3.3 Gaussian Short Rate Models

This section and the next develop methods for simulating some simple but important stochastic interest rate models. These models posit the dynamics of an instantaneous continuously compounded short rate $r(t)$. An investment in a money market account earning interest at rate $r(u)$ at time u grows from a value of 1 at time 0 to a value of

$$\beta(t) = \exp\left(\int_0^t r(u)\,du\right)$$

at time t. Though this is now a stochastic quantity, it remains the numeraire for risk-neutral pricing. The price at time 0 of a derivative security that pays X at time T is the expectation of $X/\beta(T)$, i.e.,

$$\mathsf{E}\left[\exp\left(-\int_0^T r(u)\,du\right)X\right], \tag{3.37}$$

the expectation taken with respect to the risk-neutral measure. In particular, the time-0 price of a bond paying 1 at T is given by

$$B(0,T) = \mathsf{E}\left[\exp\left(-\int_0^T r(u)\,du\right)\right]. \tag{3.38}$$

We focus primarily on the dynamics of the short rate under the risk-neutral measure.

The Gaussian models treated in this section offer a high degree of tractability. Many simple instruments can be priced in closed form in these models or using deterministic numerical methods. Some extensions of the basic models and some pricing applications do, however, require simulation for the calculation of expressions of the form (3.37). The tractability of the models offers opportunities for increasing the accuracy of simulation.

3.3.1 Basic Models and Simulation

The classical model of Vasicek [352] describes the short rate through an Ornstein-Uhlenbeck process (cf. Karatzas and Shreve [207], p.358)

$$dr(t) = \alpha(b - r(t))\,dt + \sigma\,dW(t). \tag{3.39}$$

Here, W is a standard Brownian motion and α, b, and σ are positive constants. Notice that the drift in (3.39) is positive if $r(t) < b$ and negative if $r(t) > b$; thus, $r(t)$ is pulled toward level b, a property generally referred to as *mean reversion*. We may interpret b as a long-run interest rate level and α as the speed at which $r(t)$ is pulled toward b. The mean-reverting form of the drift is

an essential feature of the Ornstein-Uhlenbeck process and thus of the Vasicek model.

The continuous-time Ho-Lee model [185] has

$$dr(t) = g(t)\,dt + \sigma\,dW(t) \tag{3.40}$$

with g a deterministic function of time. Both (3.39) and (3.40) define Gaussian processes, meaning that the joint distribution of $r(t_1), \ldots, r(t_n)$ is multivariate normal for any $t_1, \ldots, t_n$. Both define Markov processes and are special cases of the general Gaussian Markov process specified by

$$dr(t) = [g(t) + h(t)r(t)]\,dt + \sigma(t)\,dW(t), \tag{3.41}$$

with g, h, and σ all deterministic functions of time. Natural extensions of (3.39) and (3.40) thus allow σ, b, and α to vary with time. Modeling with the Vasicek model when b in particular is time-varying is discussed in Hull and White [190].

The SDE (3.41) has solution

$$r(t) = e^{H(t)}r(0) + \int_0^t e^{H(t)-H(s)}g(s)\,ds + \int_0^t e^{H(t)-H(s)}\sigma(s)\,dW(s),$$

with

$$H(t) = \int_0^t h(s)\,ds,$$

as can be verified through an application of Itô's formula. Because this produces a Gaussian process, simulation of $r(t_1), \ldots, r(t_n)$ is a special case of the general problem of sampling from a multivariate normal distribution, treated in Section 2.3. But it is a sufficiently interesting special case to merit consideration. To balance tractability with generality, we will focus on the Vasicek model (3.39) with time-varying b and on the Ho-Lee model (3.40). Similar ideas apply to the general case (3.41).

Simulation

For the Vasicek model with time-varying b, the general solution above specializes to

$$r(t) = e^{-\alpha t}r(0) + \alpha\int_0^t e^{-\alpha(t-s)}b(s)\,ds + \sigma\int_0^t e^{-\alpha(t-s)}\,dW(s). \tag{3.42}$$

Similarly, for any $0 < u < t$,

$$r(t) = e^{-\alpha(t-u)}r(u) + \alpha\int_u^t e^{-\alpha(t-s)}b(s)\,ds + \sigma\int_u^t e^{-\alpha(t-s)}\,dW(s).$$

From this it follows that, given $r(u)$, the value $r(t)$ is normally distributed with mean

$$e^{-\alpha(t-u)}r(u) + \mu(u,t), \quad \mu(u,t) \equiv \alpha \int_u^t e^{-\alpha(t-s)} b(s)\, ds \tag{3.43}$$

and variance

$$\sigma_r^2(u,t) \equiv \sigma^2 \int_u^t e^{-2\alpha(t-s)}\, ds = \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha(t-u)}\right). \tag{3.44}$$

To simulate r at times $0 = t_0 < t_1 < \cdots < t_n$, we may therefore set

$$r(t_{i+1}) = e^{-\alpha(t_{i+1}-t_i)} r(t_i) + \mu(t_i, t_{i+1}) + \sigma_r(t_i, t_{i+1}) Z_{i+1}, \tag{3.45}$$

with $Z_1, \ldots, Z_n$ independent draws from $N(0,1)$.

This algorithm is an exact simulation in the sense that the distribution of the $r(t_1), \ldots, r(t_n)$ it produces is precisely that of the Vasicek process at times $t_1, \ldots, t_n$ for the same value of $r(0)$. In contrast, the slightly simpler Euler scheme

$$r(t_{i+1}) = r(t_i) + \alpha(b(t_i) - r(t_i))(t_{i+1} - t_i) + \sigma\sqrt{t_{i+1} - t_i}\, Z_{i+1}$$

entails some discretization error. Exact simulation of the Ho-Lee process (3.40) is a special case of the method in (3.4) for simulating a Brownian motion with time-varying drift.

In the special case that $b(t) \equiv b$, the algorithm in (3.45) simplifies to

$$r(t_{i+1}) = e^{-\alpha(t_{i+1}-t_i)} r(t_i) + b(1 - e^{-\alpha(t_{i+1}-t_i)}) + \sigma\sqrt{\frac{1}{2\alpha}\left(1 - e^{-2\alpha(t_{i+1}-t_i)}\right)} Z_{i+1}. \tag{3.46}$$

The Euler scheme is then equivalent to making the approximation $e^x \approx 1 + x$ for the exponentials in this recursion.

Evaluation of the integral defining $\mu(t_i, t_{i+1})$ and required in (3.45) may seem burdensome. The effort involved in evaluating this integral clearly depends on the form of the function $b(t)$ so it is worth discussing how this function is likely to be specified in practice. Typically, the flexibility to make b vary with time is used to make the dynamics of the short rate consistent with an observed term structure of bond prices. The same is true of the function g in the Ho-Lee model (3.40). We return to this point in Section 3.3.2, where we discuss bond prices in Gaussian models.

