Draft of paper appearing in:
%A L. Ingber
%T Statistical mechanical aids to calculating term structure models
%J Phys. Rev. A
%V 42
%D 1990
%P 7057-7064

Statistical mechanical aids to calculating term structure models

Lester Ingber
Science Transfer Corporation, P.O. Box 857, McLean, VA 22101
and
Department of Mathematics, The George Washington University, Washington, DC 20052

Recent work in statistical mechanics has developed new analytical and numerical techniques to solve coupled stochastic equations. This paper describes application of the very fast simulated re-annealing and path-integral methodologies to the estimation of the Brennan and Schwartz two-factor term-structure (time-dependent) model of bond prices. It is shown that these methodologies can be utilized to estimate more complicated n-factor nonlinear models. Applications to other systems are stressed.

1. INTRODUCTION

In this paper we present an alternative methodology of very fast simulated re-annealing (VFSR) [1] to compute the parameters of term-structure models, here applied to the evolution of interest rates. The term "term-structure" here is equivalent to "time-dependent," wherein stochastic differential rate-equations are used to model these financial systems. It is also shown that the VFSR methodology is capable of handling complicated n-factor non-linear models. The advantages of using the simulated annealing methodology are: (1) Global minima in parameter space are relatively more certain than with regression fitting. (2) All parameters, including parameters in the noise, are simultaneously and equally treated in the fits, i.e., different statistical methods are not being used to estimate the deterministic parameters, then to go on to estimate noise parameters. (3) Boundary conditions on the variables can be explicitly included in the fitting process, a process not included in standard regression fits. (4) We can efficiently extend our methodology to develop n-state models, including higher order nonlinearities.

We also present an alternative method of calculating the evolution of Fokker-Planck-type equations, here in the context of describing the evolution of bond prices. Our particular non-Monte Carlo path-integral technique has proven to be extremely accurate and efficient for a variety of nonlinear systems [2,3]. To mention a few advantages: (1) A variable mesh is calculated in terms of the underlying nonlinearities. (2) Initial conditions (i.c) and boundary conditions (b.c.) typically are more easily implemented with integral, rather than with differential, equations, e.g., by using the method of images. (3) Integration is inherently a "smoothing" process, whereas differentiation is a "sharpening" process. This means that we can handle "stiff" and nonlinear problems with more ease.

2. CURRENT MODELS OF TERM STRUCTURE

There are several term-structure models of bond pricing which use interest rates as proxy variables [4-12]. For specificity, here we consider the Brennan-Schwartz (BS) model, which is developed in the variables of short- and long-term interest rates. These interest rates are assumed to follow a joint Wiener stochastic process,

$$dr = \beta_1(r, l, t)dt + \eta_1(r, l, t)dz_1 \ ,$$

$$dl = \beta_2(r, l, t)dt + \eta_2(r, l, t)dz_2 \ , \tag{1}$$

where r and l are the short- and long-term rates, respectively. β_1 and β_2 are the expected instantaneous rates of change in the short-term and long-term rates respectively. η_1 and η_2 are the instantaneous standard deviations of the processes. dz_1 and dz_2 are Wiener processes, with expected values of zero and variance of dt with correlation coefficient ρ. That is,

$$E[dz_1] = E[dz_2] = 0 \ ,$$

$$E[dz_1^2] = E[dz_2^2] = dt \ , \ E[dz_1 dz_2] = \rho dt \ , \tag{2}$$

where $E[.] \equiv <.>$ is the expectation with respect to the joint Wiener process.

BS simplified and reduced this system to

$$dr = (a_1 + b_1(l - r))dt + r\sigma_1 dz_1 \ ,$$

$$dl = l(a_2 + b_2 r + c_2 l)dt + l\sigma_2 dz_2 \ , \tag{3}$$

where $\{a_1, b_1, a_2, b_2, c_2\}$ are parameters to be estimated.

Using methods of stochastic calculus [8], BS further derived a partial differential equation for bond prices as the maturity date is approached.

$$\frac{\partial}{\partial\tau} B = (-r + f^r \frac{\partial}{\partial r} + f^l \frac{\partial}{\partial l} + g^{rr} \frac{\partial}{\partial r^2} + g^{rl} \frac{\partial}{\partial r \partial l} + g^{ll} \frac{\partial}{\partial l^2})B$$

$$= AB \ , \tag{4}$$

where the coefficients $\{f, g\}$ depend on r and l, $\tau = T - t$ for t calendar time and T the time of maturity, and A can be considered as a differential operator on B.

