Trading and Investment As A Science

Technical Analysis: A Scientific Perspective

Haksun Li

haksun.li@numericalmethod.com

www.numericalmethod.com

Speaker Profile

- Dr. Haksun Li
- CEO, Numerical Method Inc.
- (Ex-)Adjunct Professors, Industry Fellow, Advisor, Consultant with the National University of Singapore, Nanyang Technological University, Fudan University, the Hong Kong University of Science and Technology.
- Quantitative Trader/Analyst, BNPP, UBS
- PhD, Computer Sci, University of Michigan Ann Arbor
- M.S., Financial Mathematics, University of Chicago
- B.S., Mathematics, University of Chicago

Numerical Method Incorporation Limited

- A consulting firm in mathematical modeling, esp. algorithmic trading and quantitative investment.
- Products:
  - SuanShu: a modern math or numerical library
  - AlgoQuant: a financial technology for quantitative trading
- Customers:
  - brokerage houses and funds
  - multinational corporations
  - very high net worth individuals
  - gambling groups
  - academic institutions

References

- Emmanual Acar, Stephen Satchell. Chapters 4, 5 & 6, Advanced Trading Rules, Second Edition. Butterworth-Heinemann; 2nd edition. June 19, 2002.
- Algorithmic Trading: Hidden Markov Models on Foreign Exchange Data. Patrik Idvall, Conny Jonsson. University essay from Linköpings universitet/Matematiska institutionen; Linköpings universitet/Matematiska institutionen. 2008.
- A tutorial on hidden Markov models and selected applications in speech recognition. Rabiner, L.R. Proceedings of the IEEE, vol 77 Issue 2, Feb 1989.

What is Quantitative Trading?

Quantitative Trading?

- Quantitative trading is the buying and selling of assets following the instructions computed from a set of **mathematical** models.
- The differentiation from other trading approaches or the emphasis is on **how** a strategy is generated and not on what strategy is created.
- It applies (rigorous) mathematics in all steps during trading strategy construction from the start to the end.

NM Quantitative Trading Research Process

1. Translate a vague trading intuition (hypothesis) into a concrete mathematical model.
2. Translate the mathematical symbols and equations into a computer program.
3. Strategy evaluation.
4. Live execution for making money.

Step 1 - Modeling

- Where does a trading idea come from?
  - Ex-colleagues
  - Hearsays
  - Newspapers, books
  - NOW TV, e.g., Moving Average Crossover (MA)
- A quantitative trading strategy is a math function, f, that at any given time, t, takes as inputs any information that the strategy cares and that is available, F_t, and gives as output the position to take, $f(t,F_t)$.

Step 2 - Coding

- The computer code enables analysis of the strategy.
  - Most study of a strategy cannot be done analytically.
  - We must resort to simulation.
- The same piece of code used for research and investigation should go straight into the production for live trading.
  - Eliminate the possibility of research-to-IT translation errors.

Step 3 - Analysis

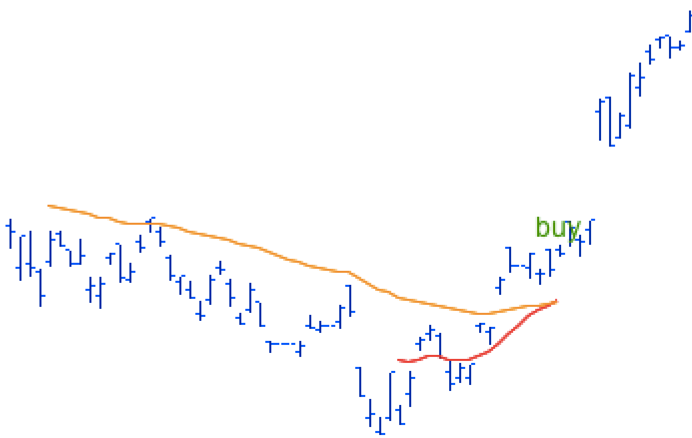
- Compute the properties of a trading strategy.
  - the P&L distribution
  - the holding time distribution
  - the stop-loss
  - the maximal drawdown
  - http://www.numericalmethod.com/trac/numericalmethod/browser/algoquant/core/src/main/java/com/numericalmethod/algoquant/execution/performance

Step 4 - Trading

- Put in capitals incrementally.
- Install safety measures.
- Monitor the performance.
- Regime change detection.

Moving Average Crossover as a TA

- A popular TA signal: Moving Average Crossover.
  - A crossover occurs when a faster moving average (i.e. a shorter period moving average) crosses above/below a slower moving average (i.e. a longer period moving average); then you buy/sell.
- In most TA book, it is then illustrated with an example of applying the strategy to a stock for a period of time to show the profits.



Technical Analysis is Not Quantitative Trading

- TA books merely describes the mechanics of strategies but seldom/never prove them.
- Appealing to common sense is not a mathematical proof.
- Conditional probabilities of outcomes are never computed. (Lo, Mamaysky, & Wang, 2000)
- Application of TA is more of an art (is it?) than a science.
  - How do you choose the parameters?

The Quantitative Trading Research Process

Moving Average Crossover as a Quantitative Trading Strategy

- There are many mathematical justifications to Moving Average Crossover.
  - weighted Sum of lags of a time series
  - Kuo, 2002
- Whether a strategy is quantitative or not depends not on the strategy itself but
  - entirely on the process to construct it;
  - or, whether there is a scientific justification to prove it.

Step 1

- Two moving averages: slower (n) and faster (m).
- Monitor the crossovers.
- $B_t = \left(\frac{1}{m}\sum_{j=0}^{m-1} P_{t-j}\right) - \left(\frac{1}{n}\sum_{j=0}^{n-1} P_{t-j}\right), n > m$
- Long when $B_t \geq 0$.
- Short when $B_t < 0$.

How to Choose n and m?

- It is an art, not a science (so far).
- They should be related to the length of market cycles.
- Different assets have different n and m.
- Popular choices:
  - (250, 5)
  - (250 , 20)
  - (20 , 5)
  - (20 , 1)
  - (250 , 1)

Two Simplifications

- Reduce the two dimensional problem to a one dimensional problem.
- Replace arithmetic averages with geometric averages.
  - This is so that we can work with log returns rather than prices.

AMA(n , 1)

- $B_t \geq 0$ iff $P_t \geq \left(\frac{1}{n}\sum_{j=0}^{n-1} P_{t-j}\right)$
- $B_t < 0$ iff $P_t < \left(\frac{1}{n}\sum_{j=0}^{n-1} P_{t-j}\right)$

GMA(n , 1)

- $B_t \geq 0$ iff $P_t \geq \left(\prod_{j=0}^{n-1} P_{t-j}\right)^{\frac{1}{n}}$
  - $R_t \geq -\sum_{j=1}^{n-2} \frac{n-(j+1)}{n-1} R_{t-j}$ (by taking log)
- $B_t < 0$ iff $P_t < \left(\prod_{j=0}^{n-1} P_{t-j}\right)^{\frac{1}{n}}$
  - $R_t < -\sum_{j=1}^{n-2} \frac{n-(j+1)}{n-1} R_{t-j}$ (by taking log)

What is n?

- $n = 2$
- $n = \infty$

Acar Framework

- Acar (1993): to investigate the probability distribution of realized returns from a trading rule, we need
  - the explicit specification of the trading rule
  - the underlying stochastic process for asset returns
  - the particular return concept involved

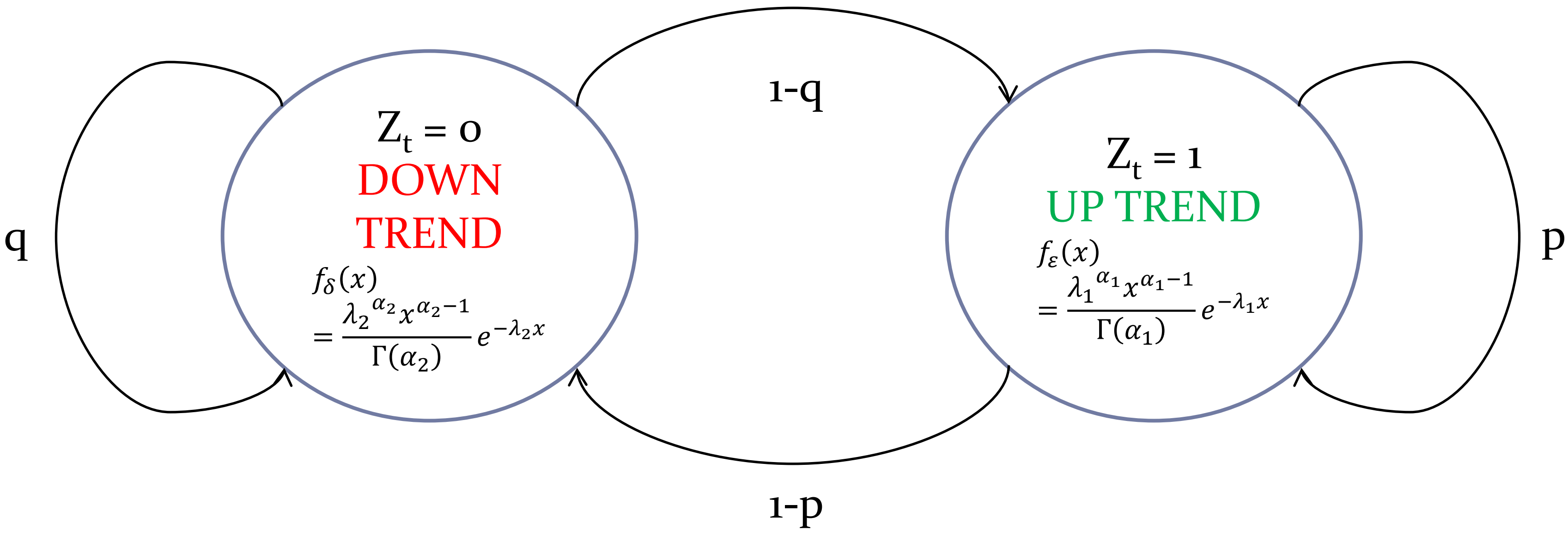
Empirical Properties of Financial Time Series

- Asymmetry
- Fat tails

Knight-Satchell-Tran Intuition

- Stock returns staying going up (down) depends on
  - the realizations of positive (negative) shocks
  - the persistence of these shocks
- Shocks are modeled by gamma processes.
- Persistence is modeled by a Markov switching process.

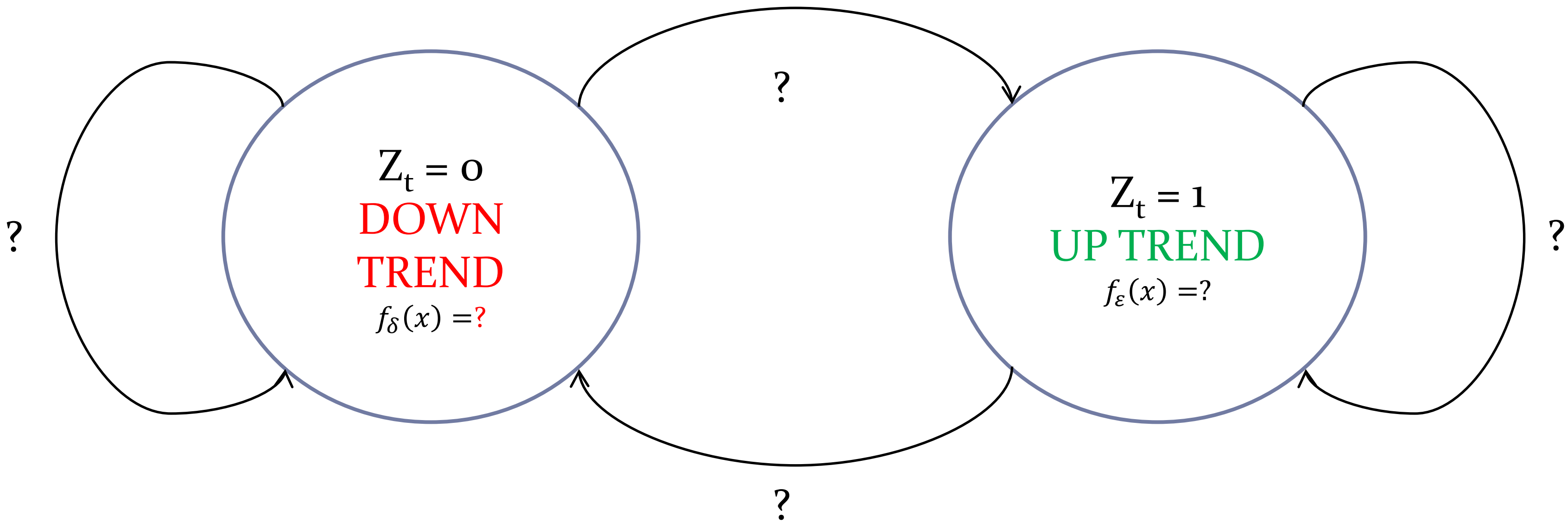
Knight-Satchell-Tran Z_t



Knight-Satchell-Tran Process

- $R_t = \mu_l + Z_t \varepsilon_t - (1 - Z_t)\delta_t$
  - μ_l: long term mean of returns, e.g., 0
  - ε_t, δ_t: positive and negative shocks, non-negative, i.i.d
- $f_\varepsilon(x) = \frac{{\lambda_1}^{\alpha_1} x^{\alpha_1 - 1}}{\Gamma(\alpha_1)} e^{-\lambda_1 x}$
- $f_\delta(x) = \frac{{\lambda_2}^{\alpha_2} x^{\alpha_2 - 1}}{\Gamma(\alpha_2)} e^{-\lambda_2 x}$

Estimation of Parameters



Markov Property

- Given the current information available at time $(t-1)$, the history, e.g., path, is irrelevant.
- $P(q_t|q_{t-1},\cdots,q_1) = P(q_t|q_{t-1})$
- Consistent with the weak form of the efficient market hypothesis.

Hidden Markov Chain

- Only observations are observable (duh).
- World states may not be known (hidden).
  - We want to model the hidden states as a Markov Chain.
- Two assumptions:
  - Markov property
  - $P(\omega_t | q_{t-1}, \cdots, q_1, \omega_{t-1}, \cdots, \omega_1) = P(\omega_t | q_t)$

Problems

- Likelihood
  - Given the parameters, λ, and an observation sequence, Ω, compute $P(\Omega|\lambda)$.
- Decoding
  - Given the parameters, λ, and an observation sequence, Ω, determine the best hidden sequence Q.
- Learning
  - Given an observation sequence, Ω, and HMM structure, learn λ.

Learning as a Maximization Problem

- Our objective is to find λ that maximizes $P(\Omega|\lambda)$.
- For any given λ, we can compute $P(\Omega|\lambda)$.
- Then solve a maximization problem.
- Algorithms:
  - Nelder-Mead
  - Baum-Welch (E-M algorithm)

Stationary State

- $\Pi = \frac{1-q}{2-p-q}$
- $R_t = \mu_l + \varepsilon_t \geq \mu_l$, with probability Π
- $R_t = \mu_l - \delta_t < \mu_l$, with probability $1 - \Pi$

Step 3

- Assume the long term mean is 0, $\mu_l = 0$.
- When $n = 2$,
  - $(B_t \geq 0) \equiv (R_t \geq 0) \equiv (Z_t = 1)$
  - $(B_t < 0) \equiv (R_t < 0) \equiv (Z_t = 0)$

GMA(2, 1) – Naïve MA Trading Rule

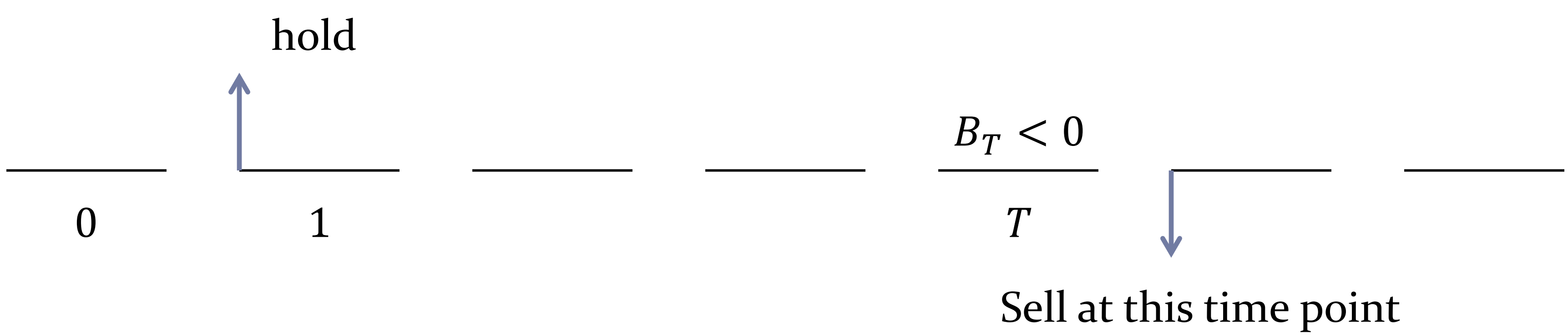
- Buy when the asset return in the present period is positive.
- Sell when the asset return in the present period is negative.

Naïve MA Conditions

- The expected value of the positive shocks to asset return >> the expected value of negative shocks.
- The positive shocks persistency >> that of negative shocks.

T Period Returns

- $RR_T = \sum_{t=1}^{T} R_t \times I_{\{B_{t-1} \geq 0\}}$



Holding Time Distribution

- $P(N = T)$
- $= P(B_T < 0, B_{T-1} \geq 0, \ldots, B_1 \geq 0, B_0 \geq 0)$
- $= P(Z_T = 0, Z_{T-1} = 1, \ldots, Z_1 = 1, Z_0 = 1)$
- $= P(Z_T = 0, Z_{T-1} = 1, \ldots, Z_1 = 1 | Z_0 = 1) P(Z_0 = 1)$
- $= \begin{cases} \Pi p^{T-1}(1-p), & \text{T} \geq 1 \\ 1 - \Pi, & \text{T}=0 \end{cases}$

Conditional Returns Distribution (1)

- $\Phi_{RR_T|N=T}(s) = \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t \times I_{\{B_{t-1}\geq 0\}}\right]s\right\}}|N = T\right]$
- $= \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t \times I_{\{B_{t-1}\geq 0\}}\right]s\right\}}|B_T < 0, B_{T-1} \geq 0, \ldots, B_0 \geq 0\right]$
- $= \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t\right]s\right\}}|Z_T = 0, Z_{T-1} = 1, \ldots, Z_1 = 1\right]$
- $= \mathrm{E}\left[e^{\{i[\varepsilon_1 + \cdots + \varepsilon_{T-1} - \delta_T]s\}}\right]$
- $= \begin{cases} {\Phi_\varepsilon}^{T-1}(s)\Phi_\delta(-s), \ \mathrm{T} \geq 1 \\ \Phi_\delta(-s), \ \mathrm{T} = 0 \end{cases}$

Unconditional Returns Distribution (2)

- $\Phi_{RR_T}(s) = \sum_{T=0}^{\infty} \mathrm{E}\left[e^{\left\{i\left[\sum_{t=1}^{T} R_t \times I_{\{B_{t-1} \geq 0\}}\right]s\right\}} | N = T\right] P(N = T)$
- $= \sum_{T=1}^{\infty} \Pi p^{T-1}(1-p){\Phi_\varepsilon}^{T-1}(s)\Phi_\delta(-s) + (1-\Pi)\Phi_\delta(-s)$
- $= (1-\Pi)\Phi_\delta(-s) + \Pi(1-p)\frac{\Phi_\delta(-s)}{1-p\Phi_\varepsilon(s)}$

Long-Only Returns Distribution

- $\Phi_{RR_T}(s|R_0 \geq 0) = \frac{(1-p)\Phi_\delta(-s)}{1-p\Phi_\varepsilon(s)}$
- Proof: make $P(Z_0 = 1) = \Pi = 1$

Expected Returns

- $\mathrm{E}(RR_T) = -i\Phi_{RR_T}{}'(0)$
- $= \frac{1}{1-p}\{\Pi p\mu_\varepsilon - (1-p)\mu_\delta\}$
- When is the expected return positive?
  - $\mu_\varepsilon \geq \frac{1-p}{\Pi p}\mu_\delta$, shock impact
  - $\mu_\varepsilon \gg \mu_\delta$, shock impact
  - $\Pi p \geq 1-p$, if $\mu_\varepsilon \approx \mu_\delta$, persistence

GMA(∞,1) Rule

- $P_t \geq \left(\prod_{j=0}^{n-1} P_{t-j}\right)^{\frac{1}{n}}$
- $\ln P_t \geq \frac{1}{n}\sum_{j=0}^{n-1} \ln P_{t-j}$
- $\ln P_t \geq \mu_1$

GMA(∞,1) Expected Returns

- $\Phi_{RR_T}(s) = (1-\Pi)q[\Phi_\delta(s)+\Phi_\delta(-s)] + [1-p(1-\Pi)][\Phi_\varepsilon(s)+\Phi_\varepsilon(-s)]$
- $\mathrm{E}(RR_T) = -[1-p(1-\Pi)][\mu_\varepsilon+\mu_\delta]$

MA Using the Whole History

- An investor will always expect to lose money using GMA(∞,1)!
- An investor loses the least amount of money when the return process is a random walk.

Optimal MA Parameters

- So, what are the optimal n and m?

A Mathematical Analysis of Linear Technical Indicators

Linear Technical Indicators

- As we shall see, a number of linear technical indicators, including the Moving Average Crossover, are really the "same" *generalized* indicator using different parameters.

The Generalized Linear Trading Rule

- A linear predictor of weighted lagged returns
  - $F_t = \delta + \sum_{j=0}^{t} d_j X_{t-j}$
- The trading rule
  - Long: $B_t = 1$, iff, $F_t > 0$
  - Short: $B_t = -1$, iff, $F_t < 0$
- (Unrealized) rule returns
  - $R_t = B_{t-1} X_t$
    - $R_t = -X_t$ if $B_{t-1} = -1$
    - $R_t = +X_t$ if $B_{t-1} = +1$

Buy And Hold

- $B_t = 1$

Predictor Properties

- Linear
- Autoregressive
- Gaussian, assuming X_t is Gaussian
- If the underlying returns process is linear, F_t yields the best forecasts in the mean squared error sense.

Returns Variance

- $\mathrm{Var}(R_t) = \mathrm{E}\left({R_t}^2\right) - (\mathrm{E}(R_t))^2$
- $= \mathrm{E}\left({B_{t-1}}^2 {X_t}^2\right) - (\mathrm{E}(R_t))^2$
- $= \mathrm{E}\left({X_t}^2\right) - (\mathrm{E}(R_t))^2$
- $= \sigma^2 + \mu^2 - (\mathrm{E}(R_t))^2$

Maximization Objective

- Variance of returns is inversely proportional to expected returns.
- The more profitable the trading rule is, the less risky this will be if risk is measured by volatility of the portfolio.
- Maximizing returns will also maximize returns per unit of risk.

Expected Returns

- $\mathrm{E}(R_t) = \mathrm{E}(B_{t-1}X_t)$
- $= \mathrm{E}(B_{t-1}(\mu + \sigma N))$
- $= \sigma\,\mathrm{E}(B_{t-1}N) + \mu\,\mathrm{E}(B_{t-1})$
- $\mathrm{E}(B_{t-1}) = 1 \times \mathrm{P}(F_{t-1} > 0) + -1 \times \mathrm{P}(F_{t-1} < 0)$
- $= \mathrm{P}(F_{t-1} > 0) - \mathrm{P}(F_{t-1} < 0)$
- $= 1 - 2 \times \mathrm{P}(F_{t-1} < 0)$
- $= 1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)$

Truncated Bivariate Moments

- Johnston and Kotz, 1972, p.116
- $\mathrm{E}(B_{t-1}N) = \iint_{F_t>0} N - \iint_{F_t<0} N$
- $= \sqrt{\frac{2}{\pi}}\rho e^{-\frac{{\mu_F}^2}{2{\sigma_F}^2}}$
- Correlation:
  - $\rho = \mathrm{Corr}(X_t, F_{t-1})$

Expected Returns As a Weighted Sum

- $\mathrm{E}(R_t) = \sigma\,\mathrm{E}(B_{t-1}N) + \mu\,\mathrm{E}(B_{t-1})$
- $= \sigma\sqrt{\frac{2}{\pi}}\rho e^{-\frac{\mu_F{}^2}{2\sigma_F{}^2}} + \mu\left(1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)$

a term for volatility

a term for drift

Praetz model, 1976

- Returns as a random walk with drift.
- $E(R_t) = \mu(1 - 2f)$, f the frequency of short positions
- $\mathrm{Var}(R_t) = \sigma^2$

Comparison with Praetz model

- Random walk implies $\rho = \mathrm{Corr}(X_t, F_{t-1}) = 0$.
- $\mathrm{E}(R_t) = \mu\left(1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)$
- $\mathrm{Var}(R_t) = \sigma^2 + \mu^2 - \left\{\mu\left(1 - 2 \times \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)\right\}^2$
- $= \sigma^2 + 4\mu^2\,\Phi\left(-\frac{\mu_F}{\sigma_F}\right)\left(1 - \Phi\left(-\frac{\mu_F}{\sigma_F}\right)\right)$

the probability of being short

increased variance

Biased Forecast

- A biased (Gaussian) forecast may be suboptimal.
- Assume underlying mean $\mu = 0$.
- Assume forecast mean $\mu_F \neq 0$.
- $\mathrm{E}(R_t) = \sigma\sqrt{\frac{2}{\pi}}\rho e^{-\frac{{\mu_F}^2}{2{\sigma_F}^2}} \leq \sigma\sqrt{\frac{2}{\pi}}\rho$

Maximizing Returns

- Maximizing the correlation between forecast and one-ahead return.
- First order condition:
  - $\frac{\mu_F}{\sigma_F} = \frac{\mu}{\sigma\rho}$

First Order Condition

- Let x $= \frac{\mu_F}{\sigma_F}$
- $\mathrm{E}(R_t) = \sigma\sqrt{\frac{2}{\pi}}\rho e^{-\frac{x^2}{2}} + \mu(1 - 2 \times \Phi(-x))$
- $\frac{d\ \mathrm{E}(R_t)}{dx} = 0$
  - $\sigma\cancel{\sqrt{\frac{2}{\pi}}}\rho(-x)\cancel{e^{-\frac{x}{2}}} + \mu\cancel{\sqrt{\frac{2}{\pi}}}\cancel{e^{-\frac{x^2}{2}}} = 0$
  - $x = \frac{\mu_F}{\sigma_F} = \frac{\mu}{\sigma\rho}$

Fitting vs. Prediction

- If X_t process is Gaussian, no linear trading rule obtained from a finite history of X_t can generate expected returns over and above F_t.
- Minimizing mean squared error $\neq$ maximizing P&L.
- In general, the relationship between MSE and P&L is highly non-linear (Acar 1993).

Technical Analysis

- Use a finite set of historical prices.
- Aim to maximize profit rather than to minimize mean squared error.
- Claim to be able to capture complex non-linearity.
- Certain rules are ill-defined.

Technical Linear Indicators

- For any technical indicator that generates signals from a finite linear combination of past prices
  - Sell: $B_t = -1$ iff $\sum_{j=0}^{m-1} a_j P_{t-j} < 0$
- There exists an (**almost**) equivalent AR rule.
  - Sell: $\widetilde{B_t} = -1$ iff $\delta + \sum_{j=0}^{m-2} d_j X_{t-j} < 0$
  - $X_t = \ln \frac{P_t}{P_{t-1}}$
  - $\delta = \sum_{j=0}^{m-1} a_j$, $d_j = -\sum_{i=j}^{m-2} a_i$

Conversion Assumption

- $1 - \frac{P_{t-j}}{P_t} \approx \ln \frac{P_t}{P_{t-j}}$
- Monte Carlo simulation:
  - 97% accurate
  - 3% error.

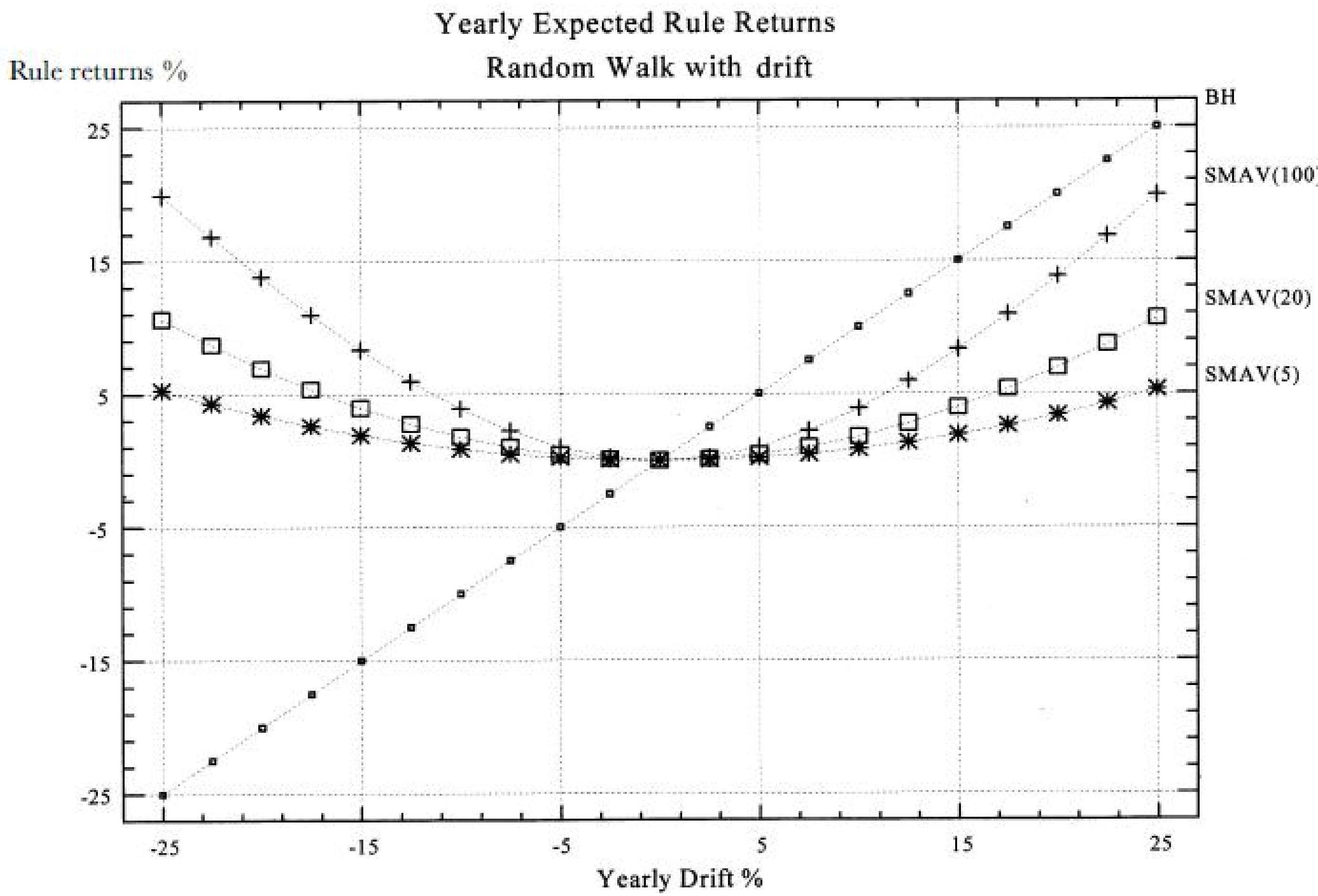
Example Linear Technical Indicators

- Simple order
- Simple MA
- Weighted MA
- Exponential MA
- Momentum
- Double orders
- Double MA

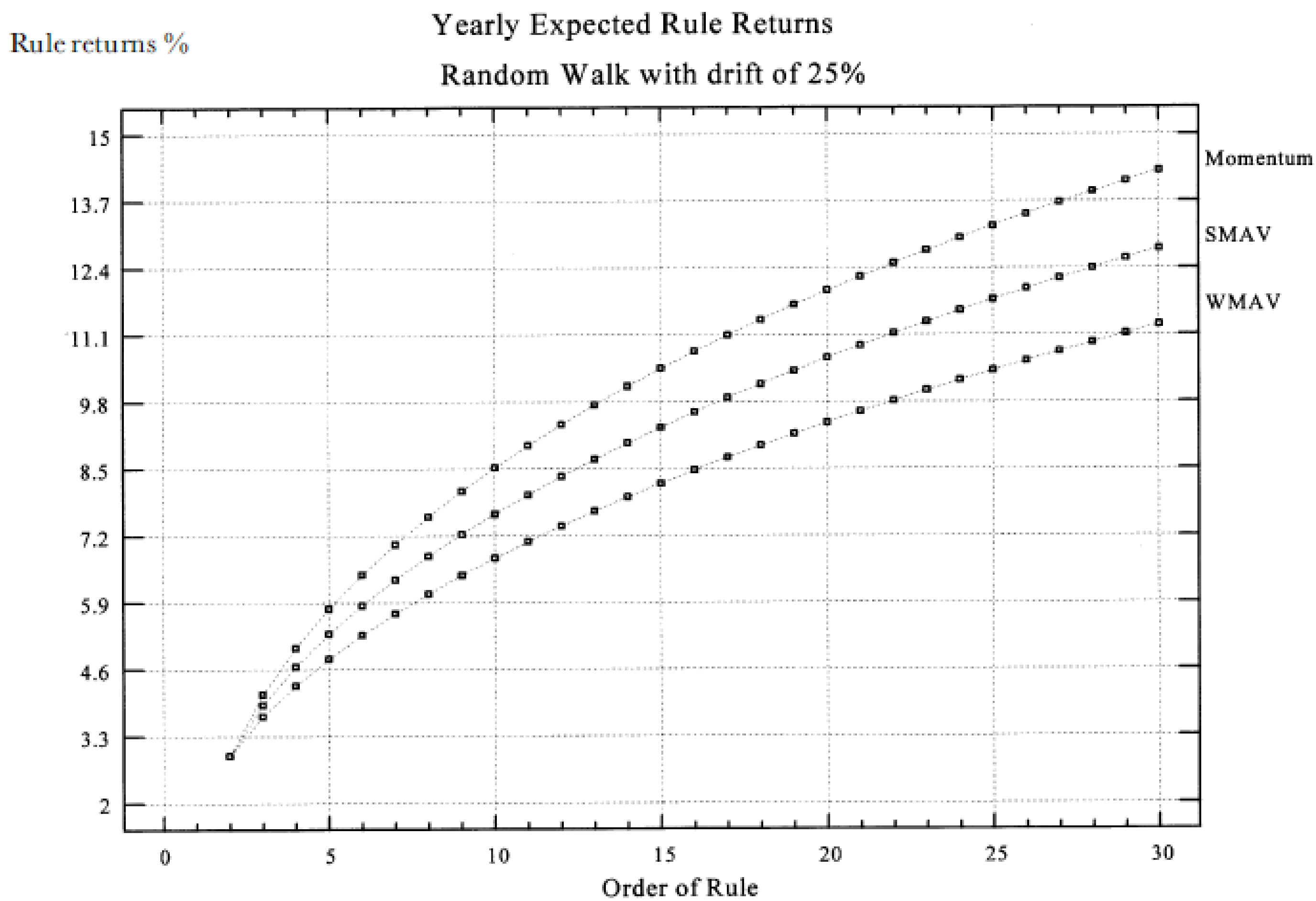
Returns: Random Walk With Drift

- $X_t = \mu + \varepsilon_t$
  - The bigger the order, the better.
  - Momentum > SMAV > WMAV
- How to estimate the *future* drift?
  - Crystal ball?
  - Delphic oracle?

