interest rate process as

$$r_{t+n\Delta t} = \left(1-\varphi^n\right)\theta(t) + \varphi^n r_t + \sigma\sqrt{\Delta t}\sum_{j=0}^{n-1}\varphi^j \varepsilon_{t+(n-j)\Delta t}.$$

The conditional cumulants of $r_{t+n\Delta t}$ are

$$\begin{aligned}
\mu_n &= \theta(t)\left(1-\varphi^n\right)+\varphi^n r_t; \\
\sigma_n^2 &= \frac{1-\varphi^{2n}}{1-\varphi^2}\sigma^2\Delta t; \\
\kappa_{3n} &= \frac{1-\varphi^{3n}}{1-\varphi^3}\sigma^3(\Delta t)^{3/2}\gamma_1; \\
\kappa_{4n} &= \frac{1-\varphi^{4n}}{1-\varphi^4}\sigma^4(\Delta t)^2\gamma_2; \\
\kappa_{jn} &= 0 \quad \text{for } j \geq 4.
\end{aligned} \tag{8}$$

The conditional skewness and kurtosis of $r_{t+n\Delta t}$ are, respectively, the variance-normalized third and fourth cumulants:

$$\begin{aligned}
\gamma_{1n} &= \gamma_1\frac{1-\varphi^{3n}}{1-\varphi^3}/\left(\frac{1-\varphi^{2n}}{1-\varphi^2}\right)^{3/2}; \\
\gamma_{2n} &= \gamma_2\frac{1-\varphi^{4n}}{1-\varphi^4}/\left(\frac{1-\varphi^{2n}}{1-\varphi^2}\right)^{2},
\end{aligned} \tag{9}$$

both of which decreases monotonically with n while γ_{2n} decreases in a faster pace than γ_1 does. In the extreme case when $\varphi = 1$ (no mean-reversion), skewness γ_1 decreases linearly with $1/\sqrt{n}$ while kurtosis γ_2 decreases with $1/n$.

Define the standardized variable

$$\omega = \frac{r_{t+n\Delta t}-\mu_n}{\sigma_n},$$

the Gram-Charlier density for ω is then given by

$$g(\omega) = f(\omega) - \gamma_{1n}\frac{1}{3!}D^3 f(\omega) + \gamma_{2n}\frac{1}{4!}D^4 f(\omega).$$

Let $n = i\tau/\Delta t$ and $m = (i+1)\tau/\Delta t$ denote the number of periods from today (time t) to the ith and $(i+1)$th payment, then the value of the ith caplet can be written as

$$c_{it} = \tau L \mathbb{E}_t\left[M_{t,t+m\Delta t}\left(R_{t+n\Delta t}-K\right)+\right]$$

where $M_{t,t+m\Delta t}$ is a multi(m)-period stochastic factor ("the pricing kernel"). Note that caplets are paid in "rears," that is, the ith caplet is settled at period n but the payment is made at period m, one tenor later. Redefine the probability space such that M and R are independent under such a probability measure, which we label, as most others do, as the "risk-neutral" measure. The caplet can then be priced under the "risk-neutral" measure $\mathbb{E}^*[\cdot]$:

$$\begin{aligned} c_{it} &= LP_t(m)\mathbb{E}_t^*\left[(\tau R_{t+n\Delta} - \tau K)^+\right] \\ &= P_t(m)\int_{\overline{\omega}}^{\infty}\left(e^{\tau(\mu_n^*+\sigma_n\omega)} - 1 - \tau K\right) g(\omega)\, d\omega \end{aligned} \tag{10}$$

where $P_t(m)$ is the current price of a zero-coupon bond with maturity of m periods, μ_n^* is the risk-adjusted drift over n periods and $\overline{\omega}$ is the standardized variable evaluated at K,

$$\overline{\omega} = \frac{\frac{1}{\tau}\ln(1+\tau K) - \mu_n^*}{\sigma_n}.$$

Moneyness d is defined as $-\overline{\omega}$.

Repeated integration of (10) and application of the parity conditions:

$$\begin{aligned} 1+\tau F_t^{n,m} i &= \exp\left(\tau\mu_n^* + \frac{1}{2}\tau^2\sigma_n^2 + \frac{\gamma_{1n}}{3!}\tau^3\sigma_n^2 + \frac{\gamma_{2n}}{4!}\tau^4\sigma_n^4\right); \\ 1+\tau K &\equiv \exp(\tau\mu_n^* - \tau\sigma_n d); \end{aligned}$$

yield the following caplet pricing formula

$$\begin{aligned} c_{it} &= P_t(m)\left[(1+\tau F_t^{n,m})\, N(d+\tau\sigma_n) - (1+\tau K)\, N(d)\right] \\ &\quad - P_t(m)\,(1+\tau K)\, f(d)\,\tau\sigma_n\left[\frac{\gamma_{1n}}{3!}(d-\tau\sigma_n)\right. \\ &\quad \left. + \frac{\gamma_{2n}}{4!}\left(1-d^2+\tau\sigma_n d - \tau^2\sigma_n^2\right)\right]. \end{aligned} \tag{11}$$

The moneyness measure d can be rewritten as

$$d = \frac{\ln\left[(1+\tau F)/(1+\tau K)\right] - \frac{1}{2}\tau^2\sigma_n^2 - \frac{\gamma_{1n}}{3!}\tau^3\sigma_n^3 - \frac{\gamma_{2n}}{4!}\tau^4\sigma_n^4}{\tau\sigma_n}. \tag{12}$$

$F_t^{n,m}$ is the simply compounded forward rate between period n and m and is given by

$$F_t^{n,m} = \frac{P_t(n) - P_t(m)}{P_t(n)}.$$

The conditional variance σ_n^2 and the conditional skewness and kurtosis $(\gamma_{1n}, \gamma_{2n})$ are given, respectively, in (8) and (9). When the innovation is conditionally normal, we have

$$c_{it} = P_t(m)\left[(1+\tau F_t^{n,m})\, N(d+\tau\sigma_n) - (1+\tau K)\, N(d)\right], \tag{13}$$

a close analogy to the Black formula.

5 Calibration

To examine the pricing accuracy and efficiency of our markov chain approximation, we calibrate the markov chain to the one-month yield of U.S. Treasury bonds and use it to price a series of five-year caplets with a tenor of 1 month. We assume, for now, that the one-month yield, or the short rate, follows the extended Vasicek model specified in (7). Since we have obtained analytical solutions to the price of caplets under such a specification, we can use the analytical solution as a natural benchmark and gauge the accuracy of our approximations. For comparison, we also calibrate the data to the Hull and White trinomial tree and use it to price the caplets. We compare the accuracy and efficiency of both approaches.

5.1 Data and calibration

For that purpose, we calibrate the one-month yield of U.S. government bonds to an AR(1) process as specified in (7) to obtain the following parameter values:

$$\begin{bmatrix} \theta & = & 6.68\% \\ \kappa & = & 0.49 \\ \sigma & = & 2.21 \times 10^{-3} \end{bmatrix}$$

The data are not observed but computed with the Smoothed Fama-Bliss method using programs supplied by Robert Bliss. The data are monthly, from January 1972 to December 1995.

