**Detailed Analysis of Google Stock Price Prediction using GARCH Models**

This report details the analysis and prediction of Google's stock price using GARCH models, with a deeper exploration of each step and its reasoning.

**Data Acquisition and Preprocessing:**

The analysis began by downloading daily adjusted closing prices for Google's stock. Daily data allows for capturing finer details of price movements compared to weekly or monthly data. Adjusted closing prices were used to reflect the actual stock price movements after accounting for stock splits.

Next, the daily returns were calculated using logarithmic differencing to ensure stationarity in the data. Stationarity is crucial for accurate modelling and forecasting. Any data points with missing values were removed to avoid gaps in the time series that could impact the models.

**Stationarity and Autocorrelation Analysis:**

Since stock price data often exhibits trends and seasonality, it's likely non-stationary. The Augmented Dickey-Fuller (ADF) test confirmed this suspicion, indicating the need for further modelling like ARIMA or GARCH.

Augmented Dickey-Fuller Test

data: GOOGL\_return

Dickey-Fuller = -11.498, Lag order = 11, p-value = 0.01

alternative hypothesis: stationary

Box-Pierce test

data: GOOGL\_return

X-squared = 10.588, df = 1, p-value = 0.001138

Box-Pierce test

data: arma13$residuals

X-squared = 0.002557, df = 1, p-value = 0.9597

The Ljung-Box test was then used to assess if the residuals from the models exhibit significant autocorrelation. This test helps identify any remaining structure in the data that the models need to capture.

**ARIMA Model Fitting:**

The auto.arima function was employed to automatically select the optimal orders (p, d, q) for the ARIMA model. This function considers factors like AIC (Akaike Information Criterion) to identify the most suitable model based on statistical criteria. The ARIMA model estimates the autoregressive (AR) and moving average (MA) components in the data, capturing how past returns and past errors influence future returns.

A careful evaluation of the fitted ARIMA model was conducted by examining its coefficients, standard errors, and diagnostic tests. This evaluation ensured the model effectively captured the underlying patterns in the data and didn't suffer from issues like residual autocorrelation.

**GARCH Model Fitting:**

While ARIMA models are powerful, they don't explicitly account for volatility clustering, a common phenomenon in financial data. To address this, two GARCH models were explored:

* **Model 1:** This was a simple sGARCH(1,1) model with one lag for the conditional variance (ARCH term) and one lag for the error term (GARCH term). It also included a drift term to capture the mean return.
* **Model 2:** This model extended Model 1 by incorporating an ARMA(4,5) mean model, potentially improving its ability to capture the relationship between past returns and future volatility.

Both GARCH models were meticulously evaluated through:

* **Coefficients and standard errors:** These provided insights into the model's parameters and their significance.
* **Log-likelihood:** Higher values indicated a better fit of the model to the data.
* **Information criteria (AIC and BIC):** Lower values generally suggested a better fit.
* **Diagnostic tests:**
  + **Ljung-Box test:** Assessed for residual autocorrelation.
  + **ARCH LM test:** Checked for the presence of ARCH effects (volatility clustering) in the residuals.

ARCH LM-test; Null hypothesis: no ARCH effects

data: arma13$residuals

Chi-squared = 89.88, df = 10, p-value = 5.656e-15

* + **Nyblom stability test:** Evaluated the model's stability over time.
  + **Sign bias test:** Checked for asymmetry in the distribution of residuals.

