A Taxonomy of Interest Rate Models and Calibration Techniques

ISI Technical Note
Inductive Solutions, Inc.

Interest rates are conveniently modeled as solutions to stochastic differential equations. This Note discusses spot rate models in Section A, forward rate models in Section B, and Calibration techniques in Section C. (Details regarding specific models can be found by looking up the model's name or author in the in the *Risk Management References*.)

A. Spot Rate Models

Suppose we know the instantaneous spot rate path $r(s)$ for all s in an interval $[t, T]$. Let $B(t;T)$ denote the price (at time t) of a riskless zero-coupon bond that matures at time T given an interest rate path. Then the value of the bond, given that path is

$$B(t;T) = e^{-\int_t^T r(s)ds}$$

Given a set of interest rate paths generated by the risk-neutral probability of occurrence, the expected value of the bond is approximated by

$$\bar{B}(t;T) = \mathrm{E}[B(t;T)\,|\,r^1(s),\ \ldots\ , r^N(s)] = \tfrac{1}{N}\sum_{r(s)} B(t;T)$$

Spot rate models are used to generate the interest rate paths. Most assume that the instantaneous spot rate r depends on a number of component rates $r_1, r_2, \ldots, r_n$. These component rates, packaged into an n-vector $\boldsymbol{r} = \langle r_1, \ldots, r_n \rangle$ typically denote some observable economic activity. Depending on the precise model, the spot rate r is either the first component $r = r_1$ of vector $\boldsymbol{r}$ or is the sum $r = r_1 + r_2 + \ldots + r_n$ of the components. In either case, the spot rate depends on the evolution of itself or other components.

This evolution is specified by the system of n-stochastic differential equations: a rule that shows how small changes in $\boldsymbol{r}$ (denoted by $\boldsymbol{dr}$) evolve in a small instant of time dt. The system takes the form

$$\boldsymbol{dr} = \boldsymbol{(m - L)}\, dt + \boldsymbol{S} \bullet \boldsymbol{dZ} \qquad \textbf{(S1)}$$

or, for each component r_j, $j = 1 \ldots n$:

$$dr_j = m_j(r_1, \ldots, r_n ; t)\, dt + s_{j1}(r_1, \ldots, r_n ; t) dZ_1 + \ldots + s_{jk}(r_1, \ldots, r_n ; t)\, dZ_k \qquad \textbf{(S2)}$$

Here $\boldsymbol{r}$, $\boldsymbol{dr}$, $\boldsymbol{m}$, and $\boldsymbol{L}$ are n-vectors; $\boldsymbol{S}$ is the *n x k diffusion* matrix (k is the number of *factors* in the model; n is the order of the model), and • denotes the usual vector inner product. The parameter $\boldsymbol{L}$, the market price of risk, is an economic function that calibrates the model parameters (see the discussion in Section 3 on calibration) to the observed market interest rates.

The parameters $\boldsymbol{m}$, $\boldsymbol{L}$, and $\boldsymbol{S}$ can all be general functions of time t and vector $\boldsymbol{r}$. A model is called *time-independent* (or *stationary*, or an *Ito diffusion*) when these parameters do not depend explicitly on time t. Note that we can eliminate an explicit dependency on time from a model by setting it to another component factor: $r_{n+1} = t$. Consequently, our taxonomy is reduced by considering only stationary multifactor models.

The random activity is concentrated in the k-vector $\boldsymbol{dZ} = \langle dZ_1, ..., dZ_k \rangle$ which denotes a k-factor Brownian motion. The n x n instantaneous *covariance matrix* of the n-interest rate components is the matrix product of $\boldsymbol{S}$ with its transpose:

$$\Sigma = (\boldsymbol{S}\,\boldsymbol{S}^{T}), \text{ with } (\Sigma_{ij}) = \mathbf{E}[dr_i \; dr_j].$$

Note that when computing covariances, $dZ_i\, dZ_j = 1$ for $i=j$ and 0 otherwise. Given a rate vector $\boldsymbol{r} = \langle r_1, ..., r_n \rangle$, the simulated spot rate r is (depending on the model) either the first component of the rate vector ($r = r_1$) or the sum of independent rate vector component rates ($r = r_1 + r_2 + \ldots + r_n$).

