Quantile Regression to Understand Interrelationship amongst Stock Exchanges Globally

MENTORED BY:
Dr. Hemant Kulkarni

Dhanashri Kanitkar (75252019009)
Hrishita Bapuram (75252019004)
Shraddha Kodavade (75252019011)

“Regression Analysis is the Hydrogen Bomb of the Statistics Arsenal”

– Charles Wheelan (*Naked Statistics*)

Key Objectives

To Establish the case for Indian Stock Indices (here, Nifty 50) being affected by Global Stock Indices (by extension, the Exchanges themselves) through statistical exploratory data analysis

To Observe the limitations and explore alternatives to Linear Regression model

To Understand the concept, scope and limitations of Quantile Regression, specifically its use in Financial Analytics

To Design a Quantile Regression Model to establish the relationship between stock exchanges by predicting the net daily change in Nifty 50 value.

To Employ the use of statistical packages and tools including R and Python in performing analysis

Contents

Introduction to Quantile Regression

Concept, Motivation, Principles, Merits and Comparison with Linear Regression

Data Snapshot

EDA, Scatterplots, Correlation Matrix, Assumption Checks, Case for Quantile Regression

Quantile Regression Analysis Results

Model fits, Key insights, Comparison with Linear fit, Limitations and Future Scope

1

Introduction to Quantile Regression

Why should we always be ***mean-minded***?

Regression Analysis

- Regression analysis is a powerful statistical method that allows you to examine the relationship between two or more variables of interest.
- Typically, a regression analysis is done for one of two purposes:
    - To predict the value of the dependent variable for individuals for whom some information concerning the explanatory variables is available
    - To estimate the effect of some explanatory variable on the dependent variable.
- Regression analysis helps an organization to understand what their data points represent and use them accordingly with the help of business analytical techniques in order to do better decision-making.

Linear Regression and its Limitations

- The central objective of regression techniques is to obtain summary of the response variable conditioning on the explanatory variables.
  - Y: Monthly Electricity Usage of family
  - X: Income, House size, number of electronic gadgets, size of family
- Model:

$$\boxed{\begin{array}{l} \textit{Response Variable} = \textit{Model Function} + \textit{Random Error} \\ Y = f(X, \theta) + \epsilon \end{array}}$$

- Linear Regression Model:

$$\boxed{Y = \beta_0 + \beta_1 X + \epsilon}$$

$\beta = (\beta_0, \beta_1)$: Model parameters
$E(\epsilon) = 0,\ V(\epsilon) = \sigma^2$

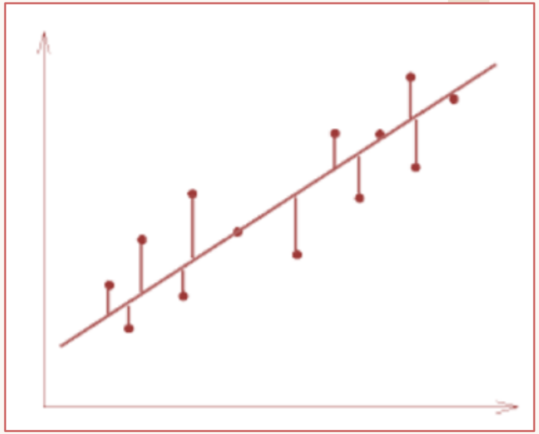
Least Squares Estimation

Objective function:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^P}{\operatorname{argmin}} \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x)^2$$

Predicted value of the response

$$\hat{Y} = \hat{E}(Y \mid X = x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

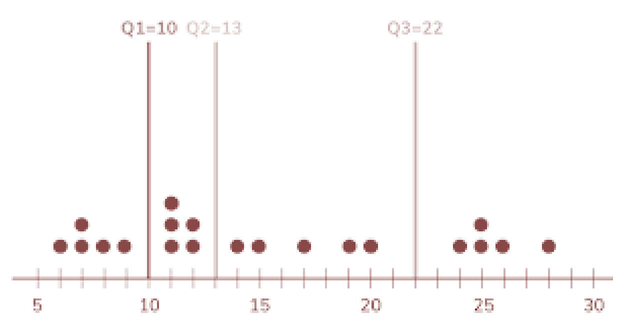


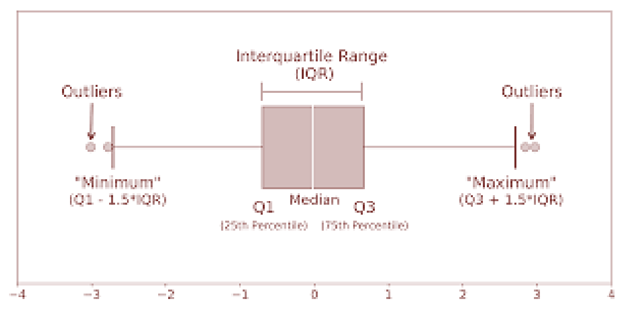
- Gives misleading results when assumptions are violated.
- Not efficient in the presence of outliers.
- Median regression based on least absolute deviation is an alternative robust method to mean regression.
- Can not capture non-central characteristics of the response variable.

Quantile Regression

- Median regression based on least absolute deviation is an alternative robust method to mean regression.
- Quantile: The tau-th quantile denotes the value of random variable bellow which proportion of population is tau: $Q_\tau(Y) = Y_\tau = F^{-1}(\tau)$
- For every given sample, the middle value of such data is called the median, middle quantile, or 50th percentile.

Quantile	Percentile
the point in a data distribution corresponding to data values in distribution divided equally in a corresponding order.	type of quantile that divides a variable into 100 equal places.



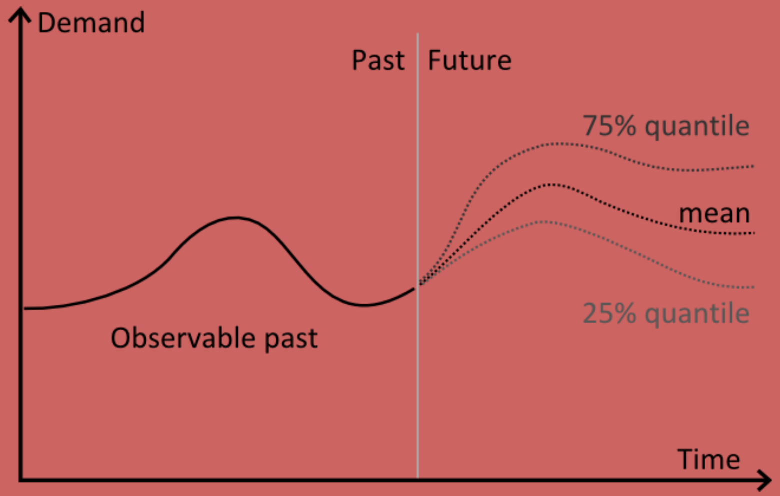


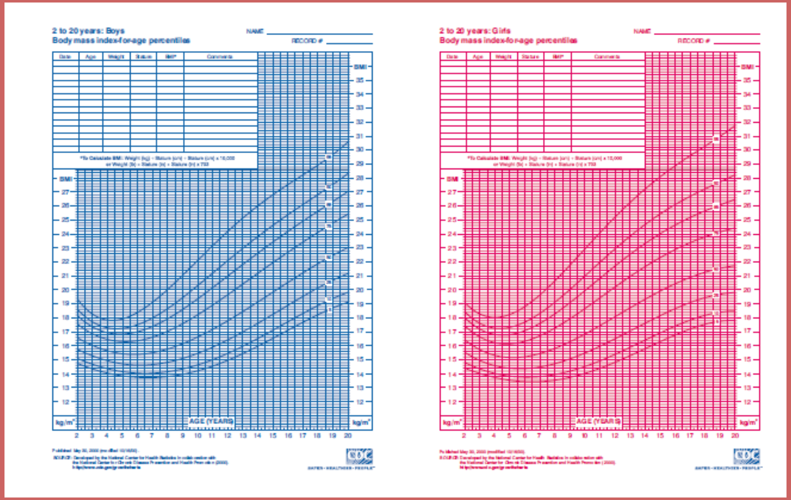
- Quantile regression explains non-central summary statistics of the response variable.

