AUGUST 1, 2019

RISK ANALYSIS OF IBM, GE AND P&G

VaR AND ES CALCULATIONS AND PORTFOLIO CREATION

DEEPTI GUPTA

ANLY515-2019/SUMMER: RISK MODELING AND ASSESSMENT

Harrisburg University of Science and Technology

Objective

The objective of this project is to improve the profitability of the trading models and use statistical techniques to identify consistent behavior in assets which can be exploited to turn a profit. To find this behavior, it must be explored that how the properties of the asset prices themselves change in time.

Introduction

As quants, we do not rely on "guesswork" or "hunches". The approach is to quantify as much as possible, both to remove any emotional involvement from the trading process and to ensure (to the extent possible) repeatability of the trading. The phenomenon that the returns of the financial market tends to exhibit is volatility clustering. There can be periods with high volatility or tranquility which cannot be predicted unless correlated with time for which we can use conditional heteroscedastic models.

ARMA (Auto-Regressive Moving Average) models are used to model the conditional expectation of a process given the past, but in an ARMA model the conditional variance given the past is constant. An ARMA model cannot capture a high volatility type of behavior because its conditional variance is constant. To cater such needs, we need better time series models to model the nonconstant volatility which is most of times exhibited by real time financial stocks data. ARCH is an acronym meaning Auto-Regressive Conditional Heteroscedasticity. GARCH stands for Generalized ARCH, the models that model conditional variances much as the conditional expectation is modeled by an ARMA model.

In this project, we will use three methods to calculate risks for the three specified assets. Method 1 will calculate VaR and ES using (Generalized Hyperbolic Distribution) GHYP distribution, Method 2 will use ARMA and GARCH method and Method 2 will determine risk using copula-GARCH method, all of these using the weights produced by Global Minimum Variance Portfolio (GMVP)

Loading the Data

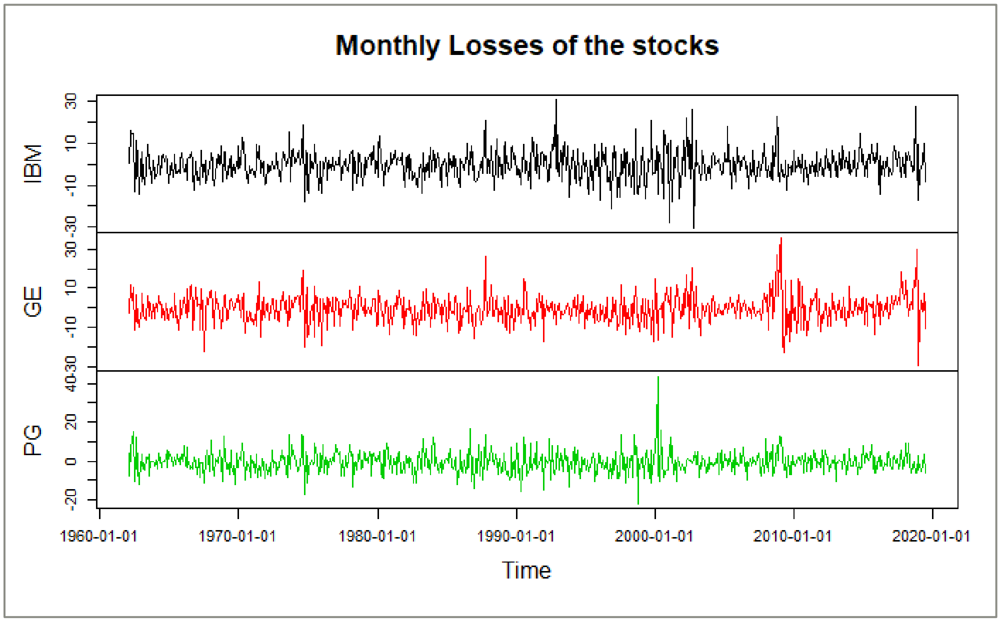
Data was loaded from yahoo.finance.com for three major market players in different sectors. The assets are IBM from technology sector, General Electric from industrial sector and Proctor and Gamble from consumer's market sector. These assets have many things in common – very long presence in consumer market as well as stock market, global presence and high brand value. General Electric (GE) Company operates as a high-tech industrial company worldwide. It operates in Power, Renewable Energy, Aviation, Oil & Gas, Healthcare, Transportation, Lighting, and Capital segments. IBM Corporation operates as an integrated technology and services company worldwide. Last but not the least, The Procter & Gamble (P&G) Company provides branded consumer packaged goods to consumers worldwide.

