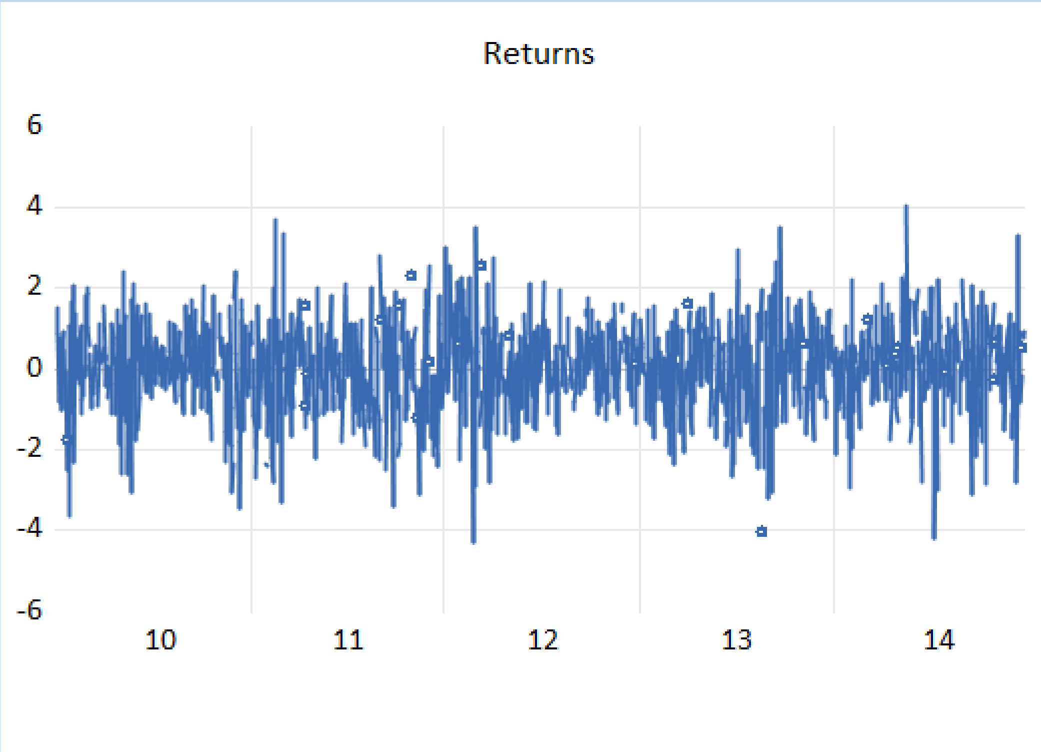
TIME SERIES FORECASTING

Log returns in percentage

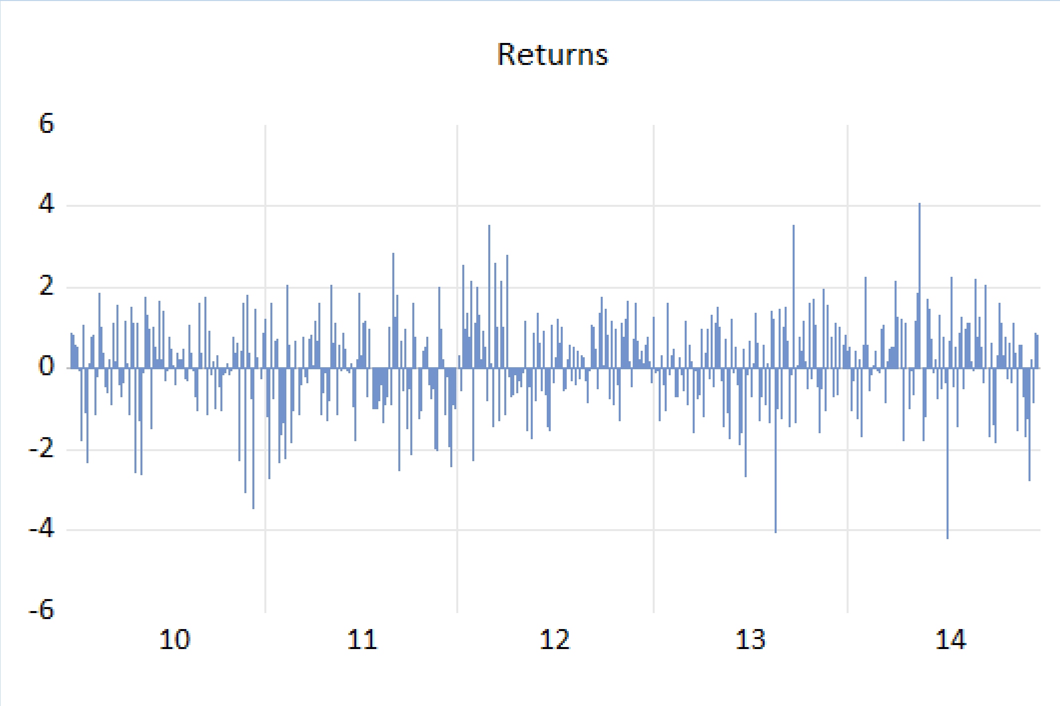
Formula used: Difference in the natural logarithm * 100

- Results obtained for the data:

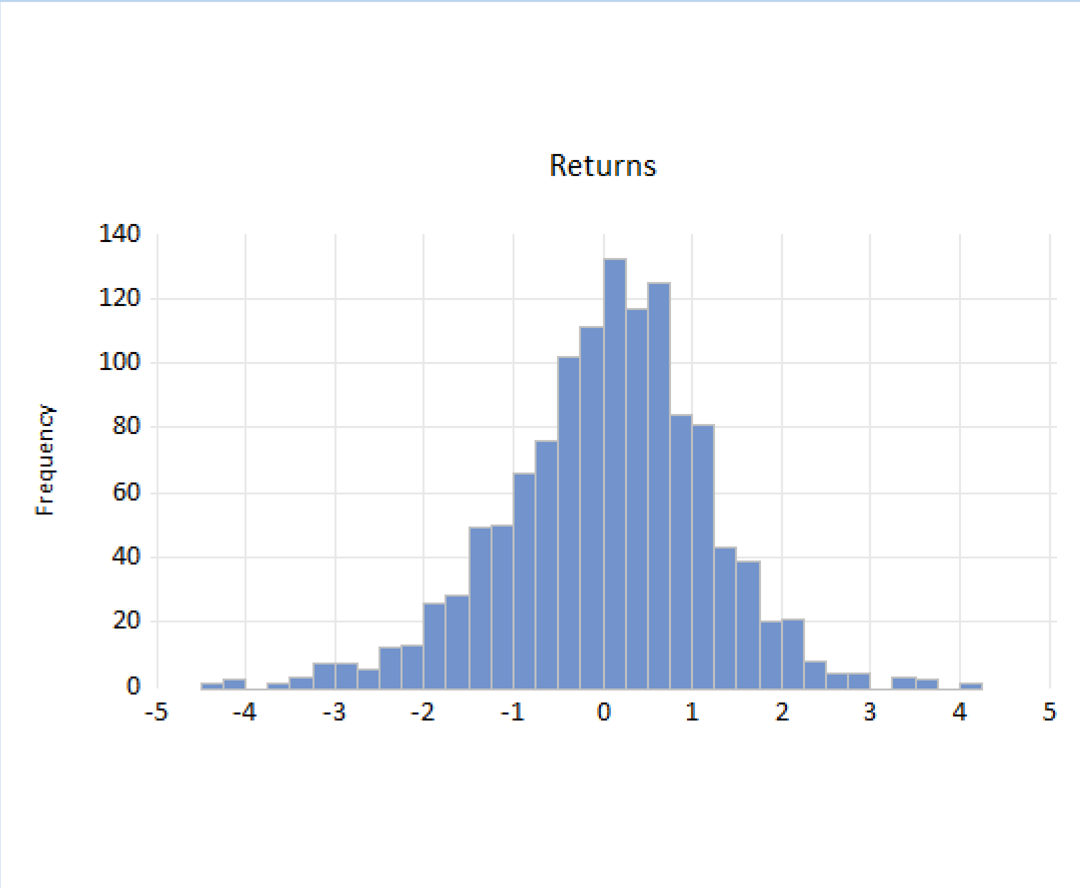
- **Log returns line graph:**



- **Log returns bar graph:**



- **Log returns histogram distribution:**

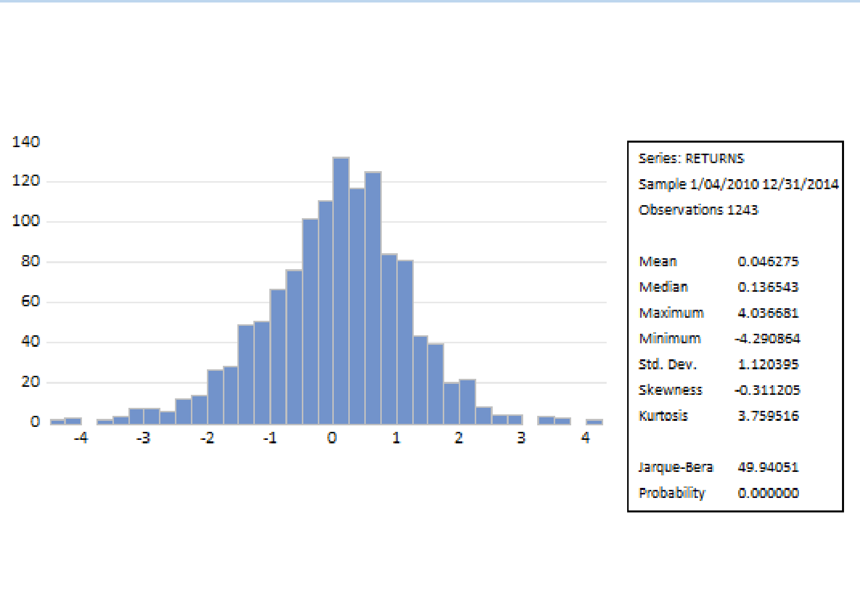


Google drive link of the data and calculated log returns
https://drive.google.com/file/d/1NhPdH6YxSgA_QwOcX-HhXswkE0TaP856/view?usp=sharing

1) Descriptive statistics:

- Results obtained from the data:

- **Histogram and stats:**



- **Stats table:**

Series: RETURNS Workfile: NIFTY NEXT 50 2010-2014::U...

