Thesis_R_Code

Pranav Tavildar

2023-04-04

Loading Libraries

```
library(rugarch)
```

```
## Loading required package: parallel
```

```
##
## Attaching package: 'rugarch'
```

```
## The following object is masked from 'package:stats':
##
##     sigma
```

```
library(readr)
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(tseries)
library(garchx)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
library(TSA)
```

```
## Registered S3 methods overwritten by 'TSA':
##   method       from
##   fitted.Arima forecast
##   plot.Arima   forecast
```

```
##
## Attaching package: 'TSA'
```

```
## The following object is masked from 'package:readr':
##
##     spec
```

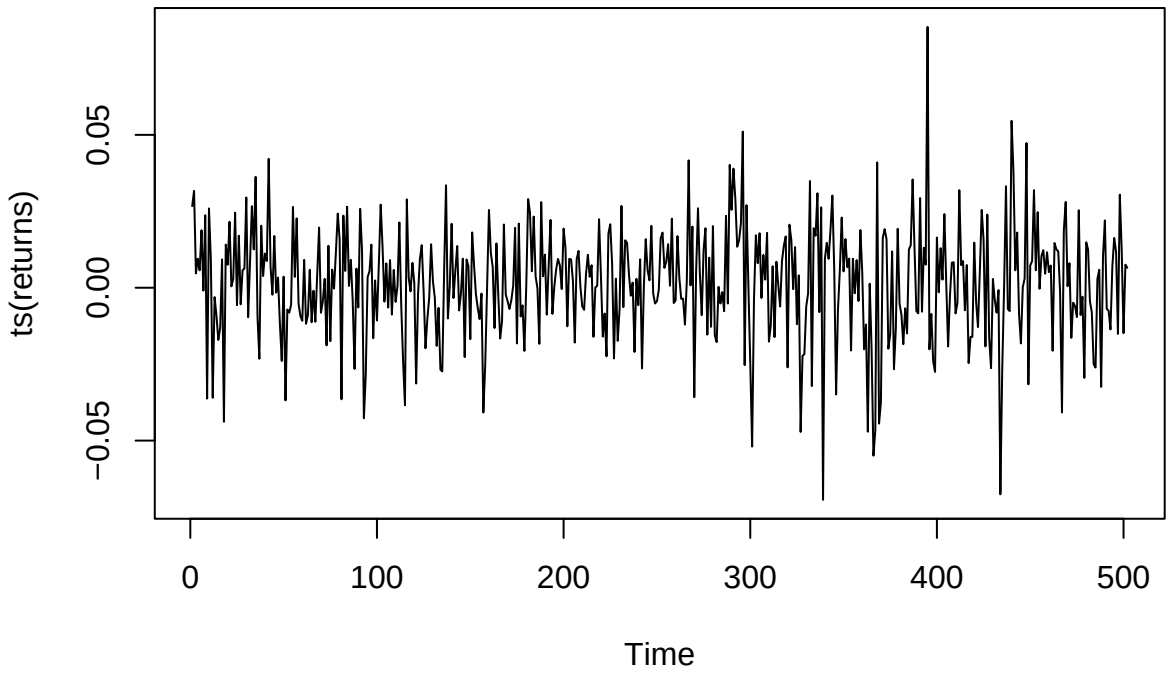
```
## The following objects are masked from 'package:stats':
##
```

```
##     acf, arima
```

```
## The following object is masked from 'package:utils':
##
##     tar
```

Loading Data and Calculating Returns

```
stock_data <- read.csv("/Users/pranavtavildar/Desktop/Senior-Thesis/data/CVX_stock_data.csv")
returns <- diff(log(stock_data$Close))
plot(ts(returns))
```



Here we use auto.arima to select the best arima model

```
arma_model <- auto.arima(returns,  max.D = 0, max.d = 0)
arma_model
```

```
## Series: returns
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       0.0015
## s.e.  0.0008
##
## sigma^2 = 0.000339:  log likelihood = 1293.56
## AIC=-2583.12   AICc=-2583.1   BIC=-2574.68
```

