Computing Value at Risk for portfolio of asset using different parametric and non parametric approaches

Mallikarjun Matti

Submitted for the Degree of Master of Science in

MSC artificial Intelligence

Department of Computer Science
Royal Holloway University of London
Egham, Surrey TW20 0EX, UK

August 31, 2022

Declaration

This report has been prepared on the basis of my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count: 11000

Student Name: Mallikarjun Matti

Date of Submission: 31/08/2022

Signature: M M

Abstract

Value at Risk (VaR) has become the standard measure that financial analysts use to quantify market risk. VaR is defined as the maximum potential change in value of a portfolio of financial instruments with a given probability over a certain horizon. VaR measures can have many applications, such as in risk management, to evaluate the performance of risk takers and for regulatory requirements, and hence it is very important to develop methodologies that provide accurate estimates.

This project focus on different methodologies for VaR computation and testing it on real financial Data. Volatility is measure of Risk, different ways for estimating the variance of returns and the covariance between returns of multiple assets is calculated using the standard formula and exponentially weighted moving average. Garch volatility models is also implemented to capture time dependent Volatility for univariate case, optimal parameters of the model is calculated using log likelihood function and parameters are analysed.

Several parametric and non-parametric methods to measure Value-at-Risk are discussed and implemented for a single Stock and for a portfolio of stocks then extended to implement Value at risk for Derivative mainly options. The non-parametric approach is represented by historical simulations, bootstrapping and Monte-Carlo methods. Variance covariance and some analytical models are used to demonstrate the parametric approach. For derivative pricing like option Binomial model building and Black sholes option pricing have been implemented. Coherent Risk measure Expected shortfall is also implemented and Results are compared against Value at Risk.

Finally, the back testing procedure is implemented and the statistical significance of various Value at risk is Analysed.

Table of Contents

1 Introduction

VAR (value at risk) traces its roots to the infamous financial disasters often early 1990s that engulfed Orange County, Barings, Metallgesellschaft, Daiwa, and so many others. The common lesson of these disasters is that billions of dollars can be lost because of poor supervision and management of financial risks. Spurred into action, financial institutions and regulators turned to value at risk, an easy-to-understand method for quantifying market risk.

What is VAR? VAR is a method of assessing risk that uses standard statistical techniques routinely used in other technical fields. Formally, VAR measures the worst expected loss over a given horizon under nor-mal market conditions at a given confidence level. Based on firm scientific foundations, VAR provides users with a summary measure of mar-kept risk. For instance, a bank might say that the daily VAR of its trading portfolio is $35 million at the 99 percent confidence level. In other words, there is only 1 chance in a 100, under normal market conditions, for a loss greater than $35 million to occur. This single number summarizes the bank's exposure to market risk as well as the probability of an adverse move. Equally important, it measures risk using the same units as the bank's bottom-line dollars. Shareholders and managers can then decide whether they feel comfortable with this level of risk. If the answer is no, the process that led to the computation of VAR can be used to decide where to trim the risk

.In contrast with traditional risk measures, VAR provides an aggregate view of a portfolio's risk that accounts for leverage, correlations, and current positions. As a result, it is truly a forward-looking risk measure. VAR, however, applies not only to derivatives but to all financial instruments. Furthermore, the methodology can also be broadened from market risk to other types of financial risks.

VAR is the maximum loss over a target horizon at an assigned confidence level, i.e. the probability that the actual loss will be larger than the VaR amount is equal to $1 - \alpha$, where α is the confidence level:

$$\Pr\left(P(t + n\Delta) - P(t) \leq -\mathrm{VaR}_{\alpha}^{p/}\,(t, t + n\Delta)\right) = 1 - \alpha. \tag{1}$$

1.1 Types of Risk

VAR was developed initially to deal with one aspect of financial risk, market risk. It should be recognized, however, that there are many other aspects of financial risk. Generally, financial risks are classified into the broad categories of market risks, credit risks, liquidity risks, operational risks, and sometimes legal risks. [1]

1.1.1 Market Risk

Market risk arises from movements in the level or volatility of market prices. VAR tools now allow users to quantify market risk in a system attic fashion. Market risk can take two forms: absolute risk, measured in dollar terms (or in the relevant currency) and relative risk, measured relative to a benchmark index. While the former focuses on the volatility of total returns, the latter measures risk in terms of tracking error, or deviation from the index. Market risk can be classified into directional and nondirectional risks. Directional risks involve exposures to the direction of movements in financial variables, such as stock prices, interest rates, exchange rates, and commodity prices. These exposures are measured by linear approximations such as beta for exposure to stock market movements, duration for exposure to interest rates, and delta for exposure of options to the underlying asset price. Nondirectional risks, then, involve the remaining risks, which con¬ sist of nonlinear exposures and exposures to hedged positions or to volatilities. Second-order or quadratic exposures are measured by convexity when dealing with interest rates and gamma when dealing with options. Basis risk is created from unanticipated movements in relative prices of assets in a hedged position, such as cash and futures or interest rate spreads. Finally, volatility risk measures exposure to movements in the actual or implied volatility. [1]

Nondirectional risks, then, involve the remaining risks, which con¬ sist of nonlinear exposures and exposures to hedged positions or to volatilities. Second-order or quadratic exposures are measured by convexity when dealing with interest rates and gamma when dealing with options. Basis risk is created from unanticipated movements in relative prices of assets in a hedged position, such as cash and futures or interest rate spreads. Finally, volatility risk measures exposure to movements in the actual or implied volatility.

1.1.2 Credit Risk

Credit risk originates from the fact that counterparties may be unwilling or unable to fulfil their contractual obligations. Its effect is measured by the cost of replacing cash flows if the other party defaults. This loss encompasses the exposure, or amount at risk, and the recovery rate, which is the proportion paid back to the lender, usually measured in terms of “cents on the dollar.”

1.1.3 Liquidity Risk

Liquidity risk takes two forms, asset liquidity risk and funding liquidity risk. Asset liquidity risk, also known as market/product liquidity risk, arises when a transaction cannot be conducted at prevailing market prices due to the size of the position relative to normal trading lots. This risk varies across categories of assets and across time as a function of prevailing market conditions. Some assets, such as major currencies or Treasury bonds, have deep markets where most positions can be liquidated easily, with very little price impact

2 Background Research

2.1 Measuring Returns

In the context of the measurement of market risk, the random variable is taken as the rate of return on a financial asset The range of possible payoffs on a security also can be described by its probability distribution function.

Define, for instance, the measurement horizon as 1 month. Returns are measured from the end of the preceding month, denoted by the t ~ 1 to the end of the current month, denoted by t. The arithmetic, or discrete, rate of return is defined as the capital gain plus any interim payment such as a dividend or coupon:

$$r_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} \tag{2}$$

To focus on long-horizon returns, the practice is to focus on the geo¬ metric rate of return, which is defined in terms of the logarithm of the price ratio

$$R_t = \ln \frac{P_t + D_t}{P_{t-1}} \tag{3}$$

There are advantage of using geometric returns, they may Be more economically meaningful than arithmetic returns. If geometric returns are distributed normally, then the distribution can never lead to price that is negative.

Log-returns take values in the interval $(-\infty, \infty)$ and Simple returns assume values in the interval $(-1, \infty)$. so, the normality assumption must be attached to log-returns rather than to simple returns. In particular, it can be shown that Returns will be distributed according to a shifted log-normal distribution. The two distributions can become quite different as we increase the time frame Δ.

The distribution of simple returns can be then obtained using the transformation rule for random variables, The two distributions can become quite different as we increase the time frame Δ. . But if returns are small (for shorter time frame) log returns and athematic returns are very close to each other R (t) ≃ r (t) using Taylors formal we get the same result

$$R(t) = e^{r(t)} - 1 \simeq (1 + r(t)) - 1 = r(t) \tag{4}$$

- For calculating Portfolio return

If $R_i(t)$ is the simple return on asset i , w_i the portfolio simple return is calculated using following formula $R_p(t) = \sum_{i=1}^{N} w_i \times R_i(t)$

If r_i (t) is the log-return on asset i and w_i is the weights of this asset in the portfolio, i = 1, ...N, the portfolio log-return r_p (t) is $r_p(t) = \ln\left(\sum_{i=1}^{N} w_i e^{r_i(t)}\right)$ Hence Log-returns do not satisfy the portfolio linearity property, i.e. the log return of a portfolio is not the value-weighted sum of the log returns of the constituent assets. Instead, this is valid for linear returns.

- Portfolio standard deviation

If σ_i^2 is the variance of asset i and w_i is the weights of this asset in the portfolio, i = 1, ...N, the portfolio standard deviation is

$$\sigma_P = \sqrt{\sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{\substack{i=1,j=1 \\ i \neq j}}^{N} w_i w_j \rho_{ij} \sigma_i \sigma_j} \tag{5}$$

2.2 Sample Estimates

In practice, the distribution of rates of return is usually estimated over a number of previous periods, assuming that all observations are identically and independently distributed (i.i.d.). If T is the number of observations, the expected return, or first moment, m = E(X) can be estimated by the sample mean: $m = \hat{\mu} = \frac{1}{T}\sum_{i=1}^{T} x_i$

and the variance, or second moment, variance = E[$((x_i - \hat{\mu})^2$] can be estimated by the sample variance: The square root of variance is the standard deviation of X, often referred to as the volatility. It measures the risk of a security as the dispersion of outcomes around its expected value. $s^2 = \hat{\sigma}^2 = \frac{1}{(T-1)}\sum_{i=1}^{T} (x_i - \hat{\mu})^2$

To measure how assets vary with each other, we calculate the covariance. The covariance between returns of two assets *X* and *Y* can be expressed as Where X_t and Y_t are returns for asset X and Y on period [t-1, t] $\text{Cov}_{XY} = \frac{1}{T-1}\sum_{t=1}^{T} (X_t - \bar{X}) \cdot (Y_t - \bar{Y})$.