	RETURNS
Mean	0.046275
Median	0.136543
Maximum	4.036681
Minimum	-4.290864
Std. Dev.	1.120395
Skewness	-0.311205
Kurtosis	3.759516
Jarque-Bera	49.94051
Probability	0.000000
Sum	57.51934
Sum Sq. Dev.	1559.064
Observations	1243

Discussion:

From the results, it is seen that it is negatively skewed and the value of kurtosis is higher than 3, so the distribution is leptokurtic,

2) AR, MA, ARMA/ARIMA modelling:

Steps used in the assignment:

Box-Jenkins Methodology is followed,

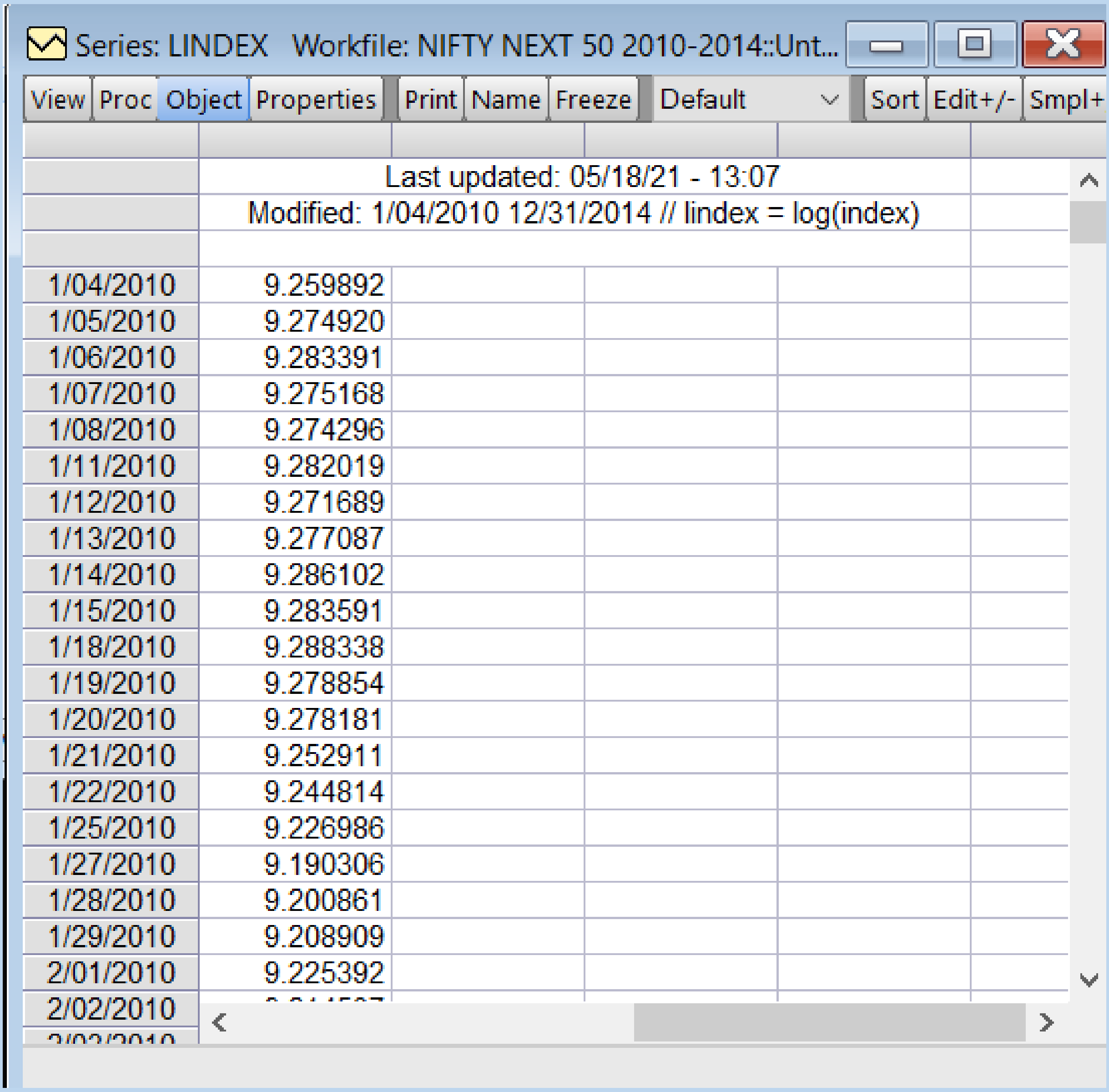
1) Identification:
    - Converting to log series
    - Unit root test at level and difference to check whether to reject the null hypothesis
    - Choosing tentative p,d,q
2) Estimation:
    - Estimating the values of p, d, q using automatic ARIMA forecasting
    - Selecting five models from the ARMA criteria table
3) Diagnostic checking:
    - Selecting the best fit model using the tests
    - Using normality test (residual diagnostics)
    - Using serial correlation test
    - Using heteroskedasticity test (ARCH)
    - Checking R squared

- Checking significant terms in the model

Based on these factors, the appropriate AR, MA, ARMA, ARIMA model is chosen

AR, MA, AR/MA Model

- **Log series of the data**



Series: LINDEX Workfile: NIFTY NEXT 50 2010-2014::Unt...

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smpl+

Last updated: 05/18/21 - 13:07

Modified: 1/04/2010 12/31/2014 // lindex = log(index)

1/04/2010	9.259892
1/05/2010	9.274920
1/06/2010	9.283391
1/07/2010	9.275168
1/08/2010	9.274296
1/11/2010	9.282019
1/12/2010	9.271689
1/13/2010	9.277087
1/14/2010	9.286102
1/15/2010	9.283591
1/18/2010	9.288338
1/19/2010	9.278854
1/20/2010	9.278181
1/21/2010	9.252911
1/22/2010	9.244814
1/25/2010	9.226986
1/27/2010	9.190306
1/28/2010	9.200861
1/29/2010	9.208909
2/01/2010	9.225392
2/02/2010	[illegible]

Discussion:

Log series of the data is generated to perform AR, MA modelling

- **Unit root test**

Phillips-Perron Unit Root Test on LINDEX

Null Hypothesis: LINDEX has a unit root
Exogenous: Constant
Bandwidth: 4 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		0.085681	0.9646
Test critical values:	1% level	-3.435398	
	5% level	-2.863657	
	10% level	-2.567947	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000125
HAC corrected variance (Bartlett kernel)	0.000154

Phillips-Perron Test Equation
Dependent Variable: D(LINDEX)
Method: Least Squares
Date: 05/18/21 Time: 13:13
Sample (adjusted): 1/05/2010 12/31/2014
Included observations: 1243 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LINDEX(-1)	0.000680	0.001890	0.359762	0.7191
C	-0.005908	0.017710	-0.333575	0.7388

R-squared	0.000104	Mean dependent var	0.000463
Adjusted R-squared	-0.000701	S.D. dependent var	0.011204
S.E. of regression	0.011208	Akaike info criterion	-6.142792
Sum squared resid	0.155890	Schwarz criterion	-6.134545
Log likelihood	3819.745	Hannan-Quinn criter.	-6.139691
F-statistic	0.129429	Durbin-Watson stat	1.696745
Prob(F-statistic)	0.719086		

Discussion:

Using the Philips – Perron Unit root test, It is estimated that it has a unit root

- **At difference 1,**

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ident

Phillips-Perron Unit Root Test on D(LINDEX)

Null Hypothesis: D(LINDEX) has a unit root
Exogenous: Constant
Bandwidth: 8 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-30.06895	0.0000
Test critical values:	1% level	-3.435402	
	5% level	-2.863659	
	10% level	-2.567948	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000122
HAC corrected variance (Bartlett kernel)	0.000112

Phillips-Perron Test Equation
Dependent Variable: D(LINDEX,2)
Method: Least Squares
Date: 05/18/21 Time: 13:21
Sample (adjusted): 1/06/2010 12/31/2014
Included observations: 1242 after adjustments

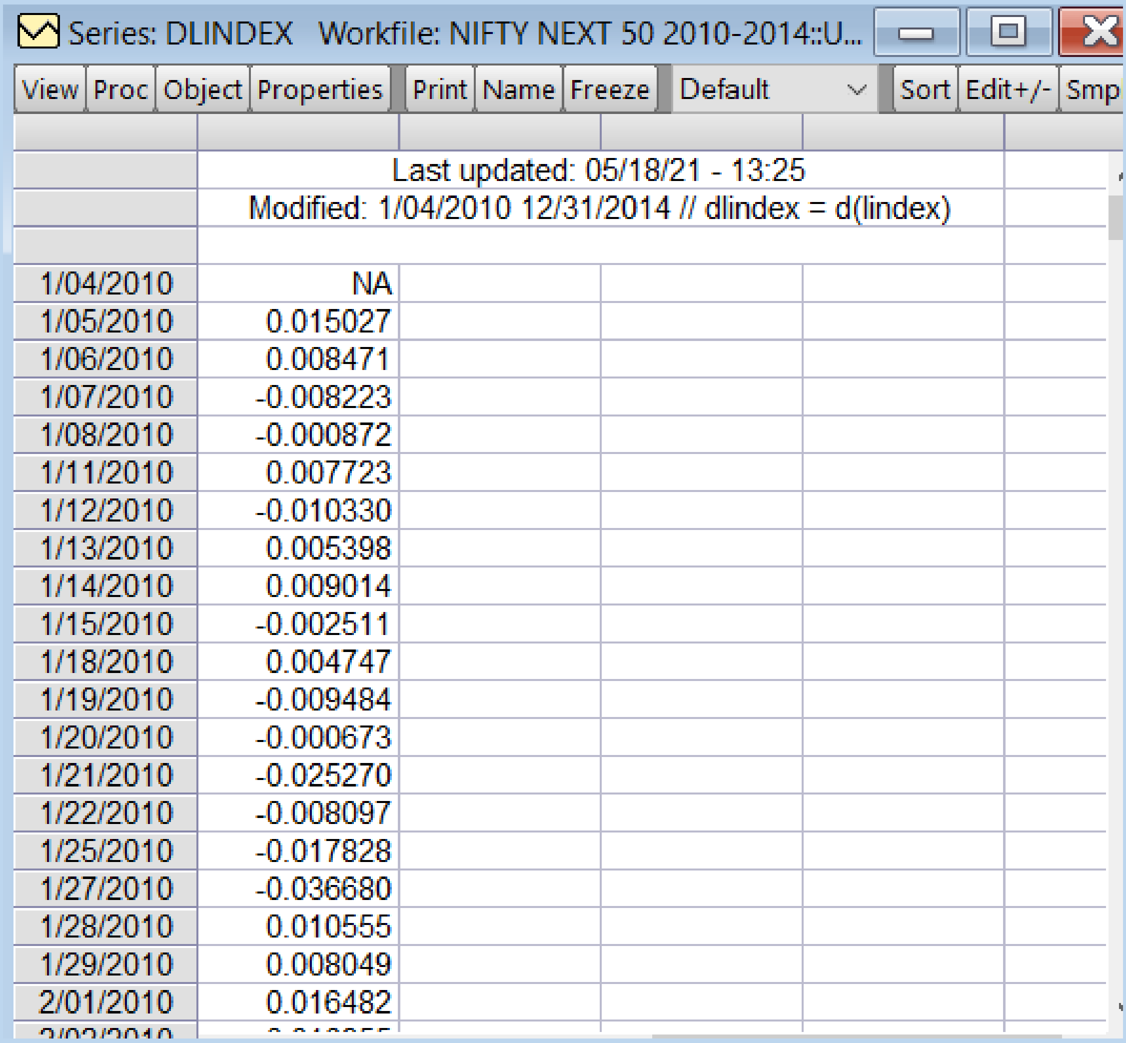
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(LINDEX(-1))	-0.848513	0.028056	-30.24362	0.0000
C	0.000382	0.000315	1.213921	0.2250

R-squared	0.424508	Mean dependent var	-5.72E-06
Adjusted R-squared	0.424044	S.D. dependent var	0.014594
S.E. of regression	0.011076	Akaike info criterion	-6.166482
Sum squared resid	0.152118	Schwarz criterion	-6.158230
Log likelihood	3831.385	Hannan-Quinn criter.	-6.163379
F-statistic	914.6767	Durbin-Watson stat	2.001891
Prob(F-statistic)	0.000000		

Discussion:

On estimating the difference of the series, It is able to reject null hypothesis as the probability is 0% and the series becomes stationary
Therefore, d = 1
So, ARIMA = (p, 1, q)
ARIMA model is formulated from the data

- **Difference the series of log (data) to convert ARIMA to ARMA**



Series: DLINDEX Workfile: NIFTY NEXT 50 2010-2014::U...