Results


Yearly Expected Rule Returns
Random Walk with drift
Rule returns %
BH
SMAV(100)
SMAV(20)
SMAV(5)
25
15
5
-5
-15
-25
-25
-15
-5
5
15
25
Yearly Drift %

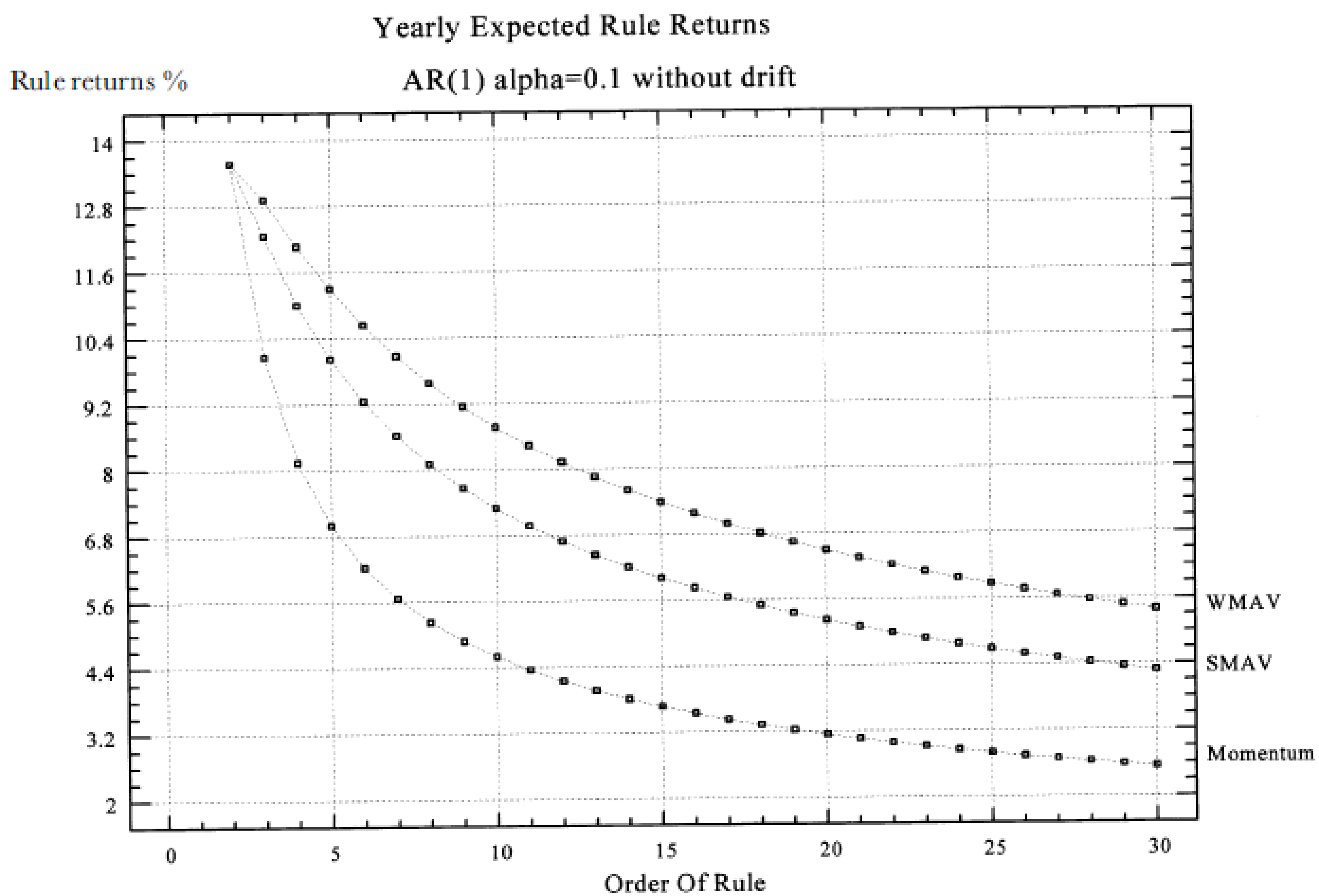
Results


Rule returns %
Yearly Expected Rule Returns
Random Walk with drift of 25%
15
13.7
12.4
11.1
9.8
8.5
7.2
5.9
4.6
3.3
2
0
5
10
15
20
25
30
Order of Rule
Momentum
SMAV
WMAV

Returns: AR(1)

- $X_t = \alpha X_{t-1} + \varepsilon_t$
  - Auto-correlation is required to be profitable.
  - The smaller the order, the better. (quicker response)

Results



Yearly Expected Rule Returns
AR(1) alpha=0.1 without drift
Rule returns %
14
12.8
11.6
10.4
9.2
8
6.8
5.6
4.4
3.2
2
0
5
10
15
20
25
30
Order Of Rule
WMAV
SMAV
Momentum

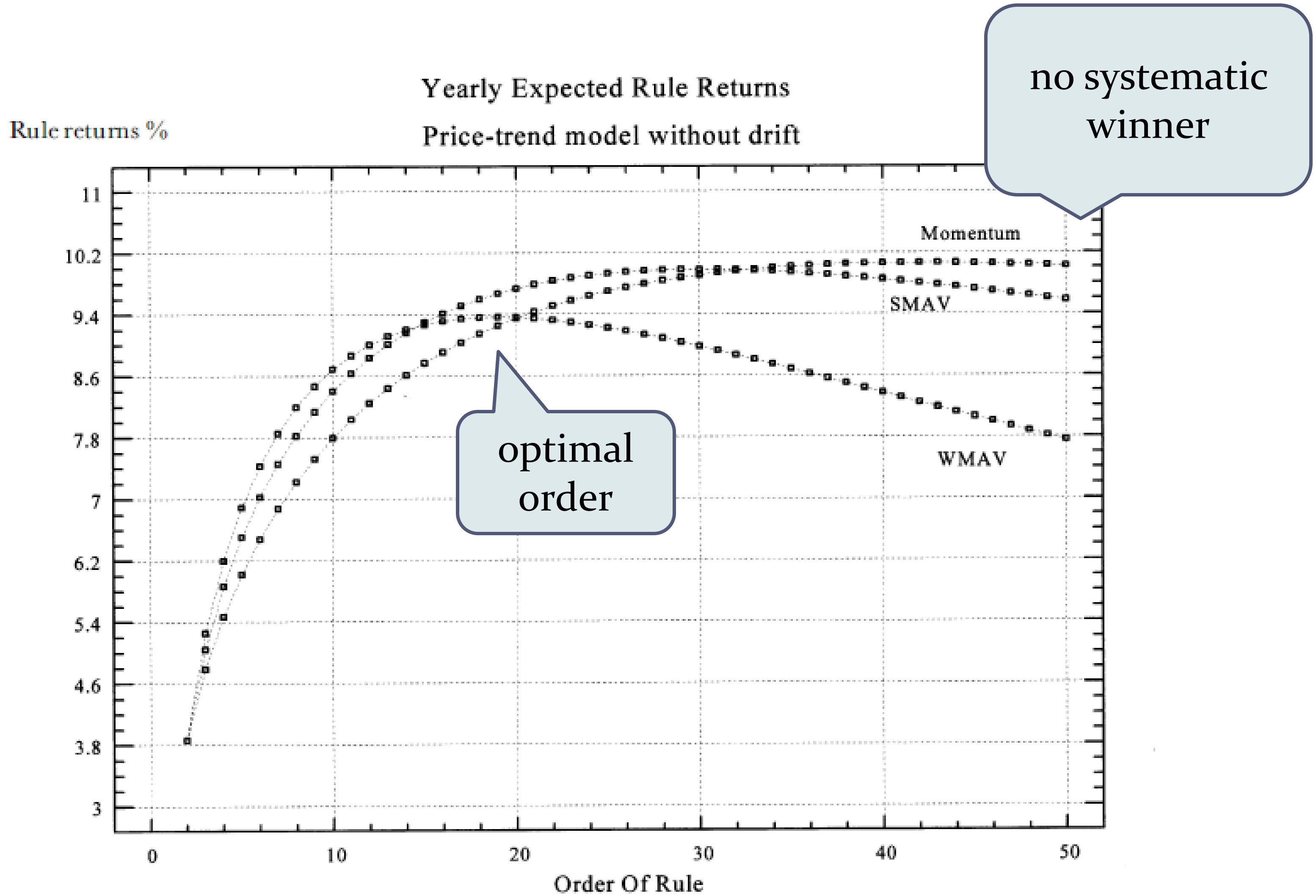
ARMA(1, 1)

AR

MA

- $(X_t - \mu) - p(X_{t-1} - \mu) = \varepsilon_t - q\varepsilon_{t-1}$
- Prices tend to move in one direction (trend) for a period of time and then change in a random and unpredictable fashion.
  - Mean duration of trends: $m_d = \frac{1}{(1-p)}$
- Information has impacts on the returns in different days (lags).
  - Returns correlation: $\rho_h = Ap^h$

Results

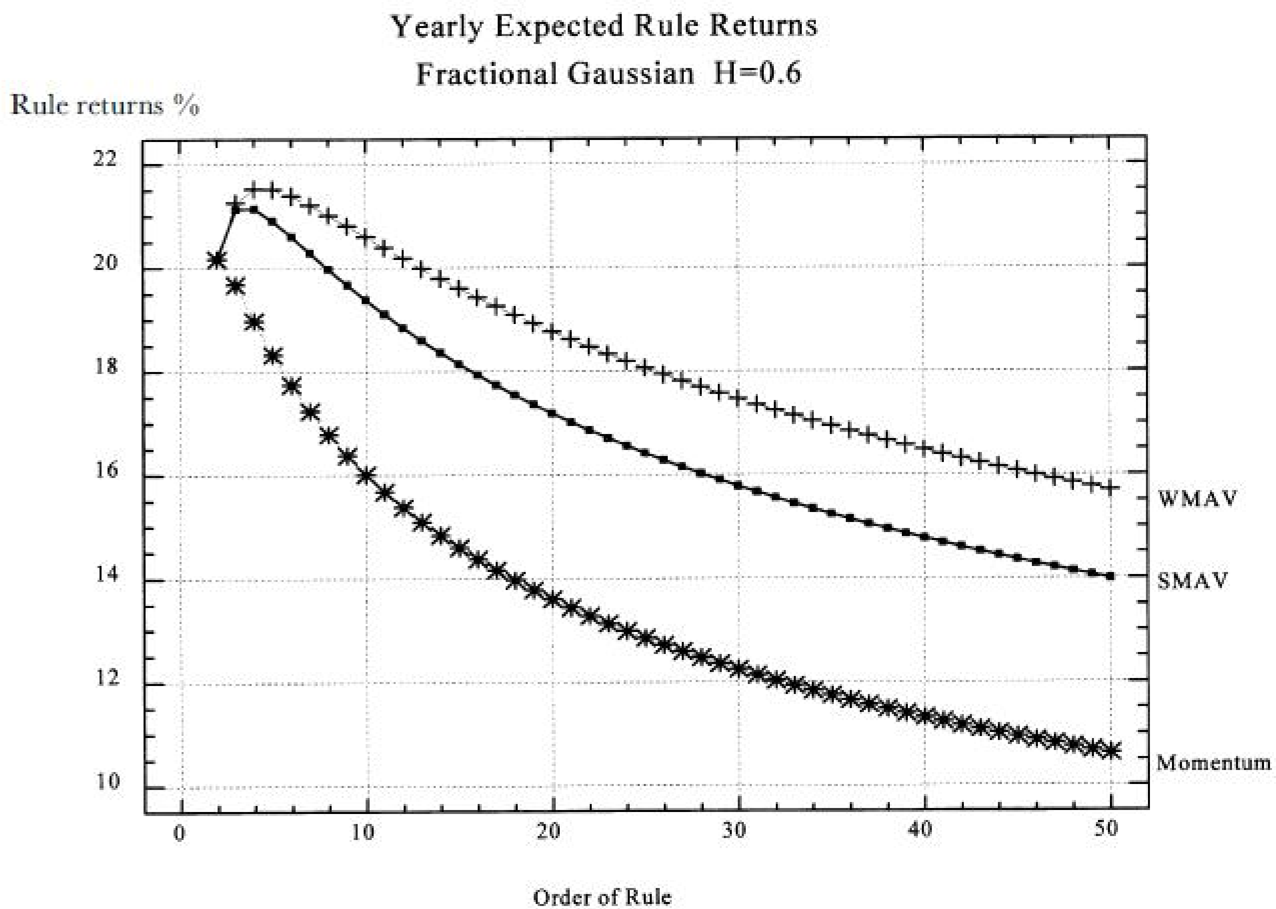


Yearly Expected Rule Returns
Price-trend model without drift
Rule returns %
11
10.2
9.4
8.6
7.8
7
6.2
5.4
4.6
3.8
3
0
10
20
30
40
50
Order Of Rule
Momentum
SMAV
WMAV
no systematic winner
optimal order

ARIMA(0, d, 0)

- $\nabla^d(X_t - \mu) = e_t$
- Irregular, erratic, aperiodic cycles.

Results



Yearly Expected Rule Returns
Fractional Gaussian H=0.6
Rule returns %
22
20
18
16
14
12
10
0
10
20
30
40
50
WMAV
SMAV
Momentum
Order of Rule

ARCH(p)

- $X_t = \mu + \left\{ \sqrt{\alpha_0 + \sum_{i=1}^{p} \alpha_i (X_{t-i} - \mu)^2} \right\} \varepsilon_t$
- $X_t - \mu$ are the residuals
- When $\mu = 0$, $\mathrm{E}(R_t) = 0$.

residual coefficients as a function of lagged squared residuals

AR(2) + GARCH(1,1)

AR(2)

GARCH(1,1)

- $X_t = a + b_1 X_{t-1} + b_2 X_{t-2} + \varepsilon_t$
- $\varepsilon_t = \sqrt{h_t} z_t$
- $h_t = \alpha_0 + \alpha_1 {\varepsilon_{t-1}}^2 + \beta h_{t-1}$

innovations

ARCH(1): lagged squared residuals

lagged variance

Results

- The presence of conditional heteroskedasticity will not drastically affect returns generated by linear rules.
- The presence of conditional heteroskedasticity, if unrelated to serial dependencies, may be neither a source of profits nor losses for linear rules.

Playing Trend Following Strategies

- Trend following model requires positive (negative) autocorrelation to be profitable.
  - What do you do when there is zero autocorrelation?
- Trend following models are profitable when there are drifts.
  - How to estimate drifts?
- It seems quicker response rules tend to work better.
- Weights should be given to the more recent data.

Conclusions

The Essential Skills

- Financial intuitions, market understanding, creativity.
- Mathematics.
- Computer programming.

An Emerging Field

- It is a financial industry where mathematics and computer science meet.
- It is an arms race to build
  - more reliable and faster execution platforms (computer sciences);
  - more comprehensive and accurate prediction models (mathematics).
- Structured products seem to be an evening industry after the financial crisis. Could quantitative trading be another gold mine?

Programming and Backtesting Quantitative Trading Strategies

AlgoQuant – A Quantitative Trading Research Toolbox

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

Speaker Profile

- Dr. Haksun Li
- CEO, Numerical Method Inc.
- (Ex-)Adjunct Professors, Industry Fellow, Advisor, Consultant with the National University of Singapore, Nanyang Technological University, Fudan University, the Hong Kong University of Science and Technology.
- Quantitative Trader/Analyst, BNPP, UBS
- PhD, Computer Sci, University of Michigan Ann Arbor
- M.S., Financial Mathematics, University of Chicago
- B.S., Mathematics, University of Chicago

The Role of Technology in Quantitative Trading

Step 2 - Coding

- After modeling, we code up the quantitative trading strategy for
  - backtesting (in-sample and out-sample),
  - computing the properties, e.g., expected P&L, max drawdown, using simulation,
  - calibrating parameters,
  - analyzing sensitivity,
  - trading live.

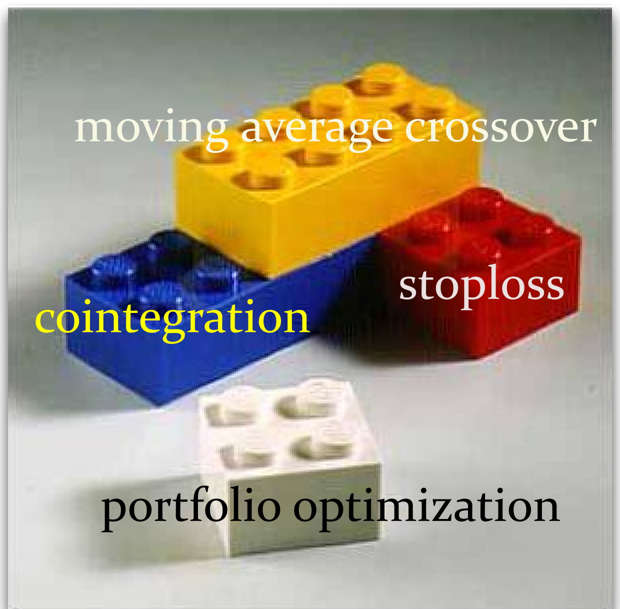
Ideal

- A trader dreams of a trading idea.
- He can quickly code it up to produce a prototype.
- He feeds the prototype to a computer system to automatically produce a report about:
  - backtesting (in-sample and out-sample)
  - computing the properties, e.g., expected P&L, max drawdown, using simulation,
  - calibrating parameters,
  - analyzing sensitivity,
- When he is happy with the report, he can trade the model live.

Building Blocks

- Moving average crossover, Acar & Satchell 2002
- Bull/bear market probabilities, Dai 2011
- Cointegration
- Pairs trading model calibration, Elliott 2005
- Mean reverting portfolio construction, d'Aspremont 2008
- Mean-variance portfolio optimization, Lai 2009
- Cone optimization of portfolio
- Factor models
- Many more……

Creating Strategy Like Building LEGO©



Reality

- Clean data
- Align time stamps
- Read Gigabytes of data
  - Retuers' EURUSD, tick-by-tick, is 1G/day
- Extract relevant information
  - PE, BM
- Handle missing data
- Incorporate events, news and announcements
- Code up the quant. strategy
- Code up the simulation
  - Bid-ask spread
  - Slippage
  - Execution assumptions
- Wait a very long time for the simulation to complete
- Recalibrate parameters and simulate again
- Wait a very long time for the simulation to complete
- Recalibrate parameters and simulate again
- Wait a very long time for the simulation to complete

- Debug
- Debug again
- Debug more
- Debug even more
- Debug patiently
- Debug impatiently
- Debug frustratingly
- Debug furiously
- Give up
- Start to trade

Research Tools – Very Primitive

- Excel
- MATLAB/ R/ other scripting languages…
- MetaTrader/ Trade Station
- RTS/ other automated trading systems…

R/ Scripting Languages Advantages

- Most people already know it.
  - There are more people who know Java/C#/C++/C than Matlab, R, etc., combined.
- It has a huge collection of math <u>functions</u> for math modeling and analysis.
  - Math libraries are also available in SuanShu (Java), NMath (C#), Boost (C++), and Netlib (C).

R Disadvantages

- TOO MANY!

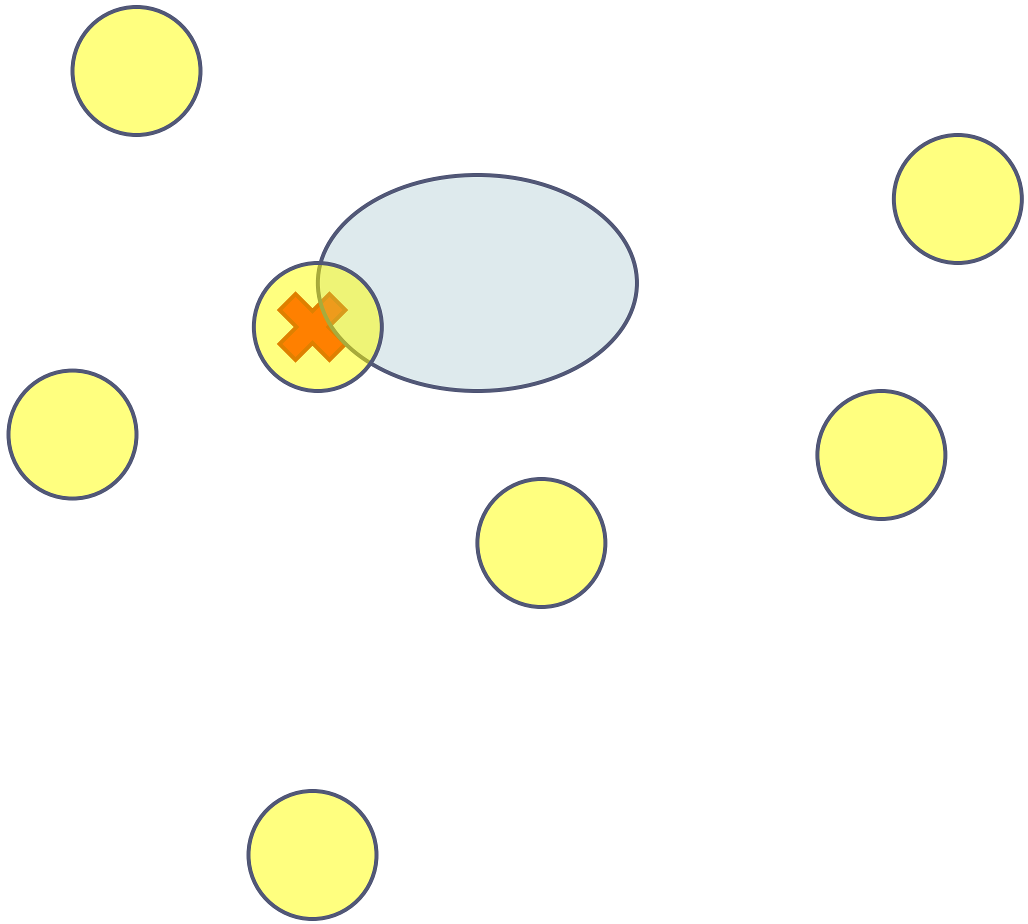
Some R Disadvantages

- Way too slow
  - Must interpret the code line-by-line
- Limited memory
  - How to read and process gigabytes of tick-by-tick data
- Limited parallelization
  - Cannot calibrate/simulate a strategy in many scenarios in parallel
- Inconvenient editing
  - No usage, rename, auto import, auto-completion
- Primitive debugging tools
  - No conditional breakpoint, disable, thread switch and resume
- Obsolete C-like language
  - No interface, inheritance; how to define $f(x)$?

R's Biggest Disadvantage

- You cannot be sure your code is right!

Productivity



Research Tools As Weapon in Trading Warfare

bare hand star trader

Excel

R/MATLAB

MT/TS

AlgoQuant

A Good Trading Research Toolbox (1)

- Allow easy strategy coding
- Allow plug-and-play multiple strategies
- Simulate using historical data
- Simulate using fake, artificial data
- Allow controlled experiments
  - e.g., bid/ask, execution assumptions, news

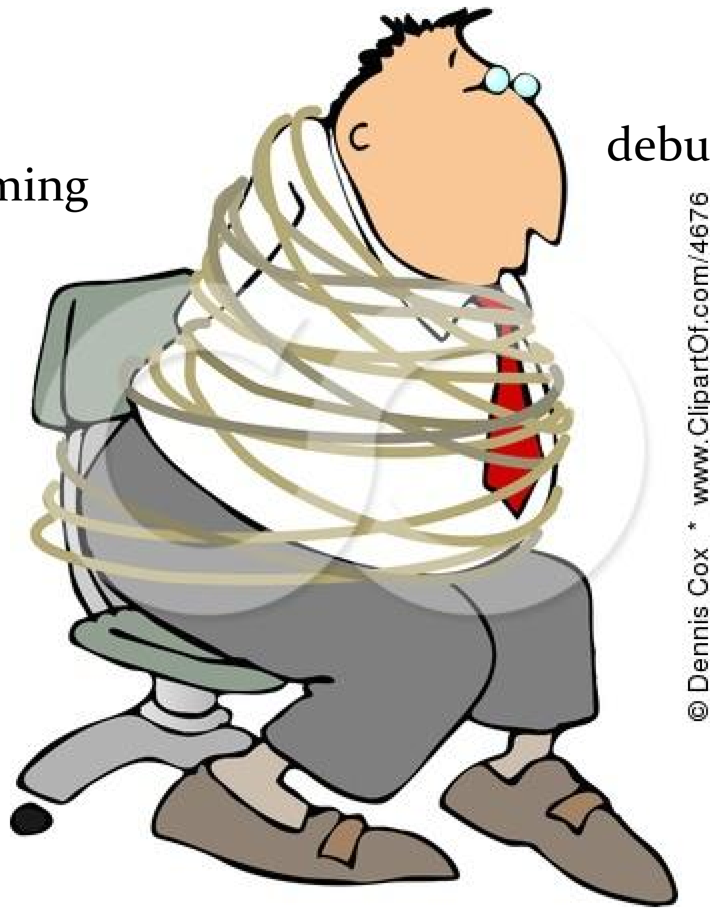
A Good Trading Research Toolbox (2)

- Generate standard and user customized statistics
- Have information other than prices
  - e.g., macro data, news and announcements
- Auto parameter calibration
- Sensitivity analysis
- Quick

Free the Trader!

programming

debugging



calibrating

data extracting

data cleaning

waiting

backtesting

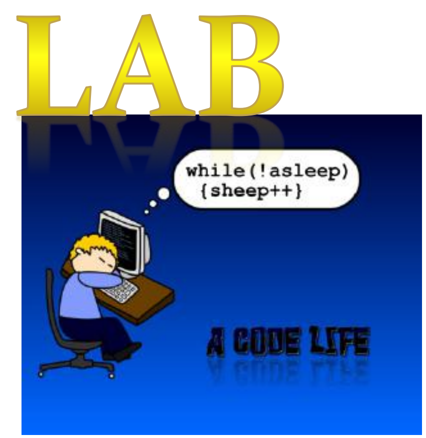
Basic Math Programming in Java

Downloads

- JDK
  - http://docs.oracle.com/javase/tutorial/
- NetBeans
- AlgoQuant
  - http://www.numericalmethod.com/trac/numericalmethod/wiki/AlgoQuant

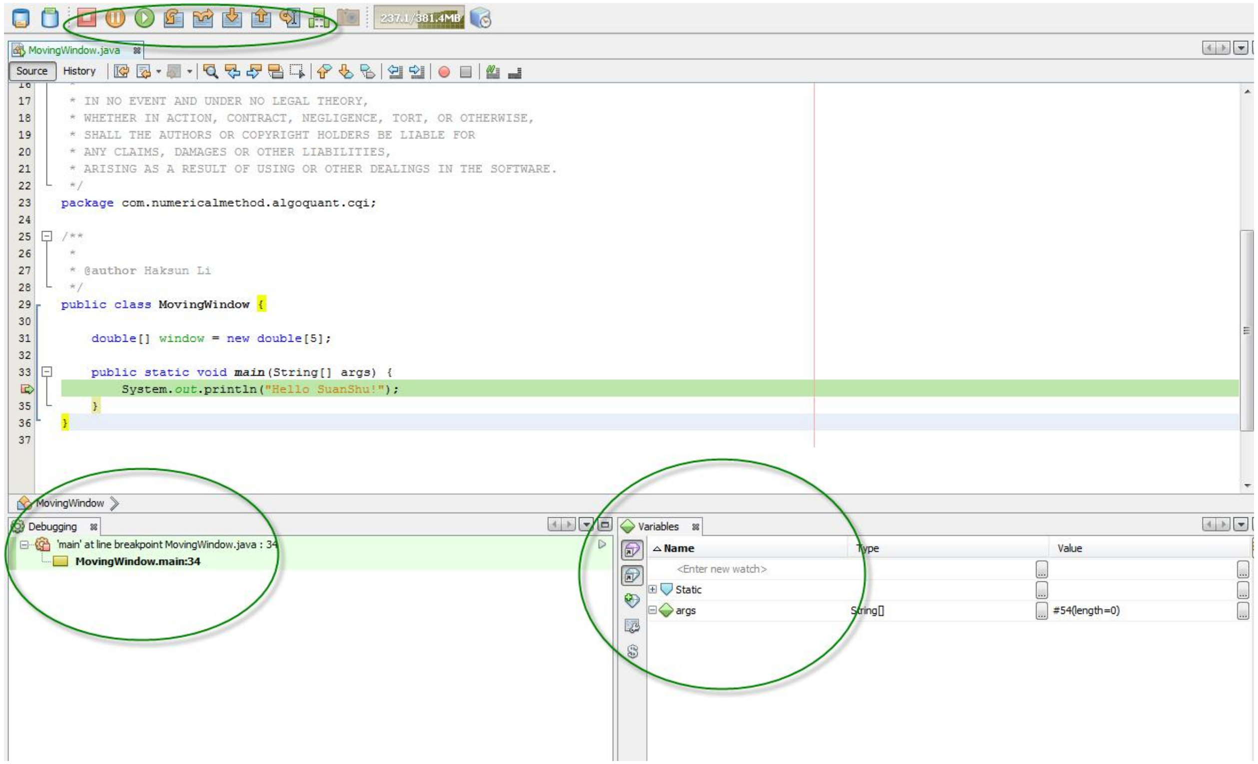
Procedural Programming

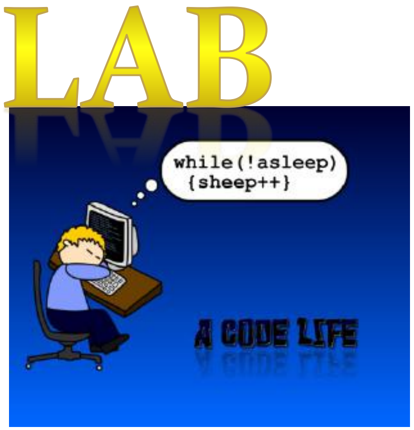
- The program is a series of computational steps to be carried out.
- The order of execution is linear from the first statement to the second and so forth
  - with occasional loops and branches.



Debugging and Testing

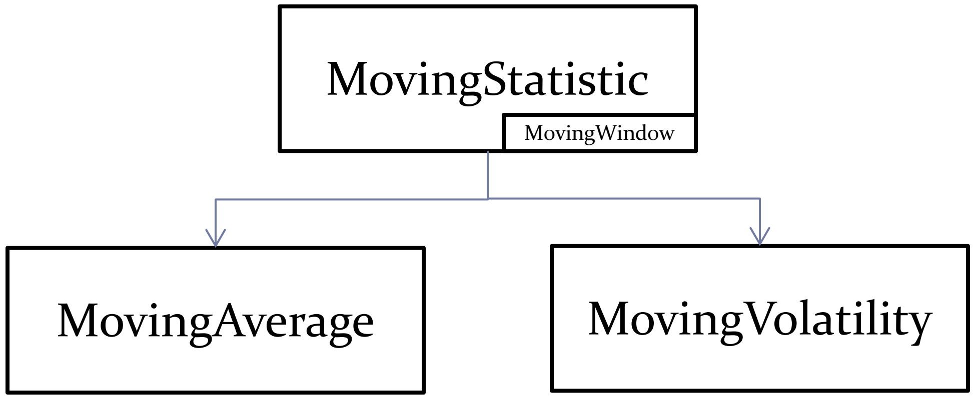
- F8/F7
- JUnit

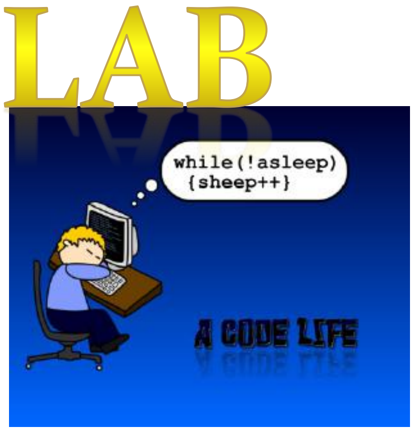




Object-Oriented Programming

- Represent "Concepts" as "Objects".
- Objects have
  - data fields,
  - methods to handle the data fields.
- Inheritance: a hierarchical relationships among objects.

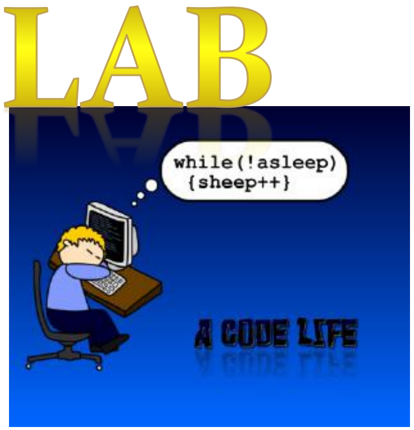




Strategy Programming

Backtesting in AlgoQuant

1. Define a data source, e.g., the asset that you want to trade.
2. Construct an instance of the strategy to be backtested.
3. Create an order book.
4. Run the simulation.
5. Analysis the performance statistics.

