The General Hull–White Model and Supercalibration

John Hull and Alan White

Term-structure models are widely used to price interest rate derivatives, such as swap options and bonds with embedded options. We describe how a general one-factor model of the short rate can be implemented as a recombining trinomial tree and calibrated to market prices of actively traded instruments. The general model encompasses most popular one-factor Markov models as special cases. The implementation and the calibration procedures are sufficiently general that they can select the functional form of the model that best fits the market prices. This characteristic allows the model to fit the prices of in- and out-of-the-money options when there is a volatility skew. It also allows the model to work well with economies characterized by very low interest rates, such as Japan, for which other models often fail.

Two major approaches to modeling the term structure of interest rates are (1) to model the evolution of either forward rates or discount bond prices and (2) to describe the evolution of the instantaneous rate of interest. The first approach was introduced by Heath, Jarrow, and Morton (HJM 1992), who modeled the behavior of instantaneous forward rates. Their method is both powerful (it encompasses many other term-structure models as special cases) and easy to understand. It exactly fits the initial term structure of interest rates, it permits as complex a volatility structure as desired, and it can readily be extended to as many sources of risk as desired.

Recently, the HJM model was modified by Brace, Gatarek, and Musiela (1997), Jamshidian (1997), and Miltersen, Sandmann, and Sondermann (1997) to apply to noninstantaneous forward rates. This modification has come to be known as the LIBOR Market Model (LMM). In one version, three-month forward rates are modeled, which allows the model to exactly replicate observed cap prices that depend on three-month forward rates.[1] In another version, forward swap rates are modeled. This modification allows the model to exactly replicate observed European swap option prices. The main difficulty with the HJM/LMM models is that they are difficult to implement by any means other than Monte Carlo simulation. As a result, they are computationally slow and difficult to use for American or Bermudan options.[2]

The other major approach to modeling the term structure is to describe the evolution of the instantaneous rate of interest, the rate that applies over the next short time interval. Short-rate models are often more difficult to understand than models of the forward rate. However, they are implemented in the form of a recombining tree similar to the stock-price tree first developed by Cox, Ross, and Rubinstein (1979). Thus, the computation is fast and the models are useful for valuing many types of interest rate derivatives.

The Generalized Model

In the general Hull–White model, some function of the short rate, $f(r)$, obeys a Gaussian diffusion process of the following form:

$$df(r) = [\theta(t) - a(t)f(r)]dt + \sigma(t)dz. \quad (1)$$

The function $\theta(t)$ is a term-structure parameter that is selected so that the model fits the initial term structure. The functions $a(t)$ and $\sigma(t)$ are volatility parameters that are chosen to fit the current market prices of a set of actively traded interest rate options. The diffusion process, dz, is a standard Wiener process with a zero mean and a variance equal to dt.

The general Hull–White model contains many popular term-structure models as special cases. For example, when $f(r) = r$, $a(t)$ is zero, and σ is

John Hull and Alan White are professors of finance at the Joseph L. Rotman School of Management, University of Toronto, Ontario.

constant, it is the Ho–Lee (1986) model. When $f(r) = r$ and $a(t)$ is not zero, it is the original Hull–White (1990) model. In both these models, future interest rates of all maturities are normally distributed and analytic solutions exist for the prices of bonds and options on bonds. When $f(r) = \sqrt{r}$, it is a model developed by Pelsser (1996), and when $f(r) = \ln r$, it is the Black–Karasinski (1991) model, which is perhaps the most popular version currently in use. In this model, the future short rate is lognormally distributed and rates of all other maturities are approximately lognormally distributed.

Implementation

In this section, we describe how the generalized model is implemented in a recombining trinomial tree. Initially, we assume that volatility parameters $a(t)$ and $\sigma(t)$ and the functional form $f(r)$ have been selected. Later, we will describe how these parameters are chosen.

First, we set the current time to zero and define a deterministic function g that satisfies

$$dg = [\theta(t) - a(t)g(t)]dt. \tag{2}$$

We then define a new variable, x, that is

$$x(r,t) = f(r) - g(t). \tag{3}$$

The new variable obeys a much simpler diffusion process than $df(r)$ in Equation 1:

$$dx = -a(t)xdt + \sigma(t)dz. \tag{4}$$

The initial value of g is chosen so that the initial value of x is zero.[3] This process is mean reverting to zero; so, if x starts at zero, the unconditional expected value of x at all future times is zero.

Building a tree for $f(r)$ involves four steps. The first step is to select the spacing of the tree nodes in the time dimension. The second step is to decide on the spacing of the nodes in the interest rate dimension. The third step is to choose the branching process for $x(r,t)$ through the grid of nodes. Once this task is completed, the fourth step involves shifting the tree by the value of g at each point in time. The result is a tree for the function $f(r)$.

Choosing the Times at Which Nodes Are Placed. When a term-structure model is implemented, it is usually for some specific purpose, such as pricing an option on a swap. As a result, the convenient approach is to construct the tree with nodes on specified dates, such as payment and exercise dates. For example, suppose we wish to build an n-step tree with nodes at times $t_0, t_1, t_2, \ldots, t_n$, where $t_0 = 0$, $t_i > t_{i-1}$, and $t_n = T$, the longest date to be considered. Because the values of all bonds, swaps, and other instruments are computed by discounting their payoffs back through the tree, T must be chosen so that no payments occur after T. We should also ensure that we have chosen our node times, t_i's, so that we have a set of nodes on every payment date. Other node times can be selected to increase the resolution of the tree.

