**Objective:**

Main Objective of this paper is to analyze different theories in improvement confining to VAR Process. We collated the knowledge based on different approaches and methodologies for Financial Risk Management using Monte Carlo Simulation on Data science model to calculate Value at Risk. This could help us to visualize, optimize and interpolate Risk in Financial Domain.

**Method:**

We are using Variance-Covariance (Parametric) method to calculate Var with help of Python.

**Result**: Analysis results based on the findings of the optimal combinations and proportion of stocks. Using historical data it can generate different combinations of portfolios in different ratios, to see how it would perform relative to each other during that time period.

**Benefits**:

If this method is implemented for VAR Calculation then it would be most powerful and fast scientific computing method which will reduce manual interpretation and also provides the more accurate results as compared to other VAR computation methodologies.

**Introduction:**

Data science is a technology which provides insight view into the understanding from unstructured and structured data based on the concept to unify statistics, analyzing data, machine learning and related techniques. Data Science being a very vast field, we are using Data Analytics (one very important component of Data Analytics) for our experiments.

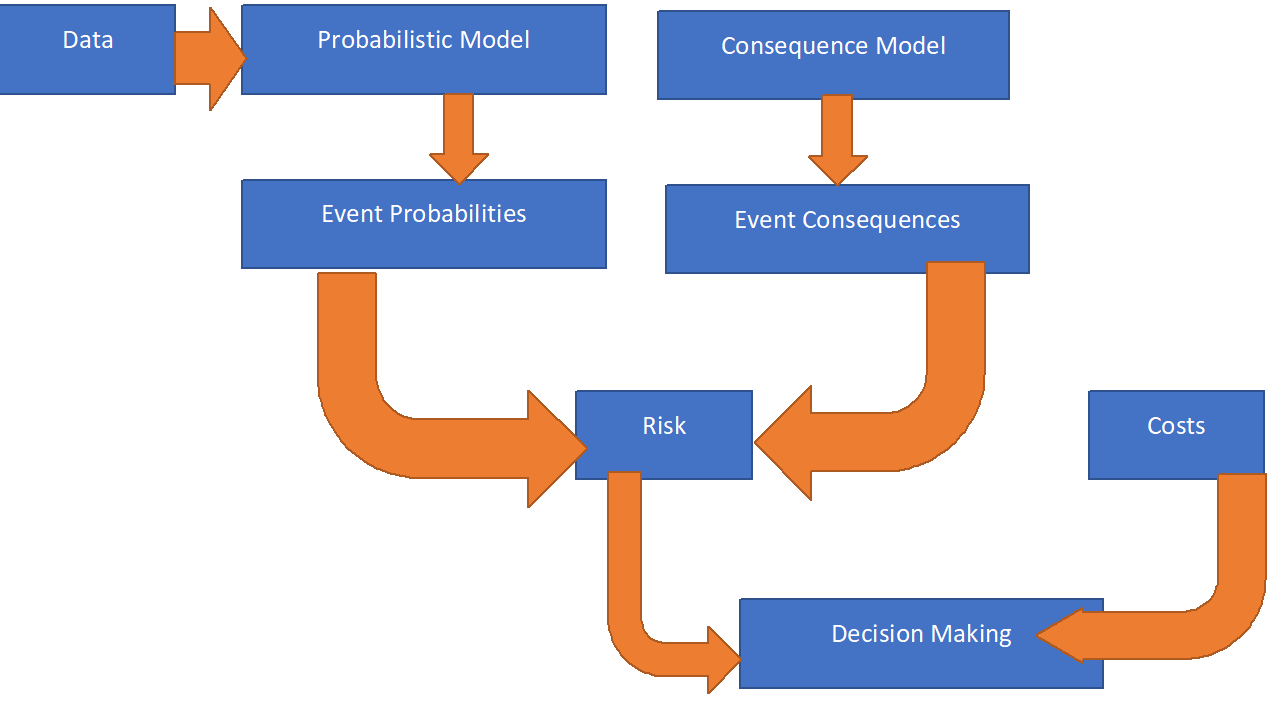
Risk in Finance can be classified as Market/Credit/Operational/Underwriting Risks which may lead to Profit or Loss. Risk Management uses indicator like VAR (Value-At-Risk), part of FRM (Financial Risk Management) estimates probability of portfolio losses using statistical analysis of historical price data and volatilities.

Financial firms like SG re-organizes actual historical returns, sorting them worst to best assuming that history repeats itself from a risk perspective.VAR is calculated daily using Parametric/Historical/Monte Carlo VAR.

Our analysis looks into the above from Data Analytics prospective.