For simplicity, we assume constant drift: $\theta(t) = \theta$ and zero market price of risk ($\lambda = 0$) and then derive bond yields with maturity n analytically:

$$y_t(n) = \frac{1}{n\Delta t}\left[A_n + B_n r(t)\right]$$

with $r(t)$ being the one-month yield, or short rate, and A_n and B_n can be computed readily through the following iteration

$$\begin{aligned} A_{n+1} &= A_n + B_n(1-\varphi)\theta\Delta t - (B_n\sigma)^2\Delta t/2 \\ B_{n+1} &= 1 + B_n\varphi \end{aligned}$$

starting with $A_0 = B_0 = 0$. Recall that the autocorrelation is given by $\varphi = \exp(-\kappa\Delta t) \approx 1 - \kappa\Delta t$. Refer to Backus, Foresi, and Telmer (1999) for an excellent review on discrete-time bond pricing. Bond prices can be computed from the yields by

$$P_t(n) = \exp\left(-y_t(n)n\Delta t\right)$$

In reality, of course, we can use bond prices observed from the market. Indeed, in the same data set, we do have bond prices of maturities ranging from one-month to 10 years. Using the hypothetical bond prices just simplifies the computation. We further assume that the current interest rate is equal to its long run mean: $r(t) = \theta$. With the assumption of zero market price, the yield curve is rather flat.

We calibrate both the markov chain and the Hull and White trinomial tree to this yield curve. Assuming conditional normality, the initial probability in the trinomial tree is $\pi_0 = 1/6$ and, based on the parameter given above with a monthly interval: $\Delta t = 1/12$, the maximum number of up states and the maximum number of down states are $j_{\max} = j_{\min} = 20$. Therefore, the maximum number of states in the trinomial tree is 41.

Recall that the maximum number of states before truncation is given by

$$j_{\max} = -j_{\min} \leq \sqrt{1 - 2\pi_0}/(1 - \varphi).$$

Hence, given π_0, the number of states depends crucially on the autocorrelation: the smaller φ is, the stronger the mean-reversion of the process, the earlier truncation has to be taken place. To increase the number of states in a trinomial tree, we need to use smaller time intervals Δt such that we can have a bigger $\varphi = \exp(-\kappa \Delta t)$. Table 1 lists, among other things, the maximum number of states that can be achieved in a trinomial tree with different time intervals.

The number of states in a markov chain, on the other hand, can be arbitrarily chosen independent of the time-interval choice. While the number of states decides the precision of the approximation, the time interval choice is solely determined by the derivative pricing needs. For comparison, we also use $N = 41$ states to calibrate the markov chain. We also choose a monthly interval for the markov chain in accordance with the tenor of the caplet. With the conditional normality assumption, we calibrate the transition matrix using the Gauss-Hermite quadrature.

We proceed to price a series of 5-year caplets with a tenor of 1 month and with different strike prices. The cap contract has a principal of \$1,000. Under such a set-up, the price of a caplet can be easily obtained analytically from equation (13). For simplicity, we also discretize the process using monthly interval: $\Delta t = 1/12$. Since the maturity of the caplet is 5 years, we have $n = 5/\Delta t = 60$ and $m = (n + \tau)/\Delta t = 61$.

The caplet price can also be computed from the calibrated markov chain or the trinomial tree:

$$\tau L \delta Q_n \max\left[R_n' - K, 0\right],$$

where K is the strike price of the caplet. $\delta = \frac{P_t(m)}{P_t(n)}$ is a discount factor adjustment since the caplet price is settled at period n while paid at period m. Q_n denotes the state price row $(1 \times N)$ at period n and R_n is the simply compounded (over τ period) interest rate state vector $(1 \times N)$ at period n. $L = 1,000$ is the principal of the contract.

The pricing accuracies of the two approximations are measured by their respective percentage pricing bias, defined as

$$\begin{aligned} Bias_M &= 100 \times \frac{C_M - C_A}{C_A}; \\ Bias_T &= 100 \times \frac{C_T - C_A}{C_A}. \end{aligned}$$

where subscript M denotes the option price obtained by markov chain, T denotes the trinomial tree, and A, the analytical solution.

5.2 Computational efficiency

All the computations are done on a Pentium II personal computer (450 MHz, 64MB RAM). Computing a 41-state Gaussian quadrature requires about 1 second. Increasing the number of states increases the computational time exponentially, as shown in Table 1. However, once the quadrature is computed, it can be saved in the computer for future use and need not to be re-computed. Indeed, some books about numerical techniques provide tables for quadratures with smaller number of states.

With the quadrature rule given, calibrating the markov chain (41 states) and computing state prices and interest rates up to 120 periods (10 years) cost about 0.72 seconds. Calibrating the trinomial tree and computing the same require 2.86 seconds.

To have better precision, we may need to increase the number of states. This can be easily done with the markov chain approximation by just using Gaussian quadrature rule with the required number of states. To obtain more states in the trinomial tree, however, we need to increase the number of time steps. To save computational time, we will just use time steps fine enough to generate the required number of states:

$$\Delta t = -\frac{\ln(1 - \sqrt{1 - 2\pi_0}/j_{\max})}{\kappa}.$$

We round off Δt such that $1/\Delta t$ is the smallest allowable integer:

$$\text{Int}[1/\Delta t] \geq 1/\Delta t.$$

Table 1 illustrate the time (in seconds) required to calibrate the markov chain and the trinomial tree and to compute up to ten years of state prices and interest rates. The listed time is an average of 10 identical runs for each case. In all cases, the markov chain approach is about twice as fast in the trinomial tree approach.

5.3 Pricing accuracy

Figure 1 compares the interest rate configurations obtained from the markov chain approach and the trinomial tree approach, with the same final number of states ($N = 41$). The configurations have three major differences. First, for the markov chain representation, the number of states are constant across periods while for the trinomial trees, the number of states increases linearly from 3 to 41, until which the tree is truncated to avoid negative probability. Actually, the trinomial tree can also be regarded as a 41-state markov chain, but with zero probabilities for transitions to non-neighboring states. Second, the distance between neighboring states is set to be equal in the trinomial tree so that the trees are recombining. In the markov chain, however, the choices of nodes are also free and are optimized. As a result of the density distribution, we have more states in the middle and fewer in the tails. Third, over long periods, the trinomial tree covers a bigger domain than that of the markov chain. The interest rate domain for the markov chain is $[5.81, 7.41]$ (in percentage) over the calibrated period while that for the trinomial tree is $[4.36, 8.82]$. As shown later in this section, the range of the interest rate in the markov chain approach depends crucially on the choice of the weighting function. These differences will be reflected in their pricing accuracy on the caplets.

The pricing results for a series of 5-year caplets at different strike prices are listed in Table 2. Within reasonable moneyness, both trinomial trees and the markov chain (with $N = 41$) result in fairly accurate option prices. Figure 2 depicts the percentage pricing bias of both approaches on a wider range of strike prices. First, we see that for close-to-the-money options, pricing biases from trinomial trees oscillate widely due to the wide equal-distance state spacing. That problem is relatively small for markov chain pricing due to its efficient spacing.