Choosing the Values of x Where Nodes Are to Be Placed. Once the times at which nodes are to be placed have been chosen, we must choose the values of x where nodes are to be placed at each time step. First, we place a node at $x = 0$. Then, at each time step t_i (where $i = 1, \ldots, n$), we place nodes at $\pm\Delta x_i, \pm 2\Delta x_i, \ldots, \pm m_i\Delta x_i$. (The determination of the value of m_i will be explained in the following section.) In choosing the Δx_i's, the only constraint is that the spacing of the nodes must be wide enough to represent the volatility of x at that time. We meet this requirement by setting the x-spacing at time t_i to[4]

$$\Delta x_i = \sigma(t_{i-1})\sqrt{3(t_i - t_{i-1})}. \tag{5}$$

The next stage of the implementation is to determine how the nodes in (x,t) space will be connected together, which will determine the m_i's, the indexes of the highest and lowest nodes that are attainable at each time step.

Choosing the Branching Process. We next choose the branching through the tree so that at every point in the tree, we are mimicking the diffusion process as closely as possible. We do so by ensuring that the expected change and the variance of the change in x seen on the tree are the same as predicted by the diffusion process for x. At each node in the tree, we select the branching process and the branching probabilities accordingly.

Suppose we are at some node $j\Delta x_i$ at time step i and propose to branch to nodes $(k-1)\Delta x_{i+1}$, $k\Delta x_{i+1}$, and $(k+1)\Delta x_{i+1}$ at time step $i+1$. From the diffusion process for x, we calculate the expected mean change in x over the next time interval, $E(dx) = M$, and the second moment of x, $E(dx^2) = V + M^2$.[5]

We let the probabilities of branching to $(k-1)\Delta x_{i+1}$, $k\Delta x_{i+1}$, and $(k+1)\Delta x_{i+1}$ be, respectively, p_d, p_m, and p_u. Matching the mean and variance gives

$$\begin{aligned} j\Delta x_i + M &= k\Delta x_{i+1} + (p_u - p_d)\Delta x_{i+1} \\ V + (j\Delta x_i + M)^2 &= k^2(\Delta x_{i+1})^2 + 2k(p_u - p_d)(\Delta x_{i+1})^2 \\ &\quad + (p_u - p_d)(\Delta x_{i+1})^2. \end{aligned} \tag{6}$$

Solving Equation 6 produces

$$p_u = \frac{V}{2\Delta x_{i+1}^2} + \frac{\varepsilon^2 + \varepsilon}{2}, \tag{7a}$$

$$p_d = \frac{V}{2\Delta x_{i+1}^2} + \frac{\varepsilon^2 - \varepsilon}{2}, \quad (7b)$$

and

$$p_m = 1 - \frac{V}{\Delta x_{i+1}^2} - \varepsilon^2, \quad (7c)$$

where $\varepsilon = [j\Delta x_i + M - k\Delta x_{i+1}]/\Delta x_{i+1}$ and is the distance from the expected value of x to the central node to which we are branching.

If $V = \sigma^2(t_i)(t_{i+1} - t_i)$ and $\Delta x_{i+1} = \sigma(t_i) \times \sqrt{3(t_{i+1} - t_i)}$, it can be shown that all the branching probabilities are positive if $-\sqrt{2/3} < \varepsilon < \sqrt{2/3}$. That is, when branching from a point $j\Delta x_i$, we should choose as the central node of the three successor nodes a node within $\sqrt{2/3}\Delta x_{i+1}$ of the expected outcome. Usually, we would choose the node closest to the expected outcome by setting k to the value of $(j\Delta x_i + M)/\Delta x_{i+1}$ rounded to the nearest integer. This process ensures that we are within $\Delta x_{i+1}/2$ of the expected outcome, and the condition for positive probabilities is satisfied.

The procedure just described determines the tree branches and branching probabilities. It also defines the highest and lowest possible node at each time step. The index of the highest node at time step $i + 1$, m_{i+1}, is determined by the branching from m_i, the index of the highest node at time step i. Similarly, $-m_{i+1}$, the index of the lowest node at time step $i + 1$, is determined by the branching from $-m_i$, the index of the lowest node at step i. At time step 0, there is only one node, $m_0 = 0$. In this way, the highest and lowest nodes at the first time step and all subsequent time steps can be determined.

We illustrate the calculation with an extreme example. Suppose that $t_0 = 0$, $t_1 = 1.5$, $t_2 = 1.6$, and $t_3 = 2.0$; so, the time steps are of widely varying lengths. (In most applications, the time steps are much more equal.) Suppose also that the volatility parameters are $a(t) = 1.0$ and $\sigma(t) = 0.30$ for all t. The node spacing at each time step is determined from Equation 5. The result is $\Delta x_1 = 0.6364$, $\Delta x_2 = 0.1643$, and $\Delta x_3 = 0.3286$. The grid of nodes on the tree is, therefore, as shown in **Table 1**.