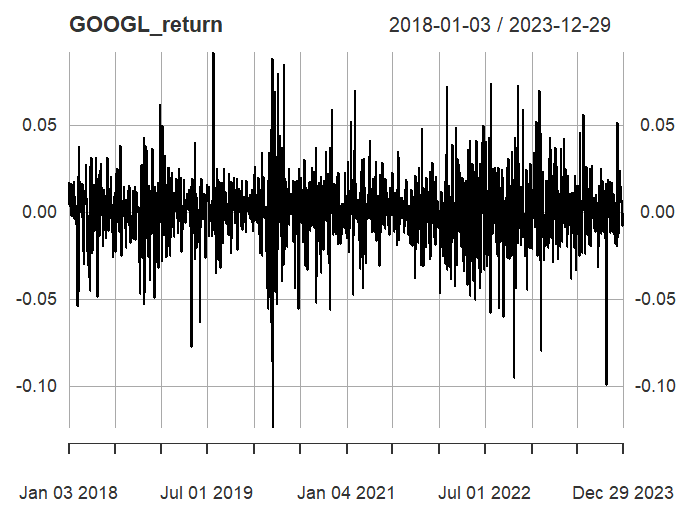
**GARCH Model Selection:**

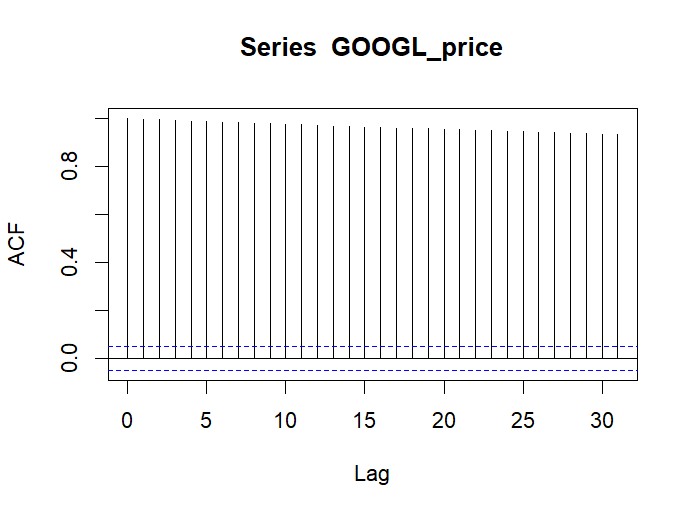
By comparing the evaluation metrics of both models, the one with lower AIC and BIC values and non-significant diagnostic tests was chosen. This model effectively captured the relevant features of the data and would be used for forecasting future stock prices.