Secondly, big pricing bias mainly comes from far out-of-the-money options for both approaches. While markov chain tends to underprice far out-of-the money options, trinomial trees tend to do the opposite. It implies that the trinomial tree assigns a over-heavy density to the tail while the markov chain does the opposite. The heavy tail of the trinomial tree is the result of the arbitrary truncation and restricted configuration at the edge. Once the tree hits the edge, it is forced to stay in the neighbor. In contrast, the quadrature method, based on optimal integration consideration, assign less weight and fewer states at the tail and results in underpricing for fare out-of the money options.

Lastly, due to the finite range of the states, neither the tree nor the markov chain is incapable of pricing far out-of-the-money options with strike price beyond their minimum states. Since the markov chain has a narrower range ($[5.81, 7.41]$) when using the conditional density (conditional on the long-run mean) as the weighting function, such failure comes at a earlier stage (when $d < -3$) while the failure for the trinomial tree comes at a much more extreme moneyness ($d > -8$). However, due to the increasingly over weights at the tail, the trinomial tree grossly overprices far out-of-the-money options and works no better than simply assigns a zero value to these options.

In short, for close-to-the-money options, e.g. for moneyness within the range $[-2, 2]$, both the markov chain and the trinomial tree approach do a passable job in terms of pricing accuracy. Among them, the markov chain approximation does a better job in terms of reducing oscillation. *The markov chain approach is not only faster but also more accurate for near at-the money options.*

Although at-the-money options are the most liquid ones in the market, far out-of-the-money options can become very useful, yet inexpensive hedging instruments. The pricing accuracy of neither method is satisfactory for far out-of-the-money options. In what follows, we propose three ways to improve the pricing accuracy and evaluate their relative efficiency.

5.3.1 Increasing number of states

Increasing number of states is the most straightforward way to increasing the pricing accuracy of all options. Increasing the number of states not only increase the fineness of the grid but also increases the range of the interest rate state for both methods. For example, when we use $N = 141$, the interest rate domain for the markov chain approach extends to $[5.08, 8.13]$, the pricing bias of the markov chain method is reduced to within 1% for moneyness $d < -4$. Similar effects happen to the binomial tree approximation. The computational cost, however, has been shown to be increasing very fast (exponentially) with increasing number of states for both methods.

5.3.2 Changing the weighting function

One way to increase the pricing range of the markov chain method without sacrificing the computational time is to change the weighting function $\omega(x)$. Instead of using the conditional density at the mean $\omega(x) = f(x|0)$, using the unconditional density $\omega(x) = f(x)$ should greatly increase the range of the states due to its larger variance. The states for z are now

$$z_i = x_i \sqrt{\frac{2\sigma^2}{1-\varphi^2}}.$$

Due to the much bigger unconditional variance, the interest rate range is widened to $[3.91, 9.31]$, even wider than that of the trinomial tree. The pricing bias result is illustrated in Figure 3. Now the markov chain model can price options as far out of the money as $d < -12$, an extremely small probability of happening ($\mathbb{P}(x < -12) = 1.78 \times 10^{-33}$) under the normality assumption. Also, as the moneyness decreases, the pricing bias is not increasing as much as in the trinomial tree case.

Nevertheless, when we use the unconditional density as the weighting function, we are essentially spreading out the states to a wider range. While that enables us to price far out-of-the money options, it also reduces the number of states in the middle and therefore the pricing bias for close-to-the-money options are bigger and the oscillation becomes as severe as in the trinomial tree approach, at least for close-to-the-money options. As found by Tauchen and Hussey (1991), using conditional density (conditional on the long-run mean) as the weighting matrix performs better for integration purposes. But to price far out-of-the-money options, we need a wider range of interest rate states and hence using the unconditional density as the weighting function performs better. This is especially true when the mean-reversion is small.

5.3.3 Moneyness-efficient quadrature

The original choice of the quadrature nodes and weights are from the consideration of performing an efficient integration. When applying to the option pricing, say a

European call option, we have the integral

$$C = \int_K^\infty (S - K) f^*(S) dS \tag{14}$$

where $f^*(S)$ is the risk-adjusted density of the underlying asset price S. The quadrature approximation of the density function in general does not depend on the integrand function $S - K$ but does depend on the range of the integration. The Gauss-Hermite quadrature we have been using is chosen on the range $S \in [-\infty, \infty]$ while the integration in (14) is on $S \in [K, \infty]$.

Therefore, for options with given strike price (or given moneyness), we can obtain a better approximation of the integral by choosing quadratures on the interval $[K, \infty]$. By approximating the integral within the range $S \in [K, \infty]$, we can allocate all the quadrature points to the relevant range and hence increase the efficiency and pricing accuracy of the markov chain approximation.

6 Further Applications

In this section, we illustrate the extreme flexibility of the markov chain approach in pricing derivatives when the underlying interest rate follows complex dynamics, such as conditional non-normality, heteroskedasticity, and multiple factors, all of which are salient features of the interest rates.

6.1 Conditional non-normality

It is well-documented that interest rates possess pronounced conditional non-normality. Section 4 illustrates how to capture conditional non-normality, particularly, skewness and kurtosis, using a Gram-Charlier A-type distribution. As illustrated in Section 3, under the original Hull and White trinomial tree set-up, one cannot capture non-zero skewness. Excess kurtosis can be captured by using a smaller initial branching probability (π_0) and a larger grid (δ):

$$\begin{aligned} \pi_0 &= \frac{1}{2(\gamma_2 + 3)} \\ \delta &= \sqrt{(\gamma_2 + 3)V} \end{aligned}$$

This, however, bring in other problems to the calibration. Using the one-month treasury yield data as described before, the conditional kurtosis of the AR(1) residual is $\gamma_2 = 9.10$. Correspondingly, if we want to match this fourth moment in the trinomial tree, the initial probability π_0 has to be reduced to 4.13%. But then, with a monthly interval, the maximum number of states are reduced to 5 ($\jmath_{\max} = -\jmath_{\min} = 2$). The small number of states severely limits the pricing accuracy of the trinomial tree. Actually, as illustrated in Figure 4, the interest rates configuration is no longer well-behaved when the conditional kurtosis is too big. The key problem is that it is difficult to capture conditional higher moments with just three equal-distance nodes.

