

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

- $r = (r_1, \ldots, r_n)'$: 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
- $x = (x_1, \ldots, x_n)'$: 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.

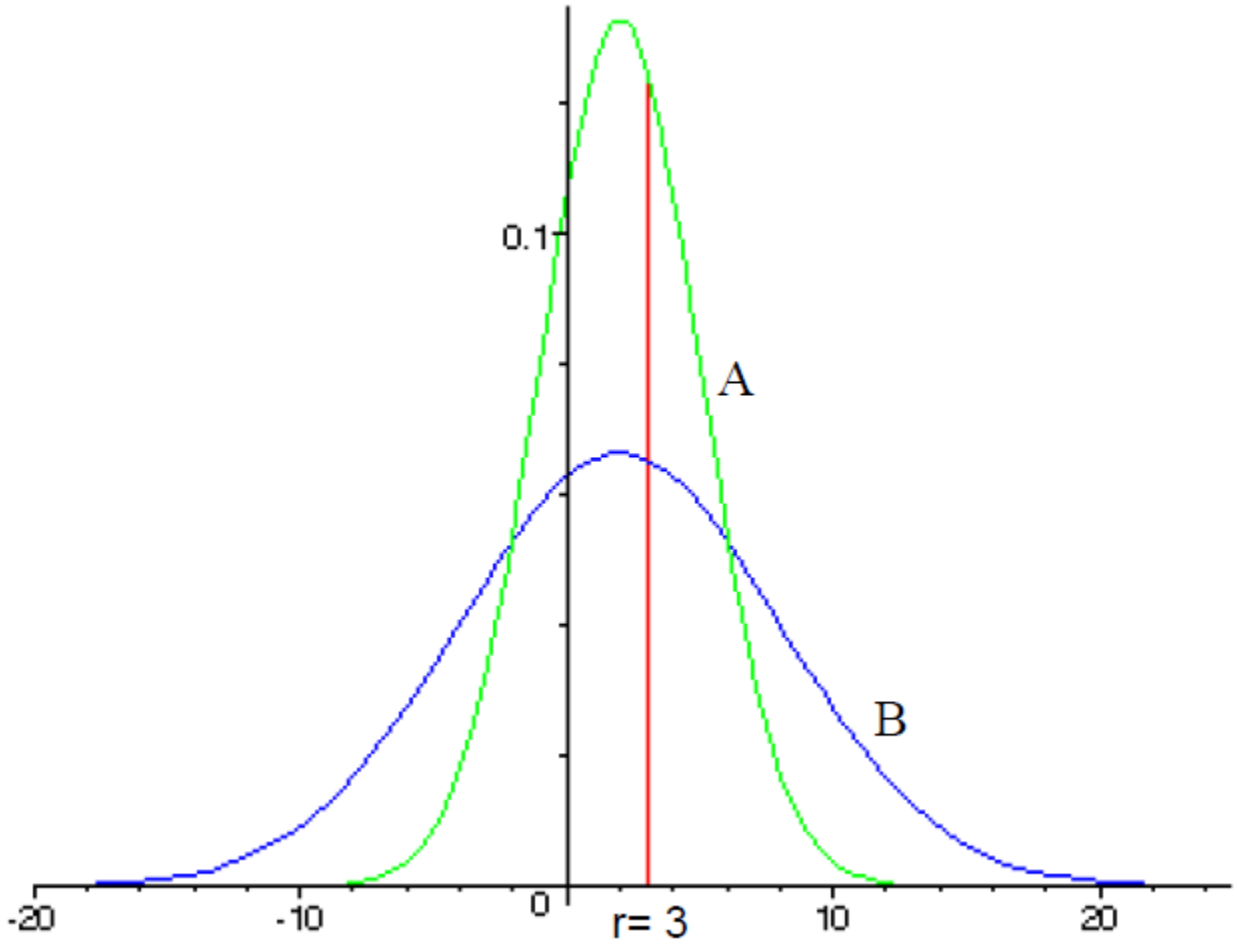


Potential for Gains

- 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.