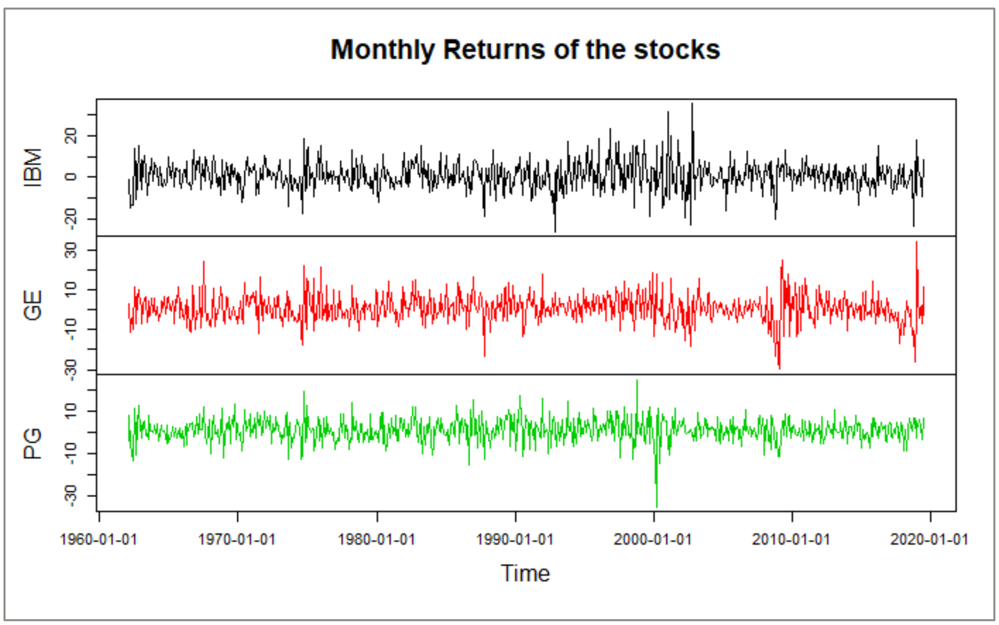
Data has been loaded for the three assets for a common period of Jan 1962 till Jul 2019 with a monthly frequency. We would be considering the closing prices of IBM, GE and PG to perform the risk analysis.

Data Preprocessing and Analysis

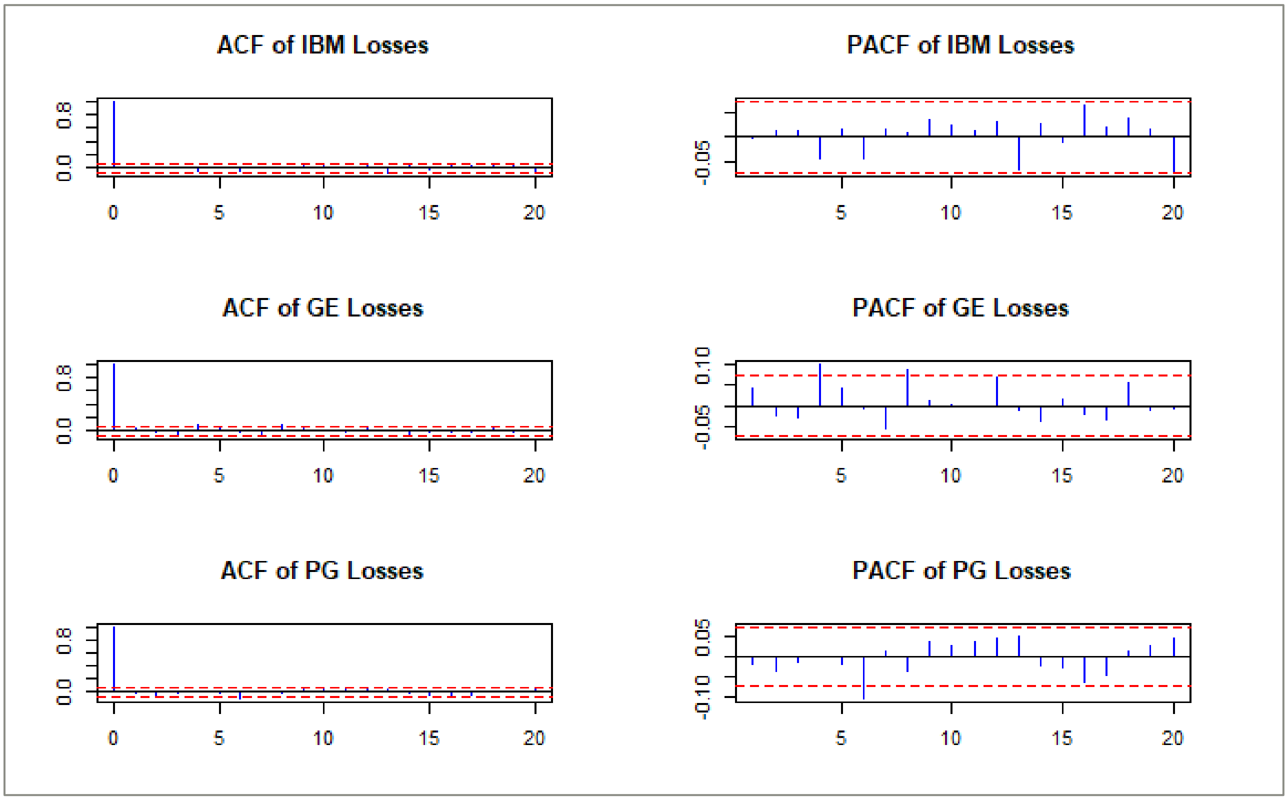
Modeling the data with NAs can severely impact the results, considering which it is important to remove all the NA values from it. The three assets contain data for the identical time span. Data preprocessing includes calculating returns, timeseries and losses for the three assets individually as well as connected in a single data frame which will be used at several part of the project and date attribute is added in these objects wherever required.

