Modelling Value at Risk

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Lakshman Teja M

Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

April, 2019

Supervisor: Uttara Naik-Nimbalkar

Certificate

This is to certify that this dissertation entitled Modelling Value at Risk towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents study/work carried out by Lakshman Teja M at Indian Institute of Science Education and Research under the supervision of Prof. Uttara Naik-Nimbalkar, Professor, Department of Mathematics , during the academic year 2018-2019.

U. Naik Nimbalkar

Prof. Uttara Naik-Nimbalkar

Committee:

Prof. Uttara Naik-Nimbalkar

Dr. Anindya Goswami

Dedicated to A & U

Declaration

I hereby declare that the matter embodied in the report entitled Modelling Value at Risk are the results of the work carried out by me at the Department of Mathematics, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Uttara Naik-Nimbalkar and the same has not been submitted elsewhere for any other degree.

Lakshman Teja M

Acknowledgments

I would like to thank my supervisor Prof. Uttara Naik-Nimbalkar for constant support during the course of the project. I express my sincerest gratitude to my TAC member Dr. Anindya Goswami for guiding me and instilling a fervour for learning. I aknowledge the role of IISER Pune community in nurturing me during the course of BS-MS.

Abstract

Various types of financial risks are studied and building blocks of finance are investigated. Time series analysis is studied with emphasis on financial data. Many models are simulated and forecasting is done. Financial risk is studied with measures to assess and manage it. Value at Risk is estimated using different techniques ranging from statistics to time-series analysis. Various models are compared and a deatiled analysis is provided comparing different models.

Contents

Chapter 1

Introduction

Financial crises have become common occurrences in economy after the advent of limited liability companies(LLC). There has been an increased incidence of crises since the beginning of 20th century with atleast one major financial crisis every decade. Recent disasters including-

- Oil price shocks starting in 1973
- Black Monday which wiped out capital of 1 trillion USD
- Japanese stock bubble of the 90s leading to a loss of nearly 3 trillion USD
- Asian turmoil in 1997 wiped out nearly three-fourths of equity capitalisation of the South East Asian economy
- Russian default of 1998 leading to a failure of Long Term Capital Management(LTCM)
- Housing credit crisis of the US starting in 2008 leading to a financial crisis comparable to only 1929 Great Depression

have led to an increase in emphasis on risk management and assessment.

1.0.1 Financial Risk

Risk refers to the probability of loss and exposure is the possibility of loss. The risk might come because of human actions such as business cycles, inflations, wars and governmental policies or natural calamities such as earthquakes and floods. Risk and willingness to take risk are paramount to the growth of an economy. Though most financial instruments have exposure, it could be turned into profit by proper exposure. Events with high probability usually have small returns(i.e. small returns are familiar) whereas those with low probabilities can have huge losses. We cannot always eliminate risk, but an understanding is necessary to manage it. There are many recipes of risk assessment and management, but they usually follow a similar framework which includes:

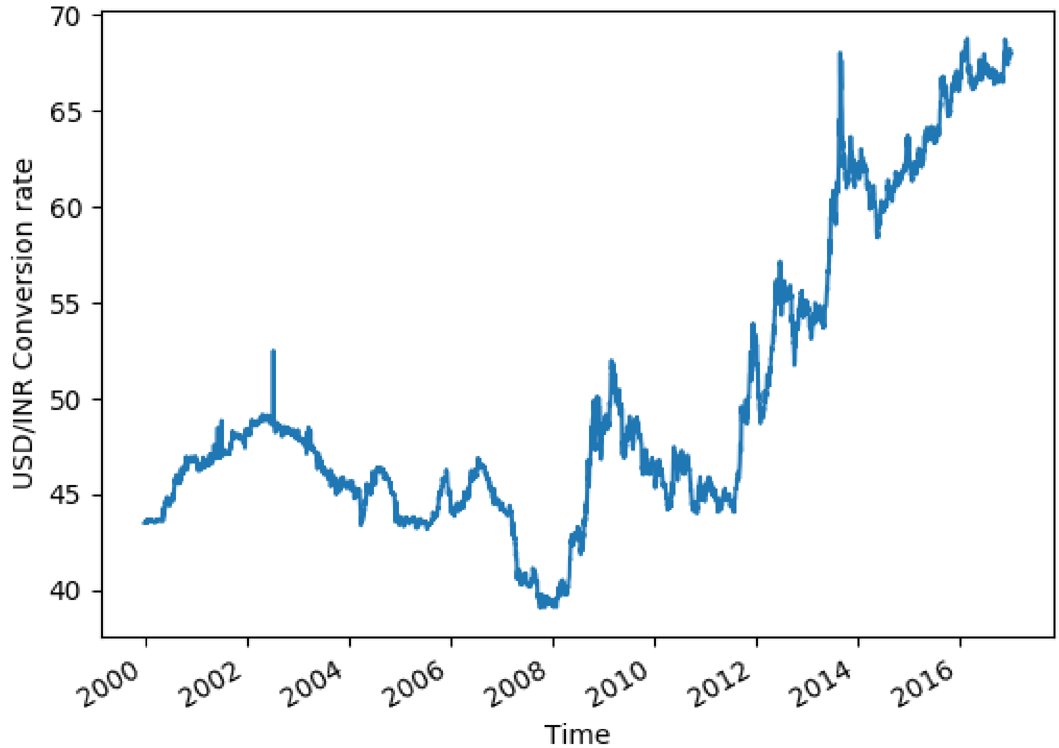
1. Identify and prioritise potential risks
2. Implement a risk management strategy for the appropriate level of tolerance
3. Measure, report, monitor and refine as needed.