View | Proc | Object | Properties | Print | Name | Freeze | Default | Sort | Edit+/- | Smp

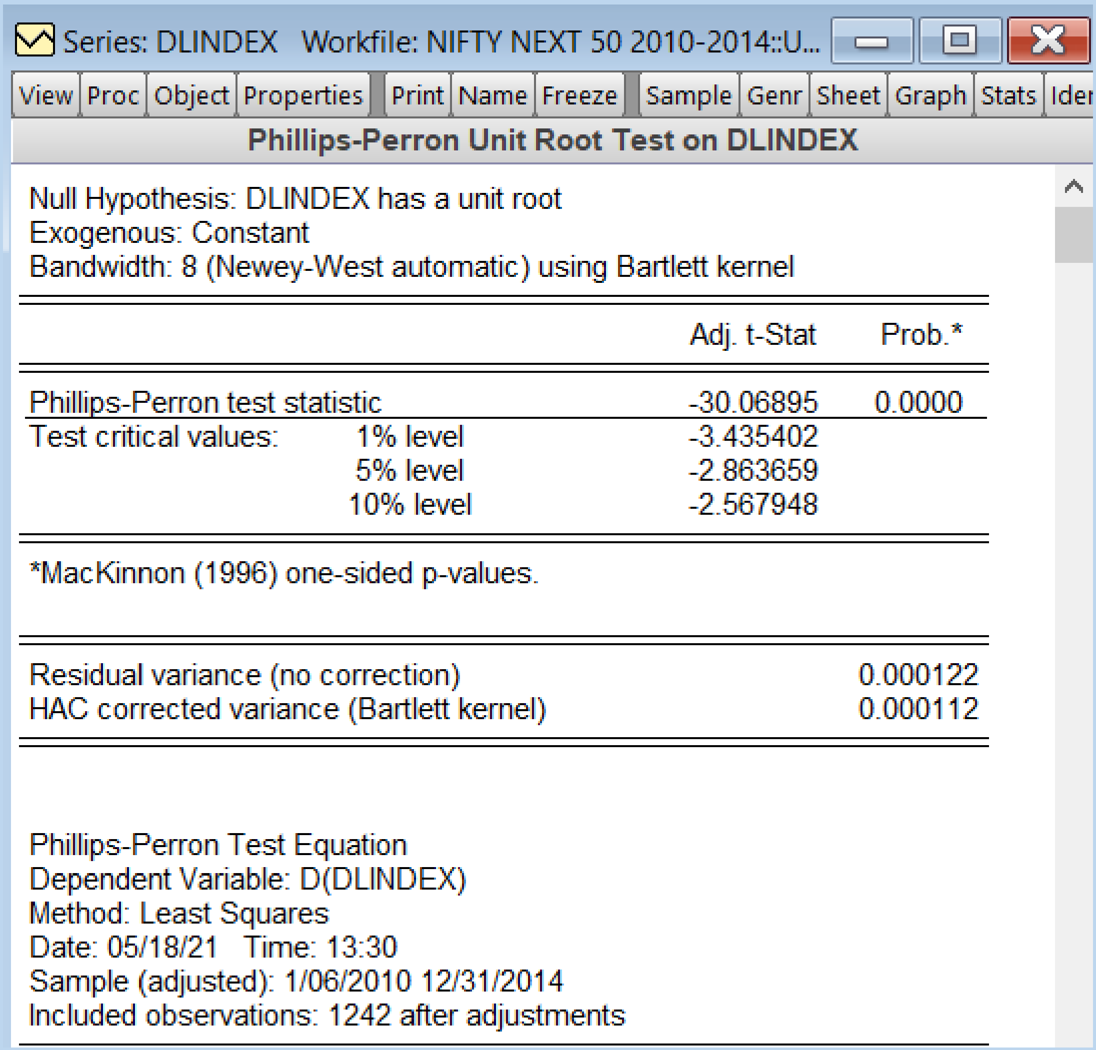
Last updated: 05/18/21 - 13:25
Modified: 1/04/2010 12/31/2014 // dlindex = d(lindex)

1/04/2010	NA
1/05/2010	0.015027
1/06/2010	0.008471
1/07/2010	-0.008223
1/08/2010	-0.000872
1/11/2010	0.007723
1/12/2010	-0.010330
1/13/2010	0.005398
1/14/2010	0.009014
1/15/2010	-0.002511
1/18/2010	0.004747
1/19/2010	-0.009484
1/20/2010	-0.000673
1/21/2010	-0.025270
1/22/2010	-0.008097
1/25/2010	-0.017828
1/27/2010	-0.036680
1/28/2010	0.010555
1/29/2010	0.008049
2/01/2010	0.016482

Discussion:

The difference of the log series is generated to convert the formulated ARIMA model to ARMA model.

- **Unit root test of d(lindex):**



Series: DLINDEX Workfile: NIFTY NEXT 50 2010-2014::U...

View | Proc | Object | Properties | Print | Name | Freeze | Sample | Genr | Sheet | Graph | Stats | Ider

Phillips-Perron Unit Root Test on DLINDEX

Null Hypothesis: DLINDEX has a unit root
Exogenous: Constant
Bandwidth: 8 (Newey-West automatic) using Bartlett kernel

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-30.06895	0.0000
Test critical values:	1% level	-3.435402	
	5% level	-2.863659	
	10% level	-2.567948	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.000122
HAC corrected variance (Bartlett kernel)	0.000112

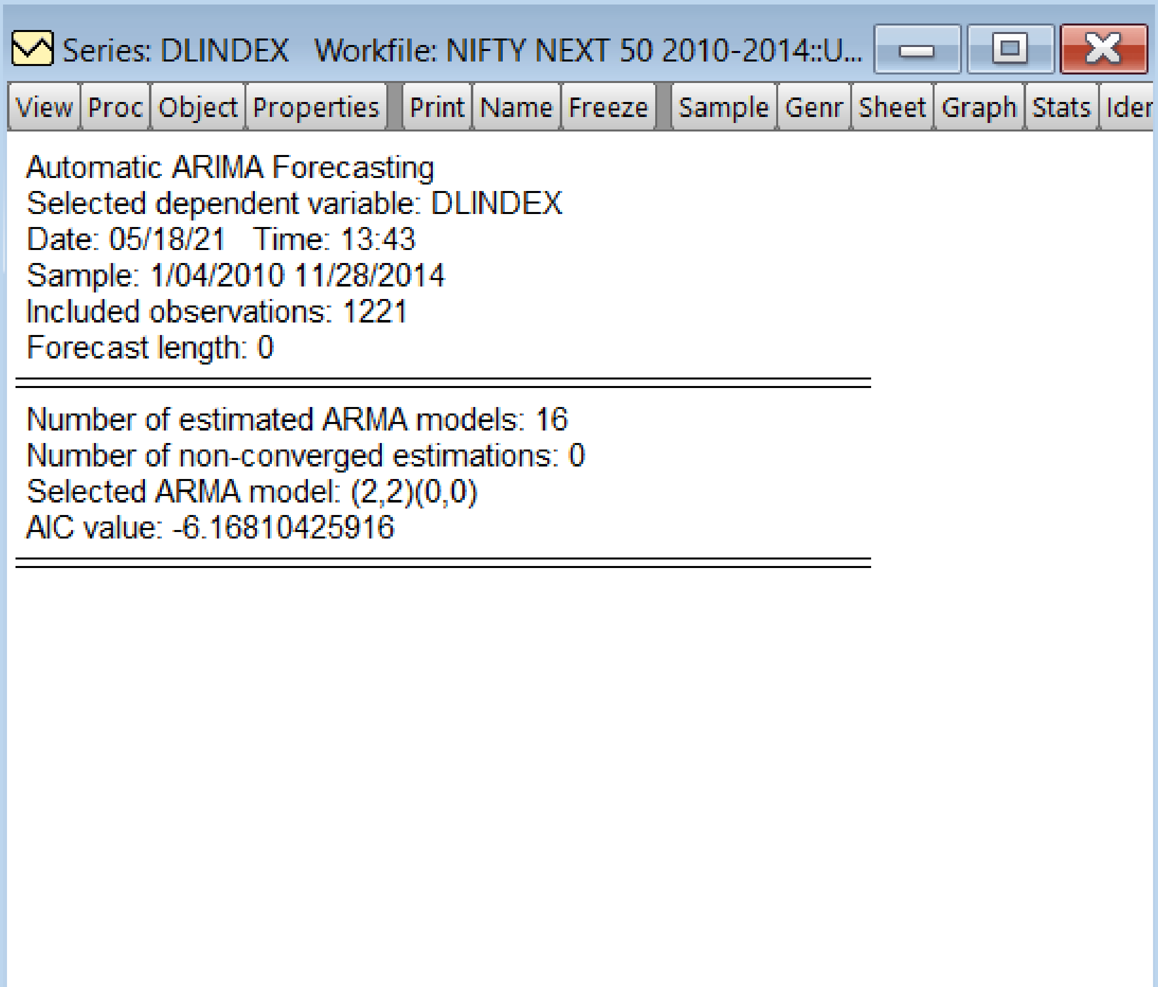
Phillips-Perron Test Equation
Dependent Variable: D(DLINDEX)
Method: Least Squares
Date: 05/18/21 Time: 13:30
Sample (adjusted): 1/06/2010 12/31/2014
Included observations: 1242 after adjustments

Discussion:

On testing the unit root, null hypothesis is rejected as the probability is 0%.

- **Automatic ARIMA forecasting**

This is the ARMA model(p, q) of dlindex



Discussion:

From the results of automatic ARIMA forecasting it can be seen that,
ARIMA(2, 1, 2) for Lindex is equivalent to ARMA(2, 2) for DLindex

d = 1

p = 2

q = 2

Estimated ARIMA Model = ARIMA(2, 1, 2)

Estimated ARMA Model = ARMA(2, 2)

- **ARMA criteria table:**

To find out the best fitting models

⊟ Summary

Automatic ARIMA Forecasting
Selected dependent variable: DLINDEX
Date: 05/18/21 Time: 13:52
Sample: 1/04/2010 12/31/2014
Included observations: 1243
Forecast length: 0

Number of estimated ARMA models: 16
Number of non-converged estimations: 0
Selected ARMA model: (2,2)(0,0)
AIC value: -6.16447548953

⊟ ARMA Criteria Table

Model Selection Criteria Table
Dependent Variable: DLINDEX
Date: 05/18/21 Time: 13:52
Sample: 1/04/2010 12/31/2014
Included observations: 1243

Model	LogL	AIC*	BIC	HQ
(2,2)(0,0)	3837.221517	-6.164475	-6.139736	-6.155172
(1,0)(0,0)	3834.112113	-6.164299	-6.151930	-6.159648
(0,1)(0,0)	3833.616640	-6.163502	-6.151132	-6.158851
(2,3)(0,0)	3837.599556	-6.163475	-6.134612	-6.152621
(3,2)(0,0)	3837.578018	-6.163440	-6.134577	-6.152586
(0,2)(0,0)	3834.232981	-6.162885	-6.146392	-6.156683
(2,0)(0,0)	3834.123081	-6.162708	-6.146215	-6.156506
(1,1)(0,0)	3834.120148	-6.162703	-6.146210	-6.156501
(3,3)(0,0)	3837.889673	-6.162333	-6.129346	-6.149928
(3,1)(0,0)	3835.694189	-6.162018	-6.137278	-6.152715
(2,1)(0,0)	3834.557729	-6.161798	-6.141182	-6.154046
(3,0)(0,0)	3834.551426	-6.161788	-6.141172	-6.154036
(0,3)(0,0)	3834.233901	-6.161277	-6.140661	-6.153525
(1,2)(0,0)	3834.233162	-6.161276	-6.140660	-6.153524
(1,3)(0,0)	3834.933642	-6.160794	-6.136054	-6.151491
(0,0)(0,0)	3819.680271	-6.142687	-6.134441	-6.139586

Discussion:

Using the ARMA table,

On testing the first 5 models from the ARMA criteria table to find the best fitting model,

(2, 2) – Model1

(1, 0) – Model2

(0, 1) – Model3

(2, 3) – Model4

(3, 2) – Model5

Model 1:

Dependent Variable: DLINDEX
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 05/18/21 Time: 14:02
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Convergence achieved after 18 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000463	0.000351	1.318634	0.1875
AR(1)	1.033308	0.171297	6.032256	0.0000
AR(2)	-0.557999	0.109483	-5.096649	0.0000
MA(1)	-0.893094	0.177982	-5.017881	0.0000
MA(2)	0.458242	0.121072	3.784869	0.0002
SIGMASQ	0.000122	4.21E-06	28.93742	0.0000