- It is a method which models quantiles instead of mean for a given set of covariates.

- It performs better in the presence of outliers and hence more robust compared to the mean regression.

- Quantile regression models response at different levels of quantiles and thus gives the information on the shape of the response distribution.





When to use Quantile Regression?

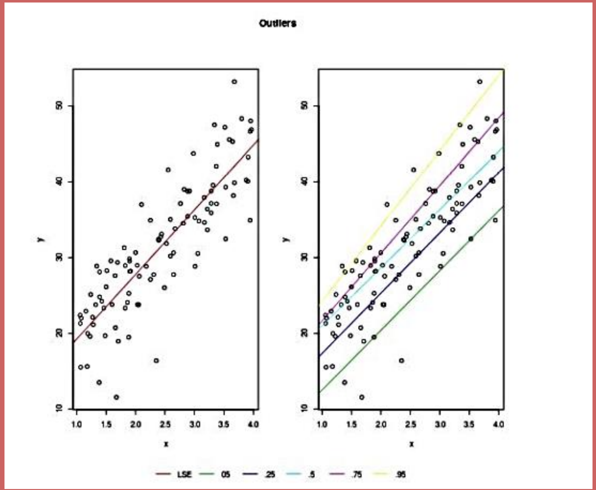
Data isn't always linear in nature, and could be distributed in a non-linear manner. Some variables may not behave normally and have an efficient linear relationship with predictable variables.

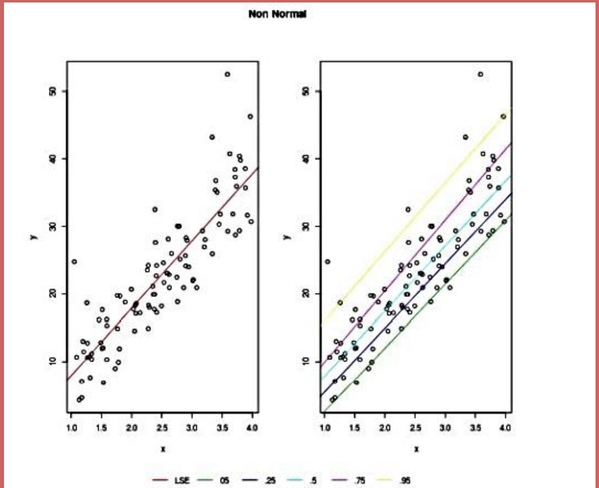
1. When the linear regression assumptions are not meeting up with each other.
2. When you have outliers in your data sample.
3. When the error variables make your outcome variable spike up (Heteroscedasticity)

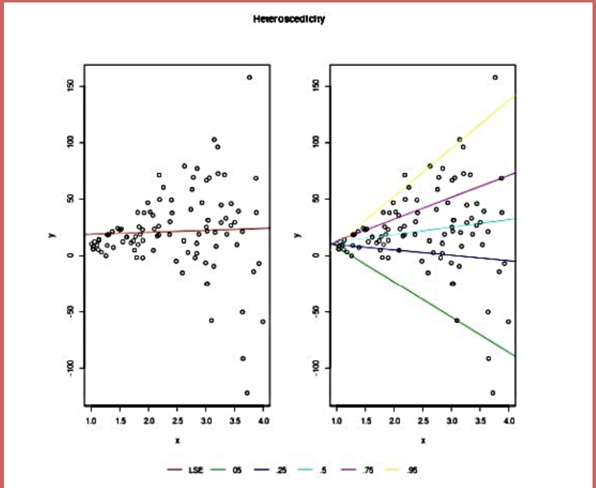
Outliers

Non-Normal

Heteroscedasticity







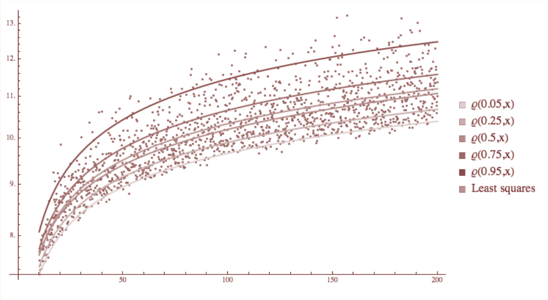
Advantages of Quantile Regression

- Quantile regression model not only gives the complete distributional picture, it has several other advantages over the linear regression method.

- The objective function of quantile regression makes it more robust in the presence of outliers.

- Most importantly, no distributional assumption is needed for the response variable in quantile regression.

- Quantile regression allows the analyst to drop the assumption that <u>**variables operate the same at the upper tails of the distribution as at the mean;**</u> and identifies the factors that are <u>**important factors**</u> of variables.



Model and Definition – Quantile Regression

- The quantile regression model equation for the τth quantile is:

  $Q_\tau(y_i) = \beta_0(\tau) + \beta_1(\tau)x_{i1} + \cdots + \beta_p(\tau)x_{ip} + \epsilon$ $\quad i = 1, \ldots, n$

- This means that instead of being constants, the beta coefficients are now **functions with a dependency on the quantile**.

OBJECTIVE FUNCTION:

For a random variable Y with observed values, $\mathbf{Y} = (y_1, y_2, \cdots, y_n)$

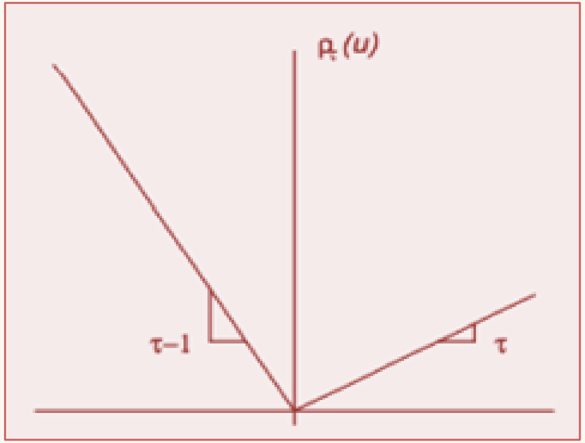
$$Mean = \underset{\xi}{\text{argmin}} \sum (y_i - \xi)^2$$

$$Median = \underset{\xi}{\text{argmin}} \sum |y_i - \xi|$$

$$Q_\tau(Y) = \underset{\xi}{\text{argmin}} \sum \rho_\tau(y_i - \xi)$$

where

$$\rho_\tau(u) = \begin{cases} (1-\tau)\,|u| & \text{if } u \leq 0 \\ \tau u & \text{if } u > 0 \end{cases}$$



Linear Quantile Regression

- Model (Koenker and Basset, 1978) for (xi, yi) is:

$$Q_\tau(y \mid X = x) = \beta_{0\tau} + \beta_{1\tau} x + \epsilon \qquad Q_\tau(\epsilon) = 0$$

- Parameters are estimated using Simplex Algorithm

$$\widehat{\beta_\tau} = \underset{\beta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{i=1}^{N} \rho_\tau(y_i - \beta_{0\tau} - \beta_{1\tau} x_i)$$

- The Fitted Conditional Quantile is:

$$\widehat{Q_\tau}(Y \mid X) = \widehat{\beta_{0\tau}} + \widehat{\beta_{1\tau}} X$$

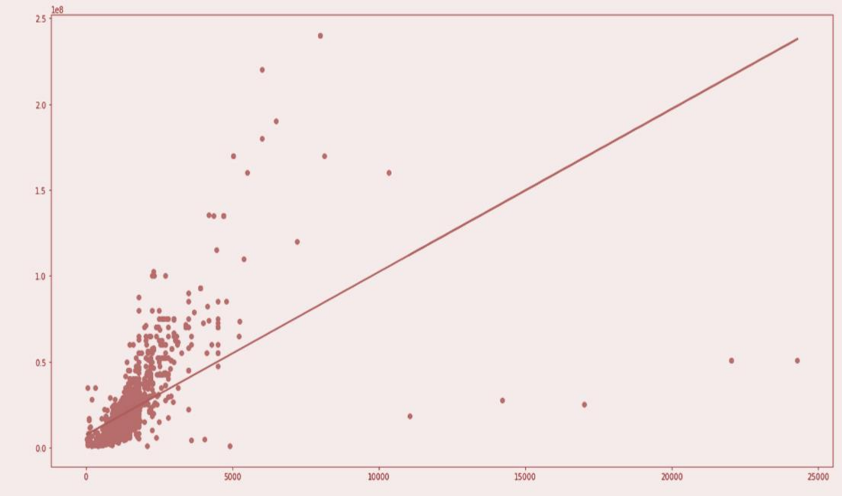
Linear Regression Vs. Quantile Regression