Stationary Version

Suppose $b(t) \equiv b$ and $\alpha > 0$. Then from (3.43) we see that

$$\mathsf{E}[r(t)] = e^{-\alpha t} r(0) + (1 - e^{-\alpha t}) b \to b \quad \text{as } t \to \infty,$$

so the process $r(t)$ has a limiting mean. It also has a limiting variance given (via (3.44)) by

$$\lim_{t\to\infty} \mathsf{Var}[r(t)] = \lim_{t\to\infty} \frac{\sigma^2}{2\alpha}\left(1 - e^{-2\alpha t}\right) = \frac{\sigma^2}{2\alpha}.$$

In fact, $r(t)$ converges in distribution to a normal distribution with this mean and variance, in the sense that for any $x \in \Re$

$$P(r(t) \le x) \to \Phi\left(\frac{x - b}{\sigma/\sqrt{2\alpha}}\right),$$

with Φ the standard normal distribution. The fact that $r(t)$ has a limiting distribution is a reflection of the stabilizing effect of mean reversion in the drift and contrasts with the long-run behavior of, for example, geometric Brownian motion.

The limiting distribution of $r(t)$ is also a stationary distribution in the sense that if $r(0)$ is given this distribution then every $r(t)$, $t > 0$, has this distribution as well. Because (3.46) provides an exact discretization of the process, the $N(b, \sigma^2/2\alpha)$ distribution is also stationary for the discretized process. To simulate a stationary version of the process, it therefore suffices to draw $r(0)$ from this normal distribution and then proceed as in (3.46).

3.3.2 Bond Prices

As already noted, time-dependent drift parameters are typically used to make a short rate model consistent with an observed set of bond prices. Implementation of the simulation algorithm (3.45) is thus linked to the *calibration* of the model through the choice of the function $b(t)$. The same applies to the function $g(t)$ in the Ho-Lee model and as this case is slightly simpler we consider it first.

Our starting point is the bond-pricing formula (3.38). The integral of $r(u)$ from 0 to T appearing in that formula is normally distributed because $r(u)$ is a Gaussian process. It follows that the bond price is the expectation of the exponential of a normal random variable. For a normal random variable $X \sim N(m, v^2)$, we have $\mathsf{E}[\exp(X)] = \exp(m + (v^2/2))$, so

$$\mathsf{E}\left[\exp\left(-\int_0^T r(t)\, dt\right)\right] = \exp\left(-\mathsf{E}\left[\int_0^T r(t)\, dt\right] + \tfrac{1}{2}\mathsf{Var}\left[\int_0^T r(t)\, dt\right]\right). \tag{3.47}$$

To find the price of the bond we therefore need to find the mean and variance of the integral of the short rate.

In the Ho-Lee model, the short rate is given by

$$r(t) = r(0) + \int_0^t g(s)\, ds + \sigma W(t)$$

and its integral by

$$\int_0^T r(u)\,du = r(0)T + \int_0^T \int_0^u g(s)\,ds\,du + \sigma \int_0^T W(u)\,du.$$

This integral has mean

$$r(0)T + \int_0^T \int_0^u g(s)\,ds\,du$$

and variance

$$\begin{aligned}\mathsf{Var}\left[\sigma \int_0^T W(u)\,du\right] &= 2\sigma^2 \int_0^T \int_0^t \mathsf{Cov}[W(u), W(t)]\,du\,dt \\ &= 2\sigma^2 \int_0^T \int_0^t u\,du\,dt \\ &= \frac{1}{3}\sigma^2 T^3. \end{aligned} \tag{3.48}$$

Substituting these expressions in (3.47), we get

$$\begin{aligned} B(0,T) &= \mathsf{E}\left[\exp\left(-\int_0^T r(u)\,du\right)\right] \\ &= \exp\left(-r(0)T - \int_0^T \int_0^u g(s)\,ds\,du + \frac{\sigma^2 T^3}{6}\right). \end{aligned}$$

If we are given a set of bond prices $B(0,T)$ at time 0, our objective is to choose the function g so that this equation holds.

To carry this out we can write

$$B(0,T) = \exp\left(-\int_0^T f(0,t)\,dt\right),$$

with $f(0,t)$ the instantaneous forward rate for time t as of time 0 (cf. Appendix C). The initial forward curve $f(0,T)$ captures the same information as the initial bond prices. Equating the two expressions for $B(0,T)$ and taking logarithms, we find that

$$r(0)T + \int_0^T \int_0^u g(s)\,ds\,du - \frac{\sigma^2 T^3}{6} = \int_0^T f(0,t)\,dt.$$

Differentiating twice with respect to the maturity argument T, we find that

$$g(t) = \frac{\partial}{\partial T} f(0,T)\bigg|_{T=t} + \sigma^2 t. \tag{3.49}$$

Thus, bond prices produced by the Ho-Lee model will match a given set of bond prices $B(0,T)$ if the function g is tied to the initial forward curve $f(0,T)$ in this way; i.e., if we specify

$$dr(t) = \left(\frac{\partial}{\partial T} f(0,T)\bigg|_{T=t} + \sigma^2 t\right) dt + \sigma\,dW(t). \tag{3.50}$$

A generic simulation of the Ho-Lee model with drift function g can be written as

$$r(t_{i+1}) = r(t_i) + \int_{t_i}^{t_{i+1}} g(s)\,ds + \sigma\sqrt{t_{i+1} - t_i}\,Z_{i+1},$$

with $Z_1, Z_2, \ldots$ independent $N(0,1)$ random variables. With g chosen as in (3.49), this simplifies to

$$r(t_{i+1}) = r(t_i) + [f(0,t_{i+1}) - f(0,t_i)] + \frac{\sigma^2}{2}[t_{i+1}^2 - t_i^2] + \sigma\sqrt{t_{i+1} - t_i}\,Z_{i+1}.$$

Thus, no integration of the drift function g is necessary; to put it another way, whatever integration is necessary must already have been dealt with in choosing the forward curve $f(0,t)$ to match a set of bond prices.

The situation is even simpler if we require that our simulated short rate be consistent only with bonds maturing at the simulation times $t_1, \ldots, t_n$. To satisfy this requirement we can weaken (3.49) to the condition that

$$\int_{t_i}^{t_{i+1}} g(s)\,ds = f(0,t_{i+1}) - f(0,t_i) + \frac{\sigma^2}{2}[t_{i+1}^2 - t_i^2].$$

Except for this constraint, the choice of g is immaterial — we could take it to be continuous and piecewise linear, for example. In fact, we never even need to specify g because only its integral over the intervals (t_i, t_{i+1}) influence the values of r on the time grid $t_1, \ldots, t_n$.

