Statistics in Finance

Arnab Kumar Laha

Indian Institute of Management, Ahmedabad

Outline of Talk

- A survey of statistical concepts applied to finance
- Some recent research:

1. Portfolio allocation with heavy tailed returns (Laha, Bhoumick & Subramaniam, 2005)
2. Multiple change point detection in NIFTY [Laha, 2004]
3. Regime switching behavior of Indian stocks using Hidden Markov Models [Laha & Keerthi Vasan, 2006]
4. Impact on macroeconomy of capital Markets [Laha & Chadha, 2006]
5. Determining relative market performance of selected Indian Firms [Laha, Bhargava, Verma, 2006]

What is Risk?

- Old Adage: No Risk, No Gain
- Gain = Return on Investment
- Return on investment is not fixed
- It varies over time
  - -prices of stock
  - -prices of real estates
  - -prices of precious metals
  - -Rs/$ exchange rate etc.

Variable Returns

- How do you measure variability of returns?
- Historical return data can be used to construct probability distribution of returns
- The expected value of the probability distribution of returns is called the Expected Return
- The standard deviation of the probability distribution of returns is a popular measure of RISK
- Most papers in financial literature use SD as a measure of risk
- Mandelbrot(1963) and Fama(1965) showed return distribution can be extremely variable and SD may not be a suitable measure of risk.
- One can use other measures of variability like MAD.

Diversification of Risk

- Suppose an investor intends to invest equal amounts in two stocks each having the same expected return
- He has two pairs of stocks (A,B) and (C,D). The pair (A,B) have correlation of 0.8 in their returns while the pair (C,D) have correlation -0.8 in their returns. Which one to choose?
- The pair (A,B) has risk $\sqrt{\sigma_1^2 + \sigma_2^2 + 1.6\sigma_1\sigma_2}$
- The pair (C,D) has risk $\sqrt{\sigma_1^2 + \sigma_2^2 - 1.6\sigma_1\sigma_2}$
- The pair (C,D) is better

Portfolio Theory

- Suppose we have certain amount of money to invest. Further, suppose we have identified two risky assets in which we may invest the money. How do we allocate the investment amount between these two assets?
- Goal : Maximize return subject to a certain maximum limit on risk
- Goal: Minimize risk subject to a certain minimum limit of return (such as the risk free rate)

Maximize $w_1\mu_1 + w_2\mu_2$

subject to

$$w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\sigma_1\sigma_2 \leq k,$$

$$w_1 + w_2 = 1, \quad w_1, w_2 \geq 0$$

Minimize $w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho\,\sigma_1\sigma_2$

subject to

$$w_1\mu_1 + w_2\mu_2 \geq c,$$

$$w_1 + w_2 = 1, \quad w_1, w_2 \geq 0$$

Portfolio Theory

- Markowitz(1952) showed how to choose the optimal portfolio
- Impact: Revolutionized finance
- Markowitz (along with Miller and Sharpe) was accorded the Nobel prize in Economics in 1990 for pioneering contributions to financial economics.

- Limitation: The theory needs that the stock returns to have finite expected return and risk

Photo courtesy: http://nobelprize.org/economics/laureates/1990/index.html

Portfolio Theory

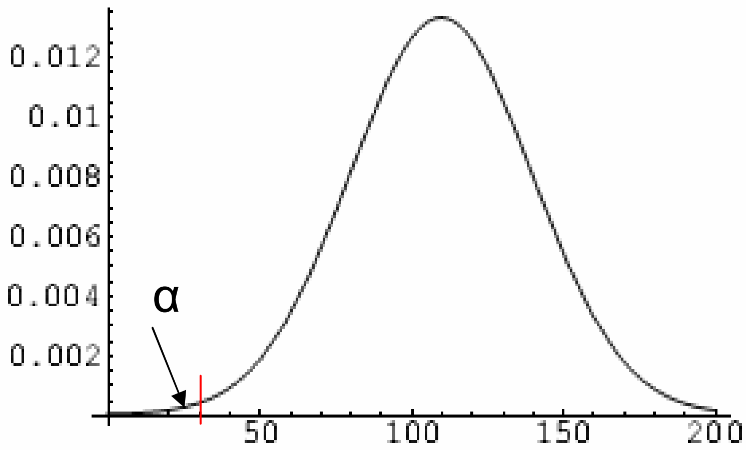
- Economists generally do not model economic decisions in terms of mean and variance
- They use a utility function instead
- A popular choice of utility function for this problem is