The markov chain modeling, on the other hand, can handle non-normality easily because it can be calibrated to match the conditional distribution up to $2N - 1$ moments. And in most scenarios, we can still use a normal distribution and thus the Gauss-Hermite quadrature as the weighting function. The only thing we need to change is to replace the normal conditional density function with any conditional density function we intend to use. That is, in the transition matrix specification,

$$\pi\left[z_j|z_i\right] = \frac{g\left(z_j|z_i\right)}{s\left(z_i\right)\omega(z_j)} w_j, j = 1, ..., N$$

we replace the normal density $f(\cdot)$ for any density function $g(\cdot)$ we want while keeping the weighting function $\omega(x)$ as Gauss-Hermite, matching the first two moments of either the conditional density conditional on the long-run mean or the unconditional density. In the case of Gram-Charlier type-A distribution, we have

$$g(z_j|z_i) = f(z_j|z_i)\left(1 + \frac{1}{6}\gamma_1 H_3 + \frac{1}{24}\gamma_2 H_4\right) \tag{15}$$

where $f(z_j|z_i)$ is the normal density function with mean and variance equal to the conditional mean and variance of z_j conditional on z_i. H_3 and H_4 are two hermite terms of the standardized variable:

$$\begin{aligned} H_3 &= x^3 - 3x; \\ H_4 &= x^4 - 6x^2 + 3; \\ x &= \frac{z_j - \varphi z_i}{\sigma}. \end{aligned}$$

Figure 5 depicts the impact of conditional non-normality on option pricing. In the analytical valuation, we match the conditional skewness and kurtosis to that of the short rate: $\gamma_1 = -0.48$, $\gamma_2 = 9.10$, which are estimated from the AR(1) regression residual. The solid line depicts the percentage bias between the markov chain pricing ignoring non-normality and the analytical price with non-normality (in equation (11)):

$$Bias = 100 \times \frac{C_M(normal) - C_A(GC)}{C_A(GC)},$$

where GC denotes the Gram-Charlier representation of the non-normality. The dashed line depicts the percentage bias between the markov chain pricing and the analytical pricing, both considering non-normality:

$$Bias = 100 \times \frac{C_M(GC) - C_A(GC)}{C_A(GC)}.$$

The solid line indicates that ignoring non-normality overprices at-the-money options while under-prices out-of-the-money options. Positive kurtosis implies higher probabilities at the tails than implied by a normal distribution. Ignoring it will under-estimate the tail probability and hence the price of out-of-the-money options. The dashed line illustrates that *taking non-normality into account in the markov chain calibration significantly reduces these biases.*

6.2 Multi-factor models

Empirical evidence has indicated that two or more factors are needed to model the term structure of interest rates. While multi-factor modeling on a tree can be extremely complicated, it can relatively easily be done on a markov chain. Tauchen and Hussey (1991) provide examples on how to model a VAR process and GARCH process. In what follows, we illustrate how the markov chain representation can be applied to the general class of affine models.

In affine models, bond yields are linear functions of a vector of state variables. This class of models include popular examples developed by, among others, Balduzzi, Das, Foresi, and Sundaram (1996), Brennan and Schwartz (1979), Chen and Scott (1993), Cox, Ingersoll, and Ross (1985), Dai and Singleton (1997), Duffie and Kan (1996), Longstaff and Schwartz (1992), Pearson and Sun (1994), and Vasicek (1977). The ease of deriving implications for long maturities makes this class extremely attractive for applied work.

Our version of the class of affine models is taken from Duffie and Kan (1996), which we translate into discrete-time. A vector of state variables z follows

$$z_{t+\Delta t} - z_t = -\left(I - \Phi\right)\left(\theta - z_t\right) + V\left(z_t\right)^{1/2}\sqrt{\Delta t}\varepsilon_{t+\Delta t}$$

where $\{\varepsilon\} \sim NID(0, I)$ but can be easily extended to incorporate conditional non-normality. $V(z)$ is a diagonal matrix with typical element

$$v_i(z) = \alpha_i + \beta_i^T z,$$

β_i has nonnegative elements, and Φ is table with positive diagonal elements. The pricing kernel is

$$-\log m_{t+\Delta t} = \delta + \gamma^T z_t + \lambda^T V(z_t)^{1/2} \sqrt{\Delta t} \varepsilon_{t+\Delta t}.$$

The process for z requires that the volatility functions v_i be positive.

The examples listed earlier are special cases of this structure. In the Vasicek (1977) model, z is a scalar, $\beta = 0$, $\gamma = 1$, and $\delta = \lambda^2 \alpha/2$. In the Cox-Ingersoll-Ross (1985) model, z is also a scalar, $\alpha = 0$, $\gamma = 1 + \lambda^2 \beta/2$ and $\delta = 0$. In both cases, z is the short rate. Since $\varepsilon \sim NID(0, I)$, we can approximate the process using a multivariate product rule by combining a set of one-dimensional Gauss rules with $N = \Pi_{j=1}^K n_j$ where K is number of factors and n_j is the number of nodes used for the jth factor.

Let $\left\{x_j^k, w_j^k\right\}_{j=1}^{n_k}$ denote the Gauss rule for the kth standardized innovation ε^k, use the conditional density conditional on the long run mean (zero) as the weighting function for each factor $\omega\left(z^k\right) = f\left(z^k|\theta\right)$, we have the nodes for z^k

$$\left\{z_j^k\right\}_{j=1}^{n_k} = \theta_k + \sqrt{2 v_k(\theta) \Delta t} \left\{x_j^k\right\}_{j=1}^{n_k},$$

from which we can form $N = \Pi_{i=1}^K n_i$ nodes of K-tiples, $\left\{Z_j = \left\{z^k\right\}_{k=1}^K\right\}_{j=1}^N$. The transition probability of the matrix $\left(\Pi^N\right)$ is then given by

$$\Pi_{ij}^N = \frac{f(Z_j|Z_i)}{s(Z_i)\, \omega(Z_j)} w_j,$$

where $w_j = \left(\Pi_{i=1}^K w^i\right)_j$, $\omega(Z_j) = \left[\Pi_{i=1}^K \omega\left(z^i\right)\right]_j$ and $f(Z_j|Z_i)$ is a multivariate normal density with mean $(I - \Phi)\theta + \Phi Z_i$ and variance $V(Z_i)\Delta t$. This normality assumption can be easily relaxed to include higher moments, as we did on the Vasicek model.

Once the nodes for state variables are constructed, so are the nodes for interest rates (yields) and bond prices, which are, respectively, linear and exponential-linear functions of the state variables:

$$\begin{aligned} y_j^n &= \frac{A_n + B_n' Z_j}{n \Delta t}; \\ P_j^n &= \exp\left(-A_n - B_n' Z_j\right), \end{aligned}$$

where A_n and B_n are given by the following iterations:

$$\begin{aligned} A_{n+1} &= A_n + \delta + B_n^T (I - \Phi)\theta\Delta t - \frac{1}{2}\sum_{j=1}^{k}(\lambda_j + B_{j,n})^2 \alpha_j \Delta t; \\ B_{n+1}^T &= \gamma^T + B_n^T \Phi - \frac{1}{2}\sum_{j=1}^{k}(\lambda_j + B_{j,n})^2 \beta_j^T, \end{aligned}$$

which can be easily computed starting with the initial condition: $A_0 = 0$ and $B_0 = 0$. Furthermore, given the current bond prices or yields, the current values of the state variables can be figured out by affine transformation.

Similarly, let state 1 denote the current state, state prices (at period i and state j) can be solved recursively from the current bond prices:

$$\begin{aligned} Q_{ij} &= \frac{P_1^i}{\Pi_1^{i-1} \exp(-B_1' Z \Delta t)} \Pi_{1j}^i \\ &= \frac{\exp(-A_i - B_i' Z_1)}{\Pi_1^{i-1} \exp(-B_1' Z \Delta t)} \Pi_{1j}^i \end{aligned}$$

where $\Pi^i = \left[\Pi D\left(e^{-B_1' Z \Delta t}\right)\right]^{i-1} \Pi$. An extended version of this is to allow δ be time-dependent so that we can fit the whole current yield curve.

