"A Foundation for Quantitative Finance: From randomness engineering to financial engineering"

R. B. Holmes, PhD
HIM&R, Inc.
rbholmes@attbi.com

Abstract

At the center of most efforts in quantitative finance is an asset price model. We address the questions:

Where do models come from?
What do you want from a model?
How do you validate a model?
What can you do with a model, once accepted?

This subject has had, of course, a long and controversial history, dating back over 40 years. Here we want to emphasize recent data-based approaches, part of what has become a new interdisciplinary field, now known as "econophysics". Indeed, it's a bold and fascinating notion, given the diversity of markets, time periods, sampling frequency, etc., that there may be some intrinsic (even universal!) properties of financial data that can in turn be captured by suitable statistical models,.

Once fixed, a model can be used to create an artificial "simworld" wherein financial strategies can be tested, derivative models priced, etc. Here there are both top-down and bottom-up approaches, the latter an attempt to "explain" the model, in much the same way as atomic structure can explain macro-observables in physics.

In either case, given that the model posits that prices are somehow made out of noise, then noise can be used to create (artificial) prices via simulation, if we can create noise. This last is a more subtle problem than perhaps one intuitively would expect…

Financial
engineering

Ideal world
of
MODELS

Simulation
and
other techniques

Randomness
engineering

The Classic Conundrum of Science

"There are two ways of forming an opinion. One is the scientific method; the other, the scholastic. To the scientific mind, experimental proof is all-important, and theory is merely a convenience in description, to be junked when it no longer fits. To the academic mind, authority is everything, and facts are junked when they do not fit theory."
-Robert A. Heinlein

Which comes first, chicken or egg, theory or data?

There are numerous instances in finance where we have or can generate plenty of data, along with relevant theory; e. g.

Fundamental asset prices --- valuation models

Derivative asset prices --- OPMs

Decision-making experiments --- vN/M E(utility) max

Positive vs. normative tension: what do we want from a theory…
a guide to action, a good explanation, a prediction,
or merely a description (a curve-fit of the data)?

Basic Questions re Models

Where do models come from?

- Experience/intuition
- Theory
- Data
- Experiments
- Simulations (Artificial markets; MMS)

In general, different models will select different problems as being of significance, hence will exhibit partial success (at best!) and unavoidable limitations...

"All models are wrong, some may be useful."
--George Box; M. Friedman

What do you want from a model?

- Which aspect(s) of reality do you wish to capture? (Assumes you have a handle on "reality"!)
- Capture genuine properties of the data
- Robust, not too complex
- Don't depend too critically on quantities difficult to measure
- Useful economic significance, esp practical significance of risk exposure...

Fundamental Price Models from Theory

DDMs (traditional or stochastic)

Equilibrium models of exp returns: CAPM, APT

Multivariate additive/multiplicative factor models

An abstract and universal formalism: $p = E(mx)$
(Here x = future payoff, m = pricing kernel/stochastic discount factor, ...)

The Follmer-Schweitzer traders

Arbitrage and entropy/information considerations

Risk-neutral distributions from option prices:

Breedon-Litzenberger-Dupire-Derman-Kani interpolations
Rubinstein-Jackwerth lattice models
Buchen-Kelly cross-entropy min
Parametric (eg, 4-par pdf families, ANN) and
non-parametric (eg, kernel) estimators

Questionable economic constructs

Arrow-Debreu market/prices
Beta
Equilibrium
Market portfolio
Market price of risk
Pricing kernel (aka SDF, EMM)
Utility function

Questionable hypotheses

EMH
Frictionless markets
Homogeneous expectations/agents
Representative investor, fully rat'l (agent aggregation)

Does any (or even all!) of this tell what you really want to know?

Inconvenient facts

(tragically killing beautiful theories!)

Prices fluctuate more than expected from external fundamentals

High levels of trading volume

Willingness to pay for information

Behavioral & cognitive biases

Difficulty in parameter estimation and/or tests

"Asset-pricing theory is both elegant and logically compelling. It is a nice piece of applied math. But this is not sufficient to conclude that it has scientific merit. To establish the latter, its predictions must be verified in a variety of contexts…

"It is fair to conclude that the predictions of asset-pricing theory have largely been rejected in empirical studies of historical data."

--Peter Bossaerts, 2002

General Questions

Given a present asset price, what is it likely to do over some future time horizon?

Why should there be any such descriptors and how could we choose among them?

How universal can such models be?
Do we need to specify a particular asset, an asset class, a time period, a sampling frequency, ...?
Does the model apply to an individual (eg, based on preferences) or to all traders?