Below figures are the plots on the monthly returns and monthly losses of the three assets over the span of years using the rate of log return changes. We observe that while the mean is approximately stationary, the variance is not stationery and volatility is varying with time too.





Next step is to analyze ACF and PACF. ACF (stands for Auto-correlation Function) is used to identify the order of a MA model while PACF (meaning Partial Auto-correlation Function) at the h_{th} lag is interpreted as the correlation between x_t and x_{t-h} where the linear dependency of the intervening lags (x_{t-1}, x_{t-2}, ... , x_{t-h+1}) has been removed, is used to estimate the AR part. Looking at the auto-correlation graphs, we do not see any significant and easily interpretable peaks at initial lags.



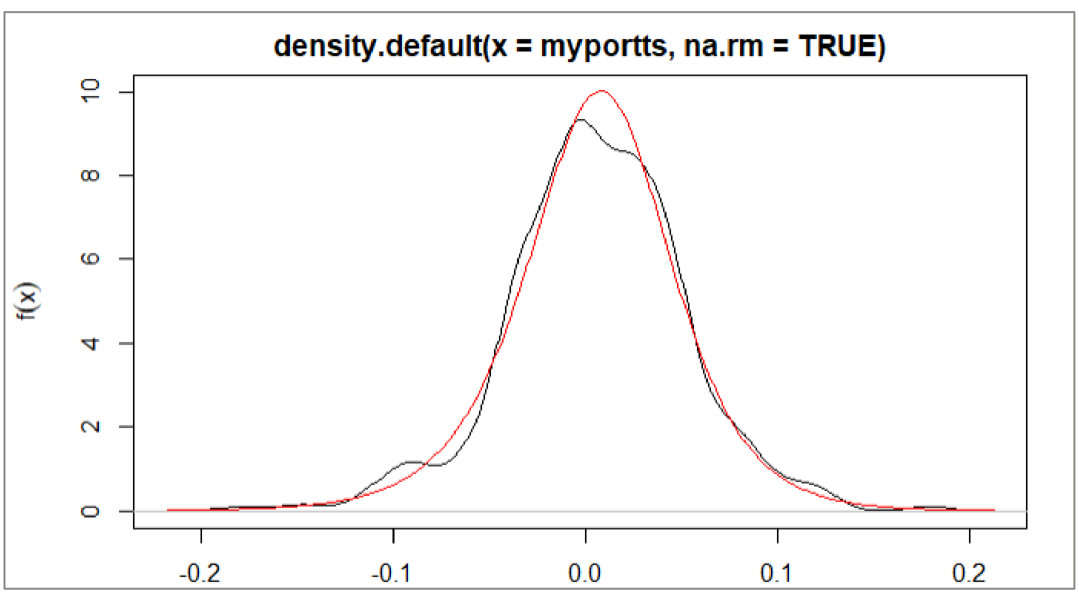
Method I – Using GHYP Distribution

There are various kinds of distributions that describe the asset returns better than the normal distribution, and additionally we know that financial data is leptokurtic (meaning that the occurrence of extreme events is more likely compared to the normal distribution) in nature and would not follow a normal distribution in real world. These distributions are capable of mirroring not only heavy-tailed behavior but also asymmetries. Assuming the time series data to be univariate, we can assume that returns are not independent and identically distributed (iid), time variant volatile and auto-correlated.

Using stepAIC function from package ‘ghyp’, we performed a model selection in the scope of the generalized hyperbolic distribution class based on the Akaike information criterion to see which distribution has the closest fit to the actual distribution of returns. The goal is that the value of AIC should be minimum which is true for ghyp (generalized hyperbolic distribution) model for symmetric = FALSE to decide on which distribution will be able to explain the returns in a best way.

	model <chr>	symmetric <lgl>	lambda <dbl>	alpha.bar <dbl>	mu <dbl>	sigma <dbl>	gamma <dbl>	aic <dbl>	llh <dbl>
8	NIG	TRUE	-0.5000000	1.886457	0.006142197	0.04658026	0.000000000	-2297.258	1151.629
7	hyp	TRUE	1.0000000	1.741664	0.006093979	0.04644043	0.000000000	-2296.784	1151.392
10	t	TRUE	-3.1107187	0.000000	0.006187733	0.04689068	0.000000000	-2296.663	1151.331