Table 1. Grid of Tree Nodes

$t = 0$	$t = 1.5$	$t = 1.6$	$t = 2.0$
		0.3286	0.6573
	0.6364	0.1643	0.3286
0.0000	0.0000	0.0000	0.0000
	–0.6364	–0.1643	–0.3286
		–0.3286	–0.6573

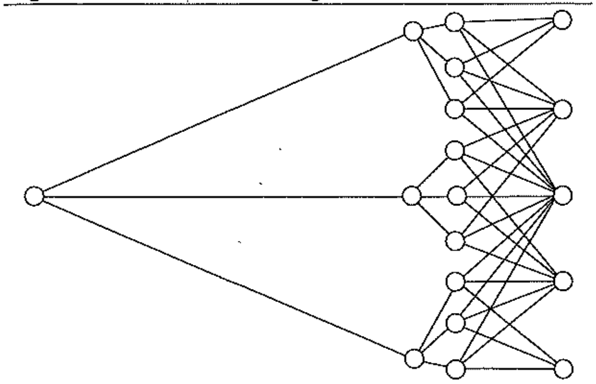
The next step is to compute the branching process. Starting at the root node ($t = 0$ and $x = 0$), we compute $x + M = x - ax \times 1.5 = 0$ and $V = 0.30^2 \times 1.5 = 0.135$. The node closest to the expected outcome is the node $k = 0$ at $t = 1.5$. For this node, $\varepsilon = 0$, and from Equations 7, the branching probabilities are $p_d = 0.1667$, $p_m = 0.6667$, and $p_u = 0.1667$. Similarly, at the highest node at the first time step ($t = 1.5$ and $x = 0.6364$), $x + M = x - ax \times 0.1 = 0.5728$, $V = 0.30^2 \times 0.1 = 0.009$, and $(x + M)/\Delta x_{i+1} = 3.486$, so $k = 3$ and $\varepsilon = (0.5728 - 3 \times 0.1643)/0.1643 = 0.4857$. The results for every node are in **Table 2**, and the shape of the tree is shown in **Figure 1**.

Adjusting the Tree. The final stage of the tree-building process involves adding the function $g(t)$ to the value of x at each node. Because $g(t)$ is a

Table 2. Tree-Branching Calculation

t	x	M	V	k	ε	p_u	p_m	p_d
0	0	0	0.135	0	0.0000	0.1667	0.6667	0.1667
1.5	0.6364	–0.0636	0.009	3	0.4857	0.5275	0.4308	0.0418
1.5	0.0000	0.0000	0.009	0	0.0000	0.1667	0.6667	0.1667
1.5	–0.6364	0.0636	0.009	–3	–0.4857	0.0418	0.4308	0.5275
1.6	0.6573	–0.2629	0.036	1	0.2000	0.2867	0.6267	0.0867
1.6	0.4930	–0.1972	0.036	1	–0.1000	0.1217	0.6567	0.2217
1.6	0.3286	–0.1315	0.036	1	–0.4000	0.0467	0.5067	0.4467
1.6	0.1643	–0.0657	0.036	0	0.3000	0.3617	0.5767	0.0617
1.6	0.0000	0.0000	0.036	0	0.0000	0.1667	0.6667	0.1667
1.6	–0.1643	0.0657	0.036	0	–0.3000	0.0617	0.5767	0.3617
1.6	–0.3286	0.1315	0.036	–1	0.4000	0.4467	0.5067	0.0467
1.6	–0.4930	0.1972	0.036	–1	0.1000	0.2217	0.6567	0.1217
1.6	–0.6573	0.2629	0.036	–1	–0.2000	0.0867	0.6267	0.2867

Figure 1. Tree-Branching Structure



function of $\theta(t)$ and the function $\theta(t)$ was selected so that the model would fit the term structure, the de facto process is to adjust the nodes in the tree so that it correctly prices discount bonds of all maturities. This adjustment is carried out in a sequential process starting at the root node.

We denote node (i,j) as the node on the tree at time t_i for which $x = j\Delta x_i$ $(0 \le i \le n; -m_i \le j \le m_i)$ and define

g_i	$= g(t_i)$
$x_{i,j}$	= value of x at node (i,j)
$f_{i,j}$	= value of $f(r)$ at node (i,j); it is $x_{i,j} + g_i$
$r_{i,j}$	= interest rate at node (i,j); it is $f^{-1}(x_{i,j} + g_i)$
$p(i,j \mid i-1,k)$	= the probability of transiting from node $(i-1,k)$ to node (i,j)
$Q(i,j \mid h,k)$	= value at node (h,k) of a security that pays off \$1 at node (i,j) and nothing at any other node[6]
$Q_{i,j}$	$= Q(i,j \mid 0,0)$

The variable $Q(i,j \mid h,k)$ is known as an Arrow–Debreu (AD) price. We will refer to the $Q_{i,j}$ price as the root AD price for node (i, j).

The root AD price for node (i, j) can be determined once the root AD prices for all nodes at time t_{i-1} have been determined. To demonstrate, we note that

$$Q(i,j \mid i-1,k) = p(i,j \mid i-1,k)\exp[-r_{i-1,k}(t_i - t_{i-1})] \tag{8a}$$

and

$$\begin{aligned} Q_{i,j} &= \sum_k Q(i,j \mid i-1,k)Q_{i-1,k} \\ &= \sum_k p(i,j \mid i-1,k)\exp[-r_{i-1,k}(t_i - t_{i-1})]Q_{i-1,k}, \end{aligned} \tag{8b}$$

where the summation is over all nodes at time step t_{i-1}.

