Statistical Mechanics of Financial Markets (SMFM): Applications to Trading Indicators and Options

Lester Ingber

ingber@ingber.com • ingber@alumni.caltech.edu
http://www.ingber.com/

MATHEMATICAL PHYSICS

NUMERICAL ALGORITHMS

APPLICATIONS TO INTEREST RATE PRODUCTS

APPLICATIONS TO TRADING INDICATORS

APPLICATIONS TO OPTIONS

$Id: markets01_lecture,v 1.40 2002/04/06 16:40:50 ingber Exp ingber $

MATHEMATICAL PHYSICS

Nonlinear Nonequilibrium Multivariate Stochastic Aggregation

Complex systems typically are in nonequilibrium, being driven by nonlinear and stochastic interactions described by many external and internal degrees of freedom. For these systems, classical thermodynamic descriptions typically do not apply. Many such systems are best treated by respecting some intermediate mesoscale as fundamental to drive larger macroscopic processes.

Often these mesoscopic scales are aptly described by Gaussian Markovian statistics. They naturally develop in physical and biological scales to maximally process information from microscopic scales up to macroscopic scales. Possibly this is true as well of some social systems such as financial markets.

For many physical systems this mesoscopic scale still has some audit trail back to its microscopic origins. Often, statistical deviations of drift variables lead to functional dependencies in diffusion variables.

Stochastic Differential Equation (SDE)

The Stratonovich (midpoint discretized) Langevin equations can be analyzed in terms of the Wiener process dW^i, which can be rewritten in terms of Gaussian noise $\eta^i = dW^i/dt$ if care is taken in the limit.

$$dM^G = f^G(t, M(t))dt + \hat{g}_i^G(t, M(t))dW^i \ ,$$

$$\dot{M}^G(t) = f^G(t, M(t)) + \hat{g}_i^G(t, M(t))\eta^i(t) \ ,$$

$$dW^i \rightarrow \eta^i dt \ ,$$

$$M = \{ \ M^G ; G = 1, \cdots, \Lambda \ \} \ ,$$

$$\eta = \{ \ \eta^i ; i = 1, \cdots, N \ \} \ .$$

$$\dot{M}^G = dM^G/dt \ ,$$

$$< \eta^j(t) >_\eta = 0 \ ,$$

$$< \eta^j(t), \eta^{j'}(t') >_\eta = \delta^{jj'}\delta(t - t') \ ,$$

η^i represents Gaussian white noise, and moments of an arbitrary function $F(\eta)$ over this stochastic space are defined by a path-type integral over η^i, folding time increments $\theta = \Delta t$,

$$< F(\eta) >_\eta = \bar{N}^{-1} \int D\eta F(\eta) \exp\left(-\frac{1}{2} \int_{t_0}^{\infty} dt \eta^i \eta^i \right),$$

$$\bar{N} = \int D\eta \exp\left(-\frac{1}{2} \int_{t_0}^{\infty} dt \eta^i \eta^i \right),$$

$$D\eta = \lim_{v \to \infty} \prod_{\alpha=0}^{v+1} \prod_{j=1}^{N} (2\pi\theta)^{-1/2} dW_\alpha^j \ ,$$

$$t_\alpha = t_0 + \alpha\theta \ ,$$

$$\frac{1}{2} \int dt \eta^i \eta^i = \frac{1}{2\theta} \sum_\beta \sum_i (W_\beta^i - W_{\beta-1}^i)^2 \ ,$$

$$< \eta^i >_\eta = 0 \ ,$$

$$< \eta^i(t) \eta^j(t') >_\eta = \delta^{ij} \delta(t - t') \ .$$

Partial Differential Equation (PDE)

If some boundary conditions are added as Lagrange multipliers, these enter as a "potential" V, creating a Schrödinger-type equation:

$$P_{,t} = \frac{1}{2}(g^{GG'}P)_{,GG'} - (g^G P)_{,G} + VP ,$$

$$P = < P_\eta >_\eta ,$$

$$g^G = f^G + \frac{1}{2}\hat{g}_i^{G'}\hat{g}_{i,G'}^G ,$$

$$g^{GG'} = \hat{g}_i^G \hat{g}_i^{G'} ,$$

$$(\cdots)_{,G} = \partial(\cdots)/\partial M^G .$$

Note that g^G replaces f^G in the SDE if the Itô (prepoint discretized) calculus is used to define that equation.

Lagrangian Probability Distribution Function (PDF)

This can be transformed to the Stratonovich representation, in terms of the Feynman Lagrangian L possessing a covariant variational principle,

$$P[M_t|M_{t_0}]dM(t) =$$

$$\int \cdots \int \underset{\sim}{D}M \exp\left(-\min \int_{t_0}^{t} dt' L \right) \delta(M(t_0) = M_0)\delta(M(t) = M_t) \ ,$$

$$\underset{\sim}{D}M = \lim_{u\to\infty} \prod_{\rho=1}^{u+1} g^{1/2} \prod_G (2\pi\theta)^{-1/2} dM_\rho^G \ ,$$

$$L(\dot{M}^G, M^G, t) = \frac{1}{2}(\dot{M}^G - h^G) g_{GG'} (\dot{M}^{G'} - h^{G'}) + \frac{1}{2} h^G{}_{;G} + R/6 - V \ ,$$

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

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

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

$$g = \det(g_{GG'}) \ ,$$

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

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

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

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

Path-Integral Riemannian Geometry

The midpoint derivation explicitly derives a Riemannian geometry induced by these statistics, with a metric defined by the inverse of the covariance matrix

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

and where R is the Riemannian curvature

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

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

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

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

but the Lagrangian $\underline{L}$ so specified does not satisfy a variational principle as useful for moderate to large noise; its associated variational principle only provides information useful in the weak-noise limit. Numerically, this often means that finer meshes are required for calculations for the prepoint representation.

Statistical Mechanics of Financial Markets (SMFM) — Microscopic Agents

Some insight may be gained into how a market becomes reasonably described by a parabolic PDE, by performing a simple statistical mechanics of risk-averting agents.

Consider the conditional probability distribution, p_j, of an agent j operating on a given market. For simplicity, assume that at time $t + \tau$, j must decide whether to buy or sell a standard increment of the market, based only on the information of the total number of buyers, M^B, and sellers, M^S, at time t. For example, take

$$p_{\sigma_j} = \frac{\exp(-\sigma_j F_j)}{[\exp(F_j) + \exp(-F_j)]}$$

$$\approx \frac{1}{2} [1 - \mathrm{erf}(\sigma_j F_j \sqrt{\pi}/2)] ,$$

$$\sigma_j = \begin{cases} +1 & \text{buy } (j \in B) \text{ or sell } (j \in S) \\ -1 & \text{do not act ,} \end{cases}$$

$$p_+ + p_- = 1 ,$$

$$F_j = F_j(M^G) ,$$

$$G = \{B, S\} .$$

F_j may be any reasonably well-behaved function of M^B and M^S, different for buyers, $F_{j \in B} \equiv F^B$, or sellers, $F_{j \in S} \equiv F^S$. F_j is considered to represent a ‘‘decision factor’’ representing a ‘‘typical’’ rational agent in the market.

A simple example of F_j for agents following market trends is obtained from

$$F^G_{\mathrm{ex1}} = a^G M^- / N ,$$

$$M^- = M^B - M^S ,$$

where a^G are constants, $a^B < 0$ and $a^S > 0$, for agents following the trends of the market. I.e., agent j acts according to a sigmoid distribution with respect to market trends: p_{σ_j} is concave with respect to gains, and convex with respect to losses. Assume that the total numbers of *potential* buyers and sellers are each constants,

$$j_S = 1, \cdots, N^S ,$$

$$j_B = 1, \cdots, N^B ,$$

$$N = N^B + N^S .$$

SMFM— Market Dynamics

The "joint" probability distribution P, joint with respect to pools of all S and B agents, but conditional with respect to time evolution, is

$$P[M(t+\tau)|M(t)] = \prod_G^{B,S} P^G[M^G(t+\tau)|M^{\bar{G}}(t)]$$

$$= \sum_{\sigma_j=\pm 1} \delta(\sum_{j\in S}\sigma_j - N^S)\delta(\sum_{j\in B}\sigma_j - N^B)\prod_j^N p_{\sigma_j}$$

$$= \prod_G (2\pi)^{-1}\int dQ^G \exp[iM^G(t+\tau)Q^G]$$

$$\times \prod_{j\in G}^{N^G} \cosh\{F_j[M(t)] + iQ^G\}\mathrm{sech}\{F_j[M(t)]\}$$

$$= \prod_G (1+E^G)^{-N^G} \binom{N^G}{\lambda^G} (E^G)^{\lambda^G} \ ,$$

$$E^G = \exp(-2F^G) \ ,$$

$$\lambda^G = [[\frac{1}{2}(M^G(t+\tau) + N^G)]] \ ,$$

$$M = \{M^G\} \ ,$$

where $M^{\bar{G}}(t)$ represents contributions from both $G = S$ and $G = B$ at time t, and λ^G is defined as the greatest integer in the double brackets.

