Monte-Carlo Methods for Single- and Multi-Factor Models

Many term-structure models do not have analytic solutions available for security prices. These prices therefore need to be computed using numerical techniques. Examples of the latter include Monte-Carlo simulation, lattice approximations, finite-difference methods and transform methods. In these notes, we will concentrate on Monte-Carlo methods.

It is not surprising that Monte-Carlo methods may be related to the finite-difference schemes for solving PDE's through the Feynmac-Kac characterization. However, while finite difference methods tend to be quicker than Monte-Carlo methods for low dimensional problems they suffer from the curse of dimensionality and quickly become impractical as the number of dimensions grow. Monte-Carlo simulation does not suffer from this drawback and is not limited to Markovian frameworks as PDE methods are. In the context of term structure models, Monte-Carlo simulation appears to be the computational tool of choice for HJM and LIBOR market models. It is also ubiquitous[1] in the related fields of credit risk, mortgage-backed securities and risk-management generally.

We mention in passing that computing security prices through lattice approximations is equivalent to using an explicit finite-difference scheme for solving the pricing PDE. The above discussion of finite-difference methods therefore also applies to lattice methods.

1 Simulating Stochastic Differential Equations

The following three examples will motivate why we often need to simulate stochastic differential equations in order to estimate quantities of interest.

Example 1 **(Geometric Brownian Motion)**

We want to compute $\theta := \mathrm{E}[f(X_T)]$ where X_t satisfies

$$dX_t \;=\; \mu X_t \; dt + \sigma X_t \; dW_t. \tag{1}$$

The solution to (1) is of course given by

$$X_T \;=\; X_0 \exp\left((\mu - \sigma^2/2)T + \sigma W_T\right). \tag{2}$$

We recognize that X_T depends on the Brownian motion only through the Brownian motion's terminal value, W_T. This implies that even if we are unable to compute θ analytically, we can estimate it by simulating W_T directly. (In this case it is also true that since the distribution of X_T is known, we could also simulate X_T directly. And of course, as an alternative to simulation, we could choose to estimate θ by evaluating the expectation numerically.) ∎

Example 2 **(OU Process)**

Suppose now that we want to compute $\theta := \mathrm{E}[f(X_T)]$ where X_t satisfies

$$dX_t = -\gamma(X_t - \alpha) \; dt + \sigma \; dW_t. \tag{3}$$

[1] See *Monte Carlo Methods in Financial Engineering* by Glasserman for a comprehensive study of Monte-Carlo methods in the context of financial engineering.

The solution to (3) is given by

$$X_T = \alpha + \exp(-\gamma T)[X_0 - \alpha] + \sigma \exp(-\gamma T) \int_0^T \exp(-\gamma s)\, dW_s \tag{4}$$

Note that unlike the previous example, X_T now depends on the entire path of the Brownian motion. This means that we cannot compute an unbiased estimate of θ by first simulating the entire path of the Brownian motion since it is only possible to simulate the latter at discrete intervals of time. It so happens, however, that we know the distribution of X_T: it is normal. In particular, this places us back in the context of Example 1 where, if θ cannot be computed analytically, we can estimate it by either simulating X_T directly or by evaluating the expectation numerically. ∎

Example 3 (CIR Model with Time Varying Parameters)

Again we want to compute $\theta := \mathrm{E}[f(X_T)]$ where X_t satisfies

$$dX_t = \alpha(\mu(t) - X_t)\, dt + \sigma\sqrt{X_t}\, dW_t. \tag{5}$$

and where $\mu(t)$ is a deterministic function of time. While it is clear that X_t follows a CIR process with time varying parameters, we do not know how to find an explicit solution[2] to the SDE in (5). Of course we do not necessarily need an explicit solution to (5) to determine the distribution of X_T (which is what we need to evaluate θ). For example, in the CIR model with constant parameters, we still do not have an explicit solution to the SDE yet it is known that X_T has a non-central χ^2 distribution from which we can easily simulate. Unfortunately, however, once we move to a CIR model with time-varying parameters as in (5), the distribution of X_T is, in general, no longer available. This then complicates the task of computing θ (either analytically or by estimating it by simulating X_T *directly*).