Now, consider a discount bond that pays \$1 at every node at time step $i+1$. Let P_{i+1} be the price at node (0,0) of this discount bond and let $V_{i,j}$ be the value of this bond at node (i,j). The process for determining adjustment g_i at time step i involves two stages. First, we determine $Q_{i,j}$ for every node j at time step i. Using these root AD prices, we then compute the value of P_{i+1}. Because the discount bond pays \$1 at every node at t_{i+1}, the value at the (i,j)th node is

$$\begin{aligned} V_{i,j} &= \exp[-r_{i,j}(t_{i+1} - t_i)] \\ &= \exp[-f^{-1}(x_{i,j} + g_i)(t_{i+1} - t_i)], \end{aligned} \tag{9}$$

and the present value is

$$\begin{aligned} P_{i+1} &= \sum_j Q_{i,j}V_{i,j} \\ &= \sum_j Q_{i,j}\exp[-f^{-1}(x_{i,j} + g_i)(t_{i+1} - t_i)]. \end{aligned} \tag{10}$$

The value of g_i is adjusted until the value computed by using Equation 10 matches the price of the discount bond computed from the current term structure.

The implementation of this two-stage process proceeds in the following way. The value of a security that pays \$1 at the root node is \$1, so $Q_{0,0} = 1$. Based on the value of $Q_{0,0}$, we use Equation 10 to compute g_0 to match the price of a discount bond maturing at t_1. This process allows us to use Equation 8b to compute $Q_{1,j}$ for every node j, which then allows us to use Equation 10 to compute g_1 and so on.

To complete the illustration of the tree-building process, we now fit our example tree to a term structure. Suppose that the term structure of continuously compounded discount bond yields is given in **Table 3** and that $x = f(r) = \ln r$. In this case, the inverse function is $r = f^{-1}(x) = e^x$. The tree-adjustment process is to first set $Q_{0,0}$ equal to 1. Then, solving Equation 10 at the root node,

$$\begin{aligned} P_1 &= Q_{0,0}\exp[-f^{-1}(x_{i,j} + g_i)(t_{i+1} - t_i)], \\ 0.9277 &= \exp[-\exp(0 + g_0)(1.5)], \end{aligned}$$

we find $g_0 = -2.9957$ and $r_{0,0} = f^{-1}(x_{0,0} + g_0) = \exp(-2.9957) = 0.05$. This rate is used to calculate

Table 3. Term-Structure of Continuously Compounded Discount Bond Yields

Time to Maturity	Yield	Bond Price
1.5 years	5.00%	\$0.9277
1.6	5.10	0.9216
2.0	5.25	0.9003
2.5	5.30	0.8759

$$Q_{1,1} = Q_{0,0}p_u\exp(-r_{0,0} \times 1.5) = 0.1546,$$

$$Q_{1,0} = Q_{0,0}p_m\exp(-r_{0,0} \times 1.5) = 0.6185,$$

and

$$Q_{1,-1} = Q_{0,0}p_d\exp(-r_{0,0} \times 1.5) = 0.1546,$$

where the probabilities p_u, p_m, and p_d are the probabilities of transiting from the root node to the three nodes at the first time step. With these solutions in hand, we use Equation 10 to find g_1 and so on. The results of the calculations are in **Table 4**.

The construction of the tree for a lognormally distributed short rate that exactly fits the term structure is now complete. Notice that the functional form $f(r)$ comes into play only at the stage when the term structure is being fit (although, as we will show, it does have an impact on the volatility parameters chosen). Prior to fitting the term structure, the tree-building process is completely generic. Also note that when the tree was being fit to the term structure, to compute the interest rates at the fourth time step, we had to specify a fifth time step at time 2.5 years. This additional specification was necessary to allow us to define the term of the rates that were being determined at the fourth time step. In this case, they were 0.5-year rates.

Calibration

Calibration is the process of determining the volatility parameters that are used in the term-structure model. It is analogous to selecting the volatility that will be used when implementing the Black–Scholes model to price equity options. In the case of the general Hull–White model, the volatility parameters that are to be chosen are the functions $a(t)$ and $\sigma(t)$. The procedure is to choose volatility parameters so that the tree implementation of the term-structure model accurately replicates the market prices of actively traded options. Specifically, we use a numerical procedure, such as the Levenberg–Marquardt algorithm (see Press, Teukolsky, Vetterling, and Flannery 1993), to find the set of volatility parameters that minimizes the sum of the squares of the differences between the model prices and market prices for these options.

Because the volatility parameters are functions, we must parameterize them before we can start the calibration process. Typically, we approximate the volatility functions with piecewise linear functions, which corresponds to selecting a set of times $T_0, T_1, T_2, \ldots, T_m$, where $T_0 = 0$, $T_i > T_{i-1}$. Then, we define the reversion rate function as

$$a(t) = \alpha_i + \beta_i t, \text{ with } T_i \leq t < T_{i+1}, \quad (11)$$

where α_i is the intercept and β_i is the slope of the ith line segment. To ensure continuity of the line segments, we require that

$$\alpha_i + \beta_i T_{i+1} = \alpha_{i+1} + \beta_{i+1} T_{i+1},$$

$$\beta_0 = 0,$$

and

$$\beta_m = 0.$$

The first condition ensures that the function is continuous, and the second and third ensure that it is constant in the first time interval and beyond the last specified date.[7] These constraints ensure that the parameter set has m degrees of freedom. The volatility function is defined in an analogous way as

$$\sigma(t) = \gamma_i + \delta_{i,t}, \text{ with } T_i \leq t < T_{i+1}, \quad (12)$$

where γ_i is the intercept and δ_i is the slope of the ith line segment. To ensure continuity of the line segments, we require that

