"Study of Recursive and Rolling Estimation Schemes by fitting a GARCH (1,1) model to a Financial Time Series"

PROJECT REPORT

submitted by

Akash Chitnis

RUID: 167003791

Department of Mathematics

Rutgers University

submitted to

Professor Norman R. Swanson

Department of Economics, Rutgers University

2nd May, 2016

ECONOMETRICS II

INDEX

A. PREDICTION BASICS

Forecasting is an important area of economics. It is the prediction of future events and conditions and is a key element in service organizations, especially banks, for management decision-making. There are typically two types of events: 1) uncontrollable external events – originating with the national economy, governments, customers and competitors and 2) controllable internal events (e.g., marketing, legal, risk, new product decisions) within the firm.

The need for forecasting stems from the time lag between awareness of an impending event or need and the occurrence of that event. Organizations constantly try to predict economic events and their impact.

The following are a few applications for forecasting modules:

- Forecasting utilization rates for credit cards: build a model based on historical data and use the model to score a current credit card portfolio to determine utilization rates.
- Forecasting financial time series like a stock market index by building a model based on historical data and using it to determine the future trends.
- Model loss rates of a group of home equity lines of credit as a function of time.
- Predicting inflation rates, GDP growth rate, etc.

There are some basic steps for creating a forecast:

- Define the problem. How will the forecasts be used, who needs the forecast and what is the purpose of the forecast?
- Gather the necessary information by obtaining historical (mathematical) data and utilizing the accumulated judgment and expertise of key personnel.
- Determine what graphical plots will best benefit management and design a preliminary analysis. Include descriptive statistics.
- Choose and fit models by using and evaluating a forecast model for decision making. Forecast errors and management response.
- Study the prediction accuracy of the model from forecast errors.

B. FINANCIALTIME SERIES

A time series is a sequence of data points made:
1) over a continuous time interval
2) out of successive measurements across that interval
3) using equal spacing between every two consecutive measurements
4) with each time unit within the time interval having at most one data point

If the data is extracted from financial markets, then such a series is called as Financial Time Series. Time series analysis comprises methods for analyzing time series data in order to extract meaningful statistics and other characteristics of the data. Time series forecasting is the use of a model to predict future values based on previously observed values.

The properties exhibited by a financial time series are as follows:

a. Trend
b. Seasonality
c. Outliers
d. Variance over time
e. Abrupt changes/Jumps

Different tests are developed to detect each of these properties and one can choose a model based on the set of properties exhibited by the time series.

We will study the financial time series *SP500 returns* per day data of year 2014 and 2015. We shall use *Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) model* for this purpose.

C. GARCH (1,1) MODEL

GARCH (p, q) indicates a conditional variance GARCH Model with GARCH degree p and ARCH degree q.
Consider the following GARCH (1,1) model,

$$y_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim iidN(0,1)$$

$$\sigma_t^2 = \omega_0 + \omega_1 \sigma_{t-1}^2 + \omega_2 y_{t-1}^2$$

With $\omega_1 + \omega_2 < 1, \omega_0 > 0, \omega_1 > 0, \omega_2 > 0$

Note that

$$\sigma_t^2 = \sigma_0^2 + \omega_0 \sum_{j=0}^{t-1} \omega_1^j + \omega_2 \sum_{j=0}^{t-1} \omega_1^j \, y_{t-1-j}^2$$

And so σ_t^2 is a measurable function of the past squared returns.

Thus the relevant information set is

$$F_{t-1} = \sigma(y_1, \ldots, y_{t-1}).$$

Now,

$$E(y_t|F_{t-1}) = E(\sigma_t \varepsilon_t|F_{t-1}) = \sigma_t E(\varepsilon_t|F_{t-1}) = 0$$

While,

$$E(y_t^2|F_{t-1}) = var(y_t|F_{t-1}) = E(\sigma_t^2 \varepsilon_t^2|F_{t-1}) = \sigma_t^2 E(\varepsilon_t^2|F_{t-1}) = \sigma_t^2$$

Hence, if the loss function is quadratic, the optimal predictor is

$$E(y_t|F_{t-1}) = 0$$

We shall use the above result for quadratic loss function in our analysis.