Parametric or not; classify par's as deterministic or stochastic

The Challenge

Nature of interactions between elements and agents comprising the system is unknown, as is way external factors affect it. Hence, as starting point resort to empirical studies to uncover regularities, since interactions generate many observables.

10s of millions of data pts now analyzed, for 5 min and daily time frames

Model Structures

Discrete or continuous time

Static or dynamic

Single, multi- or infinite dimensional random variables
Ex's = a stock or an exchange rate; a portfolio;
a term structure or a volatility surface.

One time horizon, several intermediate points or whole interval
➔ resp. a univar dist, a multivar dist or a full stochastic process.

If latter, nature of sample paths. Recall that full description of a sp requires a compatible system of finite dist's of all orders, an impossible task! So, some model structure is necessary.

What is appropriate "state of randomness" (mild, slow, wild,...)?
(Cf states of matter as gas, liquid, solid, ...)

What can the data tell us?

Basic issues of price function and time ref frame

Type of pdf at a future time, incl finiteness of moments, esp 2nd
Important for risk assessment

Sample path behavior
In gen, what can discrete data say about cont time models?
Smoothness of paths (if continuous);
Presence or not of diffusion component;
Finite or infinite "activity"; BV or not.
What do jumps say about market completeness and use of stop-loss orders?

Difference between objective and risk-neutral measures

Scaling and universality
Eg, same law applies to big and small changes, or
to different assets and/or markets;
"Data collapse"

In general, data will help discipline a theory without pinning down a specific model...

When data were scarce, logical consistency and simplicity were only guides to models (not only in finance!); now, not looking at data is just an excuse for sparing elegant but ill-fitting models.

Stylized Market Data Facts

Foundation of a data-based approach to financial modeling

The very existence of such facts, across events, markets and time frames, is both a bold hypothesis and a happy surprise. It's a consequence of several decades of empirical studies, increasing availability of massive databases, data-mining and other statistical techniques, plus a novel PoV (statistical physics).

Special cases:

Single period, single stock

Single period, multi stock

Multi-period, single stock

Multi-period, multi stock

Figure 1

A Canonical Example

A typical price graph

The corresponding returns

The empirical pdf

The fitted Gaussian pdf

Figure 2

Mandelbrot's Medley

Can you spot the fake(s)?

"The ability to imitate is a form of understanding"

Single period, ? t, the classic case:

1- Low autocorrelation

2- Heavy (“fat”) tails

3- Return asymmetry

4- Aggregational Gaussianity (? t → 8)

5- Volatility clusters

6- Volatility long memory

7- Leverage effect

8- Vol/vol correlation

<u>Example</u>: Fact 1 + finite 2nd moment ➔ CLT is asy in play
➔ Gaussian returns for long time frames ➔ support for Fact 4

The point being … some of these attributes are familiar, having been known or suspected for years, but now it is possible to exquisitely quantify them, thereby providing rather precise targets for modeling.

Room is left for dozens of parametric models, typically utilizing 4 parameters for adequate fit…

Insights from Econophysics

Physicists midway between empirical traders and formal (remote) economists…

"The mystery that finance and physics have in common is that models and mathematics work at all."
--Emanuel Derman, 1996

"Physicists attempt to scale the ivory towers of Finance"
--J. Doyne Farmer, 2000

"Economics has been a black hole for theoretical physicists … it has laid claim to the techniques of natural science when it is really a moral science…"
--Anthony Garrett, 2001

"Modern physics stands on the shoulders of giants;
modern economics stands on their toes"
--Steve Keen, 2002

"The distribution is the answer!"

Explain fluctuations beyond fundamentals…

Assorted Microscopic Market Models

MMS: design artificial market composed of various types of agent-traders, assume plausible rule-based behavior thereof, eg, market clearing and strategies, possibly with learning and/or optimization, then run many iterations and study resulting observables, such as price/return and/or wealth time series behavior and compare with empirical behavior.
Also, bubbles/crashes...

Agent types: producer/speculator/noise trader;
deterministic / adaptive / random
fundamentalist/chartist;
market maker / others

Attitudes: optimistic or pessimistic

Attributes: bounded rationality, diverse skills,
time-varying expectations;
an "ecology" of trader "species" that compete,
adapt and learn, but maybe not optimally,
while scavenging for information...

Abandon models that assume no feedback and/or investor homogeneity

Endogenous models can generate price/return activity obeying stylized empirical facts

Analogy with Boyle's Law (PV = kT): empirically derived, later explained by an underlying atomic theory, but former more useful

Insights, cont.