Bonds in the Vasicek Model

A similar if less explicit solution applies to the Vasicek model. The integral of the short rate is again normally distributed; we need to find the mean and variance of this integral to find the price of a bond using (3.47). Using (3.42), for the mean we get

$$\begin{aligned}\mathsf{E}\left[\int_0^T r(t)\,dt\right] &= \int_0^T \mathsf{E}[r(t)]\,dt \\ &= \frac{1}{\alpha}(1 - e^{-\alpha T})r(0) + \alpha \int_0^T \int_0^t e^{-\alpha(t-s)} b(s)\,ds\,dt. \end{aligned} \tag{3.51}$$

For the variance we have

$$\mathsf{Var}\left[\int_0^T r(t)\,dt\right] = 2\int_0^T \int_0^t \mathsf{Cov}[r(t), r(u)]\,du\,dt. \tag{3.52}$$

From (3.42) we get, for $u \le t$,

$$\begin{aligned}\mathsf{Cov}[r(t), r(u)] &= \sigma^2 \int_0^u e^{-\alpha(t-s)} e^{-\alpha(u-s)}\, ds \\ &= \frac{\sigma^2}{2\alpha}\left(e^{\alpha(u-t)} - e^{-\alpha(u+t)}\right). \end{aligned} \tag{3.53}$$

Integrating two more times as required for (3.52) gives

$$\mathsf{Var}\left[\int_0^T r(t)\, dt\right] = \frac{\sigma^2}{\alpha^2}\left[T + \frac{1}{2\alpha}\left(1 - e^{-2\alpha T}\right) + \frac{2}{\alpha}\left(e^{-\alpha T} - 1\right)\right]. \tag{3.54}$$

By combining (3.51) and (3.54) as in (3.47), we arrive at an expression for the bond price $B(0,T)$.

Observe that (3.54) does not depend on $r(0)$ and (3.51) is a linear transformation of $r(0)$. If we set

$$A(t,T) = \frac{1}{\alpha}\left(1 - e^{-\alpha(T-t)}\right)$$

and

$$\begin{aligned} C(t,T) = &-\alpha \int_t^T \int_t^u e^{-\alpha(u-s)} b(s)\, ds\, du \\ &+ \frac{\sigma^2}{2\alpha^2}\left[(T-t) + \frac{1}{2\alpha}\left(1 - e^{-2\alpha(T-t)}\right) + \frac{2}{\alpha}\left(e^{-\alpha(T-t)} - 1\right)\right], \end{aligned}$$

then substituting (3.51) and (3.54) in (3.47) produces

$$B(0,T) = \exp(-A(0,T)r(0) + C(0,T)).$$

In fact, the same calculations show that

$$B(t,T) = \exp(-A(t,T)r(t) + C(t,T)). \tag{3.55}$$

In particular, $\log B(t,T)$ is a linear transformation of $r(t)$. This feature has been generalized by Brown and Schaefer [71] and Duffie and Kan [101] to what is generally referred to as the *affine class* of interest rate models.

As in our discussion of the Ho-Lee model, the function $b(s)$ can be chosen to match a set of prices $B(0,T)$ indexed by T. If we are concerned only with matching a finite set of bond prices $B(0,t_1), \ldots, B(0,t_n)$, then only the values of the integrals

$$\int_{t_i}^{t_{i+1}} e^{-\alpha(t_{i+1}-s)} b(s)\, ds$$

need to be specified. These are precisely the terms $\mu(t_i, t_{i+1})$ needed in the simulation algorithm (3.45). Thus, these integrals are by-products of fitting the model to a term structure and not additional computations required solely for the simulation.

Joint Simulation with the Discount Factor

Most applications that call for simulation of a short rate process $r(t)$ also require values of the discount factor

$$\frac{1}{\beta(t)} = \exp\left(-\int_0^t r(u)\, du\right)$$

or, equivalently, of

$$Y(t) = \int_0^t r(u)\, du.$$

Given values $r(0), r(t_1), \ldots, r(t_n)$ of the short rate, one can of course generate approximate values of $Y(t_i)$ using

$$\sum_{j=1}^{i} r(t_{j-1})[t_j - t_{j-1}], \quad t_0 = 0,$$

or some other approximation to the time integral. But in a Gaussian model, the pair $(r(t), Y(t))$ are jointly Gaussian and it is often possible to simulate paths of the pair without discretization error. To carry this out we simply need to find the means, variances, and covariance of the increments of $r(t)$ and $Y(t)$.

We have already determined (see (3.45)) that, given $r(t_i)$,

$$r(t_{i+1}) \sim N\left(e^{-\alpha(t_{i+1}-t_i)} r(t_i) + \mu(t_i, t_{i+1}), \sigma_r^2(t_i, t_{i+1})\right).$$

From the same calculations used in (3.51) and (3.54), we find that, given $r(t_i)$ and $Y(t_i)$,

$$Y(t_{i+1}) \sim N(Y(t_i) + \mu_Y(t_i, t_{i+1}), \sigma_Y^2(t_i, t_{i+1})),$$

with

$$\mu_Y(t_i, t_{i+1}) = \frac{1}{\alpha}\left(1 - e^{-\alpha(t_{i+1}-t_i)}\right) r(t_i) + \alpha \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} e^{-\alpha(u-s)} b(s)\, ds\, du$$

and

$$\begin{aligned} &\sigma_Y^2(t_i, t_{i+1}) = \\ &\frac{\sigma^2}{\alpha^2}\left((t_{i+1} - t_i) + \frac{1}{2\alpha}\left(1 - e^{-2\alpha(t_{i+1}-t_i)}\right) + \frac{2}{\alpha}\left(e^{-\alpha(t_{i+1}-t_i)} - 1\right)\right). \end{aligned}$$

It only remains to determine the conditional covariance between $r(t_{i+1})$ and $Y(t_{i+1})$ given $(r(t_i), Y(t_i))$. For this we proceed as follows:

$$\begin{aligned}\mathsf{Cov}\left[r(t), Y(t)\right] &= \int_0^t \mathsf{Cov}[r(t), r(u)]\, du \\ &= \frac{\sigma^2}{2\alpha}\int_0^t e^{\alpha(u-t)} - e^{-\alpha(u+t)}\, du \\ &= \frac{\sigma^2}{2\alpha^2}\left[1 + e^{-2\alpha t} - 2e^{-\alpha t}\right].\end{aligned}$$

The required covariance is thus given by

$$\sigma_{rY}(t_i, t_{i+1}) = \frac{\sigma^2}{2\alpha}\left[1 + e^{-2\alpha(t_{i+1}-t_i)} - 2e^{-\alpha(t_{i+1}-t_i)}\right].$$

The corresponding correlation is

$$\rho_{rY}(t_i, t_{i+1}) = \frac{\sigma_{rY}(t_i, t_{i+1})}{\sigma_r(t_i, t_{i+1})\sigma_Y(t_i, t_{i+1})}.$$

With this notation, the pair (r, Y) can be simulated at times $t_1, \ldots, t_n$ without discretization error using the following algorithm:

$$\begin{aligned}r(t_{i+1}) &= e^{-\alpha(t_{i+1}-t_i)}r(t_i) + \mu(t_i, t_{i+1}) + \sigma_r(t_i, t_{i+1})Z_1(i+1) \\ Y(t_{i+1}) &= Y(t_i) + \mu_Y(t_i, t_{i+1}) + \sigma_Y(t_i, t_{i+1})[\rho_{rY}(t_i, t_{i+1})Z_1(i+1) \\ &\quad + \sqrt{1 - \rho_{rY}^2(t_i, t_{i+1})}Z_2(i+1)],\end{aligned}$$

where $(Z_1(i), Z_2(i))$, $i = 1, \ldots, n$, are independent standard bivariate normal random vectors.