Here, we are doing Garch where the order is (1,1) and we apply it to the arima model selected by auto.arima

```
garch_spec <- ugarchspec(variance.model = list(model = "sGARCH", garchOrder = c(1,1)), mean.model = lis
garch_model2 <- ugarchfit(garch_spec, data = returns)
summary(garch_model2)
```

```
##    Length     Class      Mode
##         1 uGARCHfit        S4
```

```
show(garch_model2)
```

```
##
## *---------------------------------*
## *          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      0.001503    0.000786  1.91115 0.055985
## omega   0.000003    0.000004  0.82203 0.411059
## alpha1  0.024261    0.011361  2.13550 0.032721
## beta1   0.966171    0.014672 65.85297 0.000000
##
## Robust Standard Errors:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.001503    0.000786  1.91204 0.055872
## omega   0.000003    0.000016  0.20481 0.837723
## alpha1  0.024261    0.027300  0.88868 0.374173
## beta1   0.966171    0.046123 20.94758 0.000000
##
## LogLikelihood : 1300.803
##
## Information Criteria
## ------------------------------------
##
## Akaike       -5.1665
## Bayes        -5.1329
## Shibata      -5.1667
## Hannan-Quinn -5.1534
##
## Weighted Ljung-Box Test on Standardized Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.4181  0.5179
## Lag[2*(p+q)+(p+q)-1][2]    0.4548  0.7152
## Lag[4*(p+q)+(p+q)-1][5]    1.3065  0.7873
## d.o.f=0
```

```
## H0 : No serial correlation
##
## Weighted Ljung-Box Test on Standardized Squared Residuals
## ------------------------------------
##                         statistic p-value
## Lag[1]                     0.1824  0.6693
## Lag[2*(p+q)+(p+q)-1][5]    1.7904  0.6687
## Lag[4*(p+q)+(p+q)-1][9]    3.8059  0.6219
## d.o.f=2
##
## Weighted ARCH LM Tests
## ------------------------------------
##             Statistic Shape Scale P-Value
## ARCH Lag[3]     2.137 0.500 2.000  0.1438
## ARCH Lag[5]     2.913 1.440 1.667  0.3026
## ARCH Lag[7]     4.259 2.315 1.543  0.3108
##
## Nyblom stability test
## ------------------------------------
## Joint Statistic:  11.883
## Individual Statistics:
## mu     0.05534
## omega  1.16577
## alpha1 0.18610
## beta1  0.18235
##
## 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           0.7048 0.4813
## Negative Sign Bias  0.2108 0.8331
## Positive Sign Bias  0.6407 0.5220
## Joint Effect        0.5902 0.8987
##
##
## Adjusted Pearson Goodness-of-Fit Test:
## ------------------------------------
##   group statistic p-value(g-1)
## 1    20     33.46      0.02127
## 2    30     40.99      0.06903
## 3    40     54.33      0.05226
## 4    50     70.11      0.02552
##
##
## Elapsed time : 0.08320093
```

