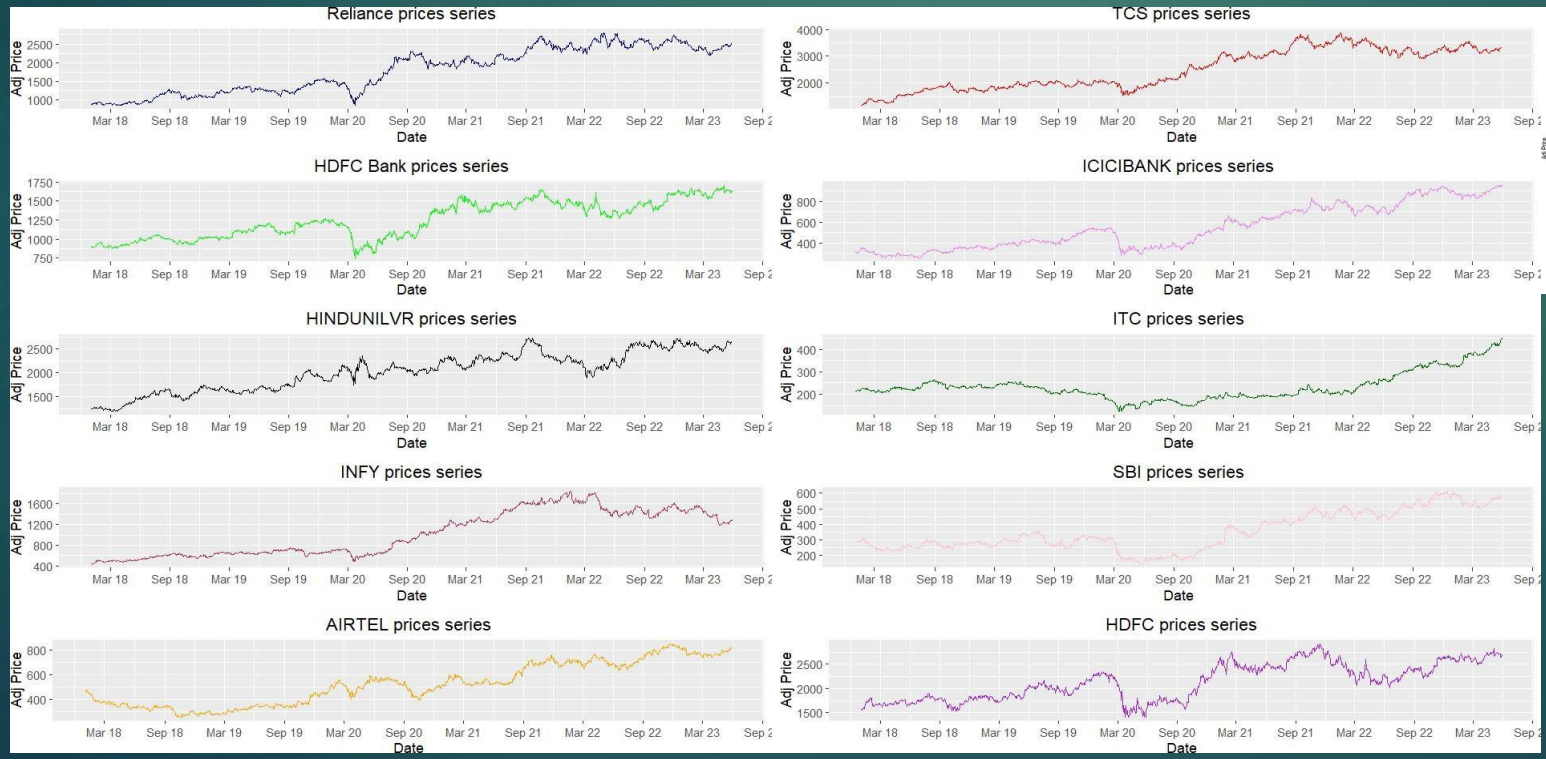
Data

- **Adjusted close values** utilized (adjusted for splits and dividends)
- **Time period**: 02-Jan-18 to 31-May-23
- Nifty50 Data
- Historic data for the below scrips
  - Reliance Industries
  - Tata Consultancy Services
  - HDFC Bank
  - ICICI Bank
  - Hindustan Unilever
  - ITC
  - Infosys
  - State Bank of India
  - Bharti Airtel
  - Housing Development Finance Corporation (HDFC)