R-squared	0.027866	Mean dependent var	0.000463
Adjusted R-squared	0.023937	S.D. dependent var	0.011204
S.E. of regression	0.011069	Akaike info criterion	-6.164475
Sum squared resid	0.151562	Schwarz criterion	-6.139736
Log likelihood	3837.222	Hannan-Quinn criter.	-6.155172
F-statistic	7.091723	Durbin-Watson stat	1.984795
Prob(F-statistic)	0.000002		

Inverted AR Roots	.52-.54i	.52+.54i
Inverted MA Roots	.45+.51i	.45-.51i

Discussion:

Based on the above results,

R-squared = 2.78 %

AR(1), AR(2), MA(1), MA(2) terms are significant

On checking the autocorrelation using GLS method, Equation: UNTITLED Workfile: NIFTY NEXT 50 2010-2014::Untitled\

View | Proc | Object | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

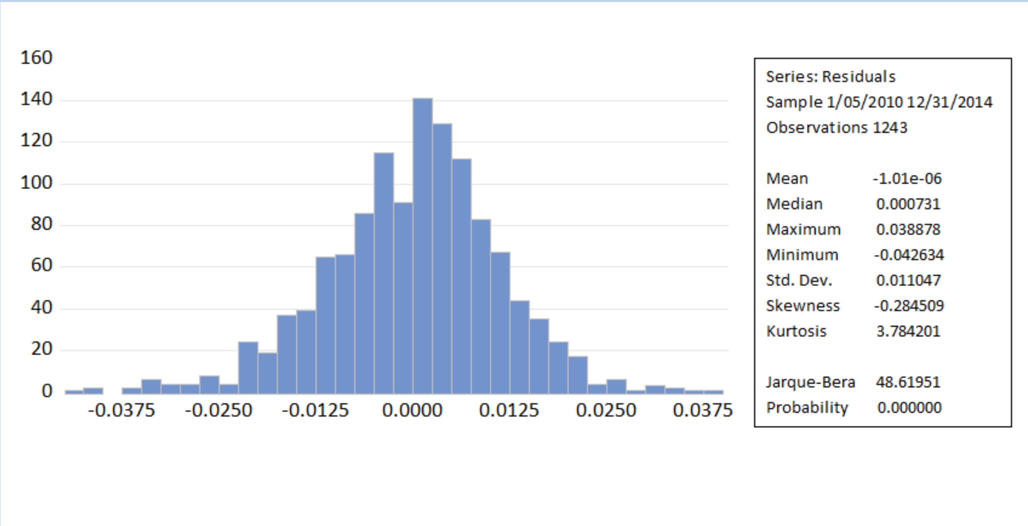
Dependent Variable: DLINDEX
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 05/18/21 Time: 14:04
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Convergence achieved after 24 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000463	0.000338	1.369758	0.1710
AR(1)	1.034292	0.188537	5.485871	0.0000
AR(2)	-0.559101	0.125235	-4.464425	0.0000
MA(1)	-0.893988	0.195102	-4.582146	0.0000
MA(2)	0.459197	0.136581	3.362092	0.0008

R-squared	0.027866	Mean dependent var	0.000463
Adjusted R-squared	0.024725	S.D. dependent var	0.011204
S.E. of regression	0.011065	Akaike info criterion	-6.166084
Sum squared resid	0.151562	Schwarz criterion	-6.145468
Log likelihood	3837.221	Hannan-Quinn criter.	-6.158332
F-statistic	8.871852	Durbin-Watson stat	1.984978
Prob(F-statistic)	0.000000		

Inverted AR Roots	.52-.54i	.52+.54i
Inverted MA Roots	.45+.51i	.45-.51i

- **Normality test:**



Discussion:

Errors terms (Residuals) follow normal distribution, so it is fine

- **Serial correlation:**

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	0.479630	Prob. F(2,1236)	0.6191
Obs*R-squared	0.963945	Prob. Chi-Square(2)	0.6176

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 05/18/21 Time: 14:11
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Coefficient covariance computed using outer product of gradients
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.09E-09	3.73E-06	-0.000827	0.9993
AR(1)	-0.000407	0.004437	-0.091732	0.9269
AR(2)	0.001371	0.003497	0.392106	0.6950
MA(1)	0.000960	0.006821	0.140804	0.8880
MA(2)	-0.001933	0.004236	-0.456381	0.6482
RESID(-1)	0.057041	0.234869	0.242862	0.8082
RESID(-2)	-0.025993	0.111018	-0.234137	0.8149

R-squared	0.000775	Mean dependent var	-1.01E-06
Adjusted R-squared	-0.004075	S.D. dependent var	0.011047
S.E. of regression	0.011069	Akaike info criterion	-6.163680
Sum squared resid	0.151444	Schwarz criterion	-6.134817
Log likelihood	3837.727	Hannan-Quinn criter.	-6.152827
F-statistic	0.159875	Durbin-Watson stat	1.998075
Prob(F-statistic)	0.987068		

There is no correlation and we cannot reject the null hypothesis

- **Heteroskedasticity test – ARCH**

Heteroskedasticity Test: ARCH

F-statistic	5.241509	Prob. F(1,1240)	0.0222
Obs*R-squared	5.227864	Prob. Chi-Square(1)	0.0222

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/18/21 Time: 14:14
Sample (adjusted): 1/06/2010 12/31/2014
Included observations: 1242 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000114	6.72E-06	16.94521	0.0000
RESID^2(-1)	0.064879	0.028339	2.289434	0.0222

R-squared	0.004209	Mean dependent var	0.000122
Adjusted R-squared	0.003406	S.D. dependent var	0.000204
S.E. of regression	0.000203	Akaike info criterion	-14.16254
Sum squared resid	5.12E-05	Schwarz criterion	-14.15429
Log likelihood	8796.940	Hannan-Quinn criter.	-14.15944
F-statistic	5.241509	Durbin-Watson stat	2.005693
Prob(F-statistic)	0.022222		

Discussion:

Heteroskedasticity test is fine and the null hypothesis can be rejected.

Model 2:

Dependent Variable: DLINDEX
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 05/18/21 Time: 15:35
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Convergence achieved after 3 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000466	0.000370	1.257928	0.2087
AR(1)	0.151695	0.028064	5.405371	0.0000

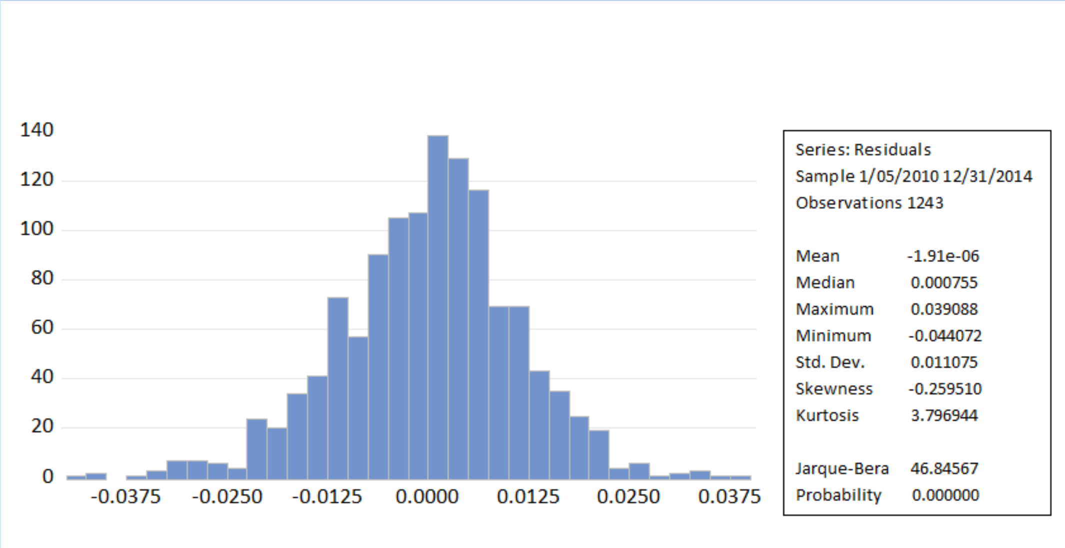
R-squared	0.022972	Mean dependent var	0.000463
Adjusted R-squared	0.022184	S.D. dependent var	0.011204
S.E. of regression	0.011079	Akaike info criterion	-6.165908
Sum squared resid	0.152325	Schwarz criterion	-6.157662
Log likelihood	3834.112	Hannan-Quinn criter.	-6.162807
F-statistic	29.17820	Durbin-Watson stat	2.000067
Prob(F-statistic)	0.000000		

Inverted AR Roots	.15

Discussion:

AR term is significant
R – squared is 2.29 %

- **Normality test (Residual diagnostics):**



Discussion:

The distribution is normal and normality is not violated

- **Serial correlation:**

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	0.426678	Prob. F(2,1239)	0.6528
Obs*R-squared	0.855522	Prob. Chi-Square(2)	0.6520

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 05/18/21 Time: 15:45
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Coefficient covariance computed using outer product of gradients
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.81E-09	4.10E-06	0.000930	0.9993
AR(1)	0.011886	0.013511	0.879719	0.3792
RESID(-1)	1.072952	1.220925	0.878803	0.3797
RESID(-2)	0.170813	0.187282	0.912061	0.3619

R-squared	0.000688	Mean dependent var	-1.91E-06
Adjusted R-squared	-0.001731	S.D. dependent var	0.011075
S.E. of regression	0.011084	Akaike info criterion	-6.163398
Sum squared resid	0.152220	Schwarz criterion	-6.146904
Log likelihood	3834.552	Hannan-Quinn criter.	-6.157196
F-statistic	0.284440	Durbin-Watson stat	2.001927
Prob(F-statistic)	0.836664		

Discussion:

There is variation in the correlation

- **Heteroskedasticity test – ARCH:**

Heteroskedasticity Test: ARCH

F-statistic	6.559105	Prob. F(1,1240)	0.0106
Obs*R-squared	6.535116	Prob. Chi-Square(1)	0.0106

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/18/21 Time: 15:47
Sample (adjusted): 1/06/2010 12/31/2014
Included observations: 1242 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000114	6.77E-06	16.78708	0.0000
RESID^2(-1)	0.072538	0.028323	2.561075	0.0106

R-squared	0.005262	Mean dependent var	0.000122
Adjusted R-squared	0.004460	S.D. dependent var	0.000205
S.E. of regression	0.000205	Akaike info criterion	-14.14896
Sum squared resid	5.19E-05	Schwarz criterion	-14.14071
Log likelihood	8788.506	Hannan-Quinn criter.	-14.14586
F-statistic	6.559105	Durbin-Watson stat	2.006799
Prob(F-statistic)	0.010552		

Discussion:

Heteroskedasticity test is fine

Model 3:

Dependent Variable: DLINDEX
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 05/18/21 Time: 15:49
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Convergence achieved after 5 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000465	0.000360	1.290437	0.1971
MA(1)	0.146488	0.028084	5.216033	0.0000