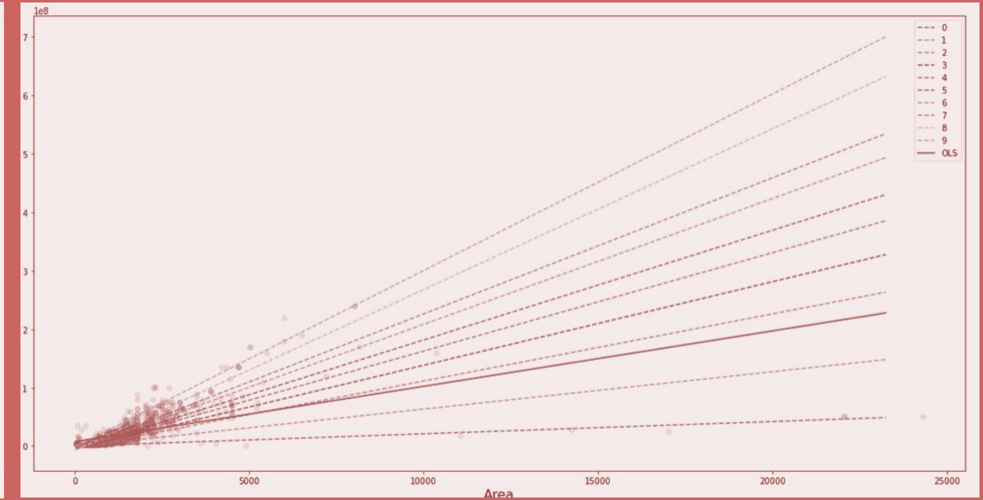
- When performing regression analysis, it isn’t enough to come up with a numerical prediction to a problem, we also need to express how confident we are in that prediction.

- Apart from this, their other loopholes in analyzing data across a specific distribution having a fixed mean across the whole data.

Linear Regression	Quantile Regression
Predicts the conditional mean $E(Y\|X)$	Predicts conditional quantiles $Q_\tau(Y\|X)$
Applies when n is small	Needs sufficient data
Often assumes normality	Is distribution agnostic
Does not preserve $E(Y\|X)$ under transformation	Preserves $Q_\tau(Y\|X)$ under transformation
Is sensitive to outliers	Is robust to response outliers
Is computationally inexpensive	Is computationally intensive

Linear Regression Vs. Quantile Regression



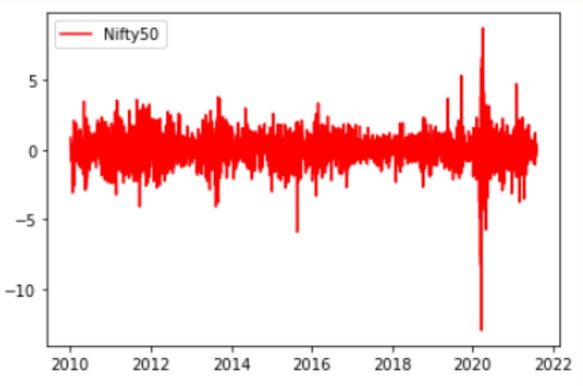


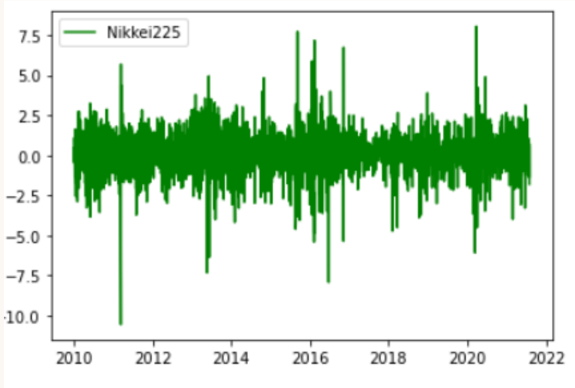
Quantile regression is capable of modeling the entire conditional distribution; this is essential for applications such as ranking the performance of students on standardized exams