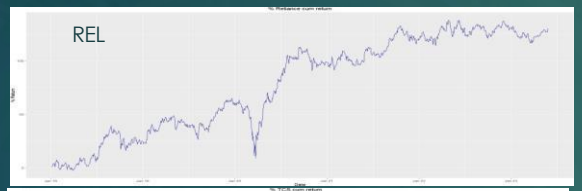
Data Pre-Processing

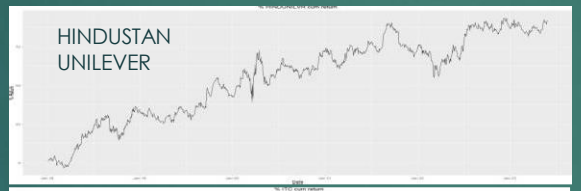
- NA value checks
- Filling missing values for Nifty50 with previous adjusted close value (3 missing values)
- Trend analysis

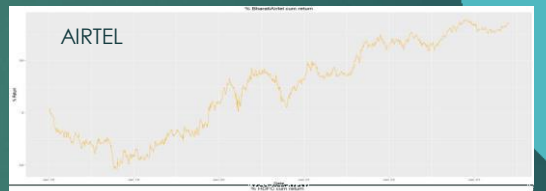


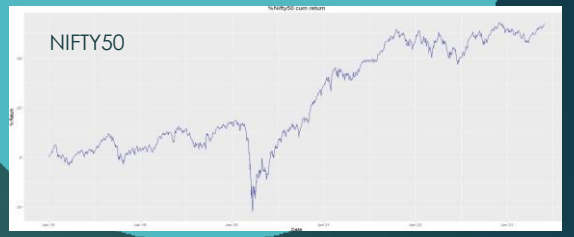
Returns Analysis

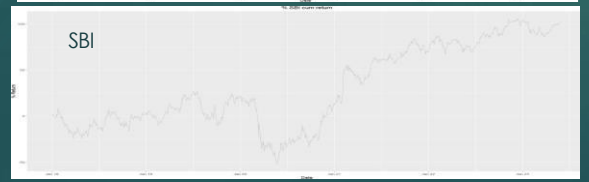
- Return calculated between start and the subsequent day for both scrip and Nifty50 index
- Absolute as well as cumulative returns exhibited a common pattern
- Considering equal weighted portfolio i.e. 10% contribution from each of the scrips, the portfolio returns exhibited a similar pattern





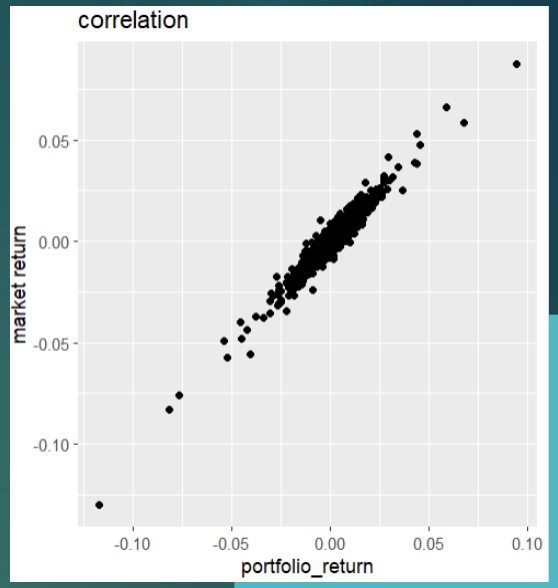






Regression

- Predictor: NIFTY50 index
- Estimated: Portfolio return
- Type of model: Linear
- Correlation: 0.96 => Strong correlation
- Augmented Dickey-Fuller score of -10.3 and -9.8 => Stationary series
- Linear model has high significance for the predictor and residuals follow normal distribution

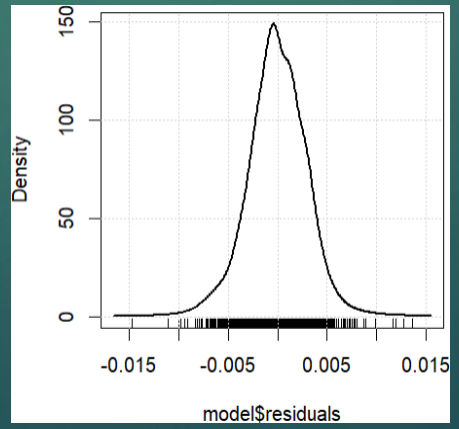



```
Call:
lm(formula = portfolio.returns ~ Nifty_adj, data = train)

Residuals:
      Min        1Q    Median        3Q       Max
-0.0146909 -0.0018427 -0.0000254  0.0019177  0.0136030

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.0001738  0.0000941   1.847   0.0651 .
Nifty_adj   0.9407631  0.0077263 121.762   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.003072 on 1065 degrees of freedom
Multiple R-squared:  0.933,     Adjusted R-squared:  0.9329
F-statistic: 1.483e+04 on 1 and 1065 DF,  p-value: < 2.2e-16
```