It may help to appreciate the importance of the BS methodology by discretizing the above partial differential equation for B, in a "mean-value" limit. That is, at a given calendar time t indexed by s, noting that $\partial/\partial\tau = -\partial/\partial t$, take

$$0 = f^r \frac{\partial B_s}{\partial r} = f^l \frac{\partial B_s}{\partial l} ,$$

$$0 = g^{rr} \frac{\partial B_s}{\partial r^2} = g^{rl} \frac{\partial B_s}{\partial r \partial l} = g^{ll} \frac{\partial B_s}{\partial l^2} ,$$

$$B_s - B_{s+1} = -r_s B_s . \tag{5}$$

This yields the popular expectations-hypothesis spot-interest estimate of bond prices, working backwards from maturity,

$$B_s = (1 + r_s)^{-1} B_{s+1} . \tag{6}$$

The important generalization afforded by BS is to include information about r and l and treat them as stochastic variables with drifts and diffusions. Then, this discretized treatment yields

$$B_{s\,rl} = (1 - A_{s\,rlr'l'})^{-1} B_{s+1\,r'l'} , \tag{7}$$

where the operator inverse of the differential operator A has been formally written, and its dependence on intermediate values of r' and l' has been explicitly portrayed. Their discretized calculation of their partial differential equation, and our discretized calculation of the path-integral representation of this model, essentially are mathematical and numerical methods of calculating this evolution of B_s.

3. DEVELOPMENT OF MATHEMATICAL METHODOLOGY

3.1. Background

The BS equations are of the more general form

$$dr/dt = f_r(r, l) + \sum_i \hat{g}_r^i(r, l)\eta_i ,$$

$$dl/dt = f_l(r, l) + \sum_i \hat{g}_l^i(r, l)\eta_i , \tag{8}$$

where the $\hat{g}$'s and f's are general nonlinear algebraic functions of the variables r and l. These equations represent differential limits of discretized stochastic difference equations, e.g., Wiener noise $dW \rightarrow \eta dt$ [13]. The resulting stochastic differential equations (s.d.e.'s) are referred to as Langevin equations [13-18]. The f's are referred to as the (deterministic) drifts, and the square of the $\hat{g}$'s are related to the diffusions (fluctuations or volatilities). In fact, the statistical mechanics can be developed for any number of variables, not just two. The η's are sources of Gaussian-Markovian noise, often referred to as "white noise." The inclusion of the $\hat{g}$'s, called "multiplicative" noise, recently has been shown to very well mathematically and physically model other forms of noise, e.g., shot noise, colored noise, dichotomic noise [19-22]. Finite-jumps diffusions also can be included [23].

These new methods of nonlinear statistical mechanics only recently have been applied to complex large-scale physical problems, demonstrating that observed data can be described by the use of these algebraic functional forms. Success was gained for large-scale systems in neuroscience, in a series of papers on statistical mechanics of neocortical interactions [24-29], and in nuclear physics [30,31]. This methodology has been used for problems in combat analyses [3,32-35]. These methods are also suggested for financial markets [36,37].

The utility of these algebraic functional forms goes further beyond their being able to fit sets of data. There is an equivalent representation to the Langevin equations, called a "path-integral" representation for the long-time probability distribution of the variables. This short-time probability distribution is driven by a "Lagrangian," which can be thought of as a dynamic algebraic "cost" function. The path-integral representation for the long-time distribution possesses a variational principle, which means that simple graphs of the algebraic cost-function give a correct intuitive view of the most likely states of the variables, and of their statistical moments, e.g., heights being first moments (likely states) and widths being second moments (uncertainties). Like a ball bouncing about a terrain of hills and valleys, one can quickly visualize the nature of dynamically unfolding r and l states.

Especially because we are trying to mathematically model sparse and poor data, different drift and diffusion algebraic functions can give approximately the same algebraic cost-function when fitting short-time probability distributions to data. The calculation of long-time distributions permits a clear choice of the best algebraic functions, i.e., those which best follow the data through a predetermined long epoch of trading. Afterwards, if there are closely competitive algebraic functions, they can be more precisely assessed by calculating higher algebraic correlation functions from the probability distribution.

The mathematical representation most familiar to other modelers is the system of stochastic rate equations, often referred to as Langevin equations. From the Langevin equations, other models may be derived, such as the times-series model and the Kalman filter method of control theory, quite popular in economics. However, in the process of this transformation, the Markovian description typically is lost by projection onto a smaller state space [38,39].

3.2. Path-Integral Lagrangian Representation

Consider a multivariate system with variance a general nonlinear function of the variables. The Einstein summation convention helps to compact the equations, whereby repeated indices in factors are to be summed over. The Itô (prepoint) discretization for a system of stochastic differential equations is defined by

$$\bar{t}_s \in [t_s, t_s + \Delta t] \equiv [t_s, t_{s+1}] \ ,$$
$$M(\bar{t}_s) = M(t_s) \ ,$$
$$dM(\bar{t}_s)/dt = M(t_{s+1}) - M(t_s) \ . \tag{9}$$

The stochastic equations are then written as

$$dM^G/dt = f^G + \hat{g}_i^G \eta^i \ ,$$
$$i = 1, \cdots, \Xi \ ,$$
$$G = 1, \cdots, \Theta \ . \tag{10}$$

The operator ordering (of the $\partial/\partial M^G$ operators) in the Fokker-Planck equation corresponding to this discretization is

$$\frac{\partial P}{\partial t} = VP + \frac{\partial(-g^G P)}{\partial M^G} + \frac{1}{2}\frac{\partial^2(g^{GG'}P)}{\partial M^G \partial M^{G'}} \ ,$$
$$g^G = f^G + \frac{1}{2}\hat{g}_i^{G'}\frac{\partial \hat{g}_i^G}{\partial M^{G'}} \ ,$$
$$g^{GG'} = \hat{g}_i^G \hat{g}_i^{G'} \ . \tag{11}$$

where a "potential" V is present in some systems.