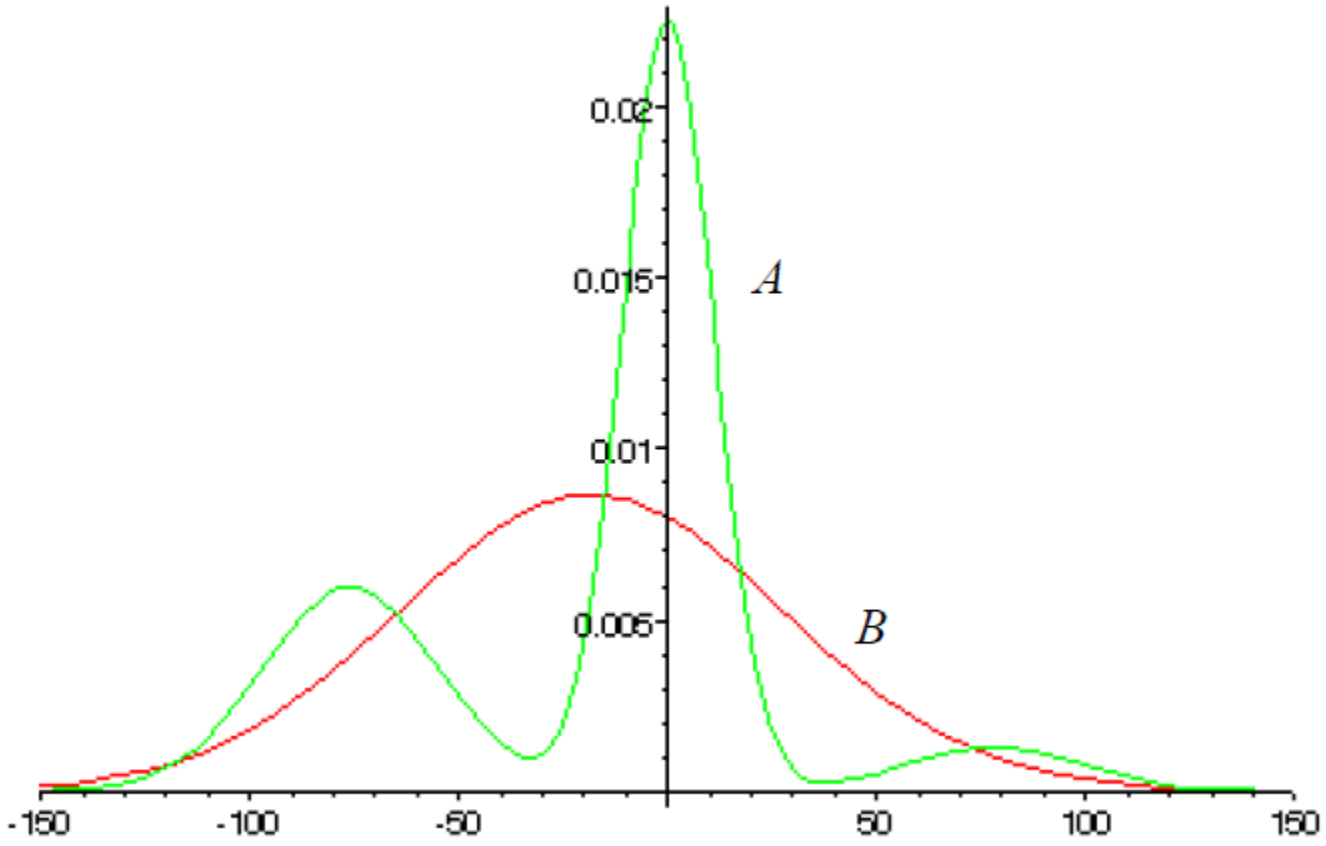


Example 1: L = 1

- 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!!!