2.3 Stationarity and ADF test

Time series data are generally characterized by their temporal nature. This temporal nature adds a trend or seasonality to the data that makes it compatible for time series analysis and forecasting. Time-series data is said to be stationary if it doesn't change with time or if they don't have a temporal structure. So, it is highly necessary to check if the data is stationary. In time series forecasting, we cannot derive valuable insights from data if it is stationary. stationarity means that the statistical properties of the process do not change over time data. Use of non-stationary variables can lead to spurious regression. If a non-stationary process must be differenced d times before it becomes stationary, then it is said to be integrated of order d . To formally establish whether a stochastic process is stationary or non-stationary, we use unit root tests Two popular tests are: the Augmented Dickey-Fuller test (ADF) and the Phillips- Perron test (PP)

2.3.1 Types of stationarity:

- **Strong stationarity** - requires the shift-invariance (in time) of the finite-dimensional distributions of a stochastic process. This means that the distribution of a finite sub-sequence of random variables of the stochastic process remains the same as we shift it along the time index axis. This is the most common definition of stationarity, and it is commonly referred to simply as stationarity. For example, all i.i.d. stochastic processes are stationary. $F_X\left(x_{t_{1+\tau}}, \dots, x_{t_{n+\tau}}\right) = F_X\left(x_{t_1}, \dots, x_{t_n}\right)$

- **Weak Stationary** - only requires the shift-invariance (in time) of the first moment and the cross moment (the auto-covariance). This means the process has the same mean at all time points, and that the covariance between the values at any two time points, *t* and *t–k*, depend only on *k*, the difference between the two times, and not on the location of the points along the time axis.
  Formally, the process $\{x_i ; I \in \mathbb{Z}\}$ is weakly stationary if:
  - The first moment of x_i is constant, i.e. $\forall t,\ E[x_i]=\mu$
  - The second moment of x_i is finite for all t; i.e. $\forall t,\ E[x_i^2]< \infty$
  - The cross moment i.e. the auto-covariance depends only on the difference u-v; i.e. $\forall u,v,a,\ cov(x_u, x_v)=cov(x_{u+a}, x_{v+a})$

  The third condition implies that every lag $\tau \in \mathbb{N}$ has a constant covariance value associated with it:cov $\left(X_{t_1}, X_{t_2}\right) = K_{XX}(t_1, t_2) = K_{XX}(t_2 - t_1, 0) = K_{XX}(\tau)$Note that this directly implies that the variance of the process is also constant, since we get that for all $t \in \mathbb{N}$

2.3.2 Augmented Dickey-Fuller test

The ADF test belongs to a category of tests called 'Unit Root Test', which is the proper method for testing the stationarity of a time series. Unit root is a characteristic of a time series that makes it non-stationary. Technically speaking, a unit root is said to exist in a time series of the value of alpha = 1 in the below equation.

$$Y_t = \alpha Y_{t-1} - Y_t = \alpha Y_{t-1} + \beta X_e + \epsilon \tag{6}$$

where, Y_t is the value of the time series at time 't' and Xe is an exogenous variable (a separate explanatory variable, which is also a time series). The test above is valid only if error is not autocorrelated the solution is to augment the previous regression including lags of the dependent variable

$$y_t = c + \beta t + \alpha y_{t-1} + \phi \Delta Y_{t-1} + e_t \tag{7}$$

Where Y_{t-1} is lag 1 of time series , $\Delta Y_{t-1}$is first difference of the series at time $t-1$ alpha is the coefficient of the first lag on Y. Fundamentally, it has a similar null hypothesis as the unit root test. That is, the coefficient of Y_{t-1} is 1, implying the presence of a unit root. If not rejected, the series is taken to be non-stationary.[3]

2.4 Cholesky decomposition

Cholesky decomposition is to decompose a positive definite matrix into the product of a lower triangular matrix and its transpose. In practice, people use it to generate correlated random variables by multiplying the lower triangular from decomposing covariance matrix by standard normal. matrix decomposition is helpful in many ways, as characterizing the matrix using hidden factors uncovers properties that are universal, and not very often we can perform matrix computation explicitly.[7]

The Cholesky decomposition of the covariance matrix $\mathbf{\Sigma}$ consists in re-writing it as $\mathbf{\Sigma} = \mathbf{AA'}$where A is a lower triangular matrix. It is also guaranteed that covariance is semidefinite positive. If in addition rank (A) = N, $\mathbf{\Sigma}$ is definite positive. If the number of stocks is 2 then,

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1-\rho^2} \end{bmatrix}$$

Linear transformations of jointly distributed Gaussian r.v. is still Gaussian. X = AZ, where is a 2x1 vector of independent Gaussian r.n, we know that X will be Gaussian with variance. $V(\mathbf{X}) = \mathbf{A}V(\mathbf{Z})\mathbf{A}' = \mathbf{AI_2A}' = \mathbf{AA}' = \mathbf{\Sigma}$. This means that in order to perform Monte Carlo simulation we need to simulate Z_1 and Z_2 independently and according to a standard normal random variable and then to set (if the means are not zero) $X_1 = \mu_1 + \sigma_1 Z_1$, and $X_2 = \mu_2 + \rho\sigma_2 Z_1 + \sigma_2\sqrt{1-\rho^2}Z_2$

For general case $\mathbf{\Sigma} = \mathbf{AA}'$ where A is a symmetric lower triangular matrix, we obtain the following formula for the entries of A:

$$A_{jj} = \sqrt{\sigma_j^2 - \sum_{k=1}^{j-1} \sigma_{jk}^2} \tag{8}$$

$$A_{ij}\& = \frac{1}{A_{jj}}\left(\sigma_{ij} - \sum_{k=1}^{j-1} \sigma_{ik}\sigma_{jk}\right), \text{for } i > k \tag{9}$$

The expression under the square root is always positive if $\mathbf{\Sigma}$ is real and positive definite.

2.5 Simulating stock returns

While there are various theories developed to predict stock movement based on historical prices, Random Walk Theory suggests that there is no relationship between the current stock prices and the future prices. The fluctuation of the stock price is simply a random event but not influenced by past events.

2.5.1 Geometric Brownian Motion

Geometric Brownian motion (GBM), which is technically a Markov process. This means the stock price follows a random and is consistent with (at the very least) the weak form of the efficient market efficient market hypothesis (EMH)—past price information is already incorporated, and the next price movement is "conditionally independent" of past price movements. $\Delta S/S = \mu\Delta t + \sigma\epsilon\sqrt{\Delta t}$. Sis the stock price , ΔS is the change in stock price , μ is the expected return , ϵ is the random variable and Δt is the time period

if we rearrange the formula to solve just for the change in stock price, we see that GBM says the change in stock price is the stock price "S" multiplied by the two terms found inside the parenthesis below:$\Delta S = S \times (\mu\Delta t + \sigma\epsilon\sqrt{\Delta t})$ the first term is a "drift" and the second term is a "shock." For each time period, our model assumes the price will "drift" up by the expected return. But the drift will be shocked (added or subtracted) by a random shock. The random shock will be the standard deviation "s" multiplied by a random number "e." This is simply a way of scaling the standard deviation. We simulate the Gaussian random innovation simulating a Uniform and via the inverse of the Gaussian CDF $\epsilon(t) = \Phi^{-1}(U(0,1))$ Stock prices are simulated recursively according to

$$S(t + \Delta) = S(t) \times e^{r_\Delta(t)} \tag{10}$$

We can simulate the returns and prices in AR(1) model , where the model is autocorrelated

$$r_\Delta(t) = \rho r_\Delta(t - \Delta) + \sigma_\Delta \Phi^{-1}(U(0,1)). \tag{11}$$

For multivariate case to stimulate corelated asset prices we can Compute the Cholesky decomposition A of the covariance matrix Σ. Then we can draw random vector Z from a multivariate standard normal random variable (to do this we just need to simulate N independent standard normal random variables) and set $X = \mu + AZ$.

2.5.2 Monte Carlo Simulation

. The Monte Carlo method is a stochastic (random sampling of inputs) method to solve a statistical problem, and a simulation is a virtual representation of a problem. This methodology has several similarities to historical simulation. The main difference is that rather than carrying out the simulation using the observed changes in the market factors over the last *N* periods to generate *N* hypothetical portfolio profits or losses, one chooses a statistical distribution that is believed to adequately capture or approximate the possible changes in the market factors. Then, a pseudo-random number generator is used to generate thousands or perhaps tens of thousands of hypothetical changes in the market factors. These are then used to construct thousands of hypothetical portfolio profits and losses on the current portfolio, and the distribution of possible portfolio profit or loss. Finally, the value at risk is then determined from this distribution.

Monte Carlo is an attempt to predict the future many times over. At the end of the simulation, thousands or millions of "random trials" produce a distribution of outcomes that can be analysed. The basics steps are as follows:

- Specify a Model - Geometric Brownian Motion as explained in section 2.5.1
- Generate Random Trials
- Process The Output

2.6 Option pricing

Black-Scholes model, aka the Black-Scholes-Merton (BSM) model, is a differential equation widely used to price options contracts. The Black-Scholes equation requires five variables. These inputs are volatility, the price of the underlying asset, the strike price of the option, the time until expiration of the option, and the risk-free interest rate

The Black-Scholes call option formula is calculated by multiplying the stock price by the cumulative standard normal probability distribution function. Thereafter, the net present value (NPV) of the strike price multiplied by the cumulative standard normal distribution is subtracted from the resulting value of the previous calculation.