Risk management could be done in various ways including-

1. Setting a threshold, called a stop-loss limit when a position is cut if the cumulative losses are more than the limit.
2. A notional amount through which we could assess the losses

The derivatives market has increased to 380 trillion USD, starting with futures in 1973. Derivative instruments are used to hedge against potential losses, but they can become the very cause of disasters without proper regulations. The Great Recession(2008) which has its roots in deregulated credit default swaps is a perfect example of this. The risk is an inherent consequence of decisions a company takes. It is broadly classified into two types *business risk* and *financial risk*. Firms willingly assume business risks like investment decisions, marketing strategies and operational structure to grow and add value to shareholders, which are necessary for the proper functioning of a firm. Risks occurring by movements in financial markets like changes in interest rates and defaults on debts are called financial risks. Most companies these days are involved in financial markets either directly through investment subsidiaries(like General Capital and Ford) or investments in financial instruments.

After the abolishment of fixed rate exchange system, currencies have become more volatile than ever. The movement of the Indian rupee against the dollar is seen in the figure.



Investments can have different risks -

- **Market risk**- Risk due to changes or movements of markets. Markets could be stock exchanges where many stocks are traded over a formalised system or trades between individuals.

  Absolute risk: Measured in terms of volatility of returns

  *Relative risk:*Measured in deviation from a benchmark

  Convexity risk: Due to the duration of the investment

  Volatility Risk: Due to changes in implied volatility of the assets

  Discount rate risk: Due to the choice of choosing a discount rate in calculating future prices of the portfolio

- **Credit risk**-Risk when an organisation is owed money or is dependent on other institution for payment that is unable or unwilling to meet its contractual obligations. It should be defined as a potential loss in mark to market value incurred during a *credit event*, which occurs when there is a change in a party's ability to meet obligations. Types of credit risk are default risk, pre-settlement risk, sovereign risk etc.,

- **Operational risk**- Risk associated with inadequate or failed investments, people or system failures from internal or external events. It could be classified into model risk, people risk and legal risk.

- **Liquidity risk**- When a corporation is not able to sell or purchase security to meet its short term goals.

  Asset-liquidity risk also called market/product-liquidity risk occurs when a transaction is possible at existing market rates because of the size of the position compared to normal trading lots.

  Funding liquidity risk also known as cash-flow risk occurs when there is early liquidation to meet financial obligations. Both interact when illiquid assets have to be sold at less than the fair market price.

Market risk is of four types: interest rate risk, equity risk, exchange rate risk and commodity risk. The risk is measured by the standard deviation of unexpected outcomes also called volatility(σ). It can be due to the volatility of financial instruments and exposure to such risk. Almost nothing can be done about the volatility of financial assets, but exposure can be hedged with derivatives. First-order measurements of exposure are known by different names-

- In stock market, exposure is called *systemic risk* or β.

- In options market, exposure to movements in underlying asset's price is called delta(δ)

- Movements of interest rates in fixed income instruments is *duration*

Second-order exposures are called convexity and gamma(γ) in fixed income and options markets respectively.

Because of the existence of various types and factors of risk, there exist many risk measures. Value at Risk **VaR** though initially developed for market risk, it is now a statistical measure common to all kinds of risk. VaR is the maximum loss that could occur at a given confidence interval and time horizon. It is quantile of the profit and loss(**P/L**) distributions for a given time horizon. Given α is the confidence level, then VAR is the $1-\alpha$ lower tail value. A higher confidence level will give fewer cases of losses greater than VaR, but it will increase the amount of VaR. Risk increases with time; hence a longer time horizon will have a larger VaR. It accounts for leverage and diversification effects. VaR is an estimate and should be supplemented by stress tests, controls and limits for a reliable measure.

1.0.2 Returns

Let S_t be the price of a stock at time t. The returns from the stock for holding it from time $t-1$ to t is

$$R_t = \frac{S_t}{S_{t-1}}.$$

To incorporate continuous compounding we use log returns,r_t

$$r_t = \ln \frac{S_t}{S_{t-1}}$$

.

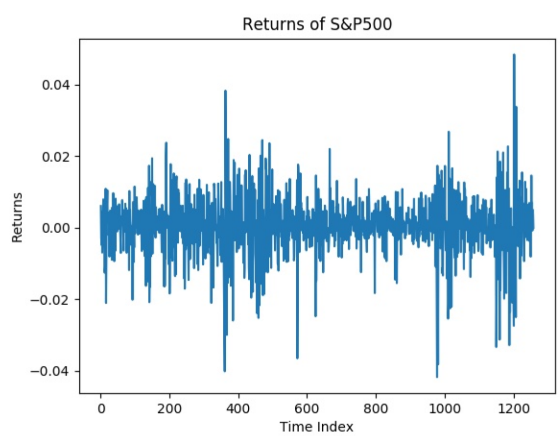


Figure 1.1: Returns of S&P

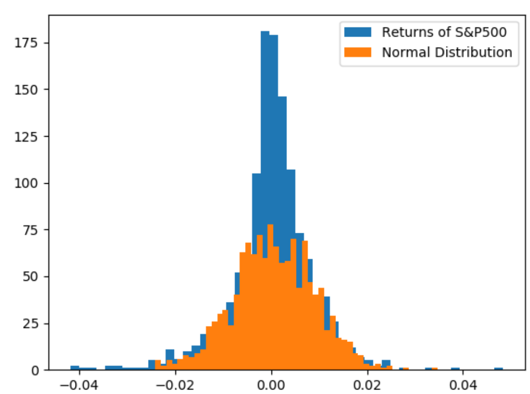


Figure 1.2: Distribution of Returns

	Returns
Number of Obs	17411.000000
NAs	0.000000
Minimum	-0.228997
Maximum	0.109572
1. Quartile	-0.004036
3. Quartile	0.004950
Mean	0.000294
Median	0.000468
Sum	5.124739
SE Mean	0.000073
LCL Mean	0.000151
UCL Mean	0.000438
Variance	0.000093
Stdev	0.009659
Skewness	-1.004394
Kurtosis	26.841302