$$\gamma_i + \delta_i T_{i+1} = \gamma_{i+1} + \delta_{i+1} T_{i+1},$$

$$\delta_0 = 0,$$

Table 4. Fitting the Tree to the Term Structure

$r_{i,j}$ (percent)				$Q_{i,j}$				$V_{i,j}$			
		10.664				0.0806				0.9582	
		9.048				0.0658				0.9645	
		7.677	10.238			0.0064	0.0302			0.9698	0.9501
	11.663	6.514	7.370		0.1546	0.1024	0.2023		0.9884	0.9743	0.9638
5.000	6.172	5.527	5.306	1.0000	0.6185	0.4098	0.4306	0.9277	0.9938	0.9781	0.9738
	3.266	4.689	3.820		0.1546	0.1024	0.2059		0.9967	0.9814	0.9811
		3.979	2.750			0.0064	0.0313			0.9842	0.9863
		3.376				0.0664				0.9866	
		2.864				0.0813				0.9886	

g_0	g_1	g_2	g_3
–2.9957	–2.7851	–2.8956	–2.9364

and

$$\delta_m = 0.$$

The choice of the number of corner points in the volatility functions and the times at which the corners should be placed is more of an art than a science. The more corner points, the more degrees of freedom and the better the fit to observed market prices. Often, the number and timing of the corner points are determined by the terms of the options used in the calibration. If we have m calibrating options with m distinct maturity dates, then holding one volatility function constant (usually the reversion rate) and choosing the corner points of the other to be the option maturity times ensures that we can fit the option prices exactly.

The most common source of option prices for calibration purposes is quotes from brokers on European-style swap options, caps, and floors. **Table 5** shows a typical panel of U.S. dollar swap option quotes. The table contains the volatilities for a range of at-the-money swap options. If these volatilities are used in the standard Black swap-option-pricing model, they result in the mid-market prices for the options. The market prices of the options range from $0.12 for the 30-day option on a $100 notional 1-year swap to $5.45 for the 5-year option on a 10-year swap.

The results of fitting both the normal and the lognormal versions of the model to these data, with only a single reversion rate and a single volatility, are shown in **Table 6**, which provides the best-fit reversion rate, the best-fit volatility, and the root mean square pricing error (RMSE).[8] The fit of the model to the option prices is moderately good for both versions of the model, although the normal version fits somewhat better than the lognormal version. The mean absolute percentage pricing error (the average of the absolute price error divided by the market price) is about 2.5 percent. Those who are not familiar with the various forms of term-structure models should also note that the magnitude of the volatility parameter depends on the functional form of the model. In the normal model, the volatility parameter corresponds to the standard deviation of annual changes in the short-term rate of interest, whereas in the lognormal model, it is the standard deviation of proportional changes in the rate. Thus, if interest rates are about 7 percent, a 1.4 percent annual standard deviation roughly corresponds to an annual standard deviation of proportional changes of 20 percent.

To improve the fit, we can use more volatility parameters. **Table 7** shows the results of increasing the parameter set so that there is a corner in both the reversion rate and volatility functions at every option maturity date. Comparing Tables 6 and 7 shows that increasing the number of volatility

Table 5. Mid-Market Volatilities for At-the-Money Swap Options, August 6, 1999

	Swap Life (years)						
Option Life	1	2	3	4	5	7	10
30 days	19.00	19.50	19.50	19.50	19.50	19.50	19.50
3 months	19.50	20.13	20.13	20.13	19.98	19.98	19.98
6 months	19.90	19.75	19.75	19.70	19.60	19.50	19.50
1 year	21.55	20.80	20.20	19.90	19.60	19.20	18.78
2 years	21.30	20.40	19.85	19.30	19.00	18.70	18.20
3 years	20.80	19.75	19.20	18.85	18.60	18.20	17.63
4 years	20.43	19.20	18.80	18.40	18.10	17.60	17.03
5 years	19.85	18.73	18.28	17.93	17.58	16.98	16.43

Note: The swap is assumed to start at the expiration of the option, so the total life of the transaction is the sum of the option life and the swap life.

Table 6. Best-Fit Volatility Parameters for Normal and Lognormal Versions of the Model: Two Volatility Parameters

Model	Reversion Rate (a)	Volatility (σ)	RMSE
Normal	0.0267	0.0146	0.0564
Lognormal	0.0243	0.2093	0.0745

Table 7. Best-Fit Volatility Parameters for Normal and Lognormal Versions of the Model: Sixteen Volatility Parameters

	Normal		Lognormal	
Date	$a(t)$	$\sigma(t)$	$a(t)$	$\sigma(t)$
9/05/99	0.1878	0.0147	0.0487	0.2144
11/05/99	0.0205	0.0135	0.0596	0.2137
2/04/00	0.0010	0.0135	0.0007	0.1669
8/05/00	0.0010	0.0136	0.0002	0.2261
8/05/01	0.0003	0.0133	0.0005	0.1513
8/05/02	0.0003	0.0132	0.0002	0.2199
8/05/03	0.0010	0.0130	0.0006	0.1436
8/04/04	0.0212	0.0130	0.0140	0.2071
RMSE	0.0310		0.0292	

parameters from 2 to 16 does improve the fit, but not dramatically so. The volatility parameter for the normal model is relatively constant, and the reversion rate changes only five times, which suggests that about the same fit could be achieved with far fewer parameters. In the lognormal model, in contrast, both $a(t)$ and $\sigma(t)$ are highly variable.