The risk-free rate and volatility of the underlying asset are known and constant, The returns of the underlying asset are normally distributed. The option is European and can only be exercised at expiration.

$$C = SN(d_1) - Ke^{-rt}N(d_2) \quad (12)$$

$$d_1 = \frac{\ln\frac{S}{K} + \left(r + \frac{\sigma_v^2}{2}t\right)}{\sigma_s\sqrt{t}} \quad (13)$$

$$d_2 = d_1 - \sigma_s\sqrt{t} \quad (14)$$

$$P = Ke^{-rt}N(-d_2) - SN(-d_1) \quad (15)$$

Where C is price of a call option, P is price of a put option , S is price of the underlying asset ,K is strike price of the option, r is the rate of interest t is time to expiration σ_v^2 is the volatility of underlying asset [5]

Binomial option pricing model is an options valuation method developed in 1979. The binomial option pricing model uses an iterative procedure, allowing for the specification of nodes, or points in time, during the time span between the valuation date and the option's expiration date.

A Multi-Step Binomial Model To produce a more practical model, we need to value the option under the assumption that there are many revaluations of the stock between the time t = 0, when the option is written, and the time t = τ , when it expires. This will give rise to a wide range of possible eventual prices. To begin the generalisation to a multistep model, we may consider a model with two steps. In the first step, the spot price of the stock moves either up to $S^u = S_0U$ or down to $S^d = S_0D$.

In the next step, the price can move up or down from these values to give $S^{uu} = S_0U^2$, $S^{ud} = S_0UD$ and $S^{dd} = S_0D^2$. The following table displays these outcomes: [5]

		$S^{uu} = S_0U^2$
	$S^u = S_0U$	
S_0		$S^{ud} = S_0UD$
	$S^d = S_0D$	
		$S^{dd} = S_0D^2$

Table 1:Up and down movement of stock

The method of pricing the call option is to start at the time of expiry and to work backwards so as to derive prices for the option at the intermediate nodes of the binomial tree. From these, one can derive the price cτ|0 of the option at the base of the tree. Let the values of the option corresponding to the outcomes S^{uu}, S^{ud} and S^{dd}be denoted by C^{uu}, $C^{ud}$and C^{dd}, respectively, and let those corresponding to the intermediate outcomes Su, Sd be cu, cd. (Here, we are omitting the temporal subscripts for ease of notation.) Then,

$$c^u = e^{-r\tau/2}\{c^{uu}p + c^{ud}(1-p)\}, c^d = e^{-r\tau/2}\{c^{ud}p + c^{dd}(1-p)\} \qquad (16)$$

From these, we derive

$$\begin{aligned} c_{\tau|0} &= e^{-r\tau/2}\{c^u p + c^d(1-p)\} \\ &= e^{-r\tau}\{[c^{uu}p + c^{ud}(1-p)]p + [c^{ud}p + c^{dd}(1-p)](1-p)\} \\ &= e^{-r\tau}\{c^{uu}p^2 + 2c^{ud}p(1-p) + c^{dd}(1-p)^2\} \end{aligned} \qquad (17)$$

The generalisation to n sub periods is as follows:

$$\begin{aligned} c_{\tau|0} &= e^{-r\tau}\left\{\sum_{j=0}^{n} \frac{n!}{(n-j)!j!} p^j (1-p)^{n-j} c^{uj,d(n-j)}\right\} \\ &= e^{-r\tau} E(c_{\tau|\tau}) \end{aligned} \qquad (18)$$

Here, $(c_{\tau|\tau})$= $c^{uj,d(n-j)}$ is the value of the option after the price has reached the value of S_τ= $S_0 U^j D^{n-j}$ by moving up j times and down n − j times. This is given by

$$\begin{aligned} c^{uj,d(n-j)} &= max(S_\tau - K_{\tau|0}, 0) \\ &= max(S_0 U^j D^{n-j} - K_{\tau|0}, 0) \end{aligned} \qquad (19)$$

By subdividing the period [0, τ] into n sub periods, we succeed in generating a range of possible outcomes for the value of S_τ , which are n+1 in number. In fact, as n → ∞, the trajectory price tends to that of a geometric Brownian motion. [6]

3 Measuring Value at Risk

The two simplest approaches for linear positions (such as stocks) consists in adopting A parametric approach, such as the log-returns are distributed according to a Gaussian r.v.. This approach requires the estimation of the parameters (mean and standard deviation) of this distribution. 2 A non-parametric approach. In this case, we do not make any assumption and we use past data to build the empirical distribution (ie the histogram).

There are three steps in VaR calculations

- First, the probability of losses exceeding VaR, $1 - \alpha$, needs to be specified, with the most common probability level being 1%. . Theory provides little guidance about the choice of α; it is mainly determined by how the user of the risk management system wishes to interpret the VaR number.
- Second, the holding period (i.e., the time period over which losses may occur). This is usually one day, but can be more or less depending on particular circumstances.
- Third, and final step is identification of the probability distribution of the profit and loss of the portfolio. This is the most difficult and important aspect of risk modelling. The standard practice is to estimate the distribution by using past observations and a statistical model. The two most common approaches are the parametric and non-parametric methods.

3.1 Non-Parametric Var:

3.1.1 Historical Var

Historical simulations represent the simplest way of estimating the Value at Risk for many portfolios. In this approach, the VaR for a portfolio is estimated by creating a hypothetical time series of returns on that portfolio, obtained by running the portfolio through actual historical data and computing the changes that would have occurred in each period. Even if this approach makes no explicit assumptions on the distribution of portfolio returns, an implicit assumption is hidden behind this procedure: the distribution of portfolio returns doesn't change within the window. From this implicit assumption several problems derive.

To run a historical simulation, we begin with time series data on each market risk factor, just as we would for the variance-covariance approach. However, we do not use the data to estimate variances and covariances looking forward, since the changes in the portfolio over time yield all the information you need to compute the Value at Risk. The approach for bootstrap one day Var is pretty simple Given a sample of size T of daily returns, $(r_1, ..., r_T)$, it could for a single assets or for portfolio of assets ,we calculate the weighted portfolio returns according equation (6), the non-parametric approach consists in estimating the VaR by using the sample $(1 - \alpha)$ percentile. If $(r_1, ..., r_T)$ returns is sample of iid returns . we can sort the returns (order statistics) $r(1) \leq r(2) \leq ...r(T-1) \leq r(T)$, with being r_1 the sample minimum and being the $r_{(T)}$ sample maximum .

The empirical VaR estimator at the desired confidence level is the order statistic of position $(1 - \alpha)$ T $-\widehat{VaR}_\alpha(t, t+\Delta) = r_{((1-\alpha)T)}$,If $(1 - \alpha)$ T is not an integer number, we use linear interpolation with weights $1 - \gamma$ and γ .$-\widehat{\mathrm{VaR}}_\alpha(t, t+\Delta) = (1-\gamma) \times r_{(\lfloor(1-\alpha)T\rfloor)} + \gamma \times r_{(\lfloor(1-\alpha)T\rfloor+1)}$

VaR is determined by the actual price movements. In other words, there are no underlying assumptions of normality driving the conclusion. The second is that each day in the time series carries an equal weight when it comes to measuring the VaR, a potential problem if there is a trend in the variability – lower in the earlier periods and higher in the later periods, for instance. The third is that the approach is based on the assumption of history repeating itself, with the period used providing a full and complete snapshot of the risks that the asset is exposed to in other periods

With Historical simulation there is no simple procedure to extrapolate a n-period VaR from the 1-period one and the square root rule is valid only for extrapolating the standard deviation under the zero-autocorrelation assumption, and it is a not valid rule to project the 1-period empirical quantile to n-periods. One possibility is to collect n-period returns and compute their sample quantile, but this implies a great loss of data. For example, in a sample of 250 daily returns we have only 12 monthly returns, so that the estimation of the monthly VaR is very inaccurate. So instead we can use bootstrapping methodology (Historical simulator) to estimate Var

3.1.2 Bootstrapping

Bootstrap is a technique introduced by Efron (1979), to make the most of limited data. The bootstrap technique draws a sample of the same size as the original data set and records the VaR from the simulated sample. This procedure is repeated over and over to obtain multiple sample VaRs, this procedure is like sampling with replacement. The best VaR estimate from the full data set is the average of all sample VaRs. it allows to generate an hypothetical distribution using past data without attaching any assumption to them.

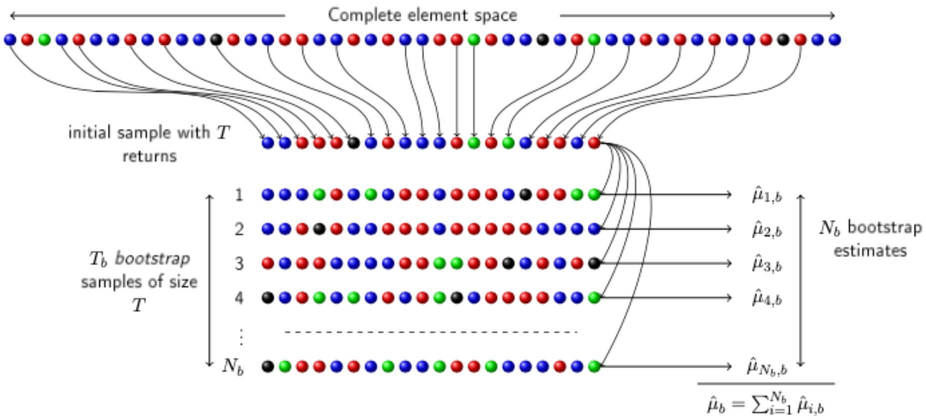


Figure 1: The Bootstrap estimation procedure of the unknown parameter m

When T (length) tend to infinity, the distribution of average values computed from bootstrap samples is equal to the distribution of average values obtained from ALL samples with T elements which can be constructed from the complete space. Thus the width of the distribution gives an evaluation of the sample quality.