$$U(w_1, w_2) = \sum_{i=1}^{2} w_i \mu_i - \alpha\left(\sum_{i=1}^{2} w_i^2 \sigma_i^2 + 2 w_1 w_2 \rho \sigma_1 \sigma_2\right)$$

- The weights are chosen so as to maximize this utility function
- It can be shown that the same solution is obtained if the return distribution of the portfolio is assumed to be normal and a CES utility function is used (Uysal, 2001)

Value at Risk (VaR)

Value at Risk is defined to be a bound such that the loss over the horizon is less than this bound with probability α



Distribution of end-of-horizon portfolio value

(Most often the distribution is assumed to be Normal but non-Normal can also occur)

Portfolio Theory
Heavy Tailed Return Distribution

- Laha, Bhoumick and Subramaniam[2005] investigates the problem of portfolio allocation when the return distribution is heavy tailed
- A new utility function is constructed involving the median and MAD for this situation

$$G(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i R_i^M - \alpha \,\text{Median}\left(\left| \sum_{i=1}^{n} x_i r_i - \text{Median}\left(\sum_{i=1}^{n} x_i r_i \right) \right| \right)$$

  where R_i^M are median returns and the second term is the MAD of the return of the portfolio.
- The optimal portfolio is chosen so as to maximize the above criterion
- Analytical solution is not possible and one has to use numerical methods to obtain the efficient portfolio.

Portfolio Theory
Heavy Tailed Return Distribution

Table 1: Comparison of the N-method with the mean-variance method for two stocks

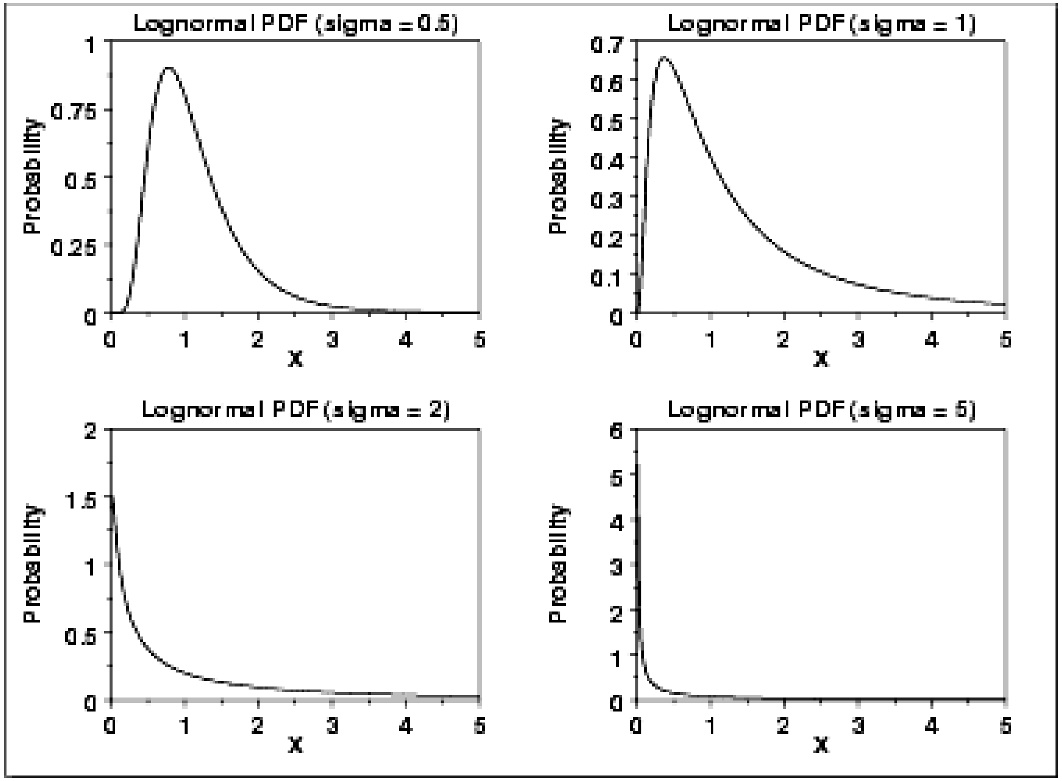
Distribution of return of Stock 1	Distribution of return of Stock 2	Superior Performance of new methodology over mean-variance approach	
		Growth of Wealth	VaR
Cauchy	Cauchy	84%	66%
Normal	Normal	8%	88%
Double Exp	Double Exp	8%	84%
Normal	Cauchy	94%	14%
Double Exp	Cauchy	42%#	100%