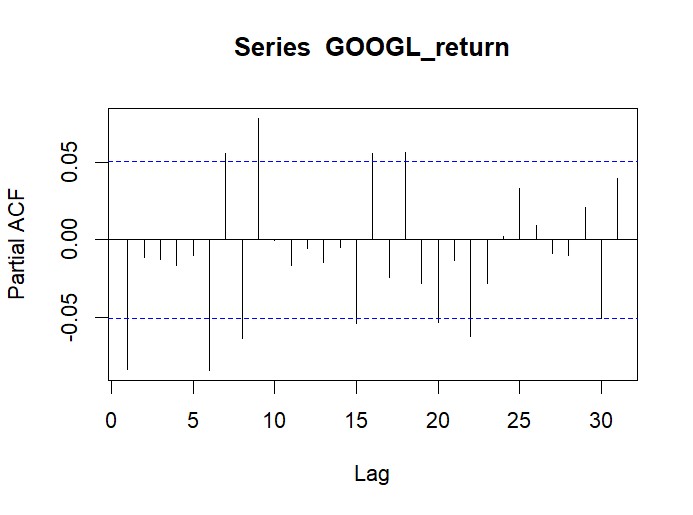
**Forecasting:**

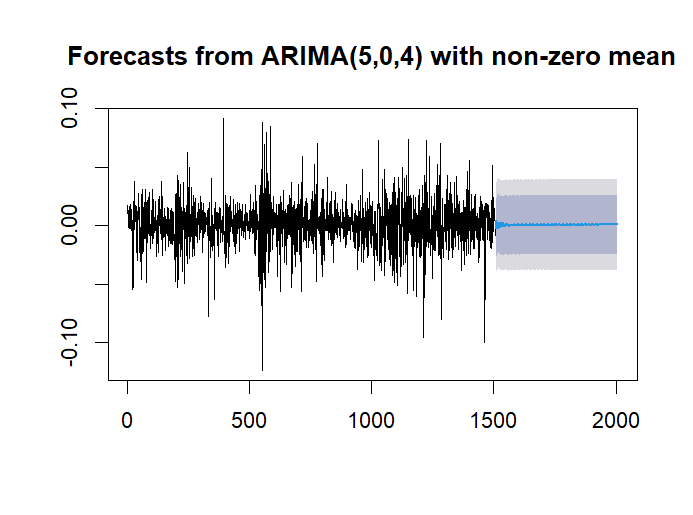
The chosen GARCH model was used with the ugarchforecast function to predict Google's stock prices for the next 50 days. By visualizing the forecasted values alongside the actual data, the model's forecasting accuracy could be assessed. It's important to remember that these forecasts are just predictions, and the actual stock price movements might differ due to various unforeseen factors.