Some experimentation reveals that fitting this full panel of option prices is not possible by using our model or, indeed, any one-factor Markov model of the term structure. As a result, when these types of models are used in practice, they are calibrated in the same way that models for pricing equity and forward exchange options are calibrated. A different volatility parameter set is used for every option or for every type of option.

Usually, the volatilities of the European options that are used to hedge the option in question will be used for calibration. For example, a common use of these models is the pricing of Bermudan swap options. To calibrate our model to price Bermudan swap options, we would use a diagonal strip of the volatilities from Table 5. If we were interested in pricing a five-year Bermudan swap option, we would note that if it is exercised at the one-year point in its life, it is similar to a one-year European option on a four-year swap. Similarly, exercising at the two-year point is similar to a two-year European option on a three-year swap, and so on. As a result, we would use the 1×4, 2×3, 3×2, and 4×1 swap option volatilities to calibrate the model and would probably use these options to hedge the Bermudan option. By using four volatility parameters, we could exactly fit the calibrating option prices with our model and achieve a good hedge—or at least a good hedge for the prices calculated by the model.

Supercalibration

In the previous section, we discussed how the volatility parameters for a particular form of the model can be determined from market prices of options. In this section, we describe how the functional form of the model can also be determined from the market prices of options.

Black's model, the market standard for caps and European swap options, assumes that interest rates are lognormally distributed. If rates really were lognormally distributed, the volatility used to price a cap or a swap option would be independent of the option strike rate. The U.S. dollar cap market has developed to the point that brokers are now able to provide volatility quotes for in- and out-of-the-money caps and floors. The usual practice is to provide at-the-money volatility quotes for the standard set of caps and to provide a table of spreads to be added to the volatilities of in- and out-of-the-money caps. A typical set of broker quotes is shown in **Table 8**.

Because the market volatilities for caps and floors are not independent of their strike rates, we can conclude that the lognormal assumption does not reflect the market perception of the distribution of rates. Table 8 shows that volatilities for in-the-money caps are significantly higher than those for at-the-money caps. Except for very long maturities, out-of-the-money caps also have somewhat higher volatilities than at-the-money caps. The market's perception is, therefore, that very low rates and (to a lesser extent) very high rates are more likely than the lognormal distribution would suggest.

The term-structure model implied by Equation 1 assumes that some function of the short rate $x = f(r)$ follows a normal mean-reverting process. To understand the role that the functional form $f(r)$ plays, note that the process that short rate r obeys is

Table 8. Volatility Adjustments for In- and Out-of-the-Money Caps and Floors for July 27, 1999

		At-the-Money (ATM) Cap Strike Rate (%)							
Cap Life (years)	ATM Volatility	–3	–2	–1	–0.5	0.5	1	2	3
1	14.88	—	—	1.00	0.50	0.00	1.00	—	—
2	18.38	3.00	2.00	1.00	0.50	0.50	1.00	1.25	1.50
3	19.19	3.15	2.15	1.15	0.75	0.70	0.75	1.10	1.10
4	19.50	3.50	2.50	1.50	0.75	0.50	0.50	1.00	1.00
5	19.50	3.00	2.00	1.20	0.80	0.00	0.50	1.00	1.00
7	18.88	3.00	2.00	1.00	0.50	0.00	0.00	0.00	0.00
10	18.19	3.00	2.00	1.00	0.50	0.00	–0.25	–0.50	–0.50

$$dr = \cdots dt + \frac{\partial h(x)}{\partial x}\sigma(t)dz, \quad (13)$$

where h is the inverse of the function f; that is, $r = h(x)$. The primary effect of the choice of the functional form is its impact on the volatility component of this process, $\sigma(t)\partial h(x)/\partial x$. This choice of the functional form determines the relationship between the level of rates and the variability of rates.

We now propose a more general model in which $\sigma(t)\partial h(x)/\partial x = \sigma(t)s(r)$ for some function of the level of rates $s(r)$. The function $\sigma(t)s(r)$ is known as the local standard deviation of the rate, and $\sigma(t)s(r)/r$ is the local volatility.

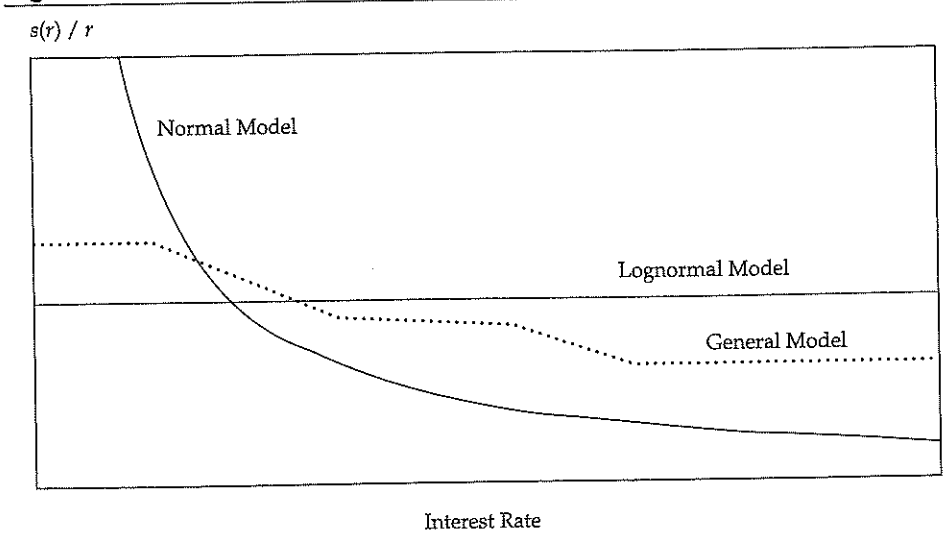
So, far, we have considered two cases:

1. $x = f(r) = r$ or $r = h(x) = x$, for which $\sigma(t)s(r) = \sigma(t)$, rates always have the same level of variability, and future rates are normally distributed. This model is the original Hull–White model.
2. $x = \ln(r)$ or $r = \exp(x)$, for which $\sigma(t)s(r) = \sigma(t)r$, the variability of rates is proportional to the level of rates, and rates are lognormally distributed. This model is the Black–Karasinski model.