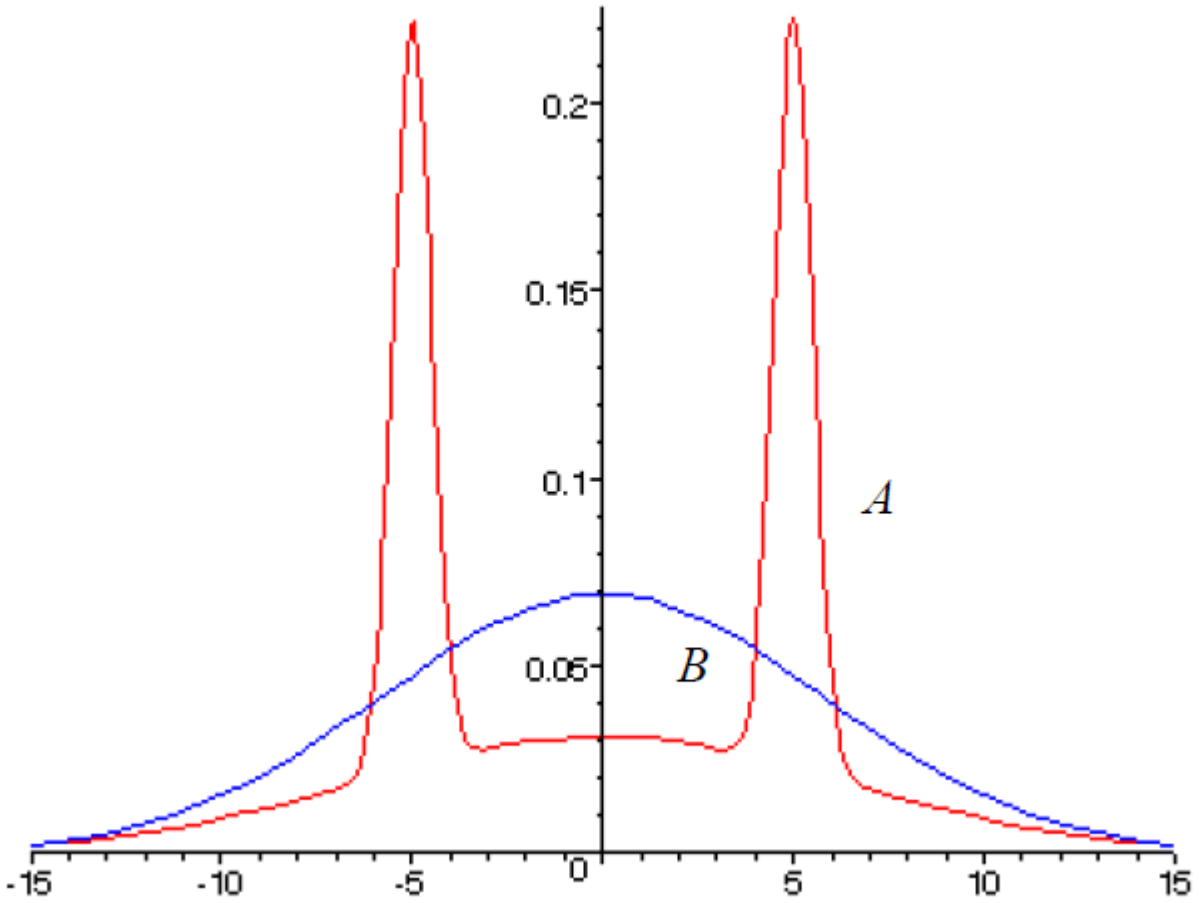


Higher Moments Are Important

- 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

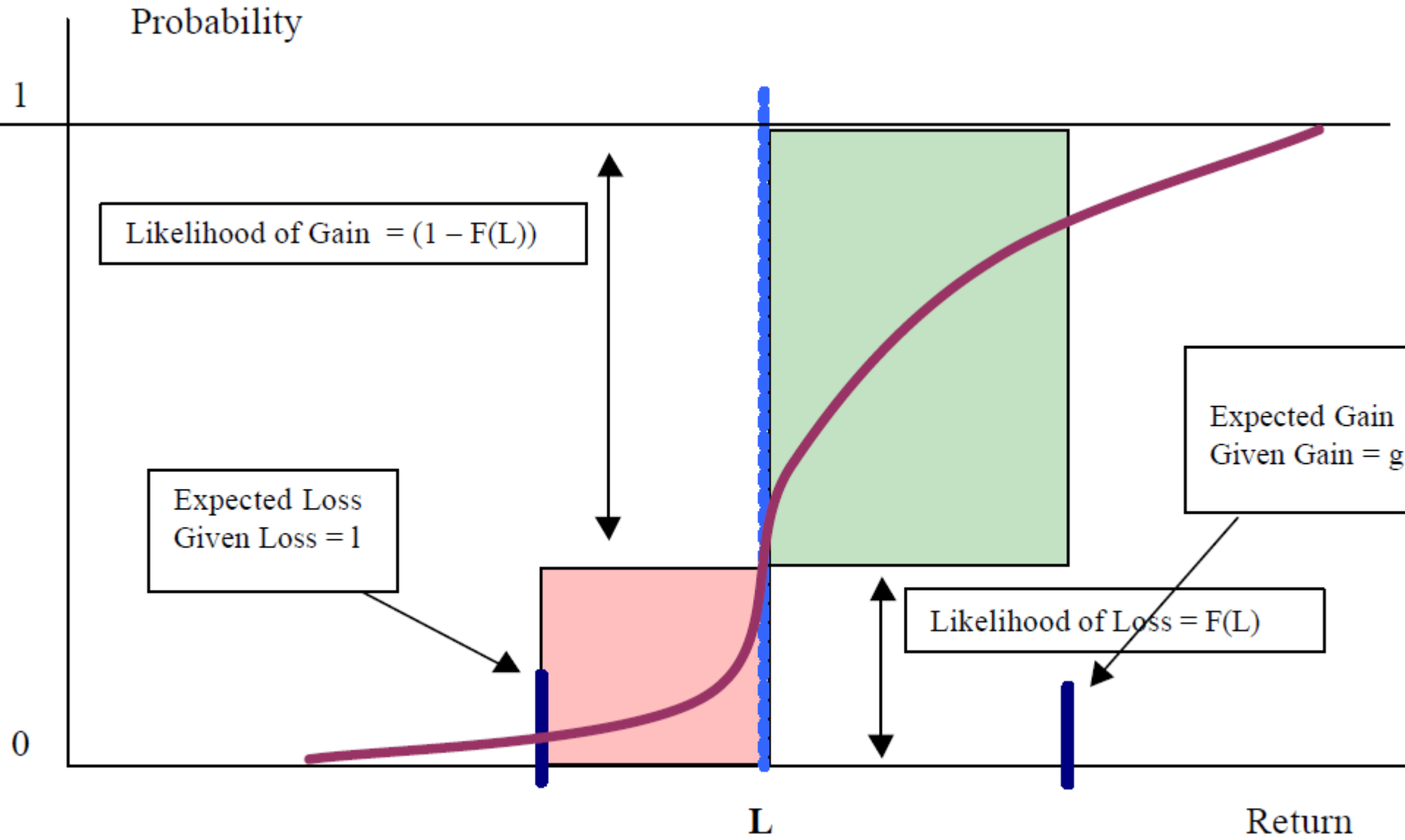