One solution to this problem is to simulate X_T *indirectly* by *simulating the SDE* in (5). ∎

Exercise 1 *Suppose we assume that the short-rate, r_t, has dynamics given by (5). What do the comments in the above example then imply about the difficulty of computing the term-structure?*

Exercise 2 *Suppose you wish to estimate $\theta := \mathrm{E}[f(\{X_t\}_{0\leq t\leq T})]$ so that $f(\cdot)$ now depends on the entire path of the process, X_t. Comment on whether or not this can be reduced to the problem of estimating $\theta = \mathrm{E}[f(Y_T)]$ for some process, Y_T.*

Remark 1 *The situation of Example 3 where we do not know the distribution of X_T is typical. As a result, it is often necessary to simulate a stochastic differential equation if we wish to estimate some associated quantity, e.g. $\theta = \mathrm{E}[f(X_T)]$. We describe how to do this below, beginning with the one-dimensional case.*

Simulating a 1-Dimensional SDE: The Euler Scheme

Let us assume that we are faced with an SDE of the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t)\, dW_t \tag{6}$$

and that we wish to simulate values of X_T but do not know its distribution. (This could be due to the fact that we cannot solve (6) to obtain an explicit solution for X_T, or because we simply cannot determine the distribution of X_T even though we do know how to solve (6)).

When we simulate an SDE, what we mean is that we simulate a *discretized* version of the SDE. In particular, we simulate a discretized process, $\{\widehat{X}_h, \widehat{X}_{2h}, \ldots, \widehat{X}_{mh}\}$, where m is the number of time steps, h is a constant and $mh = T$. The smaller the value of h, the closer our discretized path will be to the continuous-time path we wish

[2] As we did in Examples 1 and 2

to simulate. Of course this will be at the expense of greater computational effort. While there are a number of *discretization schemes* available, we will focus on the simplest and perhaps most common scheme, the *Euler* scheme.

The Euler scheme is intuitive, easy to implement and satisfies

$$\widehat{X}_{kh} \;=\; \widehat{X}_{(k-1)h} + \mu\left((k-1)h,\ \widehat{X}_{(k-1)h}\right)\ h \;+\; \sigma\left((k-1)h,\ \widehat{X}_{(k-1)h}\right)\ \sqrt{h}\ Z_k \tag{7}$$

where the Z_k's are IID $N(0,1)$. If we want to estimate $\theta := \mathrm{E}[f(X_T)]$ using the Euler scheme, then for a fixed number of paths, n, and discretization interval, h, we have the following algorithm.

Using the Euler Scheme to Estimate $\theta = \mathrm{E}[f(X_T)]$
When X_t Follows a 1-Dimensional SDE

> **for** $j = 1$ **to** n
>
> $t = 0$; $\widehat{X} = X_0$
>   **for** $k = 1$ **to** $T/h = m$
>
>   **generate** $Z \sim N(0,1)$
>   **set** $\widehat{X} = \widehat{X} + \mu(t,\widehat{X})h + \sigma(t,\widehat{X})\sqrt{h}\ Z$
>   **set** $t = t + h$
>   **end for**
>   **set** $f_j = f(\widehat{X})$
> **end for**
> **set** $\widehat{\theta}_n = (f_1 + \ldots + f_n))/n$
> **set** $\widehat{\sigma}_n^2 = \sum_{j=1}^{n}(f_j - \widehat{\theta}_n)^2/(n-1)$
> **set** approx. $100(1-\alpha)$ % CI $= \widehat{\theta}_n \pm z_{1-\alpha/2}\frac{\widehat{\sigma}_n}{\sqrt{n}}$

Remark 2 *Observe that even though we only care about X_T, we still need to generate intermediate values, X_{ih}, if we are to minimize the* discretization error*. Because of this discretization error, $\widehat{\theta}_n$ is no longer an unbiased estimator of θ.*

Remark 3 *If we wished to estimate $\theta = \mathrm{E}[f(X_{t_1},\ldots,X_{t_p})]$ then in general we would need to keep track of $(X_{t_1},\ldots,X_{t_p})$ inside the inner for-loop of the algorithm.*

Exercise 3 *Can you think of a derivative where the payoff depends on $(X_{t_1},\ldots,X_{t_p})$, but where it would* not *be necessary to keep track of $(X_{t_1},\ldots,X_{t_p})$ on each sample path?*

Simulating a Multidimensional SDE

In the multidimensional case, $\mathbf{X}_t$, $\mathbf{W}_t$ and $\mu(t,\mathbf{X}_t)$ in (6) are now vectors, and $\sigma(t,\mathbf{X_t})$ is a matrix. This situation arises when we have a *series* of SDE's in our model. This could occur in a number of financial engineering contexts. Some examples include:

(1) Modelling the evolution of multiple stocks. This might be necessary if we are trying to price derivatives whose values depend on multiple stocks or state variables, or if we are studying the properties of some portfolio strategy with multiple assets.