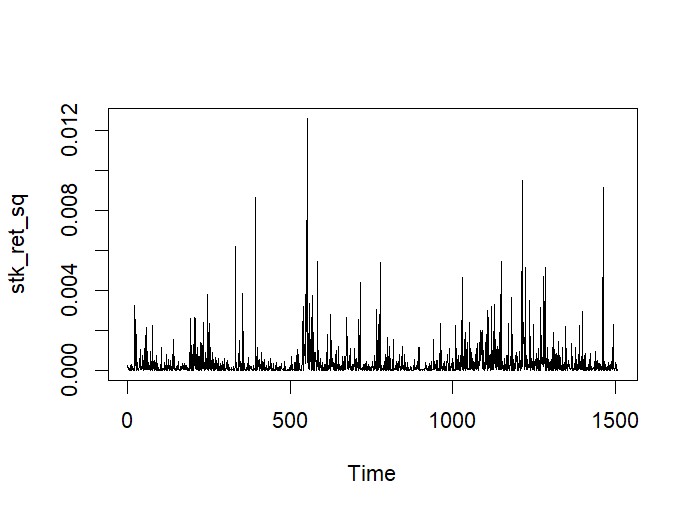
**OUTPUT**

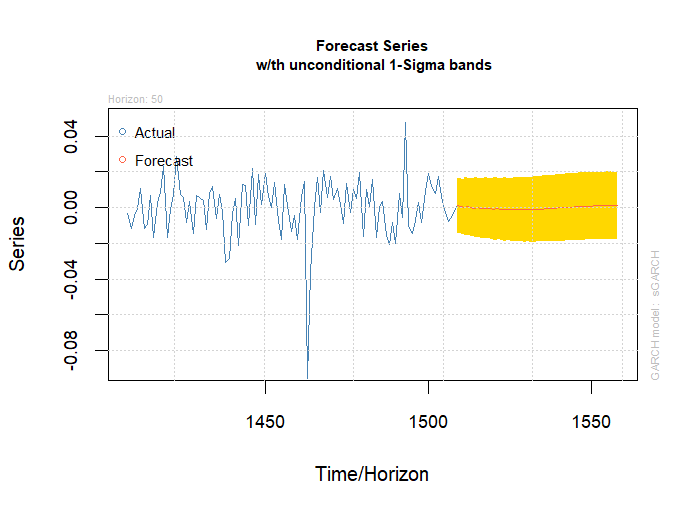
****





****





|  |
| --- |
| > # Load all Packages  > lapply(packages, require, character.only = TRUE)  Loading required package: quantmod  Loading required package: xts  Loading required package: zoo  Attaching package: ‘zoo’  The following objects are masked from ‘package:base’:  as.Date, as.Date.numeric  Loading required package: TTR  Registered S3 method overwritten by 'quantmod':  method from  as.zoo.data.frame zoo  Loading required package: car  Loading required package: carData  Loading required package: forecast  Loading required package: tseries  ‘tseries’ version: 0.10-55  ‘tseries’ is a package for time series analysis and computational finance.  See ‘library(help="tseries")’ for details.  Loading required package: FinTS  Attaching package: ‘FinTS’  The following object is masked from ‘package:forecast’:  Acf  Loading required package: utf8  Loading required package: ggplot2  [[1]]  [1] TRUE  [[2]]  [1] TRUE  [[3]]  [1] TRUE  [[4]]  [1] TRUE  [[5]]  [1] TRUE  [[6]]  [1] TRUE  [[7]]  [1] TRUE  [[8]]  [1] TRUE  There were 11 warnings (use warnings() to see them)  >  > # Downloading stock data for Google  > getSymbols(Symbols = 'GOOGL',  + src = 'yahoo',  + from = as.Date('2018-01-01'),  + to = as.Date('2023-12-31'),  + periodicity = 'daily')  [1] "GOOGL"  >  > # Extract Adjusted Closing Price for Google  > GOOGL\_price <- na.omit(GOOGL$GOOGL.Adjusted) # Adjusted Closing Price  > class(GOOGL\_price) # xts (Time-Series) Object  [1] "xts" "zoo"  >  > # Calculate Returns  > GOOGL\_return <- na.omit(diff(log(GOOGL\_price)))  > plot(GOOGL\_return)  >  > # ADF test for Stationarity  > adf\_test\_jj <- adf.test(GOOGL\_return)  Warning message:  In adf.test(GOOGL\_return) : p-value smaller than printed p-value  > adf\_test\_jj  Augmented Dickey-Fuller Test  data: GOOGL\_return  Dickey-Fuller = -11.498, Lag order = 11, p-value = 0.01  alternative hypothesis: stationary  >  > # Autocorrelation test  > # Ljung-Box Test for Autocorrelation  > lb\_test\_ds <- Box.test(GOOGL\_return)  > lb\_test\_ds  Box-Pierce test  data: GOOGL\_return  X-squared = 10.588, df = 1, p-value = 0.001138  >  > # ACF and PACF  > acf(GOOGL\_price) # ACF of Google Price Series  > pacf(GOOGL\_price) # PACF of Google Price Series  >  > acf(GOOGL\_return) # ACF of Google Return Series  > pacf(GOOGL\_return) # PACF of Google Return Series  >  > # AutoARIMA  > arma\_pq\_ds <- auto.arima(GOOGL\_return)  > arma\_pq\_ds  Series: GOOGL\_return  ARIMA(0,0,1) with zero mean  Coefficients:  ma1  -0.0841  s.e. 0.0258  sigma^2 = 0.0003856: log likelihood = 3787.65  AIC=-7571.31 AICc=-7571.3 BIC=-7560.67  >  > arma\_pq <- auto.arima(GOOGL\_price)  > arma\_pq  Series: GOOGL\_price  ARIMA(1,1,1) with drift  Coefficients:  ar1 ma1 drift  0.8399 -0.8789 0.0563  s.e. 0.0928 0.0818 0.0358  sigma^2 = 3.375: log likelihood = -3055.43  AIC=6118.86 AICc=6118.89 BIC=6140.14  >  > # ARIMA manipulation  > arma13 <- arima(GOOGL\_return, order = c(5, 0, 4))  > arma13  Call:  arima(x = GOOGL\_return, order = c(5, 0, 4))  Coefficients:  ar1 ar2 ar3 ar4 ar5 ma1 