t value of beta 1 proves that the p value is non significant

Alternative Garch implementation gives us the same result

```
garch_model <- garch(returns)
```

```
##
##  ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##
##      I     INITIAL X(I)        D(I)
##
##      1     3.050993e-04     1.000e+00
##      2     5.000000e-02     1.000e+00
##      3     5.000000e-02     1.000e+00
##
##     IT   NF      F         RELDF    PRELDF    RELDX   STPPAR   D*STEP   NPRELDF
##      0    1 -1.752e+03
##      1    8 -1.752e+03  1.17e-08  6.43e-07  1.0e-05  1.1e+09  1.0e-06  3.62e+02
##      2    9 -1.752e+03  1.70e-07  2.06e-07  5.0e-06  2.1e+00  5.0e-07  9.03e-03
##      3   10 -1.752e+03  1.52e-09  4.40e-09  1.0e-05  2.0e+00  1.0e-06  9.03e-03
##      4   13 -1.752e+03  3.98e-08  4.72e-08  1.7e-04  4.3e+00  1.7e-05  9.02e-03
##      5   18 -1.752e+03  1.38e-05  2.76e-05  7.9e-02  2.0e+00  8.6e-03  8.92e-03
##      6   22 -1.752e+03  3.24e-04  2.98e-05  7.6e-01  0.0e+00  3.2e-01  2.98e-05
##      7   24 -1.753e+03  3.28e-04  2.37e-04  1.5e-01  2.0e+00  1.3e-01  3.25e-02
##      8   26 -1.755e+03  1.11e-03  7.85e-04  2.0e-01  2.0e+00  2.6e-01  5.18e+00
##      9   28 -1.755e+03  1.03e-04  2.96e-04  3.3e-02  2.0e+00  5.2e-02  8.42e+01
##     10   30 -1.755e+03  8.86e-05  2.31e-04  1.4e-02  2.0e+00  2.4e-02  5.75e+01
##     11   31 -1.755e+03  6.85e-05  9.94e-05  1.4e-02  2.0e+00  2.4e-02  2.40e-01
##     12   32 -1.755e+03  5.91e-07  1.96e-05  1.4e-02  2.0e+00  2.4e-02  2.00e-01
##     13   34 -1.755e+03  7.46e-06  7.64e-06  7.1e-03  2.0e+00  1.2e-02  6.51e-01
##     14   40 -1.755e+03  1.39e-07  4.18e-06  2.3e-07  5.1e+00  3.8e-07  3.74e-02
##     15   41 -1.755e+03  9.96e-07  8.51e-07  1.0e-07  2.0e+00  1.9e-07  1.40e-03
##     16   52 -1.755e+03  3.08e-05  1.11e-05  2.9e-02  1.9e+00  5.0e-02  1.61e-03
##     17   54 -1.755e+03  2.09e-05  1.52e-05  5.6e-03  2.0e+00  1.0e-02  6.91e-01
##     18   55 -1.755e+03  2.97e-06  3.58e-05  1.1e-02  2.0e+00  2.0e-02  2.32e+02
##     19   61 -1.755e+03  6.46e-06  1.16e-05  1.1e-07  1.8e+01  2.0e-07  7.03e-03
##     20   69 -1.755e+03  1.33e-05  1.93e-05  1.8e-03  2.0e+00  3.3e-03  1.53e-02
##     21   72 -1.756e+03  2.46e-04  2.33e-04  6.3e-03  1.2e+00  1.3e-02  2.89e-03
##     22   73 -1.756e+03  4.45e-05  4.79e-04  1.2e-02  2.0e+00  2.6e-02  1.02e+00
##     23   74 -1.757e+03  4.49e-04  8.23e-04  4.7e-03  0.0e+00  1.2e-02  8.23e-04
##     24   76 -1.757e+03  1.07e-04  4.04e-04  1.7e-03  1.4e+00  3.2e-03  4.97e-04
##     25   77 -1.757e+03  1.46e-04  1.19e-04  1.3e-03  1.5e+00  3.2e-03  1.35e-04
##     26   79 -1.757e+03  6.37e-05  1.15e-04  5.0e-03  3.6e-01  1.1e-02  1.24e-04
##     27   80 -1.757e+03  3.71e-05  4.35e-05  1.3e-03  0.0e+00  3.0e-03  4.35e-05
##     28   81 -1.757e+03  3.06e-06  3.62e-06  8.5e-04  0.0e+00  1.7e-03  3.62e-06
##     29   82 -1.757e+03  2.48e-07  3.48e-07  1.3e-04  0.0e+00  2.6e-04  3.48e-07
##     30   83 -1.757e+03  1.29e-08  9.34e-09  1.2e-05  0.0e+00  2.5e-05  9.34e-09
##     31   84 -1.757e+03  1.62e-09  1.25e-09  9.9e-06  0.0e+00  2.4e-05  1.25e-09
##     32   85 -1.757e+03 -2.67e-10  6.95e-12  1.2e-06  0.0e+00  2.5e-06  6.95e-12
##
##  ***** RELATIVE FUNCTION CONVERGENCE *****
##
##  FUNCTION    -1.757344e+03   RELDX        1.152e-06
##  FUNC. EVALS      85         GRAD. EVALS      32
##  PRELDF       6.954e-12      NPRELDF      6.954e-12
```

```
##
##      I      FINAL X(I)        D(I)          G(I)
##
##      1    3.326084e-06     1.000e+00    -1.007e+00
##      2    2.427247e-02     1.000e+00     1.080e-03
##      3    9.658585e-01     1.000e+00     1.137e-02
```