The mean and variance of this binomial distribution yields

$$< M^G(t+\tau) >= -N^G \tanh F^G \ ,$$

$$< M^G(t+\tau)M^{G'}(t+\tau) > - < M^G(t+\tau) >< M^{G'}(t+\tau) >= \frac{1}{4}\delta^{GG'} N^G \mathrm{sech}^2 F^G \ .$$

For large N^G and large $N^G F^G$, this binomial distribution is asymptotically Gaussian, albeit with nonlinear drifts and diffusions.

Information

With reference to a steady state $\bar{P}(\tilde{M})$, when it exists, an analytic definition of the information gain $\hat{\Upsilon}$ in state $\tilde{P}(\tilde{M})$ is defined by

$$\hat{\Upsilon}[\tilde{P}] = \int \cdots \int \underline{D}\tilde{M} \; \tilde{P} \ln(\tilde{P}/\bar{P}),$$

where a path integral is defined such that all intermediate-time values of $\tilde{M}$ appearing in the folded short-time distributions $\tilde{P}$ are integrated over. This is quite general for any system that can be described as Gaussian-Markovian, even if only in the short-time limit.

As time evolves, the distribution likely no longer behaves in a Gaussian manner, and the apparent simplicity of the short-time distribution must be supplanted by numerical calculations.

Transformations Without Itô Calculus

Consider

$$V[S,t+\delta t\,|\,S,t] = (2\pi(\sigma S)^2\delta t)^{-1/2}\exp(-L\delta t)\ ,$$

$$L = \frac{(\dot{S}+rS)^2}{2(\sigma S)^2} + r\ ,$$

$$\dot{S} = \frac{\delta S}{\delta t} = \frac{S(t+\delta t)-S(t)}{\delta t}\ .$$

Some care must be taken with nonconstant drifts and diffusions. For example, for purposes of calculating volatilities, it is often convenient to transform to a variable Z (S relative to some $\bar{S}$ scale)

$$Z = \ln S\ .$$

The above distribution can be transformed into $V[Z,t+\delta t|Z,t]$,

$$dS_t V[S,t+\delta t|S,t] = dZ_t V[Z,t+\delta t|Z,t]$$

$$= dZ_t(2\pi\sigma^2\delta t)^{-1/2}\exp(-L'\delta t)\ ,$$

$$L'\delta t = \frac{([\exp(Z_{t+\delta t}-Z_t)-1]+r)^2}{2\sigma^2\delta t} + r\delta t\ .$$

This can be expanded into

$$L'\delta t \approx \frac{(Z_{t+\delta t}-Z_t+\frac{1}{2}(Z_{t+\delta t}-Z_t)^2 - r\delta t))^2}{2\sigma^2\delta t} + r\delta t$$

$$\approx \frac{(Z_{t+\delta t}-Z_t-(r-\frac{1}{2}\sigma^2)\delta t)^2}{2\sigma^2\delta t} + r\delta t\ ,$$

$$(Z_{t+\delta t}-Z_t)^2 \approx \sigma^2\delta t\ ,$$

where only terms of order δt have been kept, yielding

$$L' = \frac{\left(\dot{Z}-(r-\frac{1}{2}\sigma^2)\right)^2}{2\sigma^2} + r\ .$$

Intuitive Variables

It must be emphasized that the output need not be confined to complex algebraic forms or tables of numbers. Because L possesses a variational principle, sets of contour graphs, at different long-time epochs of the path-integral of P over its variables at all intermediate times, give a visually intuitive and accurate decision-aid to view the dynamic evolution of the scenario. For example, this Lagrangian approach permits a quantitative assessment of concepts usually only loosely defined.

Euler-Lagrange Variational Equations

The Euler-Lagrange variational equations give rise to the familiar force law

$$\text{“}F = ma\text{”}: \ \ \delta L = 0 = \frac{\partial L}{\partial M^G} - \frac{\partial}{\partial t}\,\frac{\partial L}{\partial(\partial M^G/\partial t)}\ ,$$

$$\text{“Force”} = \frac{\partial L}{\partial M^G}\ ,$$

$$\text{“Mass”} = g_{GG'} = \frac{\partial^2 L}{\partial(\partial M^G/\partial t)\partial(\partial M^{G'}/\partial t)}\ ,$$

where M^G are the variables and L is the Lagrangian. These physical entities provide another form of intuitive, but quantitatively precise, presentation of these analyses.

Canonical Momenta Indicators (CMI)

Canonical Momenta Indicators (CMI), defined by

$$\text{“Momentum”} = \Pi^G = \frac{\partial L}{\partial(\partial M^G/\partial t)} \ ,$$

can be used as financial indicators faithful to an underlying mathematics modeling markets as stochastic distributions.

NUMERICAL ALGORITHMS

Adaptive Simulated Annealing (ASA)

This algorithm fits empirical data to a theoretical cost function over a D-dimensional parameter space, adapting for varying sensitivities of parameters during the fit.

For several test problems, ASA has been shown to be orders of magnitude more efficient than other similar techniques. ASA has been applied to several complex systems, including specific problems in neuroscience, finance and combat systems.

Heuristic arguments have been developed to demonstrate that this algorithm is faster than the fast Cauchy annealing, $T_i = T_0/k$, and much faster than Boltzmann annealing, $T_i = T_0/\ln k$.

Hills and Valleys

It helps to visualize the problems presented by such complex systems as a geographical terrain. For example, consider a mountain range, with two "parameters," e.g., along the North−South and East−West directions. We wish to find the lowest valley in this terrain. SA approaches this problem similar to using a bouncing ball that can bounce over mountains from valley to valley. We start at a high "temperature," where the temperature is an SA parameter that mimics the effect of a fast moving particle in a hot object like a hot molten metal, thereby permitting the ball to make very high bounces and being able to bounce over any mountain to access any valley, given enough bounces. As the temperature is made relatively colder, the ball cannot bounce so high, and it also can settle to become trapped in relatively smaller ranges of valleys.

We imagine that our mountain range is aptly described by a "cost function." We define probability distributions of the two directional parameters, called generating distributions since they generate possible valleys or states we are to explore. We define another distribution, called the acceptance distribution, which depends on the difference of cost functions of the present generated valley we are to explore and the last saved lowest valley. The acceptance distribution decides probabilistically whether to stay in a new lower valley or to bounce out of it. All the generating and acceptance distributions depend on temperatures.

Outline of ASA Algorithm

For parameters

$$\alpha_k^i \in [A_i, B_i] \ ,$$

sampling with the random variable x^i,

$$x^i \in [-1, 1] \ ,$$

$$\alpha_{k+1}^i = \alpha_k^i + x^i (B_i - A_i) \ ,$$

define the generating function

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

in terms of parameter "temperatures"

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

The cost-functions $\underline{L}$ under consideration are of the form

$$h(M; \alpha) = \exp(-\underline{L}/T) \ ,$$

$$\underline{L} = L \Delta t + \frac{1}{2} \ln(2\pi \Delta t g_t^2) \ ,$$

where L is a Lagrangian with dynamic variables $M(t)$, and parameter-coefficients α to be fit to data. g_t is the determinant of the metric, and T is the cost "temperature."

Numerical Path Integration (PATHINT)

Given a form for L, we use the path-integral to calculate the long-time distribution of variables. This is impossible in general to calculate in closed form, and we therefore must use numerical methods. PATHINT is a code developed for calculating highly nonlinear multivariate Lagrangians.

The path-integral calculation of the long-time distribution, in addition to being a predictor of upcoming information, provides an internal check that the system can be well represented as a nonlinear Gaussian-Markovian system. The use of the path integral to compare different models is akin to comparing short- and long-time correlations. Complex boundary conditions can be cleanly incorporated into this representation, using a variant of ‘‘boundary element’’ techniques.

Outline of PATHINT Algorithm

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

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

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

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

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

which yields

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

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

Mesh Limitations

Care must be used in developing the mesh in ΔM^G, which is strongly dependent on the diagonal elements of the diffusion matrix, e.g.,

$$\Delta M^G \approx (\Delta t g^{|G||G|})^{1/2} .$$

Presently, this constrains the dependence of the covariance of each variable to be a nonlinear function of that variable, albeit arbitrarily nonlinear, in order to present a straightforward rectangular underlying mesh.

A previous paper attempted to circumvent this restriction by taking advantage of Riemannian transformations to a relatively diagonal problem.

For more than one variable, the above constraints on the mesh only suffice for diagonal elements of the $g^{GG'}$ matrix. To consider the influence on off-diagonal terms, a tiling approach should be taken to the full mesh.