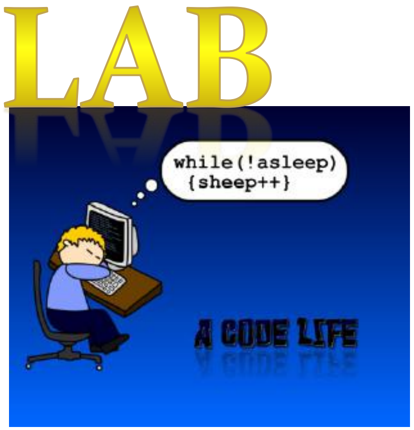


Event-Driven Programming

- An object reacts to the events that it listens to.
  - E.g., prices but not volumes.
- A trading strategy code is a set of event handlers that
  - update the internal states,
  - send orders to the (mock) market

Strategy Programming

- GMA(2,1)



Backtesting

- Historical data
- Bootstrapped data
- Simulated data

NUMERICAL METHOD

Trading Basket Construction

Mean Reversion Trading

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

Speaker Profile

- Dr. Haksun Li
- CEO, Numerical Method Inc.
- (Ex-)Adjunct Professors, Industry Fellow, Advisor, Consultant with the National University of Singapore, Nanyang Technological University, Fudan University, the Hong Kong University of Science and Technology.
- Quantitative Trader/Analyst, BNPP, UBS
- PhD, Computer Sci, University of Michigan Ann Arbor
- M.S., Financial Mathematics, University of Chicago
- B.S., Mathematics, University of Chicago

References

- Pairs Trading: A Cointegration Approach. Arlen David Schmidt. University of Sydney. Finance Honours Thesis. November 2008.
- Likelihood-Based Inference in Cointegrated Vector Autoregressive Models. Soren Johansen. Oxford University Press, USA. February 1, 1996.
- Pairs Trading. Elliot, van der Hoek, and Malcolm. 2005.
- Identifying Small Mean Reverting Portfolios. A. d'Aspremont. 2008.
- High-frequency Trading. Stanford University. Jonathan Chiu, Daniel Wijaya Lukman, Kourosh Modarresi, Avinayan Senthi Velayutham. 2011.

Paris Trading

Pairs Trading

- Intuition: The thousands of market instruments are not independent. For two closely related assets, they tend to "move together" (common trend). We want to buy the cheap one and sell the expensive one.
  - Exploit short term deviation from long term equilibrium.
- Definition: trade one asset (or basket) against another asset (or basket)
  - Long one and short the other
- Try to make money from "spread".

FOMC announcements

- Rate cut!
- Bonds react.
- FX react.
- Stocks react.
- All act!

GLD vs. SLV

Zoom: 1d 5d 1m 3m 6m YTD 1y 5y 10y All
Nov 26, 2004 - Oct 07, 2013
GLD +185.31% SLV +56.02%
300%
250%
200%
150%
100%
50%
0%
-50%
2007
2008
2009
2010
2011
2012
2013

Zoom: 1d 5d 1m 3m 6m YTD 1y 5y 10y All
Dec 31, 2012 - Oct 07, 2013
GLD -20.42% SLV -25.94%
5%
0%
-5%
-10%
-15%
-20%
-25%
-30%
-35%
-40%
Jan 2013
Feb 2013
Mar 2013
Apr 2013
May 2013
Jun 2013
Jul 2013
Aug 2013
Sep 2013

Zoom: 1d 5d 1m 3m 6m YTD 1y 5y 10y All
Jul 08, 2013 - Oct 07, 2013
GLD +8.18% SLV +18.33%
30%
25%
20%
15%
10%
5%
0%
Jul 12
Jul 19
Jul 26
Aug 2
Aug 9
Aug 16
Aug 23
Aug 30
Sep 6
Sep 13
Sep 20
Sep 26
Oct 4

Hows

- How to construct a pair?
- How to trade a pair?

Sample Pairs Trading Strategy

Spread

- $Z = X - \beta Y$
- β
  - hedge ratio
  - Cointegration coefficient
- How do you trade spread?
  - How much X to buy/sell?
  - How much Y to buy/sell?

Log-Spread

- $Z = \log X - \beta \log Y$
- How do you trade log-spread?
  - How much X to buy/sell?
  - How much Y to buy/sell?

Dollar Neutral Hedge

- Suppose ES (S&P500 E-mini future) is at 1220 and each point worth $50, its dollar value is about $61,000. Suppose NQ (Nasdaq 100 E-mini future) is at 1634 and each point worth $20, its dollar value is $32,680.
- $\beta = \frac{61000}{32680} = 1.87.$
- $Z = ES - 1.87 \times NQ$
- Buy Z = Buy 10 ES contracts and Sell 19 NQ contracts.
- Sell Z = Sell 10 ES contracts and Buy 19 NQ contracts.

Market Neutral Hedge

- Suppose ES has a market beta of 1.25, NQ 1.11.
- We use $\beta = \frac{1.25}{1.11} = 1.13$

Dynamic Hedge

- β changes with time, covariance, market conditions, etc.
- Periodic recalibration.

Distance Method

- The distance between two time series:
  - $d = \sum (x_i - y_j)^2$
- x_i, y_j are the normalized prices.
- We choose a pair of stocks among a collection with the smallest distance, d.

Distance Trading Strategy

- Sell Z if Z is too expensive.
- Buy Z if Z is too cheap.
- How do we do the evaluation?

Z Transform

- We normalize Z.
- The normalized value is called z-score.
- $z = \frac{x - \bar{x}}{\sigma_x}$
- Other forms:
- $z = \frac{x - M \times \bar{x}}{S \times \sigma_x}$
  - M, S are proprietary functions for forecasting.

A Very Simple Distance Pairs Trading

- Sell Z when $z > 2$ (standard deviations).
  - Sell 10 ES contracts and Buy 19 NQ contracts.
- Buy Z when $z < -2$ (standard deviations).
  - Buy 10 ES contracts and Sell 19 NQ contracts.

Pros of the Distance Model

- Model free.
- No mis-specification.
- No mis-estimation.
- Distance measure intuitively captures the Law of One Price (LOP) idea.

Cons of the Distance Model

- There is no reason why the model will work (or not). There is no assumption to check against the current market conditions.
- The model is difficult to analyze mathematically.
  - Cannot predict the convergence time (expected holding time).
- The model ignores the dynamic nature of the spread process, essentially treating the spread as i.i.d.
- Using more strict criterions may work for equities. In FX trading, we don’t have the luxury of throwing away many pairs.

Risks in Pairs Trading

- Long term equilibrium does not hold.
  - E.g., the company that you long goes bankrupt but the other leg does not move (one company wins over the other).
- Systematic market risk.
- Firm specific risk.
- Liquidity.

Cointegration

Stationarity

- These ad-hoc β calibration does not guarantee the single most important statistical property in trading: stationarity.
- Strong stationarity: the joint probability distribution of $\{x_t\}$ does not change over time.
- Weak stationarity: the first and second moments do not change over time.
  - Covariance stationarity

Mean Reversion

- A stationary stochastic process is mean-reverting.
- Long when the spread/portfolio/basket falls sufficiently below a long term equilibrium.
- Short when the spread/portfolio/basket rises sufficiently above a long term equilibrium.

Test for Stationarity

- An augmented Dickey–Fuller test (ADF) is a test for a unit root in a time series sample.
- It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models.
- Intuition:
  - if the series y_t is stationary, then it has a tendency to return to a constant mean. Therefore large values will tend to be followed by smaller values, and small values by larger values. Accordingly, the level of the series will be a significant predictor of next period's change, and will have a negative coefficient.
  - If, on the other hand, the series is integrated, then positive changes and negative changes will occur with probabilities that do not depend on the current level of the series.
  - In a random walk, where you are now does not affect which way you will go next.

ADF Math

- $\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \sum_{i=1}^{p-1} \Delta y_{t-i} + \epsilon_t$
- Null hypothesis $H_0: \gamma = 0$. (y_t non-stationary)
- $\alpha = 0, \beta = 0$ models a random walk.
- $\beta = 0$ models a random walk with drift.
- Test statistics = $\frac{\hat{\gamma}}{\sigma(\hat{\gamma})}$, the more negative, the more reason to reject H_0 (hence y_t stationary).
- SuanShu: AugmentedDickeyFuller.java

Cointegration

- Cointegration: select a linear combination of assets to construct an (approximately) stationary portfolio.

Objective

- Given two I(1) price series, we want to find a linear combination such that:
  - $z_t = x_t - \beta y_t = \mu + \varepsilon_t$
- ε_t is I(0), a stationary residue.
- μ is the long term equilibrium.
- Long when $z_t < \mu - \Delta$.
- Sell when $z_t > \mu + \Delta$.

Stocks from the Same Industry

- Reduce market risk, esp., in bear market.
  - Stocks from the same industry are likely to be subject to the same systematic risk.
- Give some theoretical unpinning to the pairs trading.
  - Stocks from the same industry are likely to be driven by the same fundamental factors (common trends).

Cointegration Definition

- $X_t \sim \mathrm{CI}(d, b)$ if
  - All components of X_t are integrated of same order d.
  - There exists a β_t such that the linear combination, $\beta_t X_t = \beta_1 X_{1t} + \beta_2 X_{2t} + \cdots + \beta_n X_{nt}$, is integrated of order $(d - b), b > 0$.
- β is the cointegrating vector, not unique.

Illustration for Trading

- Suppose we have two assets, both reasonably I(1), we want to find β such that
  - $Z = X + \beta Y$ is I(0), i.e., stationary.
- In this case, we have $d = 1, b = 1$.

A Simple VAR Example

- $y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$
- $z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt}$
- Theorem 4.2, Johansen, places certain restrictions on the coefficients for the VAR to be cointegrated.
  - The roots of the characteristics equation lie on or outside the unit disc.

Coefficient Restrictions

- $a_{11} = \frac{(1-a_{22})-a_{12}a_{21}}{1-a_{22}}$
- $a_{22} > -1$
- $a_{12}a_{21} + a_{22} < 1$

VECM (1)

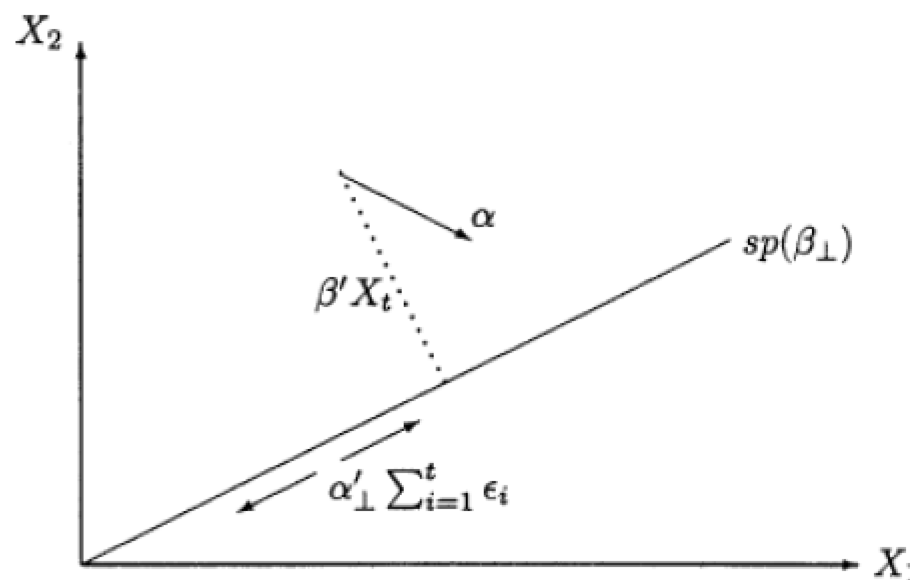
- Taking differences
  - $y_t - y_{t-1} = (a_{11} - 1)y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt}$
  - $z_t - z_{t-1} = a_{21}y_{t-1} + (a_{22} - 1)z_{t-1} + \varepsilon_{zt}$
- $\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$
- Substitution of a_{11}
  - $\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \frac{-a_{12}a_{21}}{1-a_{22}} & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$

VECM (2)

- $\Delta y_t = \alpha_y(y_{t-1} - \beta z_{t-1}) + \epsilon_{yt}$
- $\Delta z_t = \alpha_z(y_{t-1} - \beta z_{t-1}) + \epsilon_{zt}$
- $\alpha_y = \frac{-a_{12}a_{21}}{1-a_{22}}$
- $\alpha_z = a_{21}$
- $\beta = \frac{1-a_{22}}{a_{21}}$, the cointegrating coefficient
- $y_{t-1} - \beta z_{t-1}$ is the long run equilibrium, I(0).
- α_y, α_z are the speed of adjustment parameters.

Interpretation

- Suppose the long run equilibrium is 0,
  - $\Delta y_t, \Delta z_t$ responds only to shocks.
- Suppose $\alpha_y < 0$, $\alpha_z > 0$,
  - $\{y_t\}$ decreases in response to a +ve deviation.
  - $\{z_t\}$ increases in response to a +ve deviation.



Granger Representation Theorem

- If X_t is cointegrated, an VECM form exists.
- The increments can be expressed as a functions of the dis-equilibrium, and the lagged increments.
- $\Delta X_t = \alpha\beta' X_{t-1} + \sum c_t \Delta X_{t-1} + \varepsilon_t$
- In our simple example, we have
  - $\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \alpha_y \\ \alpha_z \end{bmatrix} \begin{bmatrix} 1 & -\beta \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$

Granger Causality

- $\{z_t\}$ does not Granger Cause $\{y_t\}$ if lagged values of $\{\Delta z_{t-i}\}$ do not enter the Δy_t equation.
- $\{y_t\}$ does not Granger Cause $\{z_t\}$ if lagged values of $\{\Delta y_{t-i}\}$ do not enter the Δz_t equation.

Engle-Granger Two Step Approach

- Estimate either
  - $y_t = \beta_{10} + \beta_{11} z_t + e_{1t}$
  - $z_t = \beta_{20} + \beta_{21} y_t + e_{2t}$
  - As the sample size increase indefinitely, asymptotically a test for a unit root in $\{e_{1t}\}$ and $\{e_{2t}\}$ are equivalent, but not for small sample sizes.
- Test for unit root using ADF on either $\{e_{1t}\}$ and $\{e_{2t}\}$.
- If $\{y_t\}$ and $\{z_t\}$ are cointegrated, $\{\beta\}$ super converges.

Engle-Granger Pros and Cons

- Pros:
  - simple
- Cons:
  - This approach is subject to twice the estimation errors. Any errors introduced in the first step carry over to the second step.
  - Work only for two I(1) time series.

Testing for Cointegration

- Note that in the VECM, the rows in the coefficient, Π, are **NOT** linearly independent.
  - $\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} \frac{-a_{12}a_{21}}{1-a_{22}} & a_{12} \\ a_{21} & a_{22}-1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix}$
- $\begin{bmatrix} \frac{-a_{12}a_{21}}{1-a_{22}} & a_{12} \end{bmatrix} \times \frac{-(1-a_{22})}{a_{12}} = \begin{bmatrix} a_{21} & a_{22}-1 \end{bmatrix}$
- The rank of Π determine whether the two assets $\{y_t\}$ and $\{z_t\}$ are cointegrated.

VAR & VECM

- In general, we can write convert a VAR to an VECM.
- VAR (from numerical estimation by, e.g., OLS):
  - $X_t = \sum_{i=1}^{p} A_i X_{t-i} + \varepsilon_t$
- Transitory form of VECM (reduced form)
  - $\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$
- Long run form of VECM
  - $\Delta X_t = \sum_{i=1}^{p-1} \Upsilon_i \Delta X_{t-i} + \Pi X_{t-p} + \varepsilon_t$

The Π Matrix

- Rank(Π) = n, full rank
  - The system is already stationary; a standard VAR model in levels.
- Rank(Π) = 0
  - There exists NO cointegrating relations among the time series.
- 0 < Rank(Π) < n
  - $\Pi = \alpha\beta'$
  - β is the cointegrating vector
  - α is the speed of adjustment.

Rank Determination

- Determining the rank of Π is amount to determining the number of non-zero eigenvalues of Π.
  - Π is usually obtained from (numerical VAR) estimation.
  - Eigenvalues are computed using a numerical procedure.

Trace Statistics

- Suppose the eigenvalues of Π are:$\lambda_1 > \lambda_2 > \cdots > \lambda_n$.
- For the o eigenvalues, $\ln(1 - \lambda_i) = 0$.
- For the (big) non-zero eigenvalues, $\ln(1 - \lambda_i)$ is (very negative).
- The likelihood ratio test statistics
  - $Q\big(H(r)|H(n)\big) = -T\sum_{i=r+1}^{p} \log(1 - \lambda_i)$
  - Ho: rank ≤ r; there are at most r cointegrating β.

Test Procedure

- int r = 0;//rank
- for (; r <= n; ++r) { // loop until the null is accepted
  - compute Q = $Q\left(H(r)|H(n)\right)$;
  - If (Q > c.v.) { // compare against a critical value
    - break; // fail to reject the null hypothesis; rank found
  - }
- }
- r is the rank found

Decomposing Π

- Suppose the rank of $\Pi = r$.
- $\Pi = \alpha\beta'$.
- Π is $n \times n$.
- α is $n \times r$.
- $\beta'$is r $\times$ n.

Estimating β

- β can estimated by maximizing the log-likelihood function in Chapter 6, Johansen.
  - $\log L(\Psi, \alpha, \beta, \Omega)$
- Theorem 6.1, Johansen: β is found by solving the following eigenvalue problem:
  - $\left|\lambda S_{11} - S_{10} {S_{00}}^{-1} S_{01}\right| = 0$

β

- Each non-zero eigenvalue λ corresponds to a cointegrating vector, which is its eigenvector.
- $\beta = (v_1, v_2, \cdots, v_r)$
- β spans the cointegrating space.
- For two cointegrating asset, there are only one β (v_1) so it is unequivocal.
- When there are multiple β, we need to add economic restrictions to identify β.

Trading the Pairs

- Given a space of (liquid) assets, we compute the pairwise cointegrating relationships.
- For each pair, we validate stationarity by performing the ADF test.
- For the strongly mean-reverting pairs, we can design trading strategies around them.

Problems with Using Cointegration

- The assets may be cointegrated sometimes but not always.
  - What do you do when it is not cointegrated but you are already in the market?
- Cointegration creates a dense basket – it includes every asset in the time series analyzed.
  - Incur huge transaction cost.
  - Reduce the significance of the structural relationships.
- Optimal mean reverting portfolios behave like noise and vary well inside the bid-ask spreads, hence not meaningful statistical arbitrage opportunities.
  - What about not so optimal ones?

Stochastic Spread

Ornstein–Uhlenbeck Process

- $z_t = x_t - \beta y_t$
- $dz_t = \theta(\mu - z_t)dt + \sigma dW_t$

Spread as a Mean-Reverting Process

- $x_k - x_{k-1} = (a - bx_{k-1})\tau + \sigma\sqrt{\tau}\varepsilon_k$
- $= b\left(\frac{a}{b} - x_{k-1}\right)\tau + \sigma\sqrt{\tau}\varepsilon_k$
- The long term mean = $\frac{a}{b}$.
- The rate of mean reversion = b.

Sum of Power Series

- We note that
  - $= \sum_{i=0}^{k-1} a^i = \frac{a^k - 1}{a - 1}$

Unconditional Mean

- $E(x_k) = \mu_k = \mu_{k-1} + (a - b\mu_{k-1})\tau$
- $= a\tau + (1 - b\tau)\mu_{k-1}$
- $= a\tau + (1 - b\tau)[a\tau + (1 - b\tau)\mu_{k-2}]$
- $= a\tau + (1 - b\tau)a\tau + (1 - b\tau)^2\mu_{k-2}$
- $= \sum_{i=0}^{k-1}(1 - b\tau)^i \, a\tau + (1 - b\tau)^k\mu_0$
- $= a\tau\,\frac{1-(1-b\tau)^k}{1-(1-b\tau)} + (1 - b\tau)^k\mu_0$
- $= a\tau\,\frac{1-(1-b\tau)^k}{b\tau} + (1 - b\tau)^k\mu_0$
- $= \frac{a}{b} - \frac{a}{b}(1 - b\tau)^k + (1 - b\tau)^k\mu_0$

Long Term Mean

- $\frac{a}{b} - \frac{a}{b}(1 - b\tau)^k + (1 - b\tau)^k \mu_0$
- $\to \frac{a}{b}$

Unconditional Variance

- $\mathrm{Var}(x_k) = {\sigma_k}^2 = (1 - b\tau)^2 {\sigma_{k-1}}^2 + \sigma^2\tau$
- $= (1 - b\tau)^2 {\sigma_{k-1}}^2 + \sigma^2\tau$
- $= (1 - b\tau)^2 [(1 - b\tau)^2 {\sigma_{k-2}}^2 + \sigma^2\tau] + \sigma^2\tau$
- $= \sigma^2\tau \sum_{i=0}^{k-1} (1 - b\tau)^{2i} + (1 - b\tau)^{2k} {\sigma_0}^2$
- $= \sigma^2\tau \frac{1-(1-b\tau)^{2k}}{1-(1-b\tau)^2} + (1 - b\tau)^{2k} {\sigma_0}^2$

Long Term Variance

- $\sigma^2\tau\frac{1-(1-b\tau)^{2k}}{1-(1-b\tau)^2}+(1-b\tau)^{2k}\,{\sigma_0}^2$
- $\rightarrow\frac{\sigma^2\tau}{1-(1-b\tau)^2}$

Observations and Hidden State Process

- The hidden state process is:
  - $x_k = x_{k-1} + (a - bx_{k-1})\tau + \sigma\sqrt{\tau}\varepsilon_k$
  - $= a\tau + (1 - b\tau)x_{k-1} + \sigma\sqrt{\tau}\varepsilon_k$
  - $= A + Bx_{k-1} + C\varepsilon_k$
  - $A \geq 0, 0 < B < 1$
- The observations:
  - $y_k = x_k + D\omega_k$
- We want to compute the *expected* state from observations.
  - $\hat{x}_k = \hat{x}_{k|k} = E[x_k|Y_k]$

Parameter Estimation

- We need to estimate the parameters $\vartheta = \{A, B, C, D\}$ from the observable data before we can use the Kalman filter model.
- We need to write down the likelihood function in terms of ϑ, and then maximize w.r.t. ϑ.

Likelihood Function

- A likelihood function (often simply the likelihood) is a function of the parameters of a statistical model, defined as follows: the likelihood of a set of parameter values given some observed outcomes is equal to the probability of those observed outcomes given those parameter values.
- $L(\vartheta; Y) = p(Y|\vartheta)$

Maximum Likelihood Estimate

- We find ϑ such that $L(\vartheta; Y)$ is maximized given the observation.

Example Using the Normal Distribution

- We want to estimate the mean of a sample of size N drawn from a Normal distribution.
  - $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$
- $\vartheta = \{\mu, \sigma\}$
- $L_N(\vartheta; Y) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i-\mu)^2}{2\sigma^2}\right\}$

Log-Likelihood

- $\log L_N(\vartheta; Y) = \sum_{i=1}^{N} \left\{ \log \frac{1}{\sqrt{2\pi\sigma^2}} - \frac{(y_i - \mu)^2}{2\sigma^2} \right\}$
- Maximizing the log-likelihood is equivalent to maximizing the following.
  - $-\sum_{i=1}^{N} \{(y_i - \mu)^2\}$
- First order condition w.r.t.,μ
  - $\mu = \frac{1}{N} \sum_{i=1}^{N} y_i$

Nelder-Mead

- After we write down the likelihood function for the Kalman model in terms of $\vartheta = \{A, B, C, D\}$, we can run any multivariate optimization algorithm, e.g., Nelder-Mead, to search for ϑ.
  - $\max_{\vartheta} L(\vartheta; Y)$
- The disadvantage is that it may not converge well, hence not landing close to the optimal solution.

Marginal Likelihood

- For the set of hidden states, $\{X_t\}$, we write
  - $L(\vartheta; Y) = p(Y|\vartheta) = \sum_X p(Y, X|\vartheta)$
- Assume we know the conditional distribution of X, we could instead maximize the following.
  - $\max_{\vartheta} \underset{X}{\mathrm{E}}[L(\vartheta|Y, X)]$, or
  - $\max_{\vartheta} \underset{X}{\mathrm{E}}[\log L(\vartheta|Y, X)]$
- The expectation is a weighted sum of the (log-) likelihoods weighted by the probability of the hidden states.

The Q-Function

- Where do we get the conditional distribution of $\{X_t\}$ from?
- Suppose we somehow have an (initial) estimation of the parameters, ϑ_0. Then the model has no unknowns. We can compute the distribution of $\{X_t\}$.
- $Q\left(\vartheta|\vartheta^{(t)}\right) = \underset{X|Y,\vartheta}{\mathrm{E}}[\log L(\vartheta|Y,X)]$

EM Intuition

- Suppose we know ϑ, we know completely about the mode; we can find X.
- Suppose we know X, we can estimate ϑ, by, e.g., maximum likelihood.
- What do we do if we don't know both ϑ and X?

Expectation-Maximization Algorithm

- Expectation step (E-step): compute the expected value of the log-likelihood function, w.r.t., the conditional distribution of X under Y and ϑ.
  - $Q\left(\vartheta|\vartheta^{(t)}\right) = \underset{X|Y,\vartheta}{\mathrm{E}}[\log L(\vartheta|Y,X)]$
- Maximization step (M-step): find the parameters, ϑ, that maximize the Q-value.
  - $\vartheta^{(t+1)} = \underset{\vartheta}{\operatorname{argmax}}\, Q\left(\vartheta|\vartheta^{(t)}\right)$

EM Algorithms for Kalman Filter

- Offline: Shumway and Stoffer smoother approach, 1982
- Online: Elliott and Krishnamurthy filter approach, 1999

First Passage Time

- Standardized Ornstein-Uhlenbeck process
  - $dZ(t) = -Z(t)dt + \sqrt{2}dW(t)$
- First passage time
  - $T_{0,c} = \inf\{t \geq 0, Z(t) = 0 | Z(0) = c\}$
- The pdf of $T_{0,c}$ has a maximum value at
  - $\hat{t} = \frac{1}{2}\ln\left[1 + \frac{1}{2}\left(\sqrt{(c^2-3)^2 + 4c^2} + c^2 - 3\right)\right]$

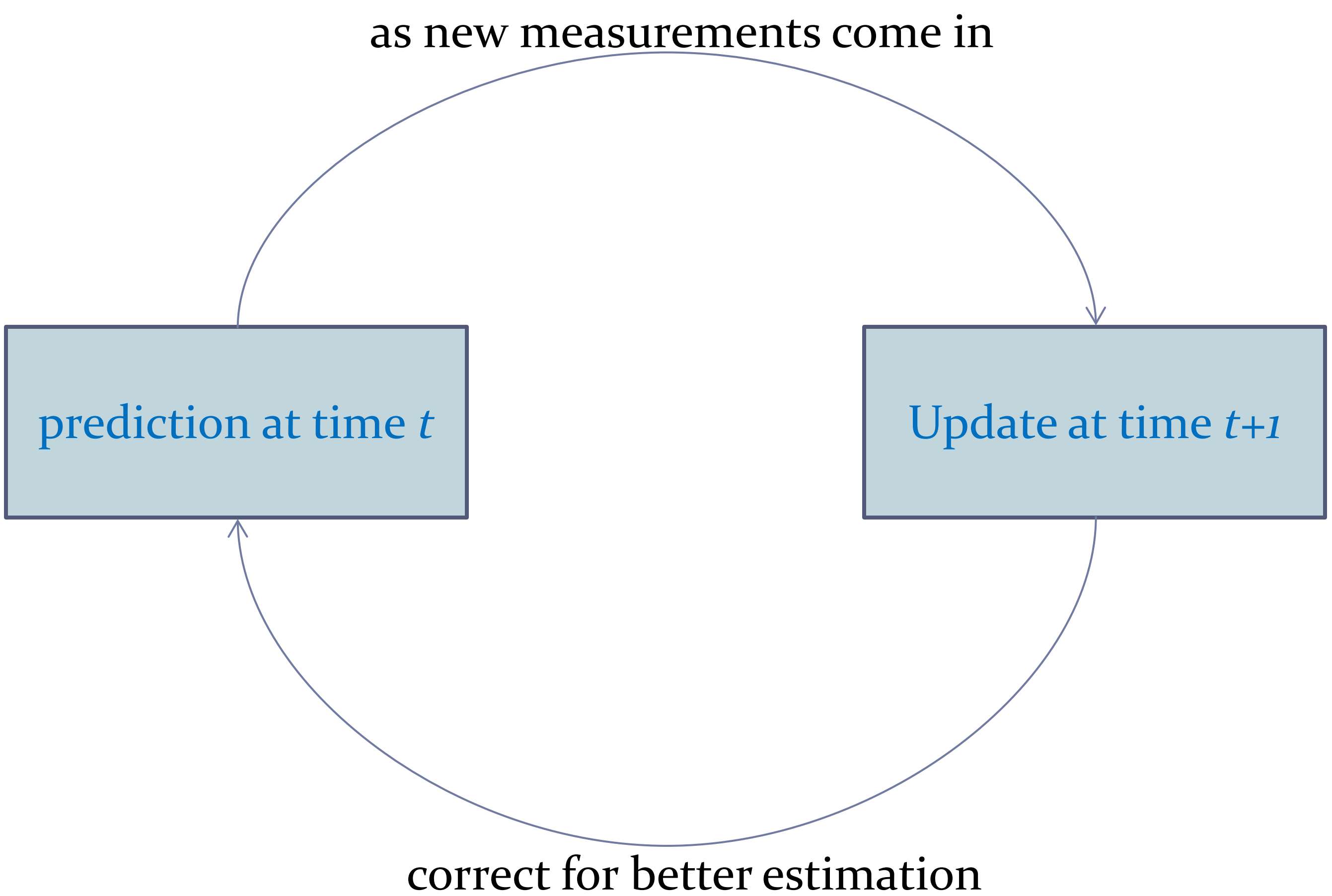
A Sample Trading Strategy

- $x_k = x_{k-1} + (a - b x_{k-1})\tau + \sigma\sqrt{\tau}\varepsilon_k$
- $dX(t) = \big(a - bX(t)\big)dt + \sigma dW(t)$
- $X(0) = \mu + c\frac{\sigma}{\sqrt{2\rho}}, X(T) = \mu$
- $T = \frac{1}{\rho}\hat{t}$
- Buy when $y_k < \mu - c\left(\frac{\sigma}{\sqrt{2\rho}}\right)$ unwind after time T
- Sell when $y_k > \mu + c\left(\frac{\sigma}{\sqrt{2\rho}}\right)$ unwind after time T

Kalman Filter

- The Kalman filter is an efficient recursive filter that estimates the state of a dynamic system from a series of incomplete and noisy measurements.

Conceptual Diagram



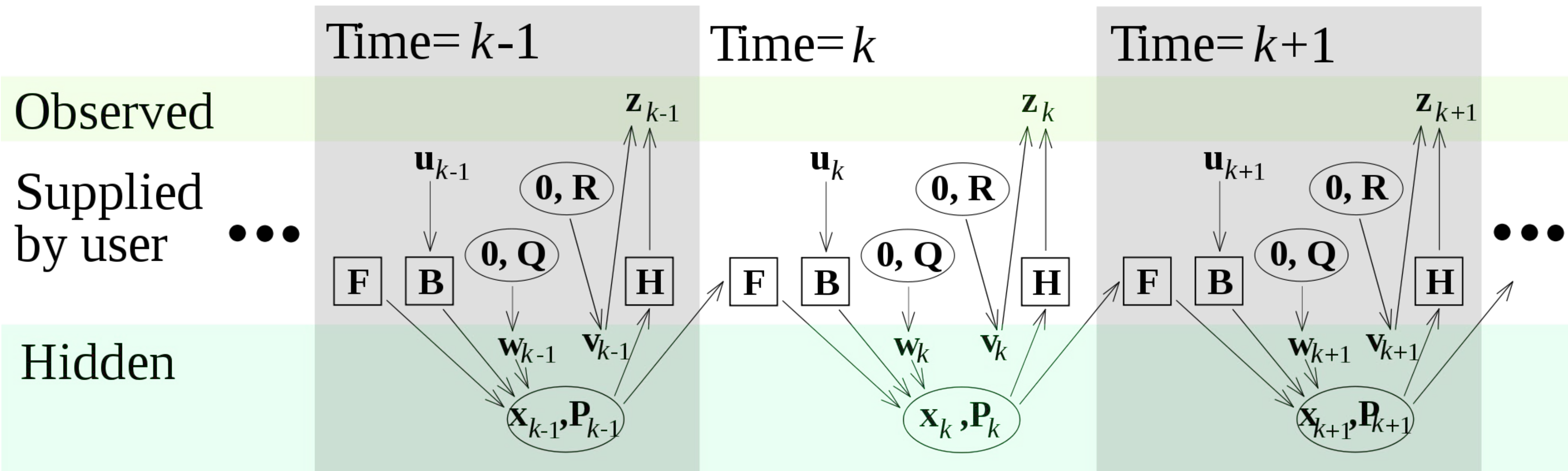
A Linear Discrete System

- $x_k = F_k x_{k-1} + B_k u_k + \omega_k$
- F_k: the state transition model applied to the previous state
- B_k: the control-input model applied to control vectors
- $\omega_k \sim N(0, Q_k)$: the noise process drawn from multivariate Normal distribution

Observations and Noises

- $z_k = H_k x_k + v_k$
- H_k: the observation model mapping the true states to observations
- $v_k \sim N(0, R_k)$: the observation noise

Discrete System Diagram



Prediction

- predicted a prior state estimate
  - $\hat{x}_{k|k-1} = F_k \hat{x}_{k-1|k-1} + B_k u_k$
- predicted a prior estimate covariance
  - $P_{k|k-1} = F_k P_{k-1|k-1} {F_k}^T + Q_k$

Update

- measurement residual
  - $\tilde{y}_k = z_k - H_k \hat{x}_{k|k-1}$
- residual covariance
  - $S_k = H_k P_{k|k-1} {H_k}^T + R_k$
- optimal Kalman gain
  - $K_k = P_{k|k-1} {H_k}^T {S_k}^{-1}$
- updated a posteriori state estimate
  - $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \tilde{y}_k$
- updated a posteriori estimate covariance
  - $P_{k|k} = (I - K_k H_k) P_{k|k-1}$

Computing the 'Best' State Estimate

- Given A, B, C, D, we define the conditional variance
  - $R_k = \Sigma_{k|k} \equiv \mathrm{E}[(x_k - \hat{x}_k)^2 | Y_k]$
- Start with $\hat{x}_{0|0} = y_0$, $R_0 = D^2$.