In what follows, we price the same series of five-year caplets on the short rate with a tenor of one month. As an example of the above affine class, we adopt a two-factor model from Backus, Telmer, and Wu (1999), the parameters of which are also estimated from the same monthly U.S. Treasury yield data described above. The model has been shown to be able to account for a number of stylized evidence of the bond yields. The parameters are:

$$\begin{aligned} \theta &= \begin{bmatrix} 3.34\% & 3.34\% \end{bmatrix}^T \\ \Phi &= \begin{bmatrix} 0.4941 & 0.8114 \\ 0 & 0.9591 \end{bmatrix} \\ \alpha &= \begin{bmatrix} 1.4252(-6) & 0 \end{bmatrix}^T \\ \beta &= \begin{bmatrix} 0 & 0 \\ 0 & 2.6096(-5) \end{bmatrix} \\ \lambda &= \begin{bmatrix} -2590.3 & -261.28 \end{bmatrix}^T \end{aligned}$$

γ and δ are set:

$$\begin{aligned}
\gamma &= 1 + \beta^T \lambda^2; \\
\delta &= \frac{1}{2} \sum_{j=1}^{k} \alpha_j \lambda_j^2 + \frac{1}{4!} \sum_{j=1}^{k} \alpha_j^2 \lambda_j^4 \gamma_{j,2},
\end{aligned}$$

such that the short rate is a sum of the two-factors: $r(t) = z_1(t) + z_2(t)$. Parameters (mean and variance) are annualized with $\Delta t = 1/12$. This example captures the key features of the affine class. The second factor follows a square-root process as in the Cox-Ingersoll-Ross model (1985), the first factor is in the spirit of Vasicek with constant conditional volatility. The two factors are correlated through the non-zero off-diagonal element of the Φ matrix. In such a model, we have incorporated not only multiple factors but also a heteroskedasticity: the conditional variance of the second factor is proportional to its current level: $Var_t(z_{2,t+\Delta t}) = \beta(2,2) z_{2t}$. Conditional non-normality can also be added to the first factor without changing the basic affine structure and, as has been shown in the previous section, without posing any extra extra difficulty in the markov chain calibration.

Specifically, we calibrate the markov chain by using a $n_j = 11$ state Gaussian quadrature rule for each factor. The total number of states for the markov chain is therefore $N = 121$. We also set the current state to the long-run mean. Calibrating such a markov chain with monthly interval and up to 10 years takes about 14.4 seconds. Table 3 presents the pricing results for a series of five-year caplets. We use the same definition for moneyness (d) as in (12). The conditional variance and higher moments of the short rate over n periods are now given by

$$\begin{aligned}
\sigma_n^2 &= \sum_{j=0}^{n-1} B_1^T \Phi^j V B_1; \\
\kappa_{3n} &= \sum_{j=0}^{n-1} B_1^T \Phi^j V^{3/2} \text{diag}(\gamma_1) B_1; \\
\kappa_{4n} &= \sum_{j=0}^{n-1} B_1^T \Phi^j V^2 \text{diag}(\gamma_2) B_1,
\end{aligned}$$

where γ_1 and γ_2 are the skewness and kurtosis of the innovation vector. In the example, we assume zero for both.

Since now we do not have an analytical form for the caplet price under the multi-factor set-up. The pricing accuracy of the markov chain for this multi-factor model cannot be readily gauged and will be left for future research. Also, in practice, it is still up to debate on, among others, (i) how to estimate the parameters of such a multi-factor model and (ii) whether one should estimate the market price of risk λ or extend the model with a deterministic but time-varying $\delta(t)$ to fit the whole current yield curve. Nevertheless, regardless of what one does with the model, the exercise fully illustrates that the markov chain approach provides a flexible, efficient, and easy-to-apply method in pricing options on underlyings with complex dynamics.

7 Concluding Remarks

Unlike the popular trees used in the financial world, the markov chain approximation proposed in this paper provides a much more natural way of modeling stationary processes. It has been shown to be efficient, accurate, and extremely flexible. Its advantage is all the more pronounced when dealing with complex dynamics such as conditional non-normality, heteroskedasticity and multiple factors, all of which are salient features of interest rates.

References

Backus, David, Silverio Foresi, and Chris Telmer, 1999, "Discrete-time models of bond pricing," manuscript, New York University.

Backus, David, Chris Telmer, and Liuren Wu, 1999, "Design and estimation of affine yield models," manuscript, New York University.

Balduzzi, Pierluigi, Sanjiv Das, Silverio Foresi, and Rangarajan Sundaram, 1996, "A simple approach to three-factor affine term structure models," *Journal of Fixed Income*, 6 (December), 43-53.

Black, F., Derman, E. and W. Toy, 1990, "A one-factor model of interest rates and its application to Treasury bond options," *Financial Analysts Journal*, 46, 33-39.

Black, F. and P. Karasinski, 1991, "Bond option pricing when the short rates are lognormal," *Financial Analysts Journal*, 47(4), 52-59.

Brennan, Michael and Eduardo Schwartz, 1979, "A continuous time approach to the pricing of bonds," *Journal of Banking and Finance*, 3, 133-155.

Chen, Ren-Raw and Louis Scott, 1993, "Maximum likelihood estimation of a multi-factor equilibrium model of the term structure of interest rates," *Journal of Fixed Income*, 3 (December), 14-31.

Cox, John, Jonathan Ingersoll, and Stephen Ross, 1985, "A theory of the term structure of interest rates," *Econometrica*, 53, 385-407.

Dai, Qiang and Kenneth Singleton, 1997, "Specification analysis of affine term structure models," manuscript, Stanford University, February.

Davis, P. J., and P. Rabinowitz, 1984, *Methods of Numerical Integration,* New York, Academic Press.

Duffie, Darrell, and Rui Kan, 1996, "A yield-factor model of interest rates," *Mathematical Finance*, 6, 379-406.

Hogan, M. and K. Weintraub, 1993, "The log-normal interest rate model and Eurodollar futures," Working paper, New York.

Hull, John, and Alan White, 1990, "Pricing interest rate derivative securities," *Review of Financial Studies*, 3(4), 573-592.

Hull, John, and Alan White, 1994, "Numerical procedures for implementing term structure models I: single-factor models," *Journal of Derivatives*, Fall, 7-16.

Hull, John and Alan White, 1996, "Using Hull-White Interest Rate Trees," *Journal of Derivatives,* 3, 26-36.

Judd, Kenneth, 1998, *Numerical Methods in Economics,* MIT Press

Kautsky, J. and S. Elhay, 1982, "Calculation of the weights of Interpolatory quadratures," *Numerical Math*, 40, 407-422.

Longstaff, Francis and Eduardo Schwartz, 1992, "Interest rate volatility and the term structure: A two-factor general equilibrium model," *Journal of Finance*, 47, 1259-1282.

Norton, R.M. and S. Arnold, 1985, "A theorem of moments," *American Statistician*, 39, 106.

Pearson, Neil and Tong-Sheng Sun, 1994, "Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model," *Journal of Finance*, 54, 1279-1304.

Tauchen, George, 1986a, "Finite state markov chain approximation to univariate and vector autoregressions," *Economic Letters*, 20, 177-181.