Probability Tree (PATHTREE)

PATHINT motivated the development of PATHTREE, an algorithm that permits extremely fast accurate computation of probability distributions of a large class of general nonlinear diffusion processes.

The natural metric of the space is used to first lay down the mesh. The evolving local short-time distributions on this mesh are then dynamically calculated.

The short-time probability density gives the correct result up to order $O(\Delta t)$ for any final point S', the order required to recover the corresponding partial differential equation.

Standard Binomial Tree

In a two-step binomial tree, the step up Su or step down Sd from a given node at S is chosen to match the standard deviation of the differential process. The constraints on u and d are chosen as

$$ud = 1 ,$$

If we assign probability p to the up step Su, and $q = (1 - p)$ to the down step Sd, the matched mean and variance are

$$pSu + (1 - p)Sd = < S(t + \Delta t) > ,$$

$$S^2 \big(pu^2 + qd^2 - (pu + qd)^2 \big) = < \big(S(t + \Delta t) - < S(t + \Delta t) > \big)^2 > .$$

The right-hand-side can be derived from the stochastic model used.

Deficiency of Standard Algorithm to Order $\sqrt{dt}$

A tree is constructed that represents the time evolution of the stochastic variable S. S is assumed to take only 2 values, u, (up value), and d (down value) at moment t, given the value S at moment $t - \Delta t$. The probabilities for the up and down movements are p and q, respectively. The 4 unknowns $\{u, d, p, q\}$ are calculated by imposing the normalization of the probability and matching the first two moments conditioned by the value S at $t - \Delta t$, using the variance of the exact probability distribution $P(S, t | S_0, t_0)$. One additional condition is arbitrary and is usually used to symmetrize the tree, e.g., $ud = 1$.

If the system to be modeled is given by a differential form, e.g.,

$$dS = f dt + g dW$$

then the noise term is only given to order $\sqrt{dt}$.

The Ornstein-Uhlenbeck (OU) process, $f = bS$ and $g = v$, for constant b and v, is special, as some higher order dt corrections in systems described by $g \propto S^x$ are zero for $x = 0$. The Black-Scholes (BS) process, $f = bS$ and $g = \sigma S$, for constant b and σ, also is special, as it can be simply transformed to a constant-diffusion lognormal process with the same $O(dt)$ simplifications.

Problems Generalizing The Standard Tree

The main problem is that the above procedure cannot be applied to a general nonlinear diffusion process, as the algorithm involves a previous knowledge of terms of $O(\Delta t)$ in the formulas of quantities $\{u, p\}$ obtained from a finite time expansion of the exact solution P sought. Otherwise, the discrete numerical approximation obtained does not converge to the proper solution.

Probability PATHTREE

In order to obtain tree variables valid up to $O(\Delta t)$, we turn to the short-time path-integral representation of the solution of the Fokker-Planck equation, which is just the multiplicative Gaussian-Markovian distribution. In the prepoint discretization relevant to the construction of a tree,

$$P(S', t'|S, t) = \frac{1}{\sqrt{2\pi \Delta t g^2}} \exp\left(- \frac{(S' - S - f dt)^2}{2 g^2 \Delta t} \right)$$

$$\Delta t = t' - t$$

valid for displacements S' from S "reasonable" as measured by the standard deviation $g\sqrt{\Delta t}$, which is the basis for the construction of meshes in the PATHINT algorithm.

The crucial aspects of this approach are: There is no a priori need of the first moments of the exact long-time probability distribution P, as the necessary statistical information to the correct order in time is contained in the short-time propagator. The mesh in S at every time step need not recombine in the sense that the prepoint-postpoint relationship be the same among neighboring S nodes, as the short-time probability density gives the correct result up to order $O(\Delta t)$ for any final point S'. Instead, we use the natural metric of the space to first lay down our mesh, then dynamically calculate the evolving local short-time distributions on this mesh.

Construction of PATHTREE

We construct an additive PATHTREE, starting with the initial value S_0, with successive increments

$$S_{i+1} = S_i + g\sqrt{\Delta t} \ , \ S_i > S_0$$

$$S_{i-1} = S_i - g\sqrt{\Delta t} \ , \ S_i < S_0 \ ,$$

where g is evaluated at S_i. We define the up and down probabilities p and q, resp., in an abbreviated notation, as

$$p = \frac{P(i+1|i;\Delta t)}{P(i+1|i;\Delta t) + P(i-1|i;\Delta t)}$$

$$q = 1 - p \ .$$

where the P's are the short-time transition probability densities. Note that in the limit of small Δt,

$$\lim_{\Delta t \to 0} p = \frac{1}{2} \ .$$

Direct Calculation of Probability

We can calculate the probability density function by first recursively computing the probabilities of reaching each node of the tree. This can be performed efficiently thanks to the Markov property. To compute the density function we need to rescale these probabilities by the distance to the neighboring nodes: the more spread the nodes are, the lower the density. First we compute the probability of reaching each final node of the tree. We do this incrementally by first computing the probabilities of reaching nodes in time slice 1, then time slice 2 and so forth. At time slice 0, we know that the middle node has probability 1 of being reached and all the others have probability 0. We compute the probability of reaching a node as a sum of two contributions from the previous time slice. We reach the node with transition *pu* from the node below at the previous slice, and with transition *pd* from the node above. Each contribution is the product of the probability at the previous node times the transition to the current node. This formula is just a discretized version of the Chapman-Kolmogorov equation

$$p(x_j, t_i) = p(x_{j-1}, t_{i-1}) pu_{j-1} + p(x_{j+1}, t_{i-1}) pd_{j+1} \ .$$

After we have computed the absolute probabilities at the final nodes, we can give a proper estimation of the density, by scaling the probabilities by the average of sizes of the two adjacent intervals, $\text{density}_i = p_i/((S_{i+2} - S_{i-2})/2)$.

Numerical Derivatives of Expectation of Probability

The probability P at time of expiration T can be calculated as a numerical derivative with respect to strike X of a European call option, taking the risk-free rate r to be zero, given an underlying S_0 evaluated at time $t = 0$, with strike X, given other variables such as volatility σ and cost of carry b. The call is the expectation of the function $\mathrm{Max}(S - X, 0)$.

$$P[S(T)|S(t_0)] \underset{S(T)\equiv X}{=} P[X|S(t_0)] = \frac{\partial^2 C}{\partial X^2}$$

This calculation of the probability distribution is dependent on the same conditions necessary for any tree algorithm, i.e., that enough nodes are processed to ensure that the resultant evaluations are a good representation of the corresponding Fokker-Planck equation, and that the number of iterations in PATHTREE are sufficient for convergence.

Alternative First Derivative Calculation of Probability

An alternative method of calculating the probability P a a first-order numerical derivative, instead of as second-order derivative, with respect to X is to define a function C_H using the Heaviside step-function $H(S, X) = 1$ if $S \geq X$ and 0 otherwise, instead of the Max function at the time to expiration. This yields

$$P[S(T)|S(t_0)] \underset{S(T) \equiv X}{=} P[X|S(t_0)] = -\frac{\partial C_H}{\partial X}$$

Sometimes this is numerically useful for sharply peaked distributions at the time of expiration, but we have found the second derivative algorithm above to work fine with a sufficient number of epochs.

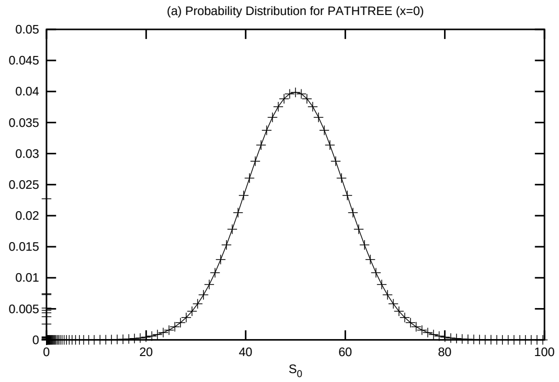
Our tests verify that the three methods give the same density. We consider the numerical-derivative calculations a very necessary baseline to determine the number of epochs required to get reasonable accuracy.

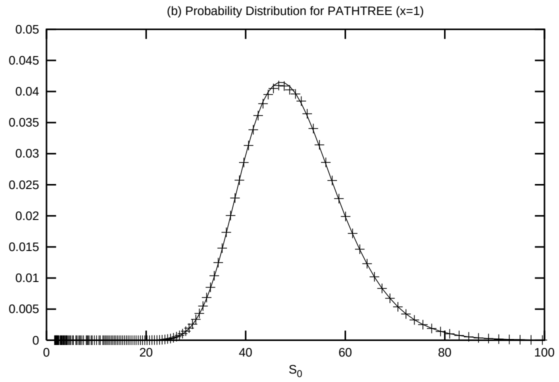
PATHTREE vs PATHINT

For PATHINT, the time and space variables are determined independently. I.e., the ranges of the space variables are best determined by first determining the reasonable spread of the distribution at the final time epoch. For PATHTREE just one parameter, the number of epochs N, determines the mesh for both time and the space variables. This typically leads to a growth of the tree, proportional to $\sqrt{N}$, much faster than the spread of the distribution, so that much of the calculation is not relevant.