D. SAMPLING SCHEMES

There are three main sampling schemes:

i. Fixed Scheme
ii. Recursive Scheme
iii. Rolling Scheme

For out of sample forecasting, we split the sample T into two subsamples, a regression period, with R observations, and a prediction period, with P observations, where $T = R + P$.

a. Fixed Estimation Scheme

In this scheme, we use the first R observations to estimate the parameters, called $\widehat{\beta_R}$, and construct a sequence of P prediction errors, defined as

$$\hat{u}_{0,t+1} = y_{t+1} - \hat{\beta}_{01,R} - \hat{\beta}_{02,R} y_t - \hat{\beta}_{03,R} x_t$$

For $t = R, \dots, R + P - 1$.

b. Recursive Estimation Scheme

In this scheme, we use the first R observations to compute $\widehat{\beta_R}$ and construct the first prediction error:

$$\hat{u}_{0,R+1} = y_{R+1} - \hat{\beta}_{01,R} - \hat{\beta}_{02,R} y_R - \hat{\beta}_{03,R} x_R$$

Then use all observations up to time $R + 1$ to construct $\hat{\beta}_{R+1}$, and get the second prediction error

$$\hat{u}_{0,R+2} = y_{R+2} - \hat{\beta}_{01,R+1} - \hat{\beta}_{02,R+1} y_{R+1} - \hat{\beta}_{03,R+!} x_{R+1}.$$

We proceed in a similar manner until we get a sequence of P prediction errors, defined as:

$$\hat{u}_{0,t+1} = y_{t+1} - \hat{\beta}_{01,t} - \hat{\beta}_{02,t} y_t - \hat{\beta}_{03,t} x_t$$

For $t = R, \dots R + P - 1$

Where $\hat{\beta}_t$ is the estimator computed using observations up to time t.

c. Rolling Estimation Scheme

In this scheme, we use the first R observations to compute $\widehat{\beta_R}$ and construct the first prediction error:

$$\hat{u}_{0,R+1} = y_{R+1} - \hat{\beta}_{01,R} - \hat{\beta}_{02,R} y_R - \hat{\beta}_{03,R} x_R$$

Then we use observations from $t = 2$ up to $t = R + 1$ to construct $\hat{\beta}_{2,R+1}$, and a second prediction error is constructed:

$$\hat{u}_{0,R+2} = y_{R+2} - \hat{\beta}_{01,2,R+1} - \hat{\beta}_{02,2,R+1} y_{R+1} - \hat{\beta}_{03,2,R+1} x_{R+1}$$

We proceed in a similar manner using the most recent R observations, until we get a sequence of P prediction errors, defined as:

$$\hat{u}_{0,t+1} = y_{t+1} - \hat{\beta}_{01,t-R+1,t} - \hat{\beta}_{02,t-R+1,t} y_t - \hat{\beta}_{03,t-R+1,t,t} x_t$$

For $t = R, \dots R + P - 1$

Where $\hat{\beta}_{t-R+1,t}$ is the estimator computed using observations from time $t - R + 1$ up to time t.

Intuitively it makes sense to use the information contained in new observation as soon as it becomes available. However, one must be aware of structural breaks due to changing data definitions, changing model specifications, etc.

We shall apply the recursive and rolling estimation schemes to the financial time series (SP500) and study the results in the following section.

E. MEHODOLOGY

As from the previous discussion, it is clear that we will be using GARCH(1,1) model for predicting the time series data (503 observations).

The methodology is explained in the following steps:

Step 1: Check for stationarity by performing Augmented Dickey-Fuller Test for stationarity on SP500 index.

Step 2: If the series exhibits non-stationarity, which it does, take the log differences of the series and confirm stationarity by again performing Augmented Dickey-Fuller Test on this newly obtained series.

Step 3: Check for autocorrelation by performing Ljung-Box test on the new time series.

Step 4: Fit a GARCH (1,1) model to the regression period observations (300 in our case) and note the coefficients.

Step 5: Use recursive scheme to Forecast the next value (horizon = 1) by using the calculated coefficients. Since we have the true values for that period, compute the true error by subtracting predicted value from true value. For the computation, consider the true parameter value and not the estimated value. Proceed similarly for the entire prediction period.

Step 6: Compute the mean square forecast error for the recursive scheme by using the quadratic loss function formula

$$MSFE_{Recursive} = \frac{1}{P}\sum_{i=1}^{P} \hat{\varepsilon}_{t+i}^{2}$$

Where $t = current\ time$

Step 7: Use rolling scheme to Forecast the next value (horizon = 1) by using the calculated coefficients. Consider the rolling window size as the entire regression period (in our case 300 observations). Since we have the true values for that period, compute the true error by subtracting predicted value from true value. For the computation, consider the true parameter value and not the estimated value. Proceed similarly for the entire prediction period.

