STATISTICAL DATA ANALYSIS

Project Report

Exploratory data analysis

STOCK PORTFOLIO DATA ANALYSIS

Group Members :

- Cherukuri Nikhilesh - S20180010040
- Kore Nithish Kumar - S20180010086
- Pulla Nagendra Babu - S20180010138
- Rishab Tripati - S20180010147

Abstract:

The goal of this project is to do exploratory data analysis on a stock portfolio performance dataset.We apply Linear Regression on the dataset and test all the assumptions of linear regression.

Methodology:

1. Loading the dataset
2. Univariate analysis
3. Check for null values
4. Check for normality of data.
5. Explore data trends
6. Influential points detection and removal
7. Construct a correlation matrix
8. Principal Component Analysis
9. Factor Analysis
10. Split data into train and test sets and apply Multiple Linear regression.
11. These of assumptions
    a. Linearity
    b. Homoscedasticity
    c. Normality of errors
    d. Uncorrelated errors

Data Sample

	Large B/P	Large ROE	Large S/P	Large Return Rate in the last quarter	Large Market Value	Small systematic Risk	Annual Return.1	Excess Return.1	Systematic Risk.1	Total Risk.1	Abs. Win Rate.1	Rel. Win Rate.1
0	1.0	0.0	0.0	0.0	0.0	0.0	0.531875	0.478116	0.738015	0.800000	0.52	0.411765
1	0.0	1.0	0.0	0.0	0.0	0.0	0.549712	0.487595	0.571579	0.412231	0.52	0.764706
2	0.0	0.0	1.0	0.0	0.0	0.0	0.692625	0.629895	0.703051	0.756879	0.44	0.376471
3	0.0	0.0	0.0	1.0	0.0	0.0	0.324351	0.255634	0.800000	0.756046	0.36	0.270588
4	0.0	0.0	0.0	0.0	1.0	0.0	0.326615	0.306501	0.432452	0.209289	0.72	0.447059

Data Summary(Univariate Analysis)

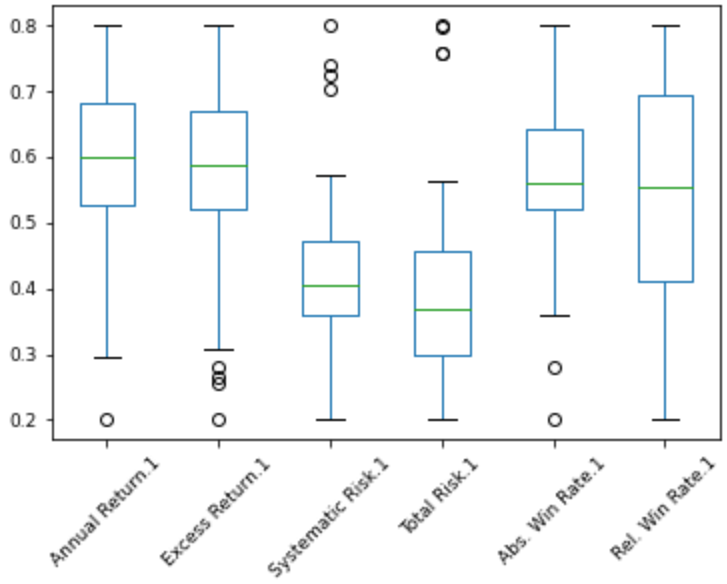
Data summary to get an idea of Mean, standard deviation, min, max etc., of the given data.

	Large B/P	Large ROE	Large S/P	Large Return Rate in the last quarter	Large Market Value	Small systematic Risk	Annual Return.1	Excess Return.1	Systematic Risk.1	Total Risk.1	Abs. Win Rate.1	Rel. Win Rate.1
count	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000	63.000000
mean	0.166619	0.166619	0.166619	0.166619	0.166619	0.166619	0.580151	0.576170	0.426494	0.391749	0.566984	0.547899
std	0.199304	0.199304	0.199304	0.199304	0.199304	0.199304	0.133358	0.137047	0.118178	0.136653	0.112803	0.159468
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.200000	0.200000	0.200000	0.200000	0.200000
25%	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.525811	0.519093	0.358600	0.297324	0.520000	0.411765
50%	0.167000	0.167000	0.167000	0.167000	0.167000	0.167000	0.598516	0.587148	0.403418	0.368958	0.560000	0.552941
75%	0.291500	0.291500	0.291500	0.291500	0.291500	0.291500	0.679636	0.669294	0.470571	0.457749	0.640000	0.694118
max	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	0.800000	0.800000	0.800000	0.800000	0.800000	0.800000

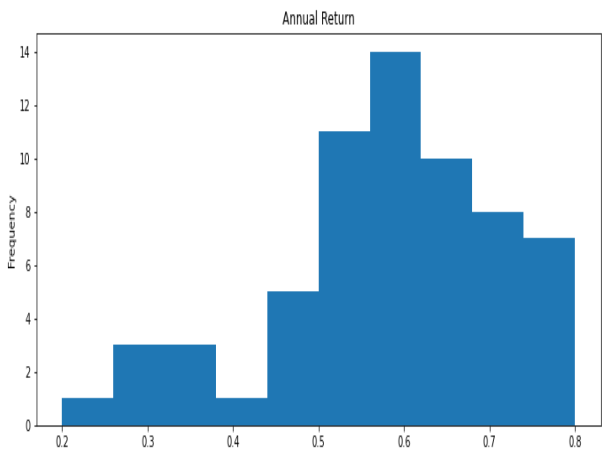
Null Values

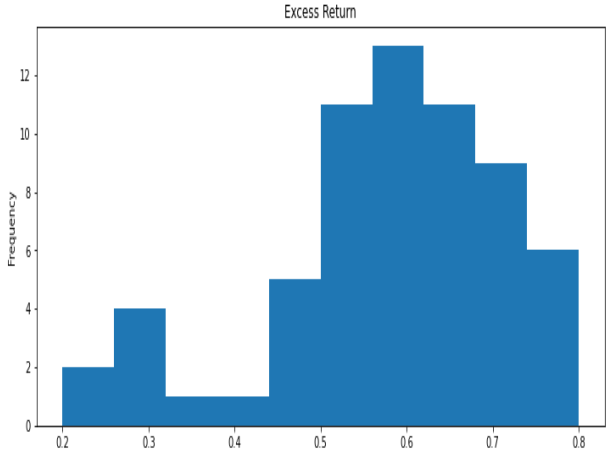
There are no Null values in the given data.

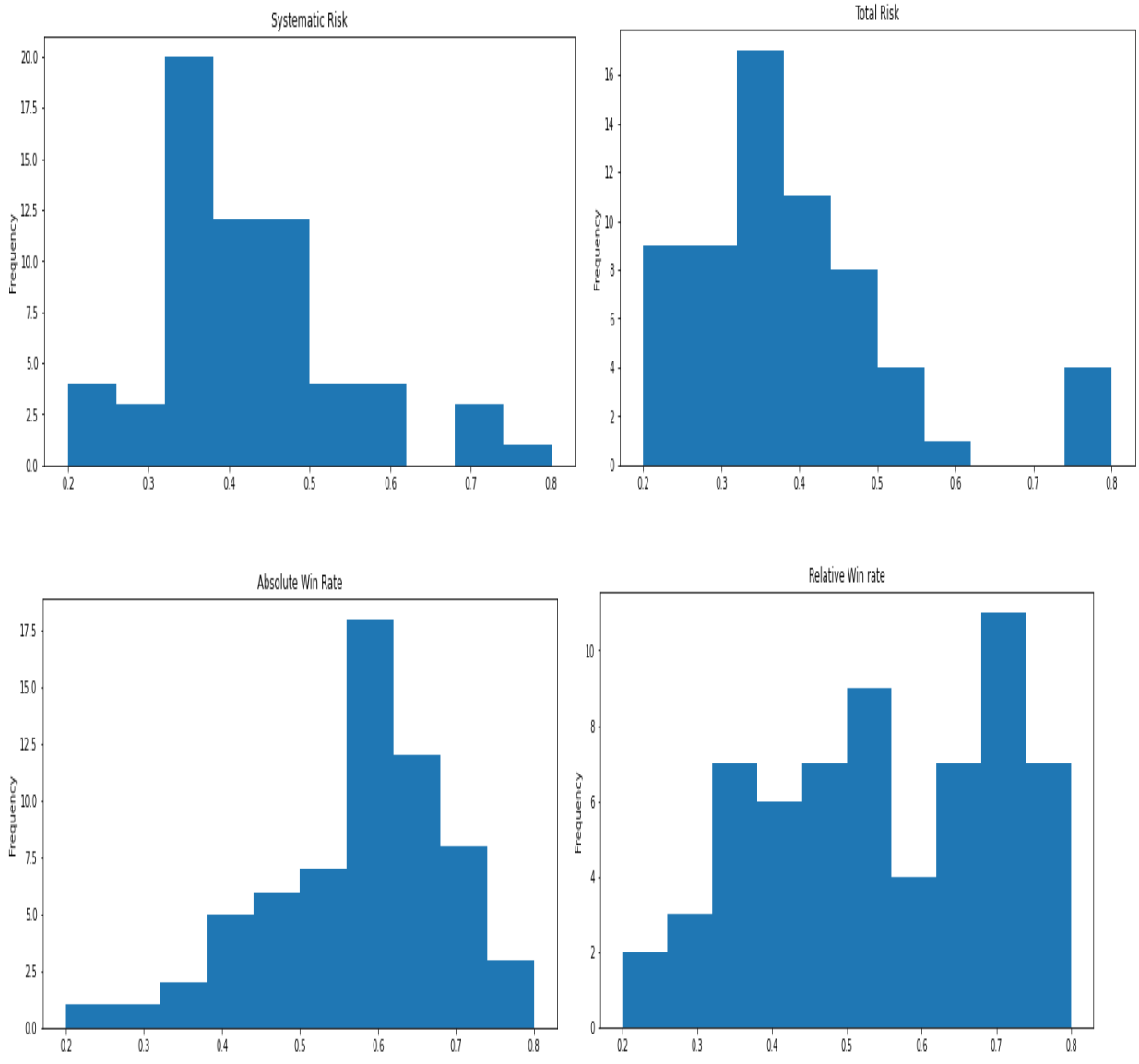
Data Trends:



The above boxplot shows the distribution and skewness of target variable quartiles in the data.





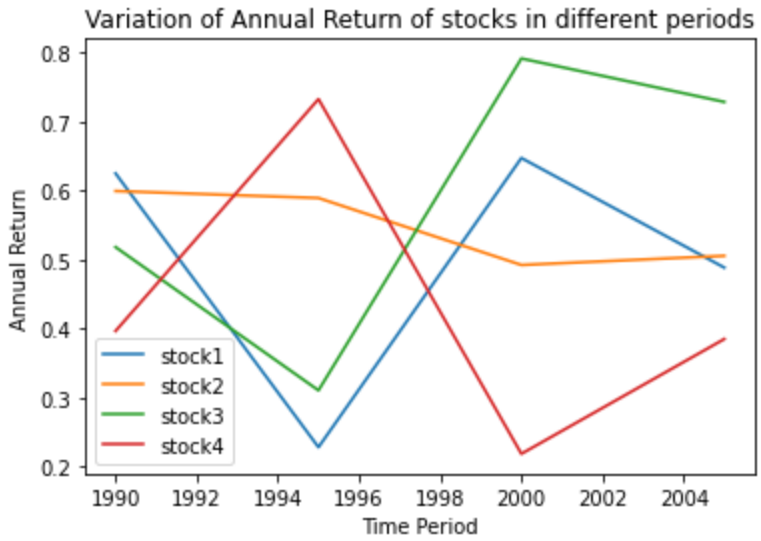


The above plots show the distribution of the target variables.We can see that all the targets are approximately normal.

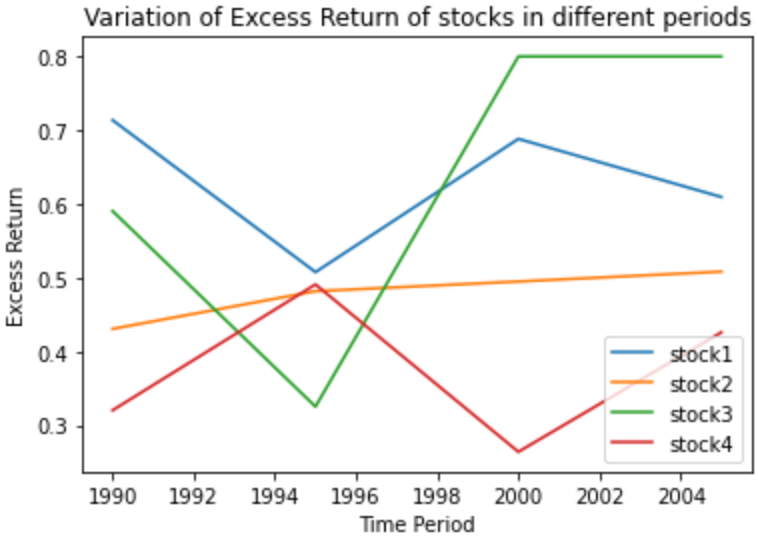
Surge of targets over time :

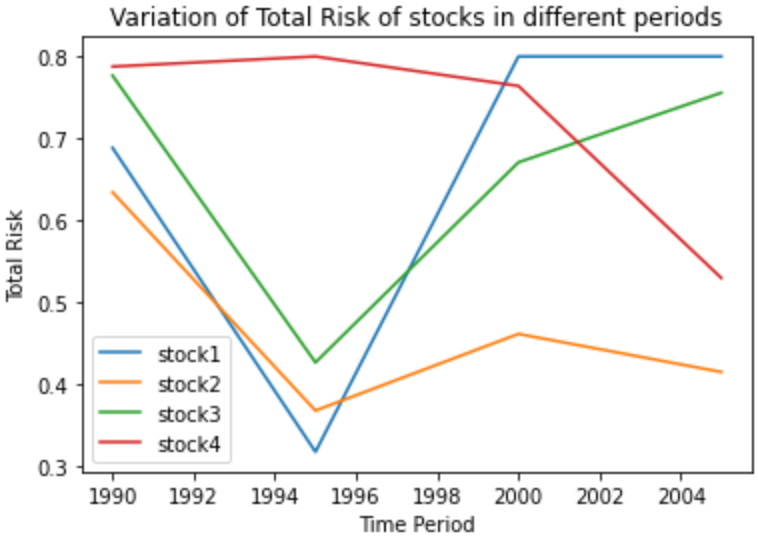
In the given dataset the corresponding rows are stocks.

Annual Income of few stocks in the given time periods are plotted below.

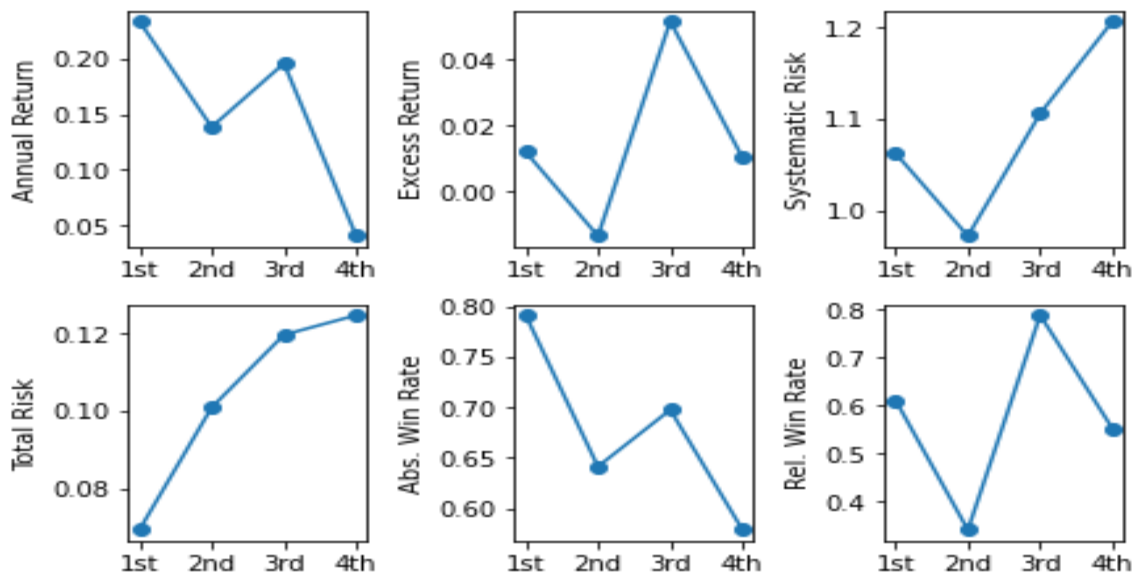


If we plot for all the stocks Maximum annual income (i.e 0.8) among the stocks is constant in every time period.