The stationary multifactor spot models as indicated in **(S1)-(S2)** include the following:

1. Independent Multifactor Spot Model: $r = r_1 + r_2 + \ldots + r_n$.
   $m_j = m_j(r_{1,\ldots,} r_n)$ (Arbitrary supplied functions)
   $S_{jm} = s_j(r_{1,\ldots,} r_n)$ (Arbitrary supplied functions)
   $L_j = L_j(r_{1,\ldots,} r_n)$ (Arbitrary supplied functions)

2. Multifactor Vasicek: $r = r_1 + r_2 + \ldots + r_n$.
   $m_j = a_j(b_j - r_j)$
   $S_{jm} = s_j$ ($\boldsymbol{S}$ is a constant diagonal matrix.)
   L_j = a constant

3. Multifactor Cox-Ingersoll-Ross: $r = r_1 + r_2 + \ldots + r_n$.
   $m_j = a_j(b_j - r_j)$
   $S_{jm} = s_j\sqrt{r_j}$ (**S** is a constant diagonal matrix.)
   L_j = a constant

4. Multifactor Black-Derman-Toy-Karasinski: $r = r_1 + r_2 + \ldots + r_n$.
   $m_j = -r_j a_j(b_j - \log r_j + \frac{1}{2}s_j^2)$
   $S_{jm} = s_j\sqrt{r_j}$ (**S** is a constant diagonal matrix.)
   L_j = a constant

5. Multifactor Rendelman-Bartter: $r = r_1 + r_2 + \ldots + r_n$.

$m_j = a_j (b_j - r_j)$
$S_{jm} = s_j r_j$ (***S*** a constant diagonal matrix)
L_j = a constant

6. Multifactor Chan-Karoly-Longstaff-Sanders: $r = r_1 + r_2 + \ldots + r_n$.

$m_j = a_j (b_j - \frac{1}{2} a_j r_j^2)$
$S_{jm} = s_j r_j \sqrt{r_j}$ (**S** is a constant diagonal matrix.)
L_j = a constant

7. Multifactor Sandmann-Sondermann: $r = r_1 + r_2 + \ldots + r_n$.

$m_j = a_j (b_j - \frac{1}{2} a_j s_j^2)$
$a_j = 1 - e^{-r_j}$
$S_{jm} = s_j$ (***S*** is a constant diagonal matrix.)
L_j = a constant

8. Non-Parametric QES/QRA: $r = r_1$

$m_j = m_j(r_1, ..., r_n)$ (Non-parametrically estimated.)
$S_{jm} = s(r_1, ..., r_n)$ (Non-parametrically estimated.)
$l_j = l_j (r_1, ..., r_n)$ (Non-parametrically estimated.)

B. Forward Rate Models

Let $B(t;T)$ denote the price (at time t) of a riskless zero-coupon bond that matures at time T. Suppose we know all instantaneous forward rates $f_j(t,T_j)$:

$$f_j(t,T_j) = \frac{\partial}{\partial T} \log B(t;T) \approx \frac{1}{\Delta} \log \frac{B(t;T+\Delta)}{B(t;T)}$$

An infinite set of instantaneous forward rates can be used to generate the entire yield curve. Given n-forward rates packaged as n-vector $\boldsymbol{f} = < f_1(t,T_1), ..., f_n(t,T_n)>$, the spot rate path r is $r(t) = f(t,t)$. The spot rate path $r(t)$ for $t = T_1, T_2, ..., T_n$ can be approximated by the evolving forward rates:

$$< r(T_1), ..., r(T_n)> = < f_1(T_1,T_1), ..., f_n(T_n, T_n)>$$

Consequently, given the spot interest rate path induced by the forward rates, the price at time t of a riskless zero-coupon bond that matures at time T is approximated by

$$B(t;T) \approx e^{-\Delta T \sum_{j=1}^{N} f_j(T_j,T_j)}$$

Given a set of spot interest rate paths generated by the risk-neutral probability of occurrence of the forward rates, the expected value of the bond is approximated by

$$\bar{B}(t;T) = \mathrm{E}[B(t;T) \,|\, r^1(s), \ \ldots\ , r^N(s)] = \tfrac{1}{N} \sum_{r(s)} B(t;T)$$

Forward rate models are used to generate the forward rates and induced spot interest rate paths.

Models for the forward rate $\boldsymbol{f}$ depend on a function of time t and maturity T. These forward models specify a rule for small changes in the yield curve: the spot rate $r(t)$ at time t is then given by $f(t,t)$. In practice, one usually divides the time interval $[0, T]$ into n segments, and form an n-vector $\boldsymbol{T} = <T_1, \ldots, T_n>$.