Step 8: Compute the mean square forecast error for the rolling scheme by using the quadratic loss function formula

$$MSFE_{Rolling} = \frac{1}{P}\sum_{i=1}^{P} \hat{\varepsilon}_{t+i}^{2}$$

Where $t = current\ time$

F. RESULTS

Table 1: Test Results for stationarity and autocorrelation of series.

Sr. No	Test	Series	Null Hypothesis	Result
1	Augmented Dickey-Fuller for stationarity	SP500	Not Rejected	Non-Stationary
2	Augmented Dickey-Fuller for stationarity	Difference of Log returns of SP500	Rejected	Stationary
3	Ljung-Box Test for Autocorrelation	Difference of Log returns of SP500	Rejected	Autocorrelation exists

Table 2: GARCH (1,1) Model Specifications

Mean: ARMAX (0,0,0)
Variance: GARCH (1,1)

Conditional Probability Distribution: Gaussian

Number of Model Parameters Estimated: 4

Sr. No	Parameter	Value	Standard Error	T Statistic
1	C	0.00063477	0.00043461	1.4605
2	K	6.5231e-06	3.2678e-06	1.9961
3	GARCH (1)	0.68269	0.10806	6.3176
4	ARCH (1)	0.21132	0.070664	2.9905

Table 3:Mean squared error values for Recursive and Rolling Scheme

Sr. No	**Parameter**	**Value**
1	Mean Squared Forecast Error for Recursive Scheme	9.9583e-05
2	Mean Squared Forecast Error for Rolling Scheme	9.9620e-05

G. OBSERVATIONS

The following observations are made by studying the above results:

1. To perform a prediction analysis on a financial time series, one needs to transform the series by taking difference of log returns to eliminate non-stationarity.
2. Most of the financial time series data exhibits autocorrelation and so it is viable to use GARCH (1, 1) model as parsimonious models tend to perform fairly well.
3. The mean squared forecast error (MSFE) obtained from both the schemes is close to zero which indicates both the schemes have strong predicting abilities for this particular series.
4. Recursive estimation scheme slightly outperforms rolling window estimation scheme for this series. In general, this can be seen from the fact that recursive scheme uses more observations as compared to the rolling scheme for the given prediction period.
5. One needs to be aware of the fact that rolling scheme may perform well in case of structural breaks or jumps seen in the early part of regression period.

H. CONCLUSION

After studying the mean squared forecast errors obtained from recursive scheme and rolling scheme, one can conclude that both schemes display fairly accurate predictions with recursive scheme performing marginally better than the rolling scheme.

I. BIBLIOGRAPHY

1. Lecture Notes: *Prediction and Simulation Based Specification Testing and Model Selection*, Valentina Corradi and Norman Rasmus Swanson
2. *Some Properties of Time Series Data and their Use in Econometric Model Specification,* C.W.J Granger, Journal of Econometrics 16 (1981) 121-130. North-Holland and Publishing Company
3. *Wikipedia and Internet*