(2) Modelling the evolution of a single stock where we assume that the volatility of the stock is itself stochastic. Such a model is termed a *stochastic volatility* model.

(3) Modelling the evolution of interest rates. For example, if we assume that the short rate, r_t, is driven by a number of factors which themselves are stochastic and satisfy SDE's, then simulating r_t amounts to simulating the SDE's that drive the factors. Such models occur in short-rate models as well as HJM and LIBOR market models.

In all of these cases, whether or not we will have to simulate the SDE's will depend on the model in question and on the particular quantity that we wish to compute. If we do need to discretize the SDE's and simulate their discretized versions, then it is very straightforward. If there are n correlated Brownian motions driving the SDE's, then at each time step, t_i, we must generate n IID $N(0,1)$ random variables. We would then use the Cholesky Decomposition to generate $X_{t_{i+1}}$. This is exactly analogous to our method of generating correlated geometric Brownian motions. In the context of simulating multidimensional SDE's, however, it is more common to use independent Brownian motions as any correlations between components of the vector, $\mathbf{X}_t$, can be induced through the matrix, $\sigma(t, \mathbf{X}_t)$.

Allocation of Computational Resources

An important issue that arises when simulating SDE's is the allocation of computational resources. In particular, we need to determine how many sample paths, n, to generate and how many steps, m, to simulate on each sample path. A smaller value of m will result in greater bias and numerical error, whereas a smaller value of n will result in greater statistical noise. If there is a fixed computational budget then it is important to choose n and m in an optimal manner. We now discuss this issue.

Suppose $dX_t = \mu(t, X_t)\ dt + \sigma(t, X_t)\ dB_t$ and that we wish to estimate $\theta := E[f(X_T)]$ using an Euler approximation scheme. Let $\widehat{\theta}_n^h$ be an estimate of θ based upon simulating n sample paths using a total of $m = T/h$ discretization points per path. In particular, we have

$$\widehat{\theta}_n^h = \frac{f(\widehat{X}_1^h) + \ldots + f(\widehat{X}_n^h)}{n}$$

where $\widehat{X}_i^h$ is the value of $\widehat{X}_T^h$ on the i^{th} path. Under some technical conditions on the process, X_t, it is known that the Euler scheme has a *weak order of convergence* equal to 1 so that

$$|E[f(X_T)] \ - \ E[f(\widehat{X}_T^h)]| \leq am^{-1} \tag{8}$$

for some constant a and all m greater than some fixed m_0.

Suppose now that we have a fixed computational budget, C, and that each simulation step costs c. We must therefore have $n = C/mc$. We would like to choose the optimal values of m (and therefore n) as a function of C. We do this by minimizing the *mean squared error* (MSE), which is the sum of the bias squared and the variance, v. In particular, (8) implies

$$MSE \approx \frac{a^2}{m^2} + \frac{v}{n} \tag{9}$$

for sufficiently large m. Substituting for n in (9), it is easy to see that it is optimal (for sufficiently large C) to take

$$m \ \propto \ C^{1/3} \tag{10}$$

$$n \ \propto \ C^{2/3}. \tag{11}$$

When it comes to estimating θ, (10) and (11) provide guidance as follows. We begin by using n_0 paths and m_0 discretization points per path to compute an initial estimate, $\widehat{\theta}_0$, of θ. If we then compute a new estimate, $\widehat{\theta}_1$, by setting $m_1 = 2m_0$, then (10) and (11) suggest we should set $n_1 = 4n_0$. We may then continue to compute new estimates, $\widehat{\theta}_i$, in this manner until the estimates and their associated confidence intervals converge. In general, if we increase m by a factor of 2 then we should increase n by a factor of 4. Although estimating θ in this way requires additional computational resources, it is not usually necessary to perform more than two or three iterations, provided we begin with sufficiently large values of m_0 and n_0.