3	NIG	FALSE	-0.5000000	1.948929	0.010716365	0.04641494	-0.005194661	-2296.277	1152.139
2	hyp	FALSE	1.0000000	1.805389	0.010926522	0.04628960	-0.005408728	-2295.757	1151.879
5	t	FALSE	-3.1956265	0.000000	0.010785018	0.04666309	-0.005280724	-2295.749	1151.875
9	VG	TRUE	2.6387985	0.000000	0.006003059	0.04634072	0.000000000	-2295.485	1150.742
6	ghyp	TRUE	-1.1499492	1.788682	0.006162806	0.04663070	0.000000000	-2295.302	1151.651
4	VG	FALSE	2.7290505	0.000000	0.011086404	0.04618652	-0.005560776	-2294.321	1151.160
1	ghyp	FALSE	-0.7896702	2.026443	0.010215791	0.04628871	-0.004655565	-2294.268	1152.134

We fitted the time series returns in generalized hyperbolic distribution and calculated the density of it. Plotting the ghyp density (red) over the empirical distribution (black) gives us the following graph which seems to be a close fit.



VaR at 95% confidence interval can be interpreted as 1 out of 1000 days (every 3 years), a loss of not greater than 7.1% should be expected. ES at 95% confidence interval can be interpreted as 1 out of 1000 days (every 3 years), a shortfall of not greater than 10.12% should be expected.

```
portvar <- abs(qghyp(p, ghypfit)) * 100
portvar           # VaR values in the vector form of quantile

# 99%    98%    97%    96%    95%
# 11.9   9.8    8.6    7.7    7.1

portes <- abs(ESghyp(p, ghypfit)) * 100
portes            # ES values in the vector form of quantile

# 99%    98%    97%    96%    95%
# 15.04  12.9   11.66  10.7   10.12
```

Method II – Using Arima and Garch Approach

Arch and Garch class of models allow you to model volatility. Because volatility seems to be a function of time, we can conclude that it is heteroskedastic in nature. We estimated ARIMA model and calculated residuals and variance of residuals follows a specific linear equation, one could use the equation to forecast future volatility which is included as univariate data.

We started with estimating ARMA+GARCH model and created 1 step ahead forecast of the standard deviation followed by estimating distribution coefficients for the return series. We simulated 100,000 random variables (returns) that follow the identified distribution with estimated parameters and multiplied each simulated return by forecasted SD. Then sorted the product obtained from smallest to largest and calculated VaR as the 5,000th largest loss and ES as the median of 5,000 largest losses.