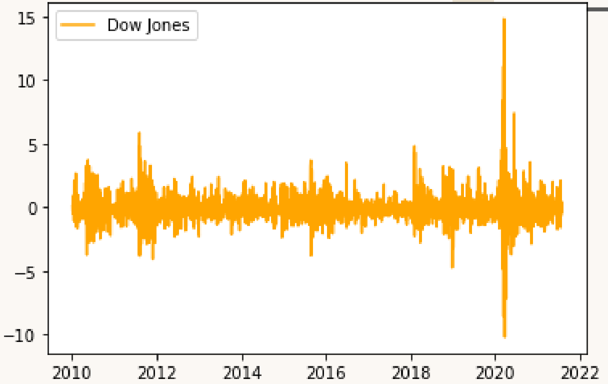
2

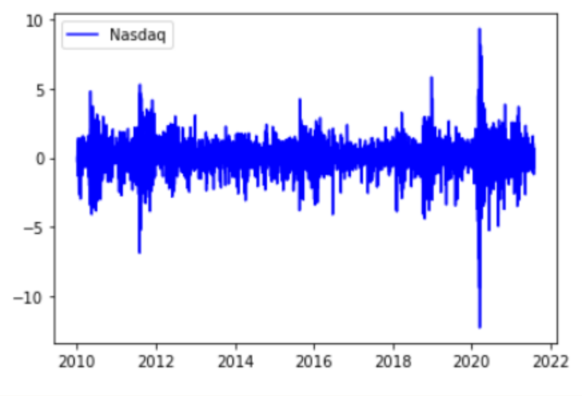
Data Snap Shot & Linear Fit

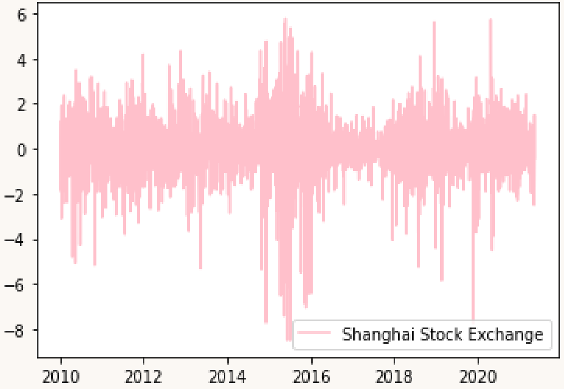
The Global Indices

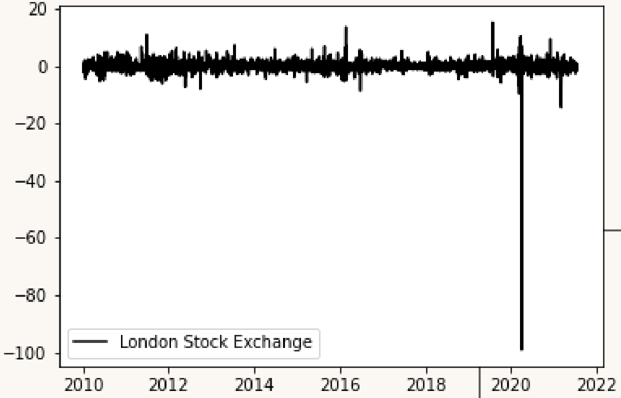






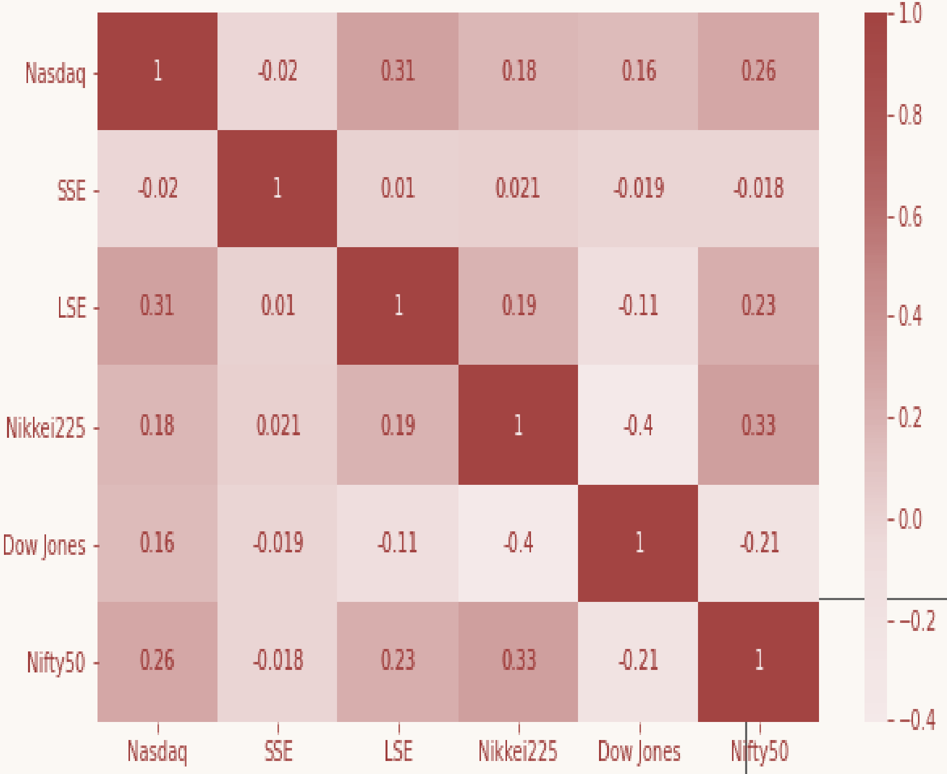






The Unclear Relationship

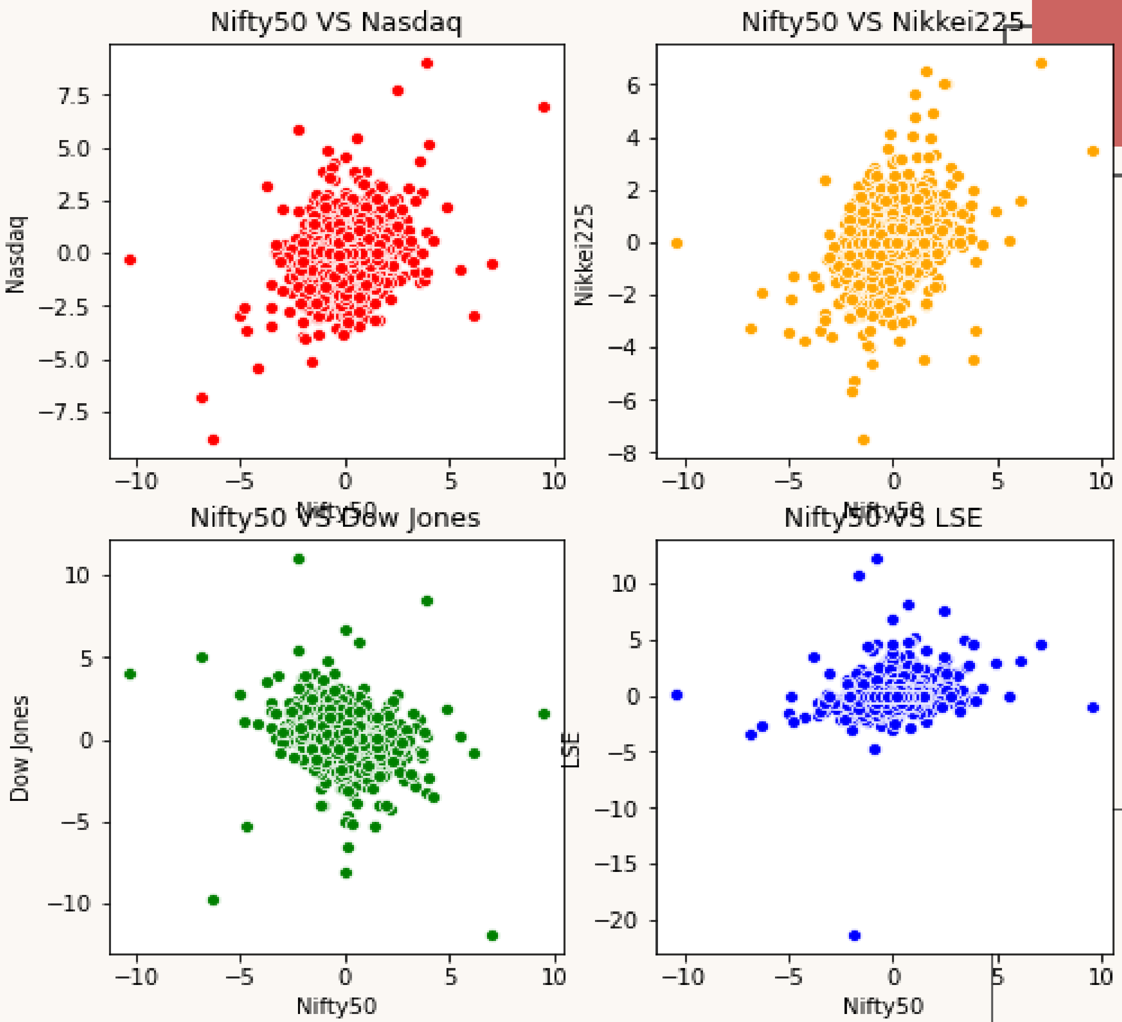
- The image represents the correlation between the dependent and independent variables.
- We can observe that Dow Jones has a negative correlation of (-0.21) with Nifty50.
- Whereas Nikkei225 and LSE have a positive correlation of 0.33 and 0.23 with Nifty50, respectively.
- Nasdaq have a positive relation of 0.26 with Nifty50.
- SSE has a negligible correlation with Nifty50