The Lagrangian corresponding to this Fokker-Planck and set of Langevin equations may be written in a Stratonovich (midpoint) representation, corresponding to

$$M(\bar{t}_s) = \frac{1}{2}[M(t_{s+1}) + M(t_s)] \ . \tag{12}$$

This discretization can be used to define a covariant Feynman Lagrangian L_F which possesses a variational principle, and which explicitly portrays the underlying Riemannian geometry induced by the metric tensor $g_{GG'}$, calculated to be the inverse of the covariance matrix [36].

$$P = \int \cdots \int \underline{D}M \exp(-\sum_{s=0}^{u} \Delta t L_{Fs}) \ ,$$

$$\underline{D}M = g_{0_+}^{1/2}(2\pi\Delta t)^{-\Theta/2} \prod_{s=1}^{u} g_{s_+}^{1/2} \prod_{G=1}^{\Theta} (2\pi\Delta t)^{-1/2} dM_s^G \ ,$$

$$\int dM_s^G \to \sum_{\iota=1}^{N^G} \Delta M_{\iota s}^G \ , M_0^G = M_{t_0}^G \ , M_{u+1}^G = M_t^G \ ,$$

$$L_F = \frac{1}{2}(dM^G/dt - h^G)g_{GG'}(dM^{G'}/dt - h^{G'}) + \frac{1}{2}h^G{}_{;G} + R/6 - V \ ,$$

$$[\cdots]_{,G} = \frac{\partial[\cdots]}{\partial M^G} \ ,$$

$$h^G = g^G - \frac{1}{2}g^{-1/2}(g^{1/2}g^{GG'})_{,G'} \ ,$$

$$g_{GG'} = (g^{GG'})^{-1} \ ,$$

$$g_s[M^G(\bar{t}_s), \bar{t}_s] = \det(g_{GG'})_s \ , g_{s_+} = g_s[M_{s+1}^G, \bar{t}_s] \ ,$$

$$h^G{}_{;G} = h^G_{,G} + \Gamma^F_{GF}h^G = g^{-1/2}(g^{1/2}h^G)_{,G} \ ,$$

$$\Gamma^F_{JK} \equiv g^{LF}[JK, L] = g^{LF}(g_{JL,K} + g_{KL,J} - g_{JK,L}) \ ,$$

$$R = g^{JL}R_{JL} = g^{JL}g^{JK}R_{FJKL} \ ,$$

$$R_{FJKL} = \frac{1}{2}(g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN}(\Gamma^M_{FK}\Gamma^N_{JL} - \Gamma^M_{FL}\Gamma^N_{JK}) \ , \tag{13}$$

where R is the Riemannian curvature, and we also have explicitly noted the discretization in the mesh of $M_{\iota s}^G$ by ι.

Because of the presence of multiplicative noise, the Langevin system differs in its Itô (prepoint) and Stratonovich (midpoint) discretizations. The midpoint-discretized covariant description, in terms of the Feynman Lagrangian L_F, is defined such that (arbitrary) fluctuations occur about solutions to the Euler-Lagrange variational equations. In contrast, the usual Itô and corresponding Stratonovich discretizations are defined such that the path integral reduces to the Fokker-Planck equation in the weak-noise limit. The term $R/6$ in L_F includes a contribution of $R/12$ from the WKB approximation to the same order of $(\Delta t)^{3/2}$ [13].

A prepoint discretization for the same probability distribution P, gives a much simpler algebraic form,

$$M(\bar{t}_s) = M(t_s) \ ,$$

$$L = \frac{1}{2}(dM^G/dt - g^G)g_{GG'}(dM^{G'}/dt - g^{G'}) - V \ , \tag{14}$$

but the the variational principle associated with the Lagrangian L is not useful for moderate to large noise. Still, this prepoint-discretized form has been quite useful in all systems examined thus far, simply requiring a somewhat finer numerical mesh.

3.3. Interest Rates

To illustrate this methodology, the BS model is summarized by:

$$dr = [a_1 + b_1(l - r)]dt + r\sigma_1 dz_1 \ ,$$

$$dl = [l(a_2 + b_2 r + c_2 l)]dt + l\sigma_2 dz_2 \ ,$$

$$< dz_i >= 0 \ , \ i = \{1, 2\} \ ,$$

$$< dz_i(t)dz_j(t') >= dt\delta(t - t') \ , \ i = j \ ,$$

$$< dz_i(t)dz_j(t') >= \rho dt\delta(t - t') \ , \ i \neq j \ ,$$

$$\delta(t - t') = \begin{cases} 0 \, , , & t \neq t' \ , \\ 1 \, , & t = t' \ , \end{cases} \qquad (15)$$

where < . > denotes expectations.