The forward rate models have the following form:

$$\boldsymbol{df}(t, \boldsymbol{T}) = \boldsymbol{a}(t,\boldsymbol{T})\, dt + \boldsymbol{S}(t, \boldsymbol{T};\ \boldsymbol{f}(t,\boldsymbol{T})\ ) \bullet \boldsymbol{dZ} \qquad \textbf{(F1)}$$

or, for each time interval T_j, $j=1..n$:

$$df_j(t,T_j) = a_j(t, T_j)\, dt + s_{j1}(t, T_j; f_1, ..., f_n)\, dZ_1 + \ldots + s_{jk}(t,T_j; f_1, ..., f_n)\, dZ_k \qquad \textbf{(F2)}$$

Here $\boldsymbol{f} = < f_1, ..., f_n>$, $\boldsymbol{df} = <df_1, ..., df_n>$, and $\boldsymbol{a} = < a_1, ..., a_n>$, are n-vectors; $\boldsymbol{S}$ is an

n x k matrix (*k* is the number of *factors* in the model; *n* is the order of the model, corresponding to the number of forward rates), and • denotes the usual vector inner product.

The long term non-random behavior is denoted by the drift ***a***. The constraint for the drift term **a**, established by Heath, Jarrow, and Morton is that

$$a_j(t;T_j) = < s_{j1}(t;T_j;\boldsymbol{f}) \ldots s_{jk}(t;T_j;\boldsymbol{f}) > \bullet \int_t^T < s_{j1}(s;T_j;\boldsymbol{f}) \ldots s_{jk}(s;T_j;\boldsymbol{f}) > ds$$

The random activity is concentrated in the *k*-vector $\boldsymbol{dZ} = \langle dZ_1, \ldots, dZ_k \rangle$ which denotes a *k*-factor Brownian motion; correlations and covariances of this motion among the interest rate components are specified by ***S***. The *n x n* covariance matrix of the *n* forward interest rate components is given by the matrix product of ***S*** with its transpose:

$$\Sigma = (\boldsymbol{S}\,\boldsymbol{S}^{\boldsymbol{T}}), \text{ with } (\Sigma_{ij}) = \mathbf{E}[df_i \; df_j].$$

In practice, once matrix ***S*** is specified, the drift term **a** should be computed dynamically.

The parameters ***S*** can be a general function of time *t* and vector ***f***.

The multifactor forward models as indicated in **(F1)-(F2)** include the following:

9. General Multifactor HJM Forward Rate Model
   $S_{jm} = s_{jm}(t, T_j; f_j(t, T_j))$ (Arbitrary supplied functions).

10. Multifactor Normal Forward Rate Model
    $S_{jm} = s_{jm}$ (***S*** is a constant: a function of $T_j - t$).

11. Multifactor Lognormal Forward Rate Model
    $S_{jm} = f_j(t, T_j) s_{jm}$ (***S*** is a constant: a function of $T_j - t$).

12. Multifactor Goldys-Musiela-Sondermann Forward Rate Model
    $S_{jm} = [1 - e^{-f_j(t,T_j)}] s_{jm}$ (***S*** is a constant: a function of $T_j - t$).

13. Multifactor Gaussian Forward Rate Model
    $S_{jm} = [1 - e^{-L_m(T_j - t)}] s_m$ (s_m and L_m are constants).

C. Calibration Techniques

The primary tool used to estimate the parameters in a system of stochastic differential equations are classical regesssion, Cholesky Decomposition, and Principal Component Analysis (that solves for eigenvalues and eigenvectors). These tools are used when the drift and diffusion parameters are constant, parametric, or even non-parametric.

Without loss of generality, we consider multifactor spot models only.

