Statistical Mechanics of Financial Markets (SMFM)

Lester Ingber

Lester Ingber Research POB 06440 Wacker Dr PO Sears Tower Chicago, IL 60606 ingber@ingber.com • ingber@alumni.caltech.edu

and

DRW Investments LLC Chicago Mercantile Exchange Center 30 S Wacker Dr Ste 1516 Chicago, IL 60606 ingber@drwtrading.com

MATHEMATICAL PHYSICS

NUMERICAL ALGORITHMS

SOME FINANCIAL APPLICATIONS

SOME OUT-OF-FINANCE LESSONS LEARNED

SOME CURRENT PROJECTS

\$Id: markets_lecture,v 1.15 1999/04/17 12:52:45 ingber Exp ingber \$

The latest Adaptive Simulated Annealing (ASA) code and related reprints can be retrieved via WWW from http://www.ingber.com/ or via FTP from ftp.ingber.com.

Statistical Mechanics of Financial Markets (SMFM)

MATHEMATICAL PHYSICS		•										4
Nonlinear Nonequilibrium Multivaria	te S	Stoc	has	stic	Ag	ggre	egat	tion	۱.			5
Stochastic Differential Equation (SDI	E)											6
Partial Differential Equation (PDE).	•											7
Lagrangian Probability Distribution F	un	ctio	n (F	PDI	F)							8
Path-Integral Riemannian Geometry			•	•	•							9
Information		•										10
Transformations Without Itô Calculus	5.											11
Intuitive Variables		•		•								12
Euler-Lagrange Variational Equations	5.											13
Canonical Momenta Indicators (CMI).	•	•		•							14
Generic Mesoscopic Neural Network	s (N	ANI	N)		•			•				15
MNN Learning	Ì.											16
MNN Prediction												17
MNN Parallel Processing												18
NUMERICAL ALGORITHMS												19
Adaptive Simulated Annealing (ASA).											20
Hills and Valleys												21
Outline of ASA Algorithm												22
Index of Pre-Compile Tuning P	Para	met	ers									23
Index of Adaptive Tuning Para	met	ers		•	•	•	•	•	•	•	•	24
Reannealing Example												25
Example: Fitting Variance	•											26
Numerical Path Integration (PATHIN)	Т)	•	•	•	•	•	•	•	•	•	•	27
Outline of PATHINT Algorithm	-, n.											$\frac{-7}{28}$
Boundary Condition Sensitivity	7.	•	•	•	•	•	•	•	•	•	•	29
Mesh Limitations	•	•	•	•	•	•	•	•	•	•	•	30
SOME FINANCIAL APPLICATIONS	•	•	•	•	•	•	•	•	•	•	•	31
Risk-Averting Agents	•	•	•	•	•	•	•	•	•	•	•	32
Aggregation of Risk-Averting Agents	•	•	•	•	•	•	•	•	•	•	•	33
Interest Rates	•	•	•	•	•	•	•	•	•	•	•	34
2-Factor Interest-Rate Model	•	•	•	•	•	•	•	•	•	•	•	35
Lagrangian Representation	•	•	•	•	•	•	•	•	•	•	•	36
ASA Fits	•	•	•	•	•	•	•	•	•	•	•	37
Bond PDF/PATHINT	•	•	•	•	•	•	•	•	•	•	•	38
S&P	•	•	•	•	•	•	•	•	•	•	•	30
S&P Interday Futures_Cash	•	•	•	•	•	•	•	•	•	•	•	<u>40</u>
Inner-Shell Dynamic Model	•	•	•	•	•	•	•	•	•	•	•	- - -0 //1
Outer-Shell Trading Model	•	•	•	•	•	•	•	•	•	•	•	+1 ∕\?
Recursive $\Delta S \Delta$ Fits	•	•	•	•	•	•	•	•	•	•	•	 Δ3
S&P Interday Futures-Cash . Inner-Shell Dynamic Model . Outer-Shell Trading Model . Recursive ASA Fits	•					•	•		• • •	•		40 41 42 43

SOME OUT-OF-FINANCE LESSONS LEARNED				•	44
Statistical Mechanics of Combat (SMC)		•		•	45
National Training Center — Janus(T)		•		•	46
Basic Equations of Aggregated Units		•		•	47
Long-Time Correlations Test Short-Time Models				•	48
СМІ				•	49
Statistical Mechanics of Neocortical Interactions (SMNI)	•	•		•	50
Basic SMNI				•	51
Multiple Scales				•	52
Microscopic Aggregation				•	53
Short-Term Memory (STM)				•	54
PDE Stability & Duration Analyses				•	55
STM PATHINT Details				•	56
Electroencephalography (EEG)	•			•	57
ASA Fits of CMI				•	58
Chaos in SMNI	•			•	59
Duffing Analog				•	60
Noise Washes Out Chaos/PATHINT				•	61
SOME CURRENT PROJECTS					62
Volatility Modeling					63
Eurodollar Spreads					64
Eurodollar Volatility of Historical Volatility				•	65
Eurodollar Volatility of Implied Volatility					66
Power-Law Model					67
Multi-Factor Volatility Model				•	68
PATHINT					69
PATHINT Two-Factor	•	•		•	70
x Market Indicators	•	•		•	71
Stochasticity: Volatility vs Exponent	•	•		•	72
BS vs CRR vs PATHINT				•	73
Tick Resolution CMI Trading				•	74

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.