These can be rewritten as Langevin equations (in the Itô prepoint discretization)

$$dr/dt = a_1 + b_1(l - r) + \sigma_1 r(\gamma^+ n_1 + \mathrm{sgn}\rho\,\gamma^- n_2) \ ,$$

$$dl/dt = l(a_2 + b_2 r + c_2 l) + \sigma_2 l(\mathrm{sgn}\rho\,\gamma^- n_1 + \gamma^+ n_2) \ ,$$

$$\gamma^{\pm} = \frac{1}{\sqrt{2}}[1 \pm (1 - \rho^2)^{1/2}]^{1/2} \ ,$$

$$n_i = (dt)^{1/2} p_i \ , \qquad (16)$$

where p_1 and p_2 are independent [0,1] Gaussian distributions.

The cost function C is defined from the equivalent short-time probability distribution, P, for the above set of equations.

$$P = g^{1/2}(2\pi dt)^{-1/2} \exp(-Ldt)$$

$$= \exp(-C) \ ,$$

$$C = Ldt + \frac{1}{2}\ln(2\pi dt) - \ln(g) \ ,$$

$$L = \frac{1}{2} F^{\dagger} \underline{g} F \ ,$$

$$F = \begin{pmatrix} dr/dt - (a_1 + b_1(l - r)) \\ dl/dt - l(a_2 + b_2 r + c_2 l) \end{pmatrix},$$

$$g = \det(\underline{g}) \ ,$$

$$k = 1 - \rho^2 \ . \qquad (17)$$

$\underline{g}$, the metric in $\{r, l\}$-space, is the inverse of the covariance matrix,

$$\underline{g}^{-1} = \begin{pmatrix} (r\sigma_1)^2 & \rho r l \sigma_1 \sigma_2 \\ \rho r l \sigma_1 \sigma_2 & (l\sigma_2)^2 \end{pmatrix}. \qquad (18)$$

As discussed below, the correct mesh for time, dt, in order that P represent the Langevin equations (to order $dt^{3/2}$) is

$$dt \le 1/\bar{L} \ , \tag{19}$$

where $\bar{L}$ is L evaluated with $ds/dt = dl/dt = 0$. If dt is greater than $1/\bar{L}$, then it is inappropriate to use P, and instead the path integral over intermediate states of folded short-time distributions must be calculated. It should be noted that the correct time mesh for the corresponding differential equations must be at least as small, since typically differentiation is a ‘‘sharpening’’ process. This will be noted in any discipline requiring numerical calculation, when comparing differential and integral representations of the same system.

3.4. Security Prices

BS [8] present arguments recognizing that the stochastic price of a discount bond for a given maturity date T can utilize straightforward stochastic calculus to derive a form in terms of coefficients appearing in their $r - l$ coupled stochastic equations. They use arbitrage arguments on portfolios of bonds with different maturity dates to derive zero risk conditions for the market prices of risks, λ_1 and λ_2, for short-term and long-term interest rates, respectively. By considering l as related to a bond’s price, they straightforwardly derive an arbitrage expression for λ_2. Their resulting p.d.e. is an equilibrium (mean value) equation for a pure discount-bond price B, at a given time until maturity $\tau = T - t$ and ‘‘continuous’’ coupon payment of c.

The above formulation of interest rates is used by BS to determine the parameters needed to calculate their derived partial differential equation (p.d.e.) for securities, i.e., bond prices B. Using some notation developed above, with $\{M^G ; G = r, l\}$, they obtain

$$\begin{aligned}
&\frac{\partial B}{\partial \tau} = VB + \frac{\partial(-g^G B)}{\partial M^G} + \frac{1}{2}\frac{\partial^2 (g^{GG'} B)}{\partial M^G \partial M^{G'}} \ , \\
&g^r = -(\beta_1 - \lambda_1 \eta_1) \\
&\quad = -a_1 - b_1(l - r) + \lambda_1 r \sigma_1 \ , \\
&g^l = -(\beta_2 - \lambda_2 \eta_2) \\
&\quad = -l(\sigma_2^2 + l - r) \ , \\
&(g^{GG'}) = (\underline{g})^{-1} \ , \\
&V = \frac{c}{B} - r \ ,
\end{aligned} \tag{20}$$

where c is the continuous coupon rate for bond B, and λ_1 is an additional parameter to be fit by the data.

The above equation can be rewritten as a Fokker-Planck-type equation with an inhomogeneous term c. However, for our numerical procedures, it is more convenient to write this equation as a ‘‘truly nonlinear’’ Fokker-Planck equation with B present in V. If c/B is a smooth function, such that

$$\begin{aligned}
\frac{V(M^G ; \tau') - V(M^G ; \tau)}{\varepsilon} &= \Delta\tau \frac{\partial V}{\partial \tau} + \Delta\tau \int \frac{\delta V}{\delta B} \frac{\partial B(M'^{G'})}{\delta \tau} \, dM'^{G'} \\
&= O(\Delta\tau^{\nu}) \ ,
\end{aligned} \tag{21}$$

for $\nu > 1$, where $\tau' = \tau + \varepsilon\Delta\tau$, then our numerical path-integral codes may be used here as well [2].