Change in average of targets over each period:

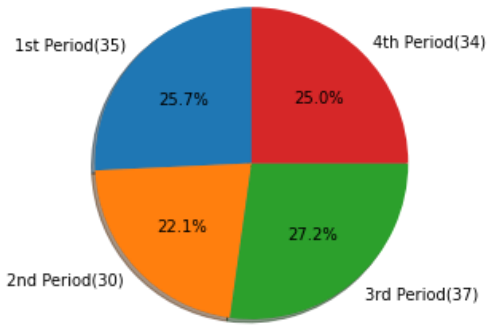


- The above plots shows how the average of each target variable changes over each period
- From the above plots we can observe that the average of Total risk is increasing for every period.

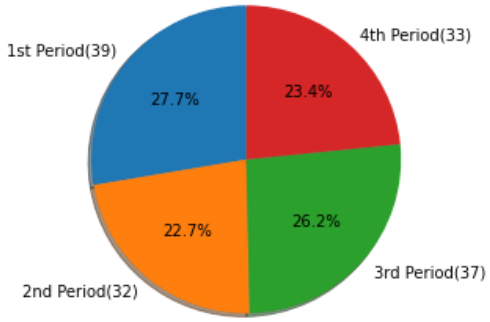
Performance Of Stocks:

- The below plots are a number of stocks which are having above average performance with respect to features in different time periods.

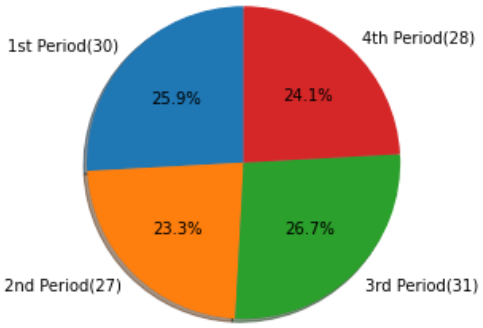
Stocks which are more than Average of Annual Return



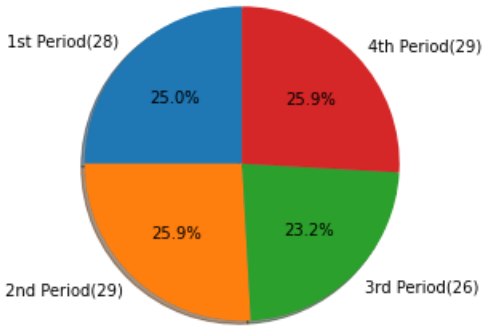
Stocks which are more than Average of Excess Return



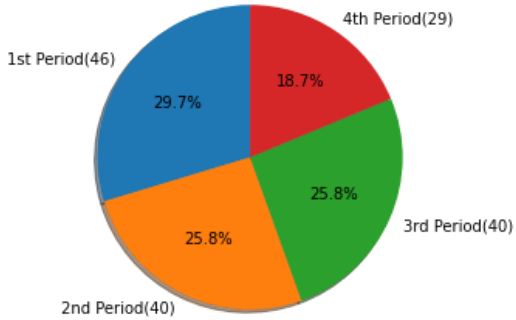
Stocks which are more than Average of Total Risk



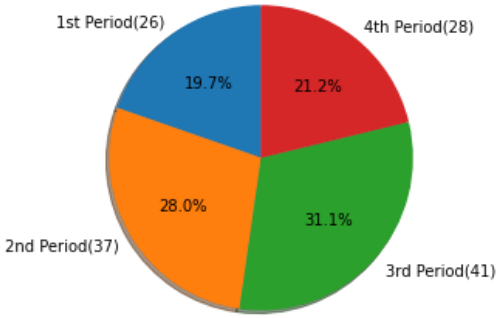
Stocks which are more than Average of Systematic Risk



Stocks which are more than Average of Abs. Win rate



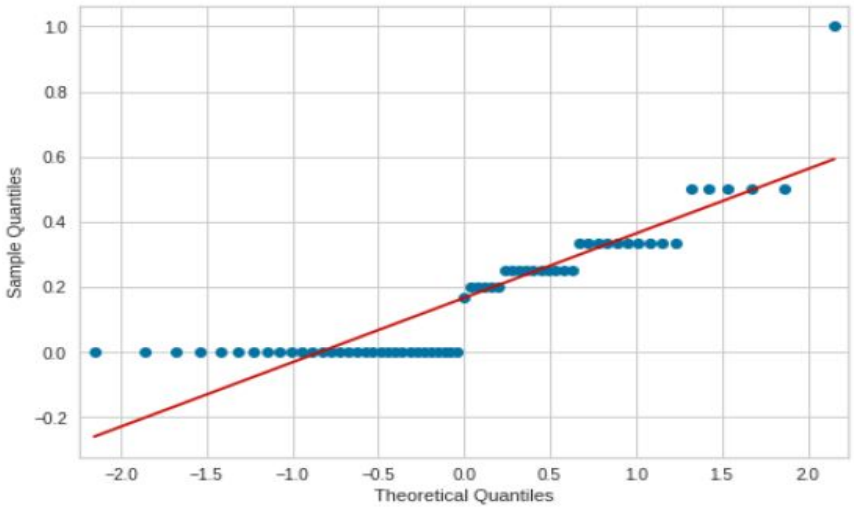
Stocks which are more than Average of Rel. Win rate



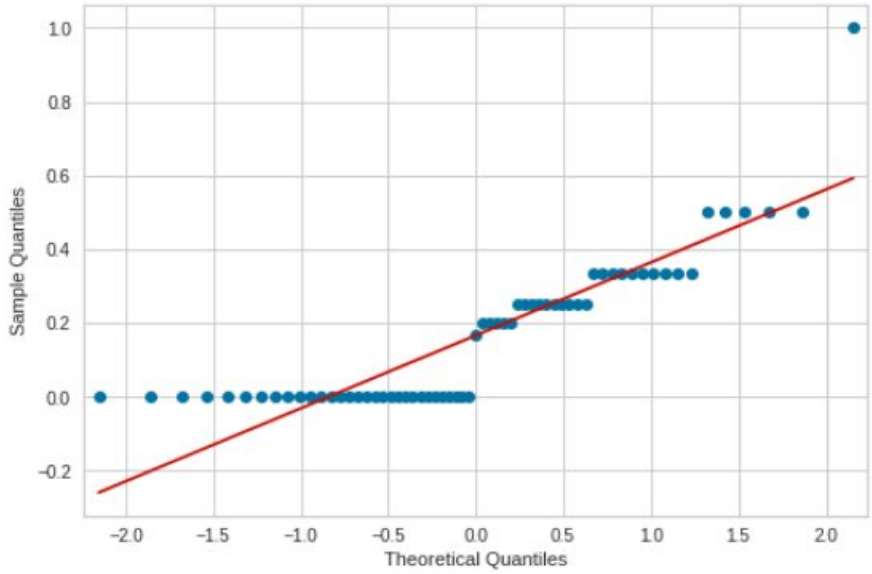
Normality of Data

The quantile-quantile (Q-Q) plot shows the distribution of the data against the expected normal distribution.The below plots shows the tendency to normal distribution of the features in the data.

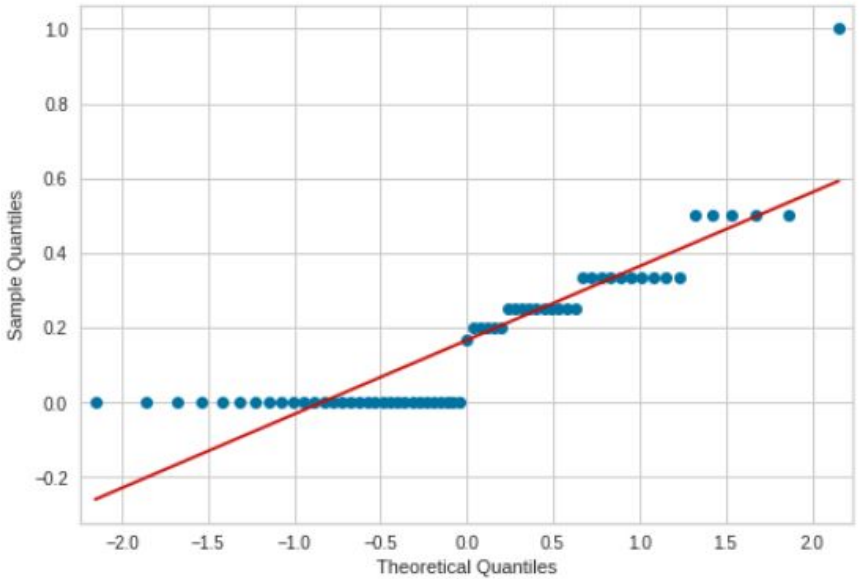
1. Large B/P



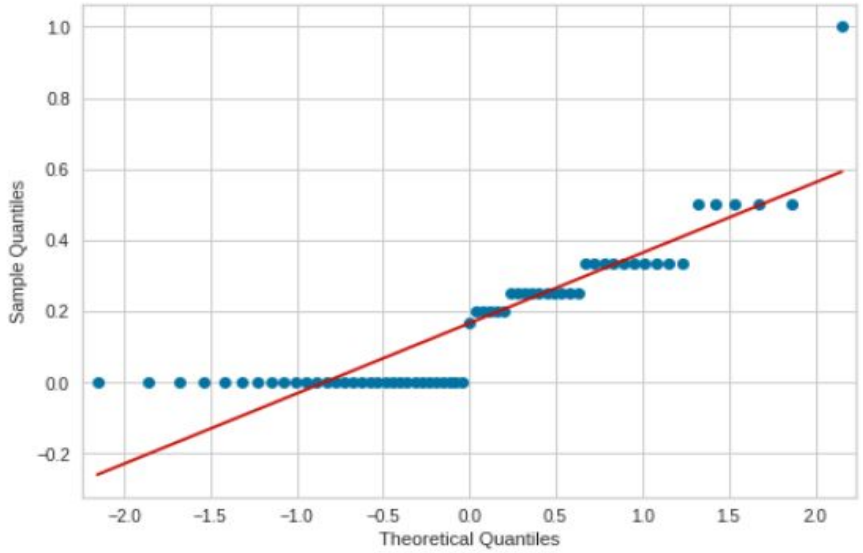
2. Large ROE



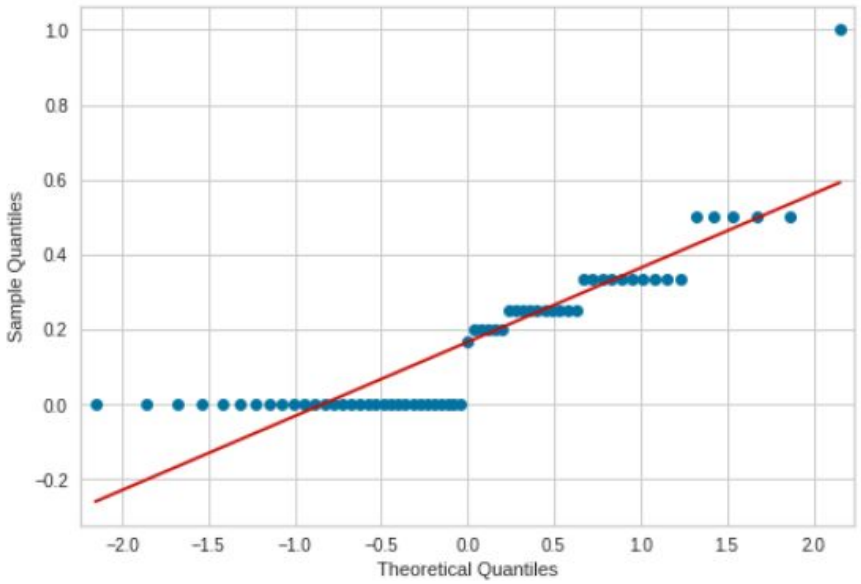
3. Large S/P



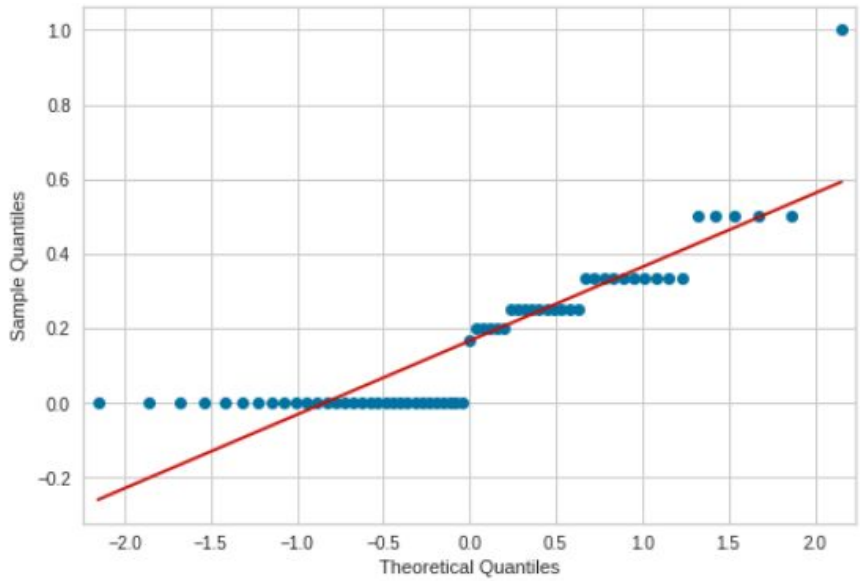
4. Large Return Rate in the last quarter



5. Large Market Value



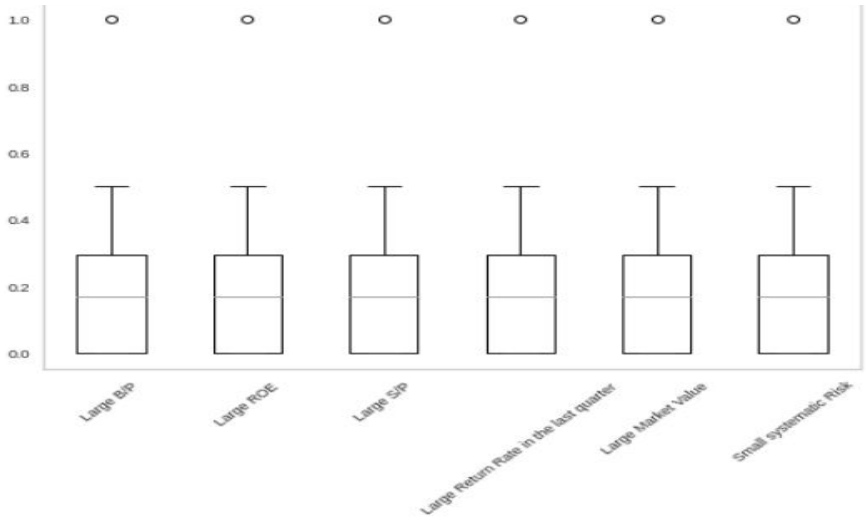
6. Small systematic Risk

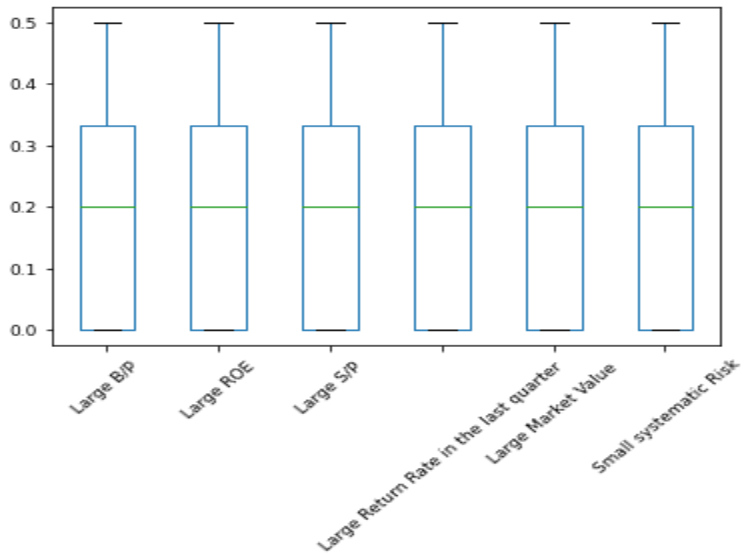


We can say that data is not normal as it is deviating from a straight line.

Influential points

Influential points are those outliers which greatly affects the slope of the regression line. It may cause the Determination coefficient very big or sometimes too small that's why it's better to remove them and then compute the regression line.The following plots show the box plots of features before and after outliers removal.