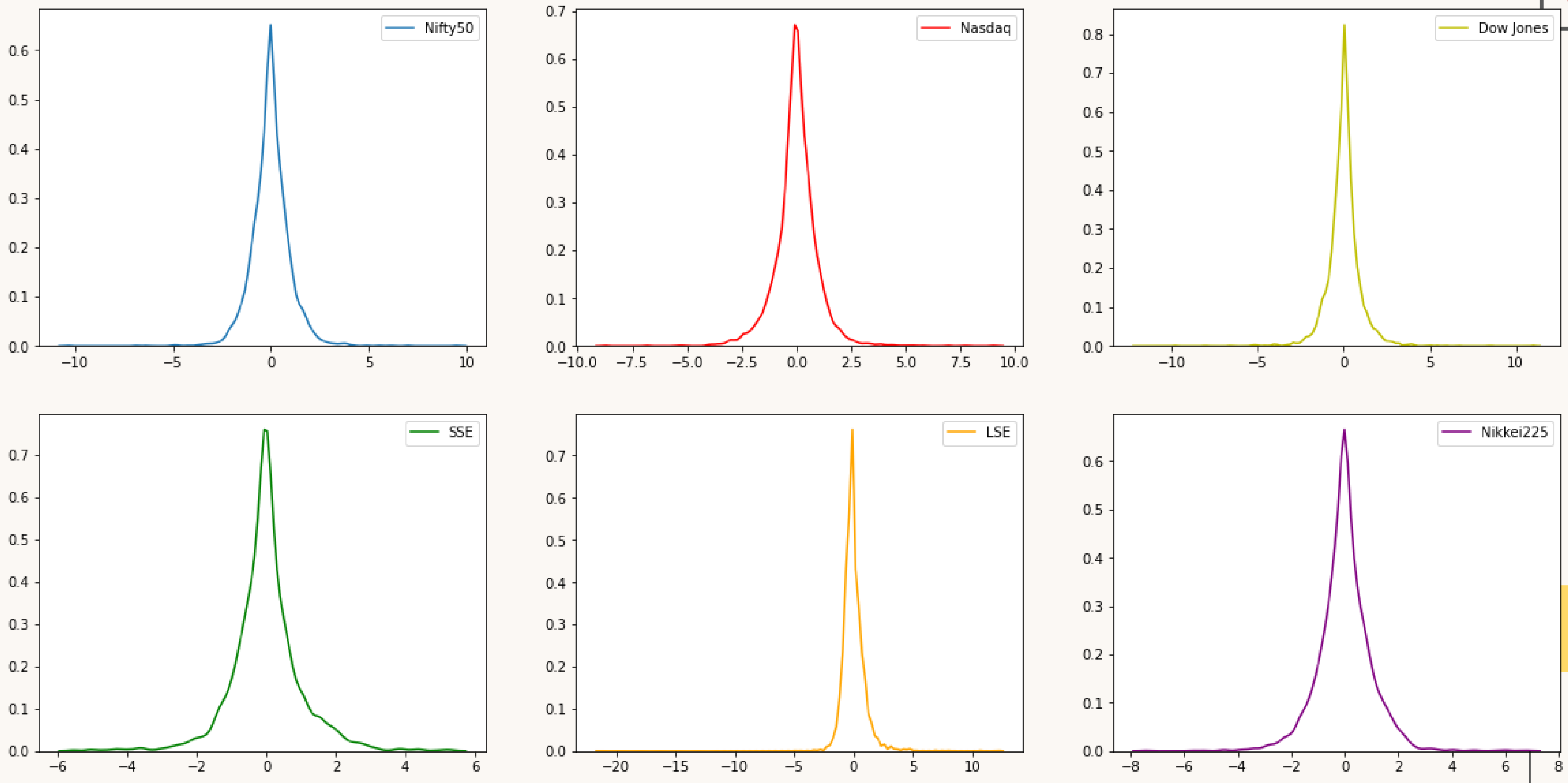
Scatter-ness

- The image is a scatterplot of Nifty50 which is the dependent variable with the independent variables like Dow Jones, Nasdaq, LSE and Nikkei

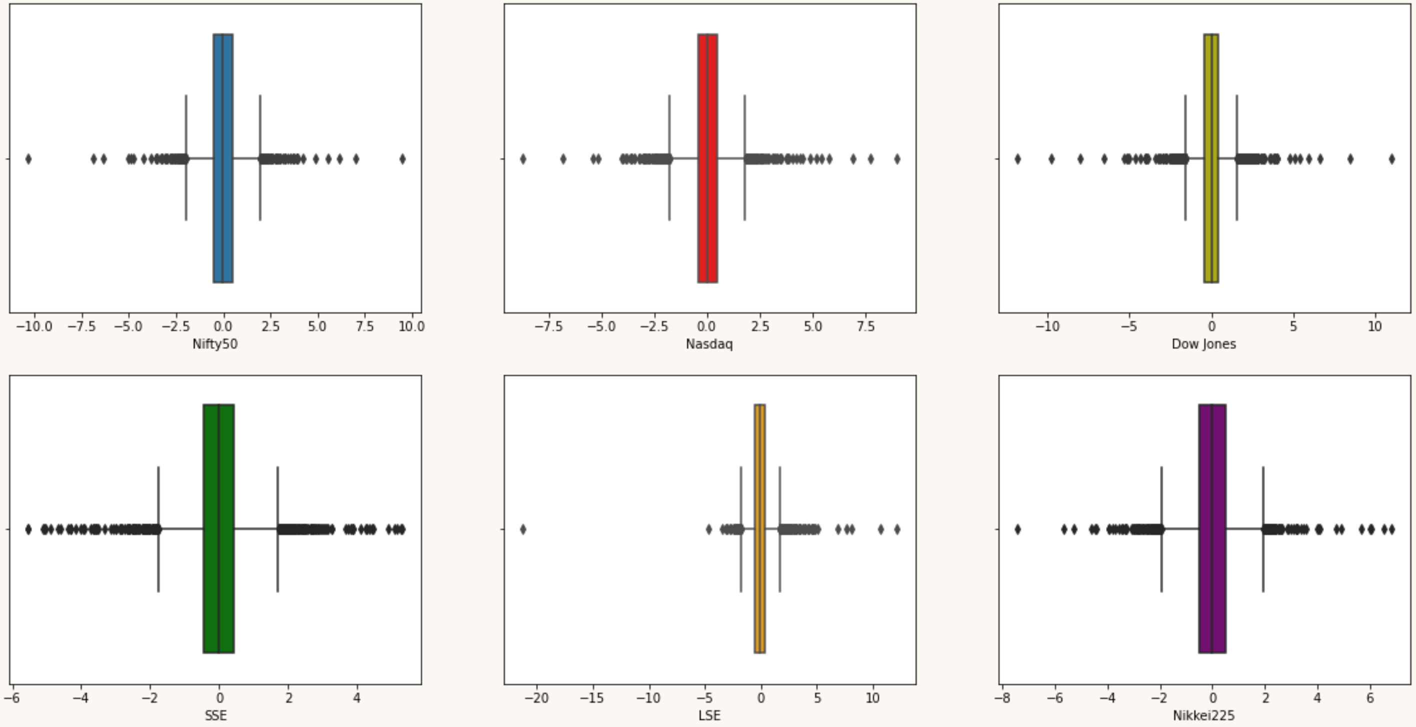
- We can see that the data is not following a linear relationship.