The procedure used for one day var is as follows ,

1. Create a T × 1 database of historical one-period returns on a stock or a on a portfolio.
2. Randomly select a date $t \in [1, T]$ and use the return for that date to generate a possible future return.
3. Repeat, with replacement, T times and obtain a resampled sample of size T, having the same information as the original sample.
4. Compute the VaR as empirical quantile of the T simulated values. We call this estimate $\widehat{Var_{b,1}}(t, t+\Delta)$
5. Repeat steps 2-4 to generate M samples of size T each and in each sample compute $\widehat{Var_{b,m}}(t, t+\Delta)$
6. The bootstrap estimate of the 1-day VaR is the average of the M simulated var

$$V_a\widehat{R_b(t}, t+\Delta\Big) = \frac{1}{M}\sum_{m=1}^{M} V_a\widehat{R_{b,m}(t}, t+\Delta\Big) \qquad (20)$$

The procedure for n period Var from 1 period returns is similar to the above procedure , instead we obtain n (no of days) , one period returns with replacement and calculate the cumulative returns , we repeat to generate T paths to obtain an estimation of the real distribution of the portfolio ,having the same information as original sample and Var is calculated as empirical quantile of T simulated Values , The point estimate of Var is the average of all Vars This relies on the independence of successive returns. We can also construct the confidence interval around our point Var estimate, for M bootstrap estimates say 90% confidence is constructed using empirical (e.g. 5% quartile) α_L, and the imperial α_H (e.g 95% of quantile) . This method works well if the original sample is small and comes from a highly asymmetric distribution. The idea is that the variability of the bootstrap estimates around the sample estimate will be similar to (or mimic) the variability of sample estimate around the true parameter value. There is good reason to believe that this will be true for large sample sizes, since as T gets larger and larger, the empirical distribution comes closer and closer to the true distribution and so sampling with replacements from the empirical distribution is almost like random sampling from the true distribution.

Advantage of bootstrap is that it is simple to implement as no parameter estimation is needed, since it is a model-free procedure, so it does not provide misleading indications if the model is poor. It also Includes fat-tailed and skewed outcomes it is possible to extend the bootstrap model to capture autocorrelation and changing volatility. But it has some drawbacks, it is very sensitive to the length of the data sample used. In practice, 250 to 1000 daily observations are often used, but a larger sample size would be needed to get the same accuracy as in the Gaussian case, Time series available on some assets may be limited and bootstrap would return a bad estimate . It does not model changes in volatility, i.e. only provides an unconditional distribution. It also react slowly to changes in the market risk environment. Since square root rule is not valid for extrapolating the quantiles at long horizons, so the procedure can become computationally expensive.

3.2 Parametric Var

3.2.1 Variance co-variance approach

Since Value at Risk measures the probability that the value of an asset or portfolio will drop below a specified value in a particular time period, it should be relatively simple to compute if we can derive a probability distribution of potential values. That is basically what we do in the variance-covariance method, an approach that has the benefit of simplicity but is limited by the difficulties associated with deriving probability distributions.

Here the assumptions are that variance-covariance method assumes that asset returns are normally distributed around the mean of the bell-shaped probability distribution. Assets may have tendency to move up and down together or against each other. This method assumes that the standard deviation of asset returns and the correlations between asset returns are constant over time.

VaR calculation for a single asset It is straightforward. From the distribution of returns calculated from daily price series, the standard deviation (σ) under a certain time horizon is estimated. The daily VaR is simply a function of the standard deviation and the desired confidence level and can be expressed as:

$$\mathrm{Va}R_{\alpha}(t, t+n\Delta) = -\left(\mu_{\Delta} n + z_{1-\alpha}\sigma_{\Delta}\sqrt{n}\right) \tag{21}$$

For Var of a portfolio We cannot model using linear returns since the liner returns have a lower bound -1, as mention in [2.1] but the Gaussian distribution has unbounded support So, the simplest approach will be to model stock log-returns using a (multivariate) Gaussian distribution but the portfolio return will not be gaussian , a common approximation that holds over a very short time horizons allows us to model portfolio return according to a Gaussian distribution , this can use to retain The common practice it to retain the Gaussian distribution at stock and portfolio level and very good estimate over a short horizon but can have shortcomings over long horizon

$$r_p(t) \simeq \sum_{i=1}^{N} w_i r_i(t)$$

$$\mathrm{Va}R_{\alpha}(t, t+n\Delta) = -\left(\mu_{\mathrm{p}} n + z_{1-\alpha}\sigma_{\mathrm{p}}\sqrt{n}\right) \tag{22}$$

Investors can estimate the probable loss value of their portfolios for different holding time periods and confidence levels. The variance–covariance approach helps us measure portfolio risk if returns are assumed to be distributed normally. However, the assumptions of return normality and constant covariances and correlations between assets in the portfolio may not hold true in real life.

This method is top down approach , we specifies the portfolio P&L distribution without reference to the constituents: The goal a model that is more parsimonious than the bottom up approach but this approach is not fine enough to identify, in the case of a VaR violation, the identity of the component that contributed the most to this violation. To calculate Var we collect past data relative to the portfolio component, given the current portfolio composition, reconstruct the past history of the portfolio and log returns of portfolio is computed under the assumption of log-return to be normally distributed, estimate the mean and the standard deviation and compute the VaR at the desired confidence level and time horizon.

For Bottom up approach The portfolio is an aggregate of the components, which the specification of the dependence structure (i.e. the joint distribution) of the components and how this builds up into the portfolio distribution. Here we assign the joint distribution of log-returns of different stocks and obtain then the portfolio distribution then we compute the Risk measure. The difficult step is to assign the joint distribution: this requires to assign all possible cross assets dependencies and then to estimate them. log-returns of different stocks are assumed to have a jointly normal distribution $\mathbf{r}(t, t+\Delta) \sim \mathcal{N}(\mu_\Delta, \Sigma_\Delta)$The distribution over n − periods will be, using the scaling rule (i.e. under the assumption of zero serial autocorrelation) $\mathbf{r}(t, t+n\Delta) \sim \mathcal{N}(\mu_\Delta \times n, \Sigma_\Delta \times n)$

4 Measuring Value at Risk

In the parametric Gaussian approach, we need to estimate the volatility parameter. A possible approach is to estimate it using historical volatility (i.e. the sample standard deviation). The limit of this approach is that, being an unconditional estimator, in principle it cannot react to market shocks: old and recent observations have the same importance. Unfortunately, an unconditional volatility estimator, such as the sample variance, cannot capture time-variation in the volatility. Let us consider the time series of the 100 largest losses (standardized returns below zero more than 3 standard deviation)

There is little evidence for non-zero autocorrelation for the returns: the uncorrelatedness of the returns occurs because given yesterday's return, today's return is equally likely to be positive or negative Significant autocorrelations at several lags for the squared and absolute returns: the correlated Ness occurs because yesterday's squared return and today's squared return are likely to be very similar in value. The autocorrelation estimates can be used to test the hypothesis that the process generating observed returns is a series of independent and identically distributed (iid) variable. The asymptotic standard error of an autocorrelation estimate is approximately $1/\sqrt{T}$

The autocorrelations for returns are not significantly different from zero and fall inside the $\pm\ 1/\sqrt{T}$ and autocorrelations for squared and absolute returns are significantly different from zero at several lags. This suggests that the series are not generated by an i.i.d. process.

Conditional and Unconditional Estimator -The Problem is that The sample variance is an unconditional estimator of the variance Reshuffling the sample does not change the variance estimate it is an very old estimate and very recent observations contribute in the equal measure to the computation of the variance. when we compute the sample variance all observations receive the same weight (1/T) If we try to estimate a changing volatility using the sample variance and a rolling sample, we only produce the so-called ghost effect, i.e. a change in the volatility estimate due to the adopted sampling procedure. Better models are needed allowing for changes

4.1 EWMA

allowing for changes for example assigning larger weights to more recent observations , generating autocorrelation in squared returns (this allows for clustering in volatility) The exponential weighted moving average (EWMA) or exponential smoothing procedure solves issues of Auto correlation . This model belongs to the class of so called generalized autoregressive conditional heteroskedastic (GARCH) models, characterized by being a conditional varying volatility model here one-period log-returns are assumed to be Gaussian with zero mean and time-varying conditional variance where σ^2 (t, t + Δ) is the conditional variance of the next-period returns

$$r(t, t+\Delta) \mid \mathcal{F}_t \sim \mathcal{N}(0, \sigma^2(t, t+\Delta))$$

$$\sigma^2(t, t+\Delta) = \mathbb{V}_t(r(t, t+\Delta))$$

In particular, given a smoothing parameter λ, $0 < \lambda < 1$, the variance is assumed to be a weighted average of past squared returns rule:

$$\begin{aligned} \sigma^2(t, t+\Delta) &= (1-\lambda)(r_{t-\Delta}^2 + \lambda r_{t-2\Delta}^2 + \lambda^2 r_{t-3\Delta}^2 + \cdots) \\ &= (1-\lambda) \textstyle\sum_{j=0}^{\infty} \lambda^j r^2(t-(j+1)\Delta, t-j\Delta) \end{aligned} \tag{23}$$

The EWMA approach has the appealing property that , the observation that is 1-day old receives a weight of $1 - \lambda$. The observation that is 2-day old receives a weight of $(1 - \lambda)\lambda$. The weight of an observation that is j + 1 days old is $(1-\lambda) \times \lambda^j$ Recent observations receive a large weight, whilst distant observations receive a small weight ,This is different from the sample variance, where all observations are equally weighted and receive a weight equal to 1/T . Here the λ is the smoothening parameter , and thsum of the weights is normalized to 1 . As we lower the value of λ. we give more weight to the most recent As we increase the value observations, and we are of λ we start to give the same weights to all back to the unconditional case.