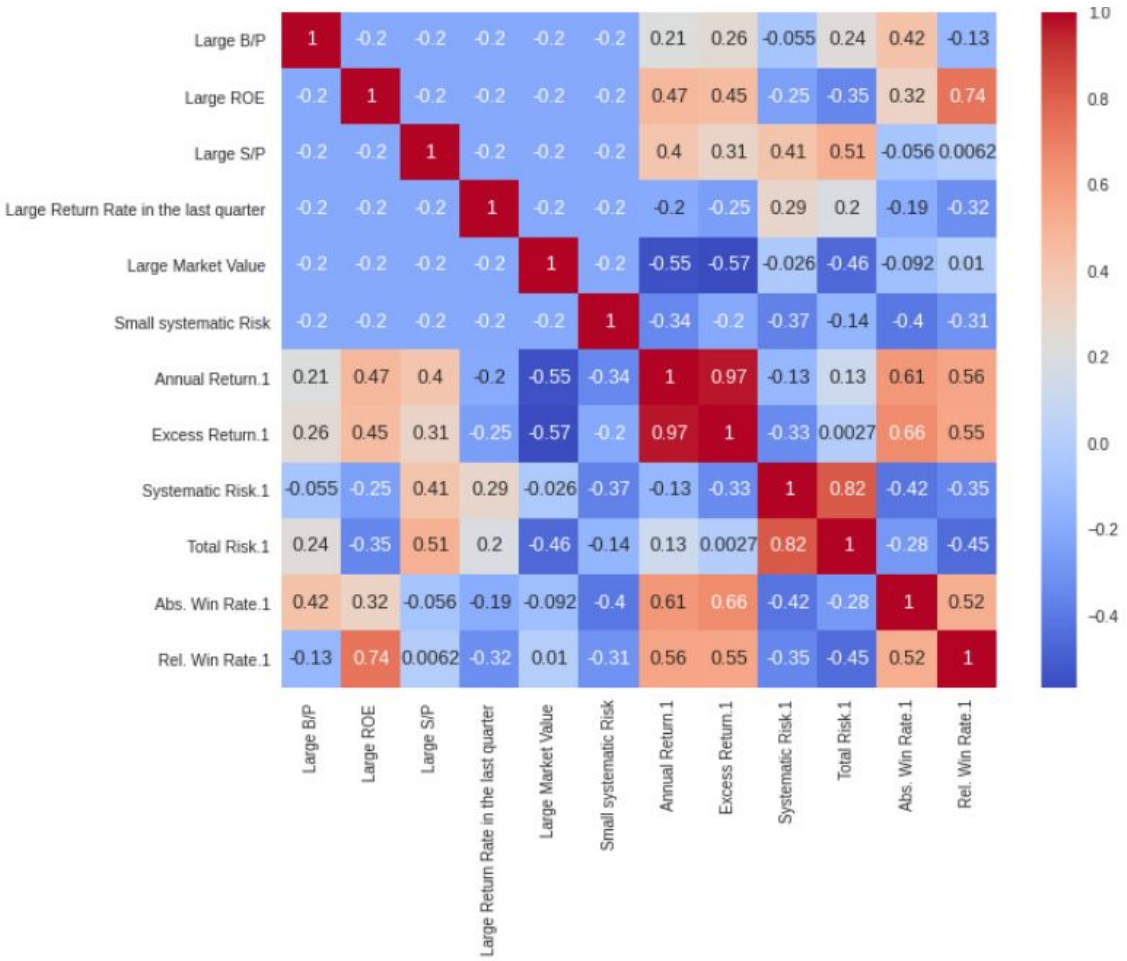




After removal of influential points the data distribution looks like this.

Correlation plot

Correlation plot is to show how much a variable is affected by all other variables. The correlation values lie between [-1,1].Perfectly correlating variables will be having value 1(positive correlation), -1(negative correlation) in the correlation matrix.



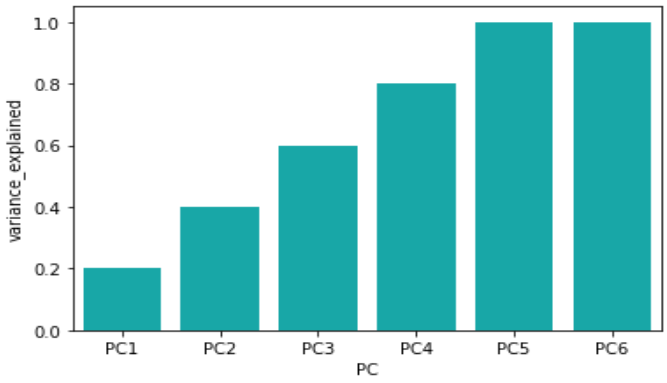
From the above matrix we can say that the correlation between features is quite low and hence we can say all that are not much linearly dependent..

Principal Component Analysis:

Principal component analysis is the process of computing the principal components and using them to perform a change of basis on the data, sometimes using only the first few principal components and ignoring the rest.

For the given feature matrix the cumulative variance explained by principal components is as follows:

```
Variance: [0.19999994 0.39999988 0.59999981 0.79999975 0.99999969 1.]
```



From the above plot we can see that 100% percent variance is explained while considering all 6 principal components of the data so we choose to consider all the principle components for building the model.

	Large B/P	Large ROE	Large S/P	Large Return Rate in the last quarter	Large Market Value	Small systematic Risk
PC-1	-0.000000	-0.280362	-0.565721	0.569912	-0.207184	0.483355
PC-2	0.912871	-0.182574	-0.182574	-0.182574	-0.182574	-0.182574
PC-3	-0.000000	0.326438	-0.661497	-0.080517	0.633831	-0.218255
PC-4	0.000000	0.768577	-0.159266	0.118150	-0.593256	-0.134205
PC-5	0.000000	0.155321	-0.130445	-0.674358	-0.058135	0.707617
PC-6	0.408248	0.408248	0.408248	0.408248	0.408248	0.408248

Principle component is the attribute which has the highest absolute value.We can see the corresponding principal components from the above dataframe.

Factor Analysis:

- We can perform Factor Analysis for the given dataset to decide factorability and to know how many factors we can have.
- Before we proceed to find the factors we need to check for the factorability of our dataset.
- Factorability can be tested using Bartlett's method.

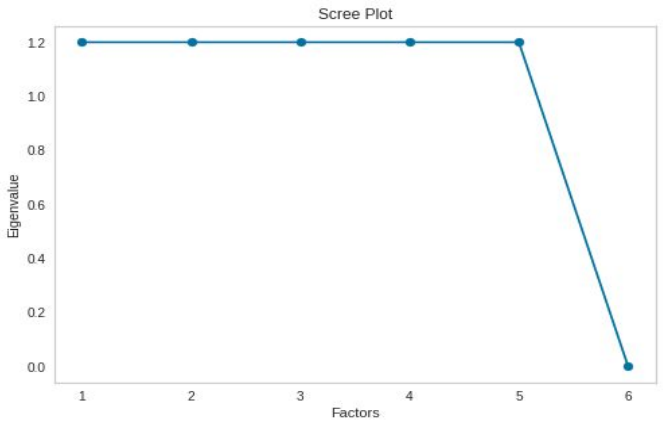
Bartlett's Method:

- In Bartlett's sphericity test, chi-square value and p-value is calculated.
- For this dataset
  - (Chi square value,p-value) is :-
    - `(652.6814476921419, 1.0907491324077153e-129)`
- As the p-value is almost 0 it is significant that factors for this dataset are possible.

Choosing the Number of Factors:

- Number of factors based on the given observed variables can be chosen by plotting Scree plot.
- Eigen values are plotted in Scree plot.
  - Eigen values are

	Original_Eigenvalues
0	1.200000
1	1.200000
2	1.200000
3	1.200000
4	1.200000
5	0.000002



- From the above plot we can conclude that 5 factors (5 unobserved variables) are possible in this dataset.

	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6
SS Loadings	1.195000	1.195000	1.195000	1.195000	1.195000	0.000000
Proportion Var	0.199167	0.199167	0.199167	0.199167	0.199167	0.000000
Cumulative Var	0.199167	0.398333	0.597500	0.796666	0.995833	0.995833

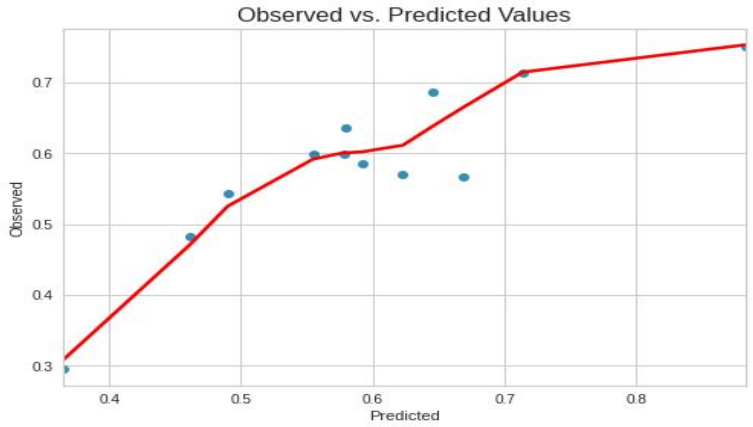
- Here, SS Loadings are the values for the factors and Proportion var is the proportion of variance contributing to the total variance and Cumulative variance is the amount of variance explained by the original data.
- We can see that maximum cumulative variance by considering all the factors, here we consider 6 features to perform a regression model.

Regression Model:

Model for **Annual Return.1** evaluation parameter

The linear model is: Y = -12.438 + 13.148*large b/p + 13.355*large ROE + 13.286*large s/p+ 12.912*large return rates + 12.719*large market sales + 12.817*small system risk

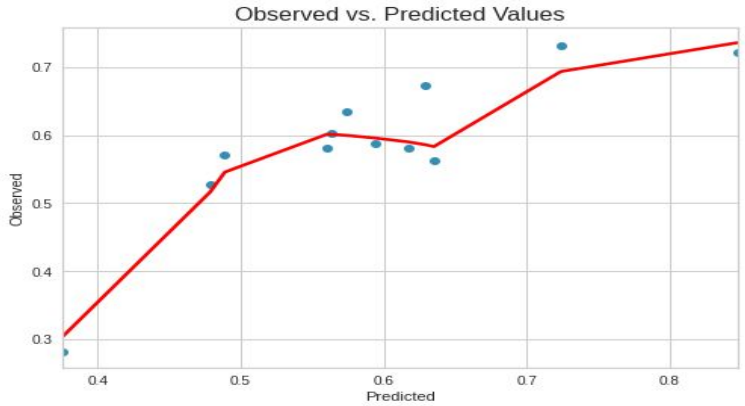
Variance score: 0.7019488999882331



Model for **Excess Return.1** evaluation parameter

The linear model is: Y = -16.664 + 17.401*large b/p + 17.568*large ROE + 17.455*large s/p+ 17.114*large return rates + 16.927*large market sales + 17.128*small system risk

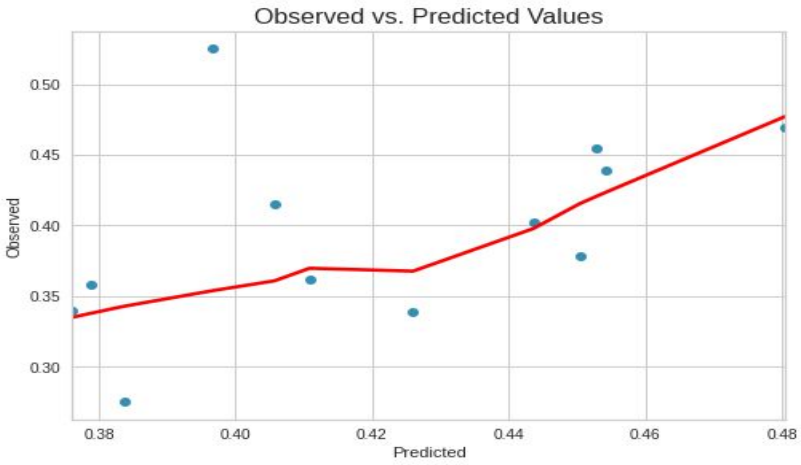
Variance score: 0.6717835969280815



Model for **Systematic Risk.1** evaluation parameter

The linear model is: Y = 14.279 + -13.893*large b/p + -13.983*large ROE + -13.669*large s/p+ -13.757*large return rates + -13.875*large market sales + -14.056*small system risk

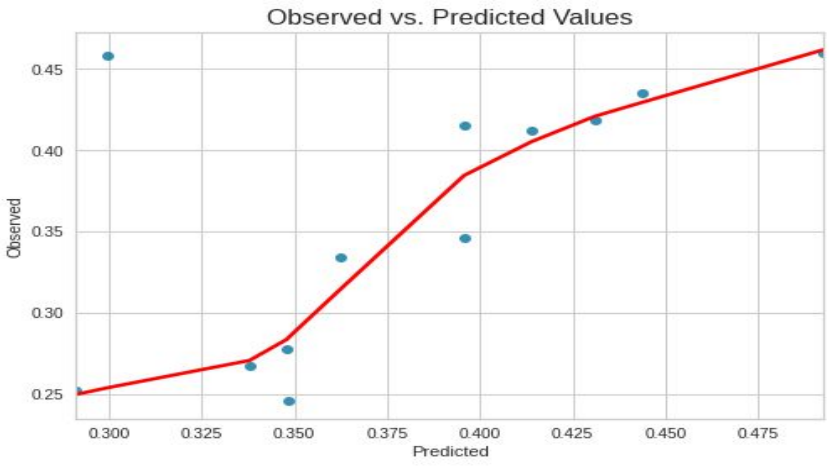
Variance score: 0.0814566258408218



Model for **Total Risk.1** evaluation parameter

The linear model is: Y = 7.0886 + -6.5722*large b/p + -6.8841*large ROE + -6.431*large s/p+ -6.6196*large return rates + -6.9572*large market sales + -6.8111*small system risk

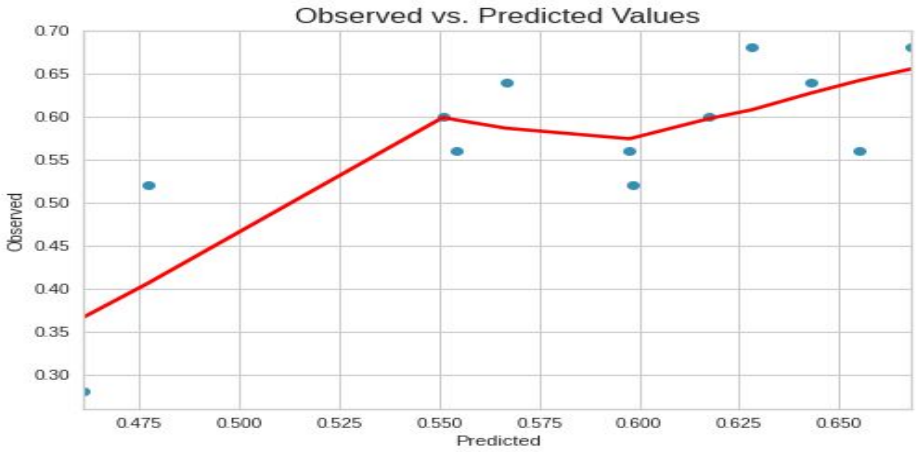
Variance score: 0.30663192303227027



Model for **Abs. Win Rate.1** evaluation parameter

The linear model is: Y = -20.205 + 20.978*large b/p + 20.954*large ROE + 20.767*large s/p+ 20.718*large return rates + 20.75*large market sales + 20.591*small system risk

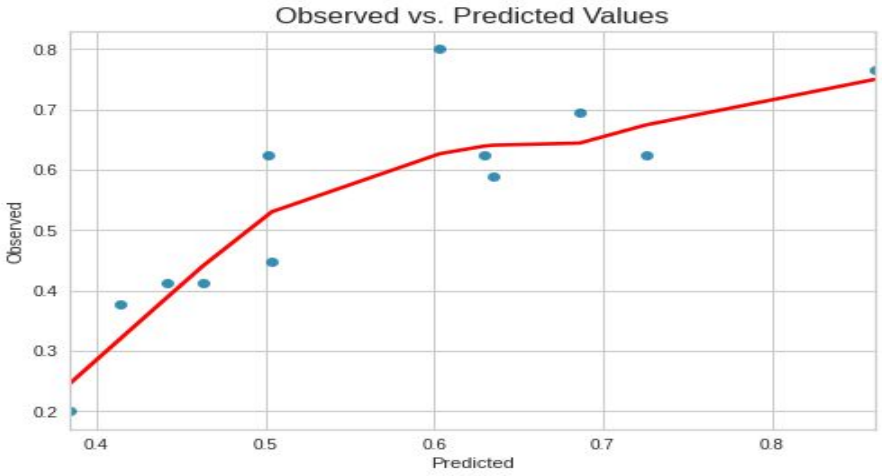
Variance score: 0.5045284905611858



Model for **Rel. Win Rate.1** evaluation parameter

The linear model is: Y = -46.851 + 47.324*large b/p + 47.982*large ROE + 47.441*large s/p+ 47.204*large return rates + 47.448*large market sales + 47.194*small system risk

Variance score: 0.6624426054758297



Ordinary least squares (OLS) regression:

It is a statistical method of analysis that estimates the relationship between one or more independent variables and a dependent variable; this method estimates the best coefficients for the independent variables by minimizing the sum of the squares of the difference between the observed and predicted value of the dependent variable configured as a straight line.