Predicted (a Priori) State Estimation

- $\hat{x}_{k+1|k}$
- $= \mathrm{E}[x_{k+1}|Y_k]$
- $= \mathrm{E}[A + Bx_k + C\varepsilon_{k+1}|Y_k]$
- $= \mathrm{E}[A + Bx_k|Y_k]$
- $= A + B\,\mathrm{E}[x_k|Y_k]$
- $= A + B\hat{x}_{k|k}$

Predicted (a Priori) Variance

- $\Sigma_{k+1|k}$
- $= \mathrm{E}[(x_{k+1} - \hat{x}_{k+1})^2 | Y_k]$
- $= \mathrm{E}[(A + Bx_k + C\varepsilon_{k+1} - \hat{x}_{k+1})^2 | Y_k]$
- $= \mathrm{E}\left[\left(A + Bx_k + C\varepsilon_{k+1} - A - B\hat{x}_{k|k}\right)^2 | Y_k\right]$
- $= \mathrm{E}\left[\left(Bx_k - B\hat{x}_{k|k} + C\varepsilon_{k+1}\right)^2 | Y_k\right]$
- $= \mathrm{E}\left[\left(Bx_k - B\hat{x}_{k|k}\right)^2 + C^2\varepsilon^2{}_{k+1} | Y_k\right]$
- $= B^2\Sigma_{k|k} + C^2$

Minimize Posteriori Variance

- Let the Kalman updating formula be
  - $\hat{x}_{k+1} = \hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K\left[y_{k+1} - \hat{x}_{k+1|k}\right]$
- We want to solve for K such that the conditional variance is minimized.
  - $\Sigma_{k+1|k} = \mathrm{E}[(x_{k+1} - \hat{x}_{k+1})^2|Y_k]$

Solve for K

- $\mathrm{E}[(x_{k+1} - \hat{x}_{k+1})^2 | Y_k]$
- $= \mathrm{E}\left[\left(x_{k+1} - \hat{x}_{k+1|k} - K\left[y_{k+1} - \hat{x}_{k+1|k}\right]\right)^2 | Y_k\right]$
- $= \mathrm{E}\left[\left(x_{k+1} - \hat{x}_{k+1|k} - K\left[x_{k+1} - \hat{x}_{k+1|k} + D\omega_{k+1}\right]\right)^2 | Y_k\right]$
- $= \mathrm{E}\left[\left[(1-K)\left(x_{k+1} - \hat{x}_{k+1|k}\right) - KD\omega_{k+1}\right]^2 | Y_k\right]$
- $= (1-K)^2\, \mathrm{E}\left[\left(x_{k+1} - \hat{x}_{k+1|k}\right)^2 | Y_k\right] + K^2 D^2$
- $= (1-K)^2\, \Sigma_{k+1|k} + K^2 D^2$

First Order Condition for k

- $\frac{d}{dK}(1-K)^2\,\Sigma_{k+1|k} + K^2D^2$
- $= \frac{d}{dK}(1-2K+K^2)\,\Sigma_{k+1|k} + K^2D^2$
- $= (-2+2K)\,\Sigma_{k+1|k} + 2KD^2$
- $= 0$

Optimal Kalman Filter

- $K_{k+1} = \frac{\Sigma_{k+1|k}}{\Sigma_{k+1|k}+D^2}$

Updated (a Posteriori) State Estimation

- So, we have the "optimal" Kalman updating rule.
  - $\hat{x}_{k+1} = \hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K\left[y_{k+1} - \hat{x}_{k+1|k}\right]$
  - $= \hat{x}_{k+1|k} + \frac{\Sigma_{k+1|k}}{\Sigma_{k+1|k}+D^2}\left[y_{k+1} - \hat{x}_{k+1|k}\right]$

Updated (a Posteriori) Variance

- $R_{k+1} = \Sigma_{k+1|k} = \mathrm{E}[(x_{k+1} - \hat{x}_{k+1})^2 | Y_{k+1}] = (1-K)^2 \, \Sigma_{k+1|k} + K^2 D^2$
- $= \left(1 - \frac{\Sigma_{k+1|k}}{\Sigma_{k+1|k} + D^2}\right)^2 \Sigma_{k+1|k} + \left(\frac{\Sigma_{k+1|k}}{\Sigma_{k+1|k} + D^2}\right)^2 D^2$
- $= \left(\frac{D^2}{\Sigma_{k+1|k} + D^2}\right)^2 \Sigma_{k+1|k} + \left(\frac{\Sigma_{k+1|k}}{\Sigma_{k+1|k} + D^2}\right)^2 D^2$
- $= \frac{D^4 \Sigma_{k+1|k} + D^2 {\Sigma_{k+1|k}}^2}{\left(\Sigma_{k+1|k} + D^2\right)^2}$
- $= \frac{D^4 \Sigma_{k+1|k} + D^2 {\Sigma_{k+1|k}}^2}{\left(\Sigma_{k+1|k} + D^2\right)^2}$
- $= \frac{\left(\Sigma_{k+1|k} D^2\right)\left(D^2 + \Sigma_{k+1|k} D^2\right)}{\left(\Sigma_{k+1|k} + D^2\right)^2}$
- $= \Sigma_{k+1|k} D^2$

A Trading Algorithm

- From $y_0, y_1, \ldots, y_N$, we estimate $\hat{\vartheta}(N)$.
- Decide whether to make a trade at $t = N$, unwind at $t = N + 1$, or some time later, e.g., $t = N + T$.
- As y_{N+1} arrives, estimate $\hat{\vartheta}(N + 1)$.
- Repeat.

Results (1)

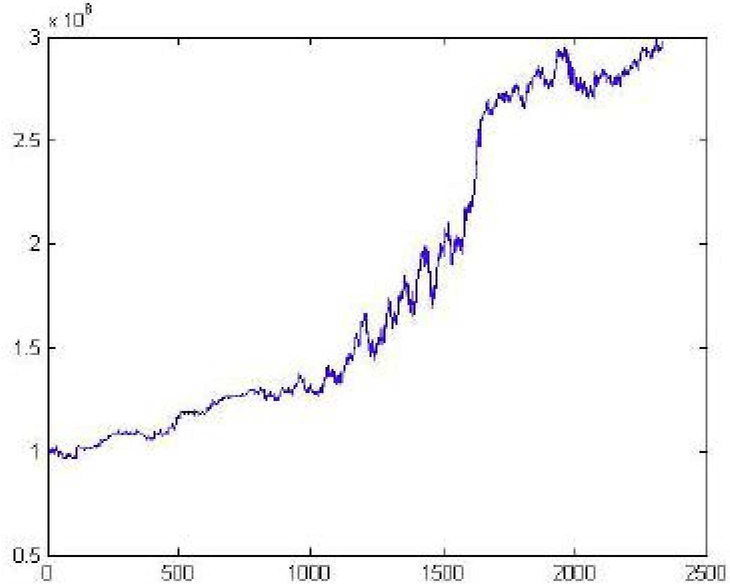


Figure 4: Backtesting Result with Optimization for weights on each pair

Results (2)

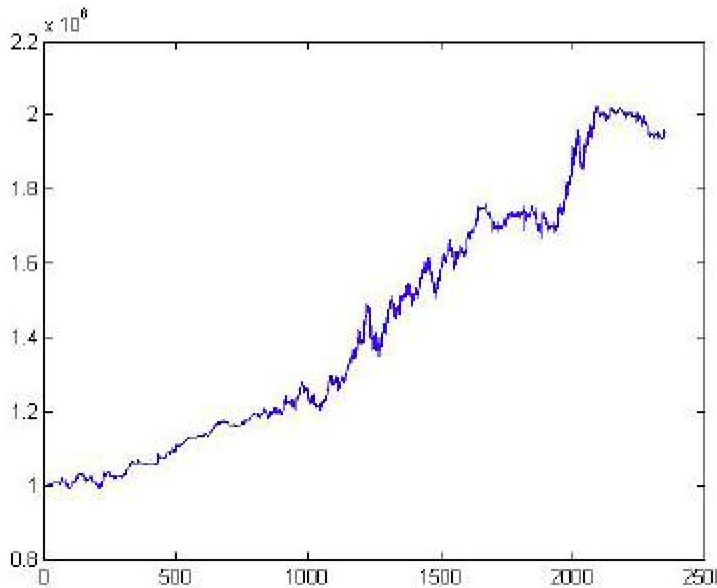


Figure 5: Backtesting Result Using Equal Weight Portfolio

Results (3)

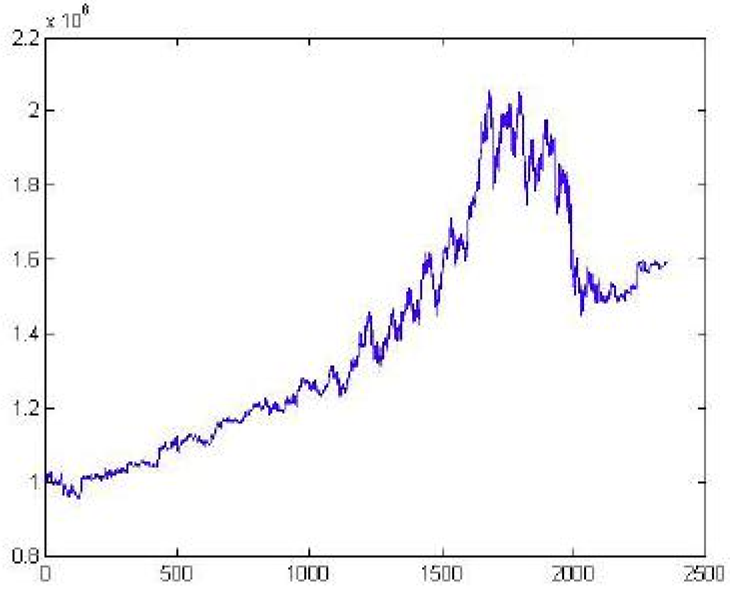


Figure 6: Backtesting Results Using 30-days Rolling Window Size

Optimal Trading Strategies

Stochastic Control

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

Speaker Profile

- Dr. Haksun Li
- CEO, [Numerical Method Inc.]
- (Ex-)Adjunct Professors, Industry Fellow, Advisor, Consultant with the National University of Singapore, Nanyang Technological University, Fudan University, the Hong Kong University of Science and Technology.
- Quantitative Trader/Analyst, BNPP, UBS
- PhD, Computer Sci, University of Michigan Ann Arbor
- M.S., Financial Mathematics, University of Chicago
- B.S., Mathematics, University of Chicago

References

- Optimal Pairs Trading: A Stochastic Control Approach. Mudchanatongsuk, S., Primbs, J.A., Wong, W. Dept. of Manage. Sci. & Eng., Stanford Univ., Stanford, CA. 2008.
- Optimal Trend Following Trading Rules. Dai, M., Zhang, Q., Zhu Q. J. 2011

Mean Reversion Trading

Stochastic Control

- We model the difference between the log-returns of two assets as an Ornstein-Uhlenbeck process.
- We compute the optimal position to take as a function of the deviation from the equilibrium.
- This is done by solving the corresponding the Hamilton-Jacobi-Bellman equation.

Formulation

- Assume a risk free asset M_t, which satisfies
  - $dM_t = rM_t dt$
- Assume two assets, A_t and B_t.
- Assume B_t follows a geometric Brownian motion.
  - $dB_t = \mu B_t dt + \sigma B_t dz_t$
- x_t is the spread between the two assets.
  - $x_t = \log A_t - \log B_t$

Ornstein-Uhlenbeck Process

- We assume the spread, the basket that we want to trade, follows a mean-reverting process.
  - $dx_t = k(\theta - x_t)dt + \eta d\omega_t$
- θ is the long term equilibrium to which the spread reverts.
- k is the rate of reversion. It must be positive to ensure stability around the equilibrium value.

Instantaneous Correlation

- Let ρ denote the instantaneous correlation coefficient between z and ω.
  - $E[d\omega_t dz_t] = \rho dt$

Univariate Ito's Lemma

- Assume
  - $dX_t = \mu_t dt + \sigma_t dB_t$
  - $f(t, X_t)$ is twice differentiable of two real variables
- We have
  - $df(t, X_t) = \left(\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{{\sigma_t}^2}{2} \frac{\partial^2 f}{\partial x^2}\right) dt + \sigma_t \frac{\partial f}{\partial x} dB_t$

Log example

- For G.B.M., $dX_t = \mu X_t dt + \sigma X_t dz_t$, $d\log X_t = ?$
- $f(x) = \log(x)$
- $\frac{\partial f}{\partial t} = 0$
- $\frac{\partial f}{\partial x} = \frac{1}{x}$
- $\frac{\partial^2 f}{\partial x^2} = -\frac{1}{x^2}$
- $d\log X_t = \left(\mu X_t \frac{1}{X_t} + \frac{(\sigma X_t)^2}{2}\left(-\frac{1}{{X_t}^2}\right)\right) dt + \sigma X_t \left(\frac{1}{X_t}\right) dB_t$
- $= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dB_t$

Multivariate Ito's Lemma

- Assume
  - $X_t = (X_{1t}, X_{2t}, \cdots, X_{nt})$ is a vector Ito process
  - $f(x_{1t}, x_{2t}, \cdots, x_{nt})$ is twice differentiable
- We have
  - $df(X_{1t}, X_{2t}, \cdots, X_{nt})$
  - $= \sum_{i=1}^{n} \frac{\partial}{\partial x_i} f(X_{1t}, X_{2t}, \cdots, X_{nt}) dX_i(t)$
  - $+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j} f(X_{1t}, X_{2t}, \cdots, X_{nt}) d[X_i, X_j](t)$

Multivariate Example

- $\log A_t = x_t + \log B_t$
- $A_t = \exp(x_t + \ \log B_t)$
- $\frac{\partial A_t}{\partial x_t} = \exp(x_t + \ \log B_t) = A_t$
- $\frac{\partial A_t}{\partial B_t} = \exp(x_t + \ \log B_t)\frac{1}{B_t} = \frac{A_t}{B_t}$
- $\frac{\partial^2 A_t}{\partial {x_t}^2} = \frac{\partial A_t}{\partial x_t} = A_t$
- $\frac{\partial^2 A_t}{\partial {B_t}^2} = \frac{\partial}{\partial B_t}\left(\frac{A_t}{B_t}\right) = 0$
- $\frac{\partial^2 A_t}{\partial B_t \partial x_t} = \frac{\partial}{\partial B_t}\left(\frac{\partial A_t}{\partial x_t}\right) = \frac{\partial A_t}{\partial B_t} = \frac{A_t}{B_t}$

What is the Dynamic of Asset A_t?

- $\partial A_t =$
$\frac{\partial A_t}{\partial x} dx_t + \frac{\partial A_t}{\partial B_t} dB_t + \frac{1}{2}\frac{\partial^2 A_t}{\partial x^2}(dx_t)^2 + \frac{\partial^2 A_t}{\partial B_t \partial x}(dx_t)(dB_t)$
- $= A_t dx_t + \frac{A_t}{B_t} dB_t + \frac{1}{2} A_t (dx_t)^2 + \frac{A_t}{B_t}(dx_t)(dB_t)$
- $= A_t[k(\theta - x_t)dt + \eta d\omega_t] + \frac{A_t}{B_t}[\mu B_t dt + \sigma B_t dz_t] +$
$\frac{1}{2} A_t \eta^2 dt + \frac{A_t}{B_t} \rho\eta\sigma B_t dt$

Dynamic of Asset A_t

- $\partial A_t = A_t[k(\theta - x_t)dt + \eta d\omega_t] + A_t[\mu dt + \sigma dz_t] + \frac{1}{2}A_t\eta^2 dt + A_t\rho\eta\sigma dt$
- $= A_t\left[k(\theta - x_t) + \mu + \frac{1}{2}\eta^2 + \rho\eta\sigma\right]dt + A_t\eta d\omega_t + A_t\sigma dz_t$
- $= A_t\left\{\left[\mu + k(\theta - x_t) + \frac{1}{2}\eta^2 + \rho\eta\sigma\right]dt + \sigma dz_t + \eta d\omega_t\right\}$

Notations

- V_t: the value of a self-financing pairs trading portfolio
- h_t:the portfolio weight for stock A
- $\widetilde{h_t} = -h_t$:the portfolio weight for stock B

Self-Financing Portfolio Dynamic

- $\frac{dV_t}{V_t} = h_t \frac{dA_t}{A_t} + \widetilde{h_t} \frac{dB_t}{B_t} + \frac{dM_t}{M_t}$
- $= h_t \left\{ \left[\mu + k(\theta - x_t) + \frac{1}{2}\eta^2 + \rho\eta\sigma \right] dt + \sigma dz_t + \eta d\omega_t \right\} - h_t \{ \mu dt + \sigma dz_t \} + r dt$
- $= h_t \left\{ \left[k(\theta - x_t) + \frac{1}{2}\eta^2 + \rho\eta\sigma \right] dt + \eta d\omega_t \right\} + r dt$
- $= \left\{ h_t \left[k(\theta - x_t) + \frac{1}{2}\eta^2 + \rho\eta\sigma \right] + r \right\} dt + h_t \eta d\omega_t$

Power Utility

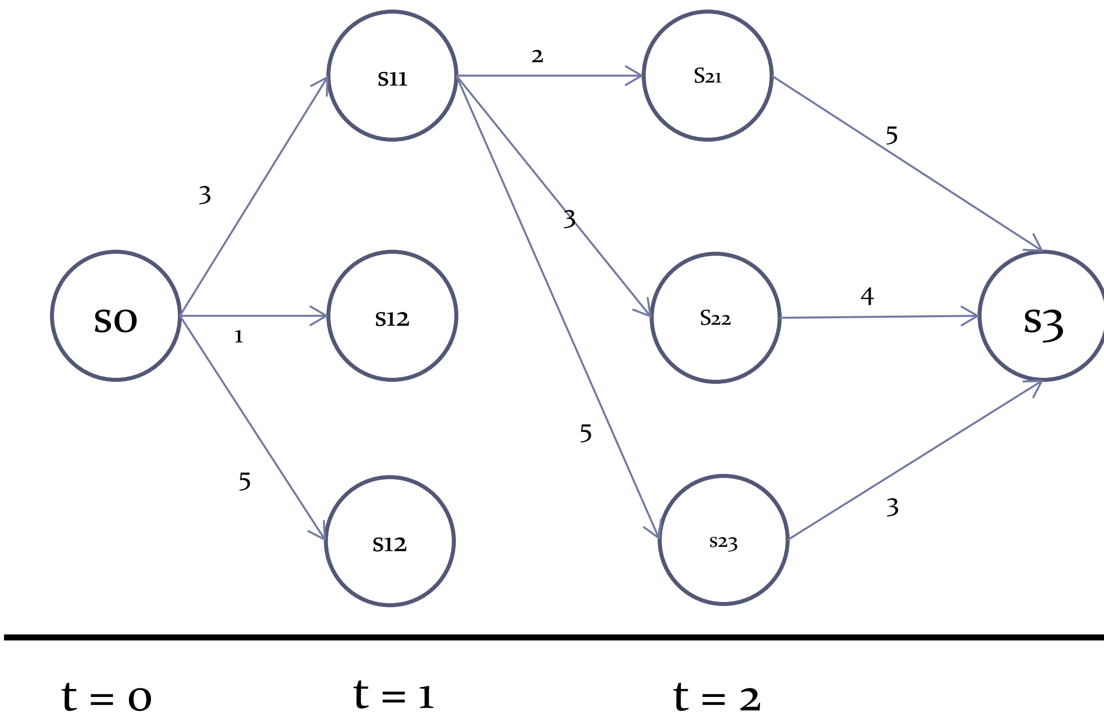
- Investor preference:
  - $U(x) = x^{\gamma}$
  - $x \geq 0$
  - $0 < \gamma < 1$

Problem Formulation

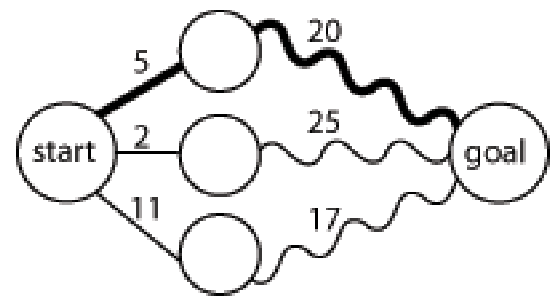
- $\max_{h_t} E[{V_T}^\gamma]$, s.t.,
- $V(0) = v_0$, $\mathrm{x}(0) = x_0$
  - $dx_t = k(\theta - x_t)dt + \eta d\omega_t$
- $dV_t = h_t dx_t = h_t k(\theta - x_t)dt + h_t \eta d\omega_t$
- Note that we simplify GBM to BM of V_t, and remove some constants.

Dynamic Programming

- Consider a stage problem to minimize (or maximize) the accumulated costs over a system path.



$$\text{Cost} = c_3 + \sum_{t=0}^{2} c_t$$



Dynamic Programming Formulation

- State change: $x_{k+1} = f_k(x_k, u_k, \omega_k)$
  - k: time
  - x_k: state
  - u_k: control decision selected at time k
  - ω_k: a random noise
- Cost: $g_N(x_N) + \sum_{k=0}^{N-1} g_k(x_k, u_k, \omega_k)$
- Objective: minimize (maximize) the expected cost.
  - We need to take expectation to account for the noise, ω_k.

Principle of Optimality

- Let $\pi^* = \{\mu_0{}^*, \mu_1{}^*, \cdots, \mu_{N-1}{}^*\}$ be an optimal policy for the basic problem, and assume that when using π^*, a give state x_i occurs at time i with positive probability. Consider the sub-problem whereby we are at x_i at time i and wish to minimize the "cost-to-go" from time i to time N.
  - $E\left\{g_N(x_N) + \sum_{k=i}^{N-1} g_k(x_k, u_k, \omega_k)\right\}$
- Then the truncated policy $\{\mu_i{}^*, \mu_{i+1}{}^*, \cdots, \mu_{N-1}{}^*\}$ is optimal for this sub-problem.

Dynamic Programming Algorithm

- For every initial state x_0, the optimal cost $J^*(x_k)$ of the basic problem is equal to $J_0(x_0)$, given by the last step of the following algorithm, which proceeds backward in time from period $N-1$ to period 0:
  - $J_N(x_N) = g_N(x_N)$
- $J_k(x_k) = \min_{u_k} E\{g_k(x_k, u_k, \omega_k) + J_{k+1}(f_k(x_k, u_k, \omega_k))\}$

Value function

- Terminal condition:
  - $G(T, V, x) = V^{\gamma}$
- DP equation:
  - $G(t, V_t, x_t) = \max_{h_t} E\{G(t + dt, V_{t+dt}, x_{t+dt})\}$
  - $G(t, V_t, x_t) = \max_{h_t} E\{G(t, V_t, x_t) + \Delta G\}$
- By Ito's lemma:
  - $\Delta G =$
    $G_t dt + G_V(dV) + G_x(dx) + \frac{1}{2} G_{VV}(dV)^2 + \frac{1}{2} G_{xx}(dx)^2 +$
    $G_{Vx}(dV)(dx)$

Hamilton-Jacobi-Bellman Equation

- Cancel $G(t, V_t, x_t)$ on both LHS and RHS.
- Divide by time discretization, Δt.
- Take limit as $\Delta t \to 0$, hence Ito.
- $0 = \max_{h_t} E\{\Delta G\}$
- $\max_{h_t} E\left\{G_t dt + G_V(dV) + G_x(dx) + \frac{1}{2}G_{VV}(dV)^2 + \frac{1}{2}G_{xx}(dx)^2 + G_{Vx}(dV)(dx)\right\} = 0$
- The optimal portfolio position is ${h_t}^*$.

HJB for Our Portfolio Value

- $\max_{h_t} E\left\{G_t dt + G_V(dV) + G_x(dx) + \frac{1}{2}G_{VV}(dV)^2 + \frac{1}{2}G_{xx}(dx)^2 + G_{Vx}(dV)(dx)\right\} = 0$

- $$\max_{h_t} E\left\{\begin{gathered} G_t dt + G_V(h_t k(\theta - x_t)dt + h_t \eta d\omega_t) + G_x(dx) + \\ \frac{1}{2}G_{VV}(h_t k(\theta - x_t)dt + h_t \eta d\omega_t)^2 + \frac{1}{2}G_{xx}(dx)^2 + \\ G_{Vx}(h_t k(\theta - x_t)dt + h_t \eta d\omega_t)(dx) \end{gathered}\right\} = 0$$

- $$\max_{h_t} E\left\{\begin{gathered} G_t dt + G_V(h_t k(\theta - x_t)dt + h_t \eta d\omega_t) + \\ G_x(k(\theta - x_t)dt + \eta d\omega_t) + \\ \frac{1}{2}G_{VV}(h_t k(\theta - x_t)dt + h_t \eta d\omega_t)^2 + \\ \frac{1}{2}G_{xx}(k(\theta - x_t)dt + \eta d\omega_t)^2 + \\ G_{Vx}(h_t k(\theta - x_t)dt + h_t \eta d\omega_t) \times (k(\theta - x_t)dt + \eta d\omega_t) \end{gathered}\right\} = 0$$

Taking Expectation

- All $\eta d\omega_t$ disapper because of the expectation operator.
- Only the deterministic dt terms remain.
- Divide LHR and RHS by dt.
- $\max_{h_t}\left\{\begin{array}{c} G_t + G_V(h_t k(\theta - x_t)) + \\ G_x\big(k(\theta - x_t)\big) + \frac{1}{2} G_{VV}(h_t\eta)^2 + \frac{1}{2} G_{xx}\eta^2 \\ +G_{Vx}(h_t\eta^2) \end{array}\right\} = 0$

Dynamic Programming Solution

- Solve for the cost-to-go function, G_t.
- Assume that the optimal policy is h_t^*.

First Order Condition

- Differentiate w.r.t. h_t.
- $G_V\left(k(\theta - x_t)\right) + {h_t}^* G_{VV}\eta^2 + G_{Vx}\eta^2 = 0$
- ${h_t}^* = -\frac{G_V\left(k(\theta - x_t)\right) + G_{Vx}\eta^2}{G_{VV}\eta^2}$
- In order to determine the optimal position, ${h_t}^*$, we need to solve for G to get G_V, G_{Vx}, and G_{VV}.

The Partial Differential Equation (1)

- $$G_t - G_V\left(\frac{G_V\big(k(\theta - x_t)\big) + G_{Vx}\eta^2}{G_{VV}\eta^2}k(\theta - x_t)\right) + G_x\big(k(\theta - x_t)\big) + \frac{1}{2}G_{VV}\eta^2\left(\frac{G_V\big(k(\theta - x_t)\big) + G_{Vx}\eta^2}{G_{VV}\eta^2}\right)^2 + \frac{1}{2}G_{xx}\eta^2 - G_{Vx}\eta^2\frac{G_V\big(k(\theta - x_t)\big) + G_{Vx}\eta^2}{G_{VV}\eta^2} = 0$$

The Partial Differential Equation (2)

$$
\begin{aligned}
&G_t - \\
&G_V k(\theta - x_t) \frac{G_V\left(k(\theta - x_t)\right) + G_{Vx}\eta^2}{G_{VV}\eta^2} + \\
&G_x\left(k(\theta - x_t)\right) + \\
&\frac{1}{2}\frac{\left[G_V\left(k(\theta - x_t)\right) + G_{Vx}\eta^2\right]^2}{G_{VV}\eta^2} + \\
&\frac{1}{2} G_{xx}\eta^2 - \\
&G_{Vx}\frac{G_V\left(k(\theta - x_t)\right) + G_{Vx}\eta^2}{G_{VV}}
\end{aligned}
= 0
$$

Dis-equilibrium

- Let $b = k(\theta - x_t)$. Rewrite:
- $G_t - G_V b \frac{G_V b + G_{Vx}\eta^2}{G_{VV}\eta^2} + G_x b + \frac{1}{2}\frac{\left[G_V b + G_{Vx}\eta^2\right]^2}{G_{VV}\eta^2} + \frac{1}{2} G_{xx}\eta^2 - G_{Vx}\frac{G_V b + G_{Vx}\eta^2}{G_{VV}} = 0$
- Multiply by $G_{VV}\eta^2$.
- $G_t G_{VV}\eta^2 - G_V b(G_V b + G_{Vx}\eta^2) + G_x b G_{VV}\eta^2 + \frac{1}{2}[G_V b + G_{Vx}\eta^2]^2 + \frac{1}{2} G_{xx} G_{VV}\eta^4 - G_{Vx}\eta^2[G_V b + G_{Vx}\eta^2] = 0$

Simplification

- Note that
- $-G_V b(G_V b + G_{Vx}\eta^2) + \frac{1}{2}[G_V b + G_{Vx}\eta^2]^2 - G_{Vx}\eta^2[G_V b + G_{Vx}\eta^2] = -\frac{1}{2}(G_V b + G_{Vx}\eta^2)^2$
- The PDE becomes
- $G_t G_{VV}\eta^2 + G_x b G_{VV}\eta^2 + \frac{1}{2}G_{xx}G_{VV}\eta^4 - \frac{1}{2}(G_V b + G_{Vx}\eta^2)^2 = 0$

The Partial Differential Equation (3)

- $G_t G_{VV} \eta^2 + G_x b G_{VV} \eta^2 + \frac{1}{2} G_{xx} G_{VV} \eta^4 - \frac{1}{2} (G_V b + G_{Vx} \eta^2)^2 = 0$

Ansatz for G

- $G(t, V, x) = f(t, x)V^\gamma$
- $G(T, V, x) = V^\gamma$
- $f(T, x) = 1$
- $G_t = V^\gamma f_t$
- $G_V = \gamma V^{\gamma-1} f$
- $G_{VV} = \gamma(\gamma - 1)V^{\gamma-2} f$
- $G_x = V^\gamma f_x$
- $G_{Vx} = \gamma V^{\gamma-1} f_x$
- $G_{xx} = V^\gamma f_{xx}$

Another PDE (1)

- $V^\gamma f_t \gamma(\gamma-1)V^{\gamma-2} f\eta^2 + V^\gamma f_x b\gamma(\gamma-1)V^{\gamma-2} f\eta^2 + \frac{1}{2}V^\gamma f_{xx}\gamma(\gamma-1)V^{\gamma-2} f\eta^4 - \frac{1}{2}(\gamma V^{\gamma-1} f b + \gamma V^{\gamma-1} f_x \eta^2)^2 = 0$
- Divide by $\gamma(\gamma-1)\eta^2 V^{2\gamma-2}$.
- $f_t f + f_x b f + \frac{1}{2} f_{xx} f \eta^2 - \frac{\gamma}{2(\gamma-1)}\left(f\frac{b}{\eta} + f_x \eta\right)^2 = 0$

Ansatz for f

- $f f_t + b f f_x + \frac{1}{2}\eta^2 f f_{xx} - \frac{\gamma}{2(\gamma-1)}\left(\frac{b}{\eta} f + \eta f_x\right)^2 = 0$
- $f(t,x) = g(t)\exp[x\beta(t) + x^2\alpha(t)] = g\exp(x\beta + x^2\alpha)$
- $f_t =$
  $g_t\exp(x\beta + x^2\alpha) + g\exp(x\beta + x^2\alpha)\,(x\beta_t + x^2\alpha_t)$
- $f_x = g\exp(x\beta + x^2\alpha)\,(\beta + 2x\alpha)$
- $f_{xx} =$
  $g\exp(x\beta + x^2\alpha)\,(\beta + 2x\alpha)^2 + g\exp(x\beta + x^2\alpha)\,(2\alpha)$
- $\frac{f_x}{f} = \beta + 2\alpha x$

Boundary Conditions

- $f(T,x) = g(T)\exp[x\beta(T) + x^2\alpha(T)] = 1$
- $g(T) = 1$
- $\alpha(T) = 0$
- $\beta(T) = 0$

Yet Another PDE (1)

- $g\exp(x\beta + x^2\alpha)\,[g_t\exp(x\beta + x^2\alpha) + g\exp(x\beta + x^2\alpha)\,(x\beta_t + x^2\alpha_t)] + bg\exp(x\beta + x^2\alpha)\,g\exp(x\beta + x^2\alpha)\,(\beta + 2x\alpha) + \frac{1}{2}\eta^2 g\exp(x\beta + x^2\alpha)\,[g\exp(x\beta + x^2\alpha)\,(\beta + 2x\alpha)^2 + g\exp(x\beta + x^2\alpha)\,(2\alpha)] - \frac{\gamma}{2(\gamma-1)}\left(\frac{b}{\eta}g\exp(x\beta + x^2\alpha) + \eta g\exp(x\beta + x^2\alpha)\,(\beta + 2x\alpha)\right)^2 = 0$
- Divide by g $\exp(x\beta + x^2\alpha)\exp(x\beta + x^2\alpha)$.
- $[g_t + g(x\beta_t + x^2\alpha_t)] + bg(\beta + 2x\alpha) + \frac{1}{2}\eta^2[g(\beta + 2x\alpha)^2 + g(2\alpha)] - \frac{\gamma}{2(\gamma-1)}g\left(\frac{b}{\eta} + \eta(\beta + 2x\alpha)\right)^2 = 0$
- $g_t + g(x\beta_t + x^2\alpha_t) + bg(\beta + 2x\alpha) + \frac{1}{2}\eta^2 g(\beta + 2x\alpha)^2 + \eta^2 g\alpha - \frac{\gamma}{2(\gamma-1)}g\left(\frac{b}{\eta} + \eta(\beta + 2x\alpha)\right)^2 = 0$

Yet Another PDE (2)

- $\lambda = \frac{\gamma}{2(\gamma-1)}$
- $g_t + g(x\beta_t + x^2\alpha_t) + bg(\beta + 2x\alpha) + \frac{1}{2}\eta^2 g(\beta + 2x\alpha)^2 + \eta^2 g\alpha - \lambda g\left(\frac{b}{\eta} + \eta(\beta + 2x\alpha)\right)^2 = 0$

Expansion in x

- $g_t + gx\beta_t + gx^2\alpha_t + bg\beta + 2x\alpha bg + \frac{1}{2}\eta^2 g(\beta^2 + 4x^2\alpha^2 + 4x\alpha\beta) + \eta^2 g\alpha - \lambda g\left(\frac{b^2}{\eta^2} + \eta^2\beta^2 + 4\eta^2 x^2\alpha^2 + 2b\beta + 4bx\alpha + 4\eta^2 x\alpha\beta\right) = 0$
- $g_t + gx\beta_t + gx^2\alpha_t + k(\theta - x)g\beta + 2x\alpha k(\theta - x)g + \frac{1}{2}\eta^2 g(\beta^2 + 4x^2\alpha^2 + 4x\alpha\beta) + \eta^2 g\alpha - \lambda g\left(\frac{k^2(\theta-x)^2}{\eta^2} + \eta^2\beta^2 + 4\eta^2 x^2\alpha^2 + 2k(\theta - x)\beta + 4k(\theta - x)x\alpha + 4\eta^2 x\alpha\beta\right) = 0$
- $g_t + gx\beta_t + gx^2\alpha_t + kg\beta\theta - kg\beta x + 2x\alpha kg\theta - 2\alpha kgx^2 + \frac{1}{2}\eta^2 g\beta^2 + 2\eta^2 gx^2\alpha^2 + 2\eta^2 gx\alpha\beta + \eta^2 g\alpha - \frac{\lambda g}{\eta^2}k^2\theta^2 + 2\frac{\lambda g}{\eta^2}k^2\theta x - \frac{\lambda g}{\eta^2}k^2x^2 - \lambda g\eta^2\beta^2 - 4\lambda g\eta^2 x^2\alpha^2 - 2\lambda gk\beta\theta + 2\lambda gk\beta x - 4\lambda gk\theta x\alpha + 4\lambda gkx^2\alpha - 4\lambda g\eta^2 x\alpha\beta = 0$

Grouping in x

- $\left(g_t + kg\beta\theta + \frac{1}{2}\eta^2 g\beta^2 + \eta^2 g\alpha - \frac{\lambda g}{\eta^2}k^2\theta^2 - \lambda g\eta^2\beta^2 - 2\lambda gk\beta\theta\right) +$
$\left(g\beta_t - kg\beta + 2\alpha kg\theta + 2\eta^2 g\alpha\beta + 2\frac{\lambda g}{\eta^2}k^2\theta + 2\lambda gk\beta - 4\lambda gk\theta\alpha - 4\lambda g\eta^2\alpha\beta\right)x +$
$\left(g\alpha_t - 2\alpha kg + 2\eta^2 g\alpha^2 - \frac{\lambda g}{\eta^2}k^2 - 4\lambda g\eta^2\alpha^2 + 4\lambda gk\alpha\right)x^2 = 0$

The Three PDE's (1)

- $g\alpha_t - 2\alpha k g + 2\eta^2 g\alpha^2 - \frac{\lambda g}{\eta^2} k^2 - 4\lambda g\eta^2\alpha^2 + 4\lambda g k\alpha = 0$
- $g\beta_t - kg\beta + 2\alpha k g\theta + 2\eta^2 g\alpha\beta + 2\frac{\lambda g}{\eta^2} k^2\theta + 2\lambda g k\beta - 4\lambda g k\theta\alpha - 4\lambda g\eta^2\alpha\beta = 0$
- $g_t + kg\beta\theta + \frac{1}{2}\eta^2 g\beta^2 + \eta^2 g\alpha - \frac{\lambda g}{\eta^2} k^2\theta^2 - \lambda g\eta^2\beta^2 - 2\lambda g k\beta\theta = 0$

PDE in α

- $\alpha_t + (2\eta^2 - 4\lambda\eta^2)\alpha^2 + (4\lambda k - 2k)\alpha - \frac{\lambda}{\eta^2}k^2 = 0$
- $\alpha_t = \frac{\lambda}{\eta^2}k^2 + 2k(1 - 2\lambda)\alpha + 2\eta^2(2\lambda - 1)\alpha^2$

PDE in β, α

- $\beta_t - k\beta + 2\eta^2\alpha\beta + 2\lambda k\beta - 4\lambda\eta^2\alpha\beta - 4\lambda k\theta\alpha + 2\frac{\lambda}{\eta^2}k^2\theta + 2\alpha k\theta = 0$
- $\beta_t = (k - 2\eta^2\alpha - 2\lambda k + 4\lambda\eta^2\alpha)\beta + \left(4\lambda k\theta\alpha - 2\frac{\lambda}{\eta^2}k^2\theta - 2\alpha k\theta\right)$

PDE in β, α, g

- $g_t + kg\beta\theta + \frac{1}{2}\eta^2 g\beta^2 + \eta^2 g\alpha - \frac{\lambda g}{\eta^2}k^2\theta^2 - \lambda g\eta^2\beta^2 - 2\lambda gk\beta\theta = 0$
- $g_t = -kg\beta\theta - \frac{1}{2}\eta^2 g\beta^2 - \eta^2 g\alpha + \frac{\lambda g}{\eta^2}k^2\theta^2 + \lambda g\eta^2\beta^2 + 2\lambda gk\beta\theta$
- $g_t = g\left(-k\beta\theta - \frac{1}{2}\eta^2\beta^2 - \eta^2\alpha + \frac{\lambda}{\eta^2}k^2\theta^2 + \lambda\eta^2\beta^2 + 2\lambda k\beta\theta\right)$

Riccati Equation

- A Riccati equation is any ordinary differential equation that is quadratic in the unknown function.
- $\alpha_t = \frac{\lambda}{\eta^2} k^2 + 2k(1-2\lambda)\alpha + 2\eta^2(2\lambda - 1)\alpha^2$
- $\alpha_t = A_0 + A_1\alpha + A_2\alpha^2$

Solving a Riccati Equation by Integration

- Suppose a particular solution, α_1, can be found.
- $\alpha = \alpha_1 + \frac{1}{z}$ is the general solution, subject to some boundary condition.