In practice we do not have to use this truly nonlinear Fokker-Planck equation to price bonds with coupons. Assume we already have fit our parameters for the entire epoch of interest. Actual bond prices with coupons may then be evaluated straightforwardly by considering a portfolio of n pure discount bonds with a series of maturity dates T_n equivalent to the dates of payment of coupons and the face value of the actual coupon bond to be modeled. This prescription requires that we integrate back such a portfolio of n pure discount bonds with maturity T_n, to various times $t_i < T$ (including only those bonds in the portfolio with maturity $T_n \ge t_i$). At each of these times, we use the observed values of $r(t_i)$ and $l(t_i)$ to calculate the bond prices $B_n(t_i)$. This portfolio of $\{B_n(t_i)\}$ is then compared to the observed coupon bond $B(t_i)$, i.e., for many such times $\{t_i\}$. For each zero-coupon bond in this portfolio, we start at its time of maturity T_n, enforcing the i.c. $B_n(r, l; T_n) = 1$, and integrate back to a given time $t < T_n$. We then weight each zero coupon bond by the actual coupon or face value paid on the coupon bond.

4. NUMERICAL CALCULATIONS

4.1. Methodology

Recently, two major computer codes have been developed, which are key tools for the use of this approach to estimate model parameters and price bonds.

The first code, very fast simulated re-annealing (VFSR) [1], fits short-time probability distributions to observed data, using a maximum likelihood technique on the Lagrangian. An algorithm of very fast simulated re-annealing has been developed to fit observed data to a theoretical cost function over a D-dimensional parameter space [1], adapting for varying sensitivities of parameters during the fit. The annealing schedule for the "temperatures" (artificial fluctuation parameters) T_i decrease exponentially in "time" (cycle-number of iterative process) k, i.e., $T_i = T_{i0} \exp(-c_i k^{1/D})$.

Heuristic arguments have been developed to demonstrate that this algorithm is faster than the fast Cauchy annealing [40], $T_i = T_0/k$, and much faster than Boltzmann annealing [41], $T_i = T_0/\ln k$. To be more specific, the kth estimate of parameter α^i,

$$\alpha^i_k \in [A_i, B_i] \ , \tag{22}$$

is used with the random variable x^i to get the $k + 1$th estimate,

$$\alpha^i_{k+1} = \alpha^i_k + x^i (B_i - A_i) \ ,$$

$$x^i \in [-1, 1] \ . \tag{23}$$

The generating function is defined as

$$g_T(x) = \prod_{i=1}^{D} \frac{1}{2 \ln(1 + 1/T_i)(|x^i| + T_i)} \equiv \prod_{i=1}^{D} g^i_T(x^i) \ ,$$

$$T_i = T_{i0} \exp(-c_i k^{1/D}) \ . \tag{24}$$

Note that the use of C, the cost function given above, is *not* equivalent to doing a simple least squares fit on $M(t + \Delta t)$.

The second code develops the long-time probability distribution from the Lagrangian fit by the first code. A robust and accurate histogram-based (non-Monte Carlo) path-integral algorithm to calculate the long-time probability distribution has been developed to handle nonlinear Lagrangians [2,3,42,43], including a two-variable code for additive and multiplicative cases.

The histogram procedure recognizes that the distribution can be numerically approximated to a high degree of accuracy as sum of rectangles at points M_i of height P_i and width ΔM_i. For convenience, just consider a one-dimensional system. The above path-integral representation can be rewritten, for each of its intermediate integrals, as

$$P(M; t + \Delta t) = \int dM' [g_s^{1/2} (2\pi \Delta t)^{-1/2} \exp(-L_s \Delta t)] P(M'; t)$$

$$= \int dM' G(M, M'; \Delta t) P(M'; t) \ ,$$

$$P(M; t) = \sum_{i=1}^{N} \pi(M - M_i) P_i(t)$$

$$\pi(M - M_i) = \begin{cases} 1\,,\ (M_i - \frac{1}{2}\Delta M_{i-1}) \le M \le (M_i + \frac{1}{2}\Delta M_i)\ , \\ 0\,,\ \text{otherwise}\,, \end{cases} \tag{25}$$

which yields

$$P_i(t + \Delta t) = T_{ij}(\Delta t) P_j(t)\ ,$$

$$T_{ij}(\Delta t) = \frac{2}{\Delta M_{i-1} + \Delta M_i} \int_{M_i - \Delta M_{i-1}/2}^{M_i + \Delta M_i/2} dM \int_{M_j - \Delta M_{j-1}/2}^{M_j + \Delta M_j/2} dM' G(M, M'; \Delta t)\ . \tag{26}$$

T_{ij} is a banded matrix representing the Gaussian nature of the short-time probability centered about the (varying) drift.

This histogram procedure has been extended to two dimensions, i.e., using a matrix T_{ijkl} [3], e.g., essentially similar to the use of the A matrix in the previous section. Explicit dependence of L on time t also can be included without complications. We see no problems in extending it to other dimensions, other than care must be used in developing the mesh in ΔM, which is dependent on the diffusion matrix.

Fitting data with the short-time probability distribution, effectively using an integral over this epoch, permits the use of coarser meshes than the corresponding stochastic differential equation. The coarser resolution is appropriate, typically required, for numerical solution of the time-dependent path-integral: By considering the contributions to the first and second moments of ΔM^G for small time slices θ, conditions on the time and variable meshes can be derived [42]. The time slice essentially is determined by $\theta \le \bar{L}^{-1}$, where $\bar{L}$ is the "static" Lagrangian with $dM^G/dt = 0$, throughout the ranges of M^G giving the most important contributions to the probability distribution P. The variable mesh, a function of M^G, is optimally chosen such that ΔM^G is measured by the covariance $g^{GG'}$, or $\Delta M^G \sim (g^{GG}\theta)^{1/2}$.