Change of Numeraire

Thus far, we have considered the dynamics of the short rate $r(t)$ only under the risk-neutral measure. Recall that the numeraire asset associated with the risk-neutral measure is $\beta(t) = \exp(\int_0^t r(u)\, du)$ and the defining feature of this probability measure is that it makes the discounted bond prices $B(t,T)/\beta(t)$ martingales. In fact, the dynamics of the bond prices under the Gaussian models we have considered are of the form (for fixed T)

$$\frac{dB(t,T)}{B(t,T)} = r(t)\, dt - A(t,T)\sigma\, dW(t) \tag{3.56}$$

with $A(t,T)$ deterministic; this follows from (3.55). The solution of this equation is

$$B(t,T) = B(0,T)\exp\left(\int_0^t [r(u) - \tfrac{1}{2}\sigma^2 A^2(u,T)]\, du - \sigma\int_0^t A(u,T)\, dW(u)\right),$$

from which it is evident that

$$\frac{B(t,T)}{\beta(t)} = B(0,T)\exp\left(-\tfrac{1}{2}\sigma^2\int_0^t A^2(u,T)\, du - \sigma\int_0^t A(u,T)\, dW(u)\right) \tag{3.57}$$

is an exponential martingale.

As discussed in Section 1.2.3, the *forward measure* for any date T_F is the measure associated with taking the T_F-maturity bond $B(t,T_F)$ as numeraire asset. The defining feature of the forward measure is that it makes the ratios $B(t,T)/B(t,T_F)$ martingales for $T < T_F$. It is defined by the likelihood ratio process

$$\left(\frac{dP_{T_F}}{dP_\beta}\right)_t = \frac{B(t,T_F)\beta(0)}{\beta(t)B(0,T_F)},$$

and this is given in (3.57) up to a factor of $1/B(0,T_F)$. From Girsanov's Theorem, it follows that the process W^{T_F} defined by

$$dW^{T_F}(t) = dW(t) + \sigma A(t,T_F)\, dt$$

is a standard Brownian motion under P_{T_F}. Accordingly, the dynamics of the Vasicek model become

$$\begin{aligned}dr(t) &= \alpha(b(t) - r(t))\, dt + \sigma\, dW(t) \\ &= \alpha(b(t) - r(t))\, dt + \sigma\left(dW^{T_F}(t) - \sigma A(t,T_F)\, dt\right) \\ &= \alpha(b(t) - \sigma^2 A(t,T_F) - r(t))\, dt + \sigma\, dW^{T_F}(t).\end{aligned} \tag{3.58}$$

Thus, under the forward measure, the short rate process remains a Vasicek process but the reversion level $b(t)$ becomes $b(t) - \sigma^2 A(t,T_F)$.

The process in (3.58) can be simulated using (3.45) with $b(t)$ replaced by $b(t) - \sigma^2 A(t,T_F)$. In particular, we simulate W^{T_F} the way we would simulate any other standard Brownian motion. The simulation algorithm does not "know" that it is simulating a Brownian motion under the forward measure rather than under the risk-neutral measure.

Suppose we want to price a derivative security making a payoff of $g(r(T_F))$ at time T_F. Under the risk-neutral measure, we would price the security by computing

$$\mathsf{E}\left[e^{-\int_0^{T_F} r(u)\, du} g(r(T_F))\right].$$

In fact, g could be a function of the path of $r(t)$ rather than just its terminal value. Switching to the forward measure, this becomes

$$\begin{aligned}&\mathsf{E}_{T_F}\left[e^{-\int_0^{T_F} r(u)\, du} g(r(T_F))\left(\frac{dP_\beta}{dP_{T_F}}\right)_{T_F}\right] \\ &= \mathsf{E}_{T_F}\left[e^{-\int_0^{T_F} r(u)\, du} g(r(T_F))\left(\frac{\beta(T_F)B(0,T_F)}{B(T_F,T_F)\beta(0)}\right)\right] \\ &= B(0,T_F)\mathsf{E}_{T_F}\left[g(r(T_F))\right],\end{aligned}$$

where E_{T_F} denotes expectation under the forward measure. Thus, we may price the derivative security by simulating $r(t)$ under the forward measure P_{T_F}, estimating the expectation of $g(r(T_F))$ and multiplying by $B(0,T_F)$. Notice that discounting in this case is deterministic — we do not need to simulate a discount factor. This apparent simplification results from inclusion of the additional term $-\sigma^2 A(t,T_F)$ in the drift of $r(t)$.

A consequence of working under the forward measure is that the simulation prices the bond maturing at T_F exactly: pricing this bond corresponds to taking $g(r(T_F)) \equiv 1$. Again, this apparent simplification is really a consequence of the form of the drift of $r(t)$ under the forward measure.

3.3.3 Multifactor Models

A general class of Gaussian Markov processes in $\Re^d$ have the form

$$dX(t) = C(b - X(t))\,dt + D\,dW(t) \tag{3.59}$$

where C and D are $d \times d$ matrices, b and $X(t)$ are in $\Re^d$, W is a standard d-dimensional Brownian motion, and $X(0)$ is Gaussian or constant. Such a process remains Gaussian and Markovian if the coefficients C, b, and D are made time-varying but deterministic. The solution of (3.59) is

$$X(t) = e^{-Ct}X(0) + \int_0^t e^{-C(t-s)}b\,ds + \int_0^t e^{-C(t-s)}D\,dW(s),$$

from which it is possible to define an exact time-discretization similar to (3.45).

A model of the short rate process can be specified by setting $r(t) = a^\top X(t)$ with $a \in \Re^d$ (or with a deterministically time-varying). The elements of $X(t)$ are then interpreted as "factors" driving the evolution of the short rate. Because each $X(t)$ is normally distributed, $r(t)$ is normally distributed. However, $r(t)$ is not in general a Markov process: to make the future evolution of r independent of the past, we need to condition on the full state information $X(t)$ and not merely $r(t)$.

Recall from (3.55) that in the Vasicek model (with constant or time-varying coefficients), bond prices are exponentials of affine functions of the short rate. A similar representation applies if the short rate has the form $r(t) = a^\top X(t)$ and $X(t)$ is as in (3.59); in particular, we have

$$B(t,T) = \exp(-A(t,T)^\top X(t) + C(t,T))$$

for some $\Re^d$-valued function $A(t,T)$ and some scalar function $C(t,T)$. In the single-factor setting, differentiating (3.55) and then simplifying leads to

$$\frac{dB(t,T)}{B(t,T)} = r(t)\,dt - A(t,T)\sigma\,dW(t),$$

with σ the diffusion parameter of $r(t)$. The instantaneous correlation between the returns on bonds with maturities T_1 and T_2 is therefore

$$\frac{A(t,T_1)\sigma \cdot A(t,T_2)\sigma}{\sqrt{A^2(t,T_1)\sigma^2}\sqrt{A^2(t,T_2)\sigma^2}} = 1.$$

In other words, all bonds are instantaneously perfectly correlated. In the multifactor setting, the bond price dynamics are given by

$$\frac{dB(t,T)}{B(t,T)} = r(t)\,dt - A(t,T)^\top D\,dW(t).$$

The instantaneous correlation for maturities T_1 and T_2 is

$$\frac{A(t,T_1)^\top DD^\top A(t,T_2)}{\|A(t,T_1)^\top D\|\|A(t,T_2)^\top D\|},$$

which can certainly take values other than 1. The flexibility to capture less than perfect instantaneous correlation between bond returns is the primary motivation for considering multifactor models.