$$\begin{split} dM^{G} &= f^{G}(t, M(t))dt + \hat{g}_{i}^{G}(t, M(t))dW^{i} ,\\ \dot{M}^{G}(t) &= f^{G}(t, M(t)) + \hat{g}_{i}^{G}(t, M(t))\eta^{i}(t) ,\\ dW^{i} &\to \eta^{i}dt ,\\ M &= \{ M^{G}; G = 1, \cdots, \Lambda \} ,\\ \eta &= \{ \eta^{i}; i = 1, \cdots, N \} .\\ \dot{M}^{G} &= dM^{G}/dt ,\\ &< \eta^{j}(t) >_{\eta} = 0 ,\\ &< \eta^{j}(t), \eta^{j'}(t') >_{\eta} = \delta^{jj'} \delta(t - t') , \end{split}$$

 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 $\theta = \Delta t$,

,

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

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

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

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

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

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

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

Partial Differential Equation (PDE)

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

$$\begin{split} P_{,t} &= \frac{1}{2} \, (g^{GG'} P)_{,GG'} - (g^G P)_{,G} + VP \ , \\ P &= < P_{\eta} >_{\eta} \ , \\ g^G &= f^G + \frac{1}{2} \, \hat{g}_i^{G'} \hat{g}_{i,G'}^G \ , \\ g^{GG'} &= \hat{g}_i^G \, \hat{g}_i^{G'} \ , \\ (\cdots)_{,G} &= \partial (\cdots) / \partial M^G \ . \end{split}$$

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,

$$\begin{split} P[M_t|M_{t_0}]dM(t) &= \\ &\int \cdots \int \underline{D}M \exp\left(-\min \int_{t_0}^t dt'L\right) \delta(M(t_0) = M_0) \delta(M(t) = M_t) , \\ DM &= \lim_{u \to \infty} \prod_{\rho=1}^{u+1} g^{1/2} \prod_G (2\pi\theta)^{-1/2} dM_{\rho}^G , \\ L(\dot{M}^G, M^G, t) &= \frac{1}{2} (\dot{M}^G - h^G) g_{GG'} (\dot{M}^{G'} - h^{G'}) + \frac{1}{2} h^G_{;G} + R/6 - V , \\ [\cdots]_{,G} &= \frac{\partial [\cdots]}{\partial M^G} , \\ h^G &= g^G - \frac{1}{2} g^{-1/2} (g^{1/2} g^{GG'})_{,G'} , \\ g_{GG'} &= (g^{GG'})^{-1} , \\ g &= \det(g_{GG'}) , \\ h^G_{;G} &= h^G_{,G} + \Gamma^F_{GF} h^G = g^{-1/2} (g^{1/2} h^G)_{,G} , \\ \Gamma^F_{JK} &\equiv g^{LF} [JK, L] = g^{LF} (g_{JL,K} + g_{KL,J} - g_{JK,L}) , \\ R &= g^{JL} R_{JL} = g^{JL} g^{JK} R_{FJKL} , \\ R_{FJKL} &= \frac{1}{2} (g_{FK,JL} - g_{JK,FL} - g_{FL,JK} + g_{JL,FK}) + g_{MN} (\Gamma^M_{FK} \Gamma^N_{JL} - \Gamma^M_{FL} \Gamma^N_{JK}) . \end{split}$$

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,

$$\begin{split} M(\bar{t}_{s}) &= M(t_{s}) \;, \\ \bar{L} &= \frac{1}{2} \, (dM^{G}/dt - g^{G}) g_{GG'}(dM^{G'}/dt - g^{G'}) - V \;, \end{split}$$

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.

Information

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

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

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 \overline{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

$$\begin{split} L' \delta t &\approx \frac{\left(Z_{t+\delta t} - Z_t + \frac{1}{2} \left(Z_{t+\delta t} - Z_t\right)^2 - r \delta t\right)\right)^2}{2\sigma^2 \delta t} + r \delta t \\ &\approx \frac{\left(Z_{t+\delta t} - Z_t - \left(r - \frac{1}{2} \sigma^2\right) \delta t\right)^2}{2\sigma^2 \delta t} + r \delta t \ , \\ \left(Z_{t+\delta t} - Z_t\right)^2 &\approx \sigma^2 \delta t \ , \end{split}$$

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

$$\label{eq:F} \begin{split} ``F = ma": \ \delta L = 0 &= \frac{\partial L}{\partial M^G} - \frac{\partial}{\partial t} \ \frac{\partial L}{\partial (\partial M^G/\partial t)} \\ ``Force" &= \frac{\partial L}{\partial M^G} \ , \\ ``Mass" &= g_{GG'} = \frac{\partial^2 L}{\partial (\partial M^G/\partial t) \partial (\partial M^{G'}/\partial t)} \ , \end{split}$$

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

"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.

Generic Mesoscopic Neural Networks (MNN)

Modern stochastic calculus permits development of alternative descriptions of path-integral Lagrangians, Fokker-Planck equations, and Langevin rate equations. The induced Riemannian geometry affords invariance of probability distribution under general nonlinear transformations.

ASA presents a powerful global optimization that has been tested in a variety of problems defined by nonlinear Lagrangians.

Parallel-processing computations can be applied to ASA as well as to a neural-network architecture.

MNN Learning

"Learning" takes place by presenting the MNN with data, and parametrizing the data in terms of the "firings," or multivariate M^G "spins." The "weights," or coefficients of functions of M^G appearing in the drifts and diffusions, are fit to incoming data, considering the joint "effective" Lagrangian (including the logarithm of the prefactor in the probability distribution) as a dynamic cost function.

The cost function is a sum of effective Lagrangians from each node and over each time epoch of data.

This program of fitting coefficients in Lagrangian uses methods of adaptive simulated annealing (ASA). This maximum likelihood procedure (statistically) avoids problems of trapping in local minima, as experienced by other types of gradient and regression techniques.