Table 1.1: Summary statistics of returns

Returns have some empirical properties which are called *stylized facts*. They are

- Linear correlations for returns are insignificant except for small intra-day time scales.
- Returns usually have leptokurtic distributions
- High volatility events usually accompanied by similar events. This is called *volatility clustering*.
- Volatility is negatively correlated with returns. It is called *leverage effect*.

Chapter 2

Time Series Analysis

The first econometric model was constructed by Jan Tinbergen in 1939. In classical time series analysis, it is assumed that residuals of estimated equations were stochastically independent. Donald Cochrane and Guy H Orcutt demonstrated in 1949 if residuals are positively correlated then variances of regressions are underestimated, and F and t statistics are overestimated, which is rectified by transforming data suitably. Box Jenkins Analysis presented a systematic use of information in data to predict the future of the variable.

Classical time series analysis(**TSA**) assumes that a time series could be differentiated into -

- a long-term development called the trend
- cyclical component with periods more than one year
- a seasonal part having ups and downs in a year

  These are called systematic components which could be explained by deterministic equations.

- a residual which could not be explained by the above three components which is a stochastic component. It is modelled as an independent or uncorrelated random variable with mean zero and constant variance. It is a pure random process.

A time series model is chosen based on statistical figures, and the parameters are estimated. These parameters are subject to statistical tests. If they satisfy our hypotheses, then the process is reiterated with a new model.

2.0.1 Lag Operators

If r_t is a time series, then a **lag-operator L** satisfies

$$\mathbf{L}^p r_t = r_{t-p}$$

Properties

- $\mathbf{L}c = c$, where c is a constant
- Distributive Law: $(\mathbf{L}^i + \mathbf{L}^j)r_t = r_{t-i} + r_{t-j}$
- Associative Law: $\mathbf{L}^i\mathbf{L}^j r_t = r_{t-i-j}$
- Lead operator is obtained when $\mathbf{L}$ is raised to negative power. $\mathbf{L}^{-i} r_t = y_{t+i}$
- For any $|\alpha| < 1, (1 + \alpha\mathbf{L} + \alpha^2\mathbf{L}^2 + \dots)r_t = r_t/(1 - \alpha\mathbf{L})$
- For $|\alpha| > 1, (1 + \alpha^{-1}\mathbf{L}^{-1} + \alpha^{-2}\mathbf{L}^{-2} + \dots)r_t = -\alpha\mathbf{L}r_t/(1 - \alpha\mathbf{L})$

Autocovariance function γ_r of two instants is given by

$$\gamma_r(s,t) = cov(r_s, r_t) = E[(r_t - \mu_t)(r_s - \mu_s)].$$

A time-series r_t is called strictly stationary if the joint distribution doesn't change with a shift in time. For a strictly stationary time-series,$r_t = r_{t+k}$ in distribution. In a *weakly stationary* time series, both the mean and covariance are invariant with a time shift. Mean=μ is constant for all t and covariance of r_t and $r_s, \gamma(s,t)$: $==\gamma(|s-t|)$ which depends only on distance between the points and not on actual points. Two important properties of covariance are (i)$\gamma_0 = Var(r_t)$ and (ii) $\gamma_{-l} = \gamma_l$.

Auto-correlation function,

$$\rho(s,t) = \frac{E[(r_t - \mu_t)(r_s - \mu_s)]}{\sigma_t \sigma_s}.$$

A **white noise** process is a set of independent and identically distributed (i.i.d) variables $\{\varepsilon_t\}$ with zero mean and constant variance

$$\mathbb{E}(\varepsilon_t) = \mathbb{E}(\varepsilon_{t-1}) = \cdots = 0$$

$$\mathbb{E}({\varepsilon_t}^2) = \mathbb{E}({\varepsilon_{t-1}}^2) = \cdots = \sigma^2$$

$$\mathbb{E}(\varepsilon_t \varepsilon_{t-s}) = 0$$

If $\varepsilon_t \sim \mathcal{N}(0,1)$ then it is called a *Gaussian white noise.*

2.1 Homoskedastic Time Series Models

Weiner-Kolmogorov prediction formula is

$$E[r_{t+1}|r_t, r_{t-1}, \dots] = \mu + \left[\frac{\psi(\mathbf{L})}{\mathbf{L}^s}\right]_+ \frac{1}{\psi(\mathbf{L})}(r_t - \mu)$$

where $[.]_+$ is the *annihilation operator* which replaces negative powers of $\mathbf{L}$ with zero.