An Attempt

- 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

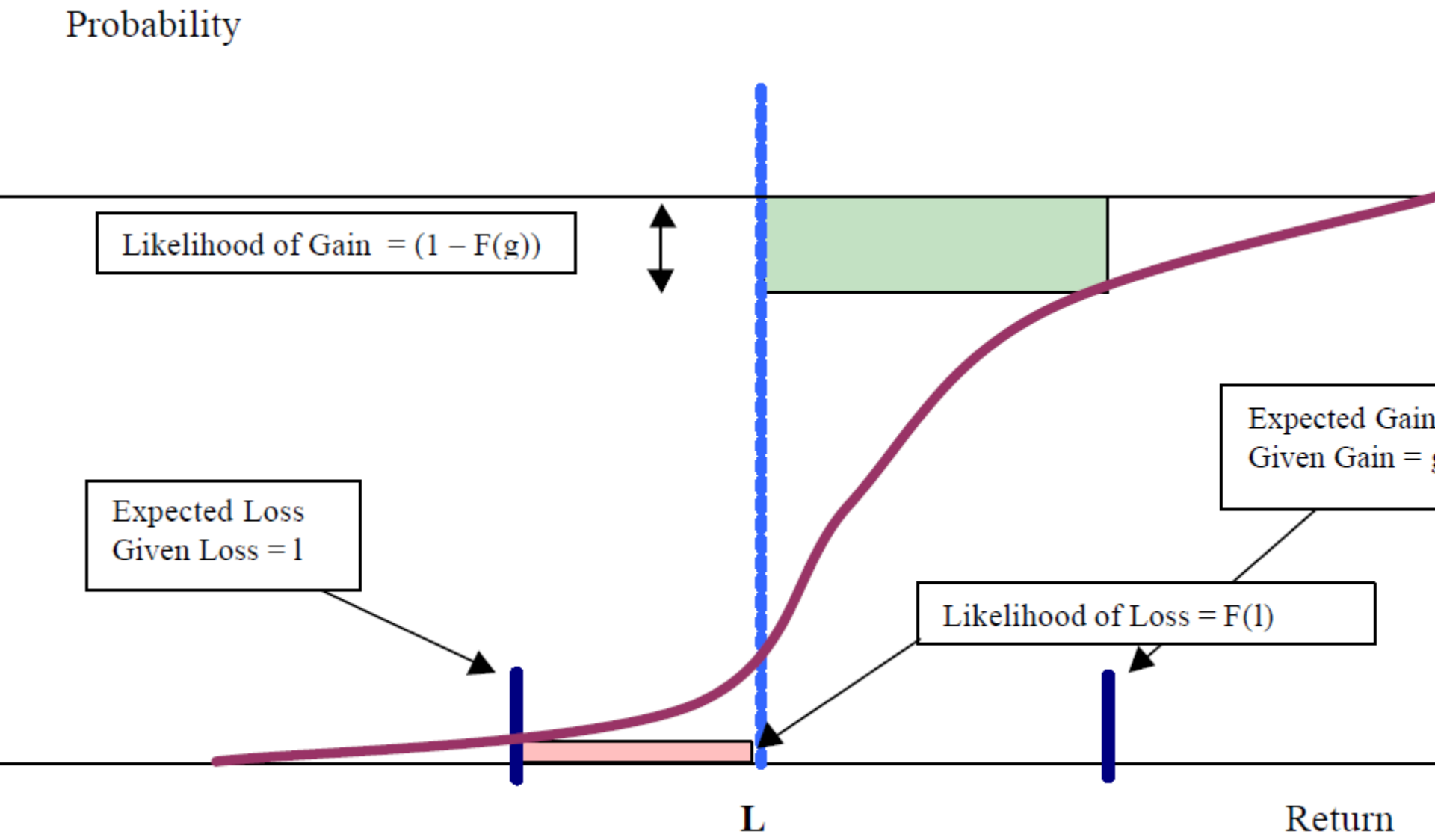


First Attempt Inadequacy