These two models have $s(r) = 1$ and $s(r) = r$. Just as the volatility functions, $a(t)$ and $\sigma(t)$, are constructed as piecewise linear functions, $s(r)$ can also be constructed as a piecewise linear function. For this construction, we select a number of different rates $r_i > 0$ for $i = 1, 2, \ldots, n$ and the corresponding values of $s(r_i)$, namely, $s_i > 0$ for $i = 1, 2, \ldots, n$. We usually force $s(r)$ to pass through the origin, which ensures that as r becomes small, the variability of rates vanishes and negative rates do not occur. The form of $s(r)/r$ for the three models is shown in **Figure 2.**

The selection of the values of s_i for $i = 1, 2, \ldots, n$ now becomes part of the calibration exercise. We will choose the values that result in a term-structure-model implementation that most closely replicates the market prices of the options. Our least-squares best-fit criterion is the same as before. Because the variability of the short rate in Equation

Figure 2. Relationship between Level of Rates and Local Volatility



13 is $\sigma(t)s(r)$, we cannot determine the forms of $\sigma(t)$ and $s(r)$ simultaneously. As a result, we will first find the $\sigma(t)$ that best fits the at-the-money options and then, holding that function fixed, find the $s(r)$ that best fits the prices of the in- and out-of-the-money options.

To illustrate the effect of calibrating the functional form to the volatility of in- and out-of-the-money options, we set $\sigma(t) = 1$ and find the best $s(r)$ to fit the prices of 3-year caps and floors. We set the corner points of $s(r)$ at the at-the-money rates of ±0.5 percent, ±1 percent, and ±2 percent. And we repeat this process for the 7-year and 10-year caps and floors. The best-fit functional form of the local volatility for each of the three maturities is shown in **Figure 3**. The overall result is not surprising. To raise the price (and implied volatility) of in- and out-of-the-money caps and floors, we have to increase the local volatility as we move away from the money. The shorter the life of the option, the more extreme the adjustment.

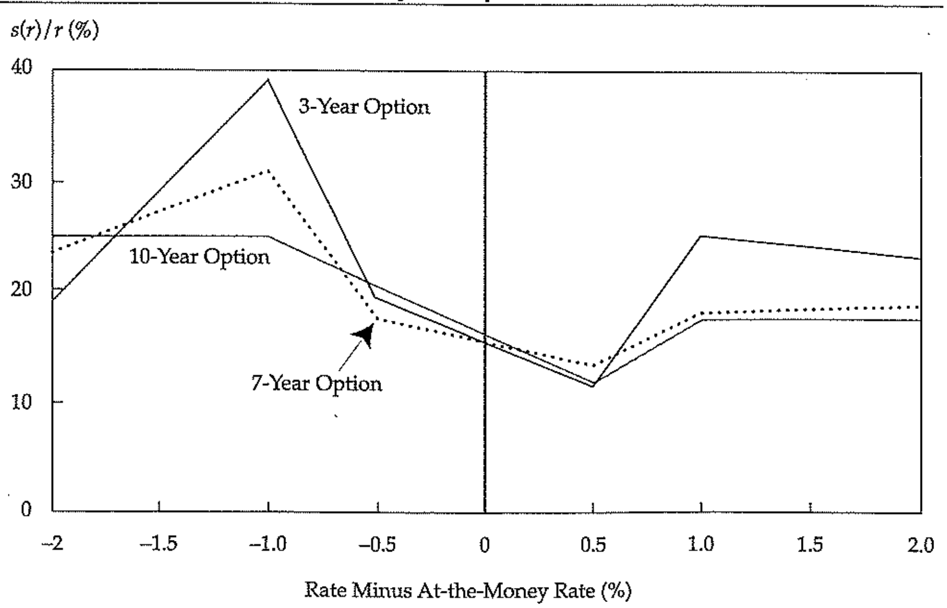
Conclusion

We have explained how a general model of the short rate can be implemented and calibrated to market data. The calibration process includes the selection of the functional form of the term-structure model that best fits the prices of in- and out-of-the-money options. Although not discussed in this article, the supercalibration process is also useful in economies like Japan's, where interest rates are very low. In this situation, if a normal model is used, the probability of rates becoming negative is large and if a lognormal model is used, the volatilities must be in excess of 100 percent to capture the observed variability of rates. A lognormal model with these large volatilities implies that rates will become extremely variable when they rise above 1 percent. This issue is discussed in more detail in Hull and White (1997).

The supercalibration procedure we described is in the same spirit as the implied-tree methodology for equity options developed by Derman, Kani, and Chriss (1996) and Rubinstein (1994). These authors made the local volatility of the stock price a function of time and the stock price, and they developed procedures to infer the local volatility from option prices. The supercalibration procedure also suffers from the same weakness as the implied-tree methodology, namely, that we are adding many free parameters to the model to force it to fit a complex data set. The result is not a model that more accurately reflects the way the term structure actually evolves; it is a model that better reproduces observed market prices.