Returning to the general formulation in (3.59), suppose that C can be diagonalized in the sense that $VCV^{-1} = \Lambda$ for some matrix V and diagonal matrix Λ with diagonal entries $\lambda_1, \ldots, \lambda_d$. Suppose further that C is nonsingular and define $Y(t) = VX(t)$. Then

$$\begin{aligned}
dY(t) &= V\,dX(t) \\
&= V[C(b - X(t))\,dt + D\,dW(t)] \\
&= (VCb - \Lambda Y(t))\,dt + VD\,dW(t) \\
&= \Lambda(\Lambda^{-1}VCb - Y(t))\,dt + VD\,dW(t) \\
&= \Lambda(Vb - Y(t))\,dt + VD\,dW(t) \\
&\equiv \Lambda(\tilde{b} - Y(t))\,dt + d\tilde{W}(t)
\end{aligned}$$

with $\tilde{W}$ a $\mathrm{BM}(0,\Sigma)$ process, $\Sigma = VDD^\top V^\top$. It follows that the components of $(Y_1, \ldots, Y_d)$ satisfy

$$dY_j(t) = \lambda_j(\tilde{b}_j - Y_j(t))\,dt + d\tilde{W}_j(t), \quad j = 1, \ldots, d. \tag{3.60}$$

In particular, each Y_j is itself a Markov process. The Y_j remain coupled, however, through the correlation across the components of $\tilde{W}$. They can be simulated as in (3.46) by setting

$$Y_j(t_{i+1}) = e^{\lambda_j(t_{i+1}-t_i)}Y_j(t_i) + (e^{\lambda_j(t_{i+1}-t_i)} - 1)\tilde{b}_j + \sqrt{\frac{1}{2\lambda_j}\left(1 - e^{-2\lambda_j(t_{i+1}-t_i)}\right)}\xi_j(i+1),$$

where $\xi(1), \xi(2), \ldots$ are independent $N(0,\Sigma)$ random vectors, $\xi(i) = (\xi_1(i), \ldots, \xi_d(i))$. Thus, when C is nonsingular and diagonalizable, simulation of (3.59) can be reduced to a system of scalar simulations.

As noted by Andersen and Andreasen [14], a similar reduction is possible even if C is not diagonalizable, but at the expense of making all coefficients time-dependent. If $V(t)$ is a deterministic $d \times d$ matrix-valued function of time and we set $Y(t) = V(t)X(t)$, then

$$\begin{aligned} dY(t) &= \dot{V}(t)X(t)\,dt + V(t)dX(t) \\ &= [\dot{V}(t)X(t) + V(t)C(b - X(t))]\,dt + V(t)D\,dW(t), \end{aligned}$$

where $\dot{V}(t)$ denotes the time derivative of $V(t)$. If we choose $V(t) = \exp([C - I]t)$, then

$$\dot{V}(t) = V(t)C - V(t)$$

and thus

$$\begin{aligned} dY(t) &= [V(t)Cb - V(t)X(t)]\,dt + V(t)D\,dW(t) \\ &= (\tilde{b}(t) - Y(t))\,dt + \tilde{D}(t)\,dW(t), \end{aligned} \tag{3.61}$$

with $\tilde{b}(t) = V(t)Cb$ and $\tilde{D}(t) = V(t)D$. Notice that the drift of each component $Y_i(t)$ depends only on that $Y_i(t)$. This transformation therefore decouples the drifts of the components of the state vector, making each Y_i a Markov process, though the components remain linked through the diffusion term. We can recover the original state vector by setting $X(t) = V(t)^{-1}Y(t)$ because $V(t)$ is always invertible. The seemingly special form of the dynamics in (3.61) is thus no less general than the dynamics in (3.59) with time-varying coefficients.

3.4 Square-Root Diffusions

Feller [118] studied a class of processes that includes the square-root diffusion

$$dr(t) = \alpha(b - r(t))\,dt + \sigma\sqrt{r(t)}\,dW(t), \tag{3.62}$$

with W a standard one-dimensional Brownian motion. We consider the case in which α and b are positive. If $r(0) > 0$, then $r(t)$ will never be negative; if $2\alpha b \geq \sigma^2$, then $r(t)$ remains strictly positive for all t, almost surely.

This process was proposed by Cox, Ingersoll, and Ross [91] as a model of the short rate, generally referred to as the CIR model. They developed a general equilibrium framework in which if the change in production opportunities is assumed to follow a process of this form, then the short rate does as well. As with the Vasicek model, the form of the drift in (3.62) suggests that $r(t)$ is pulled towards b at a speed controlled by α. In contrast to the Vasicek model, in the CIR model the diffusion term $\sigma\sqrt{r(t)}$ decreases to zero as $r(t)$ approaches the origin and this prevents $r(t)$ from taking negative values. This feature of (3.62) is attractive in modeling interest rates.

All of the coefficients in (3.62) could in principle be made time-dependent. In practice, it can be particularly useful to replace the constant b with a function of time and thus consider

$$dr(t) = \alpha(b(t) - r(t))\,dt + \sigma\sqrt{r(t)}\,dW(t). \tag{3.63}$$

As with the Vasicek model, this extension is frequently used to make the bond price function

$$T \mapsto \mathsf{E}\left[\exp\left(-\int_0^T r(u)\,du\right)\right]$$

match a set of observed bond prices $B(0,T)$.

Although we stress the application of (3.63) to interest rate modeling, it should be noted that this process has other financial applications. For example, Heston [179] proposed a stochastic volatility model in which the price of an asset $S(t)$ is governed by

$$\frac{dS(t)}{S(t)} = \mu\,dt + \sqrt{V(t)}\,dW_1(t) \tag{3.64}$$

$$dV(t) = \alpha(b - V(t))\,dt + \sigma\sqrt{V(t)}\,dW_2(t), \tag{3.65}$$

where (W_1, W_2) is a two-dimensional Brownian motion. Thus, in Heston's model, the squared volatility $V(t)$ follows a square-root diffusion. In addition, the process in (3.63) is sometimes used to model a stochastic intensity for a jump process in, for example, modeling default.

A simple Euler discretization of (3.62) suggests simulating $r(t)$ at times $t_1, \ldots, t_n$ by setting

$$r(t_{i+1}) = r(t_i) + \alpha(b - r(t_i))[t_{i+1} - t_i] + \sigma\sqrt{r(t_i)^+}\sqrt{t_{i+1} - t_i}\,Z_{i+1}, \tag{3.66}$$

with $Z_1, \ldots, Z_n$ independent $N(0,1)$ random variables. Notice that we have taken the positive part of $r(t_i)$ inside the square root; some modification of this form is necessary because the values of $r(t_i)$ produced by Euler discretization may become negative. We will see, however, that this issue can be avoided (along with any other discretization error) by sampling from the exact transition law of the process.