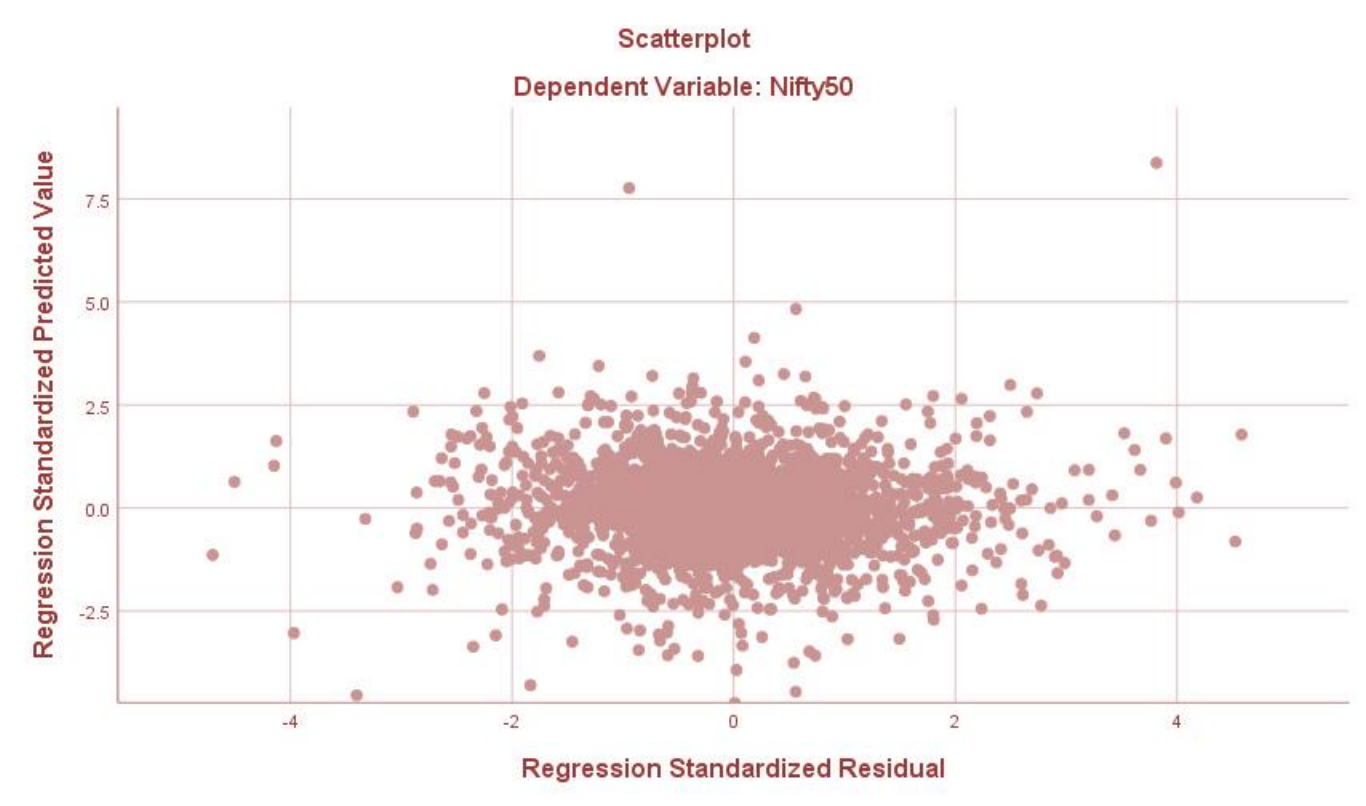
Distribution of the Data



Boxplot for each index



Homoscedasticity Check



Fitting of Linear Model

```
Call:
lm(formula = Nifty50 ~ Nasdaq + Dow.Jones + Nikkei225 + SSE +
    LSE, data = stocks)

Coefficients:
(Intercept)       Nasdaq    Dow.Jones    Nikkei225          SSE          LSE
  2.761e-11    2.201e-01   -1.556e-01    2.040e-01   -2.204e-02    1.071e-01
```

```
Residuals:
    Min      1Q  Median      3Q     Max 
-9.6718 -0.4794 -0.0217  0.4685  7.6193 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  2.761e-11  1.644e-02   0.000     1.00    
Nasdaq       2.201e-01  1.812e-02  12.149  < 2e-16 ***
Dow.Jones   -1.556e-01  1.869e-02  -8.324  < 2e-16 ***
Nikkei225    2.040e-01  1.869e-02  10.913  < 2e-16 ***
SSE         -2.204e-02  1.645e-02  -1.340     0.18    
LSE          1.071e-01  1.759e-02   6.087 1.29e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9047 on 3023 degrees of freedom
Multiple R-squared:  0.1831,	Adjusted R-squared:  0.1817 
F-statistic: 135.5 on 5 and 3023 DF,  p-value: < 2.2e-16
```

- The Adjusted R-squared value is **0.1817**.
- This shows that the linear model is **not a good fit** for the data.

3

Quantile Regression Fit

Results and Analysis

Fitting of Quantile Regression Model

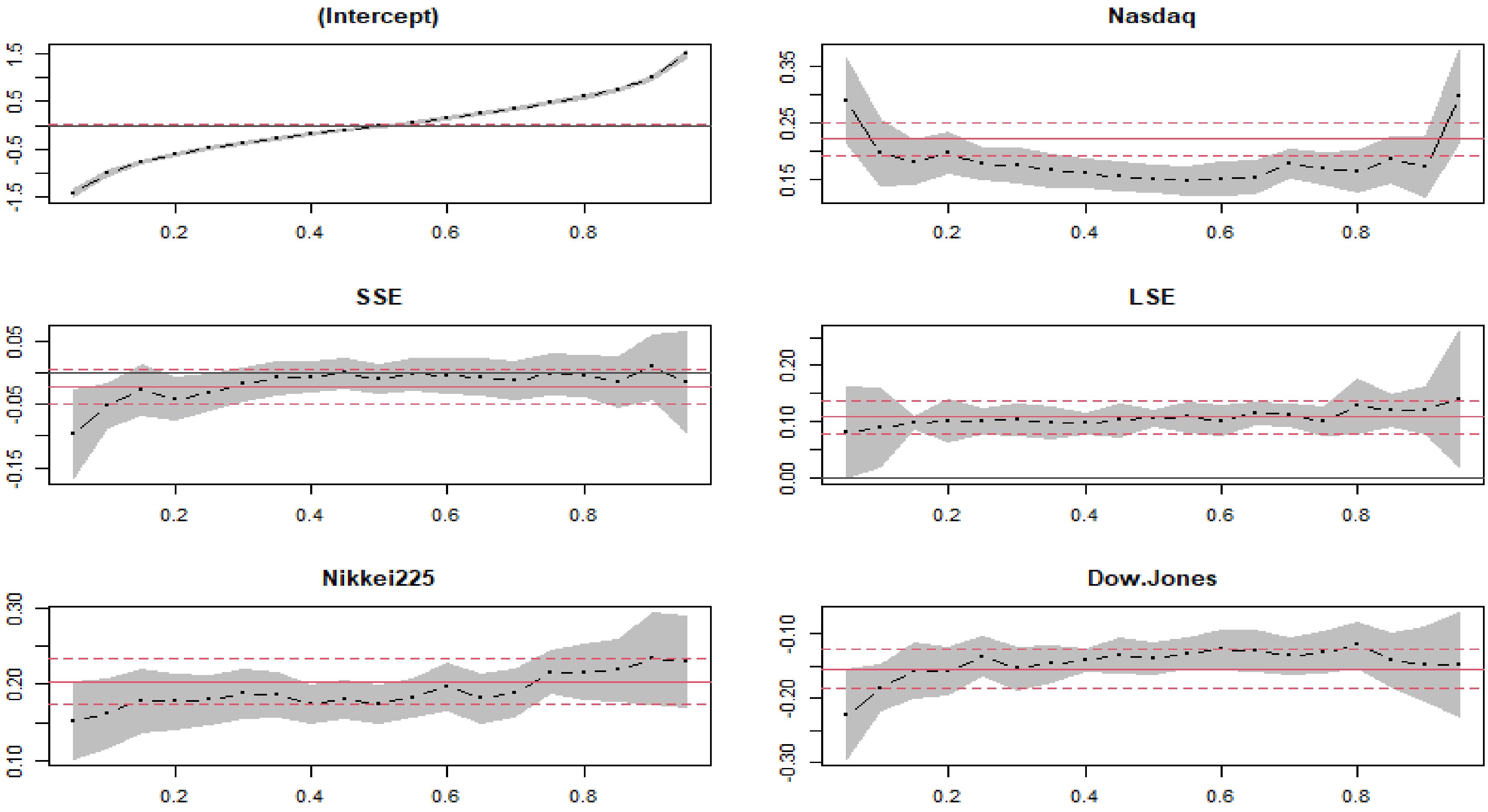
```
> qs <- 1:19/20
> Quan_fit <- rq( Nifty50 ~ Nasdaq + SSE + LSE + Nikkei225 + Dow.Jones, data = stocks, tau = qs)
> Quan_fit
Call:
rq(formula = Nifty50 ~ Nasdaq + SSE + LSE + Nikkei225 + Dow.Jones,
    tau = qs, data = stocks)

Coefficients:
              tau= 0.05   tau= 0.10   tau= 0.15   tau= 0.20  tau= 0.25   tau= 0.30    tau= 0.35    tau= 0.40    tau= 0.45
(Intercept) -1.40860943 -0.99693848 -0.75881340 -0.60098163 -0.4735619 -0.36301910 -0.269823506 -0.167777847 -0.093866507
Nasdaq       0.28792625  0.19551382  0.17933694  0.19536678  0.1755172  0.17338261  0.165283573  0.159491343  0.153657132
SSE         -0.09698647 -0.05318840 -0.02732189 -0.04221468 -0.0317753 -0.01846625 -0.008695772 -0.007082083 -0.001106207
LSE          0.08063738  0.08880332  0.09646903  0.10093982  0.1007371  0.10265213  0.096221726  0.096518871  0.101266944
Nikkei225    0.15195666  0.16139140  0.17834564  0.17770277  0.1791364  0.18751139  0.187200073  0.174444190  0.180993331
Dow.Jones   -0.22647888 -0.18439555 -0.15820452 -0.15878728 -0.1356231 -0.15425497 -0.147220959 -0.140874582 -0.134334552
              tau= 0.50   tau= 0.55    tau= 0.60    tau= 0.65   tau= 0.70    tau= 0.75    tau= 0.80   tau= 0.85    tau= 0.90
(Intercept) -0.02344914  0.05400171  0.144477643  0.243037618  0.34710049  0.473871482  0.596708597  0.74845343  0.992449627
Nasdaq       0.14986383  0.14653485  0.150260570  0.152463002  0.17705551  0.168427419  0.162457518  0.18409921  0.171479668
SSE         -0.01051515 -0.00274896 -0.005457867 -0.007111605 -0.01269808 -0.001889499 -0.004824785 -0.01456066  0.009319375
LSE          0.10568199  0.10697653  0.100831909  0.114245718  0.11178063  0.099723394  0.127655842  0.11875792  0.120287952
Nikkei225    0.17409789  0.18291447  0.197509596  0.181955127  0.18868797  0.215664707  0.216337527  0.21896118  0.233413610
Dow.Jones   -0.13840577 -0.13126155 -0.125132044 -0.126538780 -0.13485643 -0.129216534 -0.118038769 -0.14119880 -0.148011726
              tau= 0.95
(Intercept)  1.48503013
Nasdaq       0.29525166
SSE         -0.01502365
LSE          0.13947365
Nikkei225    0.23067803
Dow.Jones   -0.14860381

Degrees of freedom: 3029 total; 3023 residual
>
```