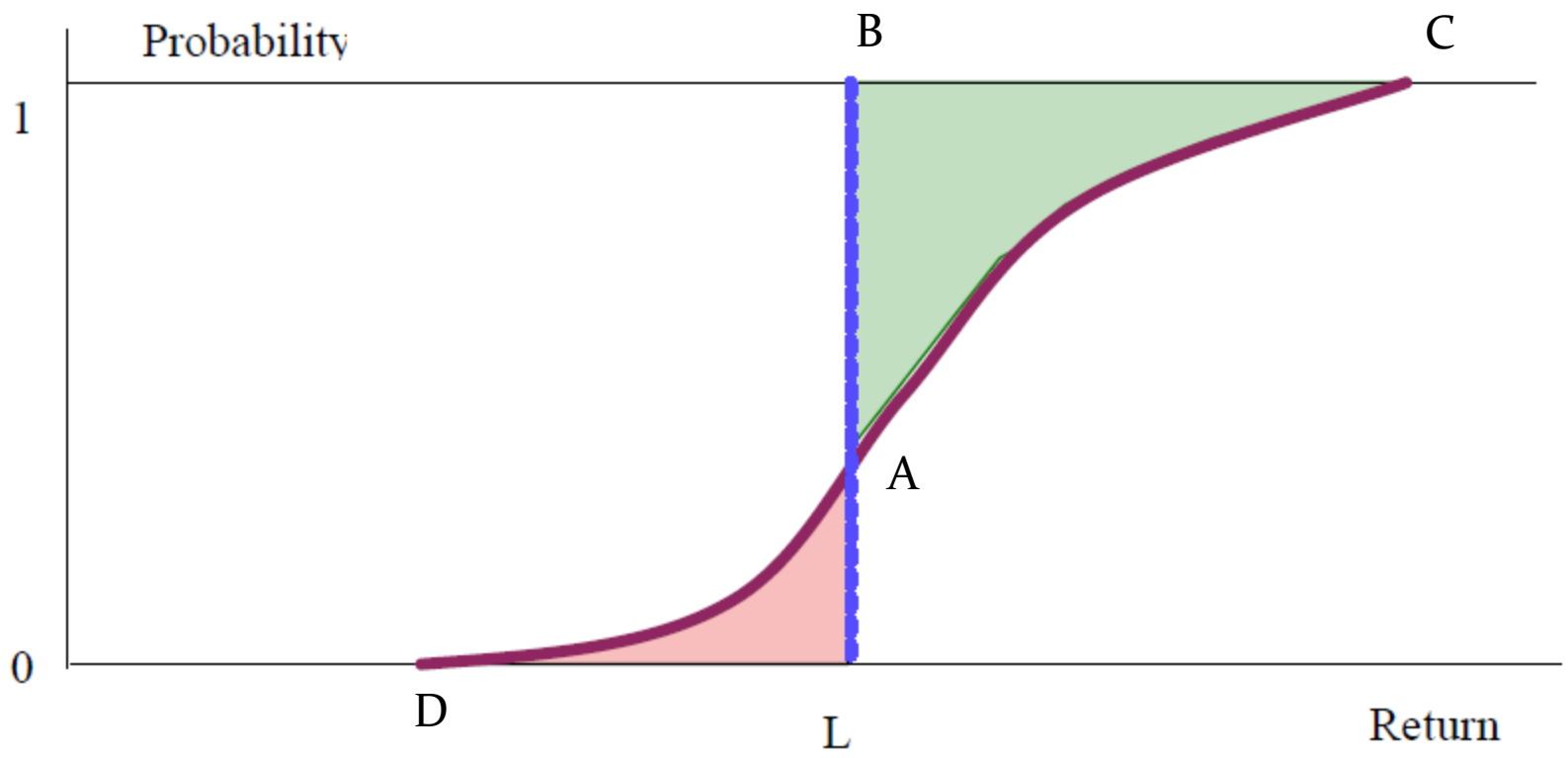
- Why F(L)?
- Not using the information from the entire distribution.
  - hence ignoring higher moments