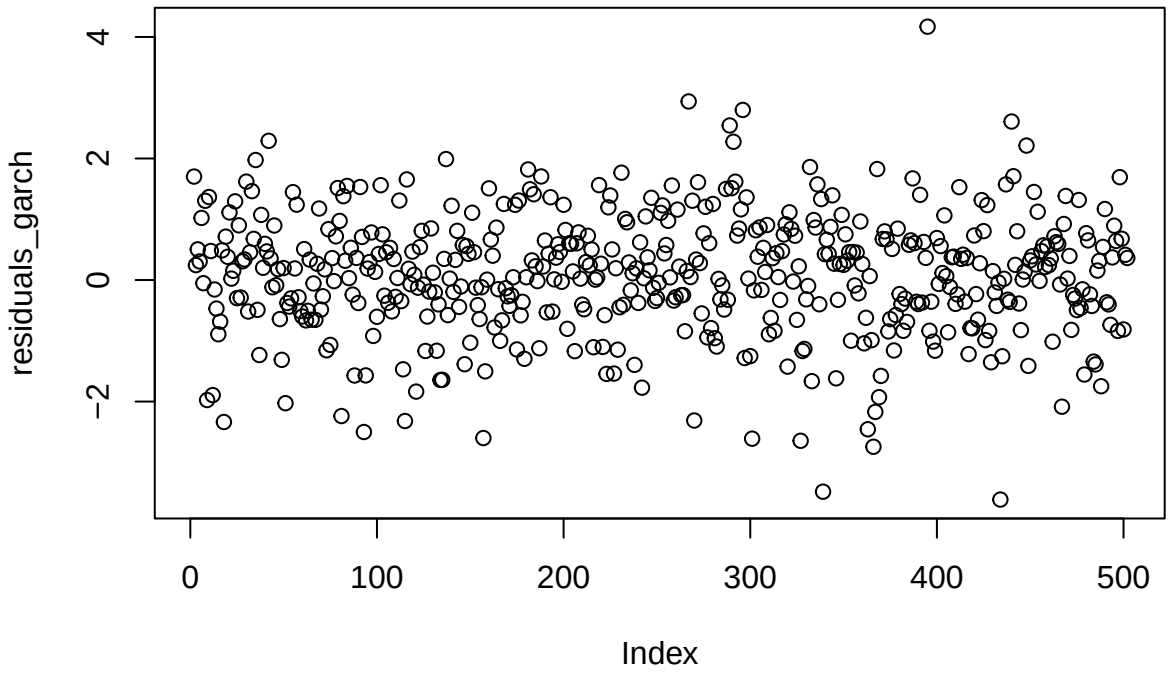
```
garch_model
```

```
##
## Call:
## garch(x = returns)
##
## Coefficient(s):
##        a0         a1         b1
## 3.326e-06  2.427e-02  9.659e-01
```

```
summary(arma_model)
```

```
## Series: returns
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       0.0015
## s.e.  0.0008
##
## sigma^2 = 0.000339:  log likelihood = 1293.56
## AIC=-2583.12   AICc=-2583.1   BIC=-2574.68
##
## Training set error measures:
##                       ME       RMSE        MAE      MPE     MAPE      MASE
## Training set 7.762973e-19 0.01839358 0.01405025 86.25959 115.9893 0.7069355
##                     ACF1
## Training set 0.03172501
```

```
residuals_garch <- residuals(garch_model)
plot(residuals_garch)
```



Adding Exogenous Variable (sentiment score)