Parameter Values

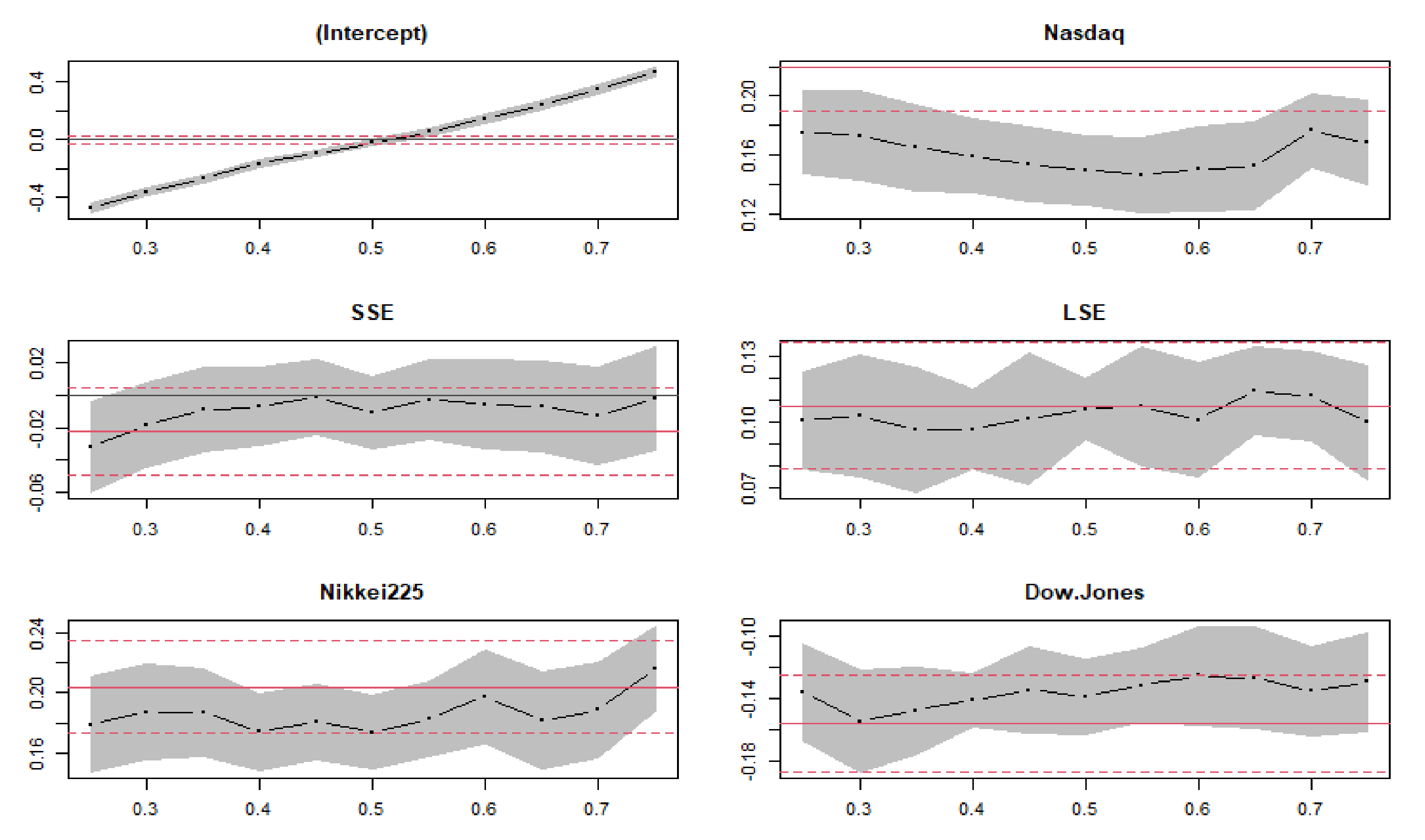
Tau Values	Intercept	Nasdaq	SSE	LSE	Nikkei225	DJI
Linear	0.00000	0.22000	-0.02000	0.11000	0.20000	-0.16000
0.05	-1.40900	0.28800	-0.09700	0.08100	0.15200	-0.22600
0.10	-0.99700	0.19600	-0.05300	0.08900	0.16100	-0.18400
0.15	-0.75881	0.17934	-0.02732	0.09647	0.17835	-0.15820
0.20	-0.60098	0.19537	-0.04221	0.10094	0.17770	-0.15879
0.25	-0.47356	0.17552	-0.03178	0.10074	0.17914	-0.13562
0.30	-0.36302	0.17338	-0.01847	0.10265	0.18751	-0.15425
0.35	-0.26982	0.16528	-0.00870	0.09622	0.18720	-0.14722
0.40	-0.16778	0.15949	-0.00708	0.09652	0.17444	-0.14087
0.45	-0.09387	0.15366	-0.00111	0.10127	0.18099	-0.13433
0.50	-0.02345	0.14986	-0.01052	0.10568	0.17410	-0.13841
0.55	0.05400	0.14653	-0.00275	0.10698	0.18291	-0.13126
0.60	0.14448	0.15026	-0.00546	0.10083	0.19751	-0.12513
0.65	0.24304	0.15246	-0.00711	0.11425	0.18196	-0.12654
0.70	0.3471	0.17706	-0.01270	0.11178	0.18869	-0.13486
0.75	0.47387	0.16843	-0.00189	0.09972	0.21566	-0.12922
0.80	0.59671	0.16246	-0.00482	0.12766	0.21634	-0.11804
0.85	0.74845	0.18410	-0.01456	0.11876	0.21896	-0.14120
0.90	0.99245	0.17148	0.00932	0.12029	0.23341	-0.14801
0.95	1.48503	0.29525	-0.01502	0.13947	0.23068	-0.14860

Tau Values	Nasdaq	SSE	LSE	Nikkei225	Dow Jones
0.05	0.00000	0.02055	0.10750	0.00000	0.00000
0.10	0.00000	0.01147	0.03652	0.00000	0.00000
0.15	0.00000	0.24971	0.00000	0.00000	0.00000
0.20	0.00000	0.03700	0.00001	0.00000	0.00000
0.25	0.00000	0.06487	0.00000	0.00000	0.00000
0.30	0.00000	0.24757	0.00000	0.00000	0.00000
0.35	0.00000	0.57808	0.00000	0.00000	0.00000
0.40	0.00000	0.62753	0.00000	0.00000	0.00000
0.45	0.00000	0.93643	0.00000	0.00000	0.00000
0.50	0.00000	0.43655	0.00000	0.00000	0.00000
0.55	0.00000	0.85391	0.00000	0.00000	0.00000
0.60	0.00000	0.74130	0.00000	0.00000	0.00000
0.65	0.00000	0.67611	0.00000	0.00000	0.00000
0.70	0.00000	0.48437	0.00000	0.00000	0.00000
0.75	0.00000	0.92247	0.00000	0.00000	0.00000
0.80	0.00000	0.80769	0.00000	0.00000	0.00000
0.85	0.00000	0.54197	0.00000	0.00000	0.00000
0.90	0.00000	0.76188	0.00000	0.00000	0.00003
0.95	0.00000	0.75476	0.05784	0.00000	0.00233