Criteria: (a) Growth of Wealth in 100 trading days

(b) VaR – 5 percentile of the distribution of the portfolio over 100 days

Lognormal Distribution

- Bachelier(1900) proposed that log returns follow a normal distribution
- This implies that the simple gross returns follow a lognormal distribution



Source: http://www.itl.nist.gov/div898/handbook/eda/section3/eda3669.htm

Brownian Motion

- Discovered by Robert Brown- a botanist, formalized by Albert Einstein – a physicist and rigorously analyzed by Norbert Weiner – a mathematician
- Bachelier(1900) was the first to propose the use of Brownian Motion in analyzing movements of stock prices.
- Black-Scholes assumed that the stock prices follow a Geometric Brownian Motion while deriving their celebrated option pricing formula.

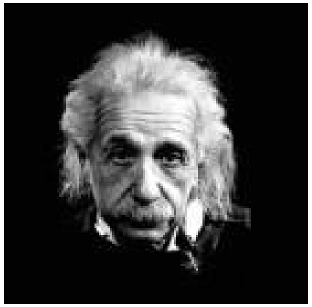


Photo Source: bradley.bradley.edu

Black-Scholes Option Pricing Formula

Scholes (along with Merton) was awarded the 1997 Nobel Prize in Economics

$$V = X_0 \Phi(g(t,x)) - Ke^{-rt}\Phi(h(t,x)) \text{ where}$$

$$g(t,x) = \frac{\ln(x/K) + (r + 0.5\sigma^2)t}{\sigma t^{1/2}},$$

$$h(t,x) = g(t,x) - \sigma t^{1/2} \text{ and}$$

Φ is the cdf of N(0,1)

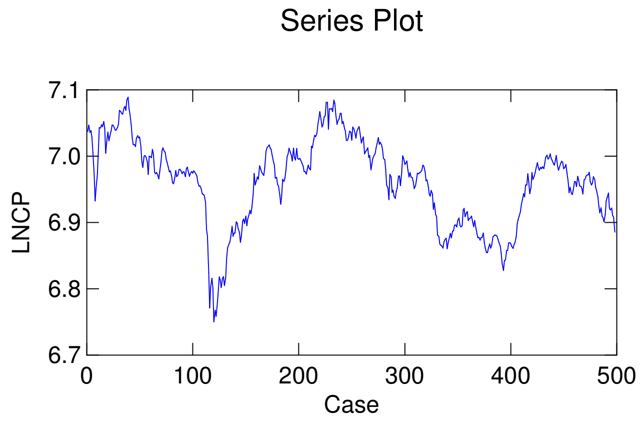
- An European Call Option gives the holder the right but not the obligation to buy a risky asset at a specified price K at a future time T.
- Question: What should be the rational price for such an option?
- No Arbitrage
- Self financing strategy
- Constant volatility