Empirical regularities uncovered

Source of universal laws: common rules of speculative markets, mixing human psychology (fear/greed) with basic mechanisms of price formation

Studies of correlation/dependency structures via

Random matrix theory
Market topology (clusters, hierarchies, trees, ultrametrics)

Game Theory

Minority games(*)

Evolutionary games

Models and studies of crashes and drawdowns

Portfolio optimization

Multitude of risk measures via cumulants
Eg, VaR vs variance

What does any such strategy say about your model?
Eg, what about mean/variance selection?

(*): "The ballparks have gotten too crowded. That's why nobody goes to see the games anymore."
--Yogi Berra

Ex.'s of Single Period (Marginal) Log-price Models

- Lognormal
- Stable
- Stochastic time changes
- Double exp
- GB2, G?, GT, Johnson, Pearson, Weibull, …
  (either direct fits or to parameterize implied models)
- Discrete normal mixtures
- Continuous normal mixtures
  - o Student–t
  - o Variance-gamma
  - o Hyperbolic
  - o NIG
- Stretched exp
- CGMY model
- Tsallis

So, this skips time series/stochastic process models (trees, ARCH-types, SDEs, multifractal, etc.) and multivariate models.

Example: ARCH assumes finite memory, hence inconsistent with long-range volatility correlations

Several approaches: modify gBm via local or stochastic volatility, or new complex (non-Wiener) process, or by a behavioral feedback drift term.

Data shuffling ➔ rapid convergence to G.

Issues of parsimony, parameter interpretation and estimation, and model selection

Purpose: forecast, option pricing, risk control, …

Multi-stock (cross-asset) Modeling

This concerns first the empirical correlation matrix, a notoriously "noisy" object, and then other (nonlinear) dependency measures.

The corr matrix C of N stocks will be hard to measure accurately unless the number of time periods $T \gg N$. For moderate sizes N this is problematic esp lacking stationarity (i.e., time varying parameters)

Correlations observed to increase during times of market stress.

Some analysis and remedies:

- Use of RMT to assess info content of C ➔
  use of shrinkage and cleansing operations on C;
- Use of skewed alt (incl semi-parametric) dist's to model, permitting greater variability yet with constant par's ➔ measure volatility as 2nd moment of fitted dist, rather than as sample moment;
- Use of copulas to model dependencies indep of marginals;
- Use of coeff of tail dependence

MSTs, a non-random and robust market topology, a measure of market self-organization. As with previous stylized facts, another kind of typical non-randomness in financial data…

"Statistical synchrony" of multiple price dynamics ➔ much more complicated than random walk theory…

Model QC
(for model validation)

In general, the entire return distribution has significance

Validate a model either via backtest or realtime monitoring

Method: compare actual returns with corresponding predicted percentile; these latter samples should be ~ U[0,1]

In turn, many GOF tests exist for this problem
(eg, chi-sq, A-D, K-S, Cramer-vonMises, Neyman,...)

Improves on testing just for VaR or other single percentile by testing whole dist...

Successes of Econophysics

What has been learned re market behavior that we didn't already know? What can we do with it?

What are the strengths & weaknesses of these new insights wrt more traditional financial economics?

What new tech tools are useful?

Examples:

Crash forecasting
http://www.ess.ucla.edu/faculty/sornette/prediction/index.asp#prediction

Most extensive look at historical databases:
empirical regularities and new models (incl MMS)

Scaling laws revealed, tails characterized, and cross-correlation SNR quantified via RMT

Portfolio selection and risk assessment in a non-G world

Option pricing in a non-B/S world

Figure 3

Illustrating "data collapse" with SPX data at intervals of

? t = 1,3,10,32,100,316 and 1000 minutes...

Figure 4

Illustrating high-frequency SPX data with fitted

"mild" and "wild" models...

{backups only}

Simulation

If prices are made out of noise, then noise can be used to create (artificial) prices via simulation, <u>if</u> we can create noise.

This last is a more subtle problem than perhaps one intuitively would expect…

What is a "random number"?

Role of info and algorithmic complexity theory to try and compress market data, equiv to ask if it is redundant;
Is the stock market a RNG?

Tests of EMH using GA/GP

But eventually we can construct both a financial laboratory and a "flight simulator", permitting experiments and tests prior to actually putting real money at risk…

First paper [Boyle, JFE, 1977]…

Role of Simulation

Once you have a model then you may want to

Create and observe the corresponding simworld to study a particular problem or strategy, or

Try to explain this model (hence really the underlying data) via an artificial market
(an MMS, analogous to atomic structure);

Hence, top-down and bottom-up artificial markets...