For updating the variance estimate We can rewrite the EWMA estimate according to the following updating rule

$$\underbrace{\sigma^2(t, t+\Delta)}_{\text{today forecat}} = (1-\lambda) \times \underbrace{r^2(t-\Delta, t)}_{\text{yesterday squared mkt shock}} + \lambda \times \underbrace{\sigma^2(t-\Delta, t)}_{\text{yesterday forecast}} \tag{24}$$

This means that higher λ, higher the persistence in the variance trough the term $\lambda\sigma 2$ (t − Δ, t) and lower λ, higher the reaction to market shocks trough the term (1−λ)^2 (t −Δ,t). In order to use the EWMA procedure we need two ingredients the starting value of the recursion, σ^2 (0, Δ) an estimate of the parameter λ. One way to is to estimate σ^2 (0, Δ) by using the sample variance relative to a pre sample of additional K observations. And f we are dealing with daily returns, to set $\lambda = 0.94$,

Another possibility it to estimate λ and σ^2 (0, Δ) by maximum likelihood In the EWMA model we are assuming that one-period ahead log-returns are Gaussian but with a conditionally time varying variance: $r_t \mid \mathcal{F}_t \sim \mathcal{N}(0, \sigma_t^2), t = 0, \Delta, \cdots, (T -$

1)Δwhere [r_t = r (t, t + Δ) (and similarly for $\sigma_t{}^2$]) and F_t represents all information available up to time t. The sample likelihood, i.e. the probability of observing the sample. Maximum likelihood means to choose the parameters in order to maximize this probability.

$$\begin{aligned} &\log \mathcal{L}\left(r_0, r_\Delta, \dots, r_{(T-1)\Delta} \mid \lambda, \sigma_0^2\right) \\ &= \sum_{j=0}^{T-1} \ln\left(\frac{1}{\sqrt{2\pi\sigma_{j\Delta}^2}} e^{-\frac{1}{2}\left(\frac{r_{j\Delta}}{\sigma_j}\right)^2}\right) \\ &= \sum_{j=0}^{T-1}\left(-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\sigma_{j\Delta}^2\right) - \frac{1}{2}\left(\frac{r_{j\Delta}}{\sigma_{j\Delta}}\right)^2\right) \end{aligned} \tag{25}$$

Then we can maximize this expression with respect to the unknown 2 parameters λ and σ_0 . In EWMA daily returns model are uncorrelated, but are not independent, because they are characterized by a serial dependence trough the variance dynamics. This implies that, one day returns are conditionally Gaussian, the n-periods log-return is not Gaussian. Therefore, the VaR formula we have obtained for one-day returns does not apply to n − periods returns in the EWMA model. The distribution of cumulative returns can be obtained only trough Monte Carlo simulation.

We can extend to the multivariate case the EWMA recursion , This allows us to take into account time-varying volatility and time varying covariance. We assume that returns are conditionally Gaussian with time varying covariance matrix

$$\mathbf{r}(t, t+\Delta) \mid \mathcal{F}_t \sim \mathcal{N}(\mathbf{0}, \Sigma^{EWMA}(t, t+\Delta))$$

$$\Sigma^{EWMA}(t, t+\Delta) = \lambda\Sigma^{EWMA}(t-\Delta, t) + (1-\lambda)\mathbf{r}(t-\Delta, t)\mathbf{r}'(t-\Delta, t) \tag{26}$$

we have a unique value of λ for all entries of the covariance matrix. This guarantees the covariance matrix to be positive definite. λ can be estimated via Maximum Likelihood.

4.2 GARCH

The ARCH or Autoregressive Conditional Heteroskedasticity method provides a way to model a change in variance in a time series that is time dependent, such as increasing or decreasing volatility. An extension of this approach named GARCH or Generalized Autoregressive Conditional Heteroskedasticity allows the method to support changes in the time dependent volatility, such as increasing and decreasing volatility in the same series. Autoregressive models can be developed for univariate time series data that is stationary (AR), has a trend (ARIMA), and has a seasonal component (SARIMA). One aspect of a univariate time series that these autoregressive models do not model is a change in the variance over time. There are some time series where the variance changes consistently over time. In the context of a time series in the financial domain, this would be called increasing and decreasing volatility. In time series where the variance is increasing in a systematic way, such as an increasing trend, this property of the series is called heteroskedasticity.It's a fancy word from statistics that means changing or unequal variance across the series. If the change in variance can be correlated over time, then it can be modelled using an autoregressive process, such as ARCH. [4]

ARCH are mean zero, serially uncorrelated processes with nonconstant variances conditional on the past, but constant unconditional variances. For such processes, the recent past gives information about the one-period forecast variance. The model should only be applied to a residual series that is uncorrelated and contains no trends or seasonal changes, such as might be obtained after fitting a satisfactory SARIMA model.Generalized Autoregressive Conditional Heteroskedasticity, or GARCH, is an extension of the ARCH model that incorporates a moving average component together with the autoregressive component. Specifically, the model includes lag variance terms (e.g. the observations if modelling the white noise residual errors of another process), together with lag residual errors from a mean process.The introduction of a moving average component allows the model to both model the conditional change in variance over time as well as changes in the time-dependent variance. [4]

As such, the model introduces a new parameter "p" that describes the number of lag variance terms:

- **p**: The number of lag variances to include in the GARCH model.
- **q**: The number of lag residual errors to include in the GARCH mode

GARCH(1,1) $\sigma^2(t) = \omega + \alpha \times r^2(t-\Delta) + \beta \times \sigma^2(t-\Delta)$, EWMA has a single parameter to control both market reaction and volatility persistence; GARCH(1,1) separates the two effects. In addition, ω controls the long-run variance level. EWMA is a nonstationary version of the GARCH(1,1) model, given that the persistence parameters sum to 1. EWMA the unconditional variance does not exist, because the stationarity condition $\alpha + \beta < 1$ is not satisfied and has a flat term structure of variance forecast while GARCH(1,1) generates mean reversion to the long run level

$$\sigma_L^2 = \frac{\omega}{1-(\alpha+\beta)} (\text{ if } \alpha + \beta < 1) \tag{27}$$

The autocorrelations of r_t^2 are found to be

$$\rho_1 = \alpha_1 + \frac{\alpha_1^2\beta_1}{1-2\alpha_1\beta_1-\beta_1^2}$$
$$\rho_k = (\alpha_1 + \beta_1)^{k-1}\rho_1, k = 2,3 \ldots \quad (28)$$

If the sum $\alpha 1$ + $\beta 1$ is close to 1, the autocorrelations will decrease only very gradually.

4.3 Coherent Risk Measure

4.3.1 Limitations of Var

Var does not describe the maximum loss or the losses in the left tail. 99% percent VAR really means that in 1% of cases (that would be 2-3 trading days in a year with daily VAR) the loss is expected to be greater than the VAR amount. Value At Risk does not say anything about the size of losses within this 1% of trading days and by no means does it say anything about the maximum possible loss. The reported VaR number is an estimate of the true VaR and therefore is affected by estimation error, i.e. the sampling variability due to limited sample size, adding new observations will change the results.

The accuracy of the estimate depends mainly on two elements:

- the sample size used to estimate the VaR: larger the sample size better the accuracy of the VaR estimate.
- the confidence level: larger the confidence level lower the accuracy of the VaR estimate;

4.3.2 Expected Shortfall

Let X and Y be two portfolios and let ρ to be a risk measure. We say that ρ, a function defined on the space of random variables and having real values, is a coherent risk measure if it satisfies the Monotonicity Subadditivity Positive omogeneity and Translation invariance

The Expected Shortfall is a coherent risk measure Expected shortfall averages the P&L scenario returns beyond the VaR threshold, so it provides additional information as to how severe the losses can be. Mathematically, Expected Shortfall is the average of the $(1-\alpha)$ % worst losses:

$$ES_\alpha^{pl} = -\mathbb{E}\big(P\&L(t,T) \mid P\&L(t,T) \leq -V_a R_\alpha^{p/}(t,T)\big) \quad (29)$$

It can be defined in terms of log-returns

$$ES_\alpha = -\mathbb{E}(r(t,T) \mid r(t,T) \leq -V_a R_\alpha(t,T))$$

5 VaR for a derivative position

If we can price the derivative using a closed form formula, such as the Black-Scholes one, and the pricing formula is a monotonic function of the underlying risk factor (price), a simple formula for the VaR is available. If the derivative price is a monotonic increasing function of the risk factor. Therefore, a loss on the risk factor will imply a loss on the derivative position and if the derivative price is a monotonic decreasing function of the risk factor. Therefore, a gain on the risk factor will imply a loss on the derivative position.