2.1.1 Auto Regressive Process

An **Autoregressive process** of order p is defined as

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \phi_p r_{t-p} + \varepsilon_t$$

In terms of lag-operator $r_t = \mu + \psi(\mathbf{L})\varepsilon_t$, where $\psi(\mathbf{L}) = (1 - \phi_1 \mathbf{L} - \cdots - \phi_p \mathbf{L}^p)^{-1}$.

$$\text{Mean}, \mu = \frac{\phi_0}{1 - \phi_1 - \cdots - \phi_p}$$

We can write $AR(p)$ process as

$$r_t - \mu = \phi_1(r_{t-1} - \mu) + \phi_1(r_{t-2} - \mu) + \cdots + \phi_p(r_{t-p} - \mu) + \varepsilon_t$$

. It is weakly stationary if roots of

$$1 - \phi_1 z - \phi_2 z - \cdots - \phi_p z^p = 0$$

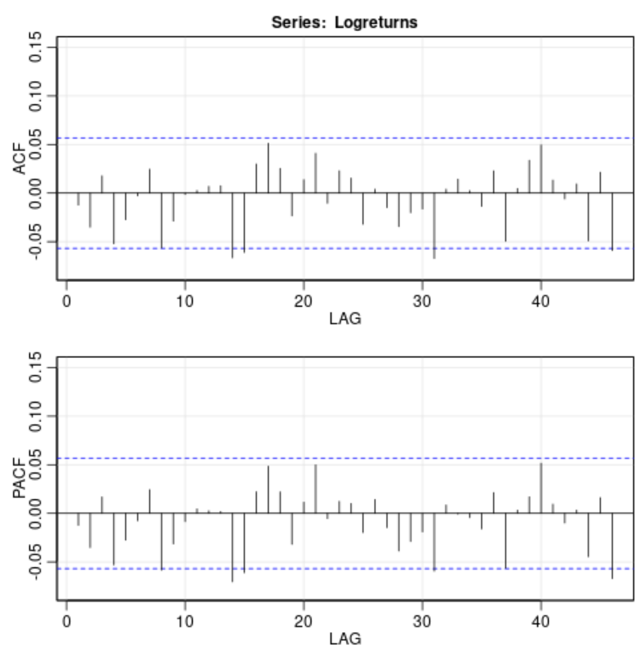
lie outside unit circle.

$$\text{Autocovariances}, \gamma_j = \begin{cases} \phi_1\gamma_{j-1} + \phi_2\gamma_{j-2} + \cdots + \phi_p\gamma_{j-p} & j = 1, 2, 3... \\ \phi_1\gamma_1 + \phi_2\gamma_2 + \cdots + \phi_p\gamma_p + \sigma^2 & j = 0 \end{cases} \tag{2.1}$$

Dividing autocovariances with γ_0 we get,

$$\rho_j = \phi_1\rho_{j-1} + \cdots + \phi_p\rho_{j-p},$$

which are called *Yule-Walker* equations. Solving these equations we get coefficients ϕ_i. Thus, both autovariances and autocorrelations follow the same pth order difference equation like $AR(p)$ process.



AR(1) model

AR(1) is written as $r_t = \phi_0 + \phi r_{t-1} + \varepsilon_t$, which is weakly stationary if $\phi_1 < 1$.

- $\mathbb{E}(r_t) = \mu = \phi_0/(1-\phi)$

- Variance=$\mathbb{E}(r_t-\mu)^2$
$$= \mathbb{E}(\varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots)^2$$
$$= (1+\phi^2+\phi^4+\dots)\sigma^2$$
$$= \sigma^2/(1-\phi^2)$$

- j-th autocovariance,$\gamma_j = \mathbb{E}(r_t-\mu)(r_{t-j}-\mu)$
$$= (\phi^j + \phi^{j+2} + \phi^{j+4} + \dots)\sigma^2$$
$$= \phi^j(1+\phi^2+\phi^4+\dots)\sigma^2$$
$$= [\phi^j/(1-\phi^2)]\sigma^2$$

- j-th autocorrelation function,
$$\rho_j = \gamma_j/\gamma_0 = \phi^j$$

Forecasting an AR(1) model

$$\psi(\mathbf{L}) = \frac{1}{1-\phi\mathbf{L}} = 1 + \phi\mathbf{L} + \phi^2\mathbf{L}^2 + \dots$$

An s-period ahead forecast is $\quad \mu + \phi^s(r_t-\mu)$

One-step ahead forecast is given by

$$\bar{E}[r_{t+1}|r_t, r_{t-1}, \dots] = \mu + \phi(r_t-\mu)$$

2.1.2 Moving Average Process

A moving average(**MA**) process of order q is defined as

$$r_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

$$\text{Mean} = \mathbb{E}(r_t) = \mu$$

$$\text{Variance}, \gamma_0 = \sigma^2 + {\theta_1}^2 \sigma^2 + {\theta_2}^2 \sigma^2 + \cdots + {\theta_q}^2 \sigma^2$$

$$\gamma_j = E[(\varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots)(\varepsilon_{t-j} + \theta_1 \varepsilon_{t-j+\dots})]$$

$$\text{As} \quad \mathbb{E}[\varepsilon_t \varepsilon_s] = 0, \gamma_j = \begin{cases} (\theta_j + \theta_{j+1}\theta_1 + \theta_{j+2}\theta_2 + \cdots + \theta_{q-j}\theta_q) & j = 1, 2, 3.. \\ 0 & j > q \end{cases}$$

MA(1) model

MA(1) model is written as $r_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1}$