Black-Scholes (BS) and Ornstein-Uhlenbeck (OU) Examples

The graphs below compare analytic solutions with second-derivative numerical ("+"s) PATHTREE calculations.





APPLICATIONS TO INTEREST RATE PRODUCTS

2-Factor Interest-Rate Model

The Brennan-Schwartz model is developed in the variables of short- and long-term interest rates, assumed to follow a joint Wiener stochastic process,

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

$$dl = \beta_2(r, l, t)dt + \eta_2(r, l, t)dz_2 \ ,$$

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.

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

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

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

BS simplified and reduced this system to

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

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

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

Lagrangian Representation

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

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

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

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

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

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

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

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

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

$$k = 1 - \rho^2 .$$

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

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

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

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

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

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

ASA Fits

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. 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 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. 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. 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. "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. 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 $c_2 > 0$.

Bond PDE/PATHINT

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

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

$$= AB \ ,$$

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 helps to appreciate the importance of the BS methodology by examining the discretized 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

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

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

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

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

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

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

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

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.

Some tentative PATHINT calculations were performed by another researcher. It would be interesting to repeat them.

APPLICATIONS TO TRADING INDICATORS

S&P Interday Futures-Cash

CMI and ASA were blended together to form a simple trading code, TRD. An example was published on inter-day trading the S&P 500, using stops for losses on short and long trades and using CMI of the coupled cash and futures data. Data for years 1989 and 1990 was used, wherein one of the years was used to train TRD, and the other year to test TRD; then the years were reversed to establish two examples of trading on two years of quite different data.

In the 1991 study, it was noted that the sensitivity of testing trades to CMI overshadowed any sensitivity to the stops. Therefore, a second study was performed on this same data, but using only CMI. Better results were obtained, but more important, this established that the CMI themselves could lead to profitable trading, taking advantage of inefficiencies in these markets. Therefore, CMI at least can be useful supplemental indicators for other trading systems.

Inner-Shell Dynamic Model

The model contains an inner-shell and an outer-shell, both of which need to be optimally fit to data. The inner shell develops CMI as trading indicators. Many traders use such indicators as price, volume, etc., to trade, but here we explored the use of CMI to see if a "true" quantitative measure of momenta could be used. This required that the CMI be fit to data, e.g., to define quantities that themselves are functions of price. The cost function for the CMI is directly related to the Lagrangian. For the S&P studies, a two-variable model was used of end-of-day cash and futures, $c(r)$ and $f(r)$, for day r, taking the variables to be ratios between days, e.g.,

$$C(r) = \frac{c(r)}{c(r-1)} ,$$

$$F(r) = \frac{f(r)}{f(r-1)} .$$

These ratio-variables were used to develop coupled SDEs,

$$\dot{C} = f_C^C C + f_C^F F + g_1^C \eta^1 + g_2^C \eta^2$$

$$\dot{F} = f_F^C C + f_F^F F + g_1^F \eta^1 + g_2^F \eta^2$$

where all eight f and g parameters were taken to be constants, and the two η's were independent sources of Gaussian-Markovian noise. This set of SDEs were recast into a Lagrangian representation to define a single cost function, whose parameters were fit by ASA to data.

Outer-Shell Trading Model

The outer shell is the set of trading rules, defined as moving averages of the momenta indicators over various sized windows. This is pretty much like many simple trading rules, but here ASA is used to find the optimal sizes of the windows and of the thresholds triggering trading actions. Here, the thresholds are in terms of the CMI.

The CMI are fit to a year's worth of data, but they are continually fine-tuned within the widest moving window used in the outer shell. The cost function for the trading rules is the net profit over a year of data. Over the course of a year, every day a trading decision is made on the CMI, but only after the CMI are tuned using optimization over the widest window. This defines the need for recursive optimization.

Recursive ASA Fits

A simple outer-shell AI-type model for trading was defined for the TRD code. A wide and a narrow window were defined, whose widths were parameters of TRD. These windows defined epochs over which moving averages of CMI were calculated for both the C and F variables. For each window, a short and a long "threshold" parameter were defined. If the CMI of both C and F were above the thresholds in both the wide and narrow windows, then a long trade was executed or maintained for the futures market. Similarly, if the CMI of C and F fell below the negative of these threshold parameters in the two windows, a short trade was executed or maintained. Otherwise, no trade was executed.

Thus, the six parameters of the outer-shell were the widths of the two windows and the two threshold parameters for each of the two variables. Each day, the CMI were determined by an inner-shell optimization: Over the length of the wide window, using the zeroth-order prior fit as a first guess, two of the CMI parameters were refit to the data in the window. At first, ASA was used recursively to establish the best fit, but it was determined for this system that only small perturbations of the CMI were regularly required, and so thereafter a faster quasi-local code was used.

Stepping through the trading decisions each trading day of a year's data determined the yearly net profit/loss as the single value of the outer-shell cost function. ASA then importance-sampled the CMI parameter space to determine the largest net profit, determining the final CMI parameters in the training set.

The CMI parameter values in TRD were then used to trade for an out-of-sample year of data. The inner-shell optimization was performed each day as before.

Tick Resolution CMI Trading

We have developed a fully electronic trading system based on a published methodology. The published results are briefly discussed.

Our S&P model for the evolution of futures price F is

$$dF = \mu\, dt + \sigma F^x dz \ ,$$

$$< dz > = 0 \ ,$$

$$< dz(t)\, dz(t') > = dt\, \delta(t - t') \ ,$$

where the exponent x of F is one of the dynamical parameters to be fit to futures data together with μ and σ.

We have used this model in several ways to fit the distribution's volatility defined in terms of a scale and an exponent of the independent variable.

Intuitive Value of CMI

In the context of other invariant measures, the CMI transform covariantly under Riemannian transformations, but are more sensitive measures of activity than other invariants such as the energy density, effectively the square of the CMI, or the information which also effectively is in terms of the square of the CMI (essentially integrals over quantities proportional to the energy times a factor of an exponential including the energy as an argument). Neither the energy or the information give details of the components as do the CMI. In oscillatory markets the relative signs of such activity can be quite important.

The CMI present single indicators for each member of a set of correlated markets, "orthogonal" in the defined metric space. Each indicator is a dynamic weighting of short-time differenced deviations from drifts (trends) divided by covariances (risks). Thus the CMI also give information complementary to just trends or standard deviations separately.

Here we report results of our one-variable trading model. However, it is straightforward to include multi-variable trading models in our approach, and we have done this, for example, with coupled cash and futures S&P markets.

Multiple Local Minima

Our criteria for the global minimum of our cost function is minus the largest profit over a selected training data set (or in some cases, this value divided by the maximum drawdown). However, in many cases this may not give us the best set of parameters to find profitable trading in test sets or in real-time trading. Other considerations such as the total number of trades developed by the global minimum versus other close local minima may be relevant. For example, if the global minimum has just a few trades, while some nearby local minima (in terms of the value of the cost function) have many trades and was profitable in spite of our slippage factors, then the scenario with more trades might be more statistically dependable to deliver profits across testing and real-time data sets.

Therefore, for the outer-shell global optimization of training sets, we have used an ASA OPTION, MULTI_MIN, which saves a user-defined number of closest local minima within a user-defined resolution of the parameters. We then examine these results under several testing sets.

Use of CMI

Based on previous work applied to daily closing data, the overall structure of the trading system consists in 2 layers, as follows: We first construct the "short-time" Lagrangian function in the Itô representation (with the notation introduced in Section 3.3)

$$L(i|i-1) = \frac{1}{2\sigma^2 F_{i-1}^{2x}} \left(\frac{dF_i}{dt} - f^F \right)^2$$

with i the post-point index, corresponding to the one factor price model

$$dF = f^F dt + \sigma F^x dz(t) ,$$

where f^F and $\sigma > 0$ are taken to be constants, $F(t)$ is the S&P future price, and dz is the standard Gaussian noise with zero mean and unit standard deviation. We perform a global, maximum likelihood fit to the whole set of price data using ASA. This procedure produces the optimization parameters $\{x, f^F\}$ that are used to generate the CMI. One computational approach was to fix the diffusion multiplier σ to 1 during training for convenience, but used as free parameters in the adaptive testing and real-time fits.

To enhance the CMI sensitivity and response time to local variations (across a certain window size) in the distribution of price increments, the momenta are generated applying an adaptive procedure, i.e., after each new data reading another set of $\{f^F, \sigma\}$ parameters are calculated for the last window of data, with the exponent x — a contextual indicator of the noise statistics — fixed to the value obtained from the global fit.