Particular Solution

- Either α_1 or α_2 is a particular solution to the ODE. This can be verified by mere substitution.
- $\alpha_{1,2} = \dfrac{-A_1 \pm \sqrt{{A_1}^2 - 4A_2A_0}}{2A_2}$

z Substitution

- Suppose $\alpha = \alpha_1 + \frac{1}{z}$.
- $\left(\dot{\frac{1}{z}}\right) = A_0 + A_1\left(\alpha_1 + \frac{1}{z}\right) + A_2\left(\alpha_1 + \frac{1}{z}\right)^2$
- $= A_0 + \left(A_1\alpha_1 + A_1\frac{1}{z}\right) + \left(A_2{\alpha_1}^2 + A_2\frac{1}{z^2} + 2A_2\frac{\alpha_1}{z}\right)$
- $= A_0 + \left(A_1\alpha_1 + A_1\frac{1}{z}\right) + \left(A_2{\alpha_1}^2 + A_2\frac{1}{z^2} + 2A_2\frac{\alpha_1}{z}\right)$
- $= \left(A_0 + A_1\alpha_1 + A_2{\alpha_1}^2\right) + \left(\frac{A_1 + 2\alpha_1 A_2}{z}\right) + \frac{A_2}{z^2}$

goes to o by the definition of α_1

Solving z

- $\dot{\left(\frac{1}{z}\right)} = \left(\frac{A_1+2\alpha_1 A_2}{z}\right) + \frac{A_2}{z^2}$
- $-\frac{1}{z^2}\dot{z} = \left(\frac{A_1+2\alpha_1 A_2}{z}\right) + \frac{A_2}{z^2}$
- 1^{st} order linear ODE
  - $\dot{z} + (A_1 + 2\alpha_1 A_2)z = -A_2$
- $z(t) = \frac{-A_2}{A_1+2\alpha_1 A_2} + C\exp(-(A_1 + 2\alpha_1 A_2)t)$

Solving for α

- $\alpha = \alpha_1 + \frac{1}{\frac{-A_2}{A_1+2\alpha_1 A_2} + c\exp(-(A_1+2\alpha_1 A_2)t)}$
- boundary condition:
  - $\alpha(T) = 0$
- $\alpha_1 + \frac{1}{\frac{-A_2}{A_1+2\alpha_1 A_2} + c\exp(-(A_1+2\alpha_1 A_2)T)} = 0$
- $C\exp(-(A_1 + 2\alpha_1 A_2)T) = -\frac{1}{\alpha_1} + \frac{A_2}{A_1+2\alpha_1 A_2}$
- $C = \exp\left((A_1 + 2\alpha_1 A_2)T\right)\left[\frac{A_2}{A_1+2\alpha_1 A_2} - \frac{1}{\alpha_1}\right]$

α Solution (1)

- $$\alpha = \alpha_1 + \frac{1}{\frac{-A_2}{A_1+2\alpha_1 A_2} + c\exp(-(A_1+2\alpha_1 A_2)t)}$$
- $$= \alpha_1 + \frac{1}{\frac{-A_2}{A_1+2\alpha_1 A_2} + \exp\left((A_1+2\alpha_1 A_2)T\right)\left[\frac{A_2}{A_1+2\alpha_1 A_2} - \frac{1}{\alpha_1}\right]\exp(-(A_1+2\alpha_1 A_2)t)}$$
- $$= \alpha_1 + \frac{1}{\frac{-A_2}{A_1+2\alpha_1 A_2} + \exp\left((A_1+2\alpha_1 A_2)(T-t)\right)\left[\frac{A_2}{A_1+2\alpha_1 A_2} - \frac{1}{\alpha_1}\right]}$$
- $$= \alpha_1 + \frac{\alpha_1(A_1+2\alpha_1 A_2)}{-\alpha_1 A_2 + \exp\left((A_1+2\alpha_1 A_2)(T-t)\right)(\alpha_1 A_2 - A_1 - 2\alpha_1 A_2)}$$

α Solution (2)

- $\alpha = \alpha_1 + \frac{\alpha_1(A_1+2\alpha_1 A_2)}{-\alpha_1 A_2+\exp\left((A_1+2\alpha_1 A_2)(T-t)\right)(-A_1-\alpha_1 A_2)}$
- $= \alpha_1 \left[1 - \frac{A_1+2\alpha_1 A_2}{A_2+\exp\left((A_1+2\alpha_1 A_2)(T-t)\right)\left(\frac{A_1}{\alpha_1}+A_2\right)}\right]$
- $= \alpha_1 \left[1 - \frac{\frac{A_1}{A_2}+2\alpha_1}{1+\left(\frac{A_1}{\alpha_1 A_2}+1\right)\exp\left((A_1+2\alpha_1 A_2)(T-t)\right)}\right]$

α Solution (3)

- $$\alpha(t) = \frac{k}{2\eta^2}\left[\left(1 - \sqrt{1-\gamma}\right) + \frac{2\sqrt{1-\gamma}}{1+\left(1-\frac{2}{1-\sqrt{1-\gamma}}\right)\exp\left(\frac{2k}{\sqrt{1-\gamma}}(T-t)\right)}\right]$$

Solving β

- $\beta_t = (k - 2\eta^2\alpha - 2\lambda k + 4\lambda\eta^2\alpha)\beta + \left(4\lambda k\theta\alpha - 2\frac{\lambda}{\eta^2}k^2\theta - 2\alpha k\theta\right)$
- Let $\tau = T - t$
- $\hat{\beta}(\tau) = \beta(T - t)$
- $\hat{\beta}_\tau(\tau) = -\beta_t(\tau)$
- $-\hat{\beta}_\tau(\tau) = \beta_t(\tau) = \left(k - 2\eta^2\alpha(\tau) - 2\lambda k + 4\lambda\eta^2\alpha(\tau)\right)\beta(\tau) + \left(4\lambda k\theta\alpha(\tau) - 2\frac{\lambda}{\eta^2}k^2\theta - 2\alpha(\tau)k\theta\right)$
- $\hat{\beta}_\tau(\tau) = (-k + 2\eta^2\hat{\alpha} + 2\lambda k - 4\lambda\eta^2\hat{\alpha})\hat{\beta} + \left(-4\lambda k\theta\hat{\alpha} + 2\frac{\lambda}{\eta^2}k^2\theta + 2\hat{\alpha}k\theta\right)$
- $\hat{\beta}_\tau(\tau) = \left((2\lambda - 1)k + 2\eta^2\hat{\alpha}(1 - 2\lambda)\right)\hat{\beta} + \left(2\hat{\alpha}k\theta(1 - 2\lambda) + 2\frac{\lambda}{\eta^2}k^2\theta\right)$

First Order Non-Constant Coefficients

- $\hat{\beta}_\tau = B_1\hat{\beta} + B_2$
  - $B_1(\tau) = (2\lambda - 1)k + 2\eta^2\hat{\alpha}(1 - 2\lambda)$
  - $B_2(\tau) = 2\hat{\alpha}k\theta(1 - 2\lambda) + 2\frac{\lambda}{\eta^2}k^2\theta$

Integrating Factor (1)

- $\hat{\beta}_\tau - B_1\hat{\beta} = B_2$
- We try to find an integrating factor $\mu = \mu(\tau)$ s.t.
  - $\frac{d}{d\tau}(\mu\hat{\beta}) = \mu\frac{d\hat{\beta}}{d\tau} + \hat{\beta}\frac{d\mu}{d\tau} = \mu B_2$
- Divide LHS and RHS by $\mu\hat{\beta}$.
  - $\frac{1}{\hat{\beta}}\frac{d\hat{\beta}}{d\tau} + \frac{1}{\mu}\frac{d\mu}{d\tau} = \frac{B_2}{\hat{\beta}}$
- By comparison,
  - $-B_1 = \frac{1}{\mu}\frac{d\mu}{d\tau}$

Integrating Factor (2)

- $\int -B_1 d\tau = \int \frac{d\mu}{\mu} = \log \mu + C$
- $\mu = \exp\left(\int -B_1 d\tau\right)$
- $\mu\hat{\beta} = \int \mu B_2 d\tau + C$
- $\hat{\beta} = \frac{\int \mu B_2 d\tau + C}{\mu}$
- $\hat{\beta} = \frac{\int \exp\left(\int -B_1 du\right) B_2 d\tau + C}{\exp\left(\int -B_1 d\tau\right)}$

$\hat{\beta}$ Solution

- $\hat{\beta} = \frac{\int_0^\tau \exp\left(\int_0^\tau -B_1(u)du\right)B_2(s)ds}{\exp\left(\int_0^\tau -B_1(u)du\right)} + C$
- $\hat{\beta}(\tau) =$
  $\exp\left(\int_0^\tau B_1(u)du\right) \int_0^\tau \left[\exp\left(\int_0^s -B_1(u)du\right) B_2(s)\right]ds + C$

B_1, B_2

- $\int_0^\tau B_1(s)ds = \int_0^\tau [(2\lambda - 1)k + 2\eta^2(1 - 2\lambda)\alpha(s)]ds$
- $I_2 = \int_0^\tau \left[\exp\left(\int_0^s -B_1(u)du\right) B_2(s)\right] ds$

β Solution

- $$\beta(t) = \frac{k\theta}{\eta^2}\left(1+\sqrt{1-\gamma}\right)\frac{\exp\left(\frac{2k}{\sqrt{1-\gamma}}(T-t)\right)-1}{1+\left[1-\frac{2}{1-\sqrt{1-\gamma}}\right]\exp\left(\frac{2k}{\sqrt{1-\gamma}}(T-t)\right)}$$

Solving g

- $g_t = g\left(-k\beta\theta - \frac{1}{2}\eta^2\beta^2 - \eta^2\alpha + \frac{\lambda}{\eta^2}k^2\theta^2 + \lambda\eta^2\beta^2 + 2\lambda k\beta\theta\right)$
- Withα and β solved, we are now ready to solve g_t.
- $\frac{g_t}{g} = G(t)$
- $\frac{d}{ds}(\log g_t) = \frac{g_t}{g} = G$
- $\log g_t = \int G ds + C$
- $g_t = C\exp\left(\int Gdt\right) = \exp\left(-\int_0^T Gds\right)\exp\left(\int_0^t Gds\right)$
- $g_t = \exp\left(-\int_t^T Gds\right)$

Computing the Optimal Position

- $h(t)^* = -\frac{G_V\left(k(\theta - x_t)\right) + G_{Vx}\eta^2}{G_{VV}\eta^2}$
- $= -\frac{\gamma V^{\gamma-1} f\left(k(\theta - x_t)\right) + \gamma V^{\gamma-1} f_x \eta^2}{\gamma(\gamma-1)V^{\gamma-2} f \eta^2}$
- $= -\frac{V f\left(k(\theta - x_t)\right) + V f_x \eta^2}{(\gamma-1) f \eta^2}$
- $= -\frac{V}{(\gamma-1)\eta^2} \frac{f\left(k(\theta - x_t)\right) + f_x \eta^2}{f}$
- $= -\frac{V}{(\gamma-1)\eta^2}\left[k(\theta - x_t) + \frac{f_x}{f}\eta^2\right]$
- $= \frac{V}{(1-\gamma)\eta^2}[k(\theta - x_t) + \eta^2(\beta + 2\alpha x)]$
- $= \frac{V}{(1-\gamma)}\left[-\frac{k}{\eta^2}(x_t - \theta) + 2\alpha x + \beta\right]$

The Optimal Position

- $h(t)^* = \frac{V_t}{(1-\gamma)}\left[-\frac{k}{\eta^2}(x_t - \theta) + 2\alpha(t)x_t + \beta(t)\right]$
- $h(t)^* \sim -\frac{k}{\eta^2}(x_t - \theta)$

P&L for Simulated Data

- The portfolio increases from $1000 to $4625 in one year.

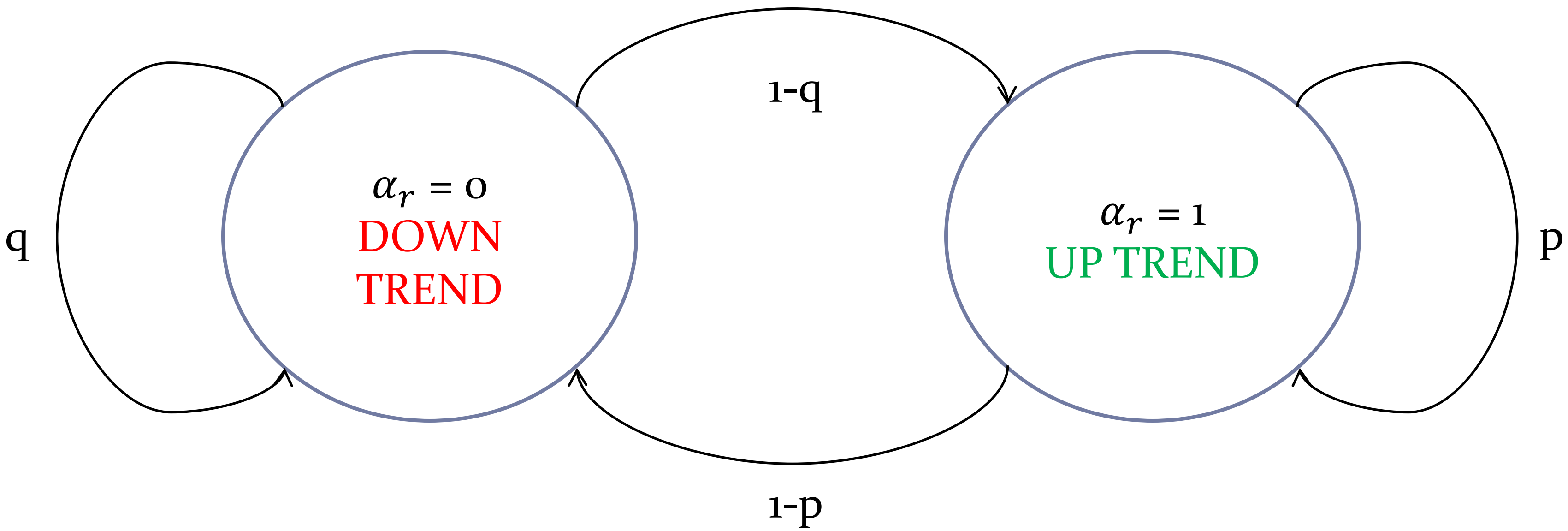
Parameter Estimation

- Can be done using Maximum Likelihood.
- Evaluation of parameter sensitivity can be done by Monte Carlo simulation.
- In real trading, it is better to be conservative about the parameters.
  - Better underestimate the mean-reverting speed
  - Better overestimate the noise
- To account for parameter regime changes, we can use:
  - a hidden Markov chain model
  - moving calibration window

Trend Following Trading

Two-State Markov Model

- $dS_r = S_r[\mu(\alpha_r)dr + \sigma dB_r]$

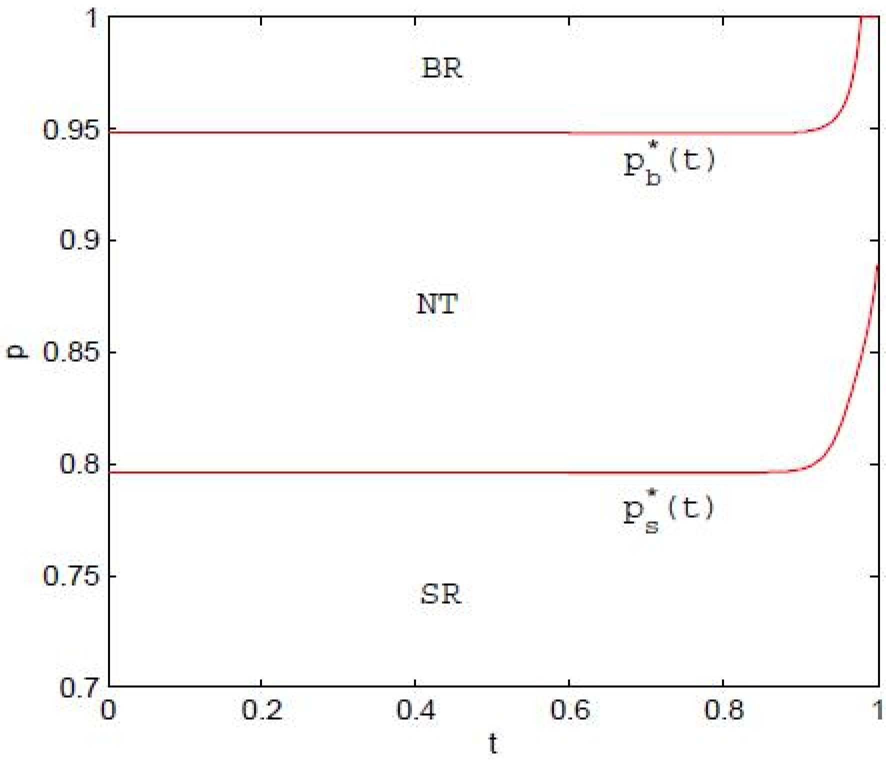


Buying and Selling Decisions

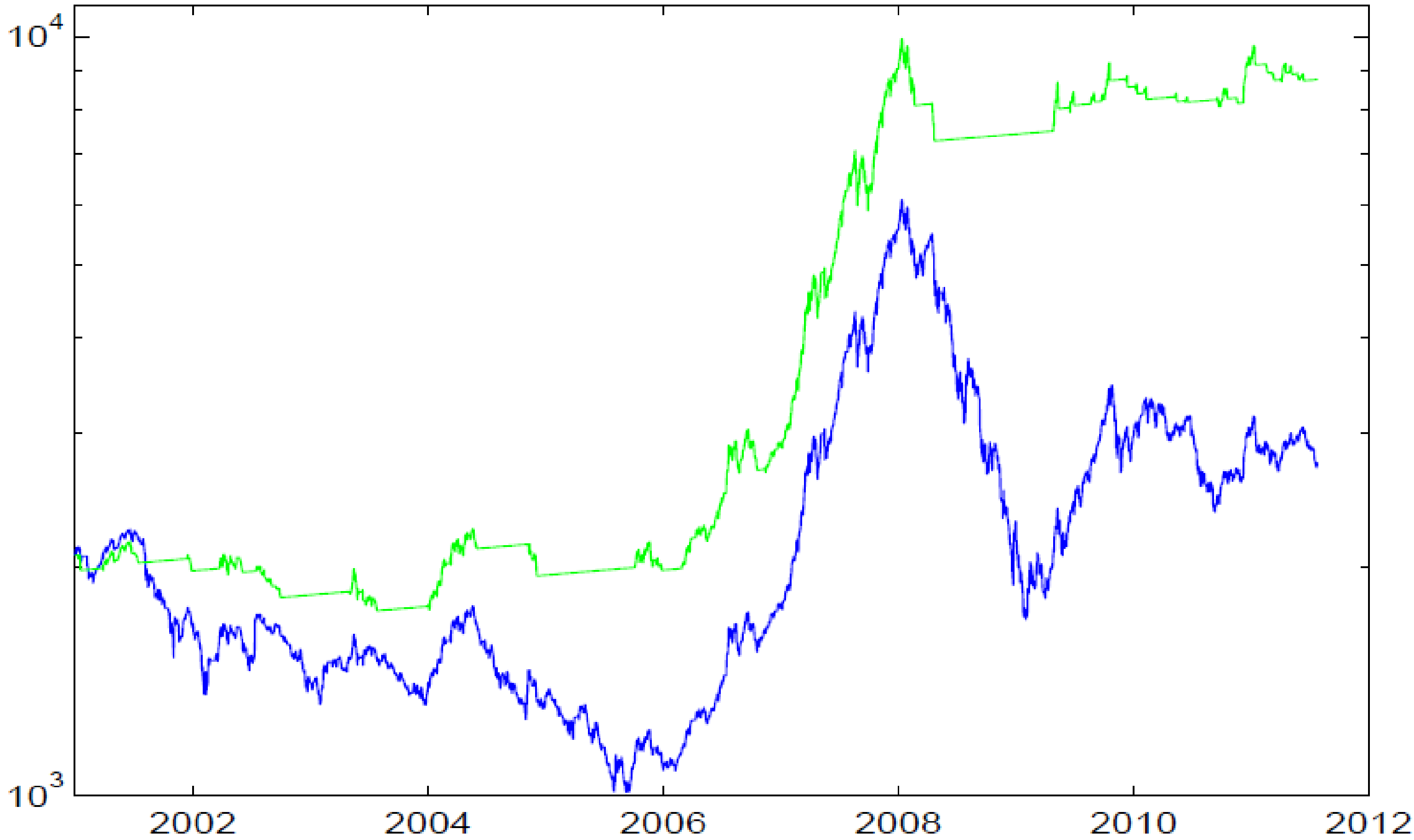
- $t \le \tau_1^0 \le \nu_1^0 \le \tau_2^0 \le \nu_2^0 \le \cdots \le \tau_n^0 \le \nu_n^0 \le \cdots \le T$

$$J_i(S,\alpha,t,\Lambda_i) = \begin{cases} E_t\left\{\log\left(e^{\rho(\tau_1-t)}\prod_{n=1}^{\infty} e^{\rho(\tau_{n+1}-v_n)}\frac{S_{v_n}}{S_{\tau_n}}\left[\frac{1-K_s}{1+K_b}\right]^{I_{\{\tau_n<T\}}}\right)\right\}, & \text{if } i=0, \\ E_t\left\{\log\left(\left[\frac{S_{v_1}}{S}e^{\rho(\tau_2-v_1)}(1-K_s)\right]\prod_{n=2}^{\infty} e^{\rho(\tau_{n+1}-v_n)}\frac{S_{v_n}}{S_{\tau_n}}\left[\frac{1-K_s}{1+K_b}\right]^{I_{\{\tau_n<T\}}}\right)\right\}, & \text{if } i=1. \end{cases}$$

Optimal Trend Following Strategy



Trading SSE 2001 – 2011



NUMERICAL METHOD

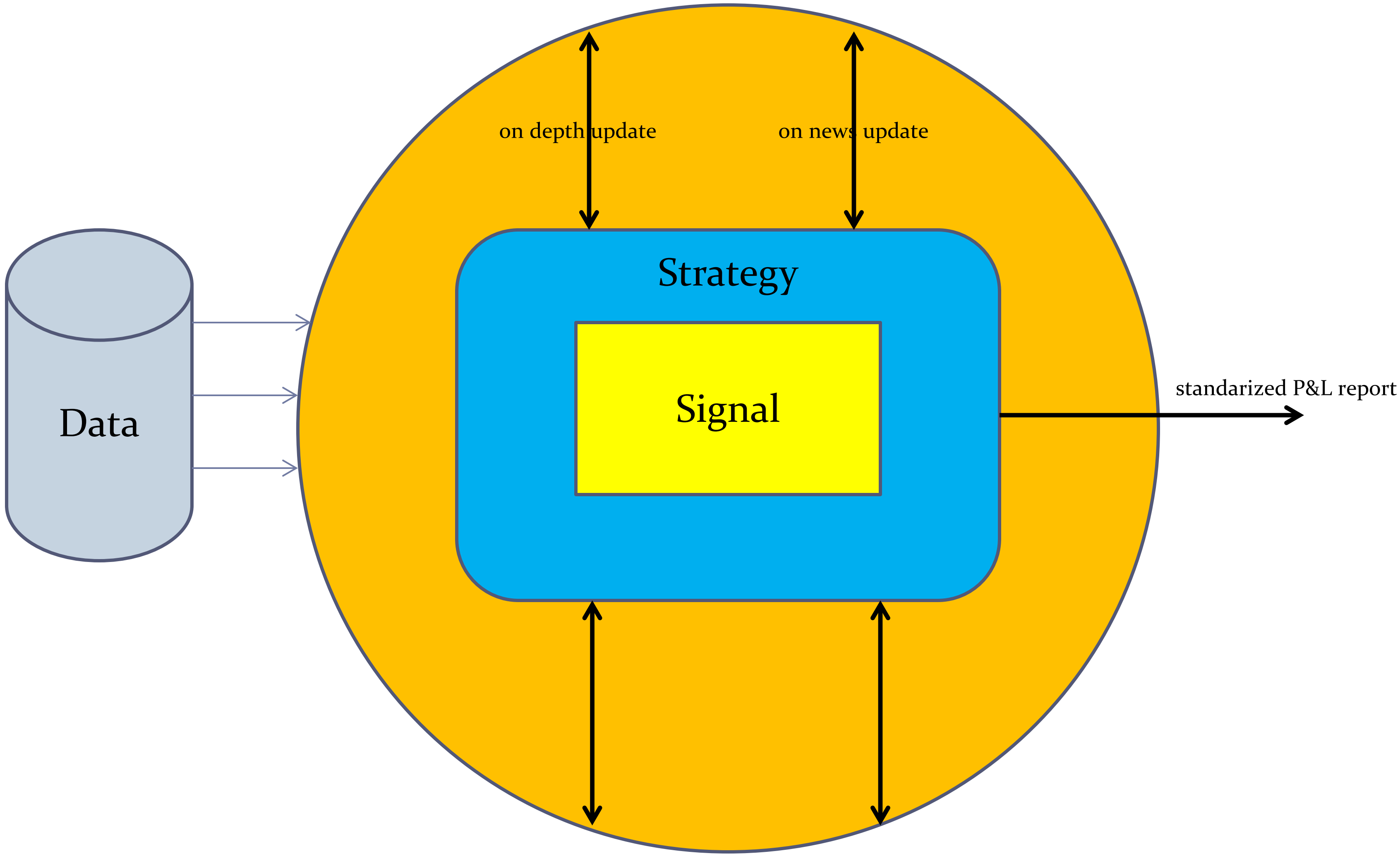
Backtesting

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

Speaker Profile

- Dr. Haksun Li
- CEO, Numerical Method Inc.
- (Ex-)Adjunct Professors, Industry Fellow, Advisor, Consultant with the National University of Singapore, Nanyang Technological University, Fudan University, the Hong Kong University of Science and Technology.
- Quantitative Trader/Analyst, BNPP, UBS
- PhD, Computer Sci, University of Michigan Ann Arbor
- M.S., Financial Mathematics, University of Chicago
- B.S., Mathematics, University of Chicago

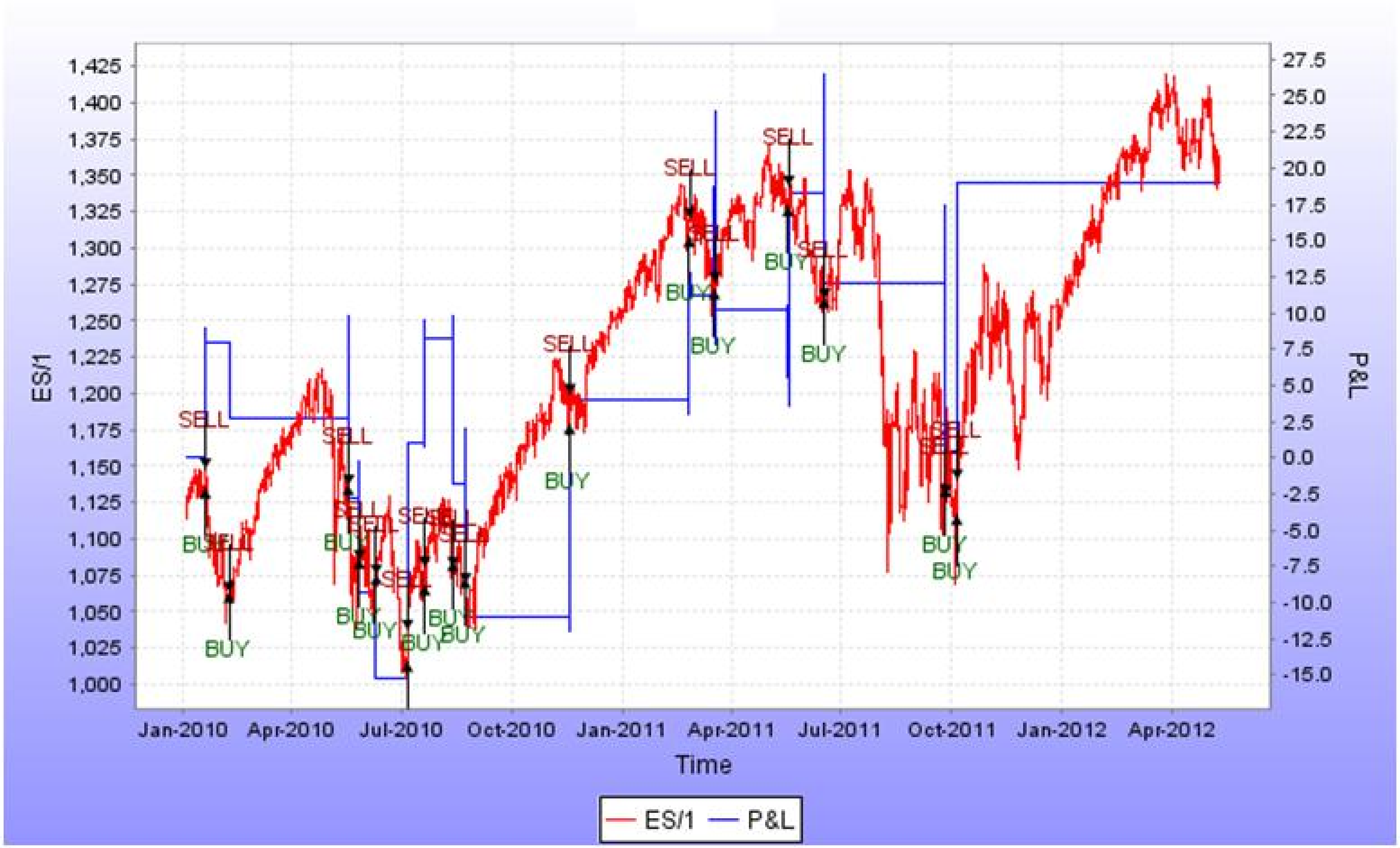
AlgoQuant Framework



Backtesting

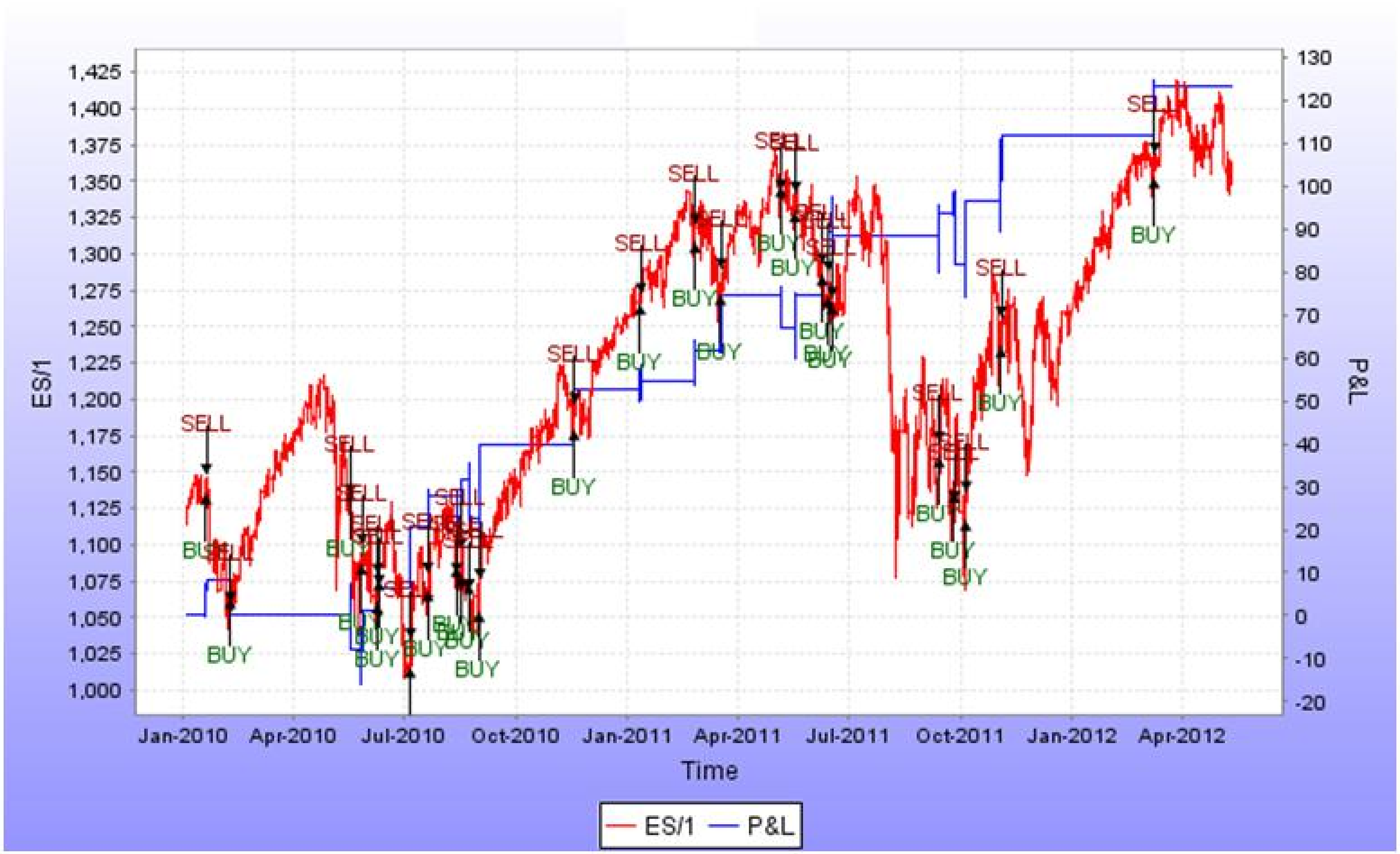
- Backtesting with historical data.
- Backtesting with optimized parameters.
- Sensitivity analysis for all parameters.
- Backtesting with carefully chosen parameters.
- Identify the P&L sources.
- Backtesting with bootstrapped data.