OLS summary as follows:

OLS Regression Results

Dep. Variable:	Annual Return.1	R-squared (uncentered):	0.990
Model:	OLS	Adj. R-squared (uncentered):	0.988
Method:	Least Squares	F-statistic:	644.1
Date:	Fri, 04 Dec 2020	Prob (F-statistic):	2.13e-37
Time:	09:18:18	Log-Likelihood:	62.057
No. Observations:	45	AIC:	-112.1
Df Residuals:	39	BIC:	-101.3
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Large B/P	0.7031	0.049	14.352	0.000	0.604	0.802
Large ROE	0.9135	0.050	18.351	0.000	0.813	1.014
Large S/P	0.8442	0.050	16.978	0.000	0.744	0.945
Large Return Rate in the last quarter	0.4679	0.049	9.498	0.000	0.368	0.567
Large Market Value	0.2811	0.048	5.896	0.000	0.185	0.377
Small systematic Risk	0.3766	0.047	7.985	0.000	0.281	0.472

Omnibus:	3.579	Durbin-Watson:	2.099
Prob(Omnibus):	0.167	Jarque-Bera (JB):	3.382
Skew:	-0.624	Prob(JB):	0.184
Kurtosis:	2.503	Cond. No.	2.53

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS Regression Results

Dep. Variable:	Excess Return.1	R-squared (uncentered):	0.986
Model:	OLS	Adj. R-squared (uncentered):	0.984
Method:	Least Squares	F-statistic:	473.7
Date:	Fri, 04 Dec 2020	Prob (F-statistic):	7.88e-35
Time:	09:18:18	Log-Likelihood:	55.222
No. Observations:	45	AIC:	-98.44
Df Residuals:	39	BIC:	-87.60
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Large B/P	0.7275	0.057	12.758	0.000	0.612	0.843
Large ROE	0.8993	0.058	15.521	0.000	0.782	1.017
Large S/P	0.7855	0.058	13.570	0.000	0.668	0.903
Large Return Rate in the last quarter	0.4423	0.057	7.714	0.000	0.326	0.558
Large Market Value	0.2629	0.055	4.737	0.000	0.151	0.375
Small systematic Risk	0.4608	0.055	8.394	0.000	0.350	0.572

Omnibus:	2.634	Durbin-Watson:	2.016
Prob(Omnibus):	0.268	Jarque-Bera (JB):	2.378
Skew:	-0.475	Prob(JB):	0.304
Kurtosis:	2.395	Cond. No.	2.53

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS Regression Results

Dep. Variable:	Systematic Risk.1	R-squared (uncentered):	0.966
Model:	OLS	Adj. R-squared (uncentered):	0.961
Method:	Least Squares	F-statistic:	187.0
Date:	Fri, 04 Dec 2020	Prob (F-statistic):	3.78e-27
Time:	09:18:18	Log-Likelihood:	51.486
No. Observations:	45	AIC:	-90.97
Df Residuals:	39	BIC:	-80.13
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Large B/P	0.3944	0.062	6.365	0.000	0.269	0.520
Large ROE	0.2997	0.063	4.761	0.000	0.172	0.427
Large S/P	0.6145	0.063	9.771	0.000	0.487	0.742
Large Return Rate in the last quarter	0.5289	0.062	8.490	0.000	0.403	0.655
Large Market Value	0.4037	0.060	6.695	0.000	0.282	0.526
Small systematic Risk	0.2259	0.060	3.788	0.001	0.105	0.347

Omnibus:	2.416	Durbin-Watson:	2.020
Prob(Omnibus):	0.299	Jarque-Bera (JB):	1.886
Skew:	0.502	Prob(JB):	0.389
Kurtosis:	3.005	Cond. No.	2.53

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS Regression Results

Dep. Variable:	Total Risk.1	R-squared (uncentered):	0.967
Model:	OLS	Adj. R-squared (uncentered):	0.962
Method:	Least Squares	F-statistic:	192.4
Date:	Fri, 04 Dec 2020	Prob (F-statistic):	2.22e-27
Time:	09:18:18	Log-Likelihood:	54.943
No. Observations:	45	AIC:	-97.89
Df Residuals:	39	BIC:	-87.05
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Large B/P	0.5203	0.057	9.068	0.000	0.404	0.636
Large ROE	0.2063	0.058	3.539	0.001	0.088	0.324
Large S/P	0.6599	0.058	11.330	0.000	0.542	0.778
Large Return Rate in the last quarter	0.4725	0.058	8.190	0.000	0.356	0.589
Large Market Value	0.1313	0.056	2.351	0.024	0.018	0.244
Small systematic Risk	0.2791	0.055	5.052	0.000	0.167	0.391

Omnibus:	5.861	Durbin-Watson:	2.038
Prob(Omnibus):	0.053	Jarque-Bera (JB):	4.851
Skew:	0.781	Prob(JB):	0.0884
Kurtosis:	3.381	Cond. No.	2.53

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS Regression Results

Dep. Variable:	Abs. Win Rate.1	R-squared (uncentered):	0.983
Model:	OLS	Adj. R-squared (uncentered):	0.980
Method:	Least Squares	F-statistic:	372.4
Date:	Fri, 04 Dec 2020	Prob (F-statistic):	7.94e-33
Time:	09:18:18	Log-Likelihood:	51.463
No. Observations:	45	AIC:	-90.93
Df Residuals:	39	BIC:	-80.09
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Large B/P	0.7621	0.062	12.294	0.000	0.637	0.887
Large ROE	0.7432	0.063	11.798	0.000	0.616	0.871
Large S/P	0.5549	0.063	8.818	0.000	0.428	0.682
Large Return Rate in the last quarter	0.5027	0.062	8.065	0.000	0.377	0.629
Large Market Value	0.5454	0.060	9.042	0.000	0.423	0.667
Small systematic Risk	0.3820	0.060	6.400	0.000	0.261	0.503

Omnibus:	0.405	Durbin-Watson:	1.902
Prob(Omnibus):	0.817	Jarque-Bera (JB):	0.067
Skew:	0.082	Prob(JB):	0.967
Kurtosis:	3.095	Cond. No.	2.53

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

OLS Regression Results

Dep. Variable:	Rel. Win Rate.1	R-squared (uncentered):	0.979
Model:	OLS	Adj. R-squared (uncentered):	0.975
Method:	Least Squares	F-statistic:	297.5
Date:	Fri, 04 Dec 2020	Prob (F-statistic):	5.76e-31
Time:	09:18:18	Log-Likelihood:	46.904
No. Observations:	45	AIC:	-81.81
Df Residuals:	39	BIC:	-70.97
Df Model:	6		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Large B/P	0.4473	0.069	6.521	0.000	0.309	0.586
Large ROE	1.1190	0.070	16.053	0.000	0.978	1.260
Large S/P	0.5754	0.070	8.263	0.000	0.435	0.716
Large Return Rate in the last quarter	0.3307	0.069	4.794	0.000	0.191	0.470
Large Market Value	0.5973	0.067	8.948	0.000	0.462	0.732
Small systematic Risk	0.3330	0.066	5.042	0.000	0.199	0.467

Omnibus:	2.566	Durbin-Watson:	1.945
Prob(Omnibus):	0.277	Jarque-Bera (JB):	2.134
Skew:	0.532	Prob(JB):	0.344
Kurtosis:	2.920	Cond. No.	2.53

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

- If $p > 0.05$, we fail to reject null hypothesis otherwise we reject null hypothesis.

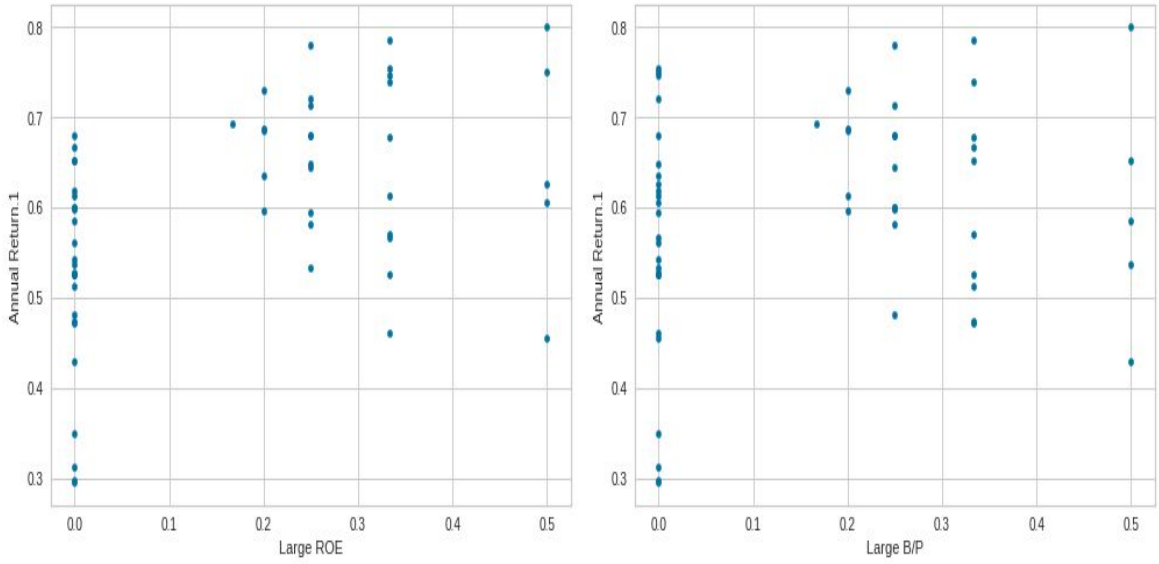
- After observing p values if p value > 0.05 we assume that features are less contributed to evaluation parameters. After observing P values in the above summary all are below 0.05 so we can assume that all are important for predicting evaluation parameters.

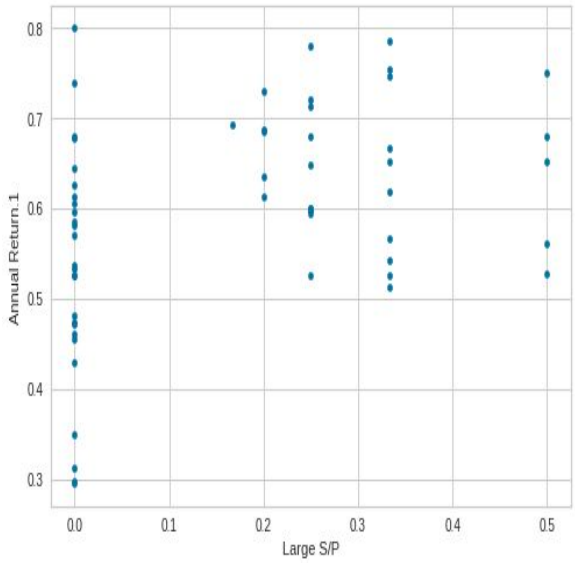
Test of assumptions:

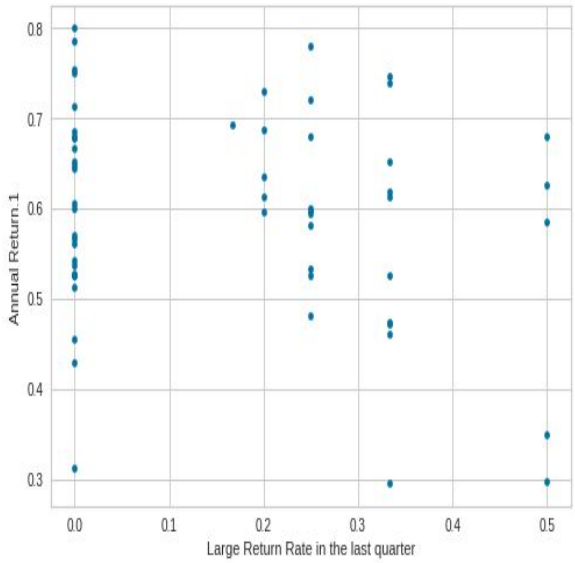
LINEARITY:

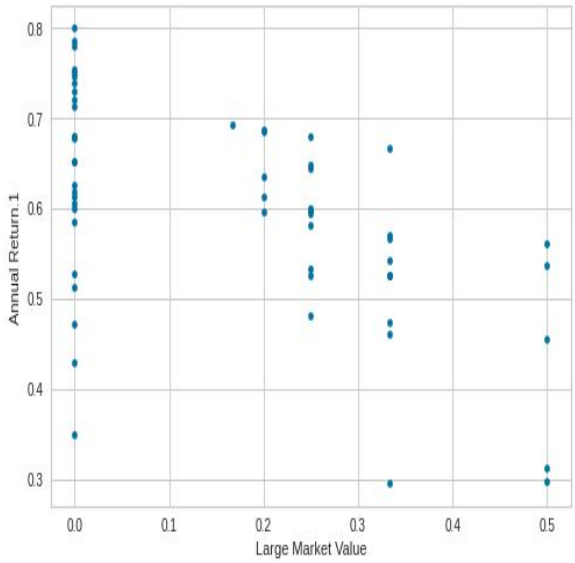
- linear regression needs the relationship between the independent and dependent variables to be linear.
- This assumption can be tested by plotting scatter plots for each dependent variable against each independent variable as shown below.