- Mean=μ
- Variance=$(1 + \theta^2)\sigma^2$

Forecasting an MA(1) model

$$r_t - \mu = (1 + \theta \mathbf{L})\varepsilon_t$$

Residual term is estimated as $\tilde{\varepsilon}_t = r_t - \mu - \theta \tilde{\varepsilon}_{t-1}$

One-step ahead forecast is $r_{t+1|t} = \mu + \theta \tilde{\varepsilon}_t$

Mixed processes

An autoregressive moving average process(**ARMA**) of order (p, q) is defined as

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \cdots + \phi_p r_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

In terms of lag-operator,

$$r_t = \mu + \psi(\mathbf{L})\varepsilon_t,$$

where

$$\psi(\mathbf{L}) = \frac{\theta(\mathbf{L})}{\phi(\mathbf{L})} = \frac{1 + \theta_1 \mathbf{L} + \theta_2 \mathbf{L}^2 + \cdots + \theta_q \mathbf{L}^q}{1 - \phi_1 \mathbf{L} - \phi_2 \mathbf{L}^2 - \cdots - \phi_p \mathbf{L}^p}$$

Given $\phi(\mathbf{L}) = 0$ has roots outside unit circle both sides are divided by $\phi(\mathbf{L})$, we get $\mu = \frac{1}{1-\phi_1-\phi_2-\cdots-\phi_p}$ and hence stationarity depends only on autoregressive part.

Autocovariances, $\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \cdots + \phi_p \gamma_{j-p}$ for $j = q+1, q+2, \ldots$

Forecasting an ARMA(1,1) model

s-step ahead forecast is given by

$$\mu + \frac{\phi^s + \theta\phi^{s-1}}{1 - \phi\mathbf{L}}(r_t - \mu)$$

One-step ahead forecast using ARMA(1,1) model is

$$r_{t+1|t} = \mu + \frac{\phi + \theta}{1 + \theta\mathbf{L}}(r_t - \mu)$$

The mean absolute percentage error(MAPE) for ARMA(1,1) model and actual observations is 1.674134

An autoregressive integrated moving average process(**ARIMA**) of order (p,q,d) is such that after differencing d times we get an ARMA(p,q) process.

$$r_t = \text{ARIMA}(p, q, d) \iff \Delta^d r_t = \text{ARMA}(p, q)$$

2.2 Heteroskedastic Time Series Models

2.2.1 ARCH model

In earlier models, the unconditional variance of the white noise process is constant σ^2. But the conditional variance can vary with time. Time-varying conditional variance is called *autoregressive heteroskedasticity*(**ARCH**) modelled by Engle in his seminal work on inflation in the UK in 1982. In ARCH models, the residuals are serially uncorrelated but are dependent and the dependence of r_t can be described by

$$r_t = \sigma_t \epsilon_t, \sigma_t^2 = \alpha_0 + \alpha_1 {r_{t-1}}^2 + \alpha_2 {r_{t-2}}^2 + \cdots + \alpha_m {r_{t-m}}^2$$

where ϵ_t are i.i.d random variables with mean 0 and variance 1, α_0, $\alpha_i > 0$ for $i \geq 1$. To be weakly stationary, the roots of the equation

$$1 - \alpha_1 z - \alpha_2 z^2 - \cdots - \alpha_m z^m = 0$$

should lie outside unit circle. Since all α_i are nonnegative $\sum_{i=1}^{m} \alpha_i = 1$. Unconditional variance is given by

$$\sigma^2 = E(r_t)^2 = \frac{\alpha_0}{1 - \alpha_1 - \alpha_2 - \cdots - \alpha_m}$$

Taking ARCH(1) model for illustration,

$$r_t = \sigma_t \epsilon_t, {\sigma_t}^2 = \alpha_0 + \alpha_1 {r_{t-1}}^2$$

Unconditional mean,$E(r_t) = E[E(r_t|r_{t-1}, r_{t-2}..)] = E[\sigma_t E(\epsilon_t)] = 0$

Unconditional variance,

$$Var(r_t) = E({r_t}^2) = E[\alpha_0 + \alpha_1 {r_{t-1}}^2] = \alpha_0 + \alpha_1 E({r_{t-1}}^2)$$

Since r_t is a stationary process, $E(r_{t-1}^2) = E(r_t)$

$$Var(r_t) = \alpha_0 + \alpha_1 Var(r_t) = \frac{\alpha_0}{1 - \alpha_1}$$

Fourth order moment is given by

$$E({r_t}^4) = \frac{3\alpha_0^2(1+\alpha_1)}{(1-\alpha_1)(1-3{\alpha_1}^2)}$$

Unconditional kurtosis,

$$\frac{E({r_t}^4)}{Var(r_t)^2} = 3\frac{1-{\alpha_1}^2}{1-3{\alpha_1}^2} > 3.$$

The excess kurtosis shows that r_t has heavier tails than normal distribution which can accommodate more outliers.

Test for ARCH effects

Engle's test for ARCH effects is based on the Lagrange multiplier principle. If T is the number of data points and m is a prespecified positive number, the regression equation ${r_t}^2 = \alpha_0 + \alpha_1 r_{t-1}^2 + \cdots + \alpha_m {r_{t-m}}^2 + \bar{e}_t$ is first fitted with ordinary least squares(OLS) and OLS residuals of $\bar{u}_t$ are saved. T times R^2 of the regression converges in χ^2 distribution with m degrees of freedom under null hypothesis that r_t is Gaussian white noise.