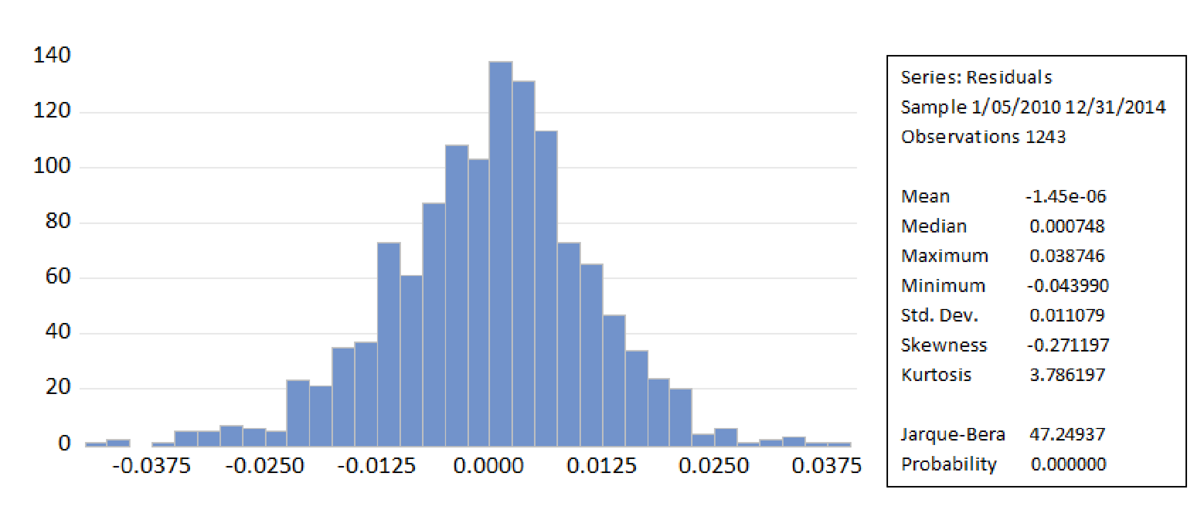
R-squared	0.022191	Mean dependent var	0.000463
Adjusted R-squared	0.021403	S.D. dependent var	0.011204
S.E. of regression	0.011083	Akaike info criterion	-6.165111
Sum squared resid	0.152447	Schwarz criterion	-6.156865
Log likelihood	3833.617	Hannan-Quinn criter.	-6.162010
F-statistic	28.16437	Durbin-Watson stat	1.990082
Prob(F-statistic)	0.000000		

Inverted MA Roots	-.15

Discussion:

MA term is significant
R – squared is 2.21 %

- **Normality test:**



Discussion:

The distribution is normal and the normality is not violated

- **Heteroskedasticity test – ARCH:**

Heteroskedasticity Test: ARCH

F-statistic	6.488486	Prob. F(1,1240)	0.0110
Obs*R-squared	6.465121	Prob. Chi-Square(1)	0.0110

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/18/21 Time: 15:59
Sample (adjusted): 1/06/2010 12/31/2014
Included observations: 1242 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000114	6.76E-06	16.81765	0.0000
RESID^2(-1)	0.072148	0.028324	2.547251	0.0110

R-squared	0.005205	Mean dependent var	0.000123
Adjusted R-squared	0.004403	S.D. dependent var	0.000205
S.E. of regression	0.000204	Akaike info criterion	-14.15117
Sum squared resid	5.18E-05	Schwarz criterion	-14.14292
Log likelihood	8789.879	Hannan-Quinn criter.	-14.14807
F-statistic	6.488486	Durbin-Watson stat	2.006981
Prob(F-statistic)	0.010977		

Discussion:

Heteroskedasticity test is fine

Model 4:

Dependent Variable: DLINDEX
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 05/18/21 Time: 16:14
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Convergence achieved after 15 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000463	0.000345	1.344247	0.1791
AR(1)	1.129400	0.193629	5.832817	0.0000
AR(2)	-0.690197	0.193777	-3.561803	0.0004
MA(1)	-0.981626	0.196495	-4.995677	0.0000
MA(2)	0.558033	0.173827	3.210276	0.0014
MA(3)	0.039646	0.049938	0.793897	0.4274

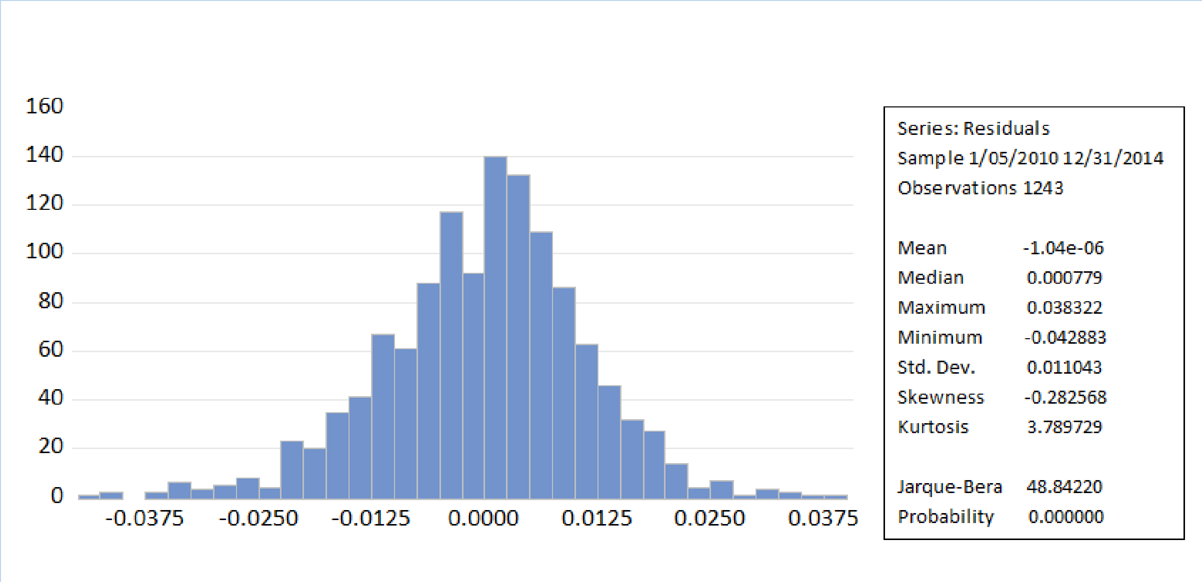
R-squared	0.028459	Mean dependent var	0.000463
Adjusted R-squared	0.024532	S.D. dependent var	0.011204
S.E. of regression	0.011066	Akaike info criterion	-6.165084
Sum squared resid	0.151469	Schwarz criterion	-6.140344
Log likelihood	3837.600	Hannan-Quinn criter.	-6.155781
F-statistic	7.246968	Durbin-Watson stat	1.999507
Prob(F-statistic)	0.000001		

Inverted AR Roots	.56+.61i	.56-.61i	
Inverted MA Roots	.52-.59i	.52+.59i	-.06

Discussion:

AR(1), AR(2), MA(1), MA(2) are significant
MA(3) is insignificant
R – squared is 2.84 %

- **Normality test:**



Discussion:

Residual distribution is normal and normality is not violated

- **Serial correlation test:**

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	0.322645	Prob. F(2,1235)	0.7243
Obs*R-squared	0.649131	Prob. Chi-Square(2)	0.7228

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 05/18/21 Time: 16:17
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Coefficient covariance computed using outer product of gradients
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.31E-08	3.81E-06	-0.019200	0.9847
AR(1)	-0.000130	0.003448	-0.037642	0.9700
AR(2)	-0.000462	0.002616	-0.176729	0.8598
MA(1)	-0.076707	0.103697	-0.739725	0.4596
MA(2)	0.080945	0.110306	0.733822	0.4632
MA(3)	-0.047866	0.065650	-0.729117	0.4661
RESID(-1)	-6.960896	9.503399	-0.732464	0.4640
RESID(-2)	0.452990	0.654529	0.692085	0.4890

R-squared	0.000522	Mean dependent var	-1.04E-06
Adjusted R-squared	-0.005143	S.D. dependent var	0.011043
S.E. of regression	0.011072	Akaike info criterion	-6.162427
Sum squared resid	0.151390	Schwarz criterion	-6.129441
Log likelihood	3837.949	Hannan-Quinn criter.	-6.150023
F-statistic	0.092183	Durbin-Watson stat	1.997812
Prob(F-statistic)	0.998714		

- **Heteroskedasticity test – ARCH:**

Heteroskedasticity Test: ARCH

F-statistic	5.059770	Prob. F(1,1240)	0.0247
Obs*R-squared	5.047335	Prob. Chi-Square(1)	0.0247

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/18/21 Time: 16:18
Sample (adjusted): 1/06/2010 12/31/2014
Included observations: 1242 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000114	6.73E-06	16.95196	0.0000
RESID^2(-1)	0.063750	0.028341	2.249393	0.0247

R-squared	0.004064	Mean dependent var	0.000122
Adjusted R-squared	0.003261	S.D. dependent var	0.000204
S.E. of regression	0.000203	Akaike info criterion	-14.16163
Sum squared resid	5.13E-05	Schwarz criterion	-14.15338
Log likelihood	8796.374	Hannan-Quinn criter.	-14.15853
F-statistic	5.059770	Durbin-Watson stat	2.005612
Prob(F-statistic)	0.024662		

Discussion:

Heteroskedasticity test is fine

Model 5:

Dependent Variable: DLINDEX
Method: ARMA Generalized Least Squares (Gauss-Newton)
Date: 05/18/21 Time: 16:19
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Convergence achieved after 19 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000464	0.000345	1.344292	0.1791
AR(1)	1.204709	0.261882	4.600204	0.0000
AR(2)	-0.765575	0.282892	-2.706243	0.0069
AR(3)	0.045085	0.059714	0.755019	0.4504
MA(1)	-1.057534	0.259438	-4.076246	0.0000
MA(2)	0.624146	0.241081	2.588946	0.0097

R-squared	0.028425	Mean dependent var	0.000463
Adjusted R-squared	0.024498	S.D. dependent var	0.011204
S.E. of regression	0.011066	Akaike info criterion	-6.165049
Sum squared resid	0.151475	Schwarz criterion	-6.140309
Log likelihood	3837.578	Hannan-Quinn criter.	-6.155746
F-statistic	7.238063	Durbin-Watson stat	1.998424
Prob(F-statistic)	0.000001		

Inverted AR Roots	.57+.61i	.57-.61i	.07
Inverted MA Roots	.53-.59i	.53+.59i	

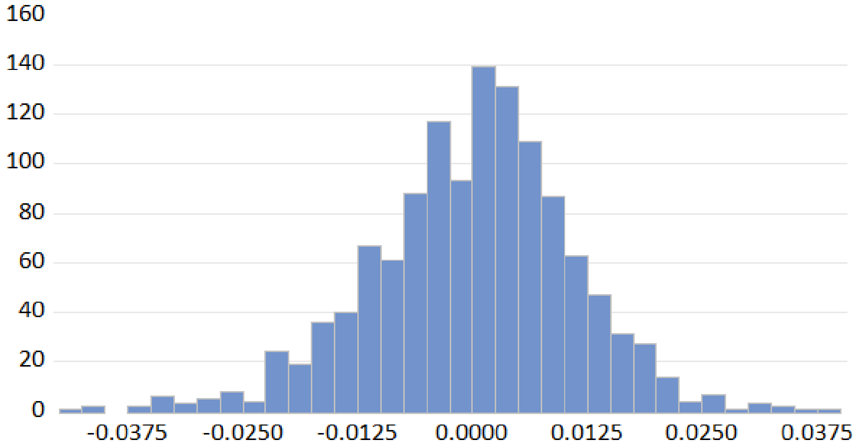
Discussion:

AR(1), AR(2), MA(1), MA(2) terms are significant

AR(3) term is insignificant

R – squared is 2.84 %

- **Normality test:**



Series: Residuals
Sample 1/05/2010 12/31/2014
Observations 1243

Mean	-1.05e-06
Median	0.000781
Maximum	0.038347
Minimum	-0.042870
Std. Dev.	0.011044
Skewness	-0.282761
Kurtosis	3.790541
Jarque-Bera	48.93121
Probability	0.000000

Discussion:

Residual distribution is normal and normality is not violated

- **Serial correlation test:**

Breusch-Godfrey Serial Correlation LM Test:
Null hypothesis: No serial correlation at up to 2 lags

F-statistic	0.399432	Prob. F(2,1235)	0.6708
Obs*R-squared	0.803519	Prob. Chi-Square(2)	0.6691

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Date: 05/18/21 Time: 16:23
Sample: 1/05/2010 12/31/2014
Included observations: 1243
Coefficient covariance computed using outer product of gradients
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.19E-08	3.81E-06	-0.018875	0.9849
AR(1)	-0.078208	0.097852	-0.799241	0.4243
AR(2)	0.088555	0.110241	0.803289	0.4220
AR(3)	-0.054038	0.066876	-0.808039	0.4192
MA(1)	-0.000189	0.004337	-0.043491	0.9653
MA(2)	0.000902	0.003011	0.299605	0.7645
RESID(-1)	-7.101628	8.749672	-0.811645	0.4172
RESID(-2)	-0.453724	0.534182	-0.849380	0.3958