Data Processing

The CMI formalism is general and by construction permits us to treat multivariate coupled markets. In certain conditions (e.g., shorter time scales of data), and also due to superior scalability across different markets, it is desirable to have a trading system for a single instrument, in our case the S&P futures contracts that are traded electronically on Chicago Mercantile Exchange (CME). The focus of our system was intra-day trading, at time scales of data used in generating the buy/sell signals from 10 to 60 secs.

A data point in our model does not necessarily mean an actual tick datum. For some trading time scales and for noise reduction purposes, data is pre-processed into sampling bins of length Δt using a standard exponential averaging procedure. Here we have used disjoint block bins and a standard average of the tick data with time stamps falling within the bin width.

Inner-Shell Optimization

In the present case there are six parameters: the time resolution Δt of price data, the length of window W used in the local fitting procedures and in computation of moving averages of trading signals, the drift f^F, volatility coefficient σ and exponent x. and a multiplicative factor M necessary for the trading rules module, as discussed below.

The training optimization occurs in two inter-related stages. An inner-shell maximum likelihood optimization over all training data is performed. The cost function that is fitted to data is the effective action constructed from the Lagrangian including the pre-factors coming from the measure element in the expression of the short-time probability distribution.

The fitted parameters were the drift coefficient f^F and the exponent x. In the case of a coupled futures and cash system, besides the corresponding values of f^F and x for the cash index, another parameter, the correlation coefficient ρ must be considered.

Trading Rules (Outer-Shell) Recursive Optimization

Three external shell optimization parameters are defined: the time resolution Δt of the data expressed as the time interval between consecutive data points, the window length W (in number of time epochs or data points) used in the adaptive calculation of CMI, and a numerical coefficient M that scales the momentum uncertainty.

At each moment a local refit of f^F and σ over data in the local window W is executed, moving the window M across the training data set and using the zeroth order optimization parameters f^F and x resulting from the inner-shell optimization as a first guess.

Here we use the one factor nonlinear model. At each time epoch we calculate the following momentum related quantities:

$$\Pi^F = \frac{1}{\sigma^2 F^{2x}} \left(\frac{dF}{dt} - f^F \right),$$

$$\Pi^F_0 = -\frac{f^F}{\sigma^2 F^{2x}} ,$$

$$\Delta\Pi^F = < (\Pi^F - < \Pi^F >)^2 >^{1/2} = \frac{1}{\sigma F^x \sqrt{dt}} ,$$

where we have used $< \Pi^F >= 0$. In the previous expressions, Π^F is the CMI, Π^F_0 is the neutral line or the momentum of a zero change in prices, and $\Delta\Pi^F$ is the uncertainty of momentum. The last quantity reflects the "uncertainty" principle,

$$\Delta F \equiv < (dF - < dF >)^2 >^{1/2} = \sigma F^x \sqrt{dt} ,$$

$$\Delta\Pi^F \, \Delta F \geq 1 ,$$

where all expectations are in terms of the exact noise distribution.

CMI Neutral Line

The neutral line divides all CMI in two classes: long signals, when $\Pi^F > \Pi^F_0$, as any CMI satisfying this condition indicates a positive price change, and short signals when $\Pi^F < \Pi^F_0$, which reflects a negative price change.

After the CMI are calculated, based on their meaning as statistical momentum indicators, trades are executed following a relatively simple model: Entry in and exit from a long (short) trade points are defined as points where the value of CMIs is greater (smaller) than a certain fraction of the uncertainty band $M\,\Delta\Pi^F$ ($-M\,\Delta\Pi^F$), where M is the multiplicative factor mentioned in the beginning of this subsection.

Training –> Testing

The values of the optimization parameters $\{\Delta t, W, M, f^F, \sigma, x\}$ resulting from a training cycle are then applied to out-of-sample test sets. During the test run, the drift coefficient f^F and the volatility coefficient σ are refitted adaptively as described previously. All other parameters are fixed. We have mentioned that the optimization parameters corresponding to the highest profit in the training set may not be the sufficiently robust across test sets. Then, for all test sets, we have tested optimization parameters related to the multiple minima (i.e., the global maximum profit, the second best profit, etc.) resulting from the training set.

Alternative Algorithms

In the previous sections we noted that there are different combinations of methods of processing data, methods of computing the CMI and various sets of trading rules that need to be tested — at least in a sampling manner — before launching trading runs in real-time:

1. Data can be preprocessed in block or overlapping bins, or forecasted data derived from the most probable transition path could be used as in one of our most recent models.

2. Exponential smoothing, wavelets or Fourier decomposition can be applied for statistical processing. We presently favor exponential moving averages.

3. The CMI can be calculated using averaged data or directly with tick data, although the optimization parameters were fitted from preprocessed (averaged) price data.

4. The trading rules can be based on current signals (no average is performed over the signal themselves), on various averages of the CMI trading signals, on various combination of CMI data (momenta, neutral line, uncertainty band), on symmetric or asymmetric trading rules, or on mixed price-CMI trading signals.

5. Different models (one and two-factors coupled) can be applied to the same market instrument, e.g., to define complementary indicators.

The selection process evidently must consider many specific economic factors (e.g., liquidity of a given market), besides all other physical, mathematical and technical considerations. In the work presented here, as we tested our system and using previous experience, we focused toward S&P500 futures electronic trading, using block processed data, and symmetric, local-model and multiple-models trading rules based on CMI neutral line and stay-in conditions.

Slippage

Generally, most trading costs can be included under the "slippage factor," although this could easily lead to poor estimates. Given that the margin of profits from exploiting market inefficiencies are thin, a high slippage factor can easily result in a non-profitable trading system.

In our situation, for testing purposes we used a $35 slippage factor per buy & sell order, a value we believe is rather high for an electronic trading environment, although it represents less than three ticks of a mini-S&P futures contract. This higher value was chosen to protect ourselves against the bid-ask spread, as our trigger price (at what price the CMI was generated) and execution price (at what price a trade signaled by a CMI was executed) were taken to be equal to the trading price. (We have changed this aspect of our algorithm in later models.) The slippage is also strongly influenced by the time resolution of the data.

Although the slippage is linked to bid-ask spreads and markets volatility in various formulas, the best estimate is obtained from experience and actual trading.

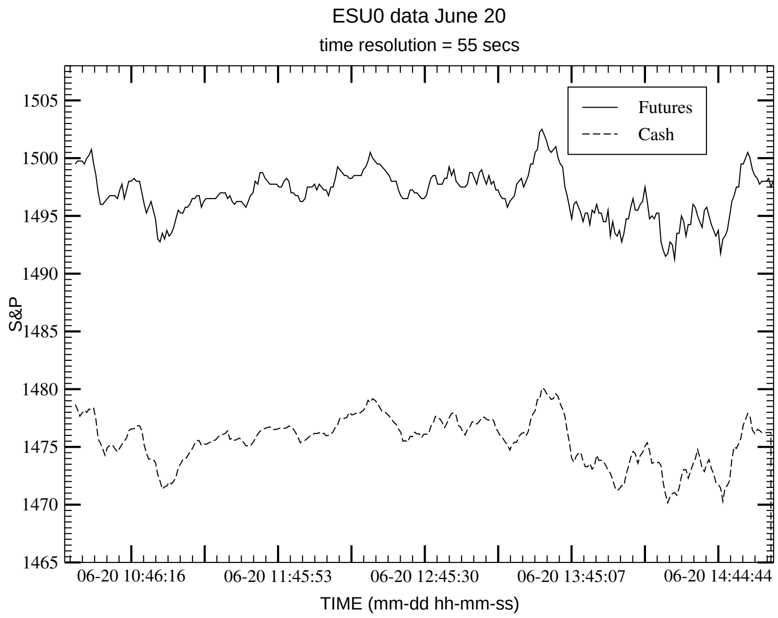
Money Management

Money management was introduced in terms of a trailing stop condition that is a function of the price volatility, and a stop-loss threshold that we fixed by experiment to a multiple of the mini-S&P contract value ($200). It is tempting to tighten the trailing stop or to work with a small stop-loss value, yet we found — as otherwise expected — that higher losses occurred as the signals generated by our stochastic model were bypassed.

Regarding the execution process, we have to account for the response of the system to various execution conditions in the interaction with the electronic exchange: partial fills, rejections, uptick rule (for equity trading), etc. Except for some special conditions, all these steps must be automated.