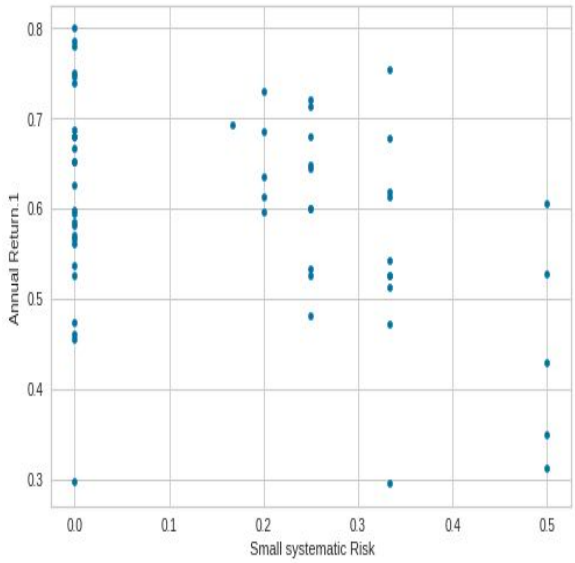
Annual return



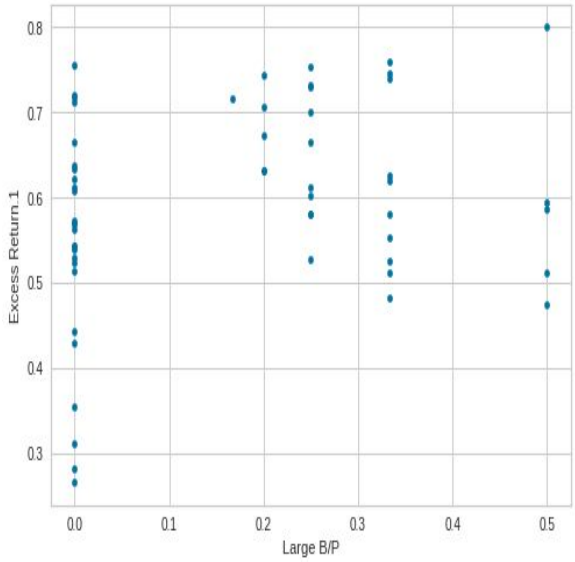

Annual Return.1
Large S/P

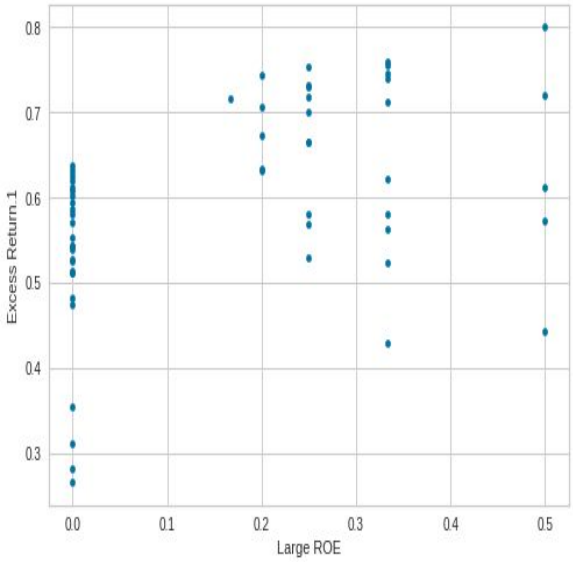

Annual Return.1
Large Return Rate in the last quarter

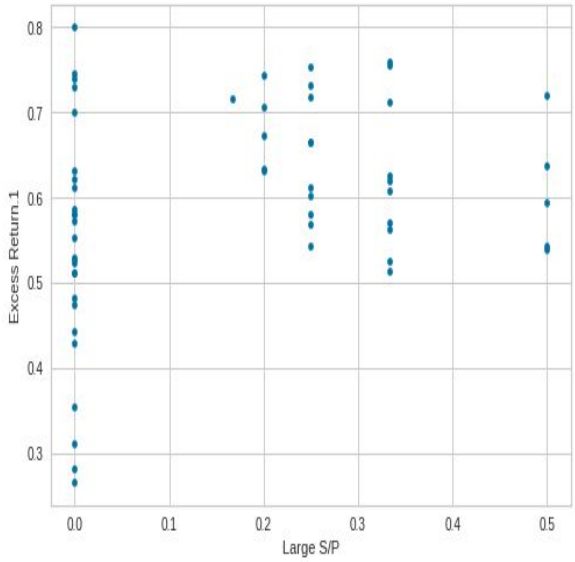

Annual Return.1
Large Market Value

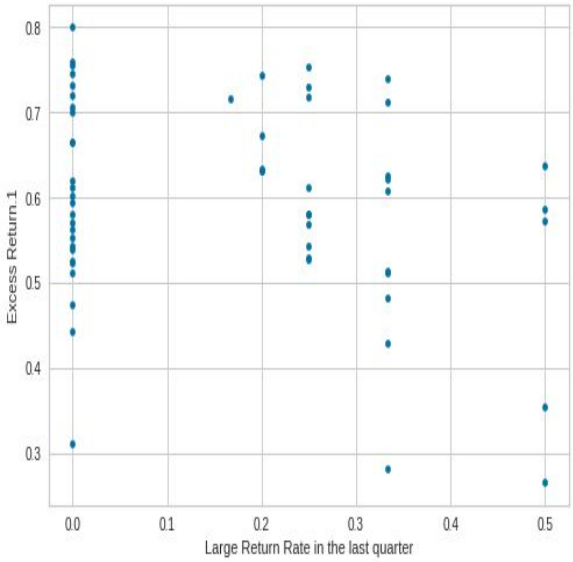

Annual Return.1
Small systematic Risk

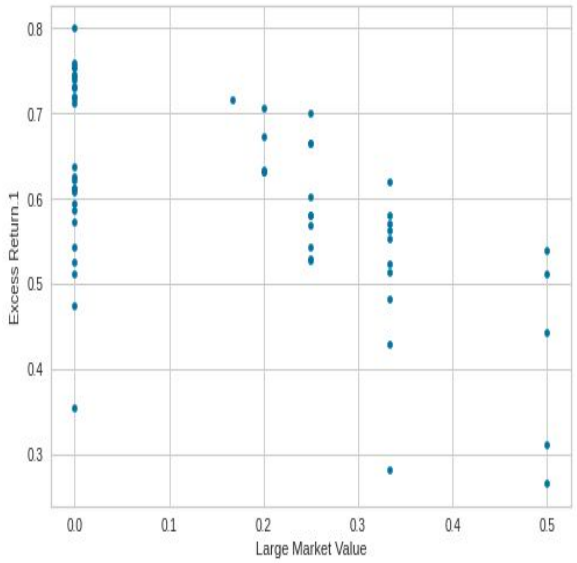
Excess return

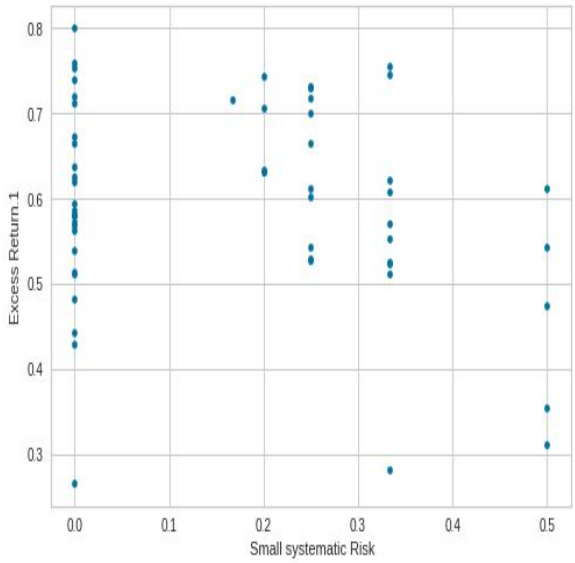




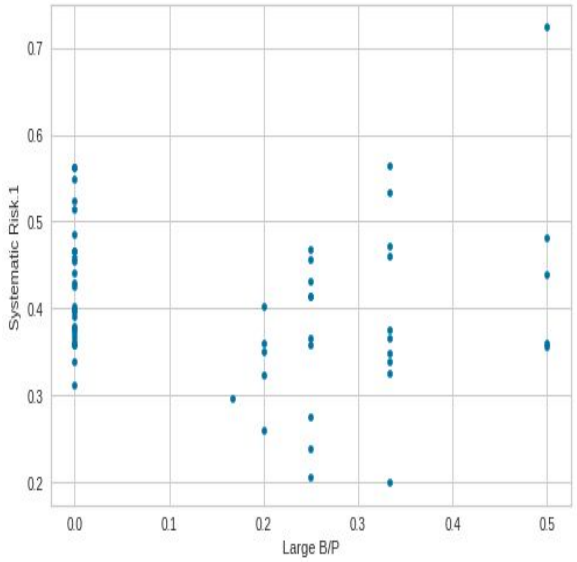


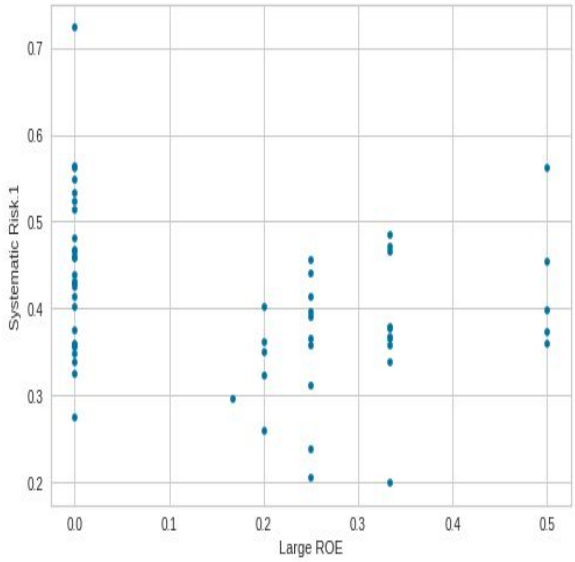


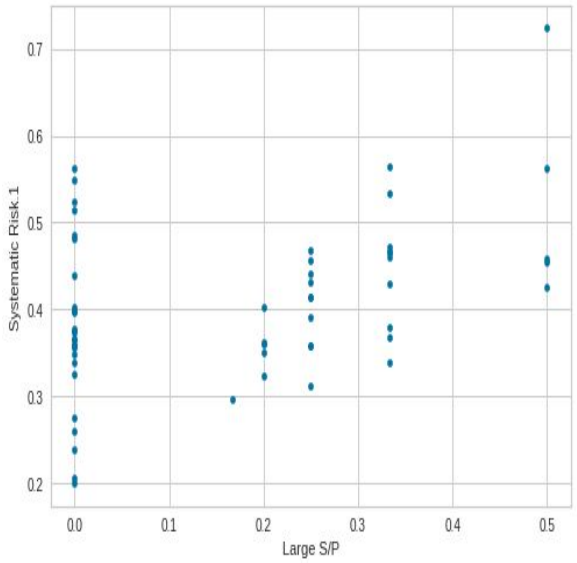


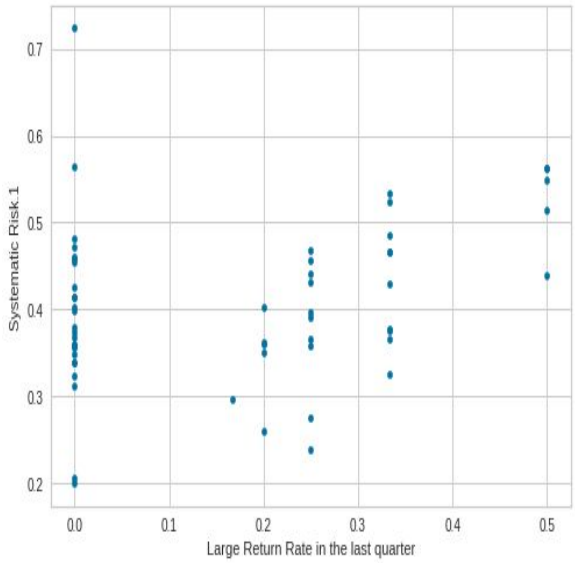


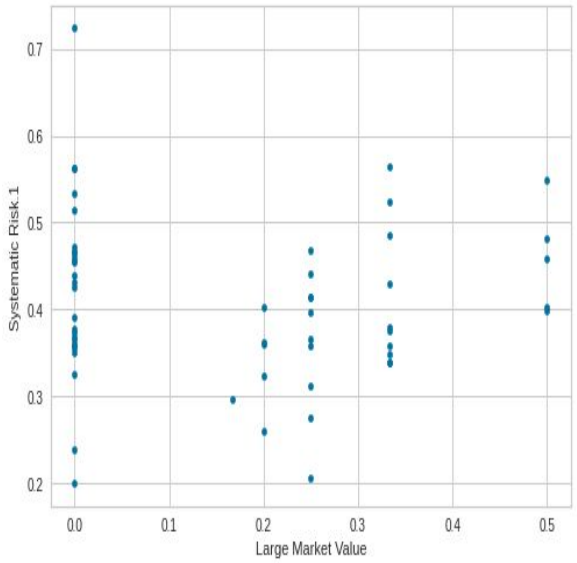
Systematic risk

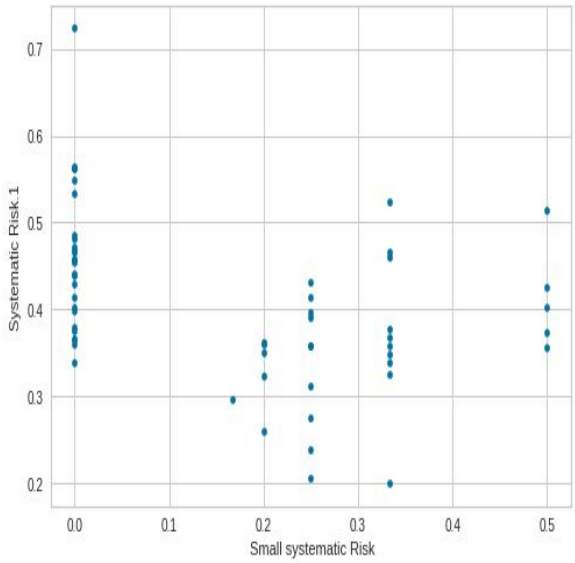




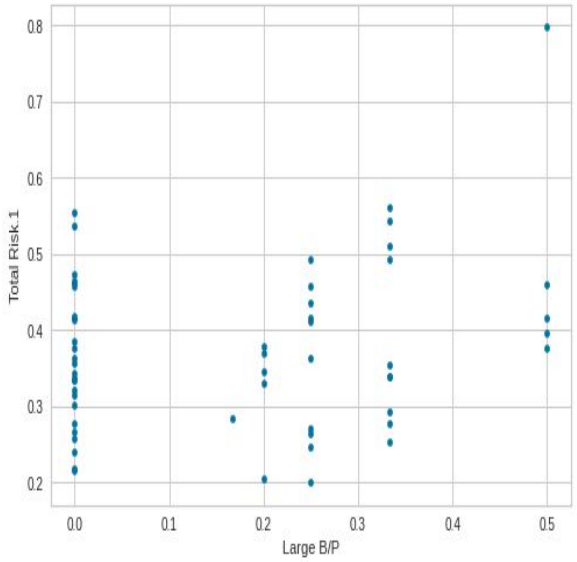

Systematic Risk.1
Large S/P

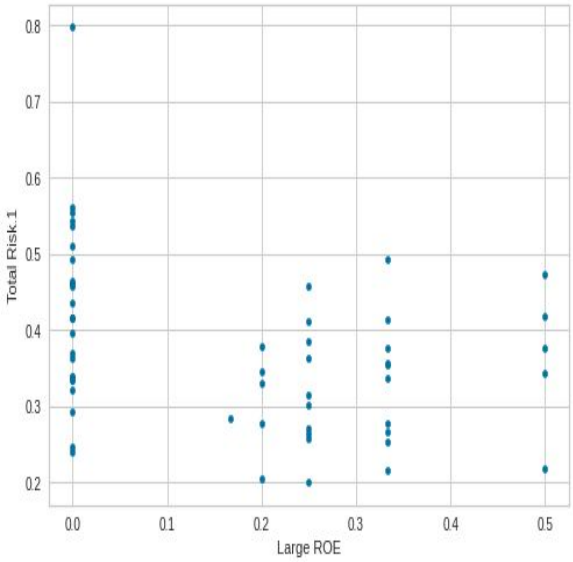

Systematic Risk.1
Large Return Rate in the last quarter

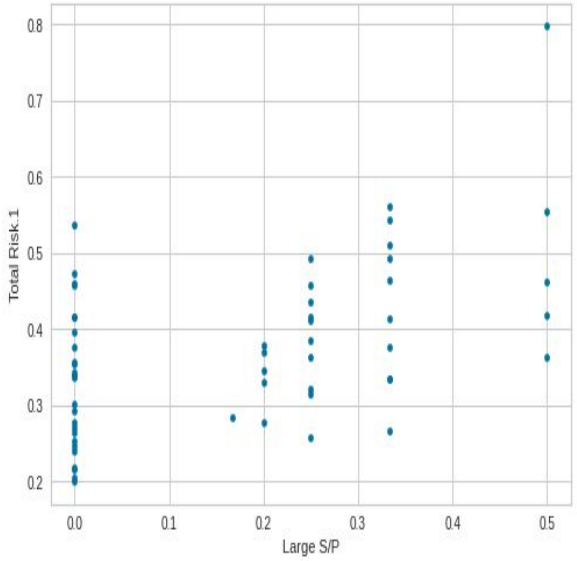

Systematic Risk.1
Large Market Value

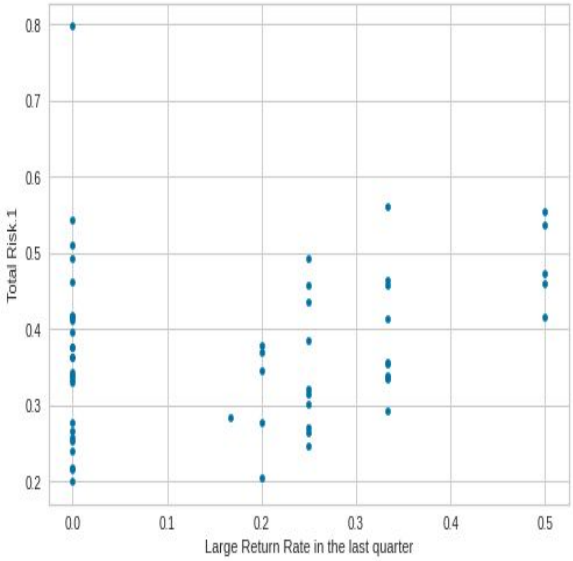

Systematic Risk.1
Small systematic Risk

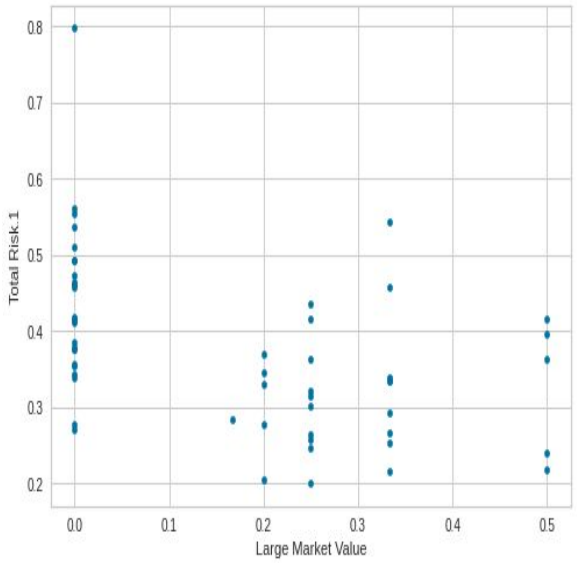
Total risk