MNN Prediction

"Prediction" takes advantage of a mathematically equivalent representation of the Lagrangian path-integral algorithm, i.e., a set of coupled Langevin rate-equations. The Itô (prepoint-discretized) Langevin equation is analyzed in terms of the Wiener process dW^i , which is rewritten in terms of Gaussian noise $\eta^i = dW^i/dt$ in the limit:

$$\begin{split} M^{G}(t + \Delta t) &- M^{G}(t) = dM^{G} = g^{G}dt + \hat{g}_{i}^{G}dW^{i} ,\\ \frac{dM^{G}}{dt} &= \dot{M}^{G} = g^{G} + \hat{g}_{i}^{G}\eta^{i} ,\\ M &= \{ M^{G}; G = 1, \cdots, \Lambda \} , \eta = \{ \eta^{i}; i = 1, \cdots, N \} ,\\ &< \eta^{j}(t) >_{\eta} = 0 , < \eta^{j}(t), \eta^{j'}(t') >_{\eta} = \delta^{jj'}\delta(t - t') . \end{split}$$

Moments of an arbitrary function $F(\eta)$ over this stochastic space are defined by a path integral over η^i . The Lagrangian diffusions are calculated as

$$g^{GG'} = \sum_{i=1}^{N} \hat{g}_{i}^{G} \hat{g}_{i}^{G'}$$
.

MNN Parallel Processing

The use of parallel processors can make this algorithm even more efficient, as ASA lends itself well to parallelization.

During "learning," blocks of random numbers are generated in parallel, and then sequentially checked to find a generating point satisfying all boundary conditions.

Advantage is taken of the low ratio of acceptance to generated points typical in ASA, to generate blocks of cost functions, and then sequentially checked to find the next best current minimum.

Additionally, when fitting dynamic systems, e.g., the three physical systems examined to date, parallelization is attained by independently calculating each time epoch's contribution to the cost function.

Similarly, during "prediction," blocks of random numbers are generated to support the Langevin-equation calculations, and each node is processed in parallel.

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, e.g., genetic algorithms. 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."

Index of Pre-Compile Tuning Parameters

OPTIONS_FILE OPTIONS_FILE_DATA **RECUR OPTIONS FILE** RECUR_OPTIONS_FILE_DATA COST_FILE ASA_LIB HAVE ANSI **IO PROTOTYPES** TIME CALC TIME_STD TIME GETRUSAGE INT LONG INT ALLOC SMALL_FLOAT MIN DOUBLE MAX DOUBLE **EPS_DOUBLE** CHECK EXPONENT NO_PARAM_TEMP_TEST NO_COST_TEMP_TEST SELF_OPTIMIZE ASA_TEST ASA_TEST_POINT MY TEMPLATE USER_INITIAL_COST_TEMP RATIO_TEMPERATURE_SCALES USER INITIAL PARAMETERS TEMPS DELTA_PARAMETERS QUENCH_PARAMETERS QUENCH COST QUENCH_PARAMETERS_SCALE QUENCH_COST_SCALE ASA TEMPLATE OPTIONAL_DATA OPTIONAL DATA INT USER_COST_SCHEDULE USER ACCEPT ASYMP EXP USER ACCEPTANCE TEST USER GENERATING FUNCTION USER REANNEAL COST USER_REANNEAL_PARAMETERS MAXIMUM REANNEAL INDEX REANNEAL SCALE ASA_SAMPLE ASA_QUEUE ASA_RESOLUTION FITLOC FITLOC ROUND FITLOC_PRINT

MULTI_MIN MULTI_NUMBER ASA_PARALLEL ASA_SAVE SYSTEM_CALL FDLIBM_POW FDLIBM_LOG FDLIBM_EXP ASA_PRINT ASA_OUT USER_ASA_OUT USER_ASA_OUT ASA_PRINT_INTERMED ASA_PRINT_MORE G_FIELD & G_PRECISION

Index of Adaptive Tuning Parameters

OPTIONS->Limit_Acceptances OPTIONS->Limit Generated **OPTIONS->Limit Invalid Generated States** OPTIONS->Accepted_To_Generated_Ratio **OPTIONS->Cost_Precision** OPTIONS->Maximum_Cost_Repeat **OPTIONS->Number** Cost Samples OPTIONS->Temperature_Ratio_Scale **OPTIONS->Cost Parameter Scale Ratio OPTIONS->Temperature** Anneal Scale OPTIONS->User_Cost_Temperature **OPTIONS->Include Integer Parameters OPTIONS->User Initial Parameters OPTIONS->Sequential_Parameters OPTIONS->Initial Parameter Temperature** OPTIONS->User_Temperature_Ratio **OPTIONS->User_Parameter_Temperature OPTIONS-**>Acceptance Frequency Modulus OPTIONS->Generated_Frequency_Modulus **OPTIONS->Reanneal** Cost **OPTIONS->Reanneal Parameters OPTIONS->Delta** X OPTIONS->User_Delta_Parameter **OPTIONS->User** Tangents **OPTIONS->Curvature** 0 OPTIONS->User_Quench_Param_Scale **OPTIONS->User** Quench Cost Scale OPTIONS->N_Accepted **OPTIONS->N** Generated **OPTIONS->Locate** Cost **OPTIONS->Immediate_Exit OPTIONS->Best_Cost OPTIONS->Best** Parameters **OPTIONS->Last** Cost **OPTIONS->Last Parameters** OPTIONS->Asa Data Dim **OPTIONS->Asa** Data OPTIONS->Asa Data Dim Int **OPTIONS->Asa Data Int OPTIONS->Asa Out File OPTIONS->Cost Schedule OPTIONS-**>Asymp_Exp_Param **OPTIONS->Acceptance** Test OPTIONS->User_Acceptance_Flag OPTIONS->Cost_Acceptance_Flag **OPTIONS->Cost** Temp Curr OPTIONS->Cost_Temp_Init **OPTIONS->Cost** Temp Scale **OPTIONS->Prob** Bias