J. MATLAB REPORT

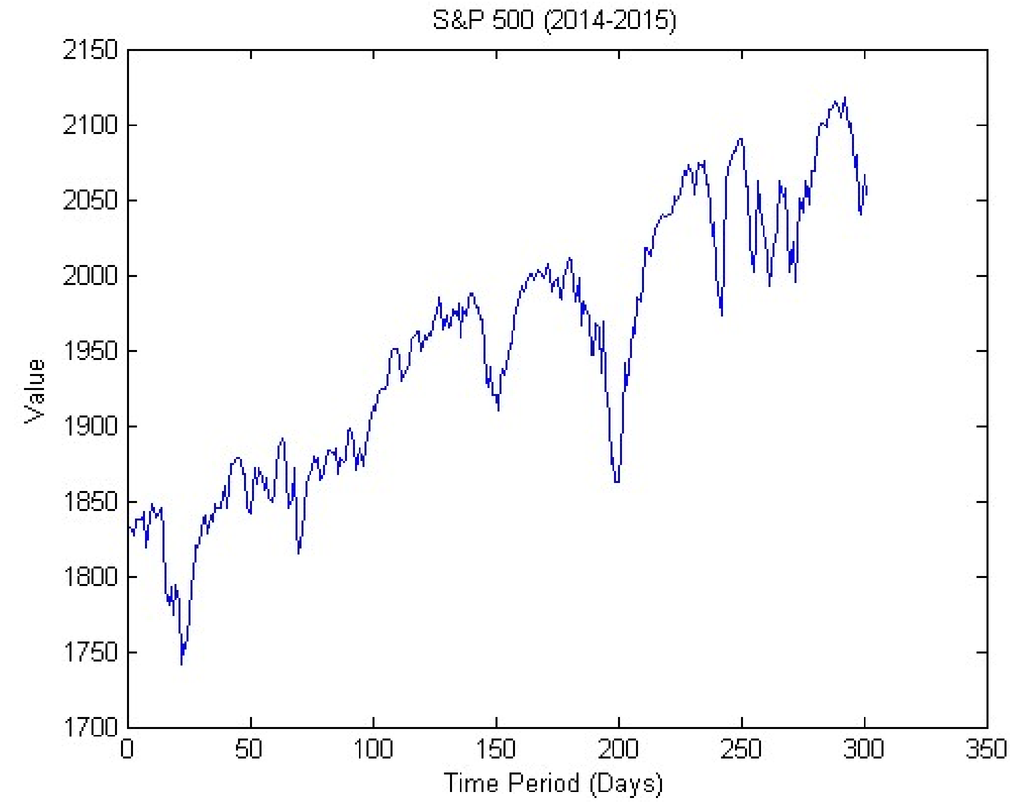
MATLAB CODE

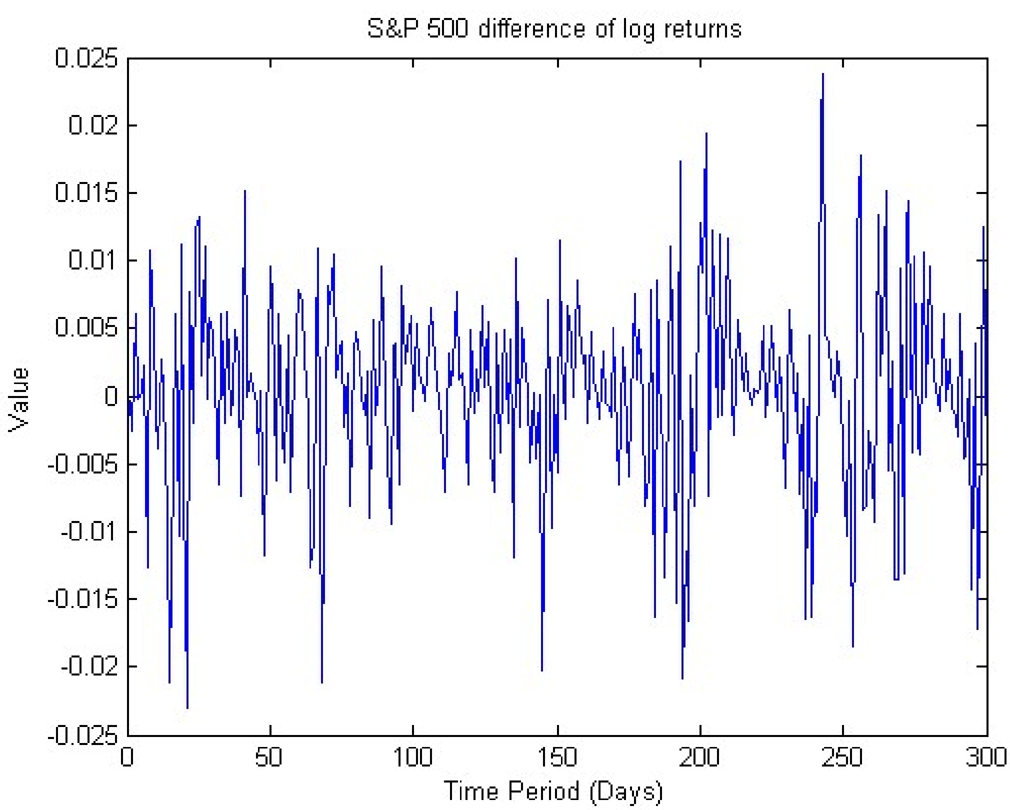
Study of Recursive and Rolling Estimation Schemes for Forecasting