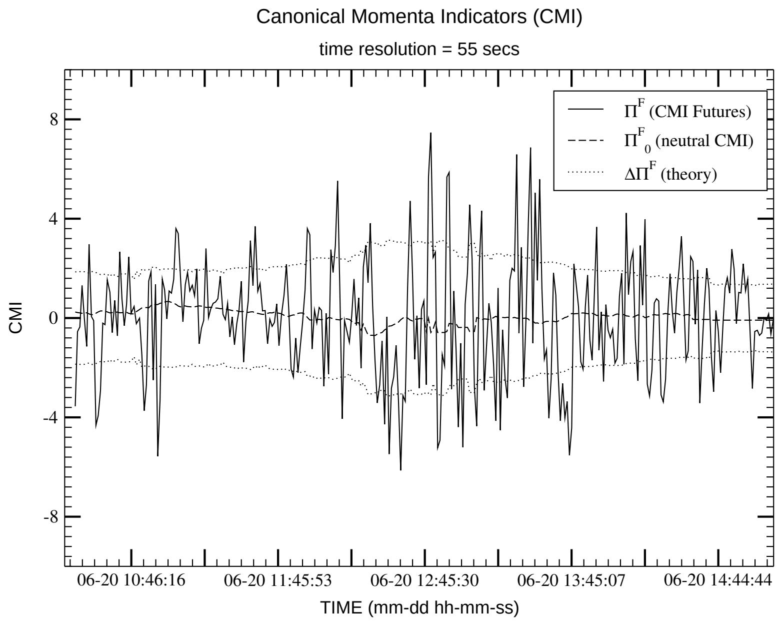
Example: Data Used

We have plotted the fastest (55 secs apart) CMI values Π^F, the neutral line Π_0^F and the uncertainty band $\Delta\Pi^F$. All CMI data were produced using the optimization parameters set $\{55 \text{ secs}, 88 \text{ epochs}, 0.15\}$ of the second-best net profit obtained with the training set "4D ESM0 0321-0324."



Example: Derived CMI

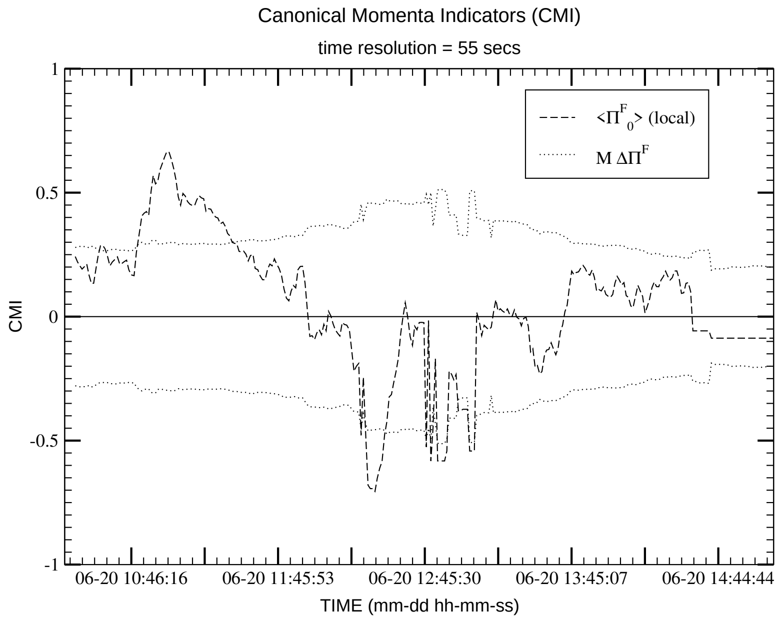
Although the CMIs exhibit an inherently ragged nature and oscillate around a zero mean value within the uncertainty band — the width of which is decreasing with increasing price volatility, as the uncertainty principle would also indicate — time scales at which the CMI average or some persistence time are not balanced about the neutral line.



Example: Long-Averaged Derived CMI

An important factor is the average length of the trades. For the type of rules presented in this work, this length is of several minutes, up to one hour, as the time scale of the local fit window mentioned above suggested.

Related to the length of a trade is the length of a winning long/short trade in comparison to a losing long/short trade. Our experience indicates that a ratio of 2:1 between the length of a winning trade and the length of a losing trade is desirable for a reliable trading system. Here, using the local-model trading rules seems to offer an advantage, although this is not as clear as one would expect.



Main Features

The main stages of building and testing this system were:

1. We developed a multivariate, nonlinear statistical mechanics model of S&P futures and cash markets, based on a system of coupled stochastic differential equations.

2. We constructed a two-stage, recursive optimization procedure using methods of ASA global optimization: An inner-shell extracts the characteristics of the stochastic price distribution and an outer-shell generates the technical indicators and optimize the trading rules.

3. We trained the system on different sets of data and retained the multiple minima generated (corresponding to the global maximum net profit realized and the neighboring profit maxima).

4. We tested the system on out-of-sample data sets, searching for most robust optimization parameters to be used in real-time trading. Robustness was estimated by the cumulative profit/loss across diverse test sets, and by testing the system against a bootstrap-type reversal of training-testing sets in the optimization cycle.

Modeling the market as a dynamical physical system makes possible a direct representation of empirical notions as market momentum in terms of CMI derived naturally from our theoretical model. We have shown that other physical concepts as the uncertainty principle may lead to quantitative signals (the momentum uncertainty band $\Delta\Pi^F$) that captures other aspects of market dynamics and which can be used in real-time trading.

We have found that keeping our model faithful to the underlying mathematical physics is not a limiting constraint on profitability of our system; quite the contrary.

APPLICATIONS TO OPTIONS

Eurodollar Volatility Modeling

The correlations we observe are specific to the time scales and the windows used for averaging out data. These scales and windows have been selected because of their widespread use in actual trading. However, this may not imply correlations at other times scales or windows. The basis-point volatility (BPV) is derived from the observed sample data of daily differenced yields ($dt = 1$), e.g.,

$$\mathrm{BPV} = < \frac{n}{n-1} (dy - < dy >)^2 >^{1/2}$$

for a sample size of n points of data. The BPV is calculated over a sample of daily data, e.g., 20 days,

$$\mathrm{BPV} \approx < y > \sigma \ .$$

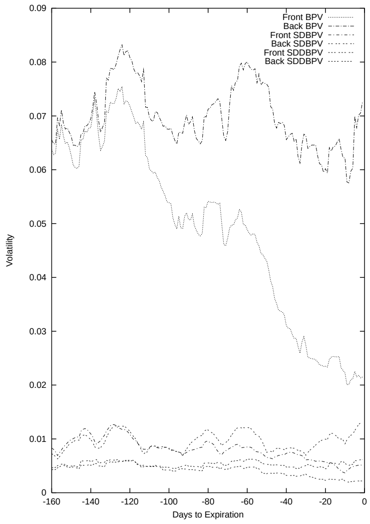
The BPV is a more natural measure of the movement of the yields and is used more by traders than by theorists. We regularly draw comparisons in our calculations between the BPV and the BS volatility (BSV), i.e., σ, to be sure that we are not inducing some effects by the choice of one volatility over the other.

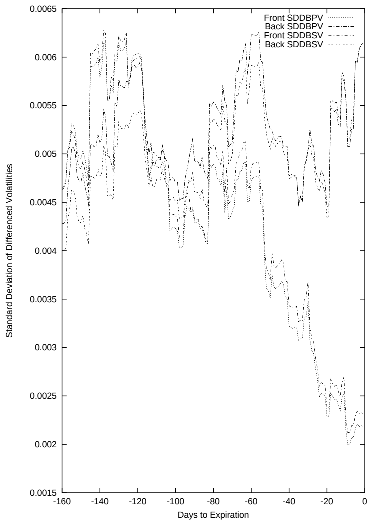
We first take standard deviations of the volatilities of each contract, then take the average over similar contracts, e.g., with the same number of days until expiration. This establishes that there is a distribution of volatilities over similar contracts, beyond the act of their aggregation.

Eurodollar Volatility of Historical Volatility

The top figure gives a comparison of Basis-Point Volatility (BPV), standard deviation of BPV (SDBPV), and standard deviation of differenced BPV (SDDBPV), for Front and Back contracts. The SDBPV illustrate that there exists a distribution of volatilities about the mean volatility. The SDDBPV illustrate that this distribution likely is a stochastic process with a constant diffusion.

The bottom figure gives a comparison of standard deviation of differenced Basis-Point Volatility (SDDBPV) and standard deviation of differenced Black-Scholes Volatility (SDDBSV), for Front and Back contracts. The SDDBSV have been scaled to the SDDBPV by multiplying them by the rounded average of the yields, i.e., 6.0. Note that after scaling, they consistently lie close to each other. Thus, both the BPV and BSV have volatilities that can be considered to be stochastic processes with constant diffusion.