The BS use of "natural" b.c., actually more general unrestricted or singular b.c. [44], is in part based on their own admittedly *ad hoc* choice of functional forms for r and l diffusions, in both their s.d.e. and p.d.e., and in the r drift in their s.d.e. Since we are using these equations in our calculations, we properly use unrestricted b.c., relying on the algebraic forms of the drifts and diffusions to enforce them. This is in contrast to BS who, when solving their pde numerically, resort to redundantly using these b.c. to define their basic transition matrix [8].

In future work, where it likely will be desirable to test other algebraic models of these drifts and diffusions, other b.c. will be appropriate, e.g., absorbing or reflecting b.c., both of which we have used in previous work in other systems [3]. We believe that the appropriate b.c. must be determined by finance considerations, permitting application of more general b.c. The path-integral methodology readily permits such inclusion in its numerical implementation. We believe it is extremely important to gain this freedom over the functional forms of the drifts and diffusions [37]. For example, our calculations with this model clearly demonstrate that the rather mild nonlinearities of the BS model only permit inflationary evolution, since those were the periods were fit to data and since the functional forms likely cannot accommodate many swings and dips, on time scales of months or years, much longer that of the fluctuations, yet shorter than the period of long-term bonds. This appears to require a higher degree of nonlinearity and/or an increase in the number of independent interest-rate variables.

We find it quite straightforward to adjust λ_2 to fit a set of bond prices. Our path-integral calculations are currently being performed on a large aggregate of bond data. This is necessary before meaningful comparisons can be made with other methodologies, in accord with other investigators who use such portfolios to average over systematic particulars of individual bonds. This data will be published in a future paper.

4.2. Fits to Interest Rates

Interest rates were developed from Treasury bill and bond yields during the period October 1974 through December 1979, the same period as one of the sets used by BS [10]. Short-term rates were determined from Treasury bills with a maturity of three months (BS used 30-day maturities), and long-term rates were determined from Treasury bonds with a maturity of twenty years (BS used at least 15-year maturities). For monthly runs, we used 63 points of data between 74-10-31 and 79-12-28. For daily runs, we used 1283 points of data between 74-10-31 and 79-12-31. We used yearly rates divided by 12 to fit the parameters.

For daily data, the actual number of days between successive trades was used; i.e., during this time period we had 1282 pieces of daily data and 62 pieces of end-of-month data. Although a rescaling in time only simply scales the deterministic parameters linearly, since that is how they appear in this model, this is not true for ρ. Then we did all subsequent runs using the scale of one day. We used yearly rates divided by 365 to fit the parameters.

The BS parameters also were run through the data, calculating the cost function they give. The single cost function bears the weight of determining all parameters. Typically, three or four significant-figure stability is required to get even one or two significant-figure stability in the parameters. (All runs were performed using double precision for all floating-point variables.) The "cost function" calculated is the sum over all Lagrangians at each short-time epoch (divided by the number of epochs, which doesn't affect its minimum, but helps to compare cost functions over different sets of data). I.e., a maximum probability fit is done by minimizing the cost functions (each the argument of the exponential representing the probability distribution of the variables) over all time epochs. The BS versus our fitted parameters are given in Table 1. Note that we have used data not quite the same as they used; we used the same time period, but a different set of bonds to determine interest rates. This likely can account for some of the apparent drastic improvements of our fits over theirs. Also note that the negative C we calculate is obtained from the negative ln term which has a very small argument. E.g., in the final column, $C = -23.83$ is obtained by adding an average (over all data points) ln contribution of -24.80 to a positive average L. "Competition" between the diminishing positive numerators in L and the diminishing diffusions in the ln term and in the denominators of L accounts for the final value of C.

Table 1

It should be noted that for all periods before October 1974, back through December 1958, using monthly data, BS found $a_1 < 0$, and for the period April 1964 through June 1969 they found $c_2 > 0$.

Fits were performed on a Hewlett Packard 9000-835SE, a "12-MIPS" computer. The VFSR code was tested by generating data from the BS Langevin differential equations, then using the VFSR code using the Lagrangian cost function to refit their parameters after they had been displaced from their values in the differential equations. A typical fit took approximately 100 CPU minutes for 1500 acceptance points, representing about 2000 generated points per 100 acceptance points at each re-annealing cycle, in this six-dimensional parameter space. It was found that once the VFSR code repeated the lowest cost function within two cycles of 100 acceptance points, e.g., typically achieving 3 or 4 significant-figure accuracy in the global minimum of the cost function, by shunting to a local fitting procedure, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [45], only several hundred acceptance points were required to achieve 7 or 8 significant-figure accuracy in the cost function. This also provided yet another test of the VFSR methodology.