OPTIONS->Random Seed OPTIONS->Generating Distrib OPTIONS->Reanneal Cost Function OPTIONS->Reanneal_Params_Function OPTIONS->Bias_Acceptance OPTIONS->Bias Generated **OPTIONS-**>Average Weights **OPTIONS->Limit_Weights OPTIONS->Queue** Size **OPTIONS->Queue Resolution OPTIONS->Coarse Resolution OPTIONS->Fit Local OPTIONS->Iter** Max **OPTIONS-**>Penalty **OPTIONS->Multi** Cost **OPTIONS->Multi** Params OPTIONS->Multi_Grid **OPTIONS->Multi Specify** OPTIONS->Gener Mov Avr **OPTIONS->Gener Block OPTIONS->Gener Block Max**





Superimposed are runs for n = 8, the case of no reannealing, 3 trajectories each for cases of Q = 1, 2, 3, 4, 5, 6, 7, 8, 16, and 24. Although the actual final cost function values are 0, they were set to 10^{-10} for purposes of these log-log plots.



Superimposed are runs for n = 8, the case including reannealing, 3 trajectories each for cases of Q = 1, 2, 3, 4, 5, 6, 7, 8, 16, and 24.

Example: Fitting Variance

Consider a one variable problem,

$$P[M_{t+\Delta t}|M_t] = (2\pi \hat{g}^2 \Delta t)^{-1/2} \exp(-\Delta t L) ,$$

$$L = (\dot{M} - f)^2 / (2\hat{g}^2) ,$$

with parameter-coefficients α n f and g to be fit to data.

The cost function to be fit to M(t) data is

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

The nonlinear entry of g into the cost function, e.g., competing influence in the denominator of L and in the logarithm term from the prefactor in P, often enables a tight fit to the absolute value of g. In most nonlinear regression methods, this is not possible.

Similar considerations hold for more than one variable. The calculation of the evolution of Langevin systems has been implemented in several systems using ASA. It has been used as an aid to debug the ASA fitting codes, by first generating data from coupled Langevin equations, relaxing the coefficients, and then fitting this data with the effective Lagrangian cost-function algorithm to recapture the original coefficients within the diffusions defined by $g^{GG'}$.

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

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

Boundary Condition Sensitivity

For derivative boundary conditions, for better numerical accuracy, it often is necessary to generalize the histogram expansion to a trapezoidal expansion to give some shape to the histograms.

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.

SOME FINANCIAL APPLICATIONS

Risk-Averting 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

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

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

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

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

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

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

 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_i for agents following market trends is obtained from

$$F_{\text{ex1}}^G = a^G M^- / N ,$$

$$M^- = M^B - M^S .$$

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,

$$j_S = 1, \cdots, N^S$$
,
 $j_B = 1, \cdots, N^B$,
 $N = N^B + N^S$.

Aggregation of Risk-Averting Agents

At any given time, any agent may belong to either pool of S or B. If each agent is considering one unit of a market's assets, then the following development becomes a microscopic model of the dynamics of the market's volume. The "joint" probability distribution P, joint with respect to pools of all S and B agents, but conditional with respect to time evolution, is

$$\begin{split} P[M(t+\tau)|M(t)] &= \prod_{G}^{B,S} P^{G}[M^{G}(t+\tau)|M^{\bar{G}}(t)] \\ &= \sum_{\sigma_{j}=\pm 1} \delta(\sum_{j\in S} \sigma_{j} - N^{S}) \delta(\sum_{j\in B} \sigma_{j} - N^{B}) \prod_{j}^{N} p_{\sigma_{j}} \\ &= \prod_{G} (2\pi)^{-1} \int dQ^{G} \exp[iM^{G}(t+\tau)Q^{G}] \\ &\times \prod_{j\in G}^{N^{G}} \cosh\{F_{j}[M(t)] + iQ^{G}\} \operatorname{sech}\{F_{j}[M(t)]\} \\ &= \prod_{G} (1+E^{G})^{-N^{G}} {N^{G} \choose \lambda^{G}} (E^{G})^{\lambda^{G}} , \\ E^{G} &= \exp(-2F^{G}) , \\ \lambda^{G} &= [[\frac{1}{2} (M^{G}(t+\tau) + N^{G})]] , \\ M &= \{M^{G}\} , \end{split}$$

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. For convenience only, $\sigma_{j}F_{j}$ was defined so that $M^{G} = 0$ is arbitrarily selected as a midpoint between agents acting and not acting on the market: $M^{G} = -N^{G}$ signifies all agents not acting, $M^{G} = N^{G}$ signifies all agents acting.

The mean and variance of this binomial distribution yields

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

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

For large N^G and large $N^G F^G$, this binomial distribution is asymptotically Gaussian. yielding a parabolic PDE.