Also, main mathematical applications of simulation are to integration, optimization, SDEs. Plus, of course, to stat models... Since, in practice, expectations of random quantities reduce to definite integrals, often of high dimension, it is clear that simulation is at least potentially likely to be relevant to the solution of FE problems.

The Quotes

“Simulation saves lives.” [Newsweek, 6/24/02]

“As the complexity of the structure of the financial claims or of the dynamics of the underlying assets increases, MC simulation often becomes the sole computationally feasible means of security pricing.”
[Fu and Su, 2002]

“The MC method is really the only viable numerical technique for high-dim problems and such problems are becoming more prevalent.”
[Boyle, Kolkiewicz & Tan, 2002]

“Monte Carlo is the standard tool for options pricing desks worldwide.”
[RISK, 6/00]

“Simulations are at the very heart of finance.”
[Paul Wilmott, 1998]

“For the case of fixed-income derivatives that include interest rate exotics and MBSs, the SDEs for int rates or bond prices are much more complicated than the B/S models. For these applications, MC simulations are considered indispensable in practice.” [Shu Tezuka, 1998]

“MC simulations have become the preferred method to project future market evolution and simulate potential changes in the value of portfolios to estimate their market risks or credit exposures.” [Till Guldimann, 1998]

“If a man will begin with certainty he will end with doubts; but if he will be content with doubts, he shall end in certainties.” [Francis Bacon, 1605]

Advantages of Simulation

- Intuitive, easy to apply, transparent, flexible
- Confidence intervals
- Unlimited experiments
- No real-world risk
- Reduced development and effort (“speed vs. brains”)
- Success in high-dim problems
- Handles less tractable more realistic models
- Only plausible method for complex options and/or SDEs;
  for former, other possibilities are finite diff’s and lattices
- Benchmark and validate analytic procedures when available

But despite these advantages the Monte Carlo method is frequently criticized for three alleged failings: first, that it underestimates the probability of large market changes; second, that it does not identify portfolio sensitivities or risk sources as easily as do simpler, parametric methods; and third, that it is prohibitively slow. These criticisms may be overstated. The Monte Carlo method can be extended to use non-normal distributions that capture large market changes, and can provide sensitivities and accurate results in a timely manner.

Basic simulation techniques (for variance reduction since efficiency speedup is key to successful use):
control variates, importance sampling, stratification.
Otherwise, speedup can only come from hardware...

Randomness engineering

In constant combat with statisticians/physicists/cryptologists who seek flaws!

Quote: “It’s a very exciting idea that randomness itself has value” [Manuel Blum, UCB, 1987]

At the least, we need some appreciation for this background field, like learning to operate some equipment without knowing all about the engine...

The questions: how is randomness defined, implemented, validated and then productively used?

What do we want to do;
How do we do it;
How do we know if we (or someone else) have done it?

Format: $x(n+1) = F(x(n))$, $F = ?$

So, we have a discrete deterministic dynamical system, where F = a function, a set of instructions, ...

Procedure or Product?

What can an end-user do? How much to take on faith?

Design own RNG
Obtain someone else's
  Agency problem of RNGs;
  necessarily general purpose algorithms
Buy a disk of bits or uniforms

Aspects of a good RNG:

- *A priori* theory (number theory, chaos,...)
- Code (esp rounding & truncation for finite number system of computer)
- Repeatability (others should be able to replicate your results)
- Portability (same results on/in different computers/languages)
- Speed
- Validation via output tests

Conundrum:

No RNG is perfect (insistence on this leads to QMC)
No finance model is either, so how accurate must a RNG be?

Procedure validation:

Theoretical figures of merit are desirable, so as to predict performance w/o need for extensive tests. Eg, period length, correlations (consecutive and/or long range) and high-dim equidistribution. Discrepancy measures for this.

Linear and nonlinear RNGs; output of former display lattice structure, latter are slower.

RNGs should be chosen on theoretical grounds, then tested empirically. These latter tests are fishing expeditions.
A “good” test will depend on the intended app. Every RNG fails some test(s).

Product validation:

Test for both aspects of iid

Test on real problems with known sol’ns, such as high-dim integrals

- Other math
- Geometry
- Physics (Ising models, RWs)
- Finance

Test on problems similar to your intended application!

Procedure examples:

Mersenne twister software; a GFSR [gen feedback shift register] generator said to have period = 2**19937 – 1 (!) (= M prime, 6K digits, verified 1971)

CA Rule #30 = basis for Mathematica RNG

Expansions of algebraic irrationals or even p (!)