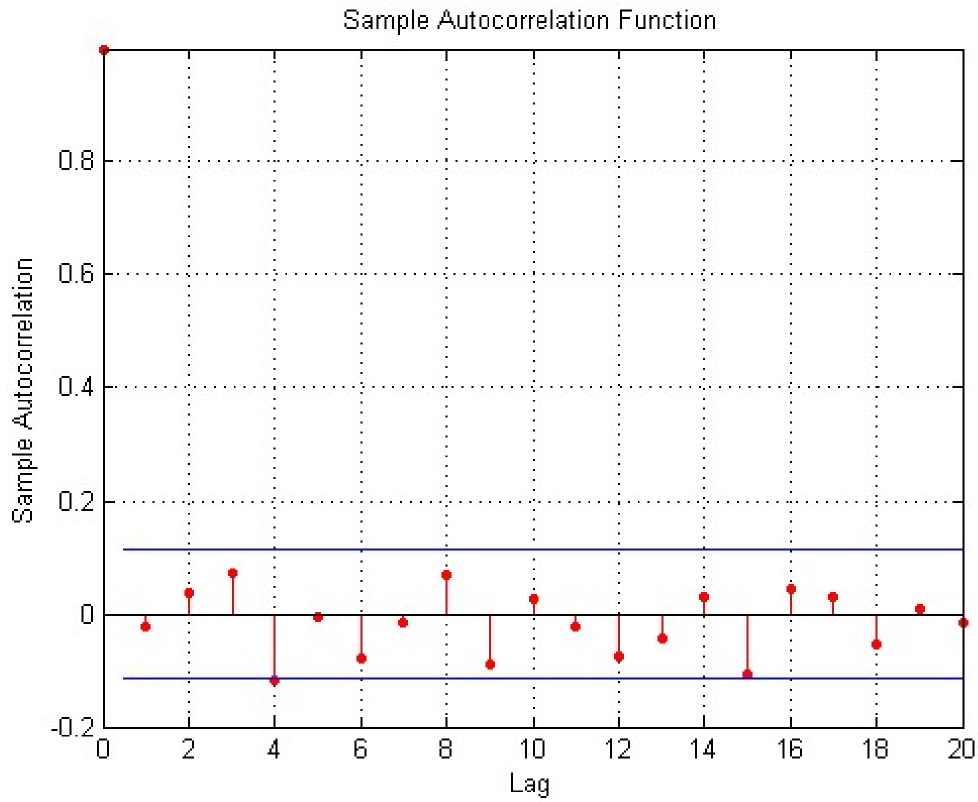
```
clc;
clear all;
warning('off','all');
A = importdata('SP500.csv');
true_delta = diff(log(A));
SP500 = A(1:301);
delta_y = diff(log(SP500));
delta_y_rolling = delta_y;

plot(SP500)
title('S&P 500 (2014-2015)')
xlabel('Time Period (Days)')
ylabel('Value')
figure
plot(delta_y)
title('S&P 500 difference of log returns')
xlabel('Time Period (Days)')
ylabel('Value')
figure

%Augmented Dickey-Fuller Test for stationarity:
fprintf('\n');
display('Augmented Dickey-Fuller Test for SP500 series:');
if (adftest(SP500))==0
display('Null hypothesis is not rejected.');
display('Data Series is non-stationary.');
fprintf('\n');
end

%Augmented Dickey-Fuller Test for stationarity:
fprintf('\n');
display('Augmented Dickey-Fuller Test for SP500 difference of log returns series:');
if (adftest(delta_y))==1
display('Null hypothesis is rejected.');
display('Log Returns Series is stationary.');
fprintf('\n');
end

%Test for Autocorrelation of difference of log returns
fprintf('\n');
autocorr(delta_y);
figure
display('Test for Autocorrelation of difference of log returns');
fprintf('\n');
display('Ljung-Box Test');
display(lbqtest(delta_y));
```