Interest Rates

2-Factor Interest-Rate Model

The Brennan-Schwartz (BS) 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 . BS simplified and reduced this system to

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

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

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)

$$\begin{aligned} dr/dt &= a_1 + b_1(l-r) + \sigma_1 r(\gamma^+ n_1 + \gamma^- n_2) ,\\ dl/dt &= l(a_2 + b_2 r + c_2 l) + \sigma_2 l(\gamma^- n_1 + \gamma^+ n_2) ,\\ \gamma^{\pm} &= \frac{1}{\sqrt{2}} \left[1 \pm (1 - \rho^2)^{1/2} \right]^{1/2} ,\\ n_i &= (dt)^{1/2} p_i , \end{aligned}$$

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

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

$$F = \begin{pmatrix} dr/dt - (a_1 + b_1(l - r)) \\ dl/dt - l(a_2 + b_2r + c_2l) \end{pmatrix},$$

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

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

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. 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*. 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

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

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 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

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

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

$$B_{s} - B_{s+1} = -r_{s}B_{s} .$$

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. Their discretized calculation of their partial differential equation, and our discretized calculation of the path-integral representation of this model, are different numerical methods of calculating this evolution of B_s .

S&P

S&P Interday Futures-Cash

Too often the management of complex systems is ill-served by not utilizing the best tools available. For example, requirements set by decision-makers often are not formulated in the same language as constructs formulated by powerful mathematical formalisms, and so the products of analyses are not properly or maximally utilized, even if and when they come close to faithfully representing the powerful intuitions they are supposed to model. In turn, even powerful mathematical constructs are ill-served, especially when dealing with multivariate nonlinear complex systems, when these formalisms are butchered into quasi-linear approximations to satisfy constraints of numerical algorithms familiar to particular analysts, but which tend to destroy the power of the intuitive constructs developed by decision-makers. These problems are present in many disciplines, including trading in financial markets. In this context, we can consider the trader as the decision maker on the nature of market data, sometimes also carrying the additional role of his or her own analyst.

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 quasilocal 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.

SOME OUT-OF-FINANCE LESSONS LEARNED

Statistical Mechanics of Combat (SMC)

National Training Center — Janus(T)

The U.S. Army National Training Center (NTC) is located at Fort Irwin, just outside Barstow, California. As of 1989, there have been about 1/4 million soldiers in 80 brigade rotations at NTC, at the level of two battalion task forces (typically about 3500 soldiers and a battalion of 15 attack helicopters), which train against two opposing force (OPFOR) battalions resident at NTC. NTC comprises about 2500 km², but the current battlefield scenarios range over about 5 km linear spread, with a maximum lethality range of about 3 km. NTC is gearing up for full brigade level exercises. The primary purpose of data collection during an NTC mission is to patch together an after action review (AAR) within a few hours after completion of a mission, giving feedback to a commander who typically must lead another mission soon afterward. Data from the field, i.e., multiple integrated laser engagement system (MILES) devices, audio communications, OCs, and stationary and mobile video cameras, is sent via relay stations back to a central command center where this all can be recorded, correlated and abstracted for the AAR. Within a couple of weeks afterwards, a written review is sent to commanders, as part of their NTC take home package.

Janus(T) is an interactive, two-sided, closed, stochastic, ground combat computer simulation. We have expanded Janus(T) to include air and naval combat, in several projects with the author's previous thesis students at the Naval Postgraduate School (NPS).

Stochastic multivariate models were developed for both NTC and Janus, to form a common language to compare the two systems to baseline the simulation to exercise data.

Basic Equations of Aggregated Units

Consider a scenario taken from our NTC study: two red systems, red T-72 tanks (RT) and red armored personnel carriers (RBMP), and three blue systems, blue M1A1 and M60 tanks (BT), blue armored personnel carriers (BAPC), and blue tube-launched optically-tracked wire-guided missiles (BTOW), where RT specifies the number of red tanks at a given time t, etc. Consider the kills suffered by BT, ΔBT , e.g., within a time epoch $\Delta t \approx 5$ min. Here, the x terms represent attrition owing to point fire; the y terms represent attrition owing to area fire. Note that the algebraic forms chosen are consistent with current perceptions of aggregated large scale combat. Now consider sources of noise, e.g., that at least arise from PD, PA, PH, PK, etc. Furthermore, such noise likely has its own functional dependencies, e.g., possibly being proportional to the numbers of units involved in the combat. We write

$$\dot{BT} = \frac{\Delta BT}{\Delta t} = x_{RT}^{BT} RT + y_{RT}^{BT} RT BT + x_{RBMP}^{BT} RBMP + y_{RBMP}^{BT} RBMP BT + z_{BT}^{BT} BT \eta_{BT}^{BT} + z_{RT}^{BT} \eta_{RT}^{BT} + z_{RBMP}^{BT} \eta_{RBMP}^{BT} \dot{RT} = \cdots R\dot{BMP} = \cdots B\dot{APC} = \cdots BTOW = \cdots$$

where the η represent sources of (white) noise (in the Itô prepoint discretization discussed below). The noise terms are taken to be log normal (multiplicative) noise for the diagonal terms and additive noise for the off-diagonal terms. The diagonal *z* term (z_{BT}^{BT}) represents uncertainty associated with the *target BT*, and the off-diagonal *z* terms represent uncertainty associated with the *shooters RT* and *RBMP*. The *x* and *y* are constrained such that each term is bounded by the mean of the KVS, averaged over all time and trajectories of similar scenarios; similarly, each *z* term is constrained to be bounded by the variance of the KVS. Equations similar to the BT equation are also written for RT, RBMP, BAPC, and BTOW. Only *x* and *y* that reflect possible nonzero entries in the KVS are free to be used for the fitting procedure. For example, since Janus(T) does not permit direct-fire fratricide, such terms are set to zero. In most NTC scenarios, fratricide typically is negligible.

Long-Time Correlations Test Short-Time Models

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 better choice of the best algebraic functions, i.e., those which best follow the data through a predetermined epoch of battle. Thus, dynamic physical mechanisms, beyond simple Lanchester "line" and "area" firing terms, can be identified. Afterwards, if there are closely competitive algebraic functions, they can be more precisely assessed by calculating higher algebraic correlation functions from the probability distribution.