Historical P&L



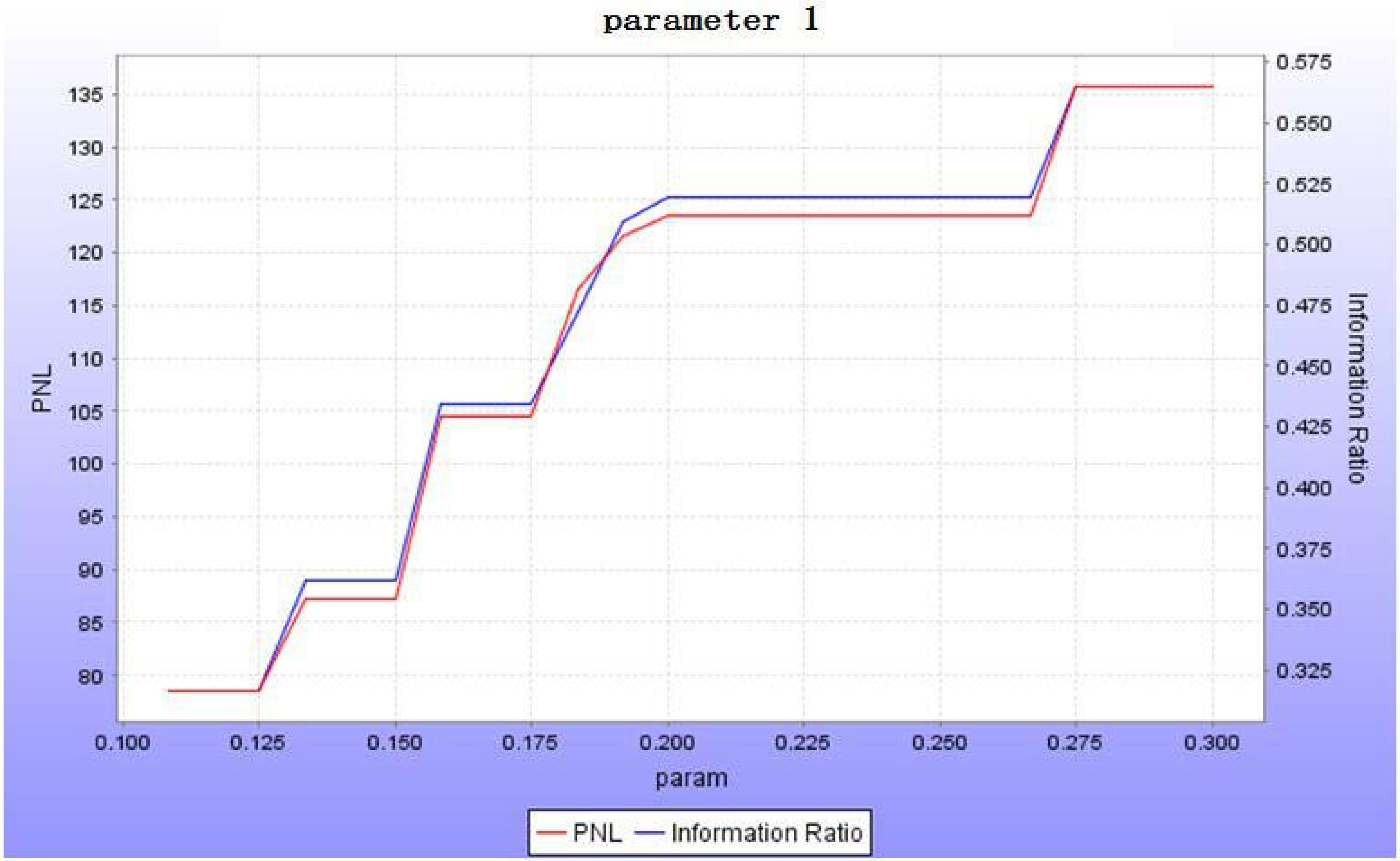
P&L = 19.00; IR = 0.076794

Optimized P&L

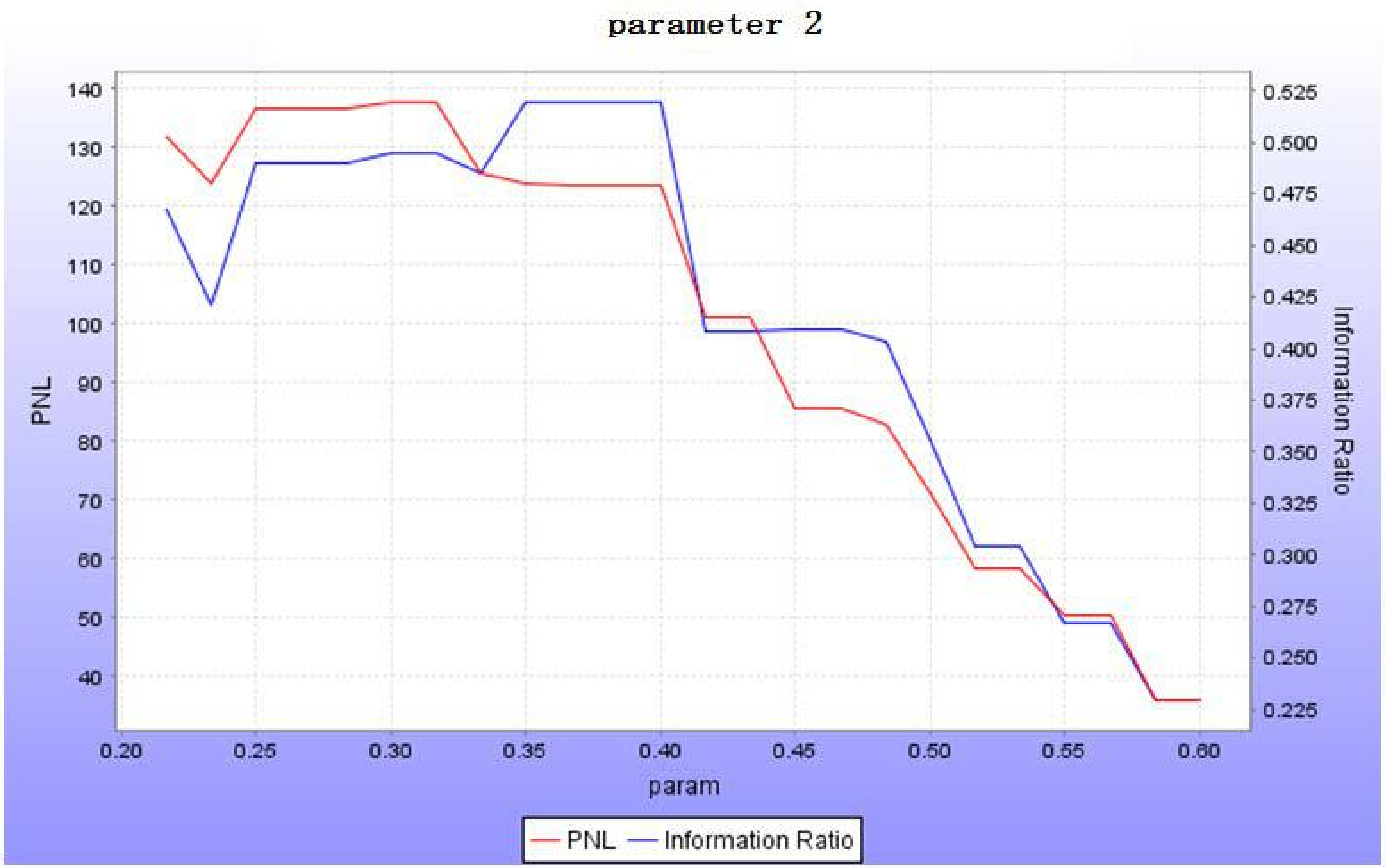


P&L = 123.5; IR = 0.519504

Sensitivity Analysis of Parameters (1)

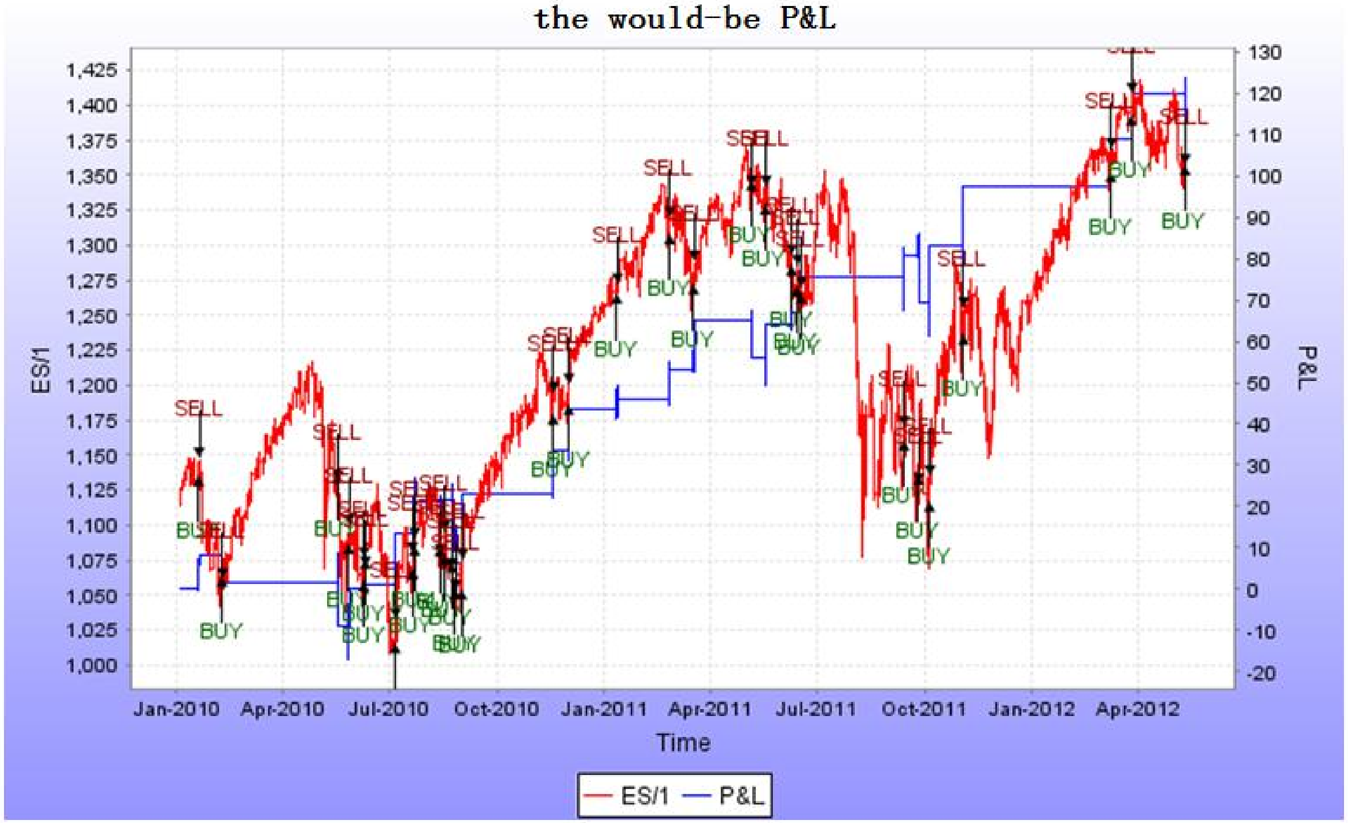


Sensitivity Analysis of Parameters (2)



Sensitivity Analysis of Parameters (3)

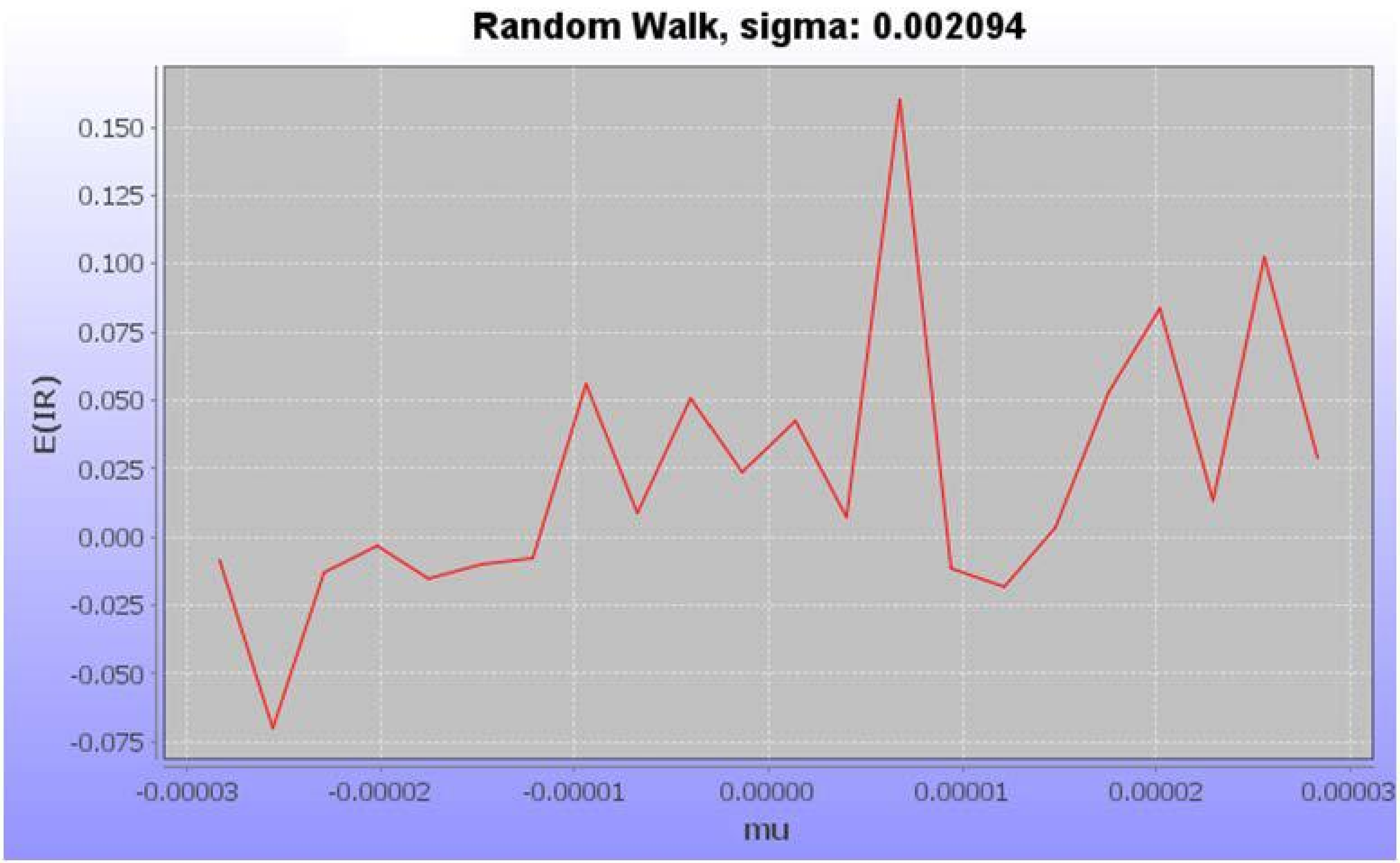
Backtesting Using Recommended Parameters



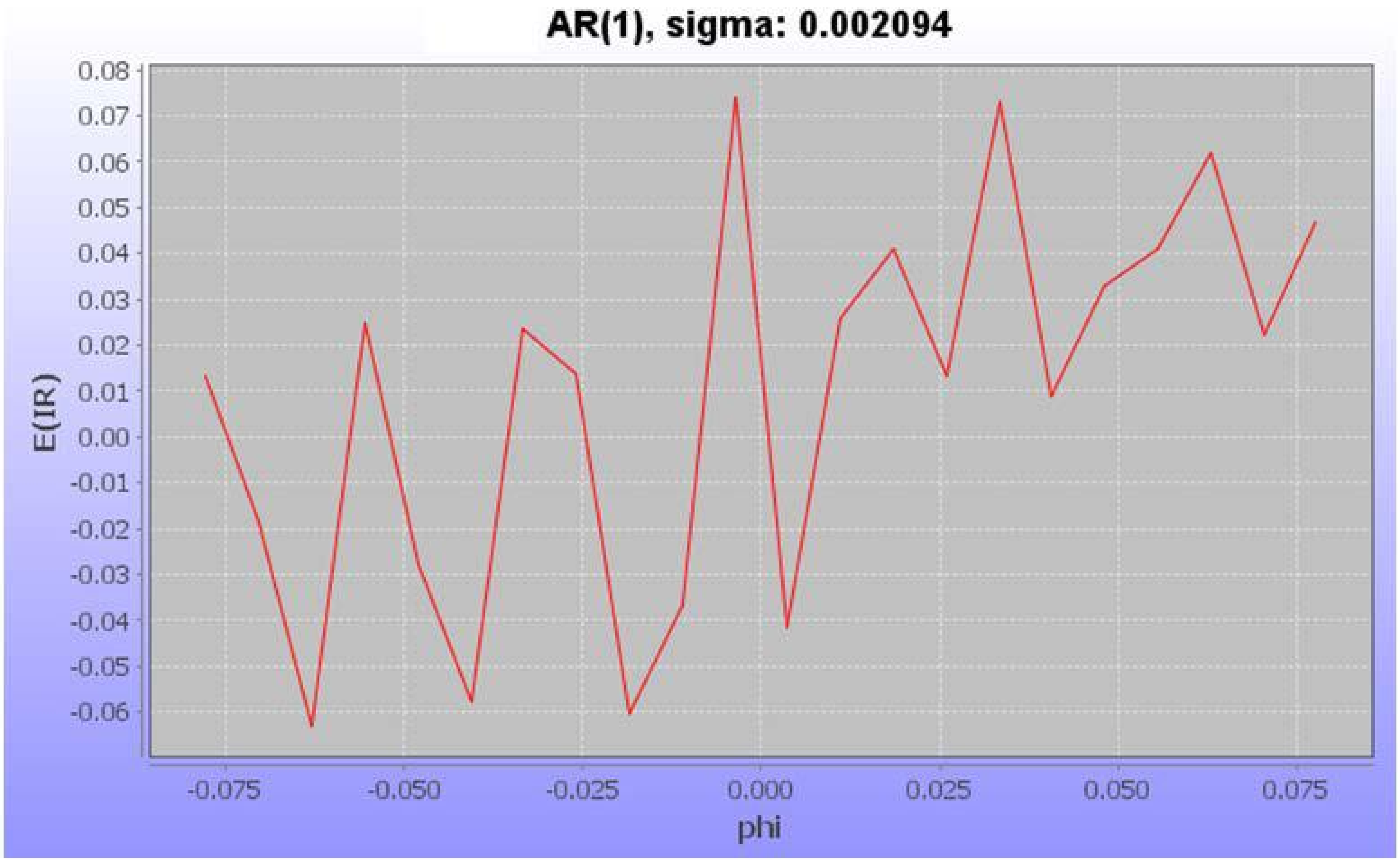
P&L = 115.50; IR = 0.450072

The performance increase is 486.08%.

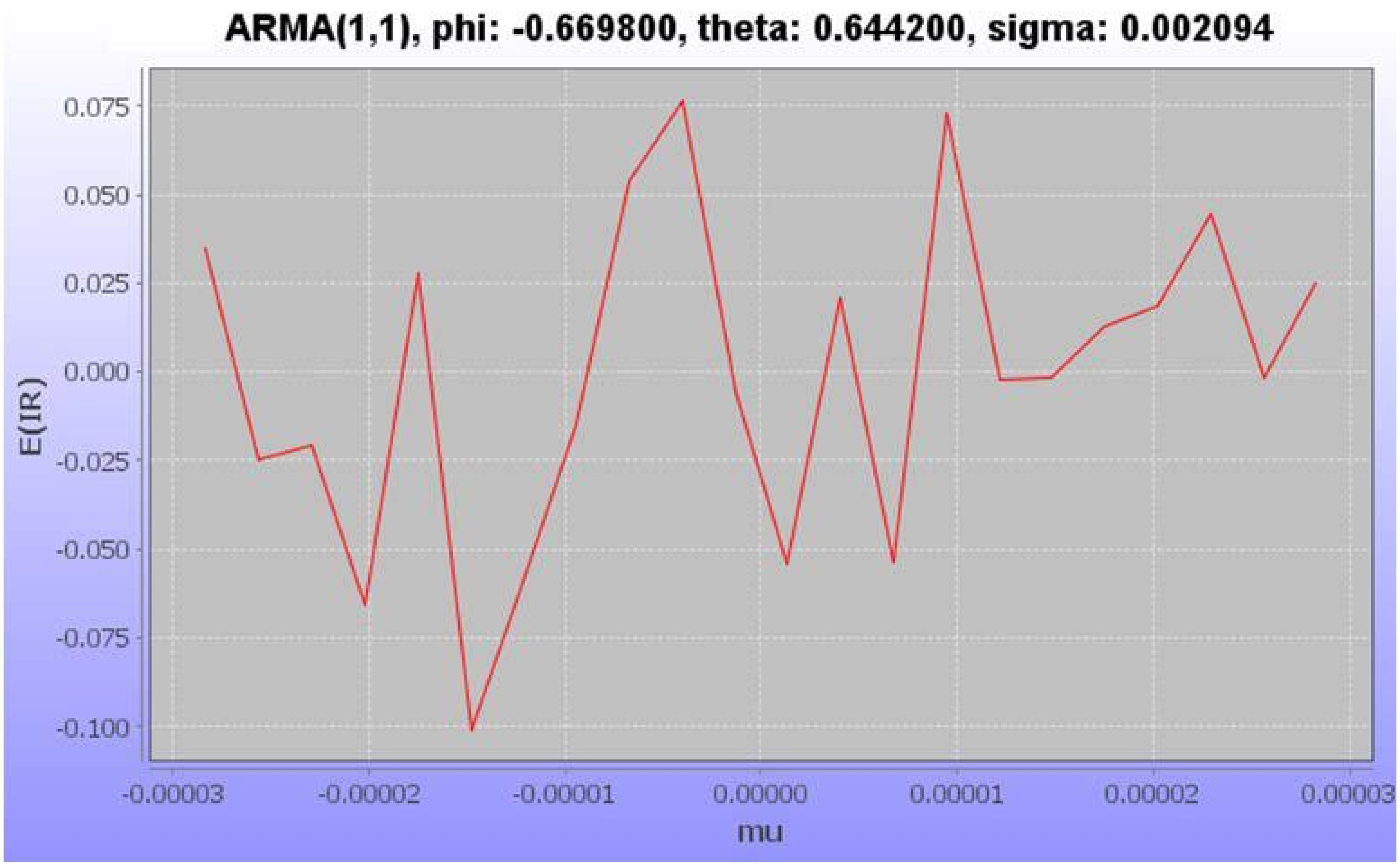
Monte Carlo Simulation – Random Walk with Drift



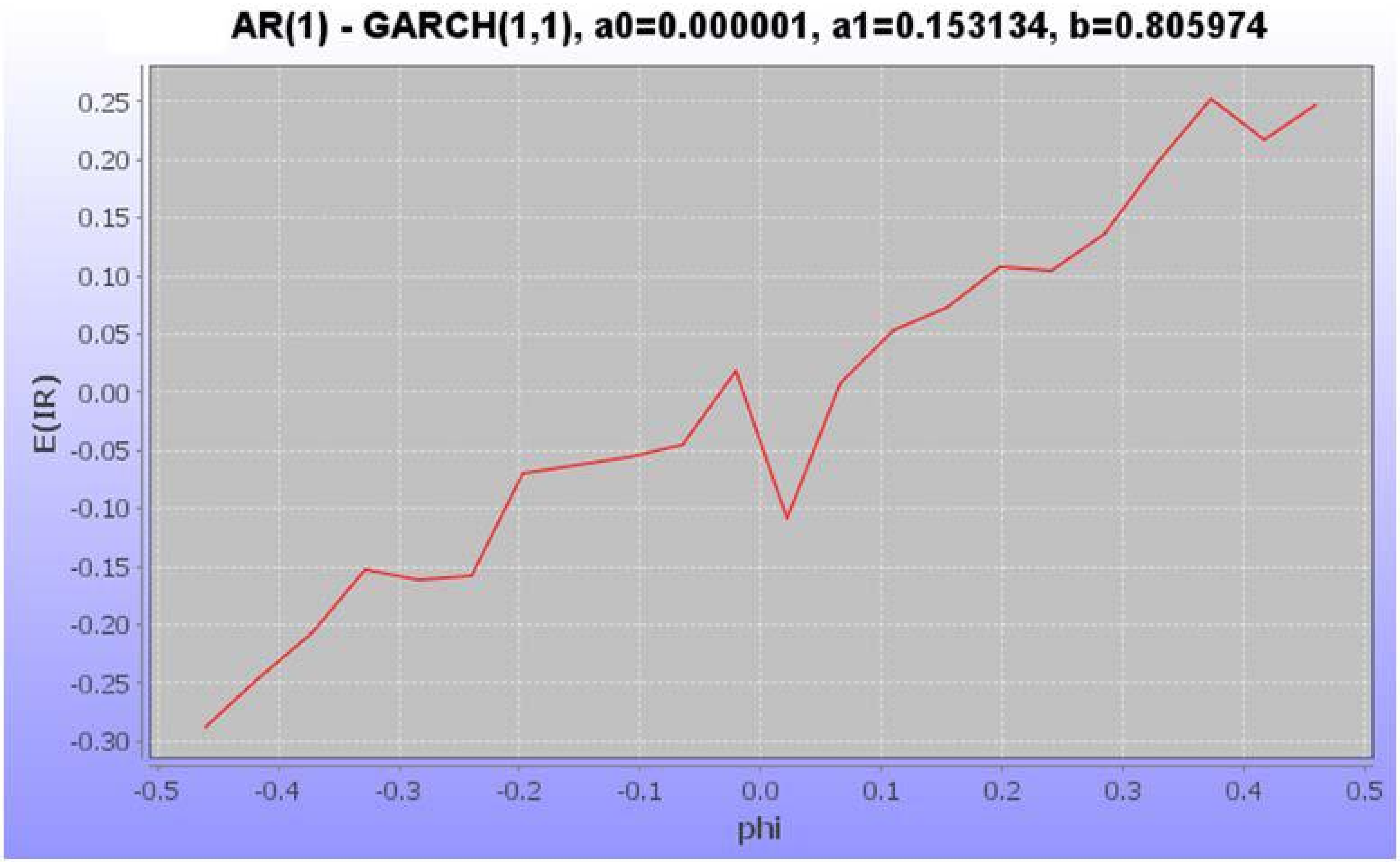
Monte Carlo Simulation – AR(1)



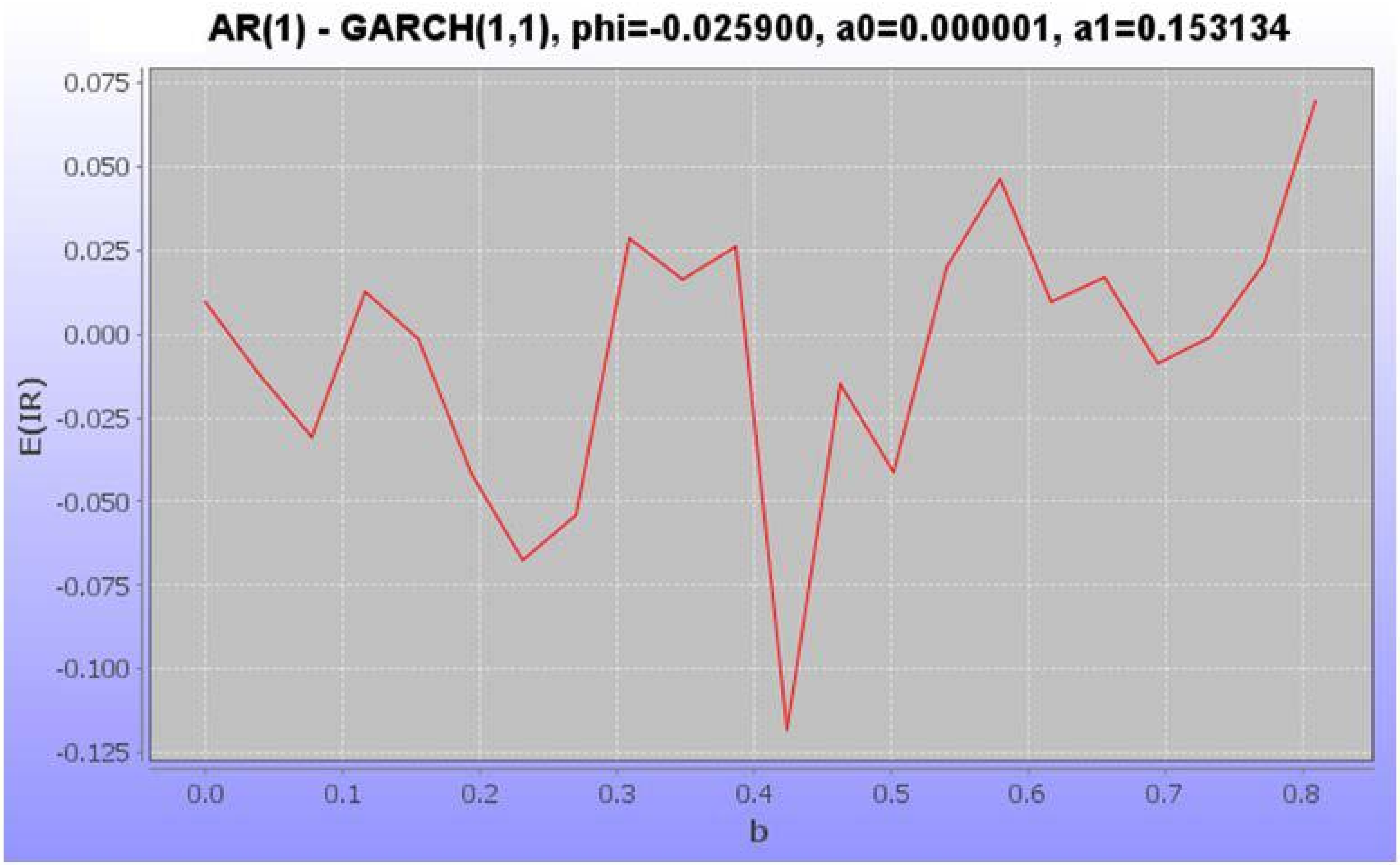
Monte Carlo Simulation – ARMA(1,1)



Monte Carlo Simulation – AR(1)+GARCH(1,1), ϕ



Monte Carlo Simulation – AR(1)+GARCH(1,1), b



Bootstrapping

- Expected P&L and Variance of P&L
- Politis, N. Dimitris, White Halbert, "Automatic Block-Length Selection for the Dependent Bootstrap", Econometric Reviews , 2004.
- Politis, D., White, H., Patton Andrew,"CORRECTION TO 'Automatic Block-Length Selection for the Dependent Bootstrap'", Econometric Reviews, 28(4):372–375, 2009.