5.1 Monotonic Increasing Case-

If the derivative price is an increasing function of the risk factor (e.g. a call option), To calculate Value at risk , we can Price the derivative contract at the current market condition C(P(t),t,T) is the price of the derivative contract expiring in T, as function of the underlying risk factor P(t), here assumed to be the price, and current time t , then we can Determine the worst risk factor scenario at the assigned horizon and confidence level $1 - \alpha$. For example, assuming Gaussian returns $P_{\text{worst}}(t + \Delta n) = P(t)e^{\mu_\Delta n + z_{1-\alpha}\sigma_\Delta\sqrt{n}}$

Then we can revaluate the derivative position at the worst case scenario above and at the VaR horizon $C(P_{\text{worst}}(t + \Delta n), t + \Delta n, T)$ The derivative VaR is obtained by the difference

$$\text{VaR}_\alpha^{\text{derivative}}(t, t + \Delta n) = |C(P(t), t, T) - C(P_{\text{worst}}(t + \Delta n), t + \Delta n, T)| \tag{29}$$

where t + nΔ is the VaR horizon. This method is possible for a limited number of contracts for which there is a monotonic relationship between risk factor and derivative contract

5.2 Full revaluation via Monte Carlo simulation

VaR can be estimated combining Monte Carlo simulation with full revaluation as follows. Let C(P,t) := C(P,t,T) the pricing function that relates the risk factor P to the derivative price at time t

We Price the derivative position using the current value of the market factors P(t). This requires the computation of C(P(t),t). Then Simulate log-returns via our preferred model, either parametric Gaussian distribution or Historical Simulation or some other more advanced model.

So we obtain M simulated scenarios then we Obtain the simulated risk-factor price at the VaR horizon $P^i(t+n\Delta) = P(t) \times e^{r^i(t,t+n\Delta)}$ then Obtain the simulated risk-factor price at the VaR horizon t + n∆ $C^i_{t+n\Delta} = C(P^i(t+n\Delta), t+n\Delta)$The VaR is calculated as the appropriate quantile of the M simulated P &L.

6 Backtracking

Backtesting is the process of comparing losses predicted by a value at risk (VaR) model to those actually experienced over the testing period. It is done to ensure that VaR models are reasonably accurate. Risk managers systematically check the validity of the underlying valuation and risk models by comparing actual to predicted levels of losses.

The overall goal of backtesting is to ensure that actual losses do not exceed the expected losses at a given level of confidence. **Exceptions** are the number of actual observations over and above the expected level. In the context of the VaR, the number of exceptions falling outside of the VaR confidence level should not exceed one minus the confidence level.

The problem of determining the accuracy of a VAR model can be reduced to the problem of determining whether the hit sequence, I_t (α) , t = 1, ..., nΔ, satisfies two key properties: Unconditional coverage property, and Independence property.

6.1 Unconditional coverage property

The probability of realizing a loss in excess of the reported VaR, must be precisely $(1-\alpha) \times 100\%$, I e$\mathrm{Pr}_t\ (I_{t+\Delta}(\alpha)=1)=1-\alpha$

If it is the case that losses in excess of the reported VAR are expected to occur more frequently than $(1-\alpha) \times 100\%$ of the time, then this would suggest that the reported VAR measure systematically understates the portfolio's risk. The opposite finding that VAR violations are expected to occur too infrequently would, alternatively, signal an overly conservative VAR measure.

A statistical test for unconditional coverage is the Kupiec Test. It examines the number of violations but it does not consider if they cluster in time or not, i.e. if the independence assumption is satisfied. The Kupiec test is also called probability of failure (POF) test. It is based on the fact that a violation can be modelled as a Bernoulli random variable with probability $1-\alpha$. If p (j) the probability of having j hits given the sample of size n:

$$p(j)=\mathrm{Pr}\ (\textstyle\sum_{i=1}^{n}\ I_{i\Delta}(\alpha)=j). \qquad (30)$$

Given that $1-\alpha$ is the theoretical probability of having a violation in each trial, and that $\sum^{n}_{i=1} I_{i\Delta}(\alpha)$ is a binomial (sum of iid Bernoulli's) random variable, it follows that the theoretical probability of having j violations out of n trials is:

$$L(j,n,\alpha) = \binom{n}{j} \times (1-\alpha)^{j} \times \alpha^{n-j}. \tag{31}$$

Let $1-\alpha$ to be the observed violation frequency It follows that the estimated probability of having j violations out of n trials is:

$$L(j,n,\hat{\alpha}) = \binom{n}{j} \times (1-\hat{\alpha})^{j} \times \hat{\alpha}^{n-j} \tag{32}$$

Then, we can construct the log-likelihood ratio Under the null hypothesis that the VaR model is good, ◎α should not be too different from α, or equivalently the LR_{uc} should have values near to 0.

$$LR_{uc} = -2\ln\left(\left(\frac{1-\alpha}{1-\hat{\alpha}}\right)^{j} \times \left(\frac{\alpha}{\hat{\alpha}}\right)^{n-j}\right) \sim \chi^2_1 \tag{33}$$

The equality can be tested, considering that for n large and under H_0: $\hat{\alpha} = \alpha$ Therefore the model is well calibrated if LR_{uc} is less than the critical value of the chi-square distribution with 1 degree of freedom. The critical level is assigned on the basis of the significance level of the test. Alternatively, we can calculate the P-value defined as the probability of getting a sample with a higher LR than the one observed:

6.2 Independence Property

Any two elements of the hit sequence, $\left(I_{t+j\Delta}(\alpha), I_{t+k\Delta}(\alpha)\right), k \neq j$ must be independent random variable. This condition requires that the previous history of VAR violations must not convey any information about whether or not an additional VAR violation If, for example, a VAR violation is more likely to occur after a previous VAR violation has occurred, then this implies that the probability that $I_{t+\Delta}(\alpha)$ conditional on the event that $I_t(\alpha) = 1$ exceeds $1-\alpha$, indicating that the reported VAR is too small and should be increased.

For a good Var model The unconditional coverage and independence properties of the hit sequence are separate and distinct and must both be satisfied by an accurate VAR model.

The probabilities of two consecutive violations is p_{11} and The probability of a violation if there was no violation on the previous day p_{01} more $p_{ij} = \Pr\left(\eta_t = j \mid \eta_{t-1} = i\right)$ where i and j are either 0 or 1 . The violation process can be represented as a Markov chain with two states . the first order transition probability matrix is defined as:

$$\Pi_1 = \begin{pmatrix} 1 - p_{01} & p_{01} \\ 1 - p_{11} & p_{11} \end{pmatrix}$$

If v_{ij} is the number of observations where j follows i The likelihood function is:

$$L_1(\Pi_1) = (1 - p_{01})^{v_{00}} p_{01}^{v_{01}} (1 - p_{11})^{v_{10}} p_{11}^{v_{11}}$$

Under the null hypothesis of no clustering, the probability of a violation tomorrow does not depend on today being a violation hen p01 = p11 = p and the transition matrix is simply:

$$\widehat{\Pi}_2 = \begin{pmatrix} 1 - \hat{p} & \hat{p} \\ 1 - \hat{p} & \hat{p} \end{pmatrix}$$

$$L_2(\Pi_2) = (1 - p)^{v_{00} + v_{10}} p^{v_{01} + v_{11}}$$

Replacing the Π by the estimated numbers $\widehat{\Pi}$ The LR test is then

$$LR = 2\left(\log\, L_1(\widehat{\Pi}_1) - \log\, L_2(\widehat{\Pi}_2)\right) \overset{\text{asymptotic}}{\sim} \chi^2_{(1)} \tag{34}$$

6.3 Stress testing

We Create artificial market outcomes and to see how risk management systems and risk models cope with the artificial event Assess the ability of a bank to survive a large shock The main aim is to come up with scenarios that are not well represented in recent historical data but are nonetheless possible and detrimental to portfolio performance

There are several ways to do that, here is a really simple approach , Suppose we have a sample 1,...,W_E,...,T where W_E , We have a VaR_{t+1} The stressed VaR is $\mathrm{VaR}^S_{t+1} = max \mathrm{VaR}_i i = W_E + 1, \ldots, T + 1$

7 Experiment Analysis of Results

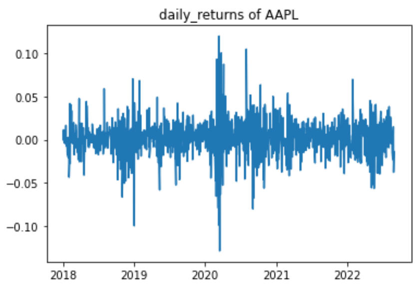
7.1 Datasets

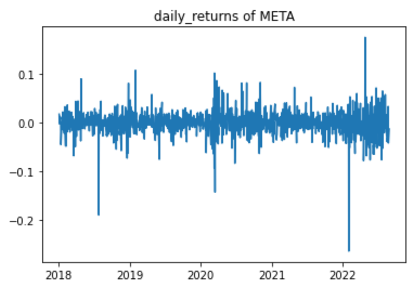
I used yahoo finance to fetch the real data on Asset prices. To fetch the data the parameters required are the Ticker Symbol or a list of Ticker Symbol (it is an abbreviation used to uniquely identify publicly traded shares of a particular stock on a particular stock market) then Start and End date for the prices. It returns a Pandas object with columns like High, Low, Close. In our analysis we are concerned about Closing price of asset. For my Analysis in the project, I have portfolio a portfolio of Apple, Facebook, Google and Amazon Stocks from year 2018 to 2021 which is four year of data and for single asset analysis I fetched data for S&P 500 (ticker symbol **^GSPC)** **.we** can construct portfolio of any stock by changing the ticker variable in the code.