Method III – Using Garch-Copula Approach

The multivariate stylized fact for financial assets deduces that correlation varies with time and to use variance-covariance matrix, we assume that all assets in the portfolio are normally and identically independently distributed (iid) which is untrue in a real time scenario. It is important to know that underestimating or overestimating the portfolio will result in inefficient (not maximum) calculation of expected returns and not minimizing the variance. Thus, it is important to verify if the assets jointly share marginal distribution.

In order to consider both the univariate stylized fact such as volatility clustering and multivariate stylized facts such as time varying correlation into a model, we will combine Garch model (which addresses univariate stylized facts) with copula (which addresses multivariate stylized facts) with the portfolio weights in our next method. We will also analyze if our returns are elliptically distributed (which means that all assets in the portfolio are normally distributed then their joint distribution represents an ellipse) because variance-covariance or correlation matrix is only appropriate for elliptically distributed random variables.

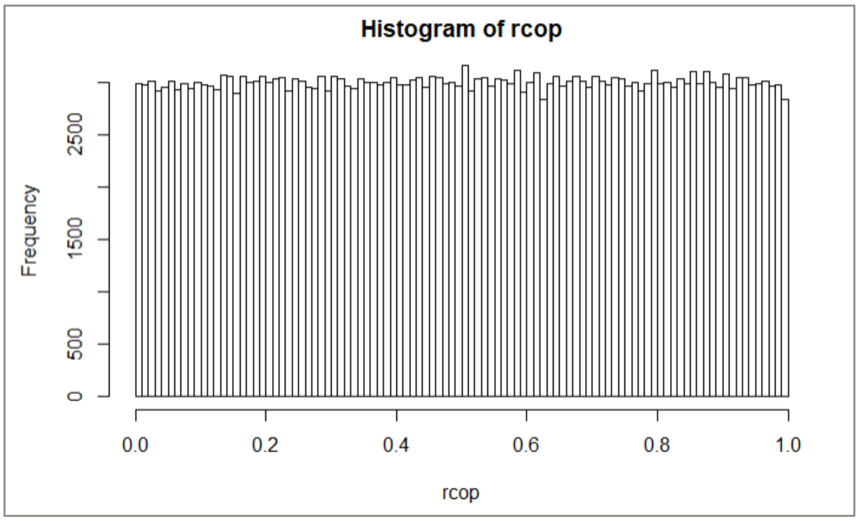
We applied garchfit to the three assets with formula=~arma(0,0)+garch(1,1) and student t-distribution followed by estimating the degrees of freedom parameter. Next, we calculated the residuals to apply Garch-copula model and standardized it by dividing it with conditional std. deviation. Then, the probabilities were calculated from pseudo-uniform variables for each risk from the standardized residuals. A copula model - Student's t copula is estimated based on Kendall's rank correlations.

Below is the Kendall rank correlation for different assets. Correlation between IBM and GE is 0.47, IBM and PG is 0.28 and, GE and PG is 0.40. The joint marginal distribution of the three assets is value of nu i.e. 0.744

```
$P
          [,1]      [,2]      [,3]
[1,] 1.0000000 0.4698119 0.2777477
[2,] 0.4698119 1.0000000 0.4029480
[3,] 0.2777477 0.4029480 1.0000000

$nu
[1] 0.7447845
```

A 100,000 random losses are then simulated for each financial instrument which follow student t-distribution and Kendall's rank correlation which can be seen in below histogram.



These pseudo uniform variables need to be converted back in the form of quantiles. The simulated portfolio losses are then determined as the outcome of the matrix-weight vector product by using weights from GMVP distribution. Sorted the product obtained from smallest to largest, VaR is the 5,000th largest loss and ES is the median of 5,000 largest losses.

Conclusion

	Method 1 Fitting GHYP Distribution		**Method 2 Garch–Arima Approach**		**Method 3 Garch-Copula Approach**	
	VaR	ES	VaR	ES	VaR	ES
Using weights from GMVP Portfolio of IBM, GE and PG	7.121	10.123	7.234	8.787	7.848	10.317

Though we can notice that VaR and ES is the least in Method 1 and Method 2 but Method 3, where we calculated VaR and ES using Garch-Copula approach gives us a better estimation because here we are considering both univariate and multivariate stylized facts, and not under or overestimating the returns.

References

Data from Proctor & Gamble, IBM and General Electrics –

PG.csv IBM.csv GE.csv

R Markdown File and PDF knitted File

RiskModelingProject_DeeptiGupta.Rmd RiskModelingProject_DeeptiGupta.pdf