Computing Power

- Very demanding!
- Need to run all different types of backtesting in parallel on a grid.

Portfolio Optimization & Risk Management

Haksun Li
haksun.li@numericalmethod.com
www.numericalmethod.com

Speaker Profile

- Dr. Haksun Li
- CEO, Numerical Method Inc.
- (Ex-)Adjunct Professors, Industry Fellow, Advisor, Consultant with the National University of Singapore, Nanyang Technological University, Fudan University, the Hong Kong University of Science and Technology.
- Quantitative Trader/Analyst, BNPP, UBS
- PhD, Computer Sci, University of Michigan Ann Arbor
- M.S., Financial Mathematics, University of Chicago
- B.S., Mathematics, University of Chicago

References

- Connor Keating, William Shadwick. A universal performance measure. Finance and Investment Conference 2002. 26 June 2002.
- Connor Keating, William Shadwick. An introduction to Omega. 2002.
- Kazemi, Scheeweis and Gupta. Omega as a performance measure. 2003.
- S. Avouyi-Dovi, A. Morin, and D. Neto. Optimal asset allocation with Omega function. Tech. report, Banque de France, 2004. Research Paper.
- AJ McNeil. Extreme Value Theory for Risk Managers. 1999.
- Blake LeBaron, Ritirupa Samanta. Extreme Value Theory and Fat Tails in Equity Markets. November 2005.

Portfolio Optimization

Notations

- $r = (r_1, \dots, r_n)'$: a *random* vector of returns, either for a single asset over n periods, or a basket of n assets
- Q : the covariance matrix of the returns
- $x = (x_1, \dots, x_n)'$: the weightings given to each holding period, or to each asset in the basket

Portfolio Statistics

- Mean of portfolio
  - $\mu(x) = x'E(r)$
- Variance of portfolio
  - $\sigma^2(x) = x'Qx$

Sharpe Ratio

- $\mathrm{sr}(x) = \frac{\mu(x) - r_f}{\sigma^2(x)} = \frac{x' E(r) - r_f}{x' Q x}$
- r_f: a benchmark return, e.g., risk-free rate
- In general, we prefer
  - a bigger excess return
  - a smaller risk (uncertainty)

Sharpe Ratio Limitations

- Sharpe ratio does not differentiate between winning and losing trades, essentially ignoring their likelihoods (odds).
- Sharpe ratio does not consider, essentially ignoring, all higher moments of a return distribution except the first two, the mean and variance.

Sharpe's Choice

- Both A and B have the same mean.
- A has a smaller variance.
- Sharpe will always chooses a portfolio of the smallest variance among all those having the same mean.
  - Hence A is preferred to B by Sharpe.

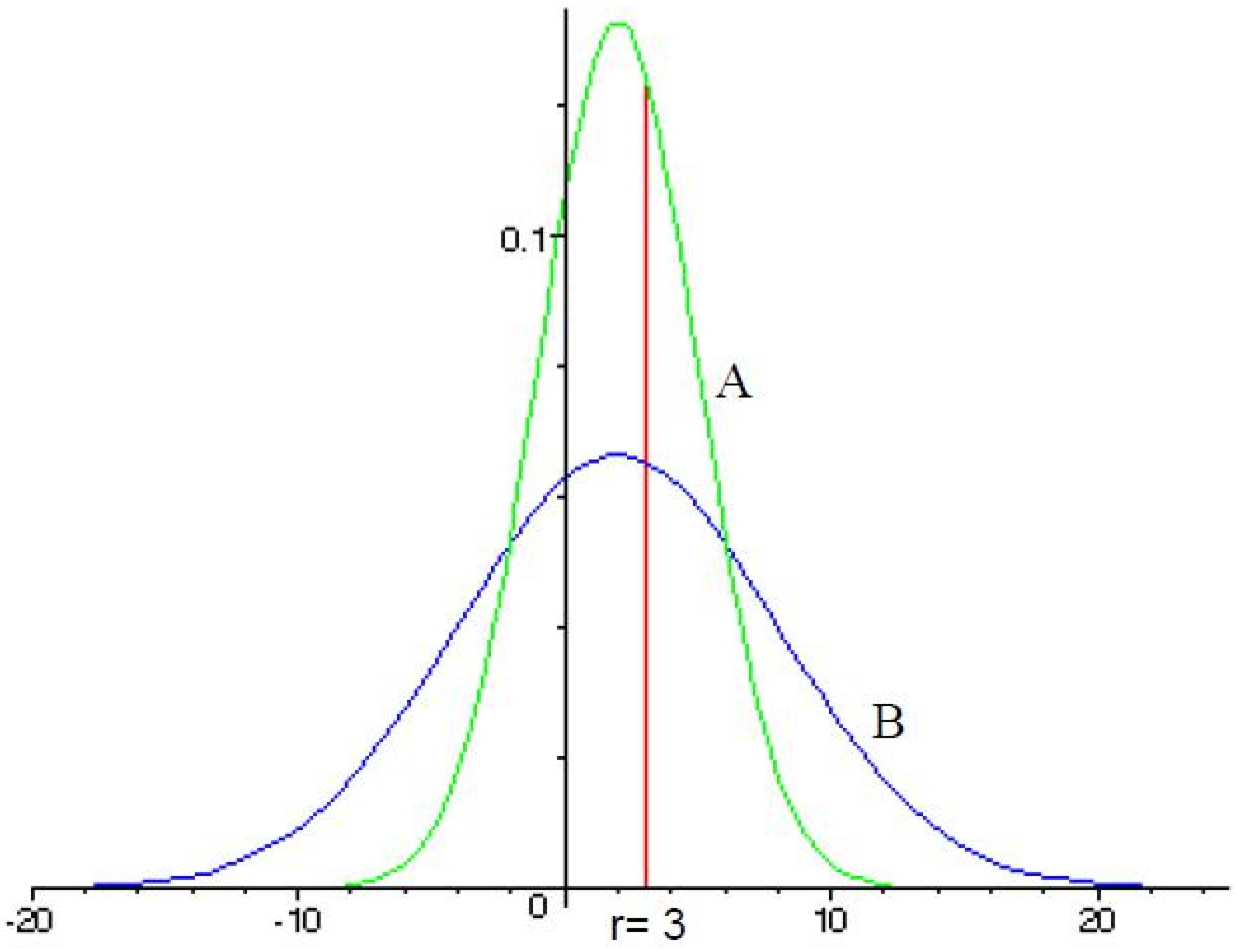
Avoid Downsides and Upsides

- Sharpe chooses the smallest variance portfolio to reduce the chance of having extreme losses.
- Yet, for a Normally distributed return, the extreme gains are as likely as the extreme losses.
- Ignoring the downsides will inevitably ignore the potential for upsides as well.

Potential for Gains

- Suppose we rank A and B by their potential for gains, we would choose B over A.
- Shall we choose the portfolio with the biggest variance then?
  - It is very counter intuitive.

Example 1: A or B?



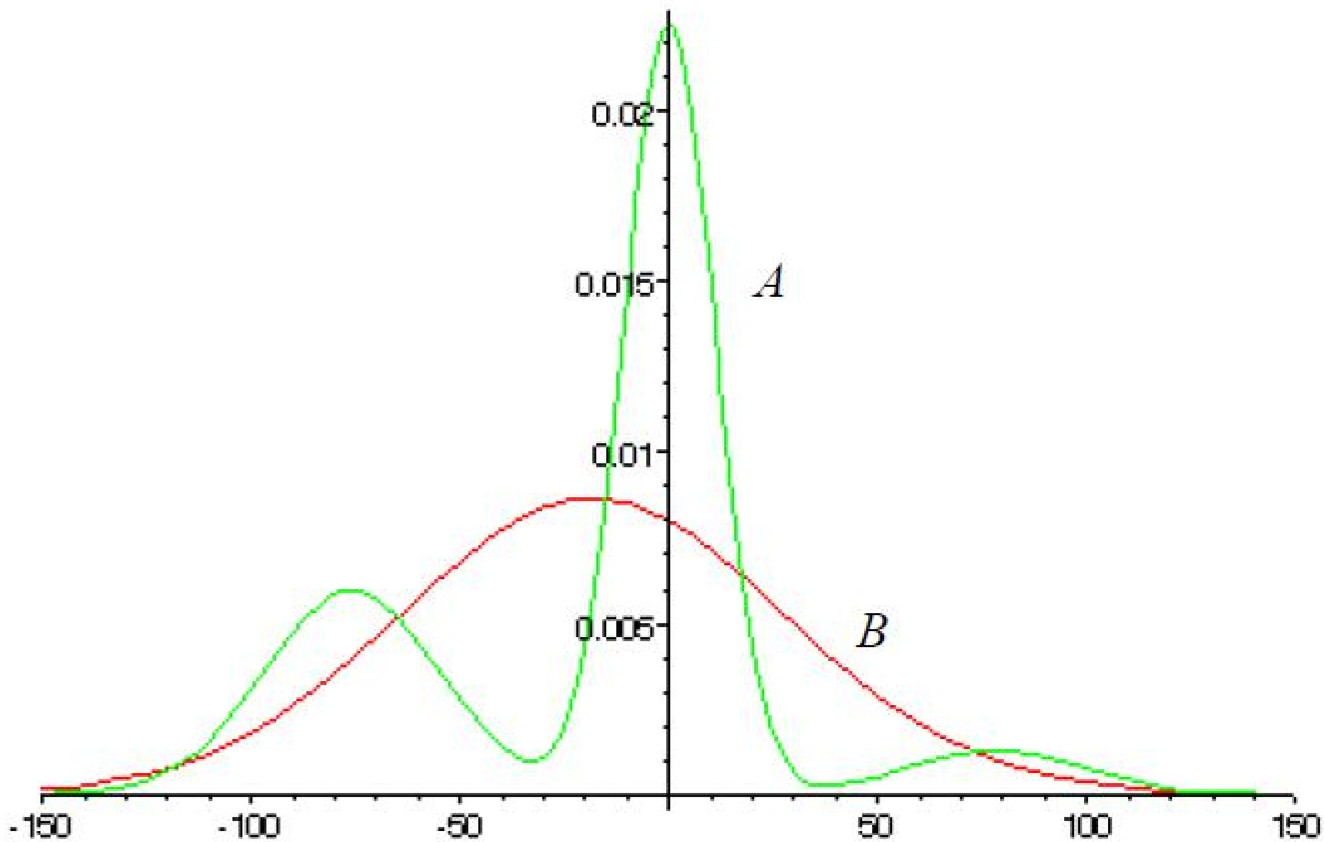
Example 1: L = 3

- Suppose the loss threshold is 3.
- Pictorially, we see that B has more mass to the right of 3 than that of A.
  - B: 43% of mass; A: 37%.
- We compare the likelihood of winning to losing.
  - B: 0.77; A: 0.59.
- We therefore prefer B to A.

Example 1: L = 1

- Suppose the loss threshold is 1.
- A has more mass to the right of L than that of B.
- We compare the likelihood of winning to losing.
  - A: 1.71; B: 1.31.
- We therefore prefer A to B.

Example 2



Example 2: Winning Ratio

- It is evident from the example(s) that, when choosing a portfolio, the likelihoods/odds/chances/potentials for upside and downside are important.
- Winning ratio $\frac{W_A}{W_B}$:
  - 2σ gain: 1.8
  - 3σ gain: 0.85
  - 4σ gain: 35

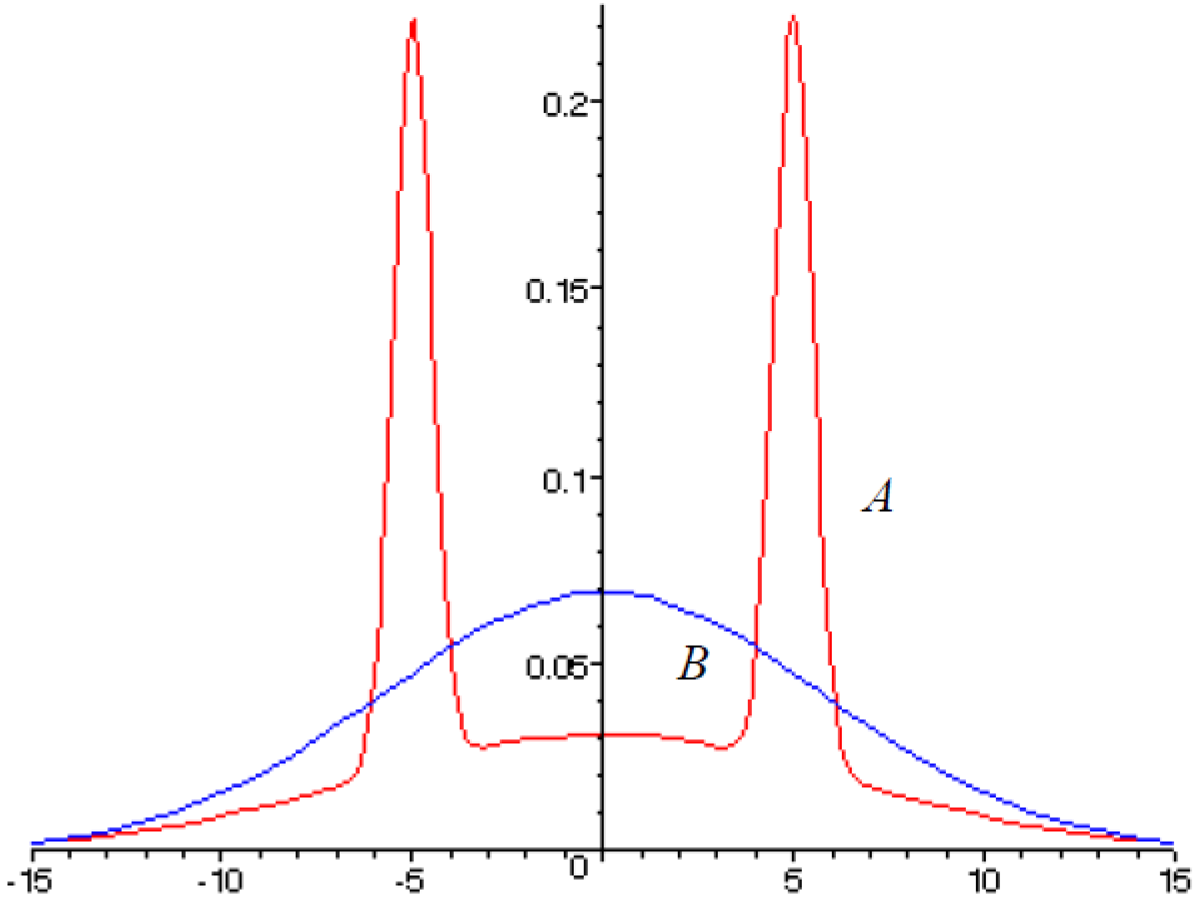
Example 2: Losing Ratio

- Losing ratio $\frac{L_A}{L_B}$:
  - 1σ loss: 1.4
  - 2σ loss: 0.7
  - 3σ loss : 80
  - 4σ loss : 100,000!!!

Higher Moments Are Important

- Both large gains and losses in example 2 are produced by moments of order 5 and higher.
  - They even shadow the effects of skew and kurtosis.
  - Example 2 has the same mean and variance for both distributions.
- Because Sharpe Ratio ignores all moments from order 3 and bigger, it treats all these very different distributions the same.

How Many Moments Are Needed?



Distribution A

- Combining 3 Normal distributions
  - N(-5, 0.5)
  - N(0, 6.5)
  - N(5, 0.5)
- Weights:
  - 25%
  - 50%
  - 25%

Moments of A

- Same mean and variance as distribution B.
- Symmetric distribution implies all odd moments (3^{rd}, 5^{th}, etc.) are 0.
- Kurtosis = 2.65 (smaller than the 3 of Normal)
  - Does smaller Kurtosis imply smaller risk?
- 6^{th} moment: 0.2% different from Normal
- *8^{th} moment: 24% different from Normal*
- *10^{th} moment: 55% bigger than Normal*

Performance Measure Requirements

- Take into account the odds of winning and losing.
- Take into account the sizes of winning and losing.
- Take into account of (all) the moments of a return distribution.

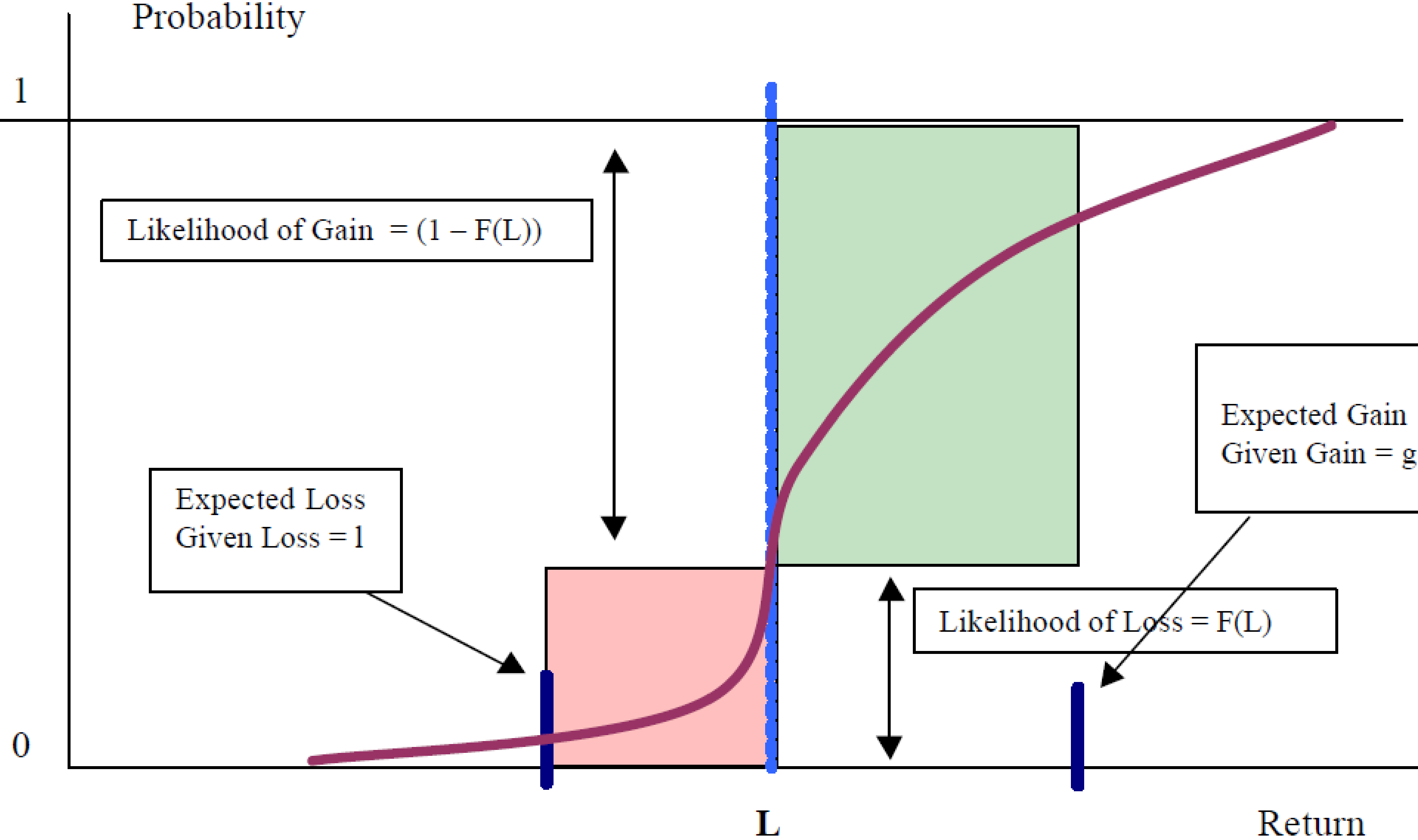
Loss Threshold

- Clearly, the definition, hence likelihoods, of winning and losing depends on how we define loss.
- Suppose L = Loss Threshold,
  - for return $< L$, we consider it a loss
  - for return $> L$, we consider it a gain

An Attempt

- To account for
  - the odds of wining and losing
  - the sizes of wining and losing
- We consider
  - $\Omega = \frac{E(r|r>L)\times P(r>L)}{E(r|r\leq L)\times P(r\leq L)}$
- $\Omega = \frac{E(r|r>L)\left(1-F(L)\right)}{E(r|r\leq L)F(L)}$

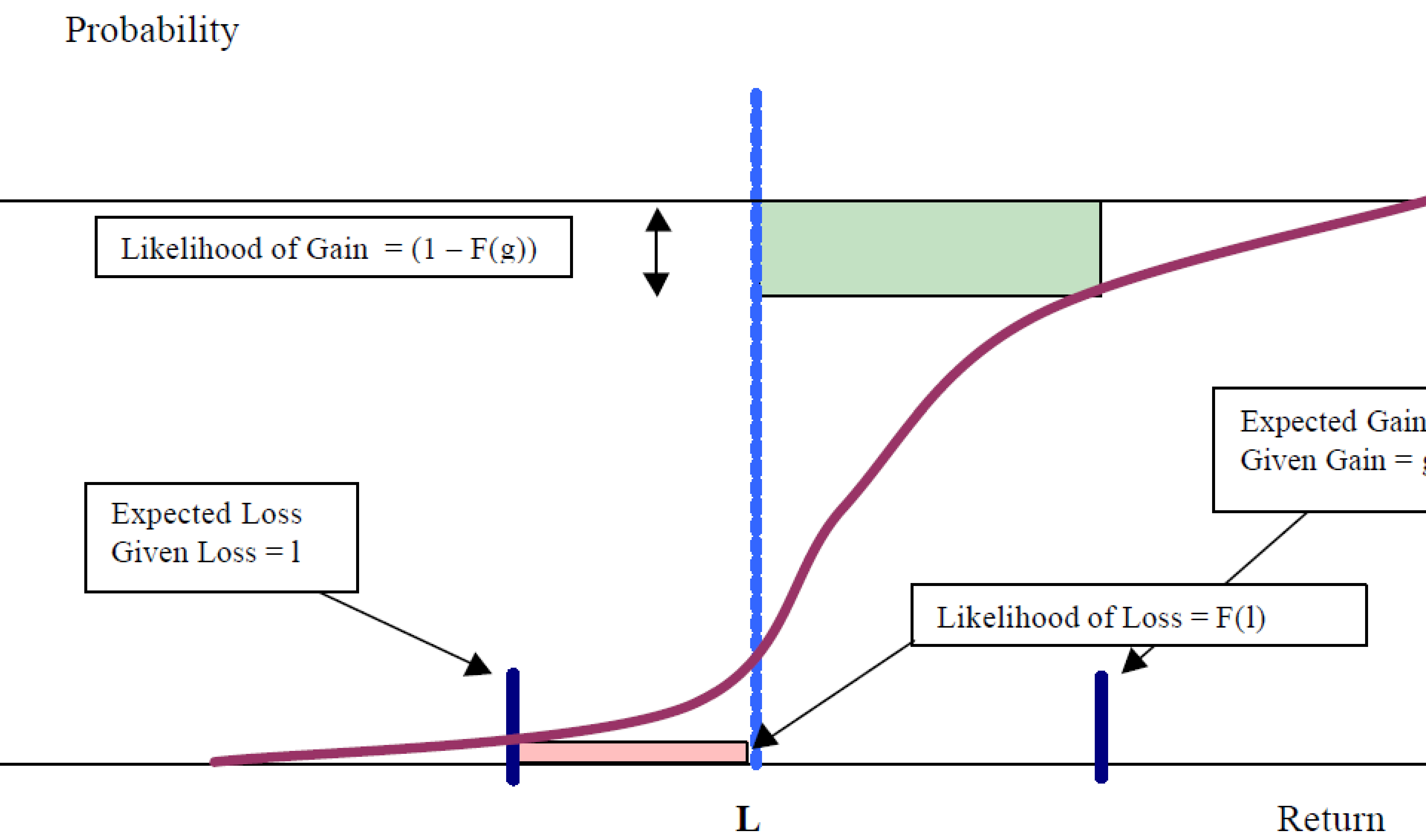
First Attempt



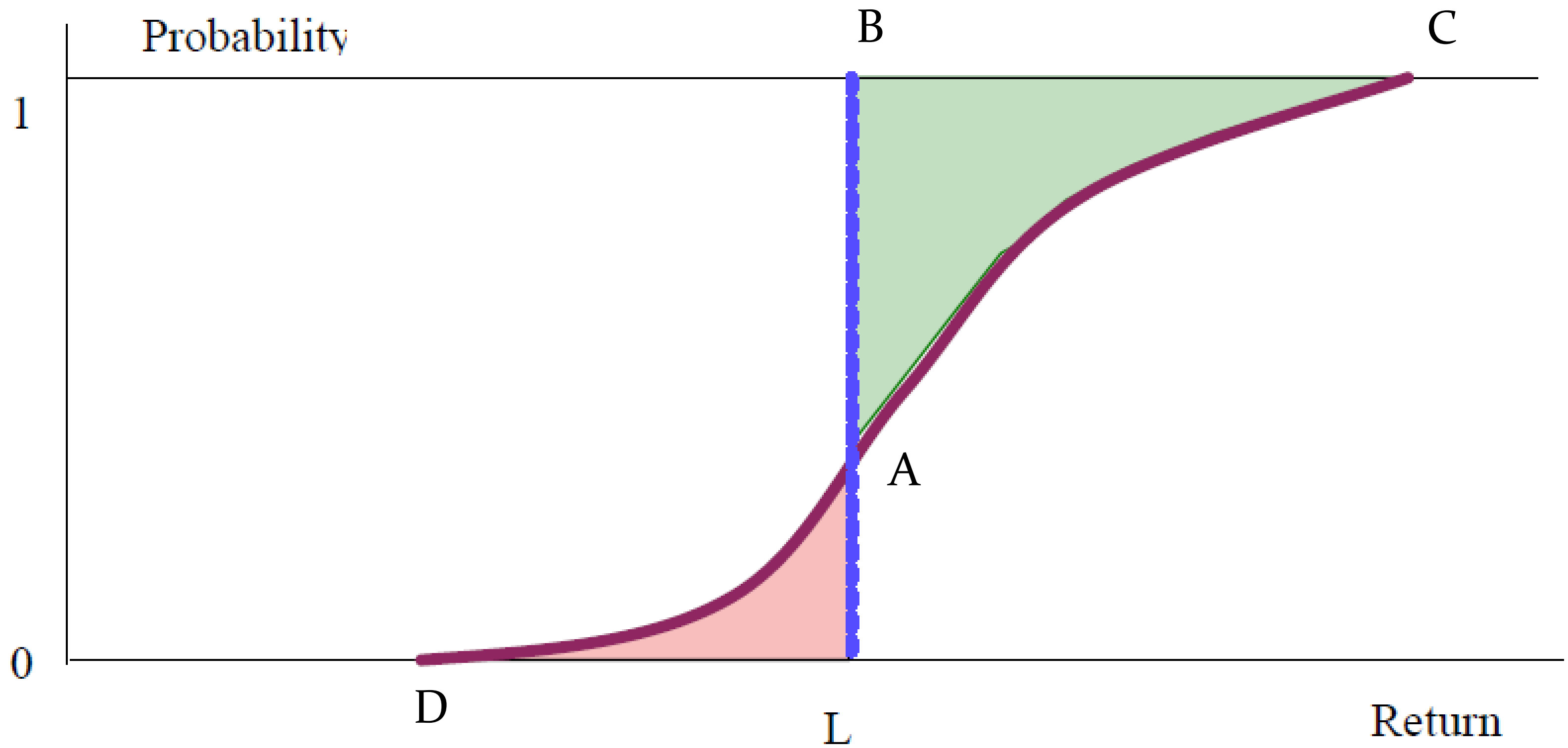
First Attempt Inadequacy

- Why F(L)?
- Not using the information from the entire distribution.
  - hence ignoring higher moments

Another Attempt



Yet Another Attempt



Omega Definition

- Ω takes the concept to the limit.
- Ω uses the whole distribution.
- Ω definition:
  - $\Omega = \frac{ABC}{ALD}$
- $\Omega = \frac{\int_{L}^{b=\max\{r\}}[1-F(r)]dr}{\int_{a=\min\{r\}}^{L} F(r)dr}$

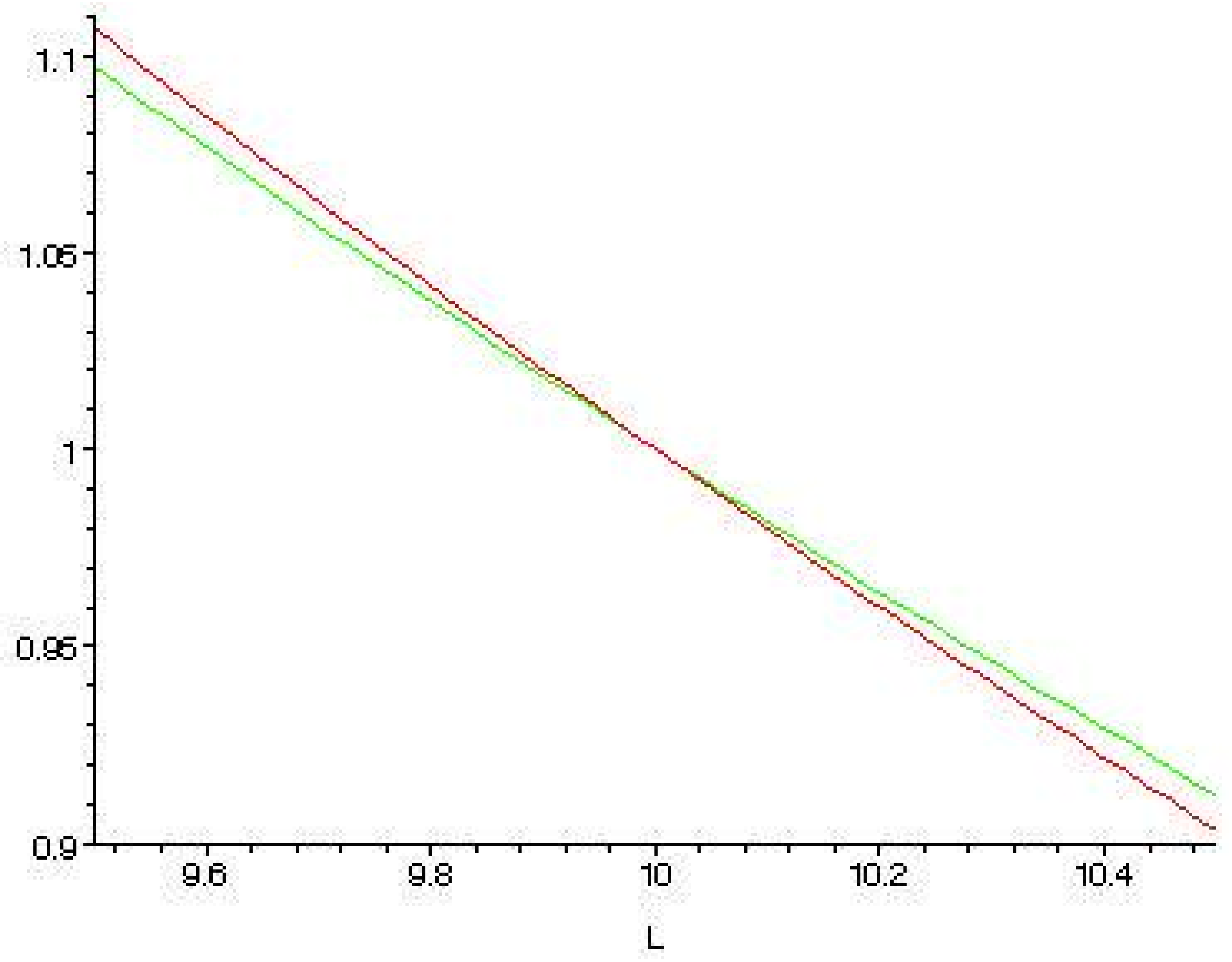
Intuitions

- Omega is a ratio of winning size weighted by probabilities to losing size weighted by probabilities.
- Omega considers size and odds of winning and losing trades.
- Omega considers all moments because the definition incorporates the whole distribution.

Omega Advantages

- There is no parameter (estimation).
- There is no need to estimate (higher) moments.
- Work with all kinds of distributions.
- Use a function (of Loss Threshold) to measure performance rather than a single number (as in Sharpe Ratio).
- It is as smooth as the return distribution.
- It is monotonic decreasing.

Omega Example



Affine Invariant

- $\varphi: r \ \rightarrow Ar + B$, iff $\widehat{\Omega}\big(\varphi(L)\big) = \Omega(L)$
- $L \rightarrow AL + B$
- We may transform the returns distribution using any invertible transformation before calculating the Gamma measure.
- The transformation can be thought of as some sort of utility function, modifying the mean, variance, higher moments, and the distribution in general.

Numerator Integral (1)

- $\int_L^b d\left[x\left(1 - F(x)\right)\right]$
- $= \left[x\left(1 - F(x)\right)\right]^b{}_L$
- $= b\left(1 - F(b)\right) - L\left(1 - F(L)\right)$
- $= -L\left(1 - F(L)\right)$

Numerator Integral (2)

- $\int_L^b d\left[x\left(1 - F(x)\right)\right]$
- $= \int_L^b \left(1 - F(x)\right)dx + \int_L^b xd\left(1 - F(x)\right)$
- $= \int_L^b \left(1 - F(x)\right)dx - \int_L^b xdF(x)$

Numerator Integral (3)

- $-L\big(1-F(L)\big) = \int_L^b \big(1-F(x)\big)dx - \int_L^b x dF(x)$
- $\int_L^b \big(1-F(x)\big)dx = -L\big(1-F(L)\big) + \int_L^b x dF(x)$
- $= \int_L^b (x-L) f(x) dx$
- $= \int_a^b \max(x-L, 0) f(x) dx$
- $= E[\max(x-L, 0)]$

undiscounted call option price

Denominator Integral (1)

- $\int_a^L d[xF(x)]$
- $= [xF(x)]^L{}_a$
- $= LF(L) - a\big(F(a)\big)$
- $= LF(L)$

Denominator Integral (2)

- $\int_a^L d[xF(x)]$
- $= \int_a^L F(x)dx + \int_a^L x dF(x)$

Denominator Integral (3)

- $LF(L) = \int_a^L F(x)dx + \int_a^L xdF(x)$
- $\int_a^L F(x)dx = LF(L) - \int_a^L xdF(x)$
- $= \int_a^L (L - x)f(x)dx$
- $= \int_a^b \max(L - x, 0)f(x)dx$
- $= E[\max(L - x, 0)]$

undiscounted put option price

Another Look at Omega

- $\Omega = \frac{\int_{L}^{b=\max\{r\}}[1-F(r)]dr}{\int_{a=\min\{r\}}^{L} F(r)dr}$
- $= \frac{E[\max(x-L,0)]}{E[\max(L-x,0)]}$
- $= \frac{e^{-r_f} E[\max(x-L,0)]}{e^{-r_f} E[\max(L-x,0)]}$
- $= \frac{C(L)}{P(L)}$

Options Intuition

- Numerator: the cost of acquiring the return above L
- Denominator: the cost of protecting the return below L
- Risk measure: the put option price as the cost of protection is a much more general measure than variance

Can We Do Better?