Remark 4 *There are other important issues that arise when simulating SDE's. For example, while we have only described the Euler scheme, there are other more sophisticated discretization schemes that can also be used. In a sense that we will not define, these schemes have superior convergence properties than the Euler scheme. However, they are sometimes more difficult to implement, particularly in the multi-dimensional setting.*

---

2 Applications to Financial Engineering

Example 4 (Option Pricing Under Stochastic Volatility)

Suppose the evolution of the stock price, S_t, under the risk-neutral probability measure is given by

$$dS_t = rS_t\ dt + \sqrt{V_t} S_t\ dW_t^{(1)} \tag{12}$$

$$dV_t = \alpha\,(b - V_t)\ dt + \sigma\sqrt{V_t}\ dW_t^{(2)}. \tag{13}$$

If we want to price a European call option on the stock with expiration, T, and strike K, then the price is given by

$$C_0 = \exp(-rT)\mathrm{E}[\max(S_T - K,\ 0)].$$

We could estimate C_0 by simulating n sample paths of $\{S_t, V_t\}$ up to time T, and taking the average of $\exp(-rT)\max(S_T - K,\ 0)$ over the n paths as our estimated call option price, $\widehat{C}_0$. ∎

Exercise 4 *Write out the details of the algorithm that you would use to estimate C_0 in Example 4.*

Example 5 (Portfolio Evaluation)

Suppose an investor trades continuously in a particular fund whose time t value is denoted by P_t. Any cash that is not invested in the the fund earns interest in a cash account at the risk-free rate, r_t. Assume that the dynamics of P_t are given by

$$\begin{aligned} dP_t &= P_t\left[\ (\mu + \lambda X_t)\ dt + \sigma_1\ dW_t^{(1)} + \sigma_2\ dW_t^{(2)}\right] \\ dX_t &= -kX_t\ dt + \sigma_{x,1}\ dW_t^{(1)} + \sigma_{x,2}\ dW_t^{(2)} \\ dr_t &= \alpha(t, r_t)\ dt + \beta(t, r_t)\ dW_t^{(3)} \end{aligned}$$

where X_t is a *state variable* that possibly represents the time t value of some relevant economic variable. Let θ_t be an adapted process that denotes the *fraction* of the investor's wealth that is invested[3] in the fund at time t, and let Y_t denote the investor's wealth at time t. We then see that Y_t satisfies

$$dY_t = [r_t + \theta_t(\mu + \lambda X_t - r)]Y_t\ dt\ +\ \theta_t Y_t[\sigma_1\ dW_t^{(1)} + \sigma_2\ dW_t^{(2)}]. \tag{14}$$

Now it may be the case that the investor wishes to compute $\mathrm{E}[u(Y_T)]$ where $u(\cdot)$ is his utility function, or that he wishes to compute $\mathbf{P}(Y_T \le a)$ for some fixed value, a. In general, however, it is not possible to perform these computations explicitly. As a result, we could instead use simulation. Noting that we will not in general be able to solve (14) for Y_T and its distribution, this means that we would have to simulate the multivariate SDE satisfied by (P_t, X_t, r_t, Y_t) in order to answer these questions. ∎

[3] This means that $1 - \theta_t$ is invested in the cash account at time t.

Exercise 5 *Write out the details of the algorithm that you would use to estimate $\theta = \mathbf{P}(Y_T \leq a)$ in Example 5.*

Example 6 (The CIR Model with Time-Dependent Parameters)
We assume the Q-dynamics of the short-rate, r_t, are given by

$$dr_t \;=\; \alpha[\mu(t) - r_t]\; dt \;+\; \sigma\sqrt{r_t}\; dW_t \tag{15}$$

where $\mu(t)$ is a deterministic function of time. This generalized CIR model is used when we want to fit a CIR-type model to the initial term-structure. (We will see later that a CIR model with constant parameters when modelled under Q becomes a CIR model with time-varying parameters under the forward measure, P^τ.)