3.4.1 Transition Density

The SDE (3.62) is not explicitly solvable the way those considered in Sections 3.2 and 3.3 are; nevertheless, the transition density for the process is known. Based on results of Feller [118], Cox et al. [91] noted that the distribution of $r(t)$ given $r(u)$ for some $u < t$ is, up to a scale factor, a noncentral chi-square distribution. This property can be used to simulate the process (3.62). We follow the approach suggested by Scott [324].

A noncentral chi-square random variable $\chi'^2_\nu(\lambda)$ with ν degrees of freedom and noncentrality parameter λ has distribution

$$P(\chi'^2_\nu(\lambda) \le y) = F_{\chi'^2_\nu(\lambda)}(y)$$
$$\equiv e^{-\lambda/2} \sum_{j=0}^{\infty} \frac{(\frac{1}{2}\lambda)^j/j!}{2^{(\nu/2)+j}\Gamma(\frac{\nu}{2}+j)} \int_0^y z^{(\nu/2)+j-1} e^{-z/2}\, dz, \quad (3.67)$$

for $y > 0$. The transition law of $r(t)$ in (3.62) can be expressed as

$$r(t) = \frac{\sigma^2(1-e^{-\alpha(t-u)})}{4\alpha} \chi'^2_d \left(\frac{4\alpha e^{-\alpha(t-u)}}{\sigma^2(1-e^{-\alpha(t-u)})} r(u) \right), \quad t > u, \quad (3.68)$$

where

$$d = \frac{4b\alpha}{\sigma^2}. \quad (3.69)$$

This says that, given $r(u)$, $r(t)$ is distributed as $\sigma^2(1-e^{-\alpha(t-u)})/(4\alpha)$ times a noncentral chi-square random variable with d degrees of freedom and noncentrality parameter

$$\lambda = \frac{4\alpha e^{-\alpha(t-u)}}{\sigma^2(1-e^{-\alpha(t-u)})} r(u); \quad (3.70)$$

equivalently,

$$P(r(t) \le y | r(u)) = F_{\chi'^2_d(\lambda)} \left(\frac{4\alpha y}{\sigma^2(1-e^{-\alpha(t-u)})} \right),$$

with d as in (3.69), λ as in (3.70), and $F_{\chi'^2_d(\lambda)}$ as in (3.67). Thus, we can simulate the process (3.62) exactly on a discrete time grid provided we can sample from the noncentral chi-square distribution.

Like the Vasicek model, the square-root diffusion (3.62) has a limiting stationary distribution. If we let $t \to \infty$ in (3.68), we find that $r(t)$ converges in distribution to $\sigma^2/4\alpha$ times a noncentral chi-square random variable with d degrees of freedom and noncentrality parameter 0 (making it an ordinary chi-square random variable). This is a stationary distribution in the sense that if $r(0)$ is drawn from this distribution, then $r(t)$ has the same distribution for all t.

Chi-Square and Noncentral Chi-Square

If ν is a positive integer and $Z_1, \ldots, Z_\nu$ are independent $N(0,1)$ random variables, then the distribution of

$$Z_1^2 + Z_2^2 + \cdots + Z_\nu^2$$

is called the chi-square distribution with ν degrees of freedom. The symbol χ^2_ν denotes a random variable with this distribution; the prime in $\chi'^2_\nu(\lambda)$ emphasizes that this symbol refers to the noncentral case. The chi-square distribution is given by

$$P(\chi^2_\nu \le y) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^y e^{-z/2} z^{(\nu/2)-1}\, dz, \quad (3.71)$$

where $\Gamma(\cdot)$ denotes the gamma function and $\Gamma(n) = (n-1)!$ if n is a positive integer. This expression defines a valid probability distribution for all $\nu > 0$ and thus extends the definition of χ^2_ν to non-integer ν.

For integer ν and constants $a_1, \ldots, a_\nu$, the distribution of

$$\sum_{i=1}^{\nu} (Z_i + a_i)^2 \quad (3.72)$$

is noncentral chi-square with ν degrees of freedom and noncentrality parameter $\lambda = \sum_{i=1}^{\nu} a_i^2$. This representation explains the term "noncentral." The distribution in (3.67) extends the definition to non-integer ν.

It follows from the representation in (3.72) that if $\nu > 1$ is an integer, then

$$\chi'^2_\nu(\lambda) = \chi'^2_1(\lambda) + \chi^2_{\nu-1},$$

meaning that the two sides have the same distribution when the random variables on the right are independent of each other. As discussed in Johnson et al. [202, p.436], this representation is valid even for non-integer $\nu > 1$. Thus, to generate $\chi'^2_\nu(\lambda)$, $\nu > 1$, it suffices to generate $\chi^2_{\nu-1}$ and an independent $N(0,1)$ random variable Z and to set

$$\chi'^2_\nu(\lambda) = (Z + \sqrt{\lambda})^2 + \chi^2_{\nu-1}. \quad (3.73)$$

This reduces sampling of a noncentral chi-square to sampling of an ordinary chi-square (and an independent normal) when $\nu > 1$.

For any $\nu > 0$, (3.67) indicates that a noncentral chi-square random variable can be represented as an ordinary chi-square random variable with a random degrees-of-freedom parameter. In more detail, if N is a Poisson random variable with mean $\lambda/2$, then

$$P(N = j) = e^{-\lambda/2} \frac{(\lambda/2)^j}{j!}, \quad j = 0, 1, 2, \ldots.$$

Consider now a random variable $\chi^2_{\nu+2N}$ with N having this Poisson distribution. Conditional on $N = j$, the random variable has an ordinary chi-square distribution with $\nu + 2j$ degrees of freedom:

$$P(\chi^2_{\nu+2N} \le y | N = j) = \frac{1}{2^{(\nu/2)+j}\Gamma((\nu/2)+j)} \int_0^y e^{-z/2} z^{(\nu/2)+j-1}\, dz.$$

The unconditional distribution is thus given by

$$\sum_{j=0}^{\infty} P(N=j) P(\chi^2_{\nu+2N} \le y | N = j) = \sum_{j=0}^{\infty} e^{-\lambda/2} \frac{(\lambda/2)^j}{j!} P(\chi^2_{\nu+2j} \le y),$$

which is precisely the noncentral chi-square distribution in (3.67). We may therefore sample $\chi'^2_\nu(\lambda)$ by first generating a Poisson random variable N and then, conditional on N, sampling a chi-square random variable with $\nu + 2N$ degrees of freedom. This reduces sampling of a noncentral chi-square to sampling of an ordinary chi-square and a Poisson random variable. We discuss methods for sampling from these distributions below. Figure 3.5 summarizes their use in simulating the square-root diffusion (3.62).