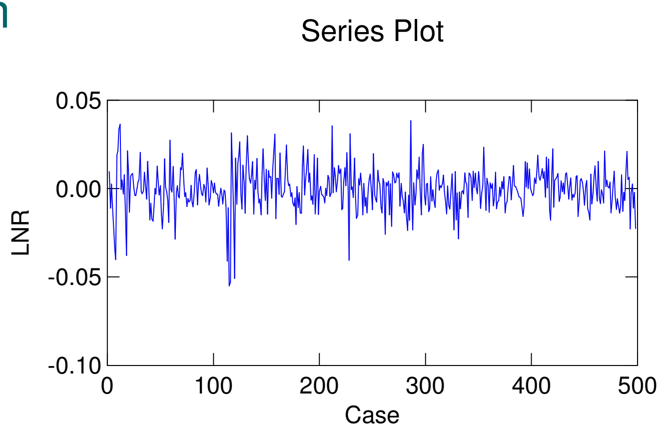
Detection of Volatility Regime Changes

- Laha(2004) discusses detection of volatility change point in a financial time series using the Binary Segmentation procedure [Vostriokova, 1981] and the LRT

- We consider the data pertaining to the daily closing values (S_t) of NIFTY from 1 April, 2001 to 31 March, 2003.
- We assume S_t follows a lognormal distribution
- The random variable $r_t = \ln S_t - \ln S_{t-1}$ follows a normal distribution with mean zero and s.d. σ. We call σ the volatility parameter.
- r_t's are assumed to be mutually independent random variables.

- We are interested to detect change points in volatility (if there are any).





Multiple Volatility Change Points

Table 1 Summary of Findings using the Binary Segmentation Procedure for detecting volatility change points in the NSE-NIFTY Stock Index 2001-2003

Segment	Change Point	p-value	Remarks
1-497	297	0.000	Significant
1-297	18	0.000	Significant
298-497	487	0.012	Not Significant
1-18	4	0.019	Not Significant
19-297	111	0.000	Significant
19-111	70	0.000	Significant
112-297	119	0.000	Significant
19-70	36	0.006	Significant
71-111	85	0.709	Not Significant
120-297	280	0.007	Significant
19-36	30	0.209	Not Significant
37-70	64	0.986	Not Significant
120-280	235	0.001	Significant
281-297	285	0.480	Not Significant
120-235	226	0.004	Significant
236-280	239	0.586	Not Significant
120-226	132	0.098	Not Significant

Period	Standard Deviation
1/4/01 - 30/4/01	0.023
1/5/01 - 25/5/01	0.008
26/5/01 - 12/7/01	0.013
13/7/01 - 11/9/01	0.006
12/9/01 - 21/9/01	0.033
22/9/01 - 27/2/02	0.012
28/2/02 - 12/3/02	0.020
13/3/02 - 17/5/02	0.009
18/5/02 - 11/6/02	0.018
12/6/02 - 31/3/03	0.009