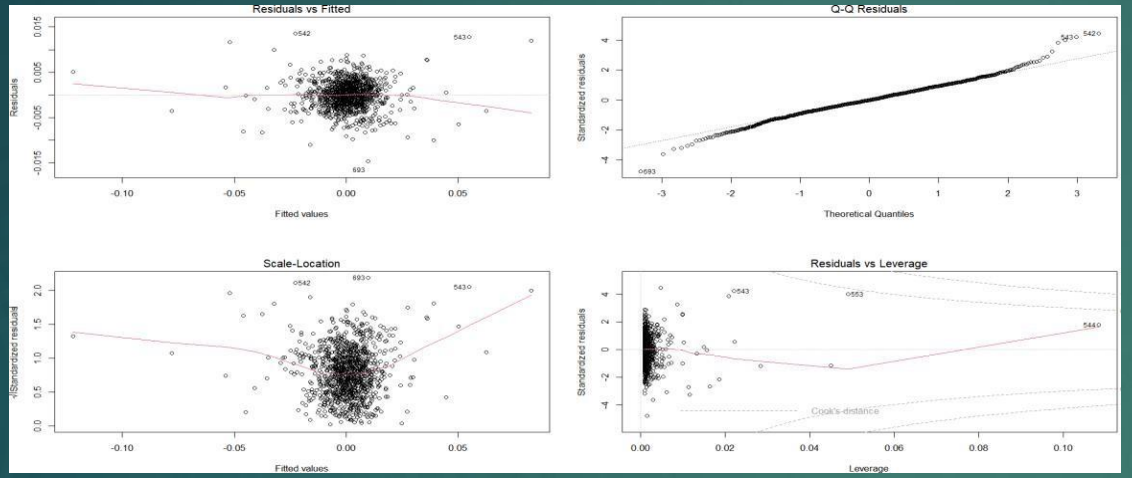
Model Evaluation

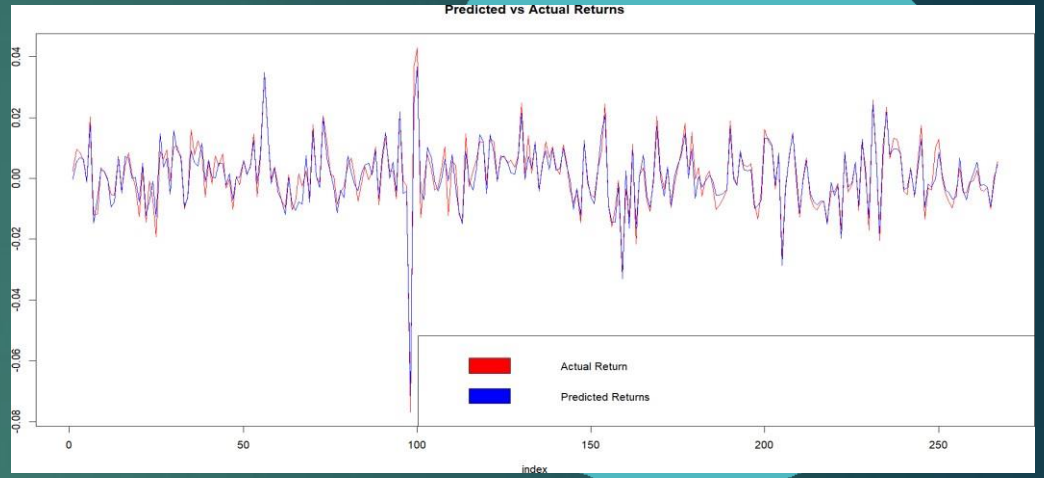
- Low p-value (to the tune of 10^-6 indicating high significance)
- NCV Test on the fitted model indicates the p-value of p = 0.0043022, which is low and so the null hypothesis that "there is no heteroscedasticity" is rejected.
- Breusch Pagan test as well as Standardized Breusch Pagan test show the p-value of 0.004302 and 0.03071. Both these are lower than 0.05 p-value and hence, null hypothesis that "the parameters are homoscedastic" is rejected.
- Durbin Watson test indicate p-value of 0.4699, which is significantly higher than 0.05. So the null hypothesis of "autocorrelation is not present" cannot be rejected
- Similar other tests signify very high significance of model correlation

	rstudent	unadjusted p-value	Bonferroni p
693	-4.836588	1.5154e-06	0.0016170
542	4.478820	8.3162e-06	0.0088734
543	4.221872	2.6298e-05	0.0280600

Model Validation

- Training conducted with random 80% and rest 20% data is used for testing





- The fitted model residual analysis shows no autocorrelation
- Similarly, the estimated and actual output closely follow each other with correlation of 0.96.