Views as to what is best in fitting a model to data range widely. At one extreme is the "academic's view" that simple, stationary models are best. Thus, the volatility parameters should not be functions of time and the functional form of the model should not change over time. The behavior of models with these properties will be the same in the future as it is now. If we restrict ourselves to stationary models, however, we can fit observed market prices only approximately. At the other

Figure 3. Best-Fit Local Volatility of Caps and Floors



extreme is the "trader's view" that the model should exactly fit all observed option prices. If this requirement is pursued, many free volatility parameters must be estimated, the model becomes highly nonstationary, and the future behavior of the model may differ a great deal from its current behavior. In particular, the future option volatilities implied by the model may be very different from the volatilities of today. Our view is that a moderate approach should be taken in fitting a model to observed option prices. Modest nonstationarity does not seriously affect the future behavior of the model and allows a good fit to today's prices.

Notes

1. A cap (floor) is an option that sets a predetermined maximum (minimum) on a floating rate of interest. For example, an interest rate swap with a cap (floor) places a maximum (minimum) on the interest rate paid on the floating rate leg.
2. An American option can be exercised on any business day after purchase through the expiration of the option, whereas a European option can be exercised only at the end of its life, on its expiration date. Bermudan options are exercisable only on specific dates; for example, many swap options can be exercised only on any swap payment date.
3. When the reversion rate is constant, the form of g is $g(t) = g(0)e^{-at} + \int_0^t \theta(s)e^{-a(t-s)}ds$. Although this equation looks ominous, we do not actually ever have to determine its exact form. The addition of this function to the process is simply a device that simplifies the implementation.
4. The node spacing can be set to $\Delta x_i = \sigma(t_{i-1})\sqrt{n(t_i - t_{i-1})}$ for a range of values of n without impairing the numerical procedure. We chose $n = 3$ because it allows the numerical procedure to exactly replicate the first five moments of the distribution of $x(t_i) \mid x(t_{i-1})$ when the reversion rate is zero. This value produces a slightly more rapid convergence than do other values of n.
5. A reasonable approximation is $M = -xa(t_i)(t_{i+1}-t_i) = -j\Delta x_i a(t_i)(t_{i+1} - t_i)$ and $V = \sigma^2(t_i)(t_{i+1} - t_i)$. When a and σ are constant, more exact calculations are possible.
6. The value of any security with deterministic payoffs can be easily computed using the Arrow–Debreu prices: Letting $C_{i,j}$ be the payment received at the (i,j)th node, the value of the security at the (h,k)th node is then $\sum_{i>h}\sum_j Q(i,j|h,k)C_{i,j}$, where the summation is taken over all time steps i later than h and all nodes j at each time step.
7. Neither of these conditions is required. We use them only because of a belief that the volatility functions should be continuous and bounded. An alternative parameterization that seems to work well is a step function in which the parameters are piecewise constant. Note that the time divisions used for the two volatility functions do not need to be the same.
8. The root mean square error is defined as $\sqrt{\sum_{i=1}^{n}(P_{model} - P_{market})^2/n}$, where n is the number of option prices being fit.

References

Black, F., and P. Karasinski. 1991. "Bond and Option Pricing When Short Rates Are Lognormal." *Financial Analysts Journal*, vol. 47, no. 4 (July/August):52–59.

Brace, A., D. Gatarek, and M. Musiela. 1997. "The Market Model of Interest Rate Dynamics." *Mathematical Finance*, vol. 7, no. 2 (April):127–155.

Cox, J., S. Ross, and M. Rubinstein. 1979. "Option Pricing: A Simplified Approach." *Journal of Financial Economics*, vol. 7, no. 3 (September):229–264.

Derman, E., I. Kani, and N. Chriss. 1996. "Implied Trinomial Trees of the Volatility Smile." *Journal of Derivatives*, vol. 3, no. 4 (Summer):7–22.

Heath, D., R. Jarrow, and A. Morton. 1992. "Bond Pricing and the Term Structure of Interest Rates: A New Methodology." *Econometrica*, vol. 60, no. 1 (January):77–105.

Ho, T.S.Y., and S.B. Lee. 1986. "Term Structure Movements and Pricing Interest Rate Contingent Claims." *Journal of Finance*, vol. 41, no. 5 (December):1011–29.

Hull, J., and A. White. 1990. "Pricing Interest Rate Derivatives Securities." *Review of Financial Studies*, vol. 3, no. 4 (Winter):573–592.

———. 1997. "Taking Rates to the Limit." *Risk*, vol. 10, no. 12 (December):168–169.

Jamshidian, F. 1997. "LIBOR and Swap Market Models and Measures." *Finance and Stochastics*, vol. 1, no. 4:293–330.

Miltersen, K., K. Sandmann, and D. Sondermann. 1997. "Closed Form Solutions for Term Structure Derivatives with Log-Normal Interest Rates." *Journal of Finance*, vol. 52, no. 1 (March):409–430.

Pelsser, A.A.J. 1996. "Efficient Methods for Valuing and Managing Interest Rate and Other Derivative Securities." Ph.D. dissertation, Erasmus University, Rotterdam, Netherlands.

Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. 1993. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. Cambridge, U.K.: Cambridge University Press.

Rubinstein, M. 1994. "Implied Binomial Trees." *Journal of Finance*, vol. 49, no. 3 (July):771–818.