P – values

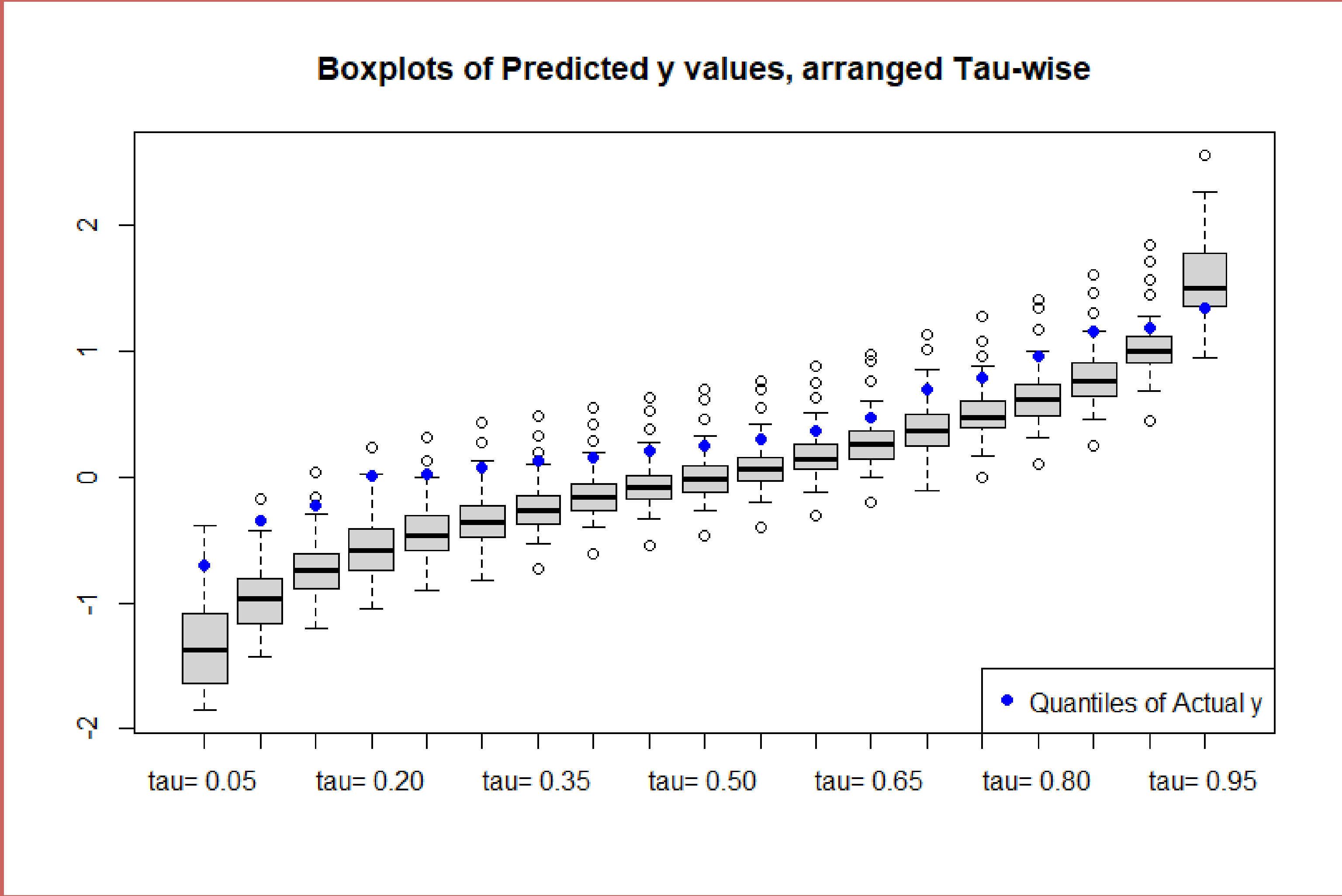

(Intercept)
Nasdaq
SSE
LSE
Nikkei225
Dow.Jones
0.2
0.4
0.6
0.8

Mid-50% Analysis



Key insights

- Nasdaq tends to have a larger impact on net change in Nifty 50 at the extreme 5% of the changes.
- Overall, Shanghai seems to be irrelevant in its impact on Nifty 50.
- Nasdaq and DJI Paradox
- LSE seems consistent and fits well within the linear model’s fit in it’s impact on Nifty 50
- The middle 50% of the net changes seem consistent around the linear model fit, with the exception of Nasdaq
- In general, stock exchanges fluctuations have an impact on the bottom 10% and top 10% of changes in Nifty 50



Box Plot of Predicted Values using QR

Limitations

Developing the mathematical intuition for a more complex set of data might lead to ambiguity in interrelation between the variables.

Not always a combative measure to when a linear model fails.

It involves very lengthy and complicated procedure of calculations and analysis.

Lack of comprehension and jargon for a non - mathematical background client.

Future Scope

- Extending the project to a machine learning dimension by finding optimal higher degree fitting curves.
- Extending the same into Bayesian statistics for more in depth understanding.
- Parallelly studying other research papers, and increasing the adaptability of the project’s problem statement with other statistical methods.
- Producing a research paper to contribute to the financial statistical domain.
- Sharing the knowledge to various fintech start ups to increase awareness of the intersection of finance, data and statistics.

References

- Comparative analysis of Indian stock market. (n.d.). Retrieved November 12, 2021, from https://www.greatlakes.edu.in/pdf/DebijanMukherjee.pdf.
- Federal Reserve Bank of Boston. (2002, November 12). A quantile regression analysis of the cross section of stock market returns. Federal Reserve Bank of Boston. Retrieved November 12, 2021, from https://www.bostonfed.org/publications/research-department-working-paper/2002/a-quantile-regression-analysis-of-the-cross-section-of-stock-market-returns.aspx.
- © 2017 JETIR February 2017, volume 4 - researchgate.net. (n.d.). Retrieved November 12, 2021, from https://www.researchgate.net/profile/Manjula-Ka/publication/344852865_A_STUDY_OF_RELATIONSHIP_BETWEEN_INDIAN_AND_ASIAN_STOCK_MARKETS_USING_MACHINE_LEARNING_TECHNIQUES/links/5f93cadc299bf1b53e4079aa/A-STUDY-OF-RELATIONSHIP-BETWEEN-INDIAN-AND-ASIAN-STOCK-MARKETS-USING-MACHINE-LEARNING-TECHNIQUES.pdf.
- Mensi, W., Hammoudeh, S., Reboredo, J. C., & Nguyen, D. K. (2014, April 13). Do global factors impact BRICS stock markets? A quantile regression approach. Emerging Markets Review. Retrieved November 12, 2021, from https://www.sciencedirect.com/science/article/abs/pii/S156601411400017X.
- Koenker, R., & Hallock, K. F. (n.d.). Quantile regression. Journal of Economic Perspectives. Retrieved November 12, 2021, from https://www.aeaweb.org/articles?id=10.1257%2Fjep.15.4.143.

Acknowledgements

We would like to acknowledge and express our gratitude to our Mentor, Dr. Hemant Kulkarni sir, for helping us structure, format and understand the concepts employed in this project.

We would like to thank NMIMS Deemed To Be University, Navi Mumbai, for providing us with this opportunity, and provide with us with the necessary tools and resources.

Lastly, we would express our gratitude to the various sources, academicians and students from other universities who contributed and helped us build this project.

Our Team

Dhanashri Kanitkar

Hrishita Bapuram

Shraddha Kodavade

Thanks!