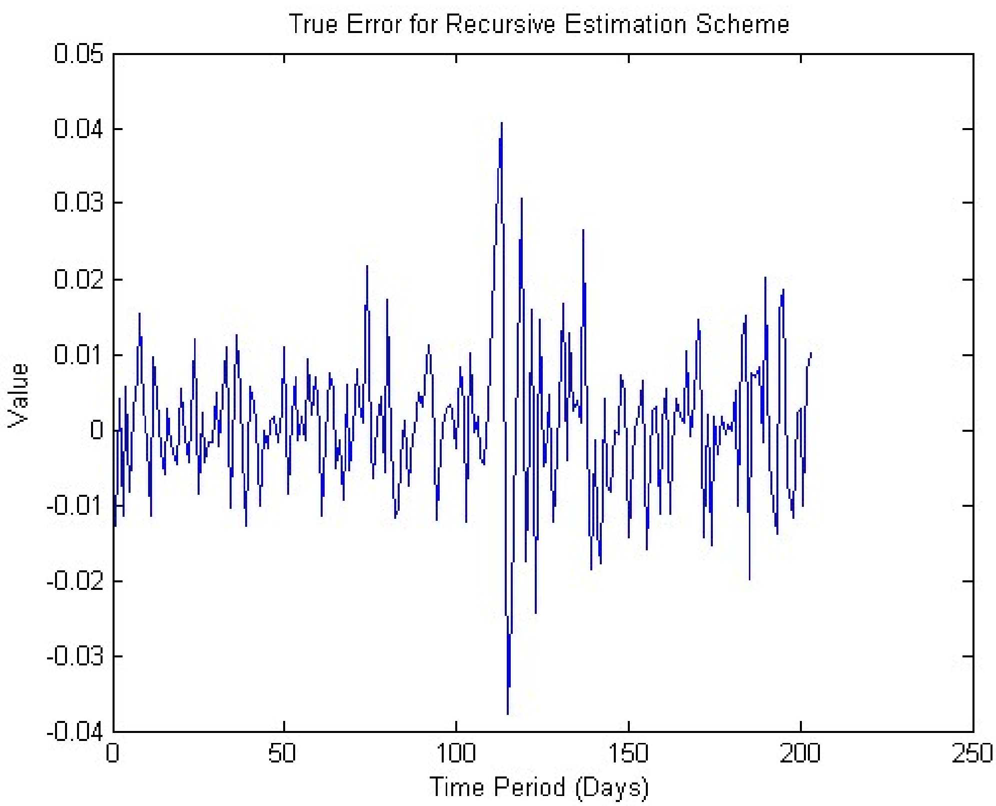
```
if (lbqtest(delta_y))==0
display('Null is rejected');
display('Returns exhibit autocorrelation for a fixed number of lags');
end
spec = garchset('VarianceModel','GARCH','P',1,'Q',1);
spec = garchset(spec, 'Distribution','Gaussian','Display','off');
[coeff, errors, LLF, efit, sFit] = garchfit(spec, delta_y);
garchdisp(coeff,errors);

%Recursive Estimation Method:
horizon = 1;
Fore_values = [];
True_error = [];
for i=1:203
    %Step 1: Forecast
        [sigmaForecast, meanForecast] = ...
       garchpred(coeff,delta_y, horizon);
        True_error = vertcat(True_error, meanForecast - true_delta(300+i));

    %Step 2: Recalculate GARCH using true value for last observation:
        delta_y = vertcat(delta_y, true_delta(300+i));
        [coeff, errors, LLF, efit, sFit] = garchfit(spec, delta_y);
end
MSFE_Recursive = (mean(True_error.^2));
fprintf('\n\n\n');
display(MSFE_Recursive);
plot(True_error)
title('True Error for Recursive Estimation Scheme')
xlabel('Time Period (Days)')
ylabel('Value')
figure

%Rolling Estimation Method:
Actual_error =[];
window_length = 300;
for j = 1:203
%Step 1: Forecast
        [sigmaForecast, meanForecast] = ...
       garchpred(coeff,delta_y_rolling, horizon);
        Actual_error = vertcat(Actual_error, meanForecast - true_delta(300+j));

%Step 2: Recalculate GARCH using true value for last observation:
        delta_y_rolling = vertcat(delta_y_rolling, true_delta(300+j));
        rolling_y = createRollingWindow(delta_y_rolling, window_length);
        roll_y = rolling_y';
        [coeff, errors, LLF, efit, sFit] = garchfit(spec, roll_y);
end
MSFE_Rolling = (mean(Actual_error.^2));
display(MSFE_Rolling);
plot(Actual_error)
title('True Error for Rolling Estimation Scheme')
xlabel('Time Period (Days)')
ylabel('Value')
```