Suppose now that we wish to price a derivative security maturing at time T with payoff $C_T(r_T)$. Then its time 0 price, C_0, is given by

$$C_0 = \mathrm{E}_0\left[e^{-\int_0^T r_s\; ds}\; C_T(r_T)\right]. \tag{16}$$

The distribution of r_t is not available in an easy-to-use closed form so perhaps the easiest way to estimate C_0 is by simulating the dynamics of r_t. Towards this end, we could either use (15) and simulate r_t directly or alternatively, we could simulate $X_t := f(r_t)$ where $f(\cdot)$ is an invertible transformation. Note that because of the discount factor in (16), it is also necessary to simulate the process, Y_t, given by

$$Y_t = \exp(-\int_0^t r_s\; ds).$$

∎

Exercise 6 *Describe in detail how you would you would estimate C_0 in Example 6. Note that there are alternative ways to do this. What way do you prefer?*

Exercise 7 *Suppose we wish to simulate the known dynamics of a zero-coupon bond. How would you ensure that the simulated process satisfies $0 < Z_t^T < 1$?*

Bias and Jensen's Inequality

We first state Jensen's Inequality.

Theorem 1 (Jensen's Inequality) *Suppose $f(\cdot)$ is a concave*[4] *function on R, $\mathrm{E}[X] < \infty$ and $\mathrm{E}[f(X)] < \infty$. Then $\mathrm{E}[f(X)] \leq f(\mathrm{E}[X])$.*

Returning to Example 6, we assumed there that the payoff function, $C_T(r_T)$, was easy to evaluate. This is true for example, if $C_T(r_T) = (Z_T^U - K)^+$ in the Vasicek or CIR model, among others. However, in many circumstances $C_T(r_T)$ will not be easy to evaluate. This is true, for example, when $C_T(r_T) = f(Z_T^U)$ in models where zero-coupon bond prices are *not* available in closed form. In such circumstances it may be necessary to estimate $f(Z_T^U)$ using an additional simulation. Staying with the call option example, we see that its price, C_0, may be written as

$$\begin{aligned} C_0 &= \mathrm{E}_0\left[e^{-\int_0^T r_s\; ds}\; (Z_T^U - K)^+\right] \\ &= \mathrm{E}_0\left[e^{-\int_0^T r_s\; ds}\; \left(\mathrm{E}_T\left[e^{-\int_T^U r_v\; dv}\right] - K\right)^+\right]. \end{aligned} \tag{17}$$

To estimate Z_T^U along a given sample path we see from (17) that it will therefore be necessary to perform an additional simulation, or a "simulation within a simulation".

[4]A function $f(\cdot)$ is concave on R if $f(\alpha x + (1-\alpha)y) \geq \alpha f(x) + (1-\alpha)f(y)$ for all $\alpha \in [0,1]$.

Exercise 8 *Use Jensen's Inequality and (17) to show that the estimate of C_0 will be biased away from the true value. In what direction will the bias be?*

The extent of the bias will depend on how accurately we can estimate Z_T^U along each simulated sample path. Accuracy can be improved by conducting a large number of "simulations within the simulation" but this is at a cost of requiring more computational resources. For reasons that will be clear later, HJM and Market LIBOR models do not not suffer from this bias problem.

---

3 Variance Reduction Techniques

Simulating SDE's is a computationally intensive task as we need to do a lot of work for each sample that we generate. As a result, variance reduction techniques are often very useful in such contexts. We give some examples.

Example 7 (Bond Prices as Control Variates)

In term-structure models we usually taken the prices of zero-coupon bonds as primitives. Indeed, a term structure model is often constructed in such a way that zero prices in the model coincide with observed zero prices in the market place. As a by-product of this, it means that zero prices are also often available to be used as control variates. Recalling that Z_0^T is given by

$$Z_0^T \;=\; \mathrm{E}_0^Q\left[\exp\left(-\int_0^T r_s\, ds\right)\right]$$

we identify two situations that may occur:

1. Due to discretization, we need to use
$$\exp\left(-\frac{T}{n}\sum_{i=1}^{n} r_{t_i}\right) - Z_0^T \tag{18}$$
as our control variate. In this situation the expression in (18) may not have mean 0 due to discretization error. However, this bias can be made arbitrarily small by taking a sufficiently fine partition of $[0, T]$.