Tauchen, George, 1986b, "Statistical properties of generalized methods of moments estimates of utility function parameters based on financial market data," *Journal of Business and Economic Statistics,* 4, 397-415.

Tauchen, George and Robert Hussey, 1991, "Quadrature-based methods for obtaining approximate solutions to nonlinear asset pricing models," *Econometrica*, 59(2), 371-396.

Vasicek, O., 1977, "An equilibrium characterization of the term structure," *Journal of Financial Economics*, 5, 177-188.

Table 1
Computational Cost Comparison

N	Markov Chain		Trinomial Tree	
	t_Q	t_M	Δt	t_{HW}
41	0.99	0.72	1/12	2.96
61	2.14	2.80	1/18	7.52
81	3.46	6.42	1/24	15.16
101	5.32	12.20	1/30	26.53
121	7.41	20.93	1/36	43.22
141	10.11	32.90	1/42	63.38
161	13.41	51.80	1/48	91.78

t_Q, t_M, and t_{HW} denote, respectively, the computational time, in seconds, of computing a Guassian quadrature, calibrating a markov chain and interest rate states up to 10 years, and calibrating a trinomial tree and interest rate states up to 10 years, with the maximum number of states equal to N. All numbers are averages of 10 identical runs. Δt denotes the time interval, in fraction of a year, required to achieve the maximum number of states in the trinomial tree. The computation is done on a Pentium II personal computer with 64MB RAM and 450 MHz. Model parameters are from the estimates of the one-month yield of the U.S. Treasury bonds: $\theta = 6.68\%$, $\sigma = 2.21 \times 10^{-3}$, and $\kappa = 0.49$. We assume conditional normality on the interest rate such that $\pi_0 = 1/6$.

Table 2
Caplet Pricing Comparison

Strike	Moneyness	Analytical	Markov Chain	HW Tree
$(K, \%)$	(d)	(C_A)	(C_M)	(C_T)
7.03	−2.00	0.001	0.001	0.001
6.91	−1.50	0.004	0.004	0.004
6.80	−1.00	0.011	0.011	0.011
6.68	−0.50	0.027	0.027	0.026
6.57	0.00	0.054	0.054	0.053
6.46	0.50	0.095	0.095	0.094
6.34	1.00	0.147	0.147	0.147
6.23	1.50	0.208	0.208	0.208
6.11	2.00	0.273	0.273	0.273

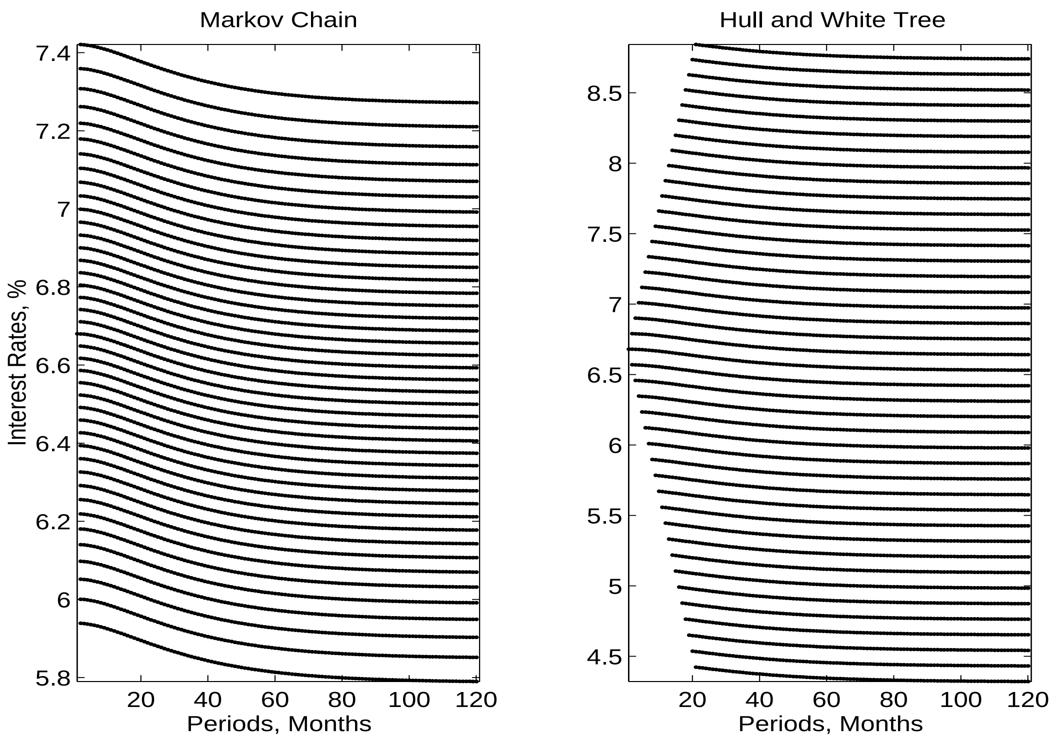
Entries are caplet prices under different strike prices (moneyness). Caplets have a maturity of 5 year and a tenor of a month $\tau = 1/12$. Both the markov chain and the trinomial tree have an interval of a month: $\Delta t = 1/12$. The number of states for the markov chain is the same as the maximum number of states for the trinomial tree $N = 41$. The interest rate process is calibrated to the one-month yield of the U.S. Treasury bonds: $\theta = 6.68\%$, $\sigma = 2.21 \times 10^{-3}$, and $\kappa = 0.49$. We assume that the current interest rate is at its long-run mean: $r(t) = \theta$ and that the market price of risk is zero. The principal for the option is $L = 1,000$.

Table 3
Markov Chain Pricing of Caplets Under Multi-factor Models

Strike	Moneyness	Caplet Price
$(K, \%)$	(d)	$(C_M, \$)$
9.17	-2.00	0.394
9.06	-1.50	0.419
8.96	-1.00	0.445
8.85	-0.50	0.471
8.75	0.00	0.498
8.64	0.50	0.527
8.53	1.00	0.556
8.43	1.50	0.586
8.32	2.00	0.618