**Study Methodology:**

VAR calculation is based on three parameters: Time Horizon or (Periods), Amount loss and Confidence Level (we used 99% Confidence Level which indicates we are 99% confident that over the given period, portfolio will not loss more than VAR calculated amount).



Diagrammatic representation of Risk based decision making

**Steps in Data Science:**

* Fetching the data

We have extracted the stock data of 5 portfolios (All portfolios are based on financial institutions) from BSEindia .

* Reading and Storing the data

Used python-pandas library to read and store the data

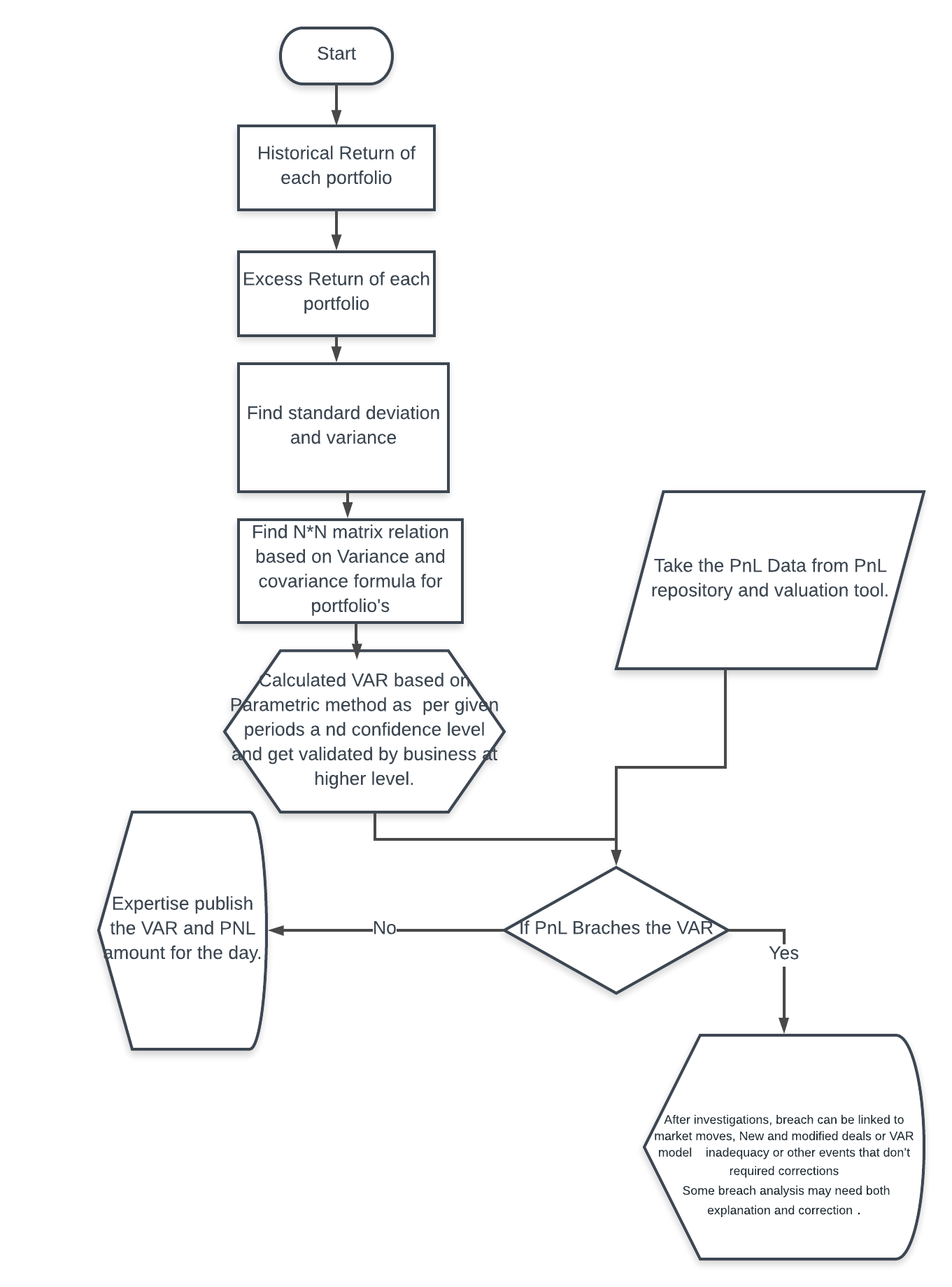
* Analyzing the data

The data analysis was done using multiple libraries supporting our use case. For example, python-scipy, python-nympy for performing different statistical analysis and seaborn, vpython, display for visualization.

* Generating inferences out of the data.