- Excess return in Sharpe Ratio is more intuitive than $C(L)$ in Omega.
- Put options price as a risk measure in Omega is better than variance in Sharpe Ratio.

Sharpe-Omega

- $\Omega_S = \frac{\bar{r}-L}{P(L)}$
- In this definition, we combine the advantages in both Sharpe Ratio and Omega.
  - meaning of excess return is clear
  - risk is bettered measured
- Sharpe-Omega is more intuitive.
- *Ω_S ranks the portfolios in exactly the same way as Ω.*

Sharpe-Omega and Moments

- It is important to note that the numerator relates only to the first moment (the mean) of the returns distribution.
- It is the denominator that take into account the variance and all the higher moments, hence the whole distribution.

Sharpe-Omega and Variance

- Suppose $\bar{r} > L$. $\Omega_S > 0$.
  - The bigger the volatility, the higher the put price, the bigger the risk, the smaller the Ω_S, the less attractive the investment.
  - We want smaller volatility to be more certain about the gains.
- Suppose $\bar{r} < L$. $\Omega_S < 0$.
  - The bigger the volatility, the higher the put price, the bigger the Ω_S, the more attractive the investment.
  - Bigger volatility increases the odd of earning a return above L.

Portfolio Optimization

- In general, a Sharpe optimized portfolio is different from an Omega optimized portfolio.
- How different?

Optimization for Sharpe

- $$\begin{cases} \min\limits_{x} x'\Sigma x \\ \sum_i^n x_i E(r_i) \geq \rho \\ \sum_i^n x_i = 1 \\ {x_i}^l \leq x_i \leq 1 \end{cases}$$
- Minimum holding: $x^l = \left({x_1}^l, \ldots, {x_n}^l\right)'$

$$\begin{aligned} \min\ & \frac{1}{2} x^T Q x + c^T x, \\ & Ax = b, \\ & x^T Q_i x + d_i \leq b_i, \qquad i = 1, \ldots, m_q \\ & x \geq 0, \end{aligned}$$

Optimization s.t. Constraints

- $\max_x \left\{ \bar{r}' x - \lambda_1 x' \Sigma x - \lambda_2 \sum_{i=1}^{n} m_i \left| x_i - w_{0_i} \right|^{\frac{3}{2}} \right\}$
- $\sum_{i=1}^{n} x = 0$, self financing
- $x_i = 0$, black list
- Many more…

$$
\begin{array}{ll}
\text{maximize} & c^\top x \\
\text{subject to} & Ax = b \\
 & x \in Q^n,
\end{array}
$$

Optimization for Omega

- $$\begin{cases} \max\limits_{x} \Omega_S(x) \\ \sum_i^n x_i E(r_i) \geq \rho \\ \sum_i^n x_i = 1 \\ {x_i}^l \leq x_i \leq 1 \end{cases}$$
- Minimum holding: $x^l = \left({x_1}^l, \dots, {x_n}^l\right)'$

Optimization Methods

- Nonlinear Programming
  - Penalty Method
- Global Optimization
  - Differential Evolution
  - Threshold Accepting algorithm (Avouyi-Dovi et al.)
  - Tabu search (Glover 2005)
  - MCS algorithm (Huyer and Neumaier 1999)
  - Simulated Annealing
  - Genetic Algorithm
- Integer Programming (Mausser et al.)

3 Assets Example

- $x_1 + x_2 + x_3 = 1$
- $R_i = x_1 r_{1i} + x_2 r_{2i} + x_3 r_{3i}$
- $= x_1 r_{1i} + x_2 r_{2i} + (1 - x_1 - x_2) r_{3i}$

Penalty Method

- $F(x_1, x_2) = -\Omega(R_i) + \rho\{[\min(0, x_1)]^2 + [\min(0, x_2)]^2 + [\min(0, 1 - x_1 - x_2)]^2\}$
- Can apply Nelder-Mead, a Simplex algorithm that takes initial guesses.
- F needs not be differentiable.
- Can do random-restart to search for global optimum.

Threshold Accepting Algorithm

- It is a local search algorithm.
  - It explores the potential candidates around the current best solution.
- It "escapes" the local minimum by allowing choosing a lower than current best solution.
  - This is in very sharp contrast to a hilling climbing algorithm.

Objective

- Objective function
  - $h: X \to R, X \in R^n$
- Optimum
  - $h_{\text{opt}} = \max_{x \in X} h(x)$

Initialization

- Initialize n (number of iterations) and $step$.
- Initialize sequence of thresholds th_k, $k = 1, \dots, step$
- Starting point: $x_0 \in X$

Thresholds

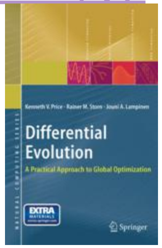
- Simulate a set of portfolios.
- Compute the distances between the portfolios.
- Order the distances from the biggest to the smallest.
- Choose the first $step$ number of them as thresholds.

Search

- $x_{i+1} \in N_{x_i}$ (neighbour of x_i)
- Threshold: $\Delta h = h(x_{i+1}) - h(x_i)$
- Accepting: If $\Delta h > th_k$ set $x_{i+1} = x_i$
- Continue until we finish the last (smallest) threshold.
  - $h(x_i) \approx h_{opt}$
- Evaluating h by Monte Carlo simulation.

Differential Evolution

- DE is a simple and yet very powerful global optimization method.
- It is ideal for multidimensional, mutilmodal functions, i.e. very hard problems.
- It works with hard-to-model constraints, e.g., max drawdown.
- DE is implemented in SuanShu.
  - $z = a + F(b - c)$ with a certain probabilty
- http://numericalmethod.com/blog/2011/05/31/strategy-optimization/



Risk Management

Risks

- Financial theories say:
  - the most important single source of profit is risk.
  - profit $\propto$ risk.
- *I personally do not agree.*

What Are Some Risks? (1)

- Bonds:
  - duration (sensitivity to interest rate)
  - convexity
  - term structure models
- Credit:
  - rating
  - default models

What Are Some Risks? (2)

- Stocks
  - volatility
  - correlations
  - beta
- Derivatives
  - delta
  - gamma
  - vega

What Are Some Risks? (3)

- FX
  - volatility
  - target zones
  - spreads
  - term structure models of related currencies

Other Risks?

- Too many to enumerate…
  - natural disasters, e.g., earthquake
  - war
  - politics
  - operational risk
  - regulatory risk
  - wide spread rumors
  - alien attack!!!
- Practically infinitely many of them…

VaR Definition

- Given a loss distribution, F, quintile $1 > \mathrm{q} \geq 0.95$,
- $\mathrm{VaR}_q = F^{-1}(q)$

Expected Shortfall

- Suppose we hit a big loss, what is its expected size?
- $\mathrm{ES}_q = E\left[X \mid X > \mathrm{VaR}_q\right]$

VaR in Layman Term

- VaR is the maximum loss which can occur with certain confidence over a holding period (of *n* days).
- Suppose a daily VaR is stated as \$1,000,000 to a 95% level of confidence.
- There is only a 5% chance that the loss the next day will *exceed* \$1,000,000.

Why VaR?

- Is it a true way to measure risk?
  - NO!
- Is it a universal measure accounting for most risks?
  - NO!
- Is it a good measure?
  - NO!
- Only because the industry and regulators have adopted it.
  - It is a widely accepted standard.

VaR Computations

- Historical Simulation
- Variance-CoVariance
- Monte Carlo simulation

Historical Simulations

- Take a historical *returns* time series as the returns distribution.
- Compute the loss distribution from the historical returns distribution.

Historical Simulations Advantages

- Simplest
- Non-parametric, no assumption of distributions, no possibility of estimation error

Historical Simulations Dis-Advantages

- As all historical returns carry equal weights, it runs the risk of over-/under- estimate the recent trends.
- Sample period may not be representative of the risks.
- History may not repeat itself.
- Cannot accommodate for new risks.
- Cannot incorporate subjective information.

Variance-CoVariance

- Assume all returns distributions are Normal.
- Estimate asset variances and covariances from historical data.
- Compute portfolio variance.
  - ${\sigma_P}^2 = \sum_{i,j} \rho_{ij} \omega_i \omega_j \sigma_i \sigma_j$

Variance-CoVariance Example

- 95% confidence level (1.645 stdev from mean)
- Nominal = $10 million
- Price = $100
- Average return = 7.35%
- Standard deviation = 1.99%
- The VaR at 95% confidence level = 1.645 x 0.0199 = 0.032736
- The VaR of the portfolio = 0.032736 x 10 million = $327,360.

Variance-CoVariance Advantages

- Widely accepted approach in banks and regulations.
- Simple to apply; straightforward to explain.
- Datasets immediately available
  - very easy to estimate from historical data
  - free data from RiskMetrics
- Can do scenario tests by twisting the parameters.
  - sensitivity analysis of parameters
  - give more weightings to more recent data

Variance-CoVariance Disadvantages

- Assumption of Normal distribution for returns, which is known to be not true.
- Does not take into account of fat tails.
- Does not work with non-linear assets in portfolio, e.g., options.

Monte Carlo Simulation

- You create your own returns distributions.
  - historical data
  - implied data
  - economic scenarios
- Simulate the joint distributions many times.
- Compute the empirical returns distribution of the portfolio.
- Compute the (e.g., 1%, 5%) quantile.

Monte Carlo Simulation Advantages

- Does not assume any specific models, or forms of distributions.
- Can incorporate any information, even subjective views.
- Can do scenario tests by twisting the parameters.
  - sensitivity analysis of parameters
  - give more weightings to more recent data
- Can work with non-linear assets, e.g., options.
- Can track path-dependence.

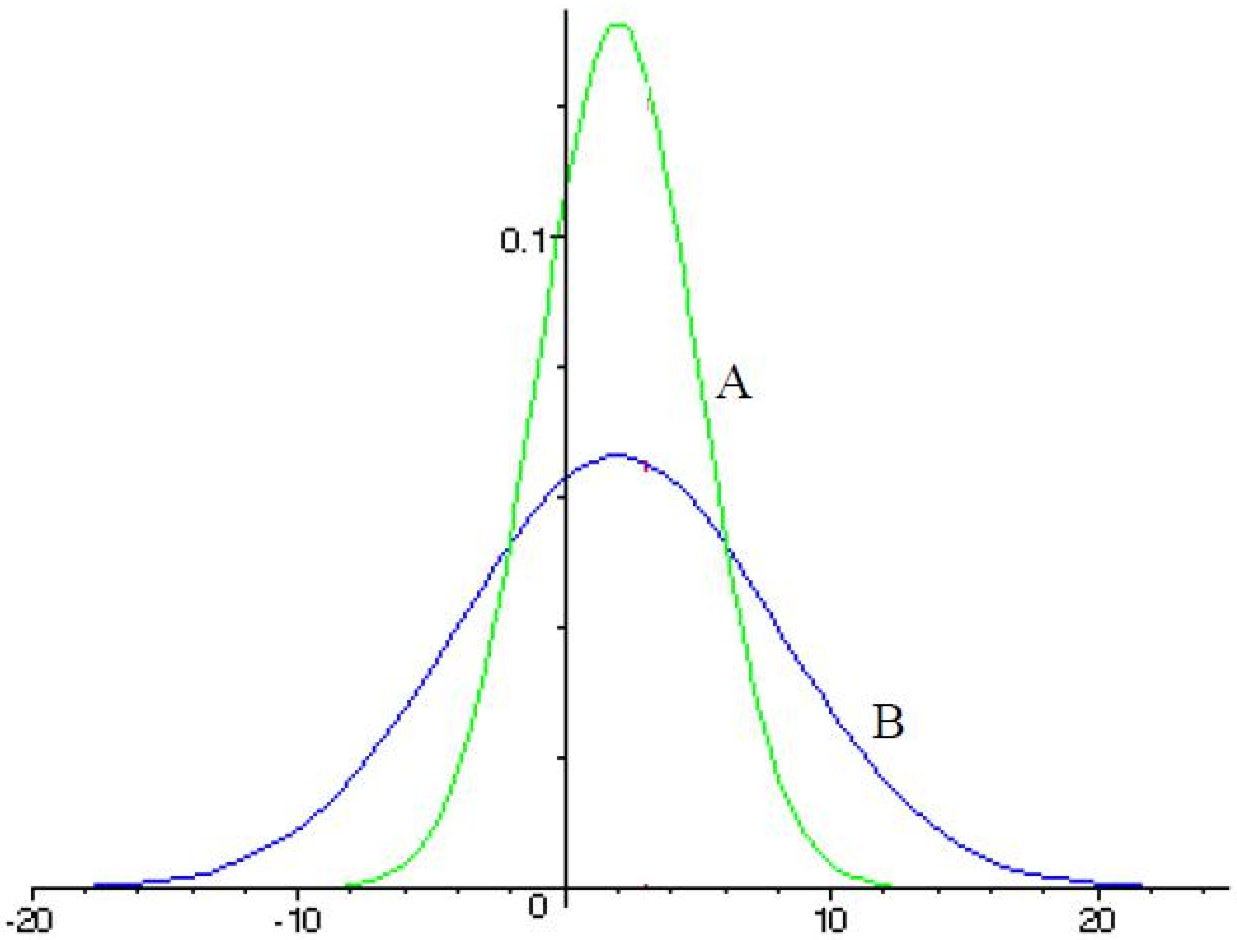
Monte Carlo Simulation Disadvantages

- Slow.
  - To increase the precision by a factor of 10, we must make 100 times more simulations.
  - Various variance reduction techniques apply.
    - antithetic variates
    - control variates
    - importance sampling
    - stratified sampling
- Difficult to build a (high) multi-dimensional joint distribution from data.

100-Year Market Crash

- How do we incorporate rare events into our returns distributions, hence enhanced risk management?
- Statistics works very well when you have a large amount of data.
- How do we analyze for (very) small samples?

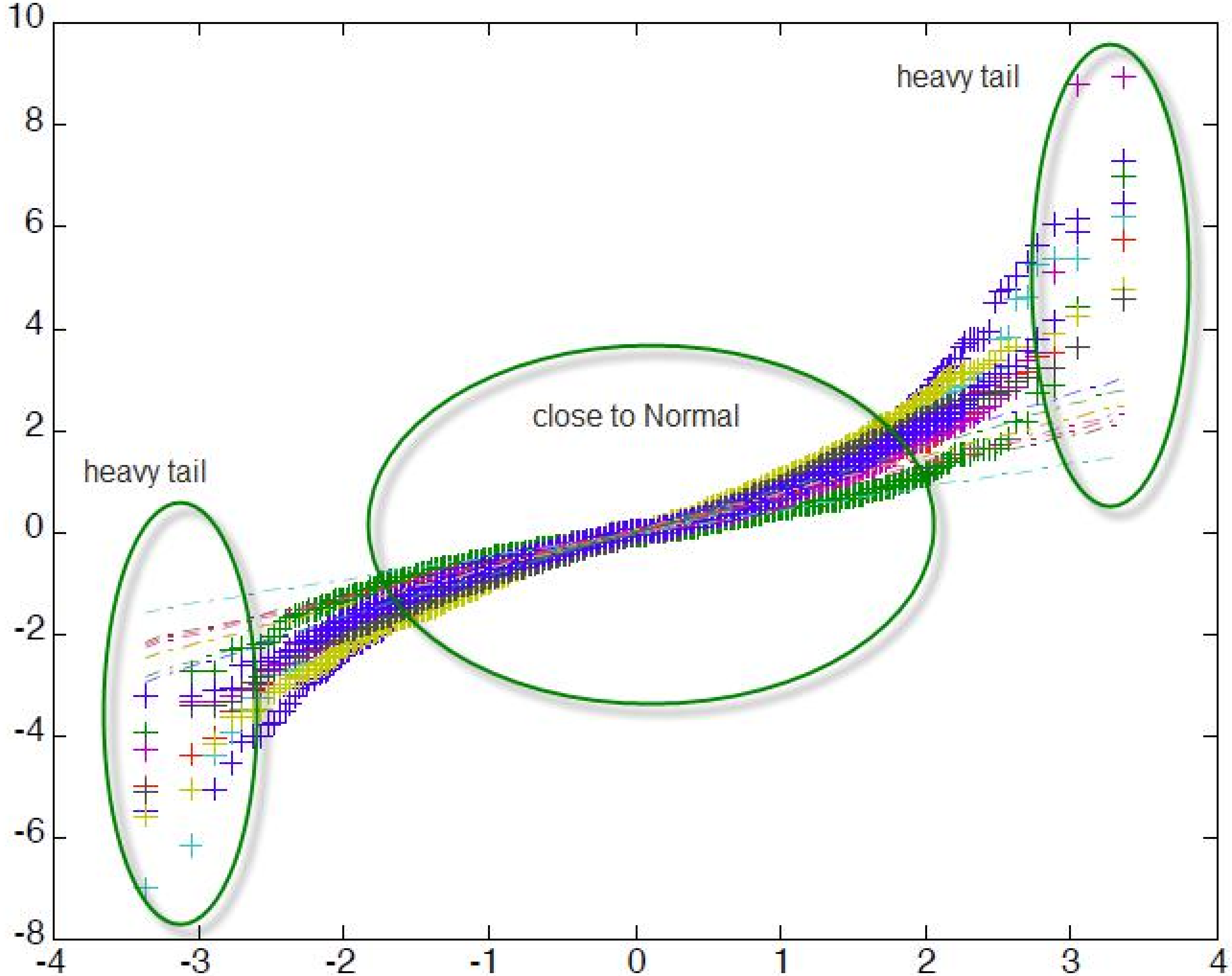
Fat Tails



QQ

- A QQ plots display the quintiles of the sample data against those of a standard normal distribution.
- This is the first diagnostic tool in determining whether the data have fat tails.

QQ Plot



heavy tail
close to Normal
heavy tail
10
8
6
4
2
0
-2
-4
-6
-8
-4
-3
-2
-1
0
1
2
3
4

Asymptotic Properties

- The (normalized) mean of a the sample mean of a large population is normally distributed, *regardless of the generating distribution*.
- What about the sample maximum?

Intuition

- Let $X_1, \ldots, X_n$ be i.i.d. with distribution $F(x)$.
- Let the sample maxima be $M_n = X_{(n)} = \max_i X_i$.
- $P(M_n \leq x) = P(X_1 \leq x, \ldots, X_n \leq x)$
- $= \prod_{i=1}^{n} P(X_i \leq x) = F^n(x)$
- What is $\lim\limits_{n \to \infty} F^n(x)$?

Convergence

- Suppose we can scale the maximums $\{c_n\}$ and change the locations (means) $\{d_n\}$.
- There may exist non-negative sequences of these such that
  - ${c_n}^{-1}(M_n - d_n) \to Y$, Y is not a point
  - $H(x) = \lim_{n\to\infty} P({c_n}^{-1}(M_n - d_n) \le x)$
  - $= \lim_{n\to\infty} P(M_n \le c_n x + d_n)$
  - $= \lim_{n\to\infty} F^n(c_n x + d_n)$

Example 1 (Gumbel)

- $F(x) = 1 - e^{-\lambda x}, x > 0.$
- Let $c_n = \lambda^{-1}, d_n = \lambda^{-1} \log n.$
- $P(\lambda(M_n - \lambda^{-1} \log n) \leq x)$
- $= P(M_n \leq \lambda^{-1}(x + \log n))$
- $= \left(1 - e^{-(x+\log n)}\right)^n$
- $= \left(1 - \frac{e^{-x}}{n}\right)^n$
- $\to e^{-e^{-x}} = e^{-e^{-x}} 1_{\{x>0\}}$

Example 2 (Fre´chet)

- $F(x) = 1 - \frac{\theta^\alpha}{(\theta+x)^\alpha} = 1 - \frac{1}{\left(1+\frac{x}{\theta}\right)^\alpha}, x > 0.$
- Let $c_n = \theta n^{\frac{1}{\alpha}}, d_n = 0.$
- $P\left(\vartheta^{-1} n^{-1/\alpha} M_n \leq x\right)$
- $= P\left(M_n \leq \vartheta n^{1/\alpha} x\right)$
- $= \left(1 - \frac{1}{\left(1+n^{1/\alpha} x\right)^\alpha}\right)^n \sim \left(1 - \frac{1}{\left(n^{1/\alpha} x\right)^\alpha}\right)^n$
- $= \left(1 - \frac{x^{-\alpha}}{n}\right)^n$
- $\rightarrow e^{-x^{-\alpha}} 1_{\{x>0\}}$

Fisher-Tippett Theorem

- It turns out that H can take only one of the three possible forms.
- Fre´chet
  - $\Phi_\alpha(x) = e^{-x^{-\alpha}} 1_{\{x>0\}}$
- Gumbel
  - $\Lambda(x) = e^{-e^{-x}} 1_{\{x>0\}}$
- Weibull
  - $\Psi_\alpha(x) = e^{-(-x)^\alpha} 1_{\{x<0\}}$

Maximum Domain of Attraction

- Fre´chet
  - Fat tails
  - E.g., Pareto, Cauchy, student t,
- Gumbel
  - The tail decay exponentially with all finite moments.
  - E.g., normal, log normal, gamma, exponential
- Weibull
  - Thin tailed distributions with finite upper endpoints, hence bounded maximums.
  - E.g., uniform distribution

Why Fre´chet?

- Since we care about fat tailed distributions for financial asset returns, we rule out Gumbel.
- Since financial asset returns are theoretically unbounded, we rule out Weibull.
- So, we are left with Fre´chet, the most common MDA used in modeling extreme risk.

Fre´chet Shape Parameter

- α is the shape parameter.
- Moments of order r greater than α are infinite.
- Moments of order r smaller than α are finite.
  - Student t distribution has $\alpha \geq 2$. So its mean and variance are well defined.

Fre´chet MDA Theorem

- $F \in$ MDA H, H Fre´chet if and only if
- the complement cdf $\bar{F}(x) = x^{-\alpha} L(x)$
- L is slowly varying function
  - $\lim_{x \to \infty} \frac{L(tx)}{L(x)} = 1,\ t > 0$
- This restricts the maximum domain of attraction of the Fre´chet distribution quite a lot, it consists only of what we would call heavy tailed distributions.

Generalized Extreme Value Distribution (GEV)

- $H_\tau(x) = e^{-(1+\tau x)^{-\frac{1}{\tau}}}, \tau \neq 0$
- $H_\tau(x) = e^{-e^{-x}}, \tau = 0$
  - $\lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^{-n} = e^{-x}$
- tail index $\tau = \frac{1}{\alpha}$
- Fre´chet: $\tau > 0$
- Gumbel:$\tau = 0$
- Weibull: $\tau < 0$

Generalized Pareto Distribution

- $G_\tau(x) = 1 - (1 + \tau x)^{-\frac{1}{\tau}}$
- $G_0(x) = 1 - e^{-x}$
  - simply an exponential distribution
- Let $Y = \beta X$, $X \sim G_\tau$.
- $G_{\tau,\beta} = 1 - \left(1 + \tau \frac{y}{\beta}\right)^{-\frac{1}{\tau}}$
- $G_{0,\beta} = 1 - e^{-\frac{y}{\beta}}$

The Excess Function

- Let u be a tail cutoff threshold.
- The excess function is defined as:
  - $F_u(x) = 1 - \bar{F}_u(x)$
  - $\bar{F}_u(x) = P(X - u > x | X > u) = \frac{P(X>u+x)}{P(X>u)} = \frac{\bar{F}(x+u)}{\bar{F}(u)}$

Asymptotic Property of Excess Function

- Let $x_F = \inf\{x: F(x) = 1\}$.
- For each τ, $F \in \mathrm{MDA}(H_\tau)$, if and only if
  - $\lim_{u \to x_F^-} \sup_{0<x<x_F-u} \left|F_u(x) - G_{\tau,\beta(u)}(x)\right| = 0$
- If $x_F = \infty$, we have
  - $\lim_{u \to \infty} \sup_{x} \left|F_u(x) - G_{\tau,\beta(u)}(x)\right| = 0$
- Applications: to determine τ, u, etc.

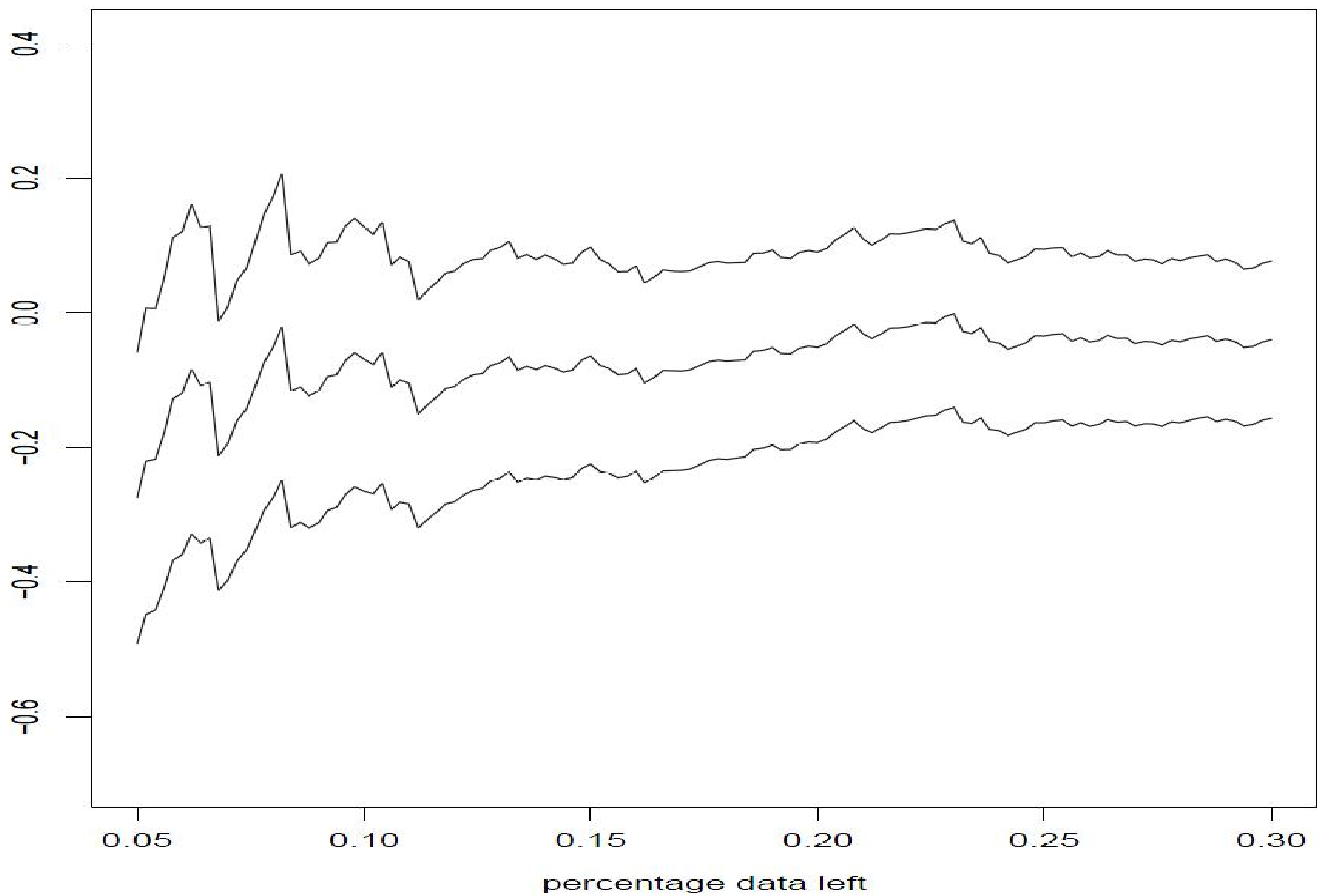
Tail Index Estimation by Quantiles

- Hill, 1975
- Pickands, 1975
- Dekkers and DeHaan, 1990

Hill Estimator

- ${\tau_{n,m}}^H = \frac{1}{m-1}\sum_{i=1}^{m-1}\left(\ln {X^*}_i - \ln {X^*}_{n-m,n}\right)$
- X^*: the order statistics of observations
- m: the number of observations in the (left) tail
- Mason (1982) shows that ${\tau_{n,m}}^H$ is a consistent estimator, hence convergence to the true value.
- Pictet, Dacorogna, and Muller (1996) show that in finite samples the expectation of the Hill estimator is biased.
- In general, bigger (smaller) m gives more (less) biased estimator but smaller (bigger) variance.

POT Plot



Pickands Estimator

- $\tau_{n,m}{}^{P} = \dfrac{\ln(X^{*}{}_{m} - X^{*}{}_{2m})/(X^{*}{}_{2m} - X^{*}{}_{4m})}{\ln 2}$

Dekkers and DeHaan Estimator

- $\tau_{n,m}{}^{D} = \tau_{n,m}{}^{H} + 1 - \frac{1}{2}\left(1 - \frac{\left(\tau_{n,m}{}^{H}\right)^{2}}{\tau_{n,m}{}^{H2}}\right)^{-1}$
- $\tau_{n,m}{}^{H2} = \frac{1}{m-1}\sum_{i=1}^{m-1}(\ln X^{*}{}_{i} - \ln X^{*}{}_{m})^{2}$

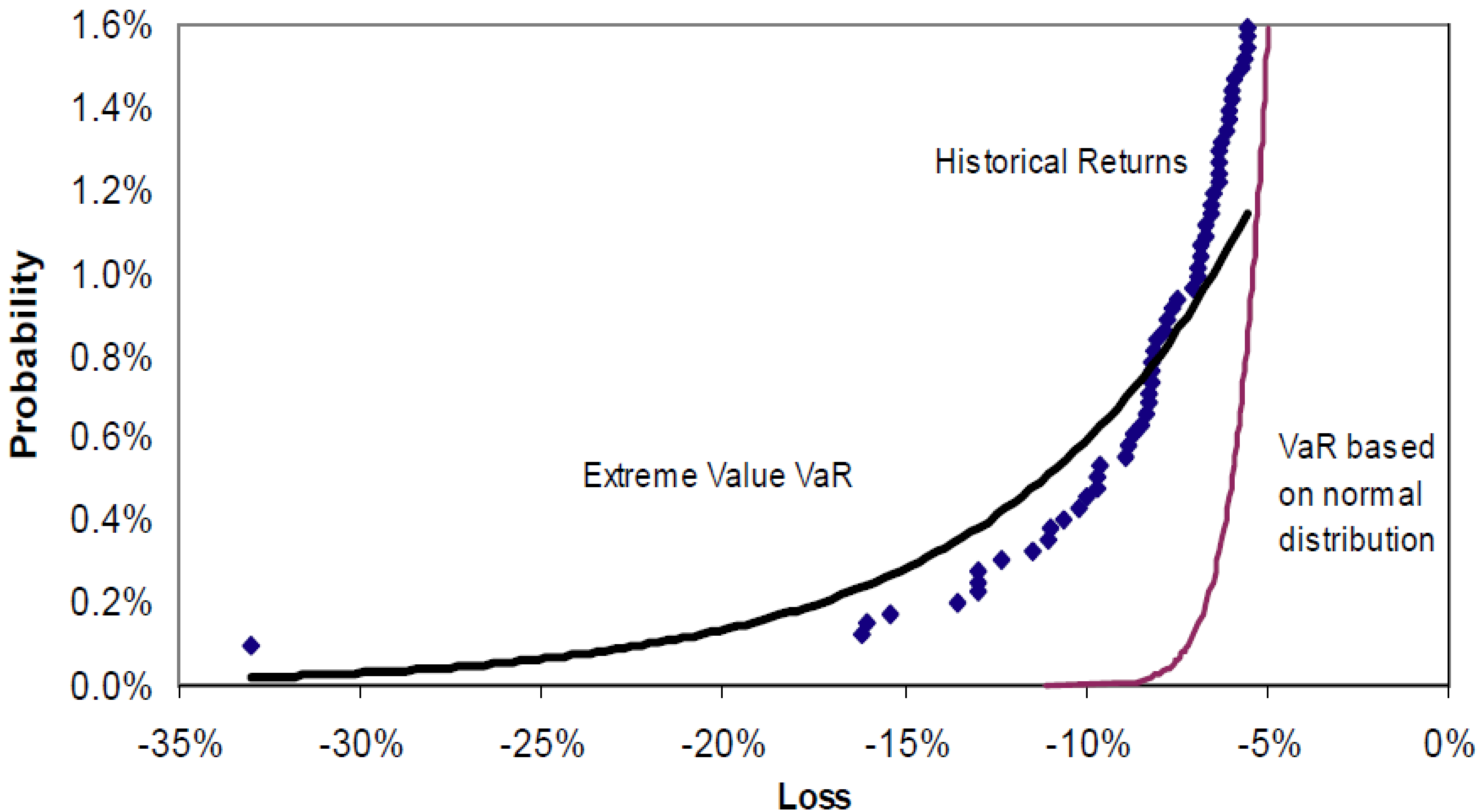
VaR using EVT

- For a given probability $q > F(u)$ the VaR estimate is calculated by inverting the excess function. We have:
- $\widehat{\mathrm{VaR}_q} = u + \frac{\hat{\beta}}{\hat{\tau}}\left(\left(\frac{n}{m}(1-q)\right)^{-\hat{\tau}} - 1\right)$
- Confidence interval can be computed using profile likelihood.

ES using EVT

- $\widehat{\mathrm{ES}}_q = \frac{\widehat{\mathrm{VaR}}_q}{1-\hat{\tau}} + \frac{\hat{\beta}-\hat{\tau}u}{1-\hat{\tau}}$

VaR Comparison



- http://www.fea.com/resources/pdf/a_evt_1.pdf