Volatility Change Points

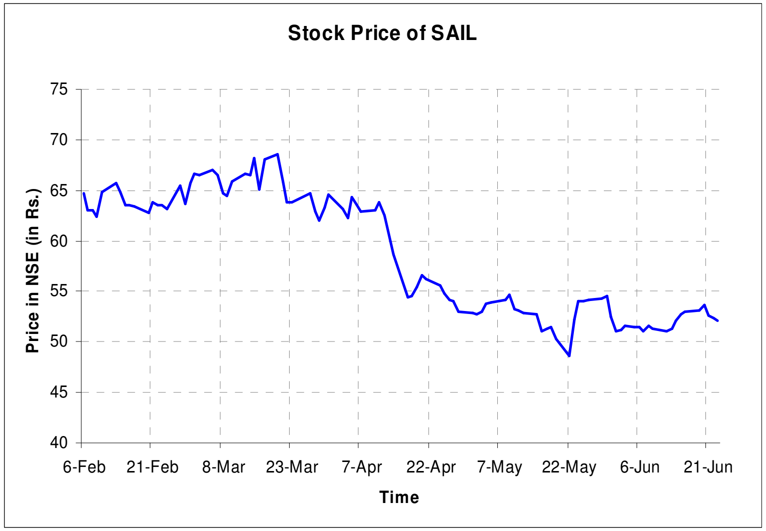
- The stock market goes through periods of high volatility followed by periods of low volatility.
- The sudden increase in volatility in April, 2001 can be attributed to the payment crisis and other market irregularities in the wake of Ketan Parekh stock scam.
- The high volatility of the period 12/9/2001 to 21/9/2001 can be attributed to the World Trade Center attack and uncertainties stemming from this incident
- The high volatility of the period 28/2/2002 to 12/3/2002 can be attributed to post-Godhra incidents.

Application of Hidden Markov Models to Regime Switching Behaviour

- Laha & Keerthi Vasan(2006) attempts to model the regime switching behaviour using HMM
- The basic idea is that the regime switching behaviour is governed by a Markov Chain which is not observed
- The advantage of this approach is that we can obtain predictive distributions of the future prices
- We take a Bayesian route using the Sampling Inspection Resampling methodology to derive the posterior distribution of the transition probability matrix.

Application

- Time variation of the stock price of SAIL during 6 Feb 2005 to 24 Jun 2005
- Assumption: The stock price is distributed as lognormal
- Expected value of posterior distribution of Initial mean =**64.8**,
- Same for the mean after the regime switch = **52.2**.
- Expected value of the posterior dsitribution of standard deviation is **0.0423**.
- The expected value of the mean after the first regime switch after 24th June 2005 is calculated to be **58.402**, a jump of **+6.202** from the existing mean of **52.2**.
- This could be done using the posterior estimated transition probability matrix.



Macroeconomy and Capital Markets

- Laha & Chadha (2006) studies the impact of macroeconomic variables on Indian capital markets along the lines of Chen(1991)
- The main statistical tools used in this study are regression analysis and time series analysis
- The findings are similar to that of Chen on the whole but differ in the details
- Few interesting findings:

  a) Dividend yield is an important explanatory variable for real return on the index

  b) Yield on Govt. securities seems to have some long range impact on the stock market

Predicting Market Performance of Companies

- Laha, Varma and Bhargava(2006) compares the present market performance of a firm with that four years later.
- In this small study 42 companies were included.
- The idea is to first identify some of the commonly used parameters which can be used as determinant for market performance measured through the price of their stocks.
- Using multiple regression analysis based on ten years data (1994-2003) these turn out to be P/E, P/B and EPS

Predicting Market Performance of Companies

- Using the values of these parameters in 1999 the 42 companies are divided into 2 clusters.
- Cluster 1: P/E = 131.5 P/B = 25.1 and EPS = 23.6 (n=5)

  Cluster 2: P/E = 23.3 P/B = 5.1 and EPS = 15.1 (n=37)
- The same companies are reclassified based on the 2003 values.

  Cluster 1: n =1; Cluster 2: n = 41
- The movement across clusters is :

		Cluster in 2003	
		1	2
Cluster in 1999	1	0	5
	2	1	36

- This can be used to estimate a transition probability matrix between the clusters

Other Applications

- Mohan, Jha, Laha and Dutta(2004) discusses the use of Aritificial Neural Networks for modeling the BSE Sensex
- Chandra, Shankar and Laha(2006) discusses application of using Chaos Theory in modeling Indian stock market.

Thank You

arnab@iimahd.ernet.in