Null-hypothesis that ARCH effects are not present is rejected, as Chi-squared = 109.28, df = 1, $p-value < 2.2e-16$

Forecasting

ARCH model forecasting is similar to AR forecasting. Consider an ARCH(m) model. At forecast horizon t, the one-step ahead forecast of σ_t^2 is

$${\sigma_{t+1}}^2 = \alpha_0 + \alpha_1 {r_t}^2 + \cdots + \alpha_m {r_{h+1-m}}^2$$

Disadvantages of ARCH models

- The model gives the same effects for positive and negative shocks because it depends on the square of previous shocks.

- Because of restrictions on ARCH models, it is hard to capture excess kurtosis in higher-order models.

- They overestimate volatility because of their slow response to largely isolated shocks.

- ARCH model doesn't give the cause of the heteroskedasticity. It only models volatility.

2.2.2 GARCH models

Because of the requirement of many parameters to estimate volatility, Bollerslev developed GARCH model in 1986. For the return series, r_t, $a_t = r_t - \mu_t$ be the innovation.

A time series $\{a_t\}$ is *generalized ARCH*(**GARCH**) model of order (p,q) if

$$a_t = \mu_t + \sigma_t \epsilon_t, {\sigma_t}^2 = \alpha_0 + \sum_{i=1}^{p} a_i {\sigma_{t-i}}^2 + \sum_{j=1}^{q} {\beta_j}^2 {\sigma_{t-j}}^2,$$

with $\alpha_0 > 0, \alpha_i \geq 0, \beta_j \geq 0, \sum_{i=1}^{max(p,q)} (\alpha_i + \beta_i) \leq 1 and \alpha_0 > 0$

Unconditional variance of the model is

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i - \sum_{j=1}^{q} \beta_j}$$

Properties of GARCH model can be understood by studying properties of GARCH(1,1) model given by

$${\sigma_t}^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 {\sigma_{t-1}}^2$$

A large a_{t-1}^2 or ${\sigma_{t-1}}^2$ gives a large a_t. This explains *volatilty clustering* in financial data first observed by Mandelbrot.

If $1 - 2{\alpha_1}^2 - (\alpha_1 + \beta_1)^2 > 0$ then

$$\frac{E({a_t}^4)}{{\sigma_t}^2} = \frac{3[1 - (\alpha_1 + \beta_1)^2]}{1 - (\alpha_1 + \beta_1)^2 - 2{\alpha_1}^2} > 3$$

Forecasting

GARCH model forecasting is similar to ARMA model forecast. One-step ahead forecast using GARCH(1,1) model is

$$\sigma_{t+1}{}^2 = \alpha_0 + \alpha_1 a_t{}^2 + \beta_1 \sigma_t{}^2$$

Drawbacks

- Similar to ARCH model, GARCH model, also does not account for *leverage effect*.

2.2.3 Modified GARCH models

There have been many modifications of GARCH models including the EWMA model, EGARCH model, TARCH model, IGARCH and others which are also called asymmetric ARCH models.

Exponentially Weighted Moving Average(**EWMA**) model was developed by Riskmetrics. Volatality forecast is $\sigma_t{}^2 = \lambda \sigma_{t-1}{}^2 + (1-\lambda) r_{t-1}{}^2$. Riskmetrics uses $\lambda = 0.94$ and goes back 75 data points for their estimation

Exponential GARCH(**EGARCH**) model was developed by Nelson and places no restrictions on model estimation.

$$ln(\sigma_t{}^2) = \alpha_0 + \sum_{i=1}^{q} \left(\alpha_i \left| \frac{r_{t-i}}{\sigma_{t-i}} \right| + \gamma_i \frac{r_{t-i}}{\sigma_{t-i}} \right) + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}{}^2)$$

Logarithm of variance ensures nonnegative forecasts of variance. γ_i allows for asymmetric effects. In real life applications, γ_i is assumed to be negative.

Threshold GARCH(**TGARCH**) model is of the form

$$\sigma_t{}^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i r_{t-i}{}^2 + \gamma_1 r_{t-1}{}^2 d_{t-1} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}{}^2$$

where $d_t = 1$ if $a_t < 0$ and $d_t = 0$ otherwise.

Chapter 3

Financial Risk Management

3.1 Risk Measures

Definition. *Let $\mathcal{G}$ be the set of all risks. A risk measure is a mapping ρ:$\mathcal{G}$→$\mathcal{R}$*

3.1.1 Coherent measures of risk

Axiom T. *Translational invariance. For all $X \in \mathcal{G}$ and all real numbers α $\rho(X + \alpha.r) = \rho(X) - \alpha$*

When a risk free-asset is added to the portfolio with weight α, then it reduces the risk proportional to that weight.

Axiom S. *Subadditivity. For all X_1 and $X_2 \in \mathcal{G}$, $\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2)$*

Risk of a portfolio is lesser than individual risks.

Axiom PH. *Positive homogeneity. For all $\lambda > 0$ and all $X \in \mathcal{G}, \rho(\lambda X) = \lambda\rho(X)$*

Risk cannot be increased or decreased by investing different amounts in the same stock.

Axiom M. *Monotonicity. For all X and $Y \in \mathcal{G}$ with $X \leq Y$ $\rho(Y) \leq \rho(X)$*

Higher risk can entail higher loss.

Axiom R. *Relevance. For all* $X \in \mathcal{G}$ *with* $X \leq 0$ *and* $X \neq 0$ $\rho(X) > 0$

Defnition. *Coherence: A risk measure which satisfies all axioms T, S, PH, M, R is coherent.*

Two most important risk measures are Value at Risk(**VaR**) and Expected Shortfall(**ES**), which depict the maximum loss incurred by a firm in case of an adverse event. Firms were advised to work on their own internal models since 1970s because it became increasingly complex to model the whole market.