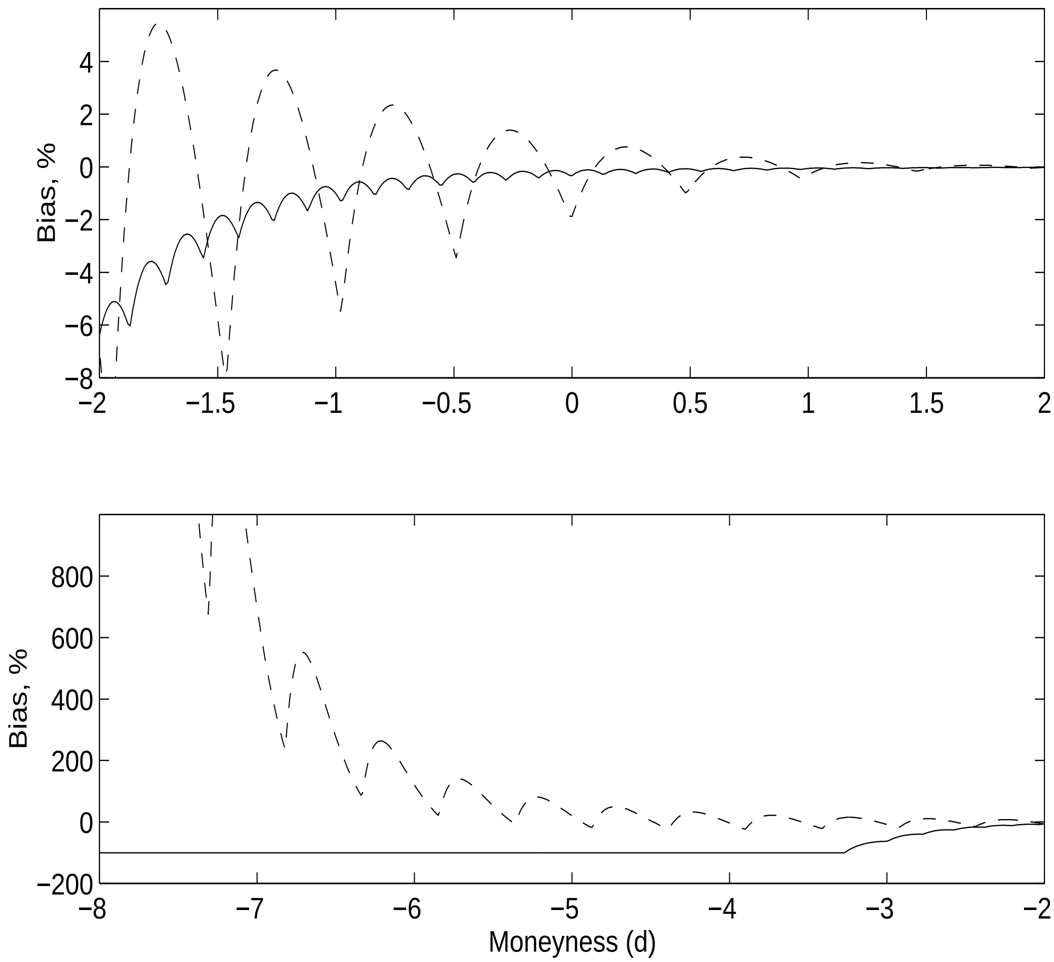
Entries are caplet prices under different strike prices (moneyness). Caplets have a maturity of 5 year and a tenor of a month $\tau = 1/12$. The markov chain is calibrated to a two-factor model described in the Backus, Telmer, and Wu (1999). The number of states used for each factor is 11. We assume that the current state is at its long-run mean: $Z(t) = \theta$. The principal for the option is $L = 1,000$.

Figure 1
Interest Rate Configuration



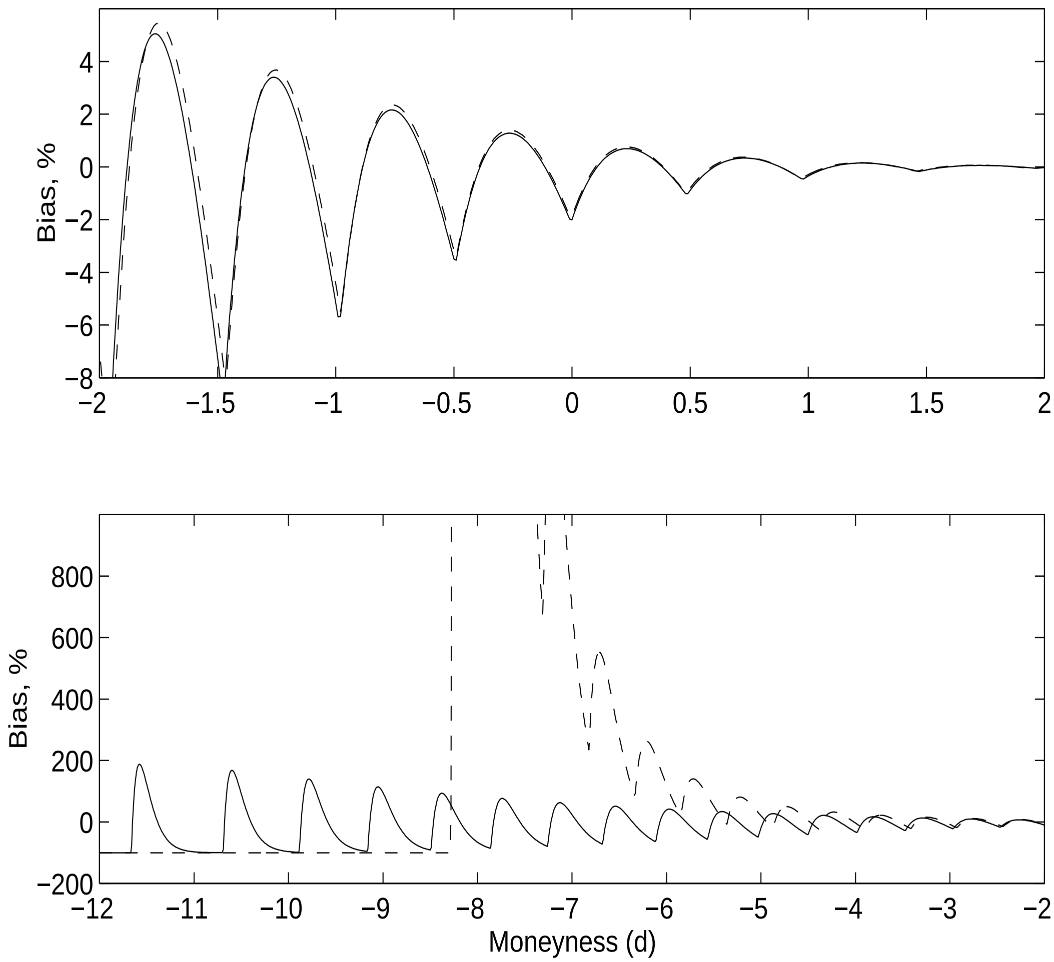
Figures illustrate the development of the one-month interest rate starting at $r = 6.68\%$. Parameters are the same as in Table 1.