The stochastic differential equation can be approximated by the discretized form, i.e.,

$$\Delta r_j(t) = m_j(r_1, \ldots, r_n; t)\Delta t + s_{j1}(r_1, \ldots, r_n; t)\Delta Z_1 + \ldots + s_{jk}(r_1, \ldots, r_n; t)\Delta Z_k$$

1. Constant and Parametric Drifts and Diffusions

Suppose we have observations of interest rate component changes over a time series over *m* small time steps. We can group the $\Delta r_j(t)$ into the following matrix and vectors:

$$\begin{bmatrix} \Delta r_1(1\Delta t) & \Delta r_2(1\Delta t) & \ldots & \Delta r_n(1\Delta t) \\ \Delta r_1(2\Delta t) & \Delta r_2(2\Delta t) & \ldots & \Delta r_n(2\Delta t) \\ \ldots & \ldots & \ldots & \ldots \\ \Delta r_1(m\Delta t) & \Delta r_2(m\Delta t) & \ldots & \Delta r_n(m\Delta t) \end{bmatrix} = \begin{bmatrix} \Delta \mathbf{r_1} & \Delta \mathbf{r}_2 & \ldots & \Delta \mathbf{r}_n \end{bmatrix}$$

The means can be estimated by computing the averages of the $\Delta \boldsymbol{r}_j$ over time. The matrix $\boldsymbol{S}$ is estimated by performing Cholesky Decomposition on an estimate of the instantaneous covariance matrix S:

$$\boldsymbol{S} = Cholesky(\Sigma)$$

where the *n x n* sample instantaneous covariance matrix S is given by

$$\Sigma_{n\times n} = Cov(\begin{bmatrix} \Delta \mathbf{r_1} & \Delta \mathbf{r}_2 & \ldots & \Delta \mathbf{r}_n \end{bmatrix}) \approx \frac{1}{m} \begin{bmatrix} \Delta \mathbf{r_1}^{\bullet} \Delta \mathbf{r_1} & \Delta \mathbf{r_1}^{\bullet} \Delta \mathbf{r}_2 & \ldots & \Delta \mathbf{r_1}^{\bullet} \Delta \mathbf{r}_n \\ \Delta \mathbf{r}_2^{\bullet} \Delta \mathbf{r_1} & \Delta \mathbf{r}_2^{\bullet} \Delta \mathbf{r}_2 & \ldots & \Delta \mathbf{r}_2^{\bullet} \Delta \mathbf{r}_n \\ \ldots & \ldots & \ldots & \ldots \\ \Delta \mathbf{r}_n^{\bullet} \Delta \mathbf{r_1} & \Delta \mathbf{r}_n^{\bullet} \Delta \mathbf{r}_2 & \ldots & \Delta \mathbf{r}_n^{\bullet} \Delta \mathbf{r}_n \end{bmatrix}$$

(Note that other "higher order" estimates for the mean and instantaneous covariance can be obtained from considering the infinitesimal generator of the stochastic differential equation.)

If we want to reduce the number of factors we need to use Principal Component Analysis. In this case, we decompose the instantaneous covariance matrix S by:

$$\Sigma = \mathrm{X}^{\mathrm{T}} \Lambda \mathrm{X}$$

where X is the matrix of eigenvectors $\mathbf{x}_j$,

$$\Sigma \mathbf{x} = l_j \mathbf{x}_j, \; j = 1..n$$

and so

$$S = \mathrm{X}^{T} \Lambda^{1/2}$$

with the diagonal matrix of eigenvalues given by

$$\Lambda = \mathrm{diag}[\, l_1 \; \ldots \; l_n \,] = \begin{bmatrix} l_1 & 0 & \ldots & 0 \\ 0 & l_2 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & 0 & l_n \end{bmatrix}$$

If we sort the eigenvalues in decreasing order and re-arrange the eigenvector matrix to correspond with this order —

$$l_{j_1} \geq l_{j_2} \geq \ldots \geq l_{j_n}$$
$$\mathrm{X} = \begin{bmatrix} \mathbf{x}_{j_1} & \mathbf{x}_{j_1} & \ldots & \mathbf{x}_{j_n} \end{bmatrix}$$

—then we can effectively capture only the "principal components" of the variance by setting the smallest eigenvalues to zero, thereby reducing the effective number of factors.

For parametric models, we need to first perform a regression on the $\Delta \boldsymbol{r}_j$ and on the sample instantaneous covariance matrix with respect to the functional forms in the particular parametric models.

2. Non-Parametric Estimation of Drifts and Diffusions

Non-parametric estimation "learns" the functional forms of the drifts and diffusions. The basic idea is that a probability density function can be inferred directly from the data, and that the drifts and diffusions can be estimated via non-parametric regression.