Simulation of $dr(t) = \alpha(b - r(t))\,dt + \sigma\sqrt{r(t)}\,dW(t)$
on time grid $0 = t_0 < t_1 < \cdots < t_n$ with $d = 4b\alpha/\sigma^2$

Case 1: $d > 1$
for $i = 0, \ldots, n-1$
  $c \leftarrow \sigma^2(1 - e^{-\alpha(t_{i+1}-t_i)})/(4\alpha)$
  $\lambda \leftarrow r(t_i)(e^{-\alpha(t_{i+1}-t_i)})/c$
  generate $Z \sim N(0,1)$
  generate $X \sim \chi^2_{d-1}$
  $r(t_{i+1}) \leftarrow c[(Z + \sqrt{\lambda})^2 + X]$
end

Case 2: $d \le 1$
for $i = 0, \ldots, n-1$
  $c \leftarrow \sigma^2(1 - e^{-\alpha(t_{i+1}-t_i)})/(4\alpha)$
  $\lambda \leftarrow r(t_i)(e^{-\alpha(t_{i+1}-t_i)})/c$
  generate $N \sim \text{Poisson}(\lambda/2)$
  generate $X \sim \chi^2_{d+2N}$
  $r(t_{i+1}) \leftarrow cX$
end

Fig. 3.5. Simulation of square-root diffusion (3.62) by sampling from the transition density.

Figure 3.6 compares the exact distribution of $r(t)$ with the distribution produced by the Euler discretization (3.66) after a single time step. The comparison is based on $\alpha = 0.2$, $\sigma = 0.1$, $b = 5\%$, and $r(0) = 4\%$; the left panel takes $t = 0.25$ and the right panel takes $t = 1$. These values for the model parameters are sensible for an interest rate model if time is measured in years, so the values of t should be interpreted as a quarter of a year and a full year, respectively. The figures suggest that the Euler discretization produces too many values close to or below 0 and a mode to the right of the true mode. The effect if particularly pronounced over the rather large time step $t = 1$.

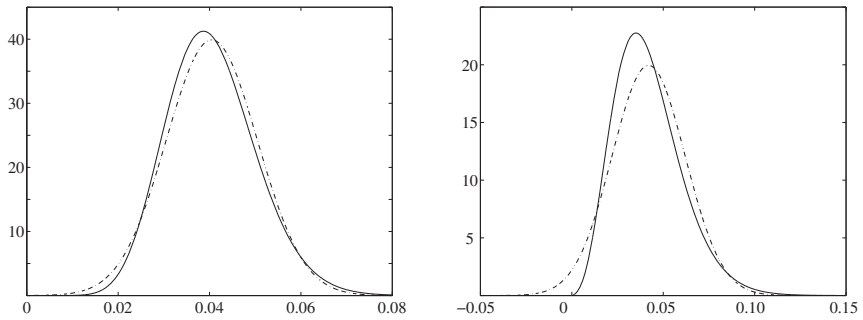

Fig. 3.6. Comparison of exact distribution (solid) and one-step Euler approximation (dashed) for a square-root diffusion with $\alpha = 0.2$, $\sigma = 0.1$, $b = 5\%$, and $r(0) = 4\%$. The left panel compares distributions at $t = 0.25$, the right panel at $t = 1$.

3.4.2 Sampling Gamma and Poisson

The discussion leading to Figure 3.5 reduces the problem of simulating the square-root diffusion (3.62) to one of sampling from a chi-square distribution and possibly also the normal and Poisson distributions. We discussed sampling from the normal distribution in Section 2.3; we now consider methods for sampling from the chi-square and Poisson distributions.

Gamma Distribution

The gamma distribution with shape parameter a and scale parameter β has density

$$f(y) = f_{a,\beta}(y) = \frac{1}{\Gamma(a)\beta^a} y^{a-1} e^{-y/\beta}, \quad y \ge 0. \tag{3.74}$$

It has mean $a\beta$ and variance $a\beta^2$. Comparison with (3.71) reveals that the chi-square distribution is the special case of scale parameter $\beta = 2$ and shape parameter $a = \nu/2$. We therefore consider the more general problem of generating samples from gamma distributions.

Methods for sampling from the gamma distribution typically distinguish the cases $a \le 1$ and $a > 1$. For the application to the square-root diffusion (3.62), the shape parameter a is given by $d/2$ with d as in (3.69). At least in the case of an interest rate model, d would typically be larger than 2 so the case $a > 1$ is most relevant. We include the case $a \le 1$ for completeness and other potential applications. There is no loss of generality in fixing the scale parameter β at 1: if X has the gamma distribution with parameters $(a, 1)$, then βX has the gamma distribution with parameters (a, β).

Cheng and Feast [83] develop a method based on a general approach to random variate generation known as the ratio-of-uniforms method. The ratio-of-uniforms method is closely related to the acceptance-rejection method discussed in Section 2.2.2. It exploits the following property. Suppose f is a

nonnegative, integrable function on $[0,\infty)$; if (X,Y) is uniformly distributed over the set $A = \{(x,y) : x \le \sqrt{f(y/x)}\}$, then the density of Y/X is proportional to f. (See p.180 of Fishman [121] or p.59 of Gentle [136].) Suppose A is contained in a bounded rectangle. Then to sample uniformly from A, we can repeatedly sample pairs (X,Y) uniformly over the rectangle and keep the first one that satisfies $X \le \sqrt{f(Y/X)}$. The ratio-of-uniforms method delivers Y/X as a sample from the density proportional to f.

To sample from the gamma density with $a > 1$, define

$$A = \left\{(x,y) : 0 \le x \le \sqrt{[(y/x)^{a-1}e^{-y/x}]}\right\}.$$

This set is contained in the rectangle $[0,\bar{x}] \times [0,\bar{y}]$ with $\bar{x} = [(a-1)/e]^{(a-1)/2}$ and $\bar{y} = [(a+1)/e]^{(a+1)/2}$. Sampling uniformly over this rectangle, the expected number of samples needed until one lands in A is given by the ratio of the area of A to that of the rectangle. As shown in Fishman [121], this ratio is $O(\sqrt{a})$, so the time required to generate a sample using this method grows with the shape parameter. Cheng and Feast [83] and Fishman [121] develop modifications of this basic approach that accelerate sampling. In Figure 3.7, which is Fishman's Algorithm GKM1, the first acceptance test is a fast check that reduces the number of logarithmic evaluations. When many samples are to be generated using the same shape parameter (as would be the case in the application to the square-root diffusion), the constants in the setup step in Figure 3.8 should be computed just once and then passed as arguments to the sampling routine. For large values of the shape parameter a, Algorithm GKM2 in Fishman [121] is faster than the method in Figure 3.7.

```
Setup: ā ← a − 1, b ← (a − (1/(6a)))/ā, m ← 2/ā, d ← m + 2
repeat
    generate U1, U2 ∼ Unif[0,1]
    V ← bU2/U1
    if mU1 − d + V + (1/V) ≤ 0, accept
    elseif m log U1 − log V + V − 1 ≤ 0, accept
until accept
return Z ← āV
```

Fig. 3.7. Algorithm GKM1 from Fishman [121], based on Cheng and Feast [83], for sampling from the gamma distribution with parameters $(a,1)$, $a > 1$.