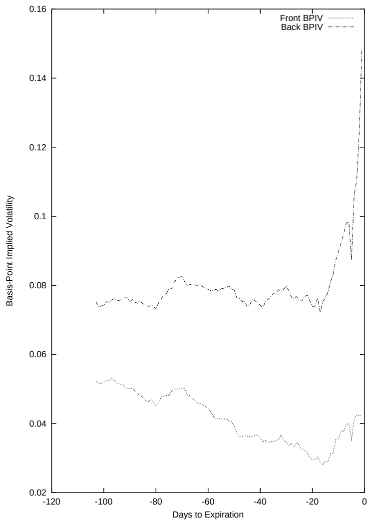


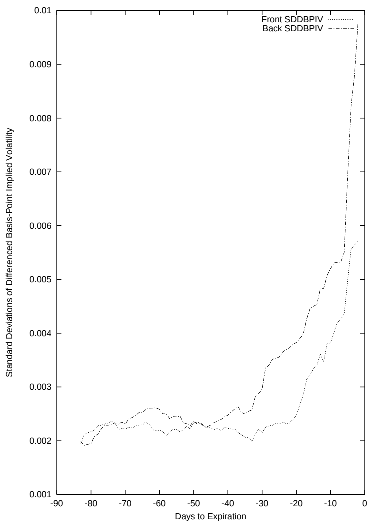


Eurodollar Volatility of Implied Volatility

The top figure shows the Basis-Point Implied Volatility (BPIV) of Front and Back contracts.

The bottom figure shows the standard deviation of differenced Basis-Point Implied Volatility (SDDBPIV) of Front and Back contracts. Note that, similar to the results with historical volatilities, this illustrates that the volatility of the implied volatilities appears to be a stochastic process with constant diffusion for times less than 30 days before expiration; afterwards, the values still are within the same scale.





Power-Law Model

There is growing evidence that the Black-Scholes lognormal distribution has been less and less descriptive of markets over the past two decades. An example of a generalization of the lognormal distribution is

$$dS/F(S, x) = \mu\, dt + \sigma\, dw_S$$

$$F(S, S_0, S_\infty, x, y) = \begin{cases} S, & S < S_0 \\ S^x S_0^{1-x}, & S_0 \le S \le S_\infty \\ S^y S_0^{1-x} S_\infty^{x-y}, & S > S_\infty \end{cases}$$

where S_0 and S_∞ are selected to lie outside the data region used to fit the other parameters, e.g., $S_0 = 1$ and $S_\infty = 20$ for fits to Eurodollar futures. We have used the Black-Scholes form $F = S$ inside $S < S_0$ to obtain the usual benefits, e.g., no negative prices as the distribution is naturally excluded from $S < 0$, preservation of put-call parity, etc. We have taken $y = 0$ to reflect total ignorance of markets outside the range of $S > S_\infty$.

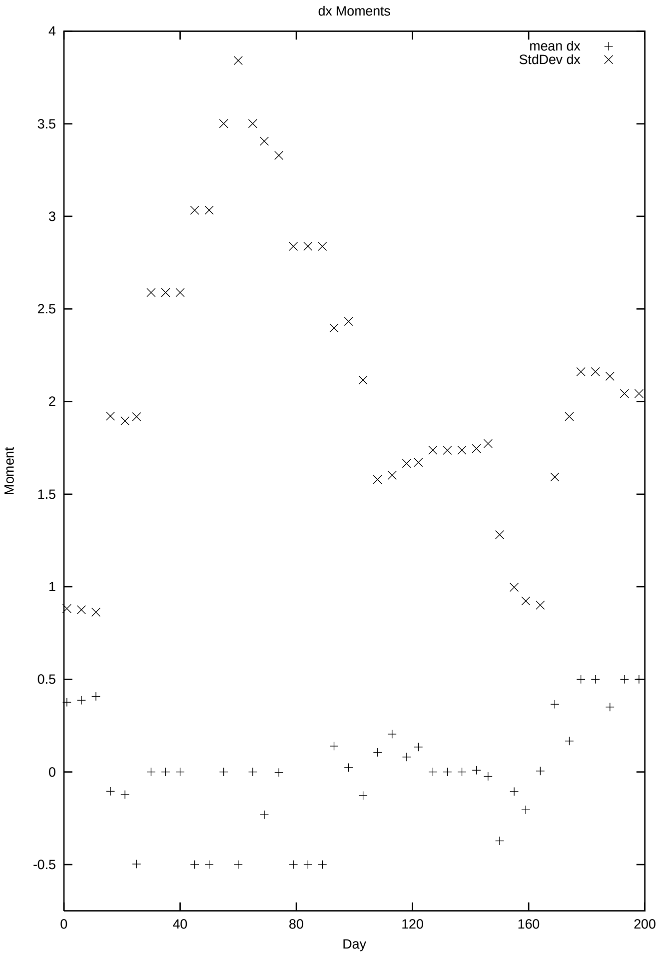
Stochasticity: Volatility vs Exponent

The one-factor exponents exhibit a random process that is approximately defined as a simple normal process η_x with mean μ_x and standard deviation σ_x,

$$\dot{x} = \frac{dx}{dt} = \mu_x + \sigma_x \eta_x$$

When averaging over a several month period, we can derive $\mu_x \approx 0$ and have σ_x essentially span all x's. However, it is clear that there are shorter periods of stochastic x which can be modeled independently, yielding a one-factor x as an indicator of market contexts.

Most important, the reasonable interpretation of our results is that suppression of stochastic volatility in the one-factor model just leaks out into stochasticity of parameters in the model, e.g., especially in x. By comparison, the x-exponents in the two-factor fits are quite stable.



Multi-Factor Volatility Model

Any study that geared to perform ASA fits of multivariate Lagrangians and PATHINT long-time calculations can also consider another variable σ, stochastic volatility, that can generalize the BS lognormal distribution:

$$dS/F(S, x) = \mu\, dt + \sigma\, dw_S$$

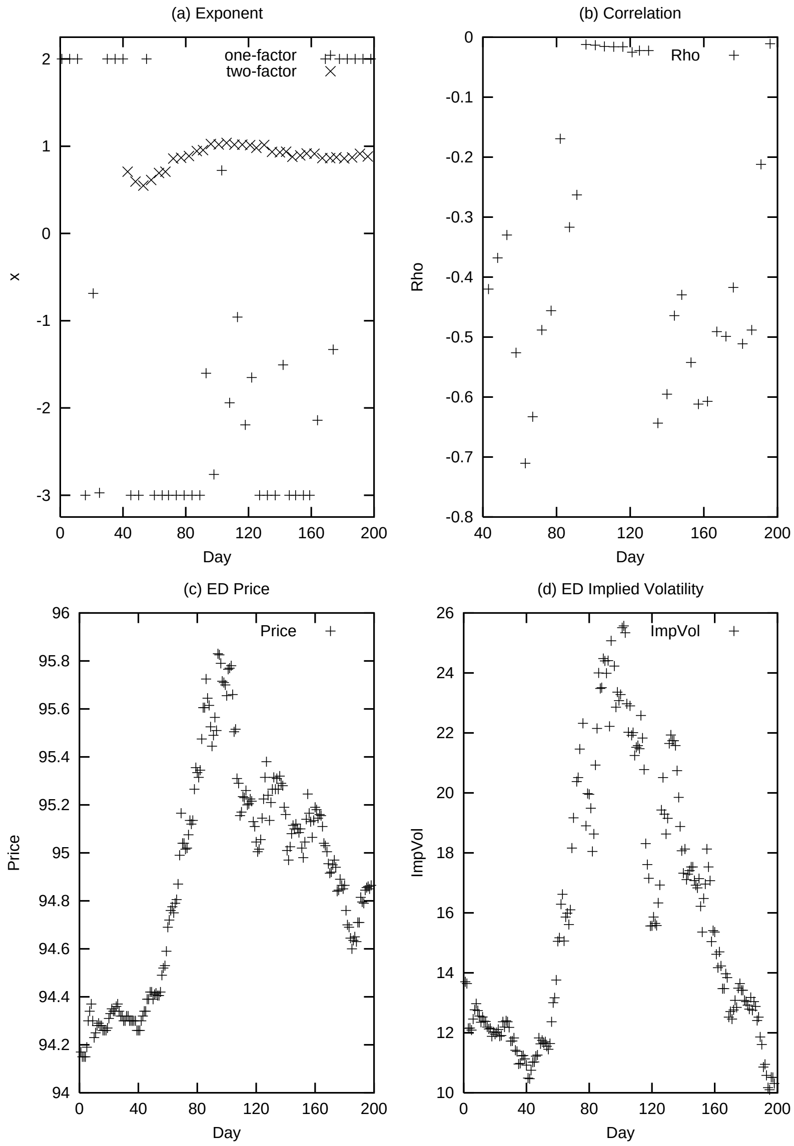
$$d\sigma = \nu + \varepsilon\, dw_\sigma$$

The current literature supports stochastic volatility as the best approach to describe these markets.

x Market Indicators

We have developed x's as indicators of different market contexts. E.g., x may be -2 for some quarter and +2 for a different quarter.