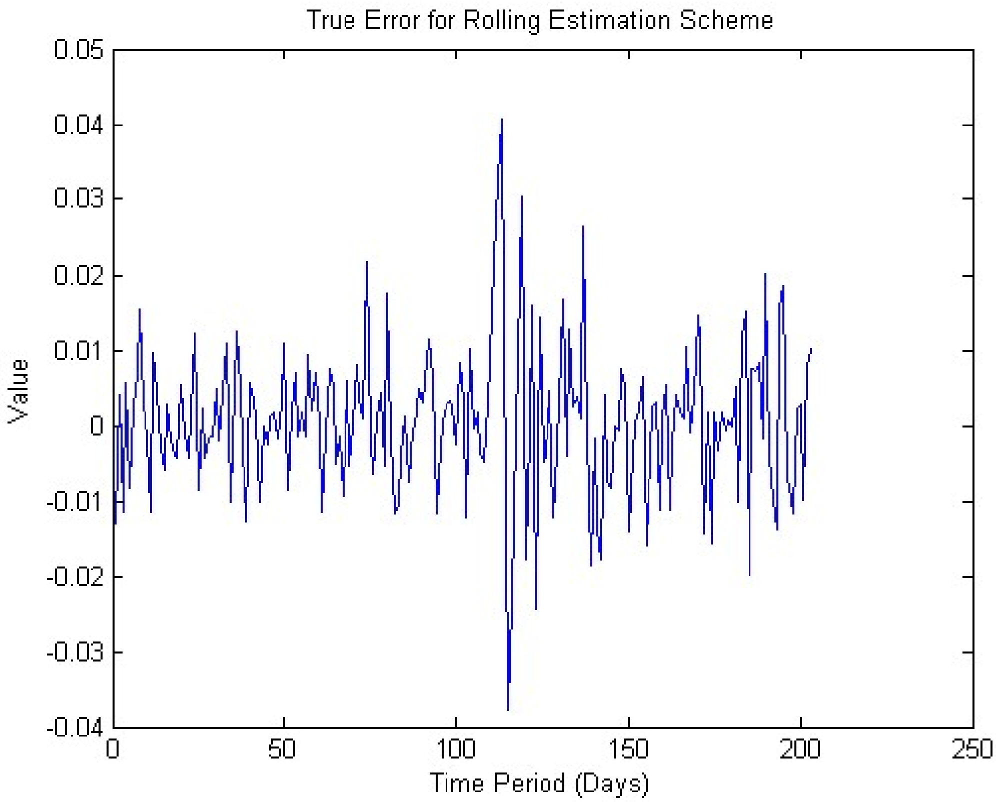
```
Augmented Dickey-Fuller Test for SP500 series:
Null hypothesis is not rejected.
Data Series is non-stationary.


Augmented Dickey-Fuller Test for SP500 difference of log returns series:
Null hypothesis is rejected.
Log Returns Series is stationary.


Test for Autocorrelation of difference of log returns

Ljung-Box Test

ans =

     0

Null is rejected
Returns exhibit autocorrelation for a fixed number of lags

  Mean: ARMAX(0,0,0); Variance: GARCH(1,1)

  Conditional Probability Distribution: Gaussian
  Number of Model Parameters Estimated: 4

                               Standard          T
  Parameter       Value          Error       Statistic
 -----------   -----------   ------------   -----------
          C    0.00063477     0.00043461       1.4605
          K    6.5231e-06     3.2678e-06       1.9961
   GARCH(1)    0.68269        0.10806          6.3176
    ARCH(1)    0.21132        0.070664         2.9905




MSFE_Recursive =

   9.9583e-05


MSFE_Rolling =

   9.9620e-05
```


S&P 500 (2014-2015)
Value
Time Period (Days)
2150
2100
2050
2000
1950
1900
1850
1800
1750
1700
0
50
100
150
200
250
300
350


S&P 500 difference of log returns
Value
Time Period (Days)
0.025
0.02
0.015
0.01
0.005
0
-0.005
-0.01
-0.015
-0.02
-0.025
0
50
100
150
200
250
300


Sample Autocorrelation Function
Sample Autocorrelation
0.8
0.6
0.4
0.2
0
-0.2
0
2
4
6
8
10
12
14
16
18
20
Lag


True Error for Recursive Estimation Scheme
Value
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
0
50
100
150
200
250
Time Period (Days)


True Error for Rolling Estimation Scheme
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
Value
0
50
100
150
200
250
Time Period (Days)