5. CHAOS OR NOISE?

Given the context of current studies in complex nonlinear systems [46], the question can be asked: What if markets have chaotic mechanisms that overshadow the above stochastic considerations? The real issue is whether the scatter in data can be distinguished between being due to noise or chaos. In this regard, we note that several studies have been proposed with regard to comparing chaos to simple filtered (colored) noise [46,47] Since we have previously derived the existence of multiplicative noise in neocortical interactions, then the previous references must be generalized, such that we must investigate whether markets scatter can be distinguished from multiplicative noise. A previously described application of this methodology follows:

In our analysis of military exercise data [3,35], we were able to fit short-time attrition epochs (determined to be about 5 minutes from mesh considerations determined by the nature of the Lagrangian) with short-time nonlinear Gaussian-Markovian probability distributions with a resolution comparable to the spread in data. When we did the long-time path-integral from some point (spread) at the beginning of the battle, we found that we could readily find a form of the Lagrangian that made physical sense and that also fit the multivariate variances as well as the means at each point in time of the rest of the exercise interval. I.e., there was not any degree of hyper-sensitivity to initial conditions that prevented us from ‘‘predicting’’ the long time means and variances of the system. Of course, since the system is dissipative, there is a strong tendency for all moments to diminish in time, but in fact this exercise was of sufficiently modest duration (typically 1 to 2 hours) that variances do increase somewhat during the middle of the battle. In summary, this battalion-regiment scale of battle does not seem to possess chaos.

Similar to serious work undertaken in several fields [47-49], here too, the impulse to cry ‘‘chaos!’’ has been premature. It is not supported by the facts, tentative as they are because of sparse data.

A more purposeful project is to compare stochastic with deterministic models of data. Today much attention is turning to the use of deterministic chaotic models for short time predictions of systems. For example, if only short time predictions are required, and if a deterministic chaotic model could well describe stochastic data within these epochs, then this model might be more computationally efficient instead of a more ‘‘correct’’ stochastic model which would be necessary for long time predictions. The scales of time involved are of course system dependent, and the deterministic chaotic modeling of data is still in its infancy [50].

For example, it has been widely noted that the correlation dimension of data is difficult to calculate; perhaps it is often not even a well founded concept [49]. It’s calculation, e.g., using the popular Grassberger-Procaccia algorithm [51], even when supplemented with finer statistical tests [52] and noise reduction techniques [53], may prove fruitful, but likely only as a sensitivity index relative to shifting contexts in some systems.

6. CONCLUSION

We have described how mathematical methodologies and numerical algorithms recently developed in the field of statistical mechanics can be brought to bear on term structure models. Specifically, methods of very fast simulated re-annealing can be used to statistically find best global fits of multivariate nonlinear stochastic term structure models, without requiring approximation of the basic models.

This new formalism also permits a fresh look at some of these models and affords comparison with other nonlinear stochastic systems. Elsewhere [37], we are publishing some numerical results on fits to daily Treasury bill and bond yields during the period October 1974 through December 1979. Another paper in progress will report on more extensive comparisons with observed bond prices. Similar studies, e.g., using these mathematical physics and computational techniques, are underway to determine behavioral correlates of electroencephalographic (EEG) data [29].

ACKNOWLEDGEMENTS

I thank Ted Barnhill for helping me to access the George Washington University financial database, George Jabbour for introducing me to the Brennan-Schwartz literature, and Michael Wehner for discussions on the path-integral boundary conditions and for work in progress calculating bond prices.

REFERENCES

[1] L. Ingber, Mathl. Comput. Modelling **12**, 967 (1989).

[2] M.F. Wehner and W.G. Wolfer, Phys. Rev. A **35**, 1795 (1987).

[3] L. Ingber, H. Fujio, and M.F. Wehner, Mathl. Comput. Modelling **15**, 65 (1991).

[4] J.C. Cox, J.E. Ingersoll, Jr., and S.A. Ross, J. Finance **36**, 769 (1981).

[5] J.C. Cox, J.E. Ingersoll, Jr., and S.A. Ross, Econometrica **53**, 363 (1985).

[6] J.C. Cox, J.E. Ingersoll, Jr., and S.A. Ross, Econometrica **53**, 385 (1985).

[7] O. Vasicek, J. Finan. Econ. **5**, 177 (1977).

[8] M.J. Brennan and E.S. Schwartz, J. Banking Finan. **3**, 133 (1979).

[9] M.J. Brennan and E.S. Schwartz, Finan. Anal. J. **Sep-Oct**, 49 (1982).

[10] M.J. Brennan and E.S. Schwartz, J. Finan. Quant. Anal. **17**, 301 (1982).

[11] S.M. Schaefer and E.S. Schwartz, J. Finan. Quant. Anal. **19**, 413 (1984).

[12] B. Dietrich-Campbell and E. Schwartz, J. Finan. Econ. **16**, 321 (1986).

[13] F. Langouche, D. Roekaerts, and E. Tirapegui, *Functional Integration and Semiclassical Expansions* (Reidel, Dordrecht, The Netherlands, 1982).

[14] H. Dekker, in *Functional Integration: Theory and Applications*, edited by J.P. Antoine and E. Tirapegui (Plenum, New York, 1980), p. 207.

[15] H. Grabert and M.S. Green, Phys. Rev. A **19**, 1747 (1979).