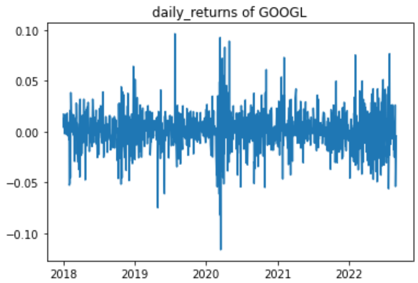
Choice of Programming Language: Python is the language of choice since python has many packages for feting real time financial data , and statistical and machine learning packages for time series analysis

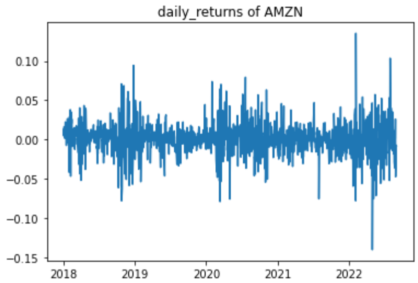
7.2 Initial Analysis for time series analysis for each stock return

Once we have fetched the data from yahoo finance the we need to perform initial analysis for each asset to get characteristics like distribution and moments of stock. We plot returns of each asset







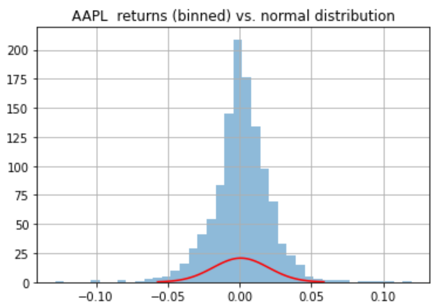


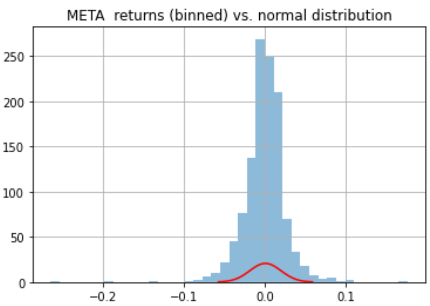
The plot shows that the returns might be stationary but there is a indication of volatility clustering (there are periods of low volatility and periods where volatility is high) . To statistically test for stationarity we can use Augmented Dickey-Fuller test .Form the test statistics , since the P-values are very low (even if our alpha is 0.01) suggest to Reject the null hypothesis (H0), and hence the data does not have a unit root and is stationary for all the assets in our portfolio

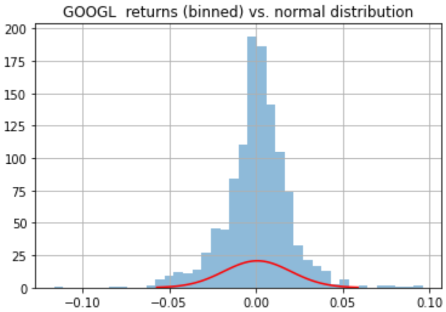
Asset	Test statistic	P-value
AAPL	-10.779	2.2125e-19
META	-37.086	0.0
GOOGL	-11.426	6.6745
AMZN	-11.561	3.2553e-21

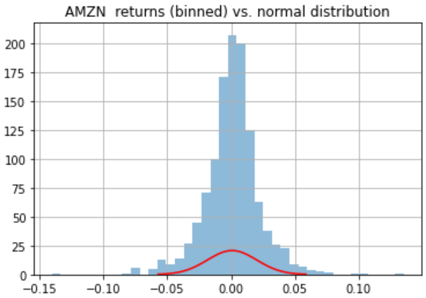
Table 2 :Dickey-Fuller test for each assest

We can plot the histogram of frequency of returns and see the distribution of the results . We can also perform normality test to see if our returns are normally distributed using statistical test









We can perform normality statistical test using Shapiro-Wilk Test for normality

Asset	Test statistic	P-value
AAPL	0.946	2.739e-20
META	0.8929	8.788e-28
GOOGL	0.9502	1.722e-19
AMZN	0.95246	4.870e-19

Table 3: Shapiro-Wilk Test for each assets

Since all the P-values are very low (even if alpha is 0.1) we can conclude the return are not normally distributed.

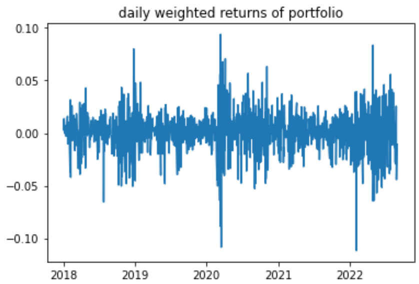
We can calculate sample moments of each asset, the mean return are AAPL 0.00133 , META 0.000220 GOOGL 0.000791 AMZN 0.000900and the variance covariance matrix is

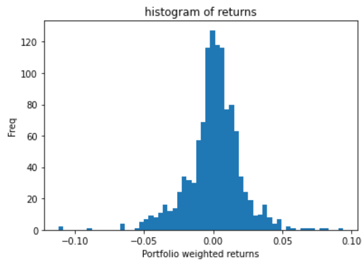
Symbols	**AAPL**	**META**	**GOOGL**	**AMZN**
Symbols				
AAPL	0.000433	0.000319	0.000278	0.000296
META	0.000319	0.000669	0.000342	0.000352
GOOGL	0.000278	0.000342	0.000375	0.000286
AMZN	0.000296	0.000352	0.000286	0.000482

Table 4 :variance co-variance of assets

7.3 Var calculations

I have used weights of .25 for Apple stock and .3 for Facebook stock and .15 for google and .3 for amazon, all these weight add up to 1. And initial investment of $1,000,000 i.e. initial value of our portfolio of one million .We can get the portfolio returns according to formula [4] Then we can find the mean returns and variance covariance matric by [5] .We can also plot the portfolio returns and perform test for stationarity





With ADF Statistic for portfolio: -11.505 very low   p-value: 4.37e-21 suggest the portfolio returns are stationary, now we can perform the sample estimates .The portfolio mean is 0.0007872 with standard deviation of 0.019289

7.3.1 Historical simulation

Algorithm:

Sorted_weighted _portfolio returns = sort(weighted portfolio returns)

Value_At_risk =quantile (sorted weighted portfolio returns, alpha)

Value_At_Risk(in $ tems) = Value_At_risk * Initial_Investment

Expected_shorfall= mean(flor all Sorted_weighted _portfolio returns < = Value at RisK)

Below is a table of value at Risk one Day using historical simulation of above portfolio at confidence of 0.05 (alpha)

Alpha	Value at Risk	Expected shortfall
0.1	$ -21867.6	$ -36844.4
0.2	$ -33709.9	$ -46513.2
0.3	$ -36496.3	$ -49458.6
0.4	$ -40253.6	$ -52893.0
0.5	$ -44235.4	$ -58136.0
0.6	$ -50273.8	$ -69099.7

Table 4 : bootstrap simulation value at risk

7.3.2 Bootstrap Approach

Algorithm:

Returns_1day(NumPy array of size M, number of days, length of returns)

=Rodom.Choice(Portfolio_weighted Returns)

Cummulative_Returns_nday= Returns_1day.sum(axis=1)

Sorted_nday_Cummulative_Returns=sort(Cummulative_Returns_nday,axis=1)

N_day_var = quantile(Sorted_nday_Cummulative_Returns,alpha ,axis=1).mean

N_day_var($ terms) = N_day_var * Initial Inverstment

We can also construct confidence interval around our point Value at risk at a certain confidence level say beta which is different from the Var confidence interval alpha by as discussed in section 3.1.2

Number of days	Point Estimate	Confidence Interval
1	-33673.15	[-30015.6, -36575.0]
2	-53094.17	[-48890.9, -57656.9]
3	-67913.41	[-62706.0, -73675.4]

Table 5 :bootstrap point var with confidence interval

7.3.3 Value at risk using Variance Covariance Approach

Value _at-risk =norm.ppg (alpha , No_days * Port_Mena, sqrt(No_days) *portfolio_standard_deviation) .If we have computed one day var we can square root rule to estimate n day Var .

The below graph shows value at risk and expected shortfall for above portfolio for different time period

```
1 day VaR and expected shortfall@ 0.05% confidence: -30941.48  -39001.87
2 day VaR and expected shortfall@ 0.05% confidence: -43296.7  -54695.8
3 day VaR and expected shortfall@ 0.05% confidence: -52594.01  -66555.0
4 day VaR and expected shortfall@ 0.05% confidence: -60308.44  -76429.2
5 day VaR and expected shortfall@ 0.05% confidence: -67011.32  -85034.88
```

Figure 4 : value at risk using VCV and expected shortfall for different time periods

7.3.4 Montecarlo simulation

To simulate the corelated stock price we first calculate the variance co variance matric for the portfolio, To simulate correlated returns we use need to calculate Cholesky decomposition, for our variance covariance matrix then we draw random variables Number of assets * time period vector of independent Gaussian r.n to simulate corelated return and commute the cumulative return of n time period . we repeat the process for M say 1000 simulation and calculate the Var at the desired confidence level. by averaging all vars calculated

```
: array([[0.02080616, 0.        , 0.        , 0.        ],
        [0.01530981, 0.02085185, 0.        , 0.        ],
        [0.01334438, 0.00659486, 0.01238155, 0.        ],
        [0.01424347, 0.00644283, 0.00432943, 0.01478912]])
```

Figure 5 : Cholesky decomposition of VCV matrix

The below graph shows value at risk and expected shortfall for above portfolio for different time period

```
1 day VaR and expected shortfall@ 0.05% confidence: 62158.34  81308.9
2 day VaR and expected shortfall@ 0.05% confidence: 68202.13  85012.21
3 day VaR and expected shortfall@ 0.05% confidence: 68480.03  85000.86
4 day VaR and expected shortfall@ 0.05% confidence: 68422.72  89981.64
5 day VaR and expected shortfall@ 0.05% confidence: 66256.17  86446.32
```

Figure 6 : value at risk and expected shortfall for different time periods

7.3.5 Var using EWMA approach

Here the initial EWMA variance covariance is calculated using standard approach but after each simulated return the EWMA variance covariance is updated using the equation (24). Then the cumulative returns are calculated by Cholesky decomposition. we repeat the process for M say 1000 simulation and calculate the Var at the desired confidence level. by averaging all vars calculated

The below graph shows value at risk and expected shortfall for above portfolio for different smoothening parameter lambda