Classification

- Linear model with UpDown variable (portfolio return is positive or negative)
- Tested for various thresholds of 0.4, 0.5, 0.6, 0.7
- Comparative performance between Linear, Logit & Probit models

Threshold	Accuracy	Sensitivity	Specificity	Class
<dbl>	<dbl>	<dbl>	<dbl>	<chr>
0.4	0.777	0.507	0.997	Linear
0.5	0.885	0.808	0.947	Linear
0.6	0.856	0.977	0.757	Linear
0.7	0.724	0.998	0.5	Linear
0.4	0.898	0.852	0.935	Logit
0.5	0.899	0.881	0.913	Logit
0.6	0.897	0.923	0.876	Logit
0.7	0.889	0.948	0.842	Logit
0.4	0.894	0.839	0.939	Probit
0.5	0.899	0.881	0.913	Probit
0.6	0.897	0.925	0.874	Probit
0.7	0.883	0.952	0.827	Probit

Threshold 0.4: Logit model has highest sensitivity, and Linear model has highest specificity

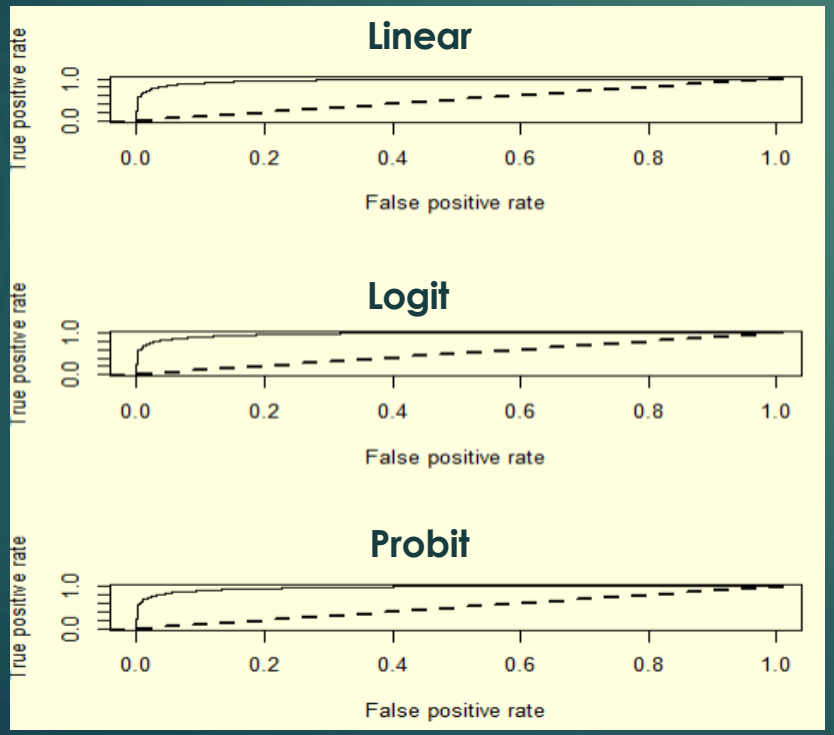
Threshold 0.5: Logit model has highest sensitivity, and Linear model has highest specificity. Logit and Probit model have comparable sensitivity (higher than Linear model), and Linear model has highest specificity (with Logit and Probit model having comparable specificity

Threshold 0.6: Linear model has highest sensitivity, and Logit model has highest specificity

Threshold 0.7: Linear model has highest sensitivity, and Logit model has highest specificity

Comparative Performance

- TPR and FPR for the models is almost similar across Linear, Logit and Probit models.
- The AUC for all Linear, Logit and Probit models is 0.9669, indicating that the model is able to make good prediction.



Further Analysis

- *Portfolio rebalancing* was performed to accommodate monthly variations in the stock prices and momentum
- Key findings
  - There is not much difference in model performance with or without rebalancing, except for heteroscedasticity. The model with rebalancing performs slightly better
  - Multiple R-squared is slightly higher at 0.9337
  - Slight improvement in AUC for ROC
  - For heteroskedasticity, using Breusch-Pagan Test, we find that p-value is greater than 0.05 so we failed to reject null hypothesis, and we conclude that there is no heteroscedasticity
- Further exploration
  - Try to do a weekly portfolio rebalancing