2. It is possible to simulate $\exp\left(-\int_0^T r_s\, ds\right)$ exactly while at the same time, simulating the SDE for r_t. This is possible, for example, in the Vasicek model where the joint distribution of $\left(-\int_0^{t_i} r_s\, ds,\; r_{t_i}\right)$ is known to be bivariate normal.

∎

Remark 5 *When continuous-time term-structure models are implemented numerically, it is necessary to discretize time as we have been doing when simulating SDE's. In such circumstances, however, it is sometimes desirable to insist that the discretized model is also arbitrage free. When we do this we can ensure that we obtain unbiased estimates of bond prices so that the control variate of (1) above is also unbiased. This is quite common in implementations of HJM and LIBOR market models.*

For an example based on conditional Monte-Carlo, consider again the stochastic volatility model of Example 4.

Example 8 (Conditional Monte-Carlo for the Stochastic Volatility Model)

As before, we will use $c(x, t, K, r, \sigma)$ to denote the Black-Scholes price of a European call option when the current stock price is x, the time to maturity is t, the strike is K, the risk-free interest rate is r and the volatility is σ.

We will also use the following fact regarding Gaussian processes: if v_t is a deterministic function of time, then

$$\int_0^T v_t \; dW_t \;\sim\; \mathsf{N}\left(0, \; \int_0^T v_t^2 \; dt\right).$$

Suppose now that the Brownian motions, $W_t^{(1)}$ and $W_t^{(2)}$ in (12) and (13), are independent. Then

$$C_0 \;=\; e^{-rT} \, \mathrm{E}[\max(S_T - K, \; 0)] \;=\; e^{-rT} \, \mathrm{E}\left[\mathrm{E}[\, \max(S_T - K, \; 0) \mid V_t, \; 0 \le t \le T] \,\right].$$

But it can be shown using the independence of $W_t^{(1)}$ and $W_t^{(2)}$ that

$$e^{-rT} \, \mathrm{E}[\, \max(S_T - K, \; 0) \mid V_t, \; 0 \le t \le T] \;=\; c(S_0, T, K, r, V)$$

where $V := \sqrt{\int_0^T V_t \; dt/T}$. In particular, this means that we can estimate C_0 by using conditional Monte-Carlo method. ∎

Exercise 9 *Write out the details of the conditional Monte-Carlo algorithm that you would use to estimate C_0.*

Remark 6 *The above example may be generalized in certain circumstances to accommodate dependence between $W_t^{(1)}$ and $W_t^{(2)}$.*

Example 9 (Antithetic Variates)

Suppose we wish to estimate $\theta = \mathrm{E}_0^Q[C_T]$ using an Euler scheme. (Note that we assume the discount factor is accounted for in C_T.) Let ϵ^+ denote the sequence of IID $N(0,1)$ random variables used to generate the path of the short rate, r_t, in the interval $[0, T]$. Likewise let C_T^+ be the payoff along this path. Setting $\epsilon^- = -\epsilon^+$, we can then use ϵ^- to construct another sample payoff, C_T^-. We then obtain an antithetic estimator by setting

$$C_T \;=\; \frac{C_T^+ + C_T^-}{2}.$$

An antithetic estimator will often provide a significant variance reduction over the naive estimator. The magnitude of the variance reduction, however, will depend on $\mathrm{Cov}(C_T^+, C_T^-)$ with a positive covariance resulting in a variance *increase*. ∎

Exercise 10 *Would the method of antithetic variates be guaranteed (by the monotonicity theorem for antithetic variates) to provide a variance reduction when estimating the value of a derivative security with date T payoff given by $|r_T - \bar{r}|$? (To be precise, we should first specify a model but the answer should be the same for most reasonable models.)*

Remark 7 *The antithetic method obviously extends to multi-factor models, as do all of the variance reduction methods.*

Example 10 (The Brownian Bridge and Stratified Sampling)

Consider a short rate model of the form

$$dr_t \;=\; \mu(t, r_t) \; dt + \sigma(t, r_t) \; dW_t.$$

When pricing a derivative that matures at time T using an Euler scheme it is necessary to generate the path $(W_h, W_{2h}, \ldots, W_{mh} = W_T)$. It will often be the case, however, that the value of W_T will be particularly significant in determining the payoff. As a result, we might want to stratify using the random variable, W_T. This is easy to do for the following two reasons.