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.

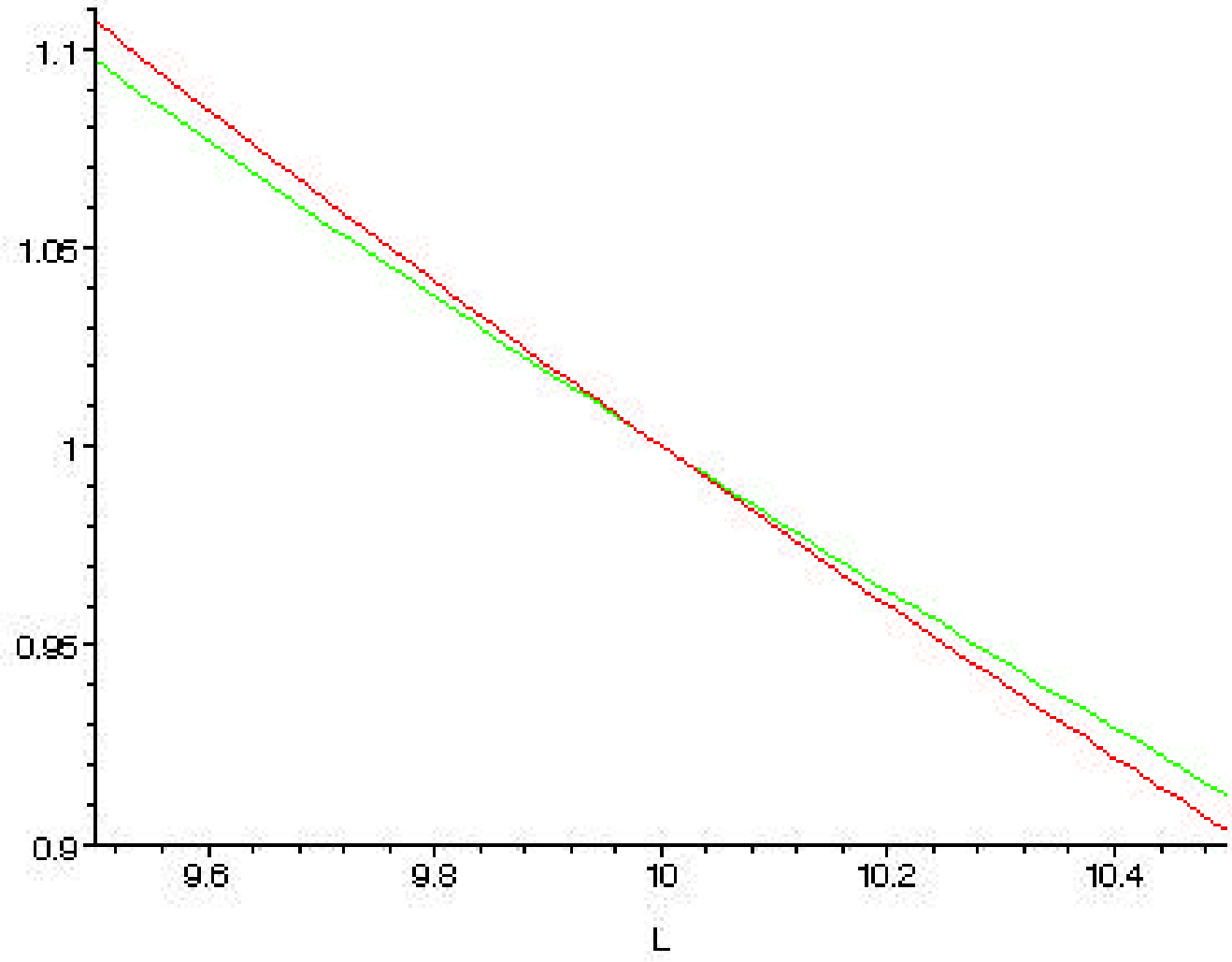


Omega Advantages

- 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)

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



Numerator Integral (3)

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

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)

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



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

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



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: $x^l = \left({x_1}^l, \ldots, {x_n}^l\right)'$

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



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…

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

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: $x^l = \left({x_1}^l, \ldots, {x_n}^l\right)'$



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
  - $h: X \rightarrow R, X \in R^n$
- 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.