3.2 Value at Risk

Defnition. *Given* $\alpha \in [0,1]$,VaR_α *of final net worth* X *with distribution* $\mathbb{P}$ *is the negative of the quantile* ${q_\alpha}^+$ *of X.*

$$VaR_\alpha(X) = -\inf\{x|\mathbb{P}[X \leq x] > \alpha\}$$

VaR is the maximum loss occurred with a confidence level $(1-\alpha)$, at time horizon T. It is rare loss under normal market conditions or minimal loss under extraordinary conditions.

$$P(r_t < -VaR_{\alpha,T}) = \alpha$$

Let $V(t)$ be the value of the portfolio at time, t. Suppose that $\Delta V(l)$ is the change in value of the portfolio after l periods. $L(l)$ be the loss function of $\Delta V(l)$ which can be positive or negative depending on position being short or long. VaR of the portfolio at time horizon t with confidence interval α is

$$\alpha = \mathrm{P}[L(l) \geq VaR] = 1 - \Pr[L(l) < \mathrm{VaR}]$$

So the probability that loss is greater than or equal to VaR is α. In case of normal distribution of returns, the VaR is straight forward to compute

$$\mathrm{VaR}(\alpha) = \Phi^{-1}(\alpha)\hat{\sigma},$$

where Φ^{-1} is the quantile function of normal distribution. VaR provides a holistic measure of risk for a portfolio. VaR has become synonymous with risk measurement after Basel

Accord(1995) which stipulated capital adequacy based on VaR. One- day VaR is related to n-day VaR as

$$VaR_{n-day} = VaR_{1-day}\sqrt{n}.$$

The BIS sets the capital requirements at three times the ten-day 1% VaR forecast.

Major differences between portfolio theory and VaR are

- Portfolio Theory(PT) measures risk in terms of standard deviation of return whereas VaR is the maximum likely loss in an adverse event.
- VaR approaches are more flexible because they can accommodate a number of possible distributions whereas PT assumes the P/L are normally or lognormally distributed.
- VaR can be applied to different types of risk such as credit risk, operational risk etc and PT is limited to market risks.
- VaR can be estimated using many methods and PT is cumbersome to interpret.

3.2.1 Estimation of VaR

- **Historical Simulation** An empirical distribution of profits and losses is obtained. VaR is determined by the associated quantile. Let $r_1, r_2, \ldots$ are returns. If there are n sample points and α is the confidence interval then VaR is the [n.α] ordered statistic, where [.] is nearest integer to be rounded off. It works only when we have a large sample.
- **Parametric Estimation** Calculation of analytic solution to assumed cumulative distribution. Not all distributions have solutions. But Extreme Value Theory(EVT) can be used. If $F(\alpha)$ is the quantile function of a distribution and σ_{t+1} is the volatality then VaR = $F(\alpha)\hat{\sigma}_{t+1}$.
- **Monte Carlo Simulation** An asset return is simulated, and the distribution of returns is obtained after many simulations. VaR can be obtained from this distribution by methods similarly used in the historical simulation.

We mainly concentrated on parametric estimation of VaR using GARCH, EGARCH, TGARCH models and provide analysis for the same.

Quantile loss(**QL**) function has the form

$$\Psi_{t+1} = \begin{cases} (r_{t+1} - VaR_{t+1|t})^2, & r_{t+1} < VaR_{t+1|t} \\ [Percentile(y, 100p)_1{}^T - VaR_{t+1|t}]^2 & r_{t+1} \geq VaR_{t+1|t} \end{cases} \tag{3.1}$$

Every time model's loss occurs, the distance between forecast and realization increases. Therefore a model which minimizes the QL function is selected. If $z_{t+1} = \Psi_{t+1}{}^A - \Psi_{t+1}{}^B$, where Ψ_A and Ψ_B are the loss functions of models A and B, respectively. A negative value of z_{t+1} indicates that model A is superior to model B. The Diebold–Mariano [20] statistic is the "t-statistic" for a regression of z_{t+1} on a constant with heteroskedastic and autocorrelated consistent standard errors (HAC).

Data Analysis

Three indices namely **S&P500**, **Nikkei225** and **Dow Jones** are used for analysis. This data is obtained from Yahoo Finance. We used data for past five years(1200 data points) to estimate value at risk, which was chosen through trial and error. Correlation times are less than a day, so we use one lag while estimation of VaR.

Model	Loss at 95%	Loss at 99%
GARCH(1,1) with normal distribution	5.385228	8.03075
GARCH(1,1) with student-t distribution	13.33769	11.14221
GARCH(1,1) with GED	5.037394	6.602765
EGARCH(1,1) with normal distribution	4.304211	6.501847
EGARCH(1,1) with student-t distribution	5.2234663	9.938402
EGARCH(1,1) with GED	3.754851	4.987685
TGARCH(1,1) with normal distribution	4.645982	6.98522
TGARCH(1,1) with student-t distribution	5.659375	10.70456
TGARCH(1,1) with GED	4.05854	5.370114

Table 3.1: Analysis of VaR estimates using different models for S&P500

A decrease in loss can be seen from GARCH to EGARCH to TGRACH. It is proved that any lag more than 1, yields no better results. These models are improvements of classical models but still a lot of refinement of parmeters should be done to obtain better results.