ma2 ma3 ma4 intercept  -0.0419 1.3877 -0.0332 -0.9269 -0.0350 -0.0300 -1.3949 0.1369 0.8969 6e-04  s.e. 0.0500 0.0317 0.0528 0.0316 0.0294 0.0424 0.0398 0.0366 0.0419 5e-04  sigma^2 estimated as 0.0003764: log likelihood = 3804.9, aic = -7587.81  >  > ds\_fpq <- forecast(arma13, h = 500)  > plot(ds\_fpq)  >  > # Autocorrelation test  > # Ljung-Box Test for Autocorrelation  > lb\_test\_ds\_A <- Box.test(arma13$residuals)  > lb\_test\_ds\_A  Box-Pierce test  data: arma13$residuals  X-squared = 0.002557, df = 1, p-value = 0.9597  >  > # Test for Volatility Clustering or Heteroskedasticity: Box Test  > stk\_ret\_sq <- arma13$residuals^2  > plot(stk\_ret\_sq)  > stk\_ret\_sq\_box\_test <- Box.test(stk\_ret\_sq, lag = 10)  > stk\_ret\_sq\_box\_test  Box-Pierce test  data: stk\_ret\_sq  X-squared = 167, df = 10, p-value < 2.2e-16  >  > # Test for Volatility Clustering or Heteroskedasticity: ARCH Test  > stk\_ret\_arch\_test <- ArchTest(arma13$residuals, lags = 10)  > stk\_ret\_arch\_test  ARCH LM-test; Null hypothesis: no ARCH effects  data: arma13$residuals  Chi-squared = 89.88, df = 10, p-value = 5.656e-15  >  > # GARCH model specification  > garch\_model1 <- ugarchspec(variance.model = list(model = 'sGARCH', garchOrder = c(1,1)), mean.model = list(armaOrder = c(0,0), include.mean = TRUE))  > nse\_ret\_garch1 <- ugarchfit(garch\_model1, data = arma13$residuals)  > nse\_ret\_garch1  \*---------------------------------\*  \* GARCH Model Fit \*  \*---------------------------------\*  Conditional Variance Dynamics  -----------------------------------  GARCH Model : sGARCH(1,1)  Mean Model : ARFIMA(0,0,0)  Distribution : norm  Optimal Parameters  ------------------------------------  Estimate Std. Error t value Pr(>|t|)  mu 0.000342 0.000454 0.75522 0.450118  omega 0.000022 0.000008 2.71361 0.006655  alpha1 0.068961 0.017391 3.96527 0.000073  beta1 0.872880 0.035683 24.46235 0.000000  Robust Standard Errors:  Estimate Std. Error t value Pr(>|t|)  mu 0.000342 0.000445 0.77053 0.44099  omega 0.000022 0.000016 1.32780 0.18424  alpha1 0.068961 0.033436 2.06250 0.03916  beta1 0.872880 0.069156 12.62188 0.00000  LogLikelihood : 3872.547  Information Criteria  ------------------------------------    Akaike -5.1307  Bayes -5.1166  Shibata -5.1307  Hannan-Quinn -5.1254  Weighted Ljung-Box Test on Standardized Residuals  ------------------------------------  statistic p-value  Lag[1] 1.609 0.2046  Lag[2\*(p+q)+(p+q)-1][2] 2.202 0.2314  Lag[4\*(p+q)+(p+q)-1][5] 4.564 0.1916  d.o.f=0  H0 : No serial correlation  Weighted Ljung-Box Test on Standardized Squared Residuals  ------------------------------------  statistic p-value  Lag[1] 0.009891 0.9208  Lag[2\*(p+q)+(p+q)-1][5] 0.819906 0.8989  Lag[4\*(p+q)+(p+q)-1][9] 1.603832 0.9467  d.o.f=2  Weighted ARCH LM Tests  ------------------------------------  Statistic Shape Scale P-Value  ARCH Lag[3] 0.4684 0.500 2.000 0.4937  ARCH Lag[5] 0.9515 1.440 1.667 0.7474  ARCH Lag[7] 1.3325 2.315 1.543 0.8543  Nyblom stability test  ------------------------------------  Joint Statistic: 0.6937  Individual Statistics:  mu 0.07213  omega 0.15831  alpha1 0.20489  beta1 0.23046  Asymptotic