Product examples:

G. Marsaglia's Random Number CDROM = 60 10MB files, totaling 4.8 billion random bits, based on the "Diehard" code

Park-Miller "min standard" used in MatLab

All RNGS in standard software pacs [C, VB, Excel, Pascal, Java,...] fail basic tests, eg, Excel fails 2-d collision test

Also, problems with short periods, esp in Excel...

Traditional tasks:

Simulating end-of-period prices

Simulating paths of Bm, as ingredients of gBm or other diffusion-type SDE for fundamental securities

Examples:

K-L or Schauder expansions;

Limits of discrete RW using bits

Use of Brownian bridge trick

Less traditional tasks:

Simulating non-diffusion models, with non-Brownian drivers

Poisson or Levy processes

Generating uniforms in high-dim regions for optimization

Major QF Problems

Portfolio optimization – needs constrained high-dim search

Risk management – needs sim of rare events

Option, CMO, structured product pricing
– needs sim of high-dim integrals

Others

Artificial markets (market ob's explained as emergent prop's)
DDMs
Portfolio rebalancing (fuzzy frontier)
Performance evaluation (POD method)
Retirement planning
    Financialengines.com
    Financeware.com
    Moneytree.com
    (Once again, the dist. is the answer!)
Trading strategies (eg, protective put rollovers)
TAA manager performance

Current frontiers of option pricing

Pricing exotics based on gBm

Pricing anything based on non-gBm

Techniques for Greeks
- Common random numbers
- Infinitesimal perturbation analysis
- Malliavin calculus

Incorporating stochastic parameters, int rates and/or vol

American-style options (cash flows depend on owner)

Current frontiers of portfolio selection

Active benchmark tracking;
non-traditional constraints: size, margin, diversified, buy-in threshold, round lots only ➔ high comp complexity, integer variables and discontinuous EF

<u>Example</u>: merging theoretical asset pricing model with simulation

Where will a stock be in T years in a B/S world?
The estimate is

$$S(T) = S(0)\exp((m - 0.5s^2)T + sN(0,T)),$$

a consequence of Ito's lemma applied to the standard SDE

$$dS(t)/S(t) = m\,dt + s\,dW(t).$$

So we need estimates of m and s. Now s can be reliably inferred from a history of log returns. What about m? As an alternative to a purely historical drift estimate (which requires at a minimum a long time series), we can try the CAPM. Recall

$$m - r_f = b(m_M - r_f),$$

where m_M = market exp return, r_f = riskfree rate and b = beta coeff. These can all be obtained is standard ways, yielding in turn an est for m. Then we have a lognormal dist for S(T) as usual. Estimating m thusly may give a more "gentle" estimate than one from a pure historical trend. Any problems? As in applying Markowitz m/v optimization for portfolio construction, it assumes parameters are known, when they're merely estimates. The estimated b, as a regression coeff, is an unbiased normally dist est of true beta. Hence future values of S(T) can be simulated by a draw from b + another from N(0,T) and from these a complete dist of S(T) can be built up...

Role of Complexity Theory

System with large no. of mutually interacting parts, often open to its environment. Results in self-organization may lead to novel (often unexpected) emergent behavior.

Impossible to predict in detail but, maybe predict major phase transitions, which often control long term behavior of system

- Earthquakes
- Volcanic eruptions
- Failure of engineering structures
- Epileptic seizures, heart attacks & SIDS
- Financial crashes

Usual problem, these extreme events are far from center of dist, hence not amenable to usual statistics and reductionist theory. View as critical singularities as studied in stat physics, eg, slow buildup of long-term correlations between interacting parts.

Over-synchronized neurons in epileptic fits, overly regular heartbeat in SIDS, as measured by *approximate entropy* ; may prove useful for market alerts.

Formalize/quantify the classic Keynesian notions of “animal spirits” and the “beauty contest” metaphor...

Summary

The construction of models and random numbers involves intellectual and philosophical issues, as well as technicalities...

Econophysics:

- Detailed study of massive market databases
- New PoV's re complex adaptive nonlinear systems
- Stylized facts & universal properties revealed
- New models to incorporate all data and some (not all) "facts", according to application
- MMS models to explain "facts" as emergent behavior, analogous to atomic/molecular theories of physics

Simulation:

- Randomness engineering (value of pure randomness)
- Simulation tricks ("swindles") for efficiency
- Simworlds created for particular financial app's

The two conundrums:

- Theory or data as sources for models
- Simulation from an inexact model via imperfect RNG...