Ahrens and Dieter [6] provide a fast acceptance-rejection algorithm for the case $a \le 1$. Their method generates candidates by sampling from distributions concentrated on $[0,1]$ and $(1,\infty)$ with appropriate probabilities. In more detail, let $p = e/(a+e)$ $(e = \exp(1))$ and define

$$g(z) = \begin{cases} paz^{a-1}, & 0 \le z \le 1 \\ (1-p)e^{-z+1}, & z > 1. \end{cases}$$

This is a probability density; it is a mixture of the densities az^{a-1} on $[0,1]$ and e^{-z+1} on $(1,\infty)$, with weights p and $(1-p)$, respectively. We can sample from g by sampling from each of these densities with the corresponding probabilities. Each of these two densities is easily sampled using the inverse transform method : for the density az^{a-1} on $[0,1]$ we can use $U^{1/a}$, $U \sim$ Unif[0,1]; for the density e^{-z+1} on $(1,\infty)$ we can use $1 - \log(U)$. Samples from g are suitable candidates for acceptance-rejection because the ratio $f_{a,1}(z)/g(z)$ with $f_{a,1}$ a gamma density as in (3.74) is bounded. Inspection of this ratio indicates that a candidate Z in $[0,1]$ is accepted with probability e^{-Z} and a candidate in $(1,\infty)$ is accepted with probability Z^{a-1}. A global bound on the ratio is given by

$$f_{a,1}(z)/g(z) \le \frac{a+e}{ae\Gamma(a)} \le 1.39;$$

recall from Section 2.2.2 that the upper bound on this ratio determines the expected number of candidates generated per accepted sample.

Figure 3.8 displays the method of Ahrens and Dieter [6]. The figure is based on Algorithm GS* in Fishman [121] but it makes the acceptance tests more explicit, if perhaps slightly slower. Notice that if the condition $Y \le 1$ fails to hold, then Y is uniformly distributed over $[1,b]$; this means that $(b-Y)/a$ has the distribution of U/e, $U \sim$ Unif[0,1] and thus $-\log((b-Y)/a)$ has the distribution of $1 - \log(U)$.

```
Setup: b ← (a + e)/e
repeat
    generate U1, U2 ∼ Unif[0,1]; Y ← bU1
    if Y ≤ 1
        Z ← Y^(1/a)
        if U2 < exp(−Z), accept
    otherwise  Z ← − log((b − Y)/a)
        if U2 ≤ Z^(a−1), accept
until accept
return Z
```

Fig. 3.8. Ahrens-Dieter method for sampling from the gamma distribution with parameters $(a,1)$, $a \le 1$.

Poisson Distribution

The Poisson distribution with mean $\theta > 0$ is given by

$$P(N=k) = e^{-\theta}\frac{\theta^k}{k!}, \quad k = 0, 1, 2, \ldots. \tag{3.75}$$

We abbreviate this by writing $N \sim \text{Poisson}(\theta)$. This is the distribution of the number of events in $[0,1]$ when the times between consecutive events are independent and exponentially distributed with mean $1/\theta$. Thus, a simple method for generating Poisson samples is to generate exponential random variables $X_i = -\log(U_i)/\theta$ from independent uniforms U_i and then take N to be the largest integer for which $X_1 + \cdots + X_N \le 1$. This method is rather slow, especially if θ is large. In the intended application in Figure 3.5, the mean of the Poisson random variable — equal to half the noncentrality parameter in the transition density of the square-root diffusion — could be quite large for plausible parameter values.

An alternative is to use the inverse transform method. For discrete distributions, this amounts to a sequential search for the smallest n at which $F(n) \le U$, where F denotes the cumulative distribution function and U is Unif[0,1]. In the case of a Poisson distribution, $F(n)$ is calculated as $P(N=0) + \cdots + P(N=n)$; rather than calculate each term in this sum using (3.75), we can use the relation $P(N=k+1) = P(N=k)\theta/(k+1)$. Figure 3.9 illustrates the method.

```
p ← exp(−θ), F ← p
N ← 0
generate U ~ Unif[0,1]
while U > F
    N ← N + 1
    p ← pθ/N
    F ← F + p
return N
```

Fig. 3.9. Inverse transform method for sampling from Poisson(θ), the Poisson distribution with mean θ.

3.4.3 Bond Prices

Cox, Ingersoll, and Ross [91] derived an expression for the price of a bond

$$B(t,T) = \mathsf{E}\left[\exp\left(-\int_t^T r(u)\,du\right)|r(t)\right]$$

when the short rate evolves according to (3.62). The bond price has the exponential affine form

$$B(t,T) = e^{-A(t,T)r(t)+C(t,T)}$$

as in a Gaussian short rate model, but with

$$A(t,T) = \frac{2(e^{\gamma(T-t)}-1)}{(\gamma+\alpha)(e^{\gamma(T-t)}-1)+2\gamma}$$

and

$$C(t,T) = \frac{2\alpha b}{\sigma^2}\log\left(\frac{2\gamma e^{(\alpha+\gamma)(T-t)/2}}{(\gamma+\alpha)(e^{\gamma(T-t)}-1)+2\gamma}\right),$$

and $\gamma = \sqrt{\alpha^2+2\sigma^2}$.

This expression for the bond price is a special case of a more general result, given as Proposition 6.2.5 in Lamberton and Lapeyre [218]. This result gives the bivariate Laplace transform of the short rate and its integral: for nonnegative λ, θ,

$$\mathsf{E}\left[\exp\left(-\lambda r(T) - \theta\int_t^T r(u)\,du\right)|r(t)\right] = \exp(-\alpha b\psi_1(T-t) - r(t)\psi_2(T-t)) \tag{3.76}$$

with

$$\psi_1(s) = -\frac{2}{\sigma^2}\log\left(\frac{2\gamma(\theta)e^{(\alpha+\gamma(\theta))s/2}}{\sigma^2\lambda(e^{\gamma(\theta)s}-1)+\gamma(\theta)-\alpha+e^{\gamma(\theta)s}(\gamma(\theta)+\alpha)}\right),$$

and

$$\psi_2(s) = \frac{\lambda(\gamma(\theta)+\alpha+e^{\gamma(\theta)s}(\gamma(\theta)-\alpha))+2\theta(e^{\gamma(\theta)s}-1)}{\sigma^2\lambda(e^{\gamma(\theta)s}-1)+\gamma(\theta)-\alpha+e^{\gamma(\theta)s}(\gamma(\theta)+\alpha)}$$

and $\gamma(\theta) = \sqrt{\alpha^2+2\sigma^2\theta}$. The bond pricing formula is the special case $\lambda = 0$, $\theta = 1$.

The bivariate Laplace transform in (3.76) characterizes the joint distribution of the short rate and its integral. This makes it possible, at least in principle, to sample from the joint distribution of $(r(t_{i+1}), Y(t_{i+1})$ given $(r(t_i), Y(t_i))$ with

$$Y(t) = \int_0^t r(u)\,du.$$

As explained in Section 3.3.2, this would allow exact simulation of the short rate and the discount factor on a discrete time grid. In the Gaussian setting, the joint distribution of $r(t)$ and $Y(t)$ is normal and therefore easy to sample; in contrast, the joint distribution determined by (3.76) is not explicitly available. Scott [324] derives the Laplace transform of the conditional distribution of $Y(t_{i+1}) - Y(t_i)$ given $r(t_i)$ and $r(t_{i+1})$, and explains how to use numerical transform inversion to sample from the conditional distribution. Through this method, he is able to simulate $(r(t_i), Y(t_i))$ without discretization error.