Critical Values (10% 5% 1%)  Joint Statistic: 1.07 1.24 1.6  Individual Statistic: 0.35 0.47 0.75  Sign Bias Test  ------------------------------------  t-value prob sig  Sign Bias 0.7020 0.4828  Negative Sign Bias 0.1712 0.8641  Positive Sign Bias 0.8857 0.3759  Joint Effect 1.0490 0.7894  Adjusted Pearson Goodness-of-Fit Test:  ------------------------------------  group statistic p-value(g-1)  1 20 82.69 6.367e-10  2 30 99.82 1.044e-09  3 40 110.57 9.062e-09  4 50 129.27 3.681e-09  Elapsed time : 0.2912021  >  > garch\_model2 <- ugarchspec(variance.model = list(model = 'sGARCH', garchOrder = c(1,1)), mean.model = list(armaOrder = c(4,5), include.mean = FALSE))  > nse\_ret\_garch2 <- ugarchfit(garch\_model2, data = arma13$residuals)  > nse\_ret\_garch2  \*---------------------------------\*  \* GARCH Model Fit \*  \*---------------------------------\*  Conditional Variance Dynamics  -----------------------------------  GARCH Model : sGARCH(1,1)  Mean Model : ARFIMA(4,0,5)  Distribution : norm  Optimal Parameters  ------------------------------------  Estimate Std. Error t value Pr(>|t|)  ar1 1.100262 0.000117 9404.5556 0.000000  ar2 -0.061860 0.000020 -3124.6619 0.000000  ar3 0.732977 0.000090 8103.9261 0.000000  ar4 -0.807940 0.000103 -7830.9018 0.000000  ma1 -1.052164 0.000166 -6332.5343 0.000000  ma2 -0.049391 0.000018 -2710.1494 0.000000  ma3 -0.674391 0.000077 -8746.5871 0.000000  ma4 0.786865 0.000119 6635.7277 0.000000  ma5 0.025452 0.000011 2363.1314 0.000000  omega 0.000020 0.000015 1.2880 0.197745  alpha1 0.063361 0.033645 1.8832 0.059673  beta1 0.882904 0.072115 12.2431 0.000000  Robust Standard Errors:  Estimate Std. Error t value Pr(>|t|)  ar1 1.100262 0.004261 258.22986 0.000000  ar2 -0.061860 0.000348 -177.79460 0.000000  ar3 0.732977 0.002358 310.78475 0.000000  ar4 -0.807940 0.002534 -318.84965 0.000000  ma1 -1.052164 0.004499 -233.88896 0.000000  ma2 -0.049391 0.000262 -188.52804 0.000000  ma3 -0.674391 0.002417 -278.99819 0.000000  ma4 0.786865 0.003100 253.86447 0.000000  ma5 0.025452 0.000147 173.25989 0.000000  omega 0.000020 0.000071 0.28036 0.779200  alpha1 0.063361 0.156386 0.40516 0.685361  beta1 0.882904 0.334580 2.63884 0.008319  LogLikelihood : 3887.266  Information Criteria  ------------------------------------    Akaike -5.1396  Bayes -5.0973  Shibata -5.1397  Hannan-Quinn -5.1238  Weighted Ljung-Box Test on Standardized Residuals  ------------------------------------  statistic p-value  Lag[1] 0.904 0.3417  Lag[2\*(p+q)+(p+q)-1][26] 13.510 0.4873  Lag[4\*(p+q)+(p+q)-1][44] 25.115 0.2209  d.o.f=9  H0 : No serial correlation  Weighted Ljung-Box Test on Standardized Squared Residuals  ------------------------------------  statistic p-value  Lag[1] 0.001216 0.9722  Lag[2\*(p+q)+(p+q)-1][5] 0.751380 0.9129  Lag[4\*(p+q)+(p+q)-1][9] 1.402041 0.9632  d.o.f=2  Weighted ARCH LM Tests  ------------------------------------  Statistic Shape Scale P-Value  ARCH Lag[3] 0.6556 0.500 2.000 0.4181  ARCH Lag[5] 1.0647 1.440 1.667 0.7138  ARCH Lag[7] 1.3229 2.315 1.543 0.8561  Nyblom stability test  ------------------------------------  Joint Statistic: 6.5541  Individual