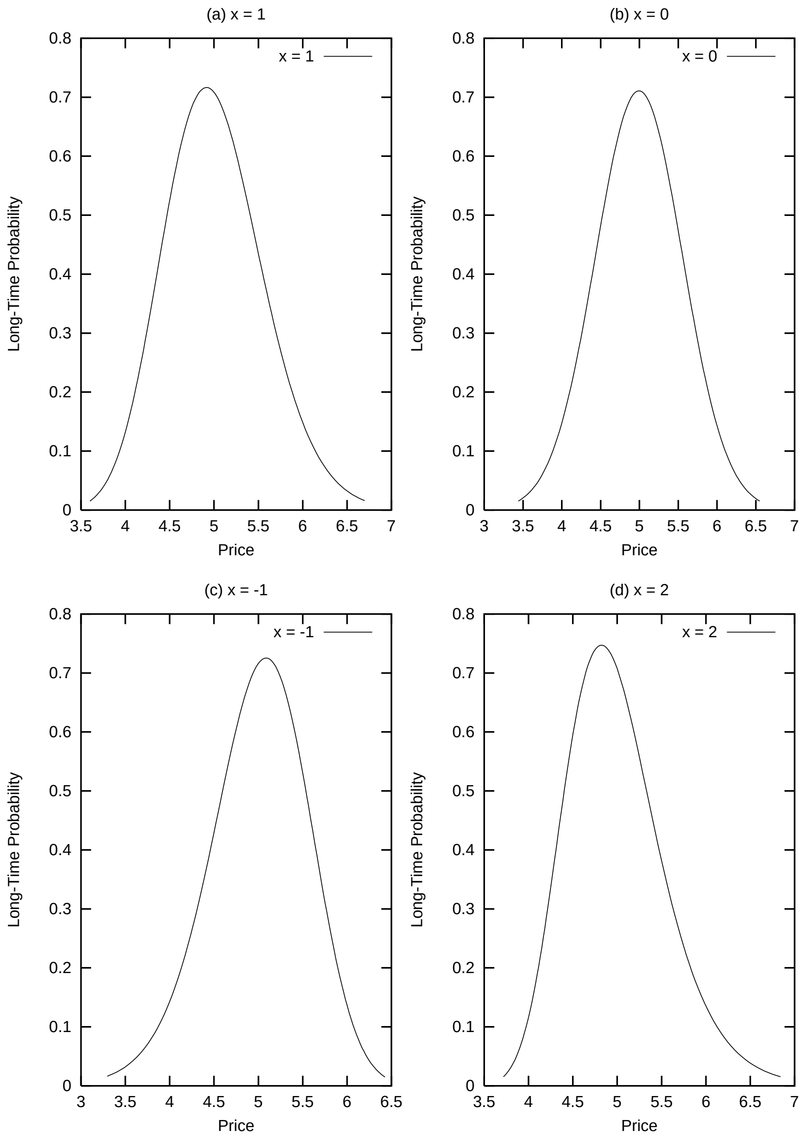
(a) Weekly two-month moving-averaged one-factor and two-factor exponents for ED contract expiring in September 1999 during the period June 1998 through March 1999. (b) Weekly two-month moving-averaged two-factor correlation ρ for this same data. (c) Raw price data used in fits for the above parameters. (d) Implied-volatility data used in fits for the above parameters.



PATHINT

PATHINT is being used to perform European and American, one-factor and two-factor, PATHINT calculations. Some examples are $F(S, S_0, S_\infty, x, y)\, dz_S$ for x in {-1, 0, 1, 2}.

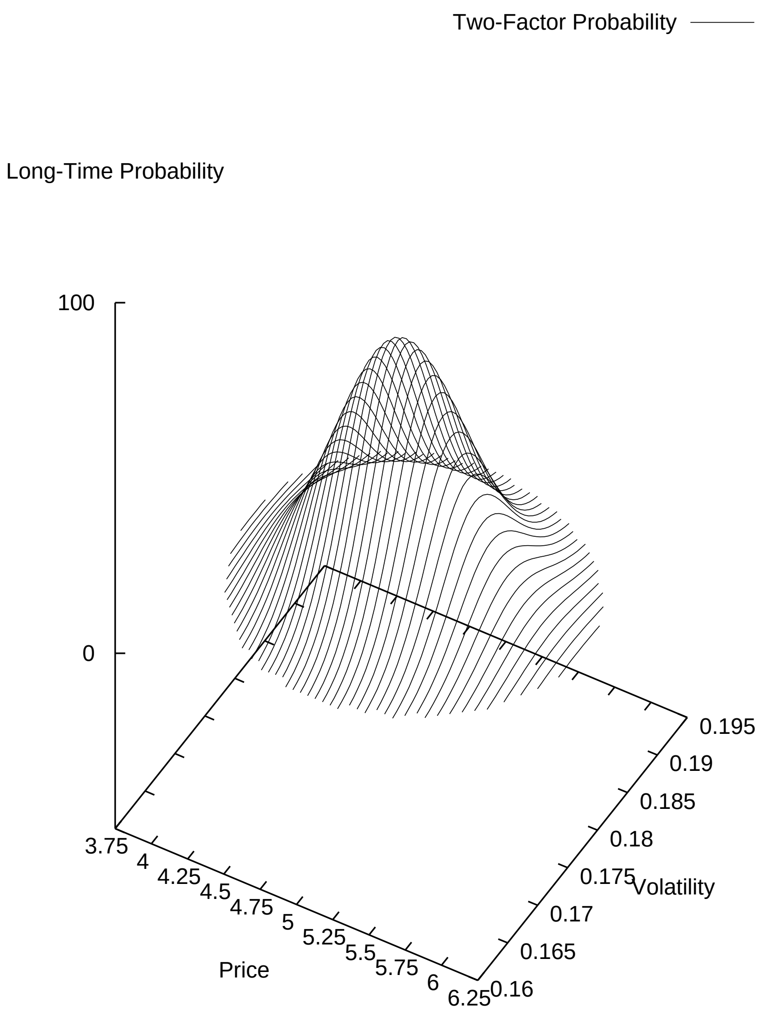
The short-time probability distribution at time $T = 0.5$ years for $x = 1$, the (truncated) Black-Scholes distribution. The short-time probability distribution at time $T = 0.5$ years for $x = 0$, the normal distribution. The short-time probability distribution at time $T = 0.5$ years for $x = -1$. The short-time probability distribution at time $T = 0.5$ years for $x = 2$.



PATHINT Two-Factor

The drawback of the two-factor PATHINT code is that it is slow. However, it is accurate and robust so we can process any diffusion for general x.

The two-factor distribution at time $T = 0.5$ years for $x = 0.7$.



BS vs CRR vs PATHINT

Calculation of prices and Greeks are given for closed form BS, binomial tree $\mathrm{CRR}_{\mathrm{European}}$, $\mathrm{CRR}_{\mathrm{American}}$, $\mathrm{CRR}_{\mathrm{variant}}$, and PATHINT. All CRR calculations include averaging over 300 and 301 iterations to minimize oscillatory errors.

Greek	BS	$\mathrm{CRR}_{\mathrm{European}}$	$\mathrm{CRR}_{\mathrm{American}}$	$\mathrm{CRR}_{\mathrm{variant}}$	PATHINT
Price	0.138	0.138	0.138	0.138	0.138
Delta	0.501	0.530	0.534	0.506	0.501
Gamma	1.100	1.142	1.159	1.116	1.100
Theta	-0.131	-0.130	-0.132	-0.133	-0.131
Rho	-0.0688	-0.0688	-0.0530	-0.0530	-0.0688
Vega	1.375	1.375	1.382	1.382	1.375

The other parameters are $S = 5$, $S_0 = 0.5$, $S_\infty = 20$, $y = 0$.

Portfolios of Options

We have used methods of nonlinear nonequilibrium multivariate statistical mechanics to develop short-time probability distributions, together with associated numerical algorithms, e.g., adaptive simulated annealing (ASA), to give a detailed methodology for trading portfolios of options.

Risk-slides are developed to assess distributions of P/Ls and of sets of specific Greeks across components that are correlated with an index, as a function of fractional StdDev moves in underlyings and volatilities. This approach also contributes to an algorithm for disbursement of component assets to be correlated with a disbursement on the index.

We develop an algorithm for realtime dynamic balancing to determine sets of reasonably optimal sub-baskets that satisfy risk constraints, while presenting alternative sorted and ranked buys and sells that maximize profits. This can be visualized as "surfing" a risk surface to benefit from profitable trades in the risk-management process.

We develop a generic value indicators, based on Training and Testing sets of "bubbles" of exiting and reentering statistical norms of indicators, e.g., prices, volatilities, ratios of other indicators, etc. Faster indicators, canonical momenta indictors (CMI) are correlated to these bubbles to give sets of sorted and ranked expected value indicators.