(i) $W_T \sim N(0,T)$ so we can easily generate a sample of W_T and

(ii) We can easily generate $(W_h, W_{2h}, \ldots, W_{T-h} \mid W_T)$ by computing the relevant conditional distributions and then simulating from them. For example, it is straightforward to see that

$$(W_t \mid W_s = x, W_v = y) \;\sim\; N\left(\frac{(v-t)x + (t-s)y}{v-s},\; \frac{(v-t)(t-s)}{v-s}\right) \quad \text{for } s < t < v \tag{19}$$

and we can use this result to generate $(W_h|W_0, W_T)$. More generally, we can use (19) to successively simulate $(W_h|W_0, W_T)$, $(W_{2h}|W_h, W_T)$, $\ldots$, $(W_{T-h}|W_{T-2h}, W_T)$.

We can in fact simulate the points on the sample path in any order we like. In particular, to simulate W_v we use (19) and condition on the two closest sample points before and after v, respectively, that have already been sampled. This method of pinning the beginning and end points of the Brownian motion is known[5] as the *Brownian bridge*. ∎

Exercise 11 *If we are working with a multi-dimensional correlated Brownian motion, W_t, (e.g. in the context of a multi-factor model of the short rate) is it still easy to use the Brownian bridge construction where we first generate the random vector, W_T?*

Remark 8 *It is clear, but perhaps worth mentioning nonetheless, that the Brownian bridge / stratification technique is not restricted to term structure applications.*

We will delay a discussion of importance sampling until the next section when we talk about changing the numeraire. This makes sense as the two concepts are clearly related.

Change of Numeraire and Importance Sampling

Let π_t denote the time t price of a contingent claim that expires at time $\tau > t$. We saw in the *Overview of Stochastic Calculus* lecture notes that we can then write

$$\pi_t \;=\; B_t \mathrm{E}_t^Q\left[\frac{\pi_\tau}{B_\tau}\right] \;=\; Z_t^\tau \mathrm{E}_t^{P^\tau}[\pi_\tau] \tag{20}$$

where P^τ is the forward measure[6] that corresponds to taking Z_t^τ as the numeraire security. We then saw in Example 1 of the *Continuous-Time Short-Rate Models* lecture notes how this change of numeraire technique can be useful for obtaining analytic expressions for derivative security prices in the Vasicek and Hull-White models. More generally, this change of numeraire technique can also be advantageous when using simulation to estimate security prices.

Example 11 (Pricing a European Option on a Zero-Coupon Bond)

Suppose we wish to price a European call option that expires at time T with payoff given by $\max(Z_T^U - k, 0)$ where $U > T$. Then (20) implies that its time 0 price, C_0, is given by

$$C_0 \;=\; \mathrm{E}_0^Q\left[e^{-\int_0^T r_s\, ds}\; \max(Z_T^U - k, 0)\right] \tag{21}$$

$$\;=\; Z_0^\tau \mathrm{E}_0^{P^\tau}[\max(Z_T^U - k, 0)]. \tag{22}$$

Note that if we estimate C_0 by simulating then there are at least two advantages that arise from using the expression in (22) rather than the expression in (21):

1. We do not need to keep track of the process $X_t := \int_0^t r_s\, ds$

[5] See Glasserman for further details.
[6] Recall also that $dP^\tau/dQ \;=\; 1/(B_\tau Z_0^\tau)$.

2. We avoid the discretization error associated with our general inability to simulate the discount factor exactly. (Note that we might still incur some discretization error in simulating Z_T^U.)

■

Remark 9 *In the Vasicek and Hull-White models we know that Z_T^U has a log-normal distribution so there is no need to use simulation to price the option in Example 11. In general, however, we will not know the distribution for Z_T^U and so simulating its SDE will be necessary.*

Exercise 12 *Can you think of any other advantage that might result from simulating under the forward measure? (Depending on the context, this might also be a disadvantage.)*

Exercise 13 *Check that a Vasicek model under Q becomes a Vasicek model with time varying parameters (i.e. a Hull-White model) under P^τ. Does a similar result hold for the CIR model?*

Relation to Importance Sampling

By necessity, a change in the martingale measure must accompany a change in the numeraire security. It is therefore clear that the change of numeraire technique is related to importance sampling. However, the motivation for changing numeraire is generally different to the motivation for using importance sampling. Possible reasons for changing the numeraire include:

1. Facilitating analytic computations (e.g. derivatives pricing in Gaussian models)
2. It might be preferable to model dynamics under a particular numeraire-EMM pair (e.g. certain LIBOR market models)
3. When using Monte-Carlo simulations, workload and / or discretization bias can be reduced.