Model	Loss at 95%	Loss at 99%
GARCH(1,1) with normal distribution	2.043907	2.476417
GARCH(1,1) with student-t distribution	2.296762	3.279197
GARCH(1,1) with GED	1.963323	2.213093
EGARCH(1,1) with normal distribution	1.997287	2.410482
EGARCH(1,1) with student-t distribution	2.106793	2.945307
EGARCH(1,1) with GED	1.832831	2.048767
TGARCH(1,1) with normal distribution	2.078126	2.524814
TGARCH(1,1) with student-t distribution	2.172477	3.060753
TGARCH(1,1) with GED	1.887095	2.1171

Table 3.2: Analysis of VaR estimates using different models for Dow Jones

Model	Loss at 95%	Loss at 99%
GARCH(1,1) with normal distribution	4.160555	6.298671
GARCH(1,1) with student-t distribution	5.90838	11.14221
GARCH(1,1) with GED	4.061348	5.373651
EGARCH(1,1) with normal distribution	3.859677	5.873134
EGARCH(1,1) with student-t distribution	5.035543	9.608109
EGARCH(1,1) with GED	3.496926	4.662886
TGARCH(1,1) with normal distribution	3.970952	6.030513
TGARCH(1,1) with student-t distribution	5.064579	9.659143
TGARCH(1,1) with GED	3.539887	4.716985

Table 3.3: Analysis of VaR estimates using different models for Nekkei

Conclusion

The VaR estimates from different models are studied in order of increasing accuracy. As we move from classical models to different *ARCH* models, the losses decrease considerably. The loss function methodology used is one among many proposed ways to select a model but it provides a better way for selecting a model. Though we have used only one lag in VaR estimation we still get very good estimates, providing an insight that financial data are dependent only to one lag and increase in lags would increase the computational complexity. As we move from normal to GED the losses decrease in some indices and increase in other, which questions the legitimacy of selected model. Estimates from GED are lesser than normal and generalised-t distributions, which is different from argument in [3]. The sampling size of 1200 or 5 years was used in estimation of VaR, which suited to be an optimal sample size. One problem observed in estimation was non-convergence of VaR when higher orders were selected for autoregressive process. An optimal strategy has to be designed after backtesting and stress testing. A generalised model cannot be designed for all data sets.

One of the major problems quoted in literature is the subadditivity of VaR. Therefore, a coherent risk measure called *estimated shortfall* is defined which is expectation of the tail beyond VaR. Expected shortfall is the expected loss conditional on the loss being greater than VaR.

$$ES_\alpha = -\inf\{E[r_t | r_t \leq -VaR_\alpha]\}$$

Appendix

Probability Distributions

Normal distribution

CDF of a normal distribution with mean μ and variance σ^2 is given by

$$N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{1}{2}(\frac{x-\mu}{\sigma})^2.$$

Log-likelihood function for T normally distributed x_i's is

$$-\frac{1}{2}\left[T\ln(2\pi) + \sum_{i=1}^{T} x^2 + \sum_{i=1}^{T} \ln({\sigma_t}^2)\right].$$

Generalised t-distribution

Density of a student t-distribution with ν degrees of freedom is

$$\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi - 2\pi}} (1 + \frac{x^2}{\nu - 1})^{-\frac{\nu+1}{2}},$$

where $\Gamma(\nu)$ is the gamma function $\Gamma(\nu) = \int_0^\infty e^{-x} x^{\nu-1}$ and ν is the shape parameter which describes the thickness of tails. For large values os ν t-distribution converges to $\mathcal{N}(0, 1)$.

Log-likelihood function for T student-t distributed x_i's is

$$T\left[\ln\Gamma\left(\frac{\nu+1}{2}\right)-\ln\Gamma(\frac{\nu}{2})-\frac{1}{2}\ln[\pi(\nu-2)]\right]-\frac{1}{2}\sum_{t=1}^{T}\left[\ln({\sigma_t}^2)+(1+\nu)\ln\left(1+\frac{x^2}{\nu-2}\right)\right].$$

General Error Distribution(GED)

A GED is given by

$$\frac{\nu\exp(-0.5|x/\lambda|^{\nu})}{\lambda 2^{(1+\frac{1}{\nu})}\Gamma(\frac{1}{\nu})}$$

$\lambda=\left[\frac{\Gamma(\frac{1}{\nu})}{2^{\frac{2}{\nu}}\Gamma(\frac{3}{\nu})}\right]^{\frac{1}{2}}$ is shape parameter.

GED converges to $\mathcal{N}(0,1)$ when $\nu = 2$ and for $\nu < 2$ it has thicker tails than normal distribution.

Log-likelihood function for GED distributed x_i's is

$$\sum_{t=1}^{T}\left[\ln(\frac{\nu}{\lambda})-0.5\left|\frac{x}{\lambda}\right|^{\nu}-(1+\nu^{-1})\ln(2)-\ln\Gamma(\frac{1}{\nu})-0.5\ln({\sigma_t}^2)\right].$$

Bibliography

[1] Philippe Jorion, Value At Risk,3rd Ed, McGraw Hill Education, 2007.

[2] Ruey Tsay, Analysis of Financial Time Series, Second Edition, Wiley NJ, 2010

[3] Timotheos Angelidis, Alexandros Benos, Stavros Degiannakis,The use of GARCH models in VaR estimation,Statistical Methodology,Volume 1, Issues 1–2,2004, Pages 105-128