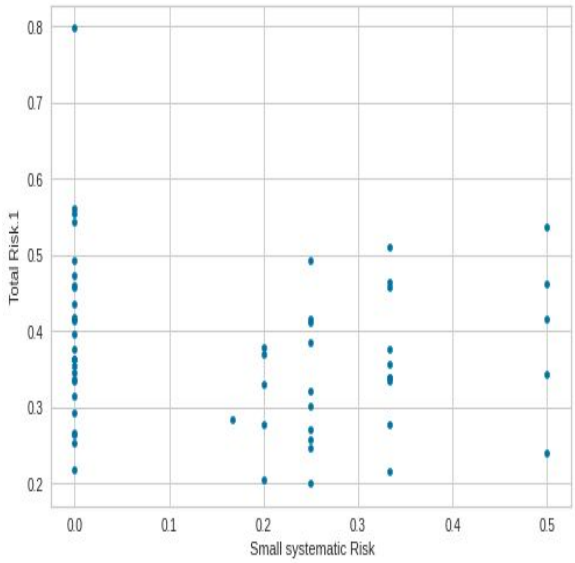

Total Risk.1
Large B/P


Total Risk.1
Large ROE

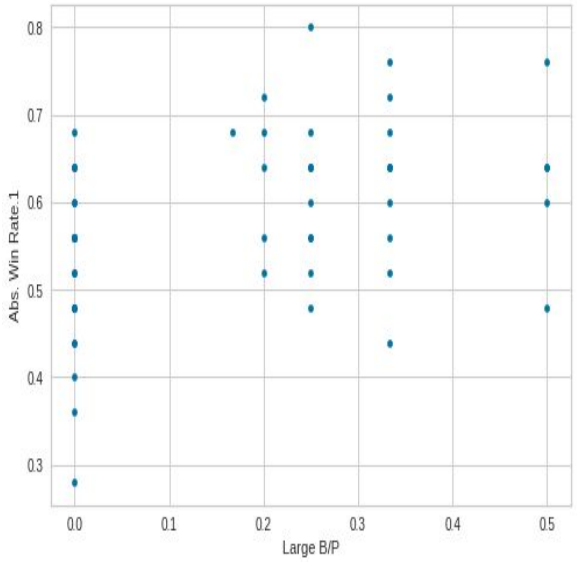

Total Risk.1
Large S/P

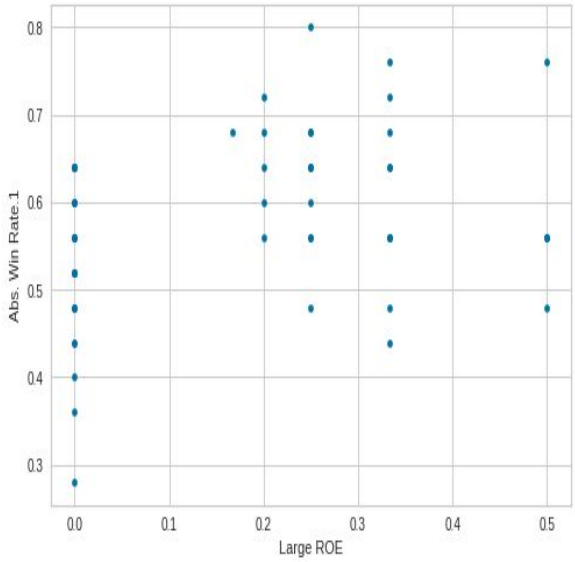

Total Risk.1
Large Return Rate in the last quarter

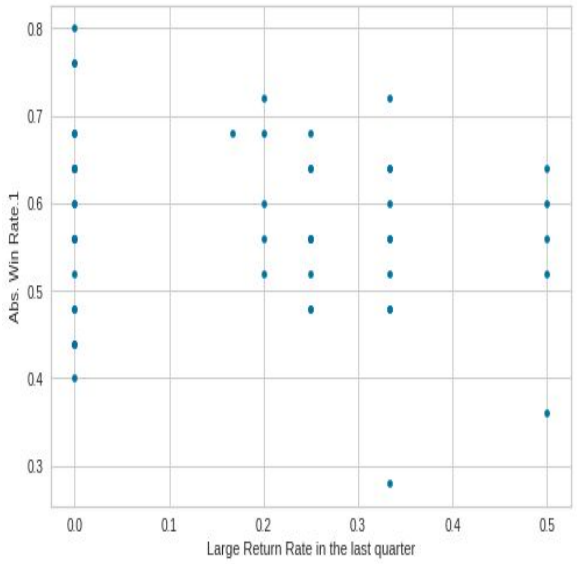

Total Risk.1
Large Market Value

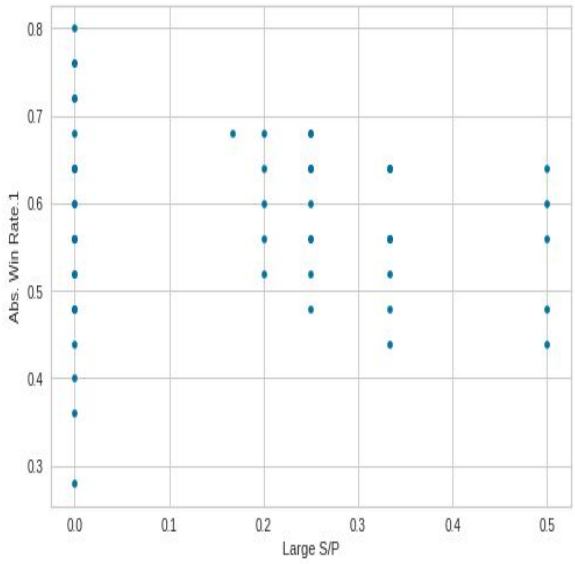

Total Risk.1
Small systematic Risk

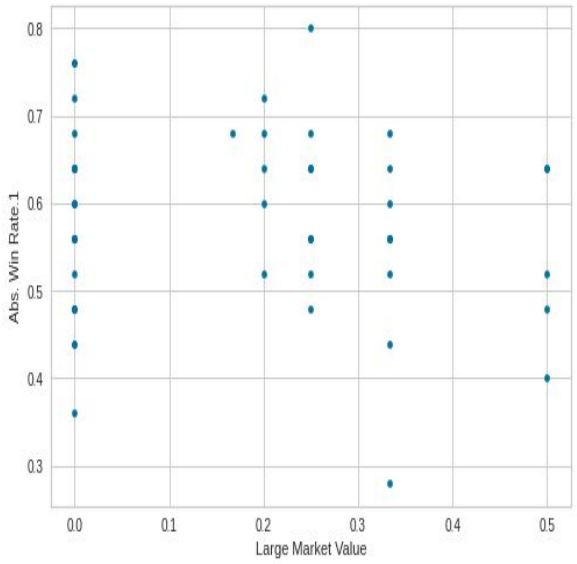
Abs. win rate

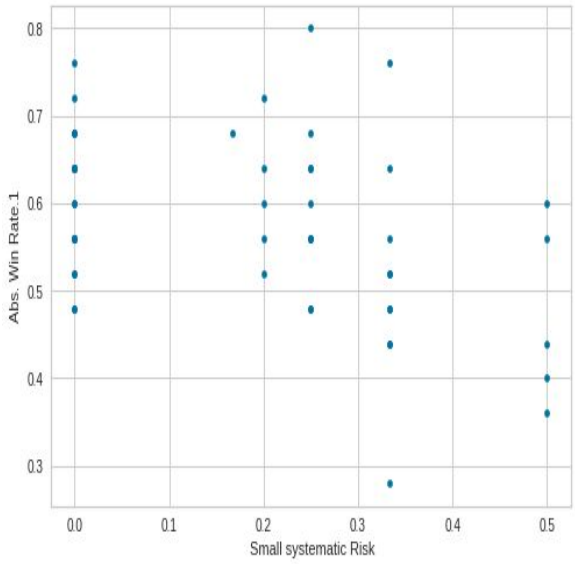

Abs. Win Rate.1
Large B/P


Abs. Win Rate.1
Large ROE

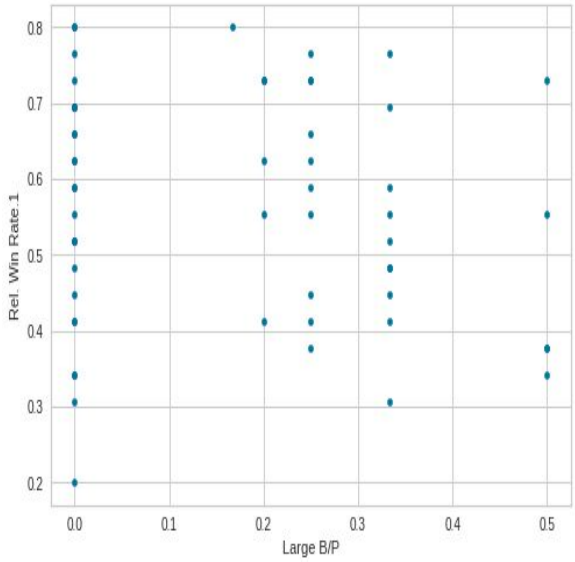

Abs. Win Rate.1
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.1
0.2
0.3
0.4
0.5
Large Return Rate in the last quarter

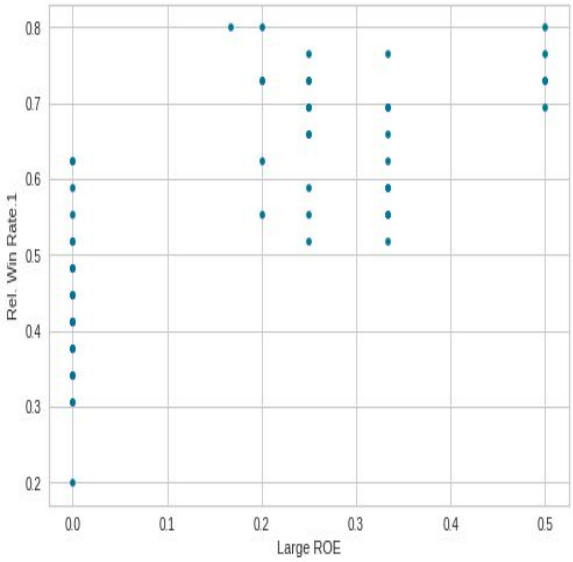

Abs. Win Rate.1
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.1
0.2
0.3
0.4
0.5
Large S/P

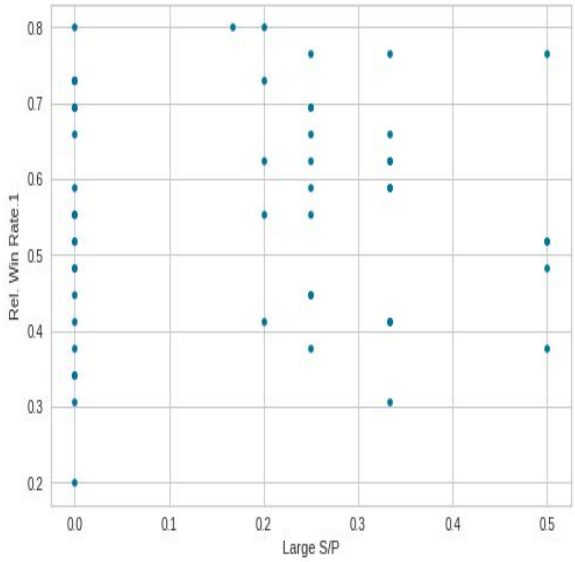

Abs. Win Rate.1
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.1
0.2
0.3
0.4
0.5
Large Market Value

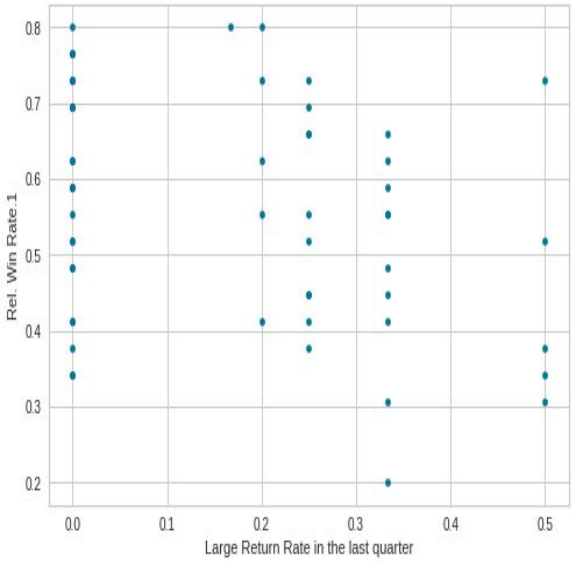

Abs. Win Rate.1
0.8
0.7
0.6
0.5
0.4
0.3
0.0
0.1
0.2
0.3
0.4
0.5
Small systematic Risk