Data from 35 to 70 minutes was used for the short-time fit. The path integral used to calculate this fitted distribution from 35 minutes to beyond 70 minutes. This serves to compare long-time correlations in the mathematical model versus the data, and to help judge extrapolation past the data used for the short-time fits. The means are fit very well by this model, even in out-of-sample time periods, something that other Lanchester modelers have not achieved, especially with such empirical data. The variances strongly suggest that the additive-noise model is inferior.



CMI

The results of Janus(T) attrition of Red and Blue units are given in the upper figure. The canonical momenta indicators (CMI) for each system are given in the lower figure.



Using the particular model considered here, the CMI are seen to be complementary to the attrition rates, being somewhat more sensitive to changes in the battle than the raw data. The coefficients fit to the combat data are modifiable to fit the current "reality" of system capabilities.

The CMI are more sensitive measures than 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). This is even more important when replenishment of forces is permitted, often leading to oscillatory variables. Neither the energy or the information give details of the components as do the CMI.

Statistical Mechanics of Neocortical Interactions (SMNI)

Basic SMNI

Multiple Scales

Multiple scales are aggregated, from synaptic dynamics, to neuronal dynamics, to minicolumnar dynamics (100's of neurons). At the level of minicolumns, neocortex seems to be well described by Gaussian-Markovian dynamics.



A derivation is given of the physics of chemical inter-neuronal and electrical intraneuronal interactions. The derivation yields a short-time probability distribution of a given neuron firing due to its just-previous interactions with other neurons. Within $\tau_j \sim 5-10$ msec, the conditional probability that neuron j fires ($\sigma_j = +1$) or does not fire ($\sigma_j = -1$), given its previous interactions with k neurons, is

$$p_{\sigma_j} \approx \Gamma \Psi \approx \frac{\exp(-\sigma_j F_j)}{\exp(F_j) + \exp(-F_j)} ,$$

$$F_j = \frac{V_j - \sum_k a_{jk}^* v_{jk}}{(\pi \sum_{k'} a_{jk'}^* (v_{jk'}^2 + \phi_{jk'}^2))^{1/2}} ,$$

$$a_{jk} = \frac{1}{2} A_{jk} (\sigma_k + 1) + B_{jk} .$$

 Γ represents the "intra-neuronal" probability distribution.

Microscopic Aggregation

A derived mesoscopic Lagrangian \underline{L}_M defines the short-time probability distribution of firings in a minicolumn, composed of ~10² neurons, given its just previous interactions with all other neurons in its macrocolumnar surround. *G* is used to represent excitatory (*E*) and inhibitory (*I*) contributions. \overline{G} designates contributions from both *E* and *I*.

$$\begin{split} P_{M} &= \prod_{G} P_{M}^{G} [M^{G}(r; t + \tau) | M^{G}(r'; t)] \\ &= \sum_{\sigma_{j}} \delta \Biggl(\sum_{jE} \sigma_{j} - M^{E}(r; t + \tau) \Biggr) \delta \Biggl(\sum_{jI} \sigma_{j} - M^{I}(r; t + \tau) \Biggr) \prod_{j}^{N} p_{\sigma_{j}} \\ &\approx \prod_{G} (2\pi\tau g^{GG})^{-1/2} \exp(-N\tau L_{M}^{G}) , \\ P_{M} &\approx (2\pi\tau)^{-1/2} g^{1/2} \exp(-N\tau L_{M}) , \\ L_{M} &= L_{M}^{E} + L_{M}^{I} = (2N)^{-1} (\dot{M}^{G} - g^{G}) g_{GG'} (\dot{M}^{G'} - g^{G'}) + M^{G} J_{G} / (2N\tau) - \underline{V}' , \\ \underline{V}' &= \sum_{G} \underline{V}''_{G'}^{G'} (\rho \nabla M^{G'})^{2} , \\ g^{G} &= -\tau^{-1} (M^{G} + N^{G} \tanh F^{G}) , \\ g^{GG'} &= (g_{GG'})^{-1} = \delta_{G}^{G'} \tau^{-1} N^{G} \mathrm{sech}^{2} F^{G} , \\ g &= \det(g_{GG'}) , \\ F^{G} &= \frac{(V^{G} - a_{G'}^{|G|} v_{G'}^{|G|} N^{G'} - \frac{1}{2} A_{G'}^{|G|} M^{G'})}{(\pi[(v_{G'}^{|G|})^{2} + (\phi_{G'}^{|G|})^{2}] (a_{G'}^{|G|} N^{G'} + \frac{1}{2} A_{G'}^{|G|} M^{G'}))^{1/2}} , \\ a_{G'}^{G} &= \frac{1}{2} A_{G'}^{G} + B_{G'}^{G} . \end{split}$$

Short-Term Memory (STM)

We choose empirical ranges of synaptic parameters corresponding to a predominately excitatory case (EC), predominately inhibitory case (IC), and a balanced case (BC) in between. For each case, also consider a "centering mechanism" (EC', IC', BC'), whereby some synaptic parameter is internally manipulated, e.g., some chemical neuromodulation or imposition of patterns of firing, such that there is a maximal efficiency of matching of afferent and efferent firings:

$$M^G \approx M^{*G} \approx 0 \; .$$

This sets conditions on other possible minima of the *static* Lagrangian \overline{L} .