[16] R. Graham, Z. Physik **B26**, 397 (1977).

[17] R. Graham, Phys. Rev. Lett. **38**, 51 (1977).

[18] L.S. Schulman, *Techniques and Applications of Path Integration* (J. Wiley & Sons, New York, 1981).

[19] L. Ingber, in *9th MIT/ONR Workshop on C^3 Systems: Naval Postgraduate School, Monterey, CA, 2-5 June 1986* (MIT, Cambridge, MA, 1986), p. 237.

[20] R.F. Fox, Phys. Rev. A **34**, 4525 (1986).

[21] C. van der Broeck, J. Stat. Phys. **31**, 467 (1983).

[22] P. Colet, H.S. Wio, and M. San Miguel, Phys. Rev. A **39**, 6094 (1989).

[23] C.W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer-Verlag, Berlin, Germany, 1983).

[24] L. Ingber, Phys. Rev. A **28**, 395 (1983).

[25] L. Ingber, Phys. Rev. A **29**, 3346 (1984).

[26] L. Ingber, IEEE Trans. Biomed. Eng. **32**, 91 (1985).

[27] L. Ingber, Phys. Rev. A **31**, 1183 (1985).

[28] L. Ingber and P.L. Nunez, Mathl. Comput. Modelling **13**, 83 (1990).

[29] L. Ingber, Phys. Rev. A **44**, 4017 (1991).

[30] L. Ingber, Phys. Rev. D **29**, 1171 (1984).

[31] L. Ingber, Phys. Rev. D **33**, 3781 (1986).

[32] L. Ingber, in *1989 JDL C^2 Symposium: National Defense University, Washington, DC, 27-29 June 1989* (SAIC, McLean, VA, 1989), p. 169.

[33] L. Ingber, in *The Science of Command and Control: Part II, Coping With Complexity*, edited by S.E. Johnson and A.H. Levis (AFCEA International, Washington, DC, 1989), p. 165.

[34] L. Ingber and D.D. Sworder, Mathl. Comput. Modelling **15**, 99 (1991).

[35] L. Ingber, in *Toward a Science of Command, Control, and Communications*, edited by C. Jones (American Institute of Aeronautics and Astronautics, Washington, D.C., 1993), p. 117.

[36] L. Ingber, Math. Modelling **5**, 343 (1984).

[37] L. Ingber, M.F. Wehner, G.M. Jabbour, and T.M. Barnhill, Mathl. Comput. Modelling **15**, 77 (1991).

[38] K. Kishida, Phys. Rev. A **25**, 496 (1982).

[39] K. Kishida, J. Math. Phys. **25**, 1308 (1984).

[40] H. Szu and R. Hartley, Phys. Lett. A **122**, 157 (1987).

[41] S. Kirkpatrick, C.D. Gelatt, Jr., and M.P. Vecchi, Science **220**, 671 (1983).

[42] M.F. Wehner and W.G. Wolfer, Phys. Rev. A **27**, 2663 (1983).

[43] M.F. Wehner and W.G. Wolfer, Phys. Rev. A **28**, 3003 (1983).

[44] N.S. Goel and N. Richter-Dyn, *Stochastic Models in Biology* (Academic Press, New York, NY, 1974).

[45] D.F. Shanno and K.H. Phua, ACM Trans. Mathl. Software **2**, 87 (1976).

[46] R. Pool, Science **243**, 25 (1989).

[47] P. Grassberger, Nature **323**, 609 (1986).

[48] W.A. Brock, J. Econ. Theory **40**, 168 (1986).

[49] P.E. Rapp, T.R. Bashore, J.M. Marinerie, A.M. Albano, I.D. Zimmerman, and A.I. Mees, Brain Topography **2**, 99 (1989).

[50] H.D.I. Abarbanel, R. Brown, and J.B. Kadtke, Phys. Rev. A **41**, 1782 (1990).

[51] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).

[52] W. Brock, J. Lakonishok, and B. LeBaron, SSRI Working Paper No. 8702, Revised, 1988.

[53] D.S. Broomhead and G.P. King, Physica D **20**, 217 (1986).

TABLE

Table 1. BS parameters were fit to data using our Lagrangian representation for their coupled $r - l$ equations, for both end-of-month and daily data between 74-10-31 and 79-12-31. The second column, designated BS Monthly, gives their published 1982 results, using somewhat different data during this period. The third column gives our monthly fits on somewhat different data during this same time period. The fourth column gives daily fits scaled to daily time. The last line gives the cost function C averaged over the number of data points. Note that the data used here is not quite the same data used by BS.

Parameter	BS Monthly	L Monthly	L Daily
a_1	0.0361	$3.02\ 10^{-5}$	$-6.33\ 10^{-9}$
b_1	0.0118	$3.89\ 10^{-4}$	0.0902
σ_1	0.0777	0.0700	0.0132
ρ	0.442	0.534	0.136
a_2	0.169	$9.73\ 10^{-3}$	$2.43\ 10^{-4}$
b_2	0.0089	0.0262	0.0320
c_2	−0.271	−0.707	−0.492
σ_2	0.0243	0.0278	$4.01\ 10^{-3}$
C	23.69	−13.87	−23.83