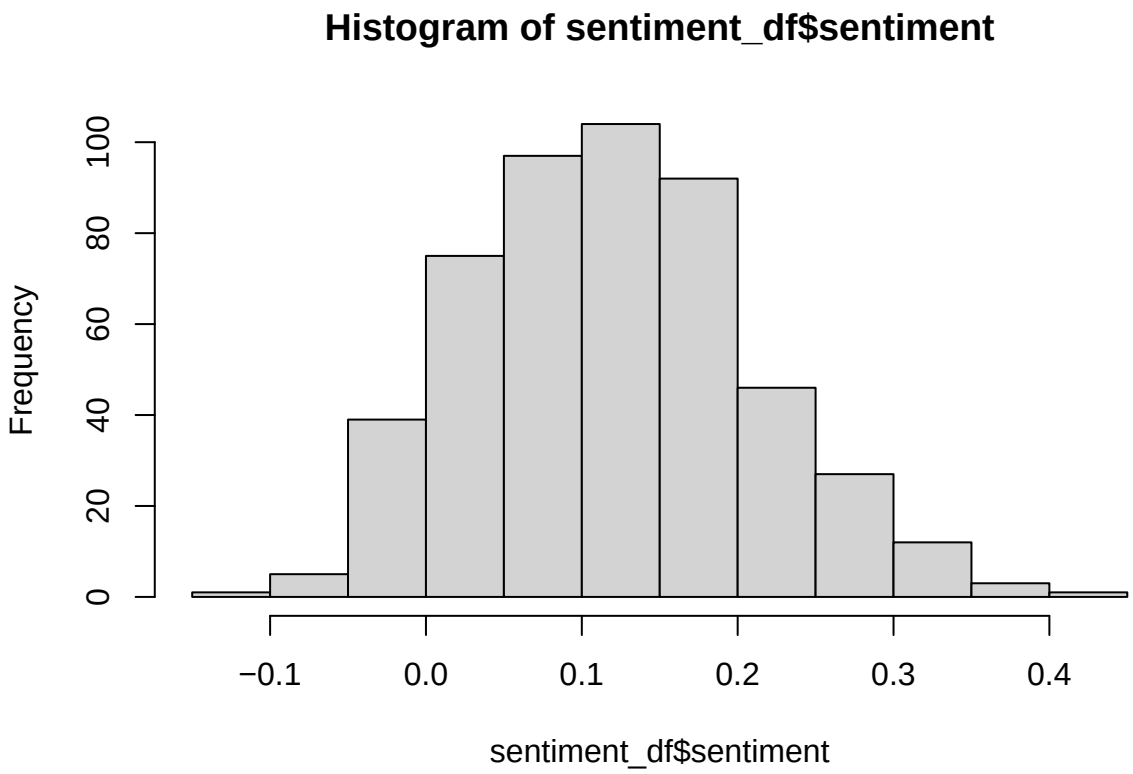
```
sentiment_df <- read.csv("/Users/pranavtavildar/Desktop/Senior-Thesis/data/daily_sentiments.csv", heade
str(sentiment_df)
```

```
## 'data.frame':    503 obs. of  2 variables:
##  $ Date     : chr  "2021-01-04" "2021-01-05" "2021-01-06" "2021-01-07" ...
##  $ sentiment: num  0.1328 0.1231 0.0734 0.0328 0.0576 ...
```

```
str(returns)
```

```
##  num [1:502] 0.02667 0.03168 0.00467 0.00949 0.00569 ...
```

```
sentiment_df = sentiment_df[-1,]
hist(sentiment_df$sentiment)
```




```
output_path <- "/Users/pranavtavildar/Desktop/Senior-Thesis/FrontiersPoster/FrontiersPoster_files/figur

png(file = output_path)
hist(sentiment_df$sentiment, main = "Sentiment Distribution", xlab = "Sentiment Score", ylab = "Frequen
dev.off()
```

```
## pdf
##   2
```

Fitting Arima Model to the new data

```
arma_model <- auto.arima(returns,  max.D = 0, max.d = 0)
arma_model
```

```
## Series: returns
## ARIMA(0,0,0) with non-zero mean
##
## Coefficients:
##         mean
##       0.0015
## s.e.  0.0008
##
## sigma^2 = 0.000339:  log likelihood = 1293.56
## AIC=-2583.12   AICc=-2583.1   BIC=-2574.68
```

```
sentiment_arma_model <- auto.arima(sentiment_df$sentiment,  max.D = 0, max.d = 0)
sentiment_arma_model
```

```
## Series: sentiment_df$sentiment
## ARIMA(0,0,0) with non-zero mean
```

```
##
## Coefficients:
##         mean
##       0.119
## s.e.  0.004
##
## sigma^2 = 0.00785:  log likelihood = 504.85
## AIC=-1005.7   AICc=-1005.67   BIC=-997.26
```