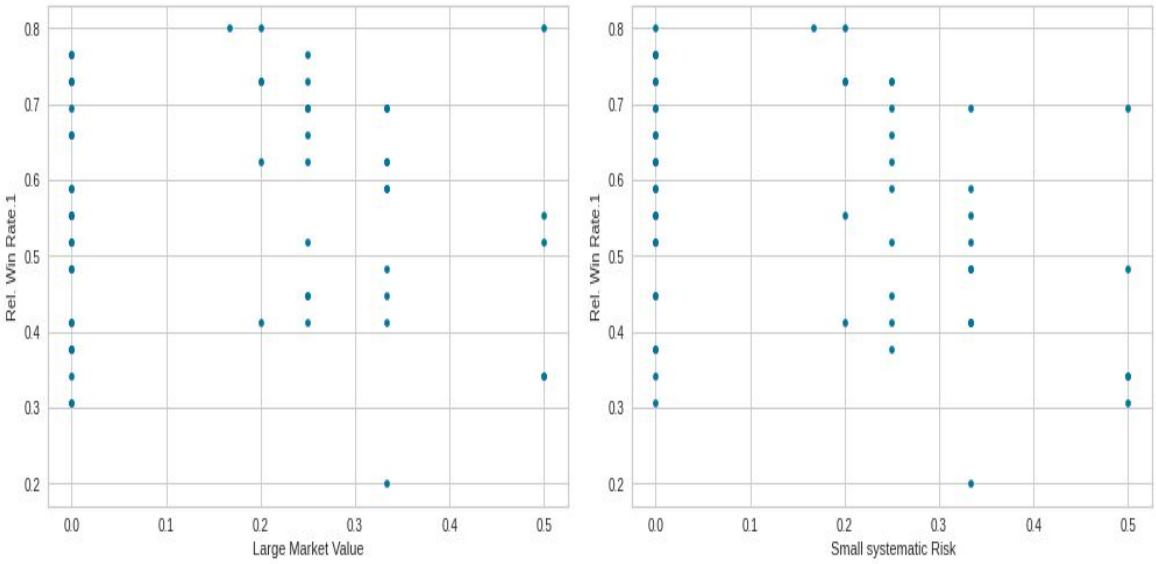
Rel. win rate


Rel. Win Rate.1
Large B/P


Rel. Win Rate.1
Large ROE


Rel. Win Rate.1
Large S/P

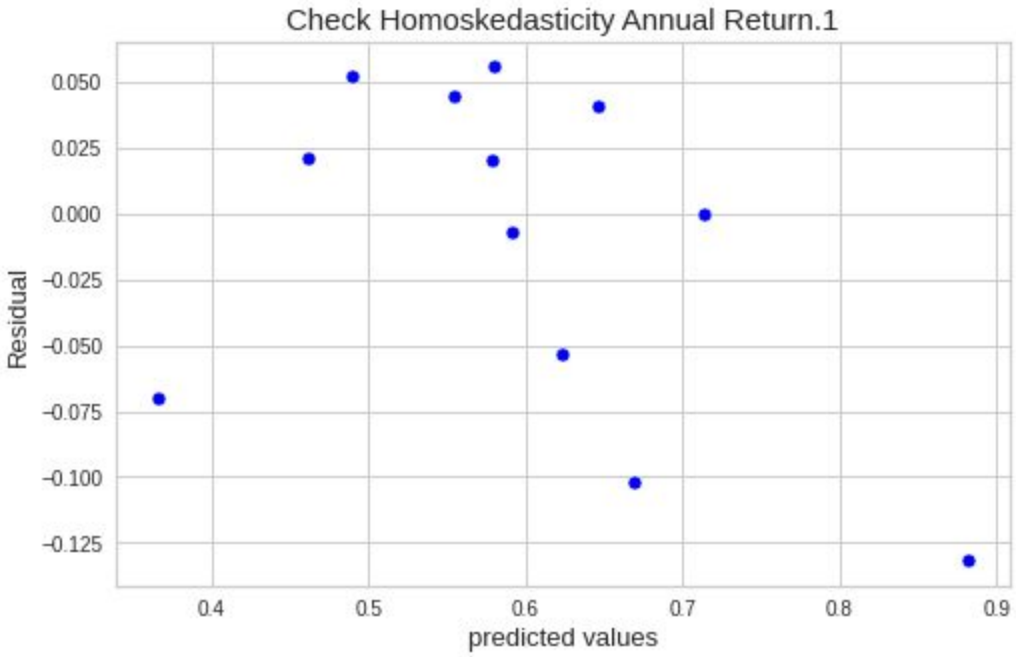

Rel. Win Rate.1
Large Return Rate in the last quarter

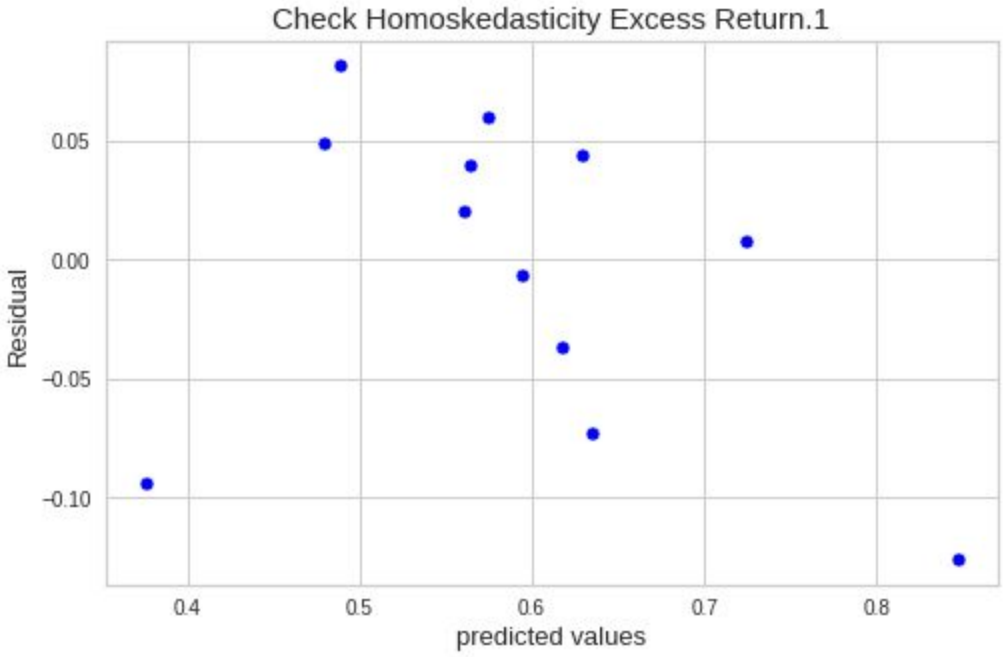


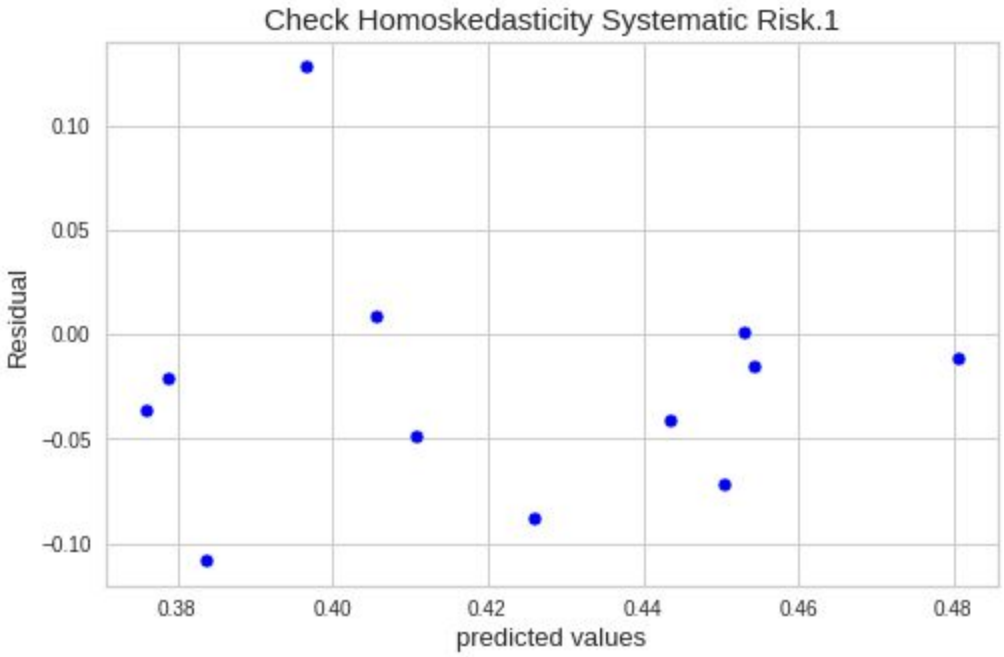
- By observing above graphs we can say that each independent variable is linearly related to the every dependent variable because they are not deviating much from the linear line between the dependent and independent variables.
- Hence linearity assumption is proved

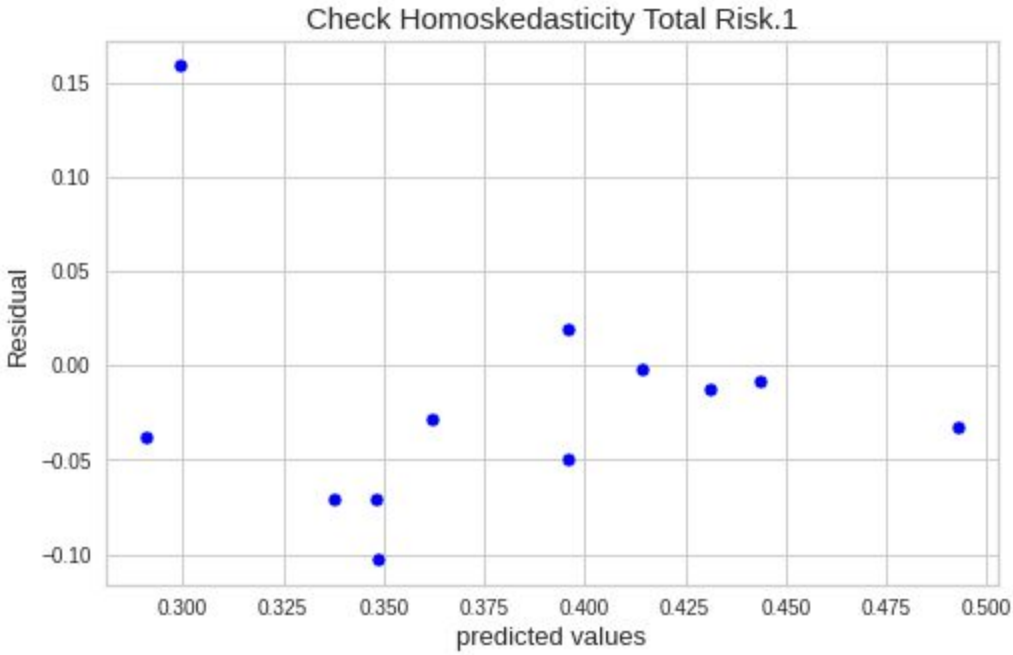
HOMOSCEDASTICITY:

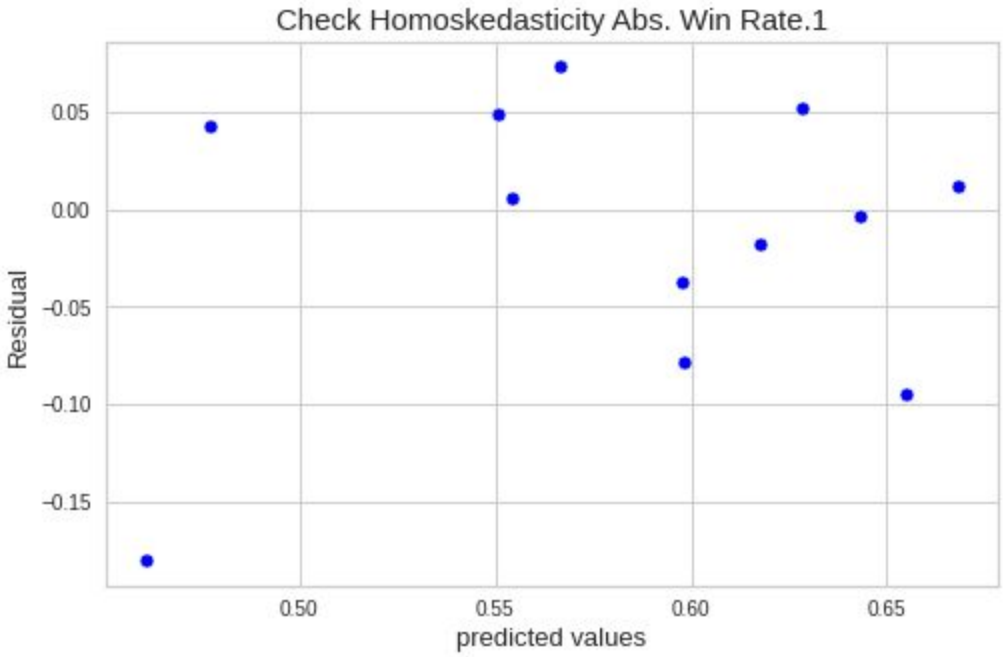
- Homoscedasticity means the residuals are equal across the regression line.
- To check homoscedasticity, we plot the residuals vs predicted values/fitted values.
- If we see any kind of funnel shape, we can say that there is heteroscedasticity

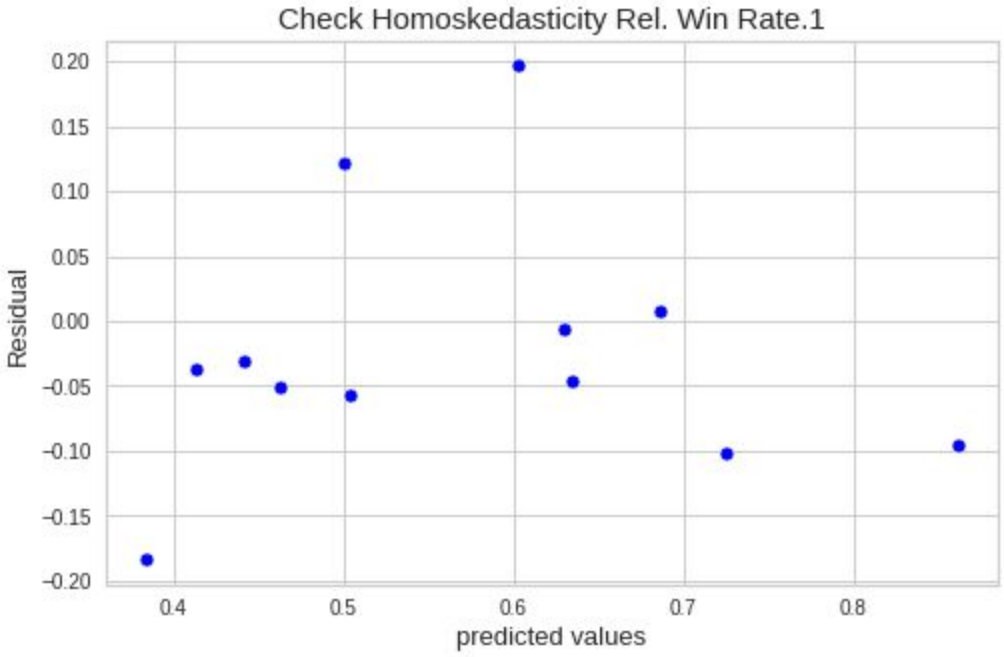

Check Homoskedasticity Annual Return.1
Residual
predicted values
0.050
0.025
0.000
−0.025
−0.050
−0.075
−0.100
−0.125
0.4
0.5
0.6
0.7
0.8
0.9


Check Homoskedasticity Excess Return.1
Residual
predicted values
0.05
0.00
−0.05
−0.10
0.4
0.5
0.6
0.7
0.8


Check Homoskedasticity Systematic Risk.1
Residual
0.10
0.05
0.00
−0.05
−0.10
0.38
0.40
0.42
0.44
0.46
0.48
predicted values


Check Homoskedasticity Total Risk.1
Residual
0.15
0.10
0.05
0.00
−0.05
−0.10
0.300
0.325
0.350
0.375
0.400
0.425
0.450
0.475
0.500
predicted values





z

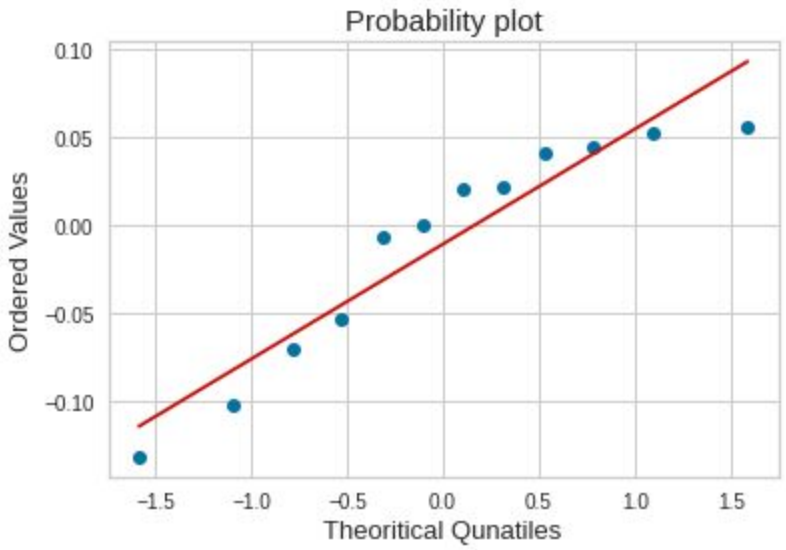
❏ The points are random. which confirms that there is homoscedasticity.

- It means that the variance of Y across all X is the same. - We can conclude that Homoscedasticity condition holds in this case.

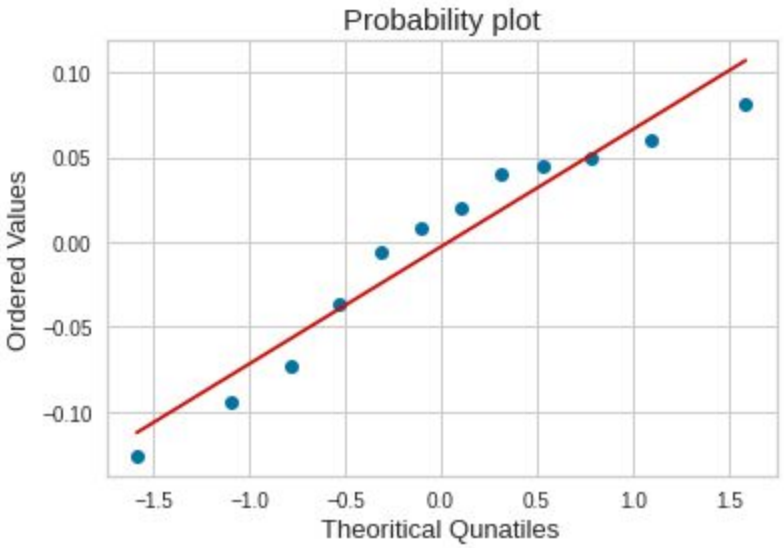
Normality Of Error terms :

- It is a test to check whether the errors follow normal distribution.
- It can be tested by plotting the probability plot or Q-Q plot. It plots against two sets of quantiles, if we see all sets of points lie on a straight line, we can conclude it follows Normal Distribution.