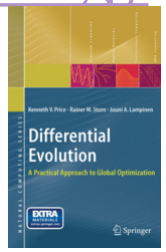
Search

- $x_{i+1} \in N_{x_i}$ (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?
- $\mathrm{ES}_q = E\left[X | X > \mathrm{VaR}_q\right]$



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.

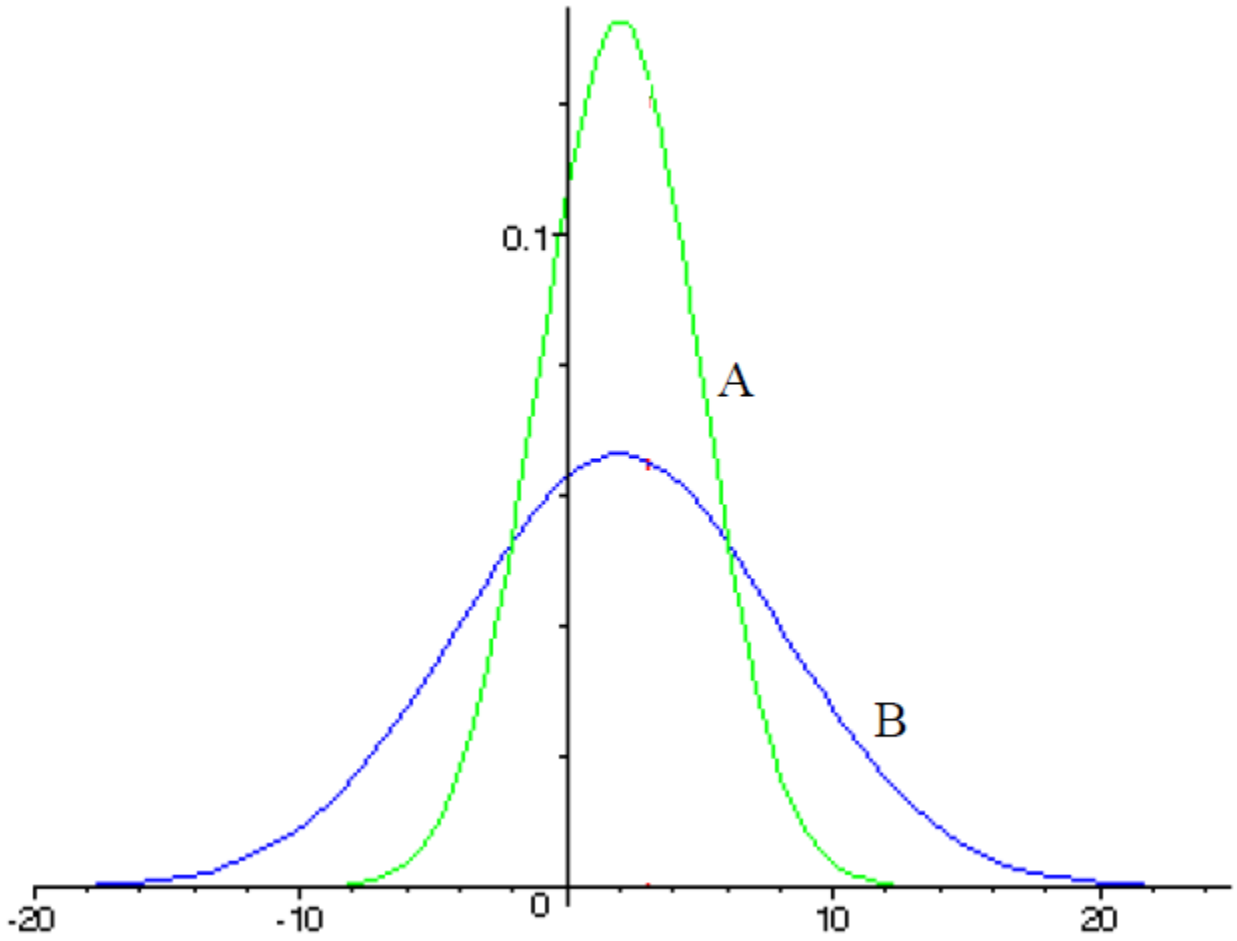


100-Year Market Crash

- 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

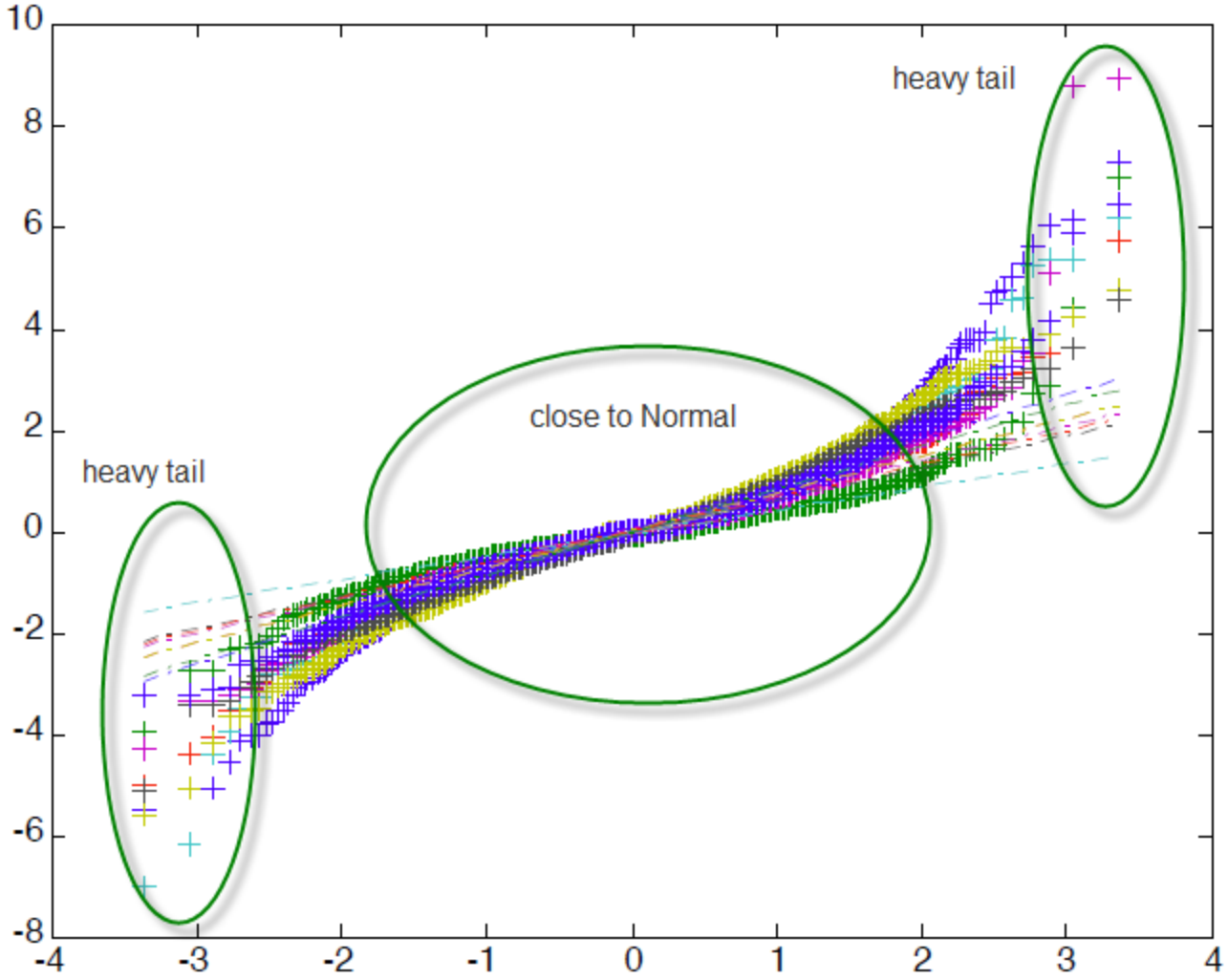


QQ

- 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 $\lim_{n \to \infty} F^n(x)$?



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
  - $H(x) = \lim_{n\to\infty} P({c_n}^{-1}(M_n - d_n) \leq x)$
  - $= \lim_{n\to\infty} P(M_n \leq c_n x + d_n)$
  - $= \lim_{n\to\infty} F^n(c_n x + d_n)$



Example 1 (Gumbel)