After analyzing the data. We can see below observation from out of the extracted data of stocks (HDFC,ICICI, Yes Bank,SBIN,PNB and Hudco) 246 scenarios from October 2017 to September 2018:

1. Standard Deviation value for each of the stocks which is a measure of Risk used by individual investor and financial institutions also it is a measure of volatility.
2. Correlation and Covariance Metrics: Correlation in financial domain is a statistical measurement between two or more securities which move in relation to each other securities/portfolio and Covariance is a statistical measurement of how two or more securities move in relation to each other. This is the important factor to mitigate the risk.
3. Display the Visualized and tabular data for variance-covariance and correlation of stocks in graphical form, which shows the correlation between each stock and formed a normal standard distribution for each stock which is a powerful metrics to calculate VAR.
4. Calculated VAR at 99% confidence level (1% significance level ) using parametric method with the help of inverse of the cumulative standardized normal distribution of 99% with respective to standard deviation.

(Workflow diagram on VAR which shown the steps of VAR and use of VAR in backtesting)

**import** **numpy** **as** **np**

**import** **pandas** **as** **pd**

**from** **IPython.display** **import** display

**import** **seaborn** **as** **sns**; sns.set()

#Scientific and mathematical calculation (SCIPY)

**import** **scipy**

**from** **vpython** **import** \*

**import** **matplotlib.pyplot** **as** **plt**

csvFile = pd.read\_csv('C:**\\**Users\Dipty Saurabh\Desktop\SG-Technical Writing\pyt.csv')

hdfc = csvFile['HDFC']

icici = csvFile['ICICI']

yesBank = csvFile['Yes Bank']

sbin = csvFile['SBIN']

pnb = csvFile['PNB']

hudco = csvFile['Hudco']

hdfc\_std = np.std(hdfc)

icici\_std = np.std(icici)

yesBank\_std = np.std(yesBank)

sbin\_std = np.std(sbin)

pnb\_std = np.std(pnb)

hudco\_std = np.std(hudco)

*#std, 25th percentile, 50th percentile etc*

display(csvFile.describe())

*#calculating correlation*

corr = csvFile.corr().as\_matrix()

mask = np.zeros\_like(corr)

mask[np.triu\_indices\_from(mask, 1)] = **True**

NORMINV = scipy.stats.norm.ppf(0.01)

VAR\_HD = hdfc\_std \* NORMINV

ICICI\_HD = icici\_std \* NORMINV

YesBank\_HD = yesBank\_std \* NORMINV

sbin\_HD = sbin\_std \* NORMINV

pnb\_HD = pnb\_std \* NORMINV

hudco\_HD = hudco\_std \* NORMINV

print ("VAR\_HDFC",VAR\_HD , "**\n**VAR\_ICICI",ICICI\_HD , "**\n**VAR\_YesBank", YesBank\_HD, "**\n**VAR\_SBIN", sbin\_HD,

"**\n**VAR\_PNB", pnb\_HD, "**\n**hudco", hudco\_HD)

*#norm.ppf(0.99, loc=0, scale = 1)*

*#VAR = np.mean[csvFile.describe()]*

*#displaying correlation matrix*

**with** sns.axes\_style("white"):

ax = sns.heatmap(corr, mask=mask, annot=**True**, fmt='+.3f')

axes = pd.plotting.scatter\_matrix(csvFile, alpha = 0.3, figsize = (14,8), diagonal = 'kde')

**for** i, j **in** zip(\*np.triu\_indices\_from(axes, k=1)):

axes[i, j].annotate("**%.3f**" %**corr**[i,j], (0.8, 0.8), xycoords='axes fraction', ha='center', va='center')

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | HDFC | ICICI | Yes Bank | SBIN | PNB | Hudco |
| count | 246 | 246 | 246 | 246 | 246 | 246 |
| mean | 0.000420 | 0.000370 | -0.002633 | 0.000216 | -0.003151 | -0.002372 |
| std | 0.009514 | 0.019888 | 0.029813 | 0.024386 | 0.039251 | 0.017357 |
| min | -0.038615 | -0.061096 | -0.338452 | -0.050894 | -0.129553 | -0.060815 |
| 25% | -0.005086 | -0.010986 | -0.011423 | -0.013852 | -0.019578 | -0.013277 |
| 50% | 0.000713 | -0.001400 | -0.000744 | -0.001078 | -0.003890 | -0.003323 |
| 75% | 0.006047 | 0.010953 | 0.009631 | 0.010915 | 0.012915 | 0.007524 |
| max | 0.042703 | 0.137090 | 0.079333 | 0.243601 | 0.379794 | 0.069412 |

|  |
| --- |
| VAR\_HDFC **-0.022088341543971355** |
| VAR\_ICICI **-0.04617197188097052** |
| VAR\_YesBank **-0.0692151269210863** |
| VAR\_SBIN **-0.05661534891767363** |
| VAR\_PNB **-0.09112484432453803** |
| hudco **-0.040295748021850195** |

**Correlation and Covariance Matrics**

It is a matrix which indicates ith row of portfolio and jth column of portfolio has correlation ( ρij) between variables i and j. If it is perfectly correlated, the diagonal of correlation matrix is 1. This type of correlation matrix is symmetric.