PDE Stability & Duration Analyses

An estimate of a stationary solution P_{stat} to the Fokker-Planck differential equation for the probability distribution P of M^G firings for an uncoupled mesocolumn, i.e., V' = 0, is given by the stationary limit of the short-time propagator,

$$P_{\text{stat}} \approx N_{\text{stat}} g^{1/2} \exp(-CN\tau \underline{L}bar) ,$$
$$g = \det(g^{GG'})^{-1} \equiv \det(g_{GG'}) = g_{EE}g_{II} ,$$

where N_{stat} and C are constant factors. An estimate of the approximation made is made by seeking values of constants C, such that the stationary Fokker-Planck equation is satisfied exactly. Contour plots of C versus \overline{M}^G demonstrate that there exists real positive C which may only range from $\sim 10^{-1}$ to ~ 1 , for which there exists unbroken contours of C which pass through or at least border the line of minima. At each point \overline{M}^{G} , this leaves a quadratic equation for C to be solved. Dropping the $g^{1/2}$ factor results in C not being real throughout the domain of \overline{M}^G . Thus, this defines a solution with potential $N^2 \overline{L} = \int A dM$, drift A, and diffusion N/τ . Stability of transient solutions, defined for δM^G about a stationary state by $\delta \dot{M}^G \approx -A_G \delta M^G = -N^2 \overline{L}_G \delta M^G$, is therefore equivalent to $\ll \overline{M} \gg$ being minima of \overline{L} . This stationary solution is also useful for calculating the time of first passage, t_{yp} , to fluctuate out of a valley in one minima over a peak to another minima. It turns out that the values of $\tau L \sim 10^{-2}$ for which the minima exist are just right to give t_{vp} on the order of tenths a second for about 9 of the minima when the maximum of 10—11 are present. The other minima give t_{vp} on the order of many seconds, which is large enough to cause hysteresis to dominate single jumps between other minima. Thus, 7 ± 2 is the capacity of STM, for memories or new patterns which can be accessed in any order during tenths of a second, all as observed empirically.

This is a very sensitive calculation. If N were a factor of 10 larger, or if $\tau L < 0.1$ at the minima, then t_{vp} is on the order of hours instead of seconds, becoming unrealistic for STM durations. Oppositely, if t_{vp} were much smaller, i.e., less than $\sim 5\tau$, this would be inconsistent with empirical time scales necessary for formation of any memory trace.

The statistical nature of this storage and processing also explains the primacy vs. recency effect in STM serial processing, wherein first-learned items are recalled most error-free, with last-learned items still more error-free than those in the middle.

SMFM

STM PATHINT Details

The interior of M^G -space of model BC' is examined at 0.01 seconds = τ .



PATHINT STM BC' t=1

Electroencephalography (EEG)

A coarse calculation begins by considering the Lagrangian L_F , the Feynman midpoint-discretized Lagrangian. The Euler-Lagrange variational equation associated with L_F leads to a set of 12 coupled first-order differential equations, in M^G , in nonlinear the with coefficients 12 variables { $M^{G}, \dot{M}^{G}, \ddot{M}^{G}, \nabla M^{G}, \nabla^{2} M^{G} \}$ in (r; t) space. In the neighborhood of extrema $\ll \overline{M}^G \gg$, L_F can be expanded as a Ginzburg-Landau polynomial, i.e., in powers of M^E and M^I . To investigate first-order linear oscillatory states, only powers up to 2 in each variable are kept, and from this the variational principle leads to a relatively simple set of coupled linear differential equations with constant coefficients:

$$\begin{split} 0 &= \delta \underline{L}_F = \underline{L}_{F,\dot{G}:t} - \delta_G \underline{L}_F \\ &\approx - \underline{f}_{|G|} \underline{\ddot{M}}^{|G|} + \underline{f}_G^1 \underline{\dot{M}}^{G^{\neg}} - \underline{g}_{|G|} \nabla^2 \underline{M}^{|G|} + \underline{b}_{|G|} \underline{M}^{|G|} + \underline{b} \underline{M}^{G^{\neg}} , \ G^{\neg} \neq G , \\ (\cdots)_{,\dot{G}:t} &= (\cdots)_{,\dot{G}G'} \underline{\dot{M}}^{G'} + (\cdots)_{,\dot{G}\dot{G}'} \underline{\ddot{M}}^{G'} , \\ \underline{M}^G &= M^G - \ll \overline{M}^G \gg , \ \underline{f}_E^1 = -\underline{f}_I^1 \equiv \underline{f} . \end{split}$$

These equations are then Fourier transformed and the resulting dispersion relation is examined to determine for which values of the synaptic parameters and of the normalized wave-number ξ , the conjugate variable to r, can oscillatory states, $\omega(\xi)$, persist.

For instance, a typical example is specified by extrinsic sources $J_E = -2.63$ and $J_I = 4.94$, $N^E = 125$, $N^I = 25$, $V^G = 10$ mV, $A_E^G = 1.75$, $A_I^G = 1.25$, $B_{G'}^G = 0.25$, and $v_{G'}^G = \phi_{G'}^G = 0.1$ mV. The synaptic parameters are within observed ranges, and the J_G 's are just those values required to solve the Euler-Lagrange equations at the selected values of M^G . The global minimum is at $\overline{M}^E = 25$ and $\overline{M}^I = 5$. This set of conditions yields (dispersive) dispersion relations

$$\omega\tau = \pm \{ -1.86 + 2.38(\xi\rho)^2; -1.25i + 1.51i(\xi\rho)^2 \} ,$$

where $\xi = |\xi|$. The propagation velocity defined by $d\omega/d\xi$ is ~1 cm/sec, taking typical wavenumbers ξ to correspond to macrocolumnar distances ~ 30ρ . Calculated frequencies ω are on the order of EEG frequencies ~ 10^2 sec^{-1} , equivalent to $v = \omega/(2\pi) = 16 \text{ cps}$ (Hz). These mesoscopic propagation velocities permit processing over several minicolumns ~ 10^{-1} cm, simultaneous with the processing of mesoscopic interactions over tens of centimeters via association fibers with propagation velocities ~600–900 cm/sec. I.e., both intraregional and interregional information processing can occur within ~ 10^{-1} sec.