Given a set of *m d*-dimensional vectors $\mathbf{x_1}$, ..., $\mathbf{x_m}$, a multivariate probability density function p(**x**) can be approximated by the multivariate kernel density estimator:

$$\hat{p}(\mathbf{x}) = \hat{p}(< x_1 \; \ldots \; x_D >) = \frac{1}{m h_1 h_2 \; \ldots \; h_D} \sum_{t=1}^{m} \prod_{d=1}^{D} K\left(\frac{x_d - x_d^t}{h_d} \right)$$

Here, $K(x)$ is a kernel density function and the h_d are kernel bandwidths that correspond to bin sizes in a histogram.

Choices for the kernel density function include the Gaussian Kernel

$$K(x) = \frac{1}{\sqrt{2p}} e^{-x^2/2}$$

and the Epanechnikov Kernel

$$K(x) = \frac{3}{4}(1-x^2) \text{ for } |x \le 1|; \text{ 0 otherwise}$$

One practical formula for the bandwidth (that is optimal for a normal distribution) is usually estimated in terms of the sample standard deviation s_d:

$$h_d = s_d m^{-1/(D+4)}$$

Given a set of data pairs $\{(y^t, \mathbf{x}^t)\}$, the non-parametric multivariate regression function $y = R(\mathbf{x})$ is given by

$$\begin{aligned} R(\mathbf{x}) &= R(< x_1 \ldots x_D >) \\ &= E[y \mid \mathbf{x} =< x_1 \ldots x_D >] \\ &= \frac{\sum_{t=1}^{m} y^t \prod_{d=1}^{D} K\left(\frac{x_d - x_d^t}{h_d}\right)}{\sum_{t=1}^{m} \prod_{d=1}^{D} K\left(\frac{x_d - x_d^t}{h_d}\right)} \end{aligned}$$

For estimating the parameters of spot models, we set

$$\mathbf{x}^t = \mathbf{r}(t)$$

$$y^t = \Delta r_j(t)$$

for estimating the drift, and

$$\mathbf{x}^t = \mathbf{r}(t)$$

$$y^t = \Delta r_j(t) \Delta r_l(t)$$

for estimating the instantaneous covariance. The diffusion $\boldsymbol{S}$ is estimated via Cholesky Decomposition of the instantaneous covariance estimate.

Techniques exist for reducing the dimension D. The procedure associated with Sliced Inverse Regression first uses histogram techniques to bin the y^t and effectively reduces the number of y samples from m to b:

$$\left\{y^1, y^2, \ \ldots, y^m\right\} \overset{binned}{\Rightarrow} \left\{\bar{y}^1, \bar{y}^2, \ \ldots, \bar{y}^b\right\}$$

These binned values are used to create new pairs of associated **x** values: all **x** values associated with a particular y^t are averaged together to produce a new set of b observations:

$$\left\{(\mathbf{x}^1, y^1), (\mathbf{x}^2, y^2), \ \ldots, (\mathbf{x}^m, y^m)\right\} \overset{binned}{\Rightarrow} \left\{(\bar{\mathbf{x}}^1, \bar{y}^1), (\bar{\mathbf{x}}^2, \bar{y}^2), \ \ldots, (\bar{\mathbf{x}}^b, \bar{y}^b)\right\}$$

We now look at the covariance matrix formed from b sample vectors:

$$\Sigma_{D\times D} = Cov(\bar{\mathbf{x}}^1, \bar{\mathbf{x}}^2, \ \ldots, \bar{\mathbf{x}}^b)$$

and perform Principal Component Analysis. We find that

$$\Sigma_{D\times D} = \mathrm{A}^T \Lambda \mathrm{A}$$

and use the largest eigenvalues to reduce the dimension from D to E, E<D. This transformation is used in a new regression function:

$$\begin{aligned} R_{SIR}(\mathbf{x}) &= R_{SIR}(\mathrm{A}\mathbf{x}) \\ &= R_{SIR}(\mathrm{A}_{E\times D}[x_1 \ \ldots \ x_D]) \\ &= R_{SIR}([z_1 \ \ldots \ z_E]) \\ &= E\left[y \mid \mathbf{z} = [z_1 \ \ldots \ z_E]\right] \\ &= \frac{\sum_{t=1}^{m} \bar{y}^t \prod_{e=1}^{E} K\left(\frac{z_e - z_e^t}{h_e}\right)}{\sum_{t=1}^{m} \prod_{e=1}^{E} K\left(\frac{z_e - z_e^i}{h_e}\right)} \end{aligned}$$