The Covariance between variables i and j ,

|  |
| --- |
| **COVij = σi σj ρij** |

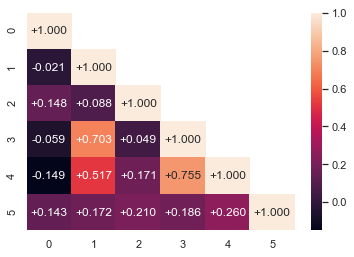
**Table of Variance-Covariance Matrix**

var1 cov12 cov13 … cov1n

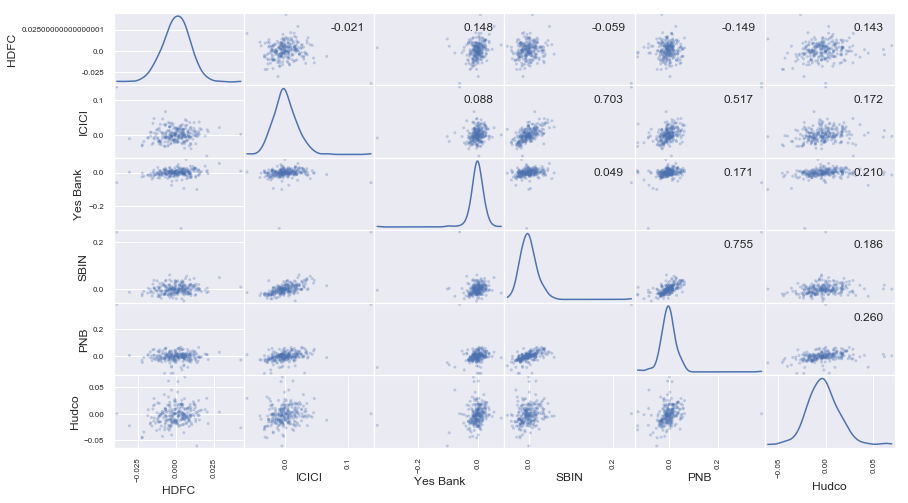
cov21 var2 cov23 … cov2n

cov31 var32 var3 … cov3n

Below results are through Python simulation which indicates the Variance-Covariance relation between all the portfolios.



Variance-Covariance Heatmap between the stocks chosen. Index: 1. HDFC, 2. ICICI, 3. Yes Bank, 4. SBIN, 5. PNB, 6. Hudco



**Comparison between commonly used VAR methods**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Parametric / Variance-Covariance | Historical Simulation | Monte Carlo |
| Feature | Delta-Normal (Parametric) | Compute simulated level  using historical returns(Ranking) | Simulation |
| Valuation | Linear | Full | Full |
| Distribution | | | |
| Shape | Normal Standard Distribution | Actual (Histogram) | General |
| Implementation | | | |
| Easy of Compute | Yes | Intermediate | No |
| Communicability | Easy | Easy | Difficult |
| VAR Precision | Excellent | Poor with short window | Good with many iteration |
| Major Pitfalls | Non linearity | Time variation in risk (unusual event) | Model Risk |

**Benefits:**

* Parametric simulation is a computerized mathematical technique that allows users to account for variability in their process to enhance quantitative analysis and decision making
* Data analytics is rich scientific commutating libraries which give statistical measures, visualization presentation of data.
* It is a powerful, fast scientific computing method which will reduce manual interpretation and also provides the optimized, interpolated curve fitting output.

**Summary:**

* Financial Risk is a major metric in Investment Baking where VAR is one of the important indicator to mitigate the Risk.
* Our study based calculating VAR on 247 Historical Scenarios (consists of 5 different bank portfolios) using Parametric simulation.
* Parametric is more productive and speedier as we can easily load Historical data using python library data.
* We used statistical Model to calculate VAR and plotted graphs which give clear visualization of result.
* Also we analyzed python provides more extensive support libraries and open source.
* Data analytics gives a better chance of mitigating with the risk than traditional method (eg: Histogram using excel).

**Easy of Application in SG Context:**

Based on our current Project under FRM VAR calculation is based on Historical method. Whereas we found parametric will be more productive, user Friendly involving less manual intervention in system.

As per our study, parametric simulation is more statically aligned, as deviation is calculated at every juncture, unlike historical where only an aggregated picture of portfolios are considered.