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



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



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

- $F \in \text{MDA}\, H$, 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

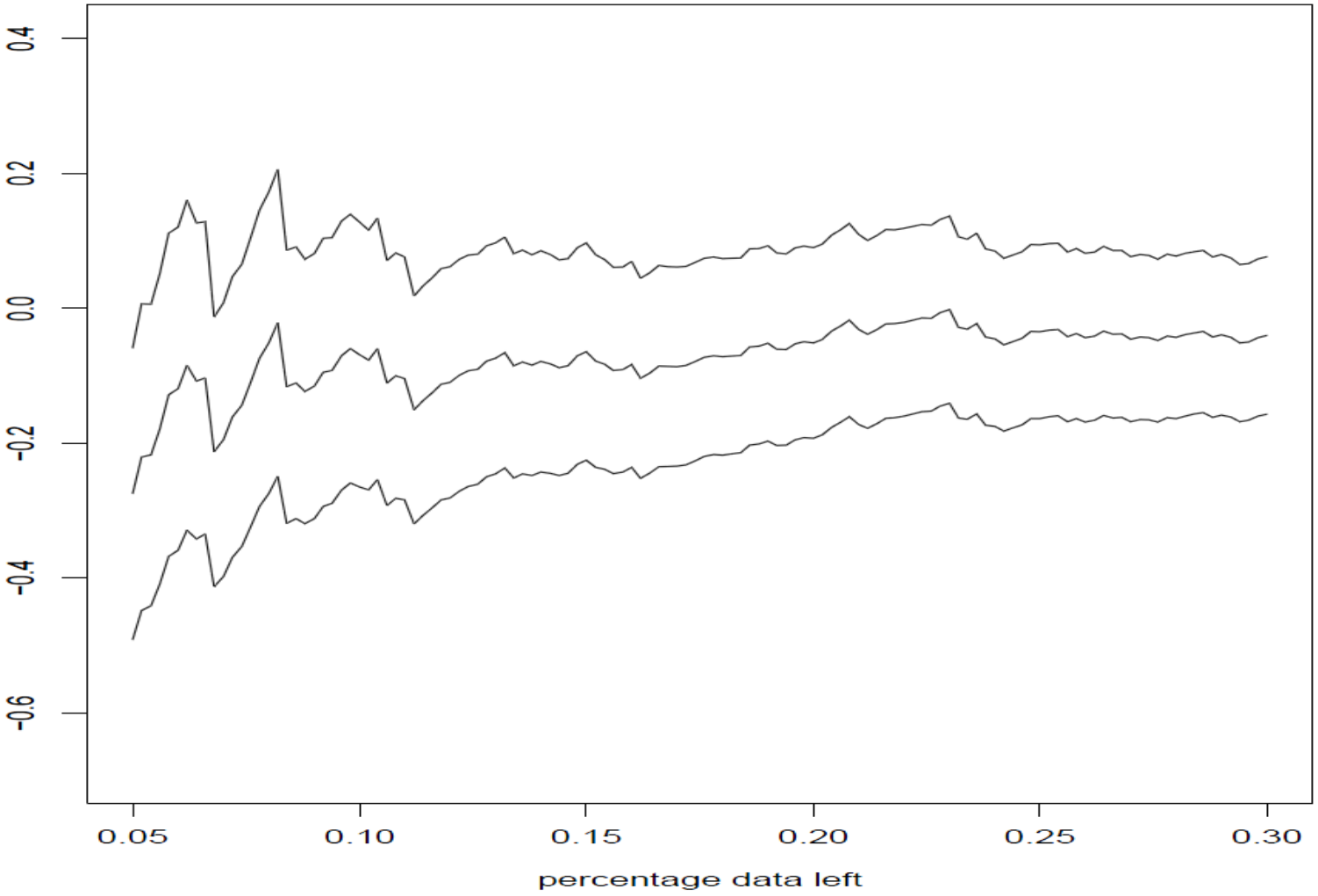


Hill Estimator

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

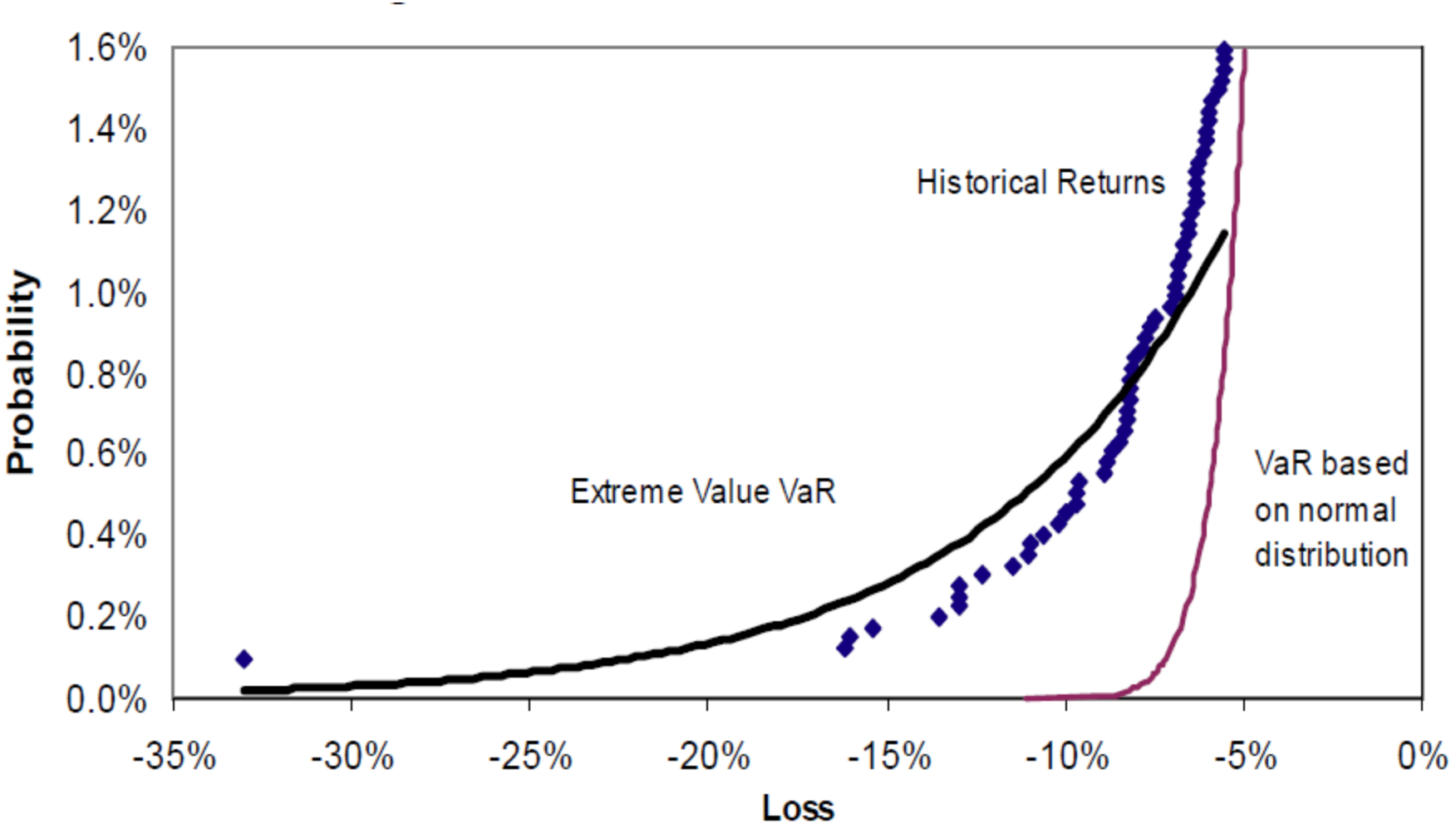


ES using EVT

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



VaR Comparison



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