Adjusting Scale for Garchx

```
adj_sentiment <- sentiment_df$sentiment/100
```

Running arimax and Garchx

```
###armax
exarma_model <- arimax(x = returns, xreg = sentiment_df$sentiment, order= c(1,0,1))
exarma_model
```

```
##
## Call:
## arimax(x = returns, order = c(1, 0, 1), xreg = sentiment_df$sentiment)
##
## Coefficients:
##          ar1      ma1  intercept    xreg
##       0.6004  -0.5787     0.0008  0.0059
## s.e.  0.5483   0.5661     0.0014  0.0093
##
## sigma^2 estimated as 0.0003378:  log likelihood = 1293.98,  aic = -2579.95
```

```
###garchx
centered_returns <- scale(returns, center = TRUE, scale = FALSE)
garchx_model <- garchx(centered_returns, xreg=adj_sentiment, order = c(1,1), initial.values = c(coef(ga
garchx_model
```

```
## Warning in sqrt(diag(vcovmat)): NaNs produced
```

```
##
## Date: Tue Apr 11 12:00:53 2023
## Method: normal ML
## Coefficient covariance: ordinary
## Message (nlminb): false convergence (8)
## No. of observations (fitted): 501
## Sample: 1 to 502
##
##                intercept      arch1    garch1        xreg1
## Estimate:   3.309414e-06 0.02427247 0.9658585 9.931742e-11
## Std. Error:          NaN        NaN       NaN 1.542782e-03
##
## Log-likelihood: 1298.663
```

Repeating Process with Weighted Sentiment Score

```
### weighted sentiment score
weighted_df <- read.csv("/Users/pranavtavildar/Desktop/Senior-Thesis/data/weighted_daily_sentiments.csv
str(weighted_df)
```

```
## 'data.frame':    503 obs. of  3 variables:
##  $ Date              : chr  "2021-01-04" "2021-01-05" "2021-01-06" "2021-01-07" ...
##  $ weighted_sentiment: num  0.18486 0.37177 0.22083 0.03277 0.00859 ...
##  $ adjusted_likes    : int  18 99 18 10 67 12 30 56 74 15 ...
```

```
str(returns)
```

```
##  num [1:502] 0.02667 0.03168 0.00467 0.00949 0.00569 ...
```

```
weighted_df = weighted_df[-1,]

###armax
exarma_model <- arimax(x = returns, xreg = weighted_df$weighted_sentiment)
exarma_model
```

```
##
## Call:
## arimax(x = returns, xreg = weighted_df$weighted_sentiment)
##
## Coefficients:
##       intercept     xreg
##          0.0023  -0.0069
## s.e.     0.0011   0.0063
##
## sigma^2 estimated as 0.0003375:  log likelihood = 1294.16,  aic = -2584.32
```

```
###garchx
garchx_model <- garchx(returns, xreg=sentiment_df$weighted_sentiment)
summary(garchx_model)
```

```
##                 Length Class  Mode
## date              1    -none- character
## sys.call          3    -none- call
## y.name            1    -none- character
## y.n               1    -none- numeric
## recursion.n       1    -none- numeric
## y.coredata      502    -none- numeric
## y.index         502    -none- numeric
## y2              502    -none- numeric
## y2mean            1    -none- numeric
## order             3    -none- numeric
## archK             1    -none- numeric
## arch              1    -none- numeric
## archOrder         1    -none- numeric
## garchK            1    -none- numeric
## garch             1    -none- numeric
## garchOrder        1    -none- numeric
## asymK             1    -none- numeric
## asymOrder         1    -none- numeric
## xregK             1    -none- numeric
## maxpqr            1    -none- numeric
## maxpqrpluss1      1    -none- numeric
```

```
## archIndx          1    -none- numeric
## garchIndx         1    -none- numeric
## asymIndx          1    -none- numeric
## xregIndx          1    -none- numeric
## backcast.values   1    -none- numeric
## y2short         501    -none- numeric
## y2matrix        502    -none- numeric
## sigma2          502    -none- numeric
## initial.values    3    -none- numeric
## upper             1    -none- numeric
## lower             1    -none- numeric
## control           0    -none- list
## hessian.control   0    -none- list
## solve.tol         1    -none- numeric
## c.code            1    -none- logical
## sigma2.min        1    -none- numeric
## objective.fun     1    -none- numeric
## penalty.value     1    -none- numeric
## par               3    -none- numeric
## objective         1    -none- numeric
## convergence       1    -none- numeric
## iterations        1    -none- numeric
## evaluations       2    -none- numeric
## message           1    -none- character
## fitted          501    zoo    numeric
## residuals       501    zoo    numeric
## hessian           9    -none- numeric
## vcov              9    -none- numeric
```