ASA Fits of CMI

These momenta indicators should be considered as supplemental to other clinical indicators. This is how they are being used in financial trading systems. The CMI are more sensitive measures of neocortical 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. Neither the energy or the information give details of the components as do the CMI. EEG is measuring a quite oscillatory system and the relative signs of such activity are quite important. Each set of results is presented with 6 figures, labeled as [{alcoholic|control}, {stimulus 1|match|no-match}, subject, {potential|momenta}], where match or no-match was performed for stimulus 2 after 3.2 sec of a presentation of stimulus 1. For each subjects run, after fitting 28 parameters with ASA, epoch by epoch averages are developed of the raw data and of the multivariate SMNI canonical momenta. There are fits and CMI calculations using data sets from 10 control and 10 alcoholic subjects for each of 3 paradigms.



Chaos in SMNI

Duffing Analog

Some aspects of EEG can be approximately cast as a model of chaos, the Duffing oscillator.

$$\ddot{x} = f(x, t) ,$$

$$f = -\alpha \dot{x} - \omega_0^2 x + B \cos t .$$

This can be recast as

$$\begin{split} \dot{x} &= y , \\ \dot{y} &= f(x,t) , \\ f &= -\alpha y - \omega_0^2 x + B \cos t \end{split}$$

Note that this is equivalent to a 3-dimensional autonomous set of equations, e.g., replacing $\cos t$ by $\cos z$, defining $\dot{z} = \beta$, and taking β to be some constant.

We studied a model embedding this deterministic Duffing system in moderate noise, e.g., as exists in such models as SMNI. Independent Gaussian-Markovian ("white") noise is added to both \dot{x} and \dot{y} , η_i^j , where the variables are represented by $i = \{x, y\}$ and the noise terms are represented by $j = \{1, 2\}$,

$$\begin{split} \dot{x} &= y + \hat{g}_{x}^{1} \eta_{1} , \\ \dot{y} &= f(x,t) + \hat{g}_{y}^{2} \eta_{2} , \\ &< \eta^{j}(t) >_{\eta} = 0 , \\ &< \eta^{j}(t), \eta^{j'}(t') >_{\eta} = \delta^{jj'} \delta(t-t') . \end{split}$$

In this study, we take moderate noise and simply set $\hat{g}_i^j = 1.0\delta_i^j$. The equivalent short-time conditional probability distribution *P*, in terms of its Lagrangian *L*, corresponding to these Langevin rate-equations is

$$P[x, y; t + \Delta t | x, y, t] = \frac{1}{(2\pi\Delta t)(\hat{g}^{11}\hat{g}^{22})^2} \exp(-L\Delta t) ,$$
$$L = \frac{(\dot{x} - y)^2}{2(\hat{g}^{11})^2} + \frac{(\dot{y} - f)^2}{2(\hat{g}^{22})^2} .$$

Noise Washes Out Chaos/PATHINT

No differences were seen in the stochastic system, comparing regions of Duffing parameters that give rise to chaotic and non-chaotic solutions. More calculations must be performed for longer durations to draw more definitive conclusions.



Path Integral Evolution of Chaotic Stochastic Duffing Oscillator



SOME CURRENT PROJECTS

Volatility Modeling

Eurodollar Spreads

It should be clear that 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.,

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

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

$$BPV \approx < y > \sigma .$$

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, 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.

Trading on the ratios of Front/Back contracts often presents less risk than trading on the separate contracts. Let w(x, y) be a function of 2 random variables (say, w = x/y, like Front and Back contracts). For small perturbations,

$$w = \langle w \rangle + (x - \langle x \rangle) \frac{\partial w}{\partial x} + (y - \langle y \rangle) \frac{\partial w}{\partial y} + \cdots,$$

$$\langle w \rangle = w(\langle x \rangle, \langle y \rangle) + \cdots,$$

$$\operatorname{Var}(w^{2}) = \langle (w - \langle w \rangle)^{2} \rangle$$

$$= \left(\frac{\partial w}{\partial x}\right)^{2} \operatorname{Var}(x^{2}) + \left(\frac{\partial w}{\partial y}\right)^{2} \operatorname{Var}(y^{2}) + 2 \frac{\partial w}{\partial x} \frac{\partial w}{\partial y} \operatorname{Var}(xy),$$

$$\operatorname{Var}((x/y)^{2}) = 1/y^{2} \operatorname{Var}(x^{2}) + x^{2}/y^{4} \operatorname{Var}(y^{2}) - 2x/y^{3} \operatorname{Var}(xy).$$

Thus, the standard deviation of the ratio is reduced by the correlation Var(xy) between the two contracts.

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) with 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 $S > S_{\infty}$.

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:

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

 $d\sigma = v + \varepsilon \, dw_{\sigma}$

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.

PATHINT

PATHINT is being used to perform European and American, one-factor and twofactor, 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 = -1. The short-time probability distribution at time T = 0.5 years for x = 2.



PATHINT Two-Factor

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

Two-Factor Probability -----



Long-Time Probability

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.



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.


BS vs CRR vs PATHINT

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

Greek	BS	CRR _{European}	CRR _{American}	CRR _{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.

Tick Resolution CMI Trading

We have tested a prototype of CMI trading on S&P using minute resolution data. This looks promising enough to consider for current research projects.