In contrast, the motivation for changing the probability measure when we use importance sampling is to reduce the variance of a particular estimator. In fact, even when the numeraire-EMM pair is changed with a view to simulating the corresponding SDE's, it might be the case that estimator variances *increase*.

Exercise 14 *Consider pricing an out-of-the-money European option with payoff $\max(Z_T^U - k, 0)$ where we need to simulate an SDE to generate samples of Z_T^U. Do you think working under the forward measure would tend to increase or decrease the variance of your estimator? What if you wanted to price a put option?*

Exercise 15 *Is it possible to work with $\left(Z_t^T, P^T\right)$ as the numeraire-EMM pair and then use importance sampling without foregoing the advantages of working under the forward measure if we wish to price a security using simulation?*

---

4 Exact Simulation of Term Structure Models

Suppose the short-rate, r_t, satisfies

$$dr_t \;=\; \alpha(t, r_t)\; dt + \beta(t, r_t)\; dW_t$$

where W_t is a Q-Brownian motion. Then a derivative security with payoff at time T given by $C(r_T)$ has time 0 price, C_0, given by

$$C_0 \;=\; \mathrm{E}_0^Q \left[e^{-\int_0^T r_s\, ds}\; C(r_T) \right].$$

In general, if we are to estimate C_0 using simulation then it will be necessary to simulate the SDE's satisfied by r_t and $Y_t := \int_0^t r_s\; ds$. However, it is worth mentioning that in some circumstances it is possible to simulate without bias from the joint distribution of (Y_T, r_T). This would then imply that we could estimate C_0 without any statistical bias.

Example 12 (Gaussian Models)

The joint distribution of (Y_T, r_T) in the Vasicek model is a bivariate normal distribution from which we can easily simulate. This result holds more generally[7] for other Gaussian processes whether they are univariate or multivariate, or whether they have constant or time-varying parameters. In particular for the Vasicek model where the Q-dynamics of the short rate satisfy

$$dr_t \;=\; \alpha(\mu - r_t)\; dt + \sigma \; dW_t$$

it may be seen[8] that

$$\begin{aligned}
\mathrm{E}^Q\left[r_T \mid r_t\right] &= \mu + (r_t - \mu)e^{-\alpha\tau} \\
\mathrm{E}^Q\left[\int_t^T r_s \; ds \mid r_t\right] &= \mu\tau + (r_t - \mu)\frac{(1 - e^{-\alpha\tau})}{\alpha} \\
\mathrm{Var}^Q(r_T \mid r_t) &= \sigma^2 \frac{(1 - e^{-2\alpha\tau})}{2\alpha} \\
\mathrm{Var}^Q\left(\int_t^T r_s \; ds \mid r_t\right) &= \frac{\sigma^2}{2\alpha^3}\left[2\alpha\tau - 3 + 4e^{-\alpha\tau} - e^{-2\alpha\tau}\right] \\
\mathrm{Cov}^Q\left(R_T, \int_t^T r_s \; ds \mid r_t\right) &= \frac{\sigma^2}{2\alpha^2}(1 - e^{-\alpha\tau})^2
\end{aligned}$$

where $\tau := T - t$. ∎

Example 13 (CIR Models)

It is also possible to simulate (Y_T, r_T) without bias in the CIR model but this is considerably more complicated than the Gaussian case. This is due to the fact that the distribution of (Y_T, r_T) is not explicitly available and generating a sample of (Y_T, r_T) is therefore more difficult (though it is possible). Generating a sample of r_t on the other hand is straightforward as it is known to have a non-central χ^2 distribution from which it is easy to simulate.

When we move to a CIR model with time-varying coefficients, however, it is no longer possible to simulate (Y_T, r_T) directly without bias and it becomes necessary to simulate the SDE's using, for example, the Euler scheme. ∎

---

[7] See *Monte Carlo Methods in Financial Engineering* by Glasserman.
[8] See Cairns for a derivation based on computing the joint moment generating function.