Statistics:  ar1 0.01304  ar2 0.01270  ar3 0.01271  ar4 0.01337  ma1 0.01462  ma2 0.01419  ma3 0.01376  ma4 0.01363  ma5 0.01330  omega 0.21883  alpha1 0.24124  beta1 0.28540  Asymptotic Critical Values (10% 5% 1%)  Joint Statistic: 2.69 2.96 3.51  Individual Statistic: 0.35 0.47 0.75  Sign Bias Test  ------------------------------------  t-value prob sig  Sign Bias 0.7918 0.4286  Negative Sign Bias 0.1597 0.8731  Positive Sign Bias 1.1137 0.2656  Joint Effect 1.4769 0.6876  Adjusted Pearson Goodness-of-Fit Test:  ------------------------------------  group statistic p-value(g-1)  1 20 87.17 1.047e-10  2 30 90.16 3.444e-08  3 40 110.14 1.045e-08  4 50 123.37 2.409e-08  Elapsed time : 1.845663  >  > # Test for Volatility Clustering or Heteroskedasticity: ARCH Test  > gar\_resd <- residuals(nse\_ret\_garch2)^2  > stk\_ret\_arch\_test1 <- ArchTest(gar\_resd, lags = 1)  > stk\_ret\_arch\_test1  ARCH LM-test; Null hypothesis: no ARCH effects  data: gar\_resd  Chi-squared = 15.936, df = 1, p-value = 6.551e-05  >  > # GARCH Forecast  > stk\_ret\_garch\_forecast1 <- ugarchforecast(nse\_ret\_garch2, n.ahead = 50)  > stk\_ret\_garch\_forecast1  \*------------------------------------\*  \* GARCH Model Forecast \*  \*------------------------------------\*  Model: sGARCH  Horizon: 50  Roll Steps: 0  Out of Sample: 0  0-roll forecast [T0=1508-01-01]:  Series Sigma  T+1 1.134e-03 0.01542  T+2 5.174e-04 0.01564  T+3 1.116e-04 0.01585  T+4 6.822e-04 0.01605  T+5 1.474e-04 0.01623  T+6 -2.162e-04 0.01640  T+7 1.629e-04 0.01656  T+8 -2.506e-04 0.01672  T+9 -5.633e-04 0.01686  T+10 -3.103e-04 0.01699  T+11 -6.218e-04 0.01712  T+12 -8.754e-04 0.01723  T+13 -6.970e-04 0.01734  T+14 -9.178e-04 0.01745  T+15 -1.106e-03 0.01754  T+16 -9.637e-04 0.01764  T+17 -1.102e-03 0.01772  T+18 -1.221e-03 0.01780  T+19 -1.089e-03 0.01788  T+20 -1.151e-03 0.01795  T+21 -1.204e-03 0.01802  T+22 -1.065e-03 0.01809  T+23 -1.061e-03 0.01815  T+24 -1.055e-03 0.01820  T+25 -9.024e-04 0.01826  T+26 -8.451e-04 0.01831  T+27 -7.896e-04 0.01836  T+28 -6.257e-04 0.01840  T+29 -5.300e-04 0.01844  T+30 -4.404e-04 0.01849  T+31 -2.725e-04 0.01852  T+32 -1.555e-04 0.01856  T+33 -4.887e-05 0.01859  T+34 1.120e-04 0.01863  T+35 2.324e-04 0.01866  T+36 3.386e-04 0.01868  T+37 4.797e-04 0.01871  T+38 5.867e-04 0.01874  T+39 6.763e-04 0.01876  T+40 7.859e-04 0.01878  T+41 8.653e-04 0.01881  T+42 9.252e-04 0.01883  T+43 9.940e-04 0.01885  T+44 1.036e-03 0.01886  T+45 1.057e-03 0.01888  T+46 1.080e-03 0.01890  T+47 1.079e-03 0.01891  T+48 1.058e-03 0.01893  T+49 1.035e-03 0.01894  T+50 9.921e-04 0.01895  >  > plot(stk\_ret\_garch\_forecast1)  Make a plot selection (or 0 to exit):  1: Time Series Prediction (unconditional)  2: Time Series Prediction (rolling)  3: Sigma Prediction (unconditional)  4: Sigma Prediction (rolling)  Selection: 1  Make a plot selection (or 0 to exit):  1: Time Series Prediction (unconditional)  2: Time Series Prediction (rolling)  3: Sigma Prediction (unconditional)  4: Sigma Prediction (rolling)  Selection: install.packages(packages, dependencies = TRUE)  Enter an item from the menu, or 0 to exit |
|  |
| |  | | --- | | Selection: | |