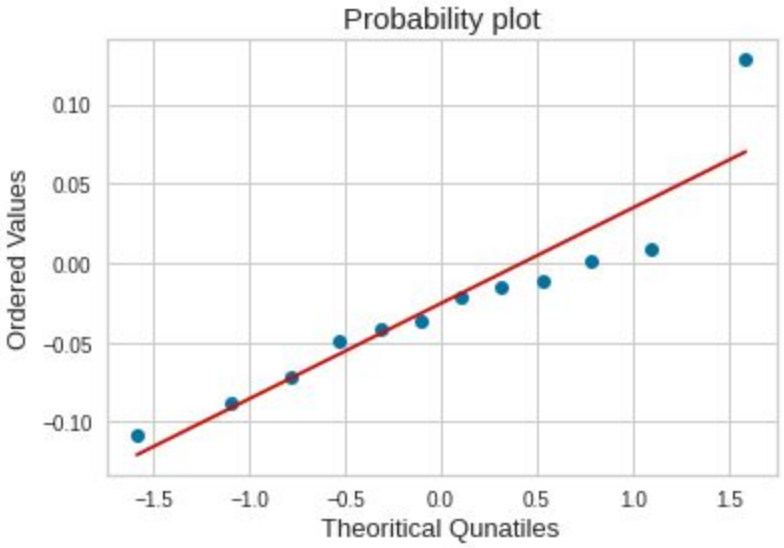
Annual Return



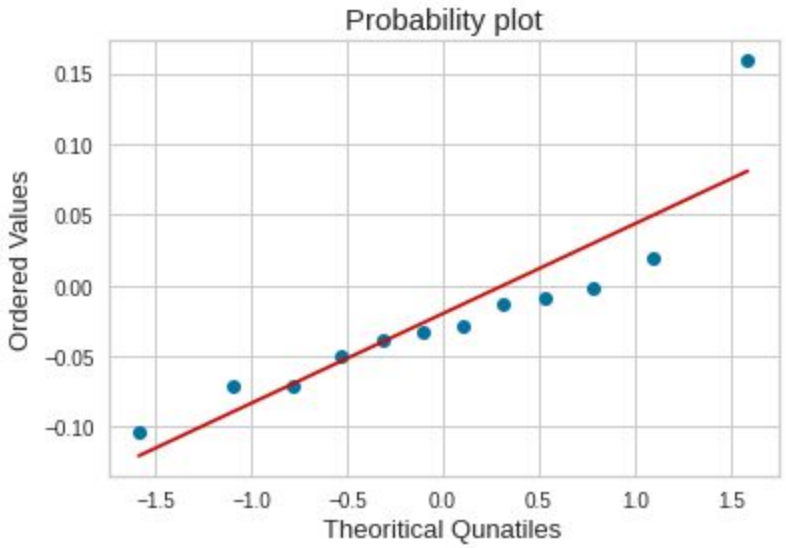
Excess Return



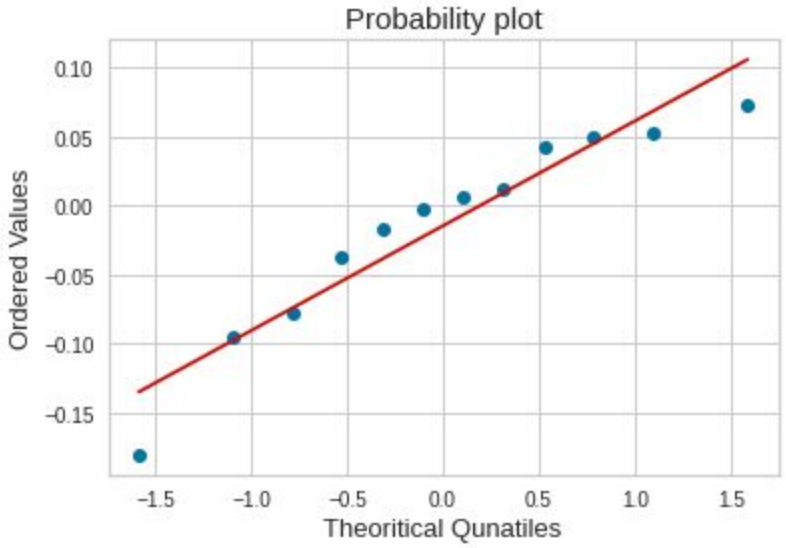
Systematic Risk



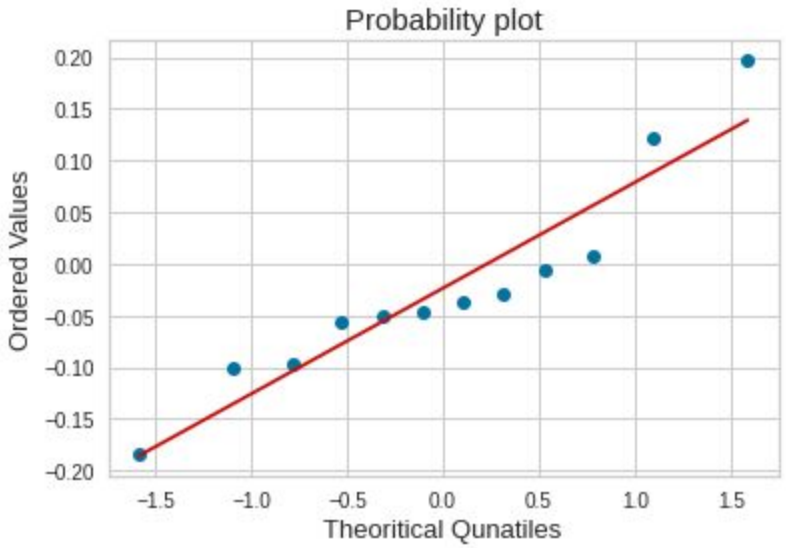
Total Risk



Absolute win Rate



Relative win Rate

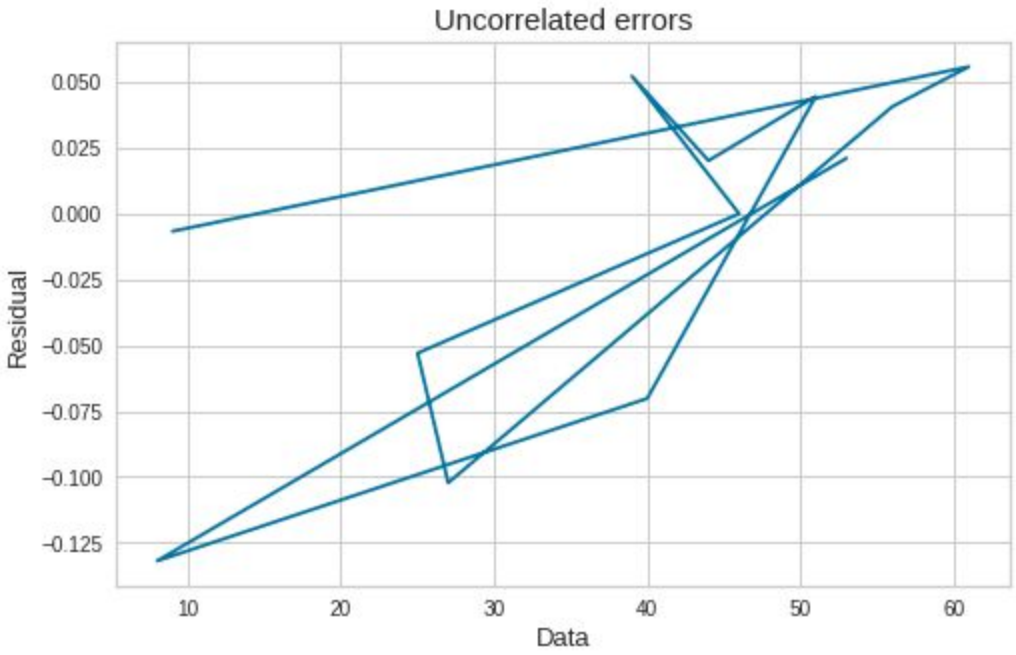


- The above plots are between Quantiles from the Standard Normal Distribution and Ordered responses of residuals.
- If we observe the above plots, all the data points are fitted well in the straight line, so we can conclude that residuals follow Normal distribution.

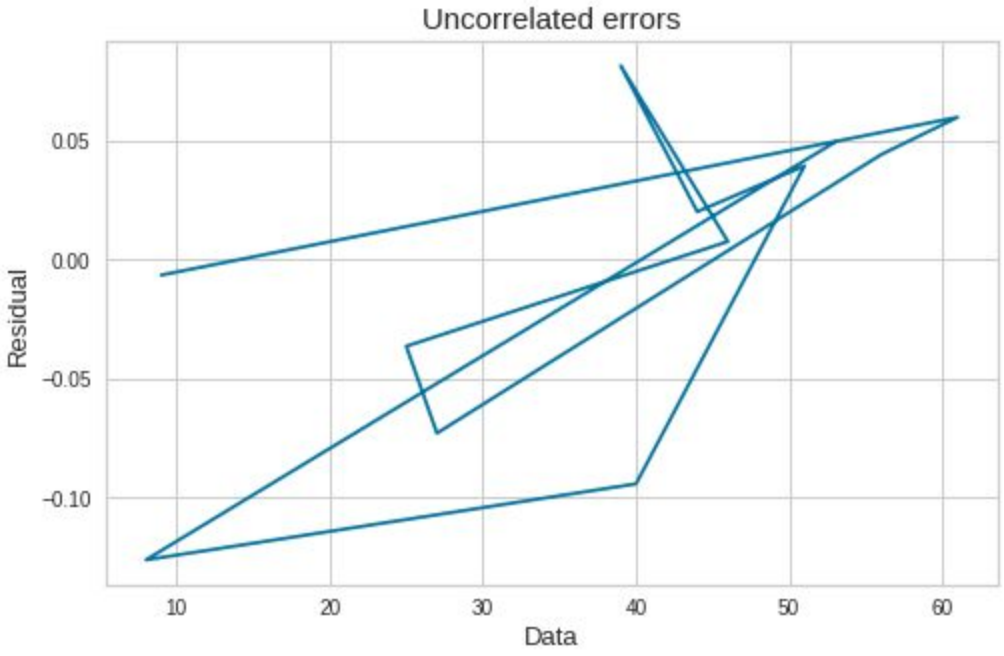
Uncorrelated Errors:

- It is a test to check the linear relationship between the variables.
- Two variables are said to be uncorrelated if the covariance between them is zero. Here, Durbin watson test can be performed to check for the uncorrelation.

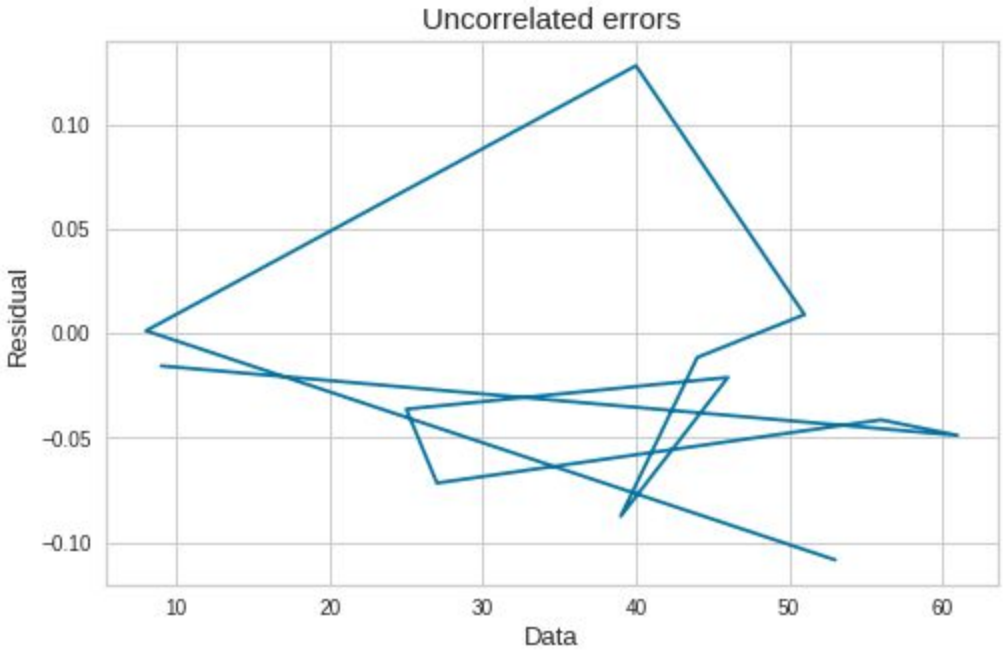
Annual Return



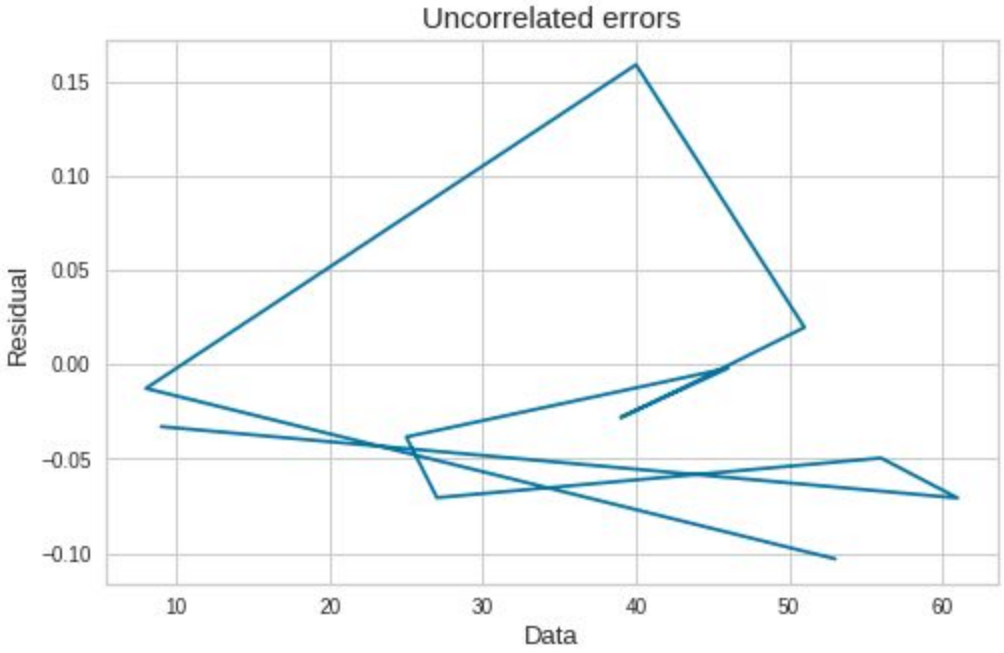
Excess Return



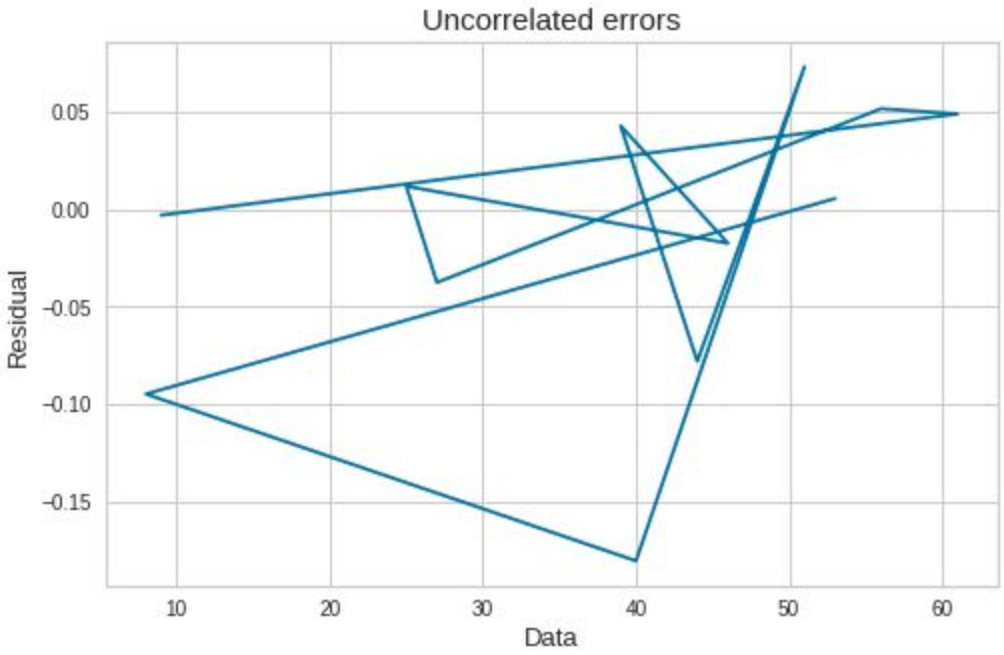
Systematic Risk



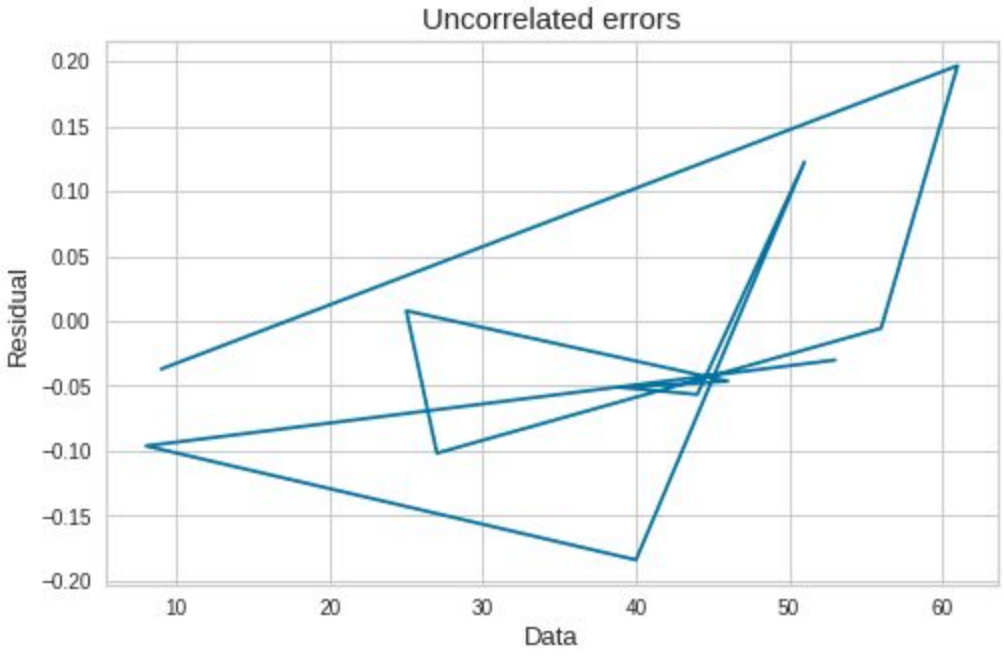
Total Risk



Abs. win Rate



Rel. win Rate



❏ The above plots are between Observation values and Residuals.

- If we observe, there exists correlation/pattern between errors.
- In detail we can also check for autocorrelation by Durbin-Watson test.
- In the Durbin Watson Test there is a factor called DW(values are scaled from 0-4).
    - If DW = 2, then there is no correlation.
    - If DW < 2, then the errors are positively correlated.
    - If DW > 2, then the errors are negatively correlated.
- If we perform Durbin-Watson Test, the values of DW are
    - `1.6215799780769313 -- for Annual Return`
    - `1.6819424442764814 -- for Excess return`
    - `1.1953060659793606 -- for Systematic Risk`
    - `1.215215373547574  --  for Total Risk`
    - `2.2147484787943488 -- for Abs. win rate`
    - `2.1946702508707543 -- for Rel. win rate`
- According to the results obtained from the test, we can say that for the first 4 evaluation parameters errors are positively correlated and remaining are negatively correlated.
- However, this is a point estimate for perfect un- correlation of errors (DW=2). So, we won't get DW as 2 on real data. If it is around 2, then we can conclude that the errors are uncorrelated.