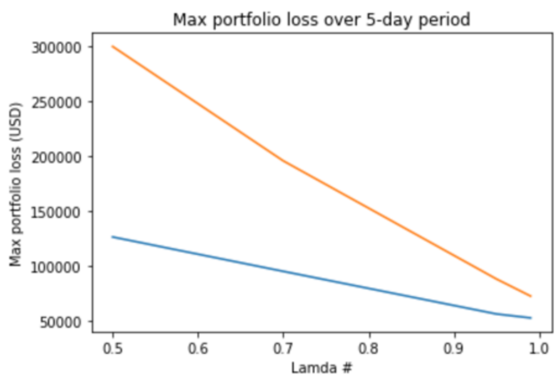


Figure 6 : value at risk and expected shortfall for different smoothening parameter lamda

The below graph shows value at risk and expected shortfall for above portfolio for different time periods

```
1 day VaR and expected shortfall@ 0.1% confidence: 55507.33  85520.2
2 day VaR and expected shortfall@ 0.1% confidence: 59675.14  86628.99
3 day VaR and expected shortfall@ 0.1% confidence: 51550.37  80024.77
4 day VaR and expected shortfall@ 0.1% confidence: 51327.11  74938.74
5 day VaR and expected shortfall@ 0.1% confidence: 52522.81  81116.05
```

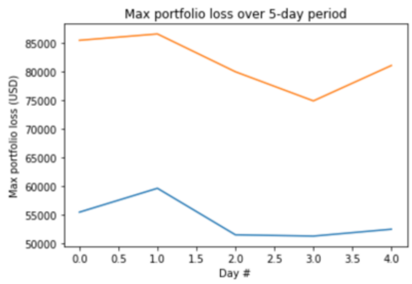


Figure 7 : value at risk and expected shortfall for different time periods using EWMA

7.3.6 Var using GARCH approach

To check for serial correlation we can use The Ljung-Box test , the module I have used also calculates Box-Pierce test which is similar to Ljung-Box . Both the models statistics are compared to against a chi-squared distribution To determine if the series is white noise. We might use the Ljung-Box test on the residuals of our model to look for autocorrelation, ideally our residuals would be white noise. The null hypothesis is data are independently distributed, no autocorrelation and alternate hypothesis is The data are not independently distributed; they exhibit serial correlation. if p value is less than alpha of (0.05) we reject the null hypothesis.

I have used a lag of 40 , and used both model to check for serial corelation , the test statistics , the p values were high for some lags ,we reject the null hypothesis and assume there is serial corelation in our data .

For garch volatility I have used garch(1,1) model om SP500 stock index returns from Jan 2019 to Dec 2021 the parameter of omega was 1.7296e-06 the index is not that volatile . Alpha of 0.1000 Alpha measures the extent to which a volatility shock today feeds through into next period's volatility. beta of 0.8800 According the Market Risk Analysis the usual range for alpha in a stable market is $0.05<\alpha<0.1$ and the parameter beta is $0.85<\beta<0.98$ the above model confirms it .

Then used rolling forecast to calculate updated garch variance and then to calculate the Var at desired confidence interval .

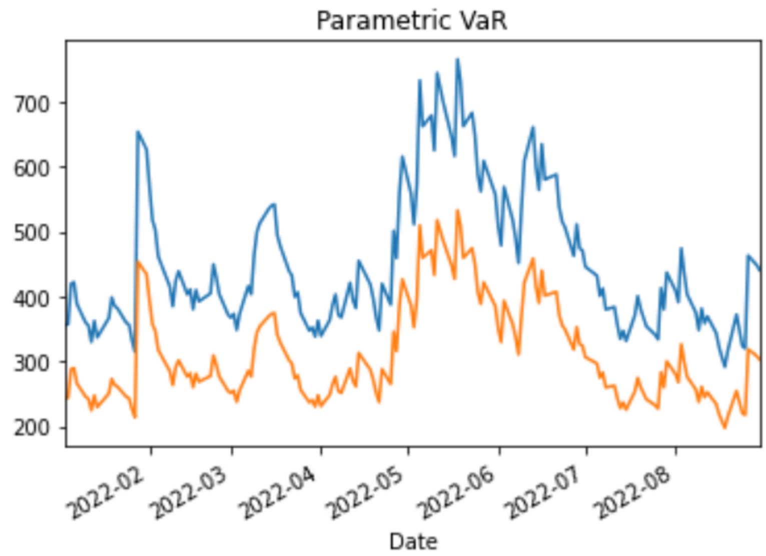


Figure 8: Value at risk forecast using garch model

The yellow line is the var for alpha = 0.05 and The Blue line is the var for alpha = 0.05

Then I performed the back testing for both the confidence level the below table is the test statis for unconditional coverage or Bernoulli test , (the 1 % critical values is 6.63, and for 5% critical value is 3.84) Both independence and unconditional coverage tests can be accepted at 1% but the 5 % fails the test . the conditional test which is sum of independence and conditional coverage fails the test for both the significance level

	Statistic	p-value
Unconditional	5.945424	0.014756
Independence	4.180858	0.040883
Conditional	10.126282	0.006326

Table 6 -back test statistics for var confidence level of 0.05

7.3.7 Var on options

The code I have implemented Var value at Risk for option is as follows To create a object of stock option , the parameter required are strike price K , risk free rate r , time to maturity of the option T , no of steps in binominal model N , Number of options, start date , end date ,Boolean variable is call if the derivative is call option and apha the confidence level of the Var. To set type of volatility we can call the method mean sigma( )or garch sigma() . Initial price of the option has been calculated with the current stock price.

To calculate Var I have used monte carlo simulation to stimulate the prices of the underlying stock , then recalculated the option price based on both binomial and Black sholes option pricing .To calculate profit and loss distribution the difference between recalculated price and initial option price is performed then profit and losses are sorted and var is calculated for the confidence interval .

The table shows value at risk and expected shortfall for above apple stock for different strike price

Strike Price	Value at Risk	Expected shortfall
140	$ -17.24	$ -17.24
165	$ -25.80	$ -25.0
170	$ -34.7	$ -34.5
200	$ -37.89	$ -37.65

Table 7 option var for different strike price

8 Code structure

The main file is the Jupiter file , portfolio.py contain the class and methods to calculate var for portfolio of assets . single.py contains the class and methods to calculate var for single assets . ew.py contains class and methods that have been implemented to calculate EWMA variance . price and var of option . Backtest.py contain all the class and methods to calculate the backtesing statistics. The main Jupiter file has the above calculated values and analysis.

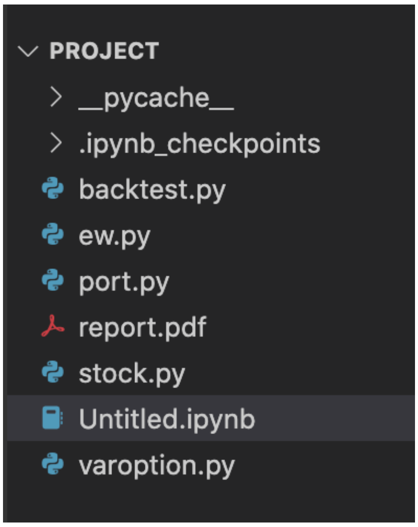


Figure 8: code structure of the project

9 Future directions, Conclusions

If it is assumed that financial returns, follow a normal distribution the mean returns and the variance-covariance matrix are sufficient to describe the full distribution of financial returns. There is strong evidence to suggest, however, that financial returns are not normally distributed. All the normality statistical test done using Shapiro-Wilk Test for each asset failed the normality test.

Stock returns usually have volatility clustering , ARCH model and GARCH model the conditional modelling ,many models like VCC GARCH (varying conditional corelation) Baba, Engle GARCH (BEEK) model the multivariate conditional covariance volatility and increasing availability of intraday data has led to the development of so-called High-frequency-based Volatility like DCC-HEAVY or DECO-HEAVY [11] this is challenging with many parameters to estimate for too many assets and with non-linear derivatives .

Having mentioned the limitation of Value at risk section 4.3.1 , it can be still useful too , it must be complemented with other tools especially those taking care of the 1 % worst case area which VAR ignores .

10 References

[1] Jorion, P. (2007). Value at Risk, McGraw-Hill, 3rd Edition.
[2] Aswath Damodaran paper of Value at Risk URL https://pages.stern.nyu.edu/~adamodar/pdfiles/papers/VAR.pdf
[3] Shay Palachy Stationarity in time series analysis url:https://towardsdatascience.com/stationarity-in-time-series-analysis-90c94f27322
[4] Volatility with ARCH and GARCH for Time Series Forecasting by Jason Brownlee url -https://machinelearningmastery.com/develop-arch-and-garch-models-for-time-series-forecasting-in-python/
[5]Black- Scholes Model url-
https://www.investopedia.com/terms/b/blackscholes.asp
[6] Binomial Option Pricing Model Lse url https://www.le.ac.uk/users/dsgp1/COURSES/DERIVATE/BINOPTION.PDF
[7] Yitong Ren The Significance and Applications of Covariance Matrix https://towardsdatascience.com/the-significance-and-applications-of-covariance-matrix-d021c17bce82
[8] A review of backtesting for value at risk Authors:Y. ZhangS. Nadarajah 2017 Journal Communications in Statistics - Theory and Methods
[9] Christoffersen, P., (2012). Elements of Financial Risk Management, Academic Press, 2nd edition.
[10] RDP 1999-04: Value at Risk: On the Stability and Forecasting of the Variance-Covariance Matrix James Engel and Marianne Gizycki
[11] DCC- and DECO-HEAVY: Multivariate GARCH models based on realized variances and correlations BY LucBauwensaAND YongdengXub