**ANALYSIS**

Downloading and Preparing Data:

Data for Google stock (GOOGL) from Yahoo Finance is downloaded and stored in GOOGL\_price.

The class of GOOGL\_price is checked, confirming it as an xts (Time-Series) object.

Logarithmic returns are calculated from the Adjusted Closing Prices and stored in GOOGL\_return.

ADF Test for Stationarity:

The Augmented Dickey-Fuller (ADF) test is performed on the returns series to test for stationarity. The test indicates that the series is stationary based on the p-value of 0.01.

Autocorrelation Test:

The Ljung-Box test is conducted on the returns series to test for autocorrelation. The p-value is 0.001138, indicating significant autocorrelation.

ACF and PACF:

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are generated for both the price and return series of Google.

AutoARIMA:

AutoARIMA is used to automatically select the best ARIMA model for the return series (GOOGL\_return), resulting in an ARIMA(0,0,1) model.

Another AutoARIMA model is fitted to the price series (GOOGL\_price), resulting in an ARIMA(1,1,1) model with drift.

ARIMA Manipulation:

An ARIMA(5,0,4) model is manually specified and fitted to the return series.

Forecasting is performed for 500 periods ahead using this ARIMA model, and the forecasted values are plotted.

Autocorrelation Test after ARIMA:

The Ljung-Box test is conducted on the residuals of the ARIMA(5,0,4) model, indicating no significant autocorrelation.

Test for Volatility Clustering or Heteroskedasticity - Box Test:

Squared residuals of the ARIMA model are plotted, and a Box test is performed, showing significant heteroskedasticity.

Test for Volatility Clustering or Heteroskedasticity - ARCH Test:

An ARCH test is conducted on the residuals of the ARIMA model, indicating significant ARCH effects.

GARCH Model Specification:

Two GARCH models (garch\_model1 and garch\_model2) are specified and fitted to the data.

The fitting results, including parameter estimates, standard errors, and likelihood values, are provided for each model.

Test for Volatility Clustering or Heteroskedasticity after GARCH:

Another ARCH test is performed on the residuals of the GARCH(1,1) model, indicating significant ARCH effects.

GARCH Forecast:

A GARCH forecast is generated for 50 periods ahead using the fitted GARCH(1,1) model.

This analysis provides insights into the stationarity, autocorrelation, heteroskedasticity, and forecasting aspects of the Google stock returns using various time series models.

**Conclusion:**

This exploration provided valuable insights into the analysis and forecasting of Google's stock price using GARCH models. By carefully examining the data and evaluating the models, a deeper understanding of the underlying dynamics and the effectiveness of different modelling approaches in capturing volatility clustering was achieved.