R-squared	0.000646	Mean dependent var	-1.05E-06
Adjusted R-squared	-0.005018	S.D. dependent var	0.011044
S.E. of regression	0.011071	Akaike info criterion	-6.162517
Sum squared resid	0.151377	Schwarz criterion	-6.129530
Log likelihood	3838.004	Hannan-Quinn criter.	-6.150113
F-statistic	0.114122	Durbin-Watson stat	1.997472
Prob(F-statistic)	0.997442		

- **Heteroskedasticity test – ARCH:**

Heteroskedasticity Test: ARCH

F-statistic	5.054539	Prob. F(1,1240)	0.0247
Obs*R-squared	5.042139	Prob. Chi-Square(1)	0.0247

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 05/18/21 Time: 16:23
Sample (adjusted): 1/06/2010 12/31/2014
Included observations: 1242 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.000114	6.73E-06	16.95072	0.0000
RESID^2(-1)	0.063717	0.028341	2.248230	0.0247

R-squared	0.004060	Mean dependent var	0.000122
Adjusted R-squared	0.003257	S.D. dependent var	0.000204
S.E. of regression	0.000203	Akaike info criterion	-14.16127
Sum squared resid	5.13E-05	Schwarz criterion	-14.15302
Log likelihood	8796.147	Hannan-Quinn criter.	-14.15816
F-statistic	5.054539	Durbin-Watson stat	2.005597
Prob(F-statistic)	0.024737		

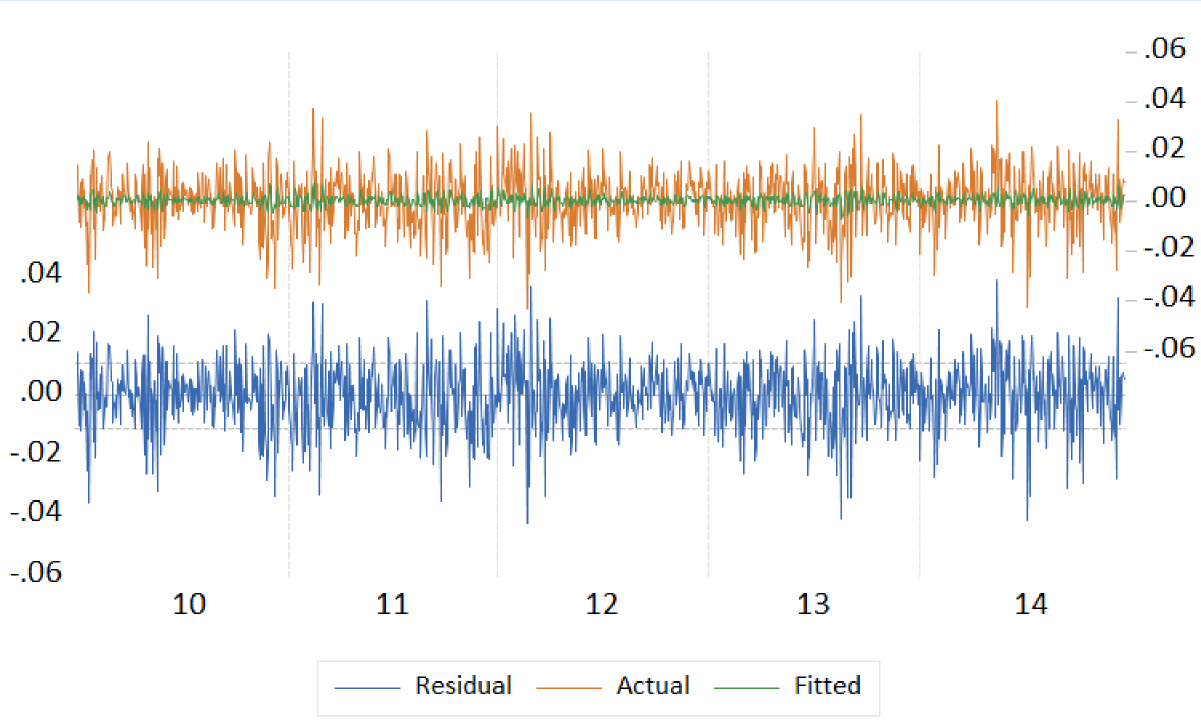
Discussion:

Heteroskedasticity test is fine

Residual graphs:

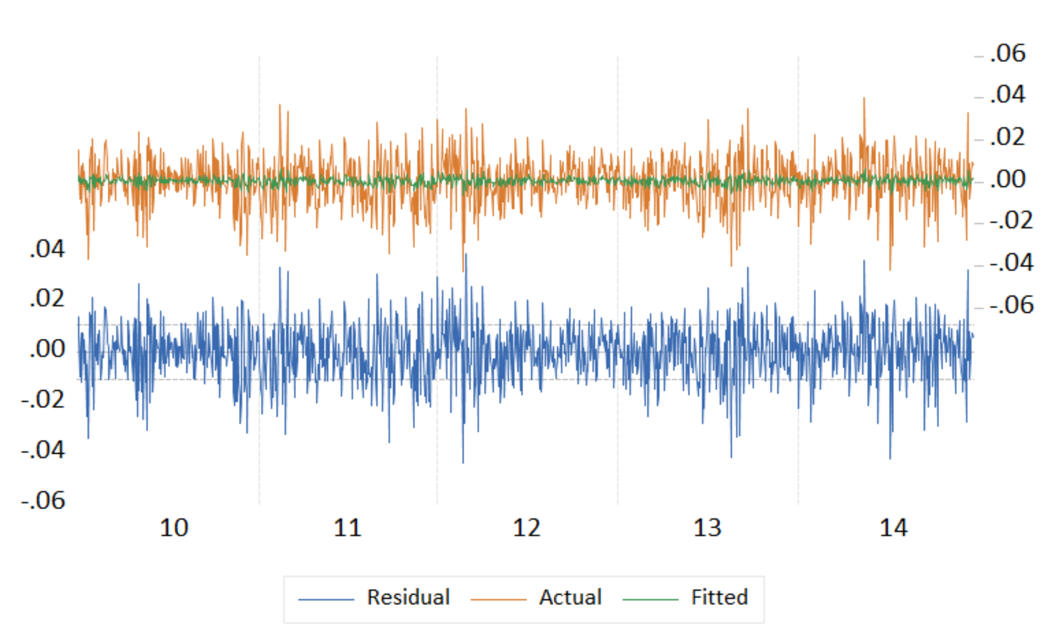
Residual graphs are drawn for all the models to check the residual distribution so as to estimate the best fitting model.

Model 1:



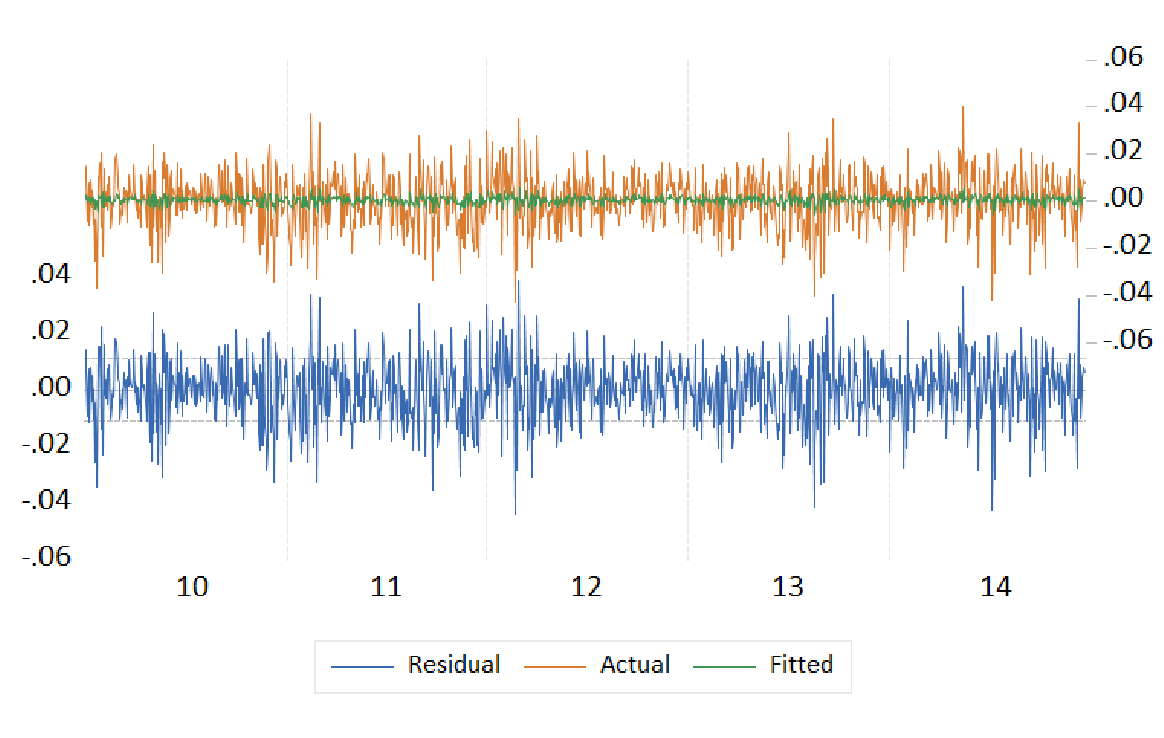
obs	Actual	Fitted	Residual	Residual Plot
1/05/2010	0.01503	0.00067	0.01436	
1/06/2010	0.00847	0.00261	0.00586	
1/07/2010	-0.00822	0.00184	-0.01006	
1/08/2010	-0.00087	-0.00134	0.00047	
1/11/2010	0.00772	-0.00108	0.00880	
1/12/2010	-0.01033	0.00106	-0.01139	
1/13/2010	0.00540	-0.00054	0.00594	
1/14/2010	0.00901	0.00107	0.00795	
1/15/2010	-0.00251	0.00217	-0.00468	
1/18/2010	0.00475	0.00044	0.00431	
1/19/2010	-0.00948	0.00056	-0.01004	
1/20/2010	-0.00067	-0.00127	0.00059	
1/21/2010	-0.02527	-0.00029	-0.02498	
1/22/2010	-0.00810	-0.00291	-0.00518	
1/25/2010	-0.01783	-0.00084	-0.01699	
1/27/2010	-0.03668	-0.00086	-0.03582	
1/28/2010	0.01055	-0.00351	0.01406	
1/29/2010	0.00805	0.00265	0.00540	
2/01/2010	0.01648	0.00430	0.01219	
2/02/2010	-0.01085	0.00438	-0.01523	
2/03/2010	0.02080	-0.00099	0.02179	
2/04/2010	-0.01908	0.00135	-0.02043	
2/05/2010	-0.02324	-0.00285	-0.02039	
2/06/2010	0.01346	-0.00428	0.01774	
2/08/2010	0.00108	0.00194	-0.00086	
2/09/2010	0.00387	0.00275	0.00113	
2/10/2010	-0.00216	0.00224	-0.00441	
2/11/2010	0.00724	0.00030	0.00694	
2/15/2010	-0.00357	0.00071	-0.00428	
2/16/2010	0.00771	-0.00048	0.00819	
2/17/2010	0.01045	0.00092	0.00953	

Model 2:



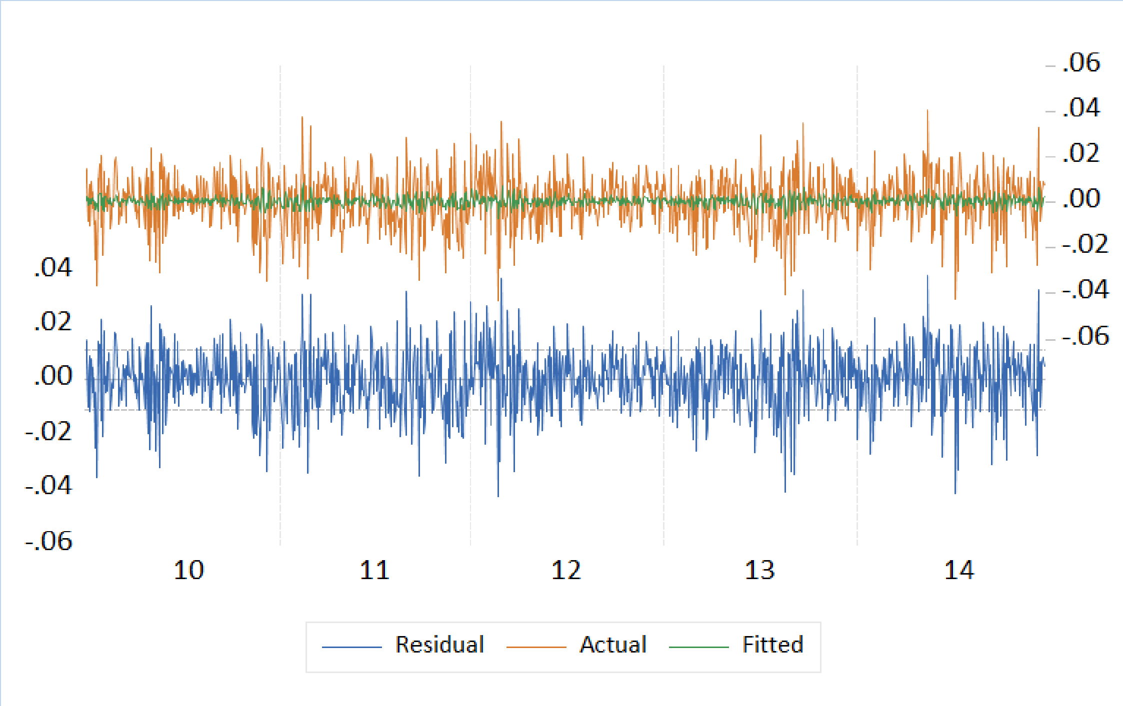
obs	Actual	Fitted	Residual	Residual Plot
1/05/2010	0.01503	0.00063	0.01439	
1/06/2010	0.00847	0.00267	0.00580	
1/07/2010	-0.00822	0.00168	-0.00990	
1/08/2010	-0.00087	-0.00085	-2.0E-05	
1/11/2010	0.00772	0.00026	0.00746	
1/12/2010	-0.01033	0.00157	-0.01190	
1/13/2010	0.00540	-0.00117	0.00657	
1/14/2010	0.00901	0.00121	0.00780	
1/15/2010	-0.00251	0.00176	-0.00427	
1/18/2010	0.00475	1.4E-05	0.00473	
1/19/2010	-0.00948	0.00112	-0.01060	
1/20/2010	-0.00067	-0.00104	0.00037	
1/21/2010	-0.02527	0.00029	-0.02556	
1/22/2010	-0.00810	-0.00344	-0.00466	
1/25/2010	-0.01783	-0.00083	-0.01700	
1/27/2010	-0.03668	-0.00231	-0.03437	
1/28/2010	0.01055	-0.00517	0.01572	
1/29/2010	0.00805	0.00200	0.00605	
2/01/2010	0.01648	0.00162	0.01487	
2/02/2010	-0.01085	0.00290	-0.01375	
2/03/2010	0.02080	-0.00125	0.02205	
2/04/2010	-0.01908	0.00355	-0.02263	
2/05/2010	-0.02324	-0.00250	-0.02074	
2/06/2010	0.01346	-0.00313	0.01659	
2/08/2010	0.00108	0.00244	-0.00136	
2/09/2010	0.00387	0.00056	0.00331	
2/10/2010	-0.00216	0.00098	-0.00315	
2/11/2010	0.00724	6.7E-05	0.00717	
2/15/2010	-0.00357	0.00149	-0.00506	
2/16/2010	0.00771	-0.00015	0.00785	

Model 3:

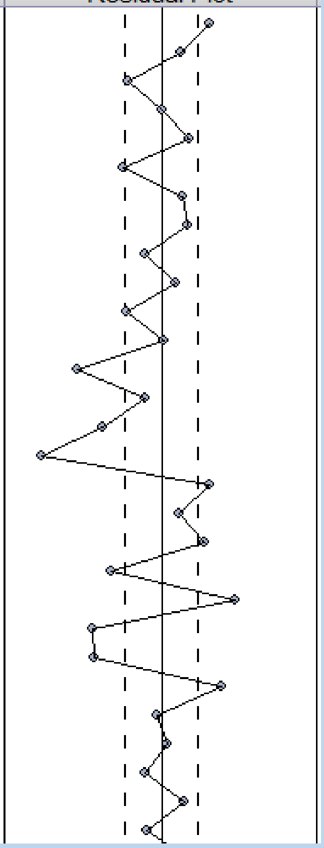


obs	Actual	Fitted	Residual	Residual Plot
1/05/2010	0.01503	0.00062	0.01441	
1/06/2010	0.00847	0.00255	0.00592	
1/07/2010	-0.00822	0.00133	-0.00955	
1/08/2010	-0.00087	-0.00093	6.3E-05	
1/11/2010	0.00772	0.00047	0.00725	
1/12/2010	-0.01033	0.00153	-0.01186	
1/13/2010	0.00540	-0.00127	0.00667	
1/14/2010	0.00901	0.00144	0.00757	
1/15/2010	-0.00251	0.00157	-0.00409	
1/18/2010	0.00475	-0.00013	0.00488	
1/19/2010	-0.00948	0.00118	-0.01066	
1/20/2010	-0.00067	-0.00110	0.00042	
1/21/2010	-0.02527	0.00053	-0.02580	
1/22/2010	-0.00810	-0.00331	-0.00478	
1/25/2010	-0.01783	-0.00024	-0.01759	
1/27/2010	-0.03668	-0.00211	-0.03457	
1/28/2010	0.01055	-0.00460	0.01515	
1/29/2010	0.00805	0.00268	0.00536	
2/01/2010	0.01648	0.00125	0.01523	
2/02/2010	-0.01085	0.00270	-0.01355	
2/03/2010	0.02080	-0.00152	0.02232	
2/04/2010	-0.01908	0.00373	-0.02281	
2/05/2010	-0.02324	-0.00288	-0.02037	
2/06/2010	0.01346	-0.00252	0.01598	
2/08/2010	0.00108	0.00281	-0.00173	
2/09/2010	0.00387	0.00021	0.00366	
2/10/2010	-0.00216	0.00100	-0.00317	
2/11/2010	0.00724	1.2E-06	0.00724	
2/15/2010	-0.00357	0.00152	-0.00509	
2/16/2010	0.00771	-0.00028	0.00799	
2/17/2010	0.01045	0.00163	0.00882	
2/18/2010	-0.00788	0.00176	-0.00964	
2/19/2010	-0.01139	-0.00095	-0.01045	

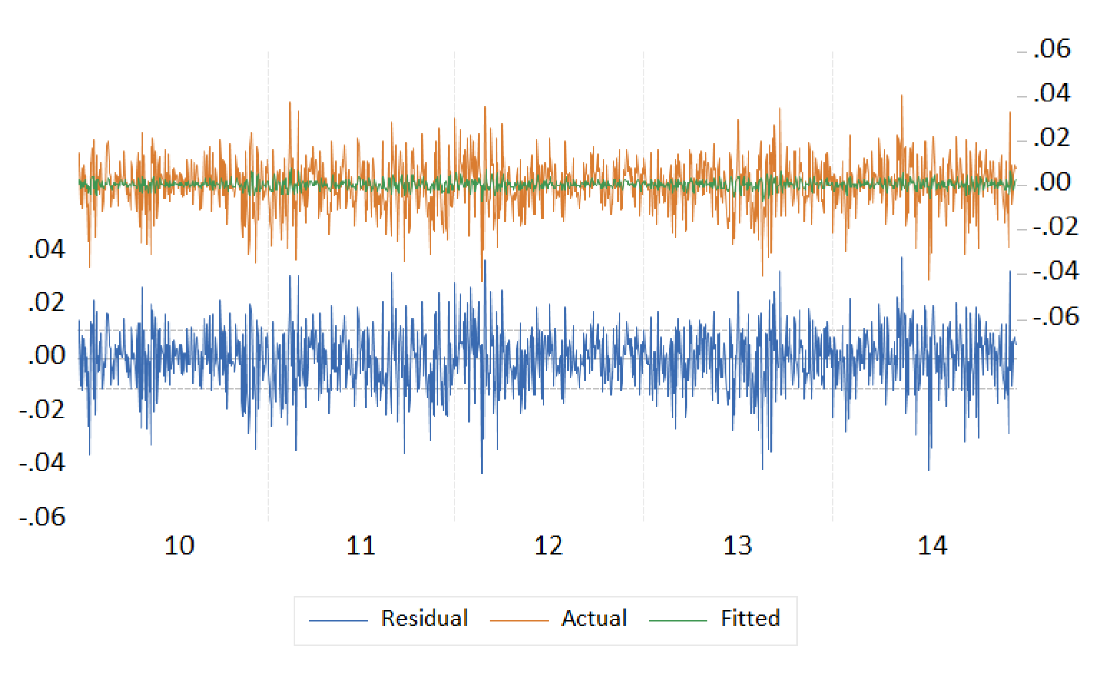
Model 4:



obs	Actual	Fitted	Residual	Residual Plot
1/05/2010	0.01503	0.00067	0.01435	
1/06/2010	0.00847	0.00270	0.00578	
1/07/2010	-0.00822	0.00174	-0.00996	
1/08/2010	-0.00087	-0.00133	0.00046	
1/11/2010	0.00772	-0.00081	0.00853	
1/12/2010	-0.01033	0.00106	-0.01139	
1/13/2010	0.00540	-0.00078	0.00618	
1/14/2010	0.00901	0.00141	0.00761	
1/15/2010	-0.00251	0.00224	-0.00475	
1/18/2010	0.00475	0.00036	0.00439	
1/19/2010	-0.00948	0.00069	-0.01018	
1/20/2010	-0.00067	-0.00148	0.00080	
1/21/2010	-0.02527	-0.00025	-0.02502	
1/22/2010	-0.00810	-0.00321	-0.00489	
1/25/2010	-0.01783	-0.00058	-0.01725	
1/27/2010	-0.03668	-0.00107	-0.03561	
1/28/2010	0.01055	-0.00373	0.01428	
1/29/2010	0.00805	0.00292	0.00513	
2/01/2010	0.01648	0.00359	0.01289	
2/02/2010	-0.01085	0.00409	-0.01495	
2/03/2010	0.02080	-0.00131	0.02210	
2/04/2010	-0.01908	0.00171	-0.02079	
2/05/2010	-0.02324	-0.00349	-0.01975	
2/06/2010	0.01346	-0.00416	0.01762	
2/08/2010	0.00108	0.00236	-0.00128	
2/09/2010	0.00387	0.00250	0.00138	
2/10/2010	-0.00216	0.00252	-0.00469	
2/11/2010	0.00724	0.00046	0.00678	
2/15/2010	-0.00357	0.00071	-0.00428	