Figure 2
Pricing Bias



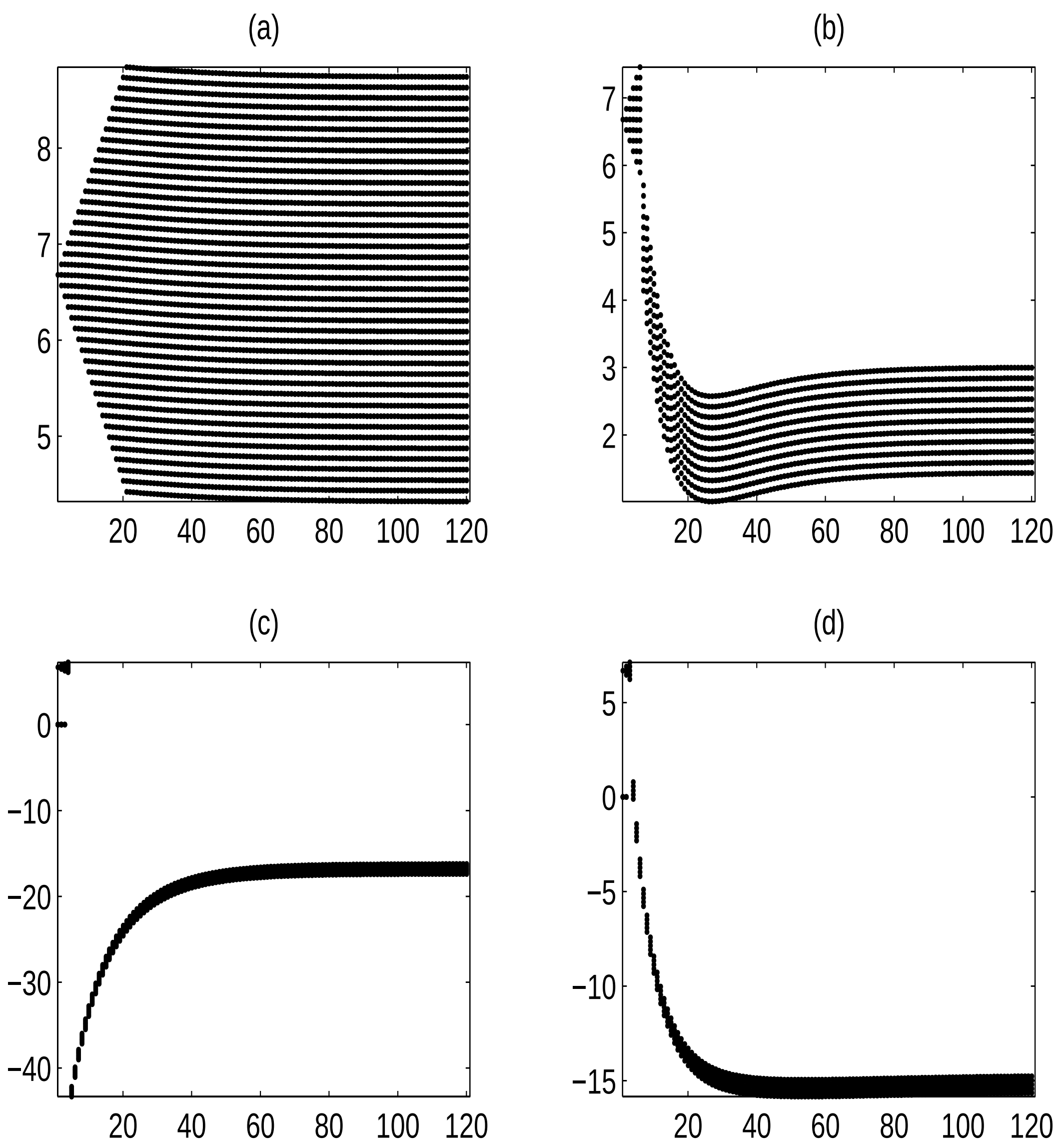
Solid lines illustrate the percentage pricing bias of the markov chain approach while dashed lines denote that of the trinomial tree. Parameters are the same as in Table 1. The bottom graph is an extension of the top graph to far out-of-the-money options.

Figure 3
Pricing Bias



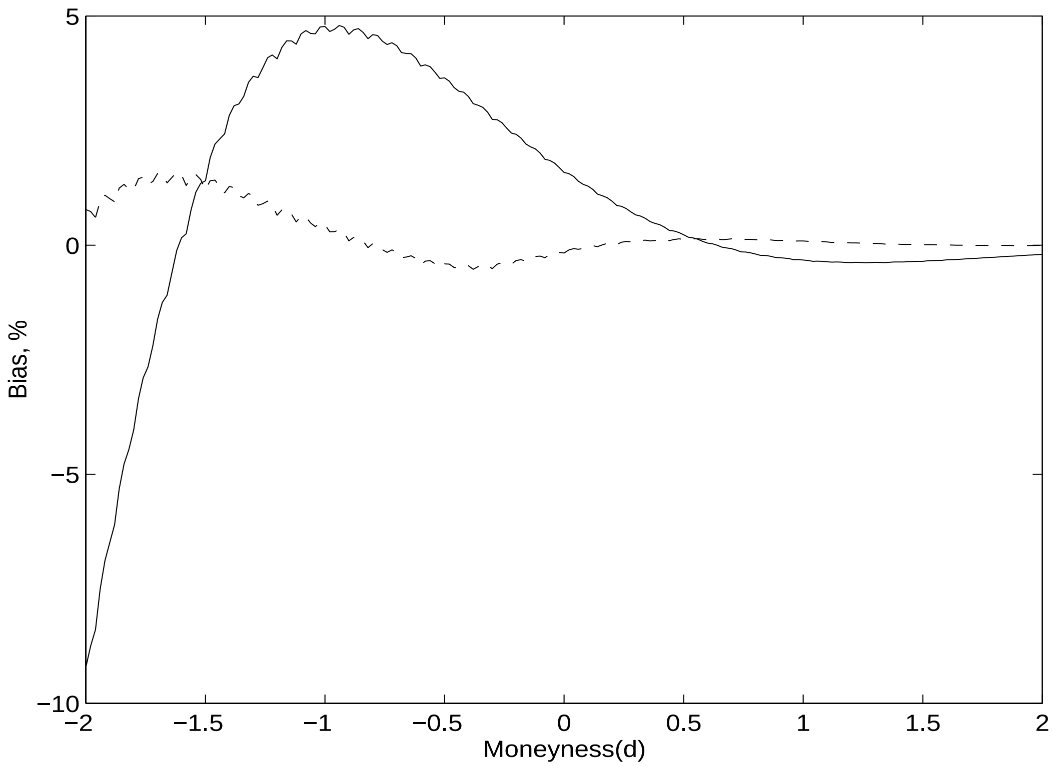
Solid lines illustrate the percentage pricing bias of the markov chain approach while dashed lines denote that of the trinomial tree. The bottom graph is an extension of the top graph to far out-of-the-money options. Parameters are the same as in Table 1, except that the weighting function is changed from the conditional density conditional on the long-run mean to the unconditional density.

Figure 4
The Ill Behavior of The Trinomial Tree With Non-normality



Figures depict the interest rate configuration implied by Hull and White trinomial tress when incorporating different degrees of conditional non-normality. From (a) to (d), the conditional kurtosis γ_2 is set to 0, 3, 6, 9, respectively. Conditional skewness is set to zero. The maximum number of states is 41, 11, 7, 5, respectively. Other parameters are the same as in Table 1.

Figure 5
Option Pricng Bias With Non-Normality



The solid line depicts the percentage bias between the markov chain pricing without considering non-normality and the analytical price with non-normality (in equation (11)):

$$Bias = 100 \times \frac{C_M(normal) - C_A(GC)}{C_A(GC)}.$$

The dashed line depicts the percentage pricing bias between the markov chain pricing and the analytical pricing in (13), both considering non-normality:

$$Bias = 100 \times \frac{C_M(GC) - C_A(GC)}{C_A(GC)}.$$

$N = 141$. Conditional non-normality parameters are: $\gamma_1 = -0.48$ and $\gamma_2 = 9.10$, which are estimated from the AR(1) regression residuals of the one-month yield. Normal density conditional on the long-run mean is used as the weighting function $\omega(z)$. All other parameters are the same as in Table 1.