Model 5:



obs	Actual	Fitted	Residual	Residual Plot
obs	Actual	Fitted	Residual	Residual Plot
1/05/2010	0.01503	0.00067	0.01435	
1/06/2010	0.00847	0.00269	0.00578	
1/07/2010	-0.00822	0.00175	-0.00998	
1/08/2010	-0.00087	-0.00134	0.00046	
1/11/2010	0.00772	-0.00083	0.00855	
1/12/2010	-0.01033	0.00108	-0.01141	
1/13/2010	0.00540	-0.00076	0.00616	
1/14/2010	0.00901	0.00137	0.00764	
1/15/2010	-0.00251	0.00226	-0.00477	
1/18/2010	0.00475	0.00037	0.00438	
1/19/2010	-0.00948	0.00068	-0.01016	
1/20/2010	-0.00067	-0.00145	0.00078	
1/21/2010	-0.02527	-0.00027	-0.02500	
1/22/2010	-0.00810	-0.00319	-0.00491	
1/25/2010	-0.01783	-0.00061	-0.01722	
1/27/2010	-0.03668	-0.00104	-0.03564	
1/28/2010	0.01055	-0.00372	0.01427	
1/29/2010	0.00805	0.00289	0.00516	
2/01/2010	0.01648	0.00366	0.01283	
2/02/2010	-0.01085	0.00406	-0.01492	
2/03/2010	0.02080	-0.00131	0.02211	
2/04/2010	-0.01908	0.00166	-0.02073	
2/05/2010	-0.02324	-0.00343	-0.01981	
2/06/2010	0.01346	-0.00421	0.01767	
2/08/2010	0.00108	0.00234	-0.00126	
2/09/2010	0.00387	0.00255	0.00133	
2/10/2010	-0.00216	0.00250	-0.00466	
2/11/2010	0.00724	0.00047	0.00676	
2/15/2010	-0.00357	0.00073	-0.00430	
2/16/2010	0.00771	-0.00093	0.00864	

Discussion on the choice of AR, MA, ARMA / ARIMA models:

On analysing and observing the five models based on the tests,

a) AR model
Model 2 is chosen (1, 0)
P = 1, q = 0
AR(1)
Based on the unit root test it is chosen as the null hypothesis is rejected at first difference
Residual diagnostics test is an important test and based on that it can be seen that it works it and the normality assumption is not violated.
Heteroskedasticity tests are also fine with the value of R squared to be 2.29 %
There is one term - AR(1) term and it is significant
So, based on observing and checking it can be estimated to be the appropriate AR model.

b) MA model
Model 3 is chosen (0,1)
P = 0, q = 1
MA(1)
Based on the unit root test it is chosen as the null hypothesis is rejected at first difference
Residual diagnostics test is an important test and based on that it can be seen that it works it and the normality assumption is not violated.
Heteroskedasticity tests are also fine with the value of R squared to be 2.21 %
There is one term – MA(1) term and it is significant
So, based on observing and checking it can be estimated to be the appropriate MA model.

c) ARMA model
Model 1 is chosen (2, 2)
P = 2, q = 2
Residual diagnostics test is an important test and based on that it can be seen that it works it and the normality assumption is not violated.
Heteroskedasticity tests are also fine with the value of R squared to be 2.78 %
All the AR and MA terms are significant
Model 1 is chosen over Model 4 and 5 because all the terms are significant in Model 1 and it is best fit with the data.
ARMA(2, 3) is the best fit model through automatic selection as well as by comparing and checking with other models.
So, based on observing and checking it can be estimated to be the appropriate ARMA model.

Appropriate ARIMA model - ARIMA (2,1,2)
Appropriate ARMA model – ARMA(2, 2)

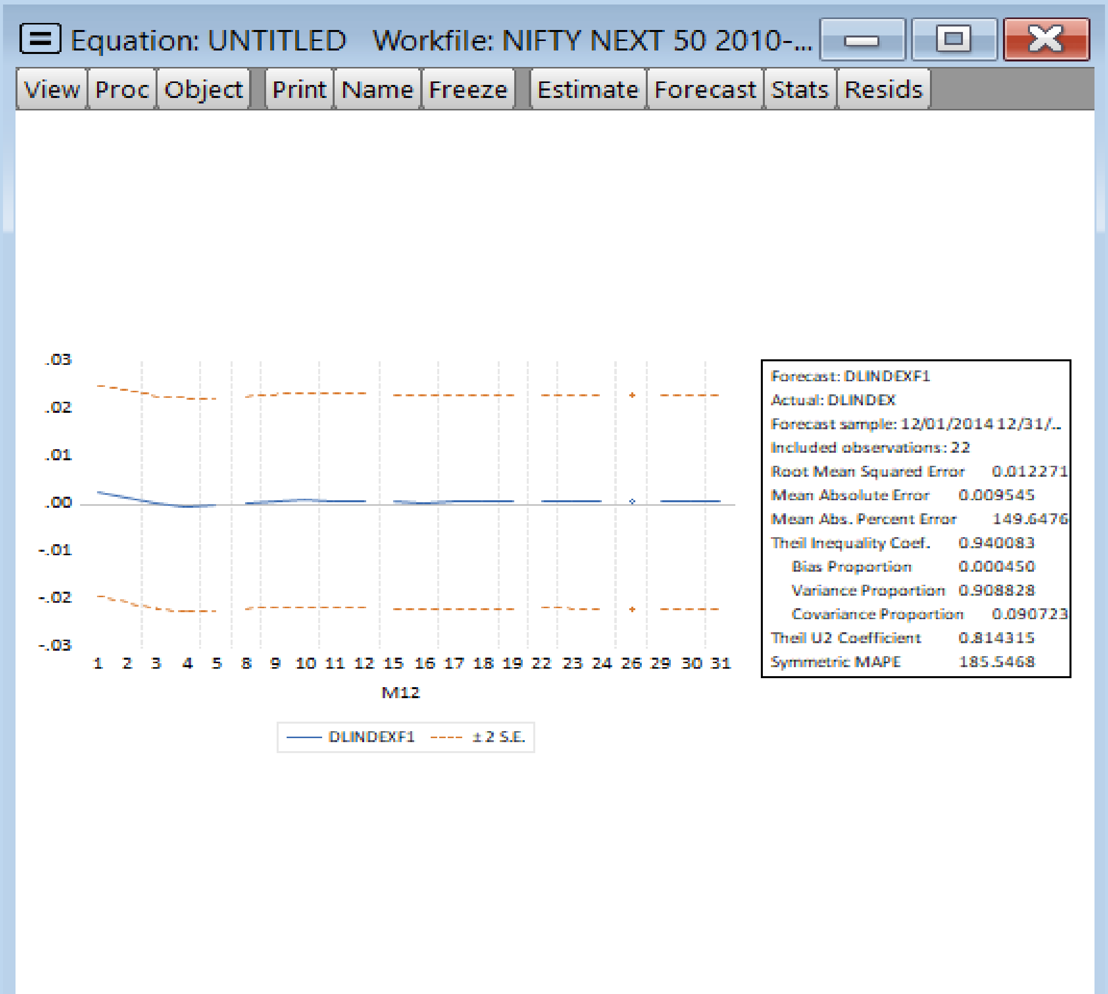
3) ARMA / ARIMA forecasting

Removing the last 20 observations and forecasting it for the most appropriate model,

Model 1:

ARMA(2, 2)

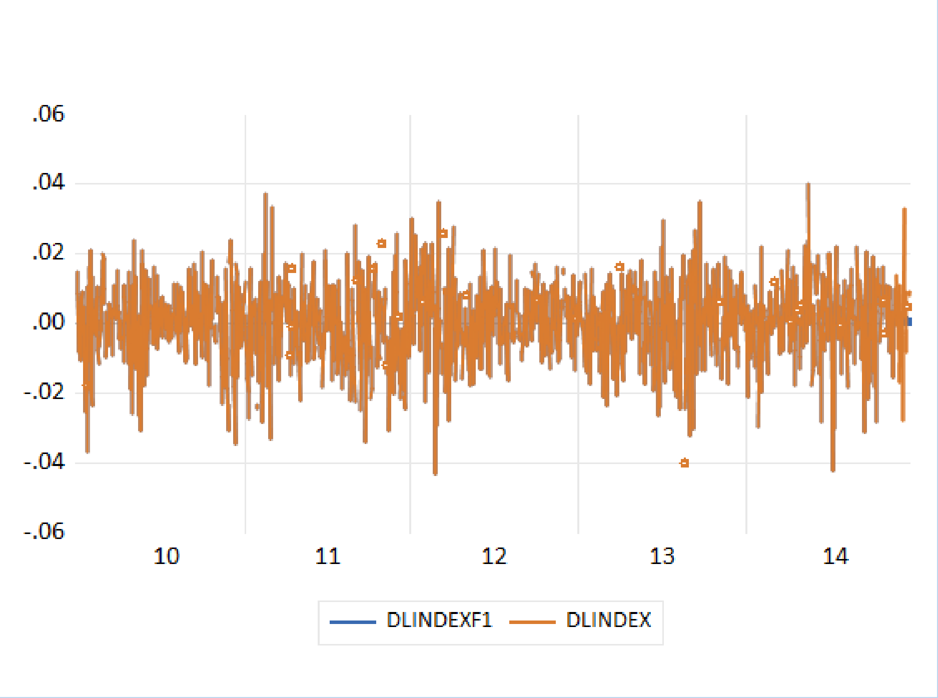
Forecast graphs



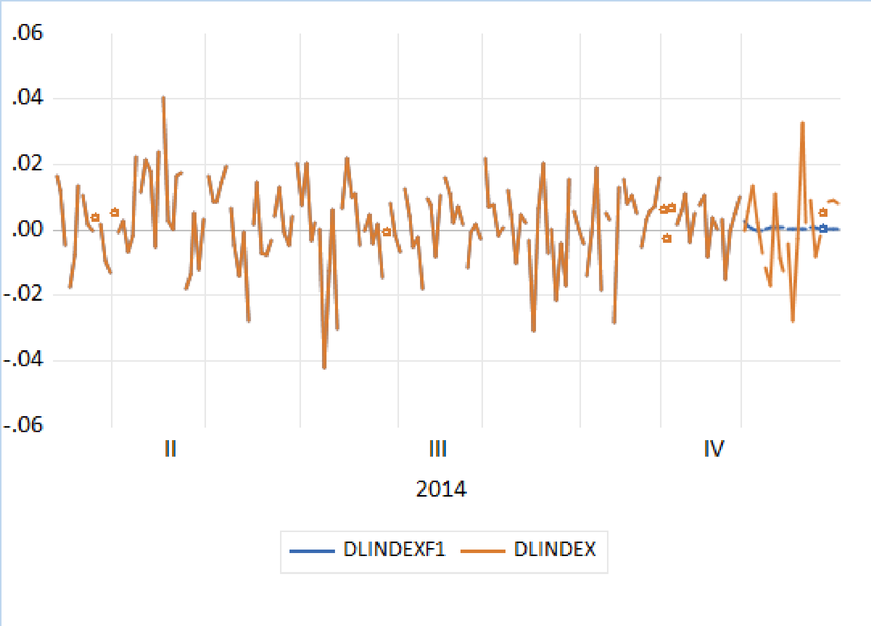
The table showing the forecasted values which is DLINDEXF1, it can be observed that the values of DLINDEXF1 and DLINDEX are same till 11/28/2014 and after that the values are forecasted to get the best possible ARMA forecasting after trimming the last 22 obervations.

	DLINDEXF1	DLINDEX		
11/28/2014	0.010201	0.010201		
12/01/2014	0.002668	-0.000245		
12/02/2014	0.001290	0.005573		
12/03/2014	6.56E-05	0.013711		
12/04/2014	-0.000437	0.002393		
12/05/2014	-0.000261	-0.007062		
12/08/2014	0.000218	-0.011404		
12/09/2014	0.000625	-0.016933		
12/10/2014	0.000781	0.011007		
12/11/2014	0.000710	-0.008642		
12/12/2014	0.000544	-0.012382		
12/15/2014	0.000409	-0.003999		
12/16/2014	0.000361	-0.027812		
12/17/2014	0.000389	-0.002714		
12/18/2014	0.000446	0.032806		
12/19/2014	0.000491	0.002023		
12/22/2014	0.000505	0.009107		
12/23/2014	0.000495	-0.008242		
12/24/2014	0.000475	-0.001792		
12/26/2014	0.000460	0.004991		
12/29/2014	0.000456	0.008398		
12/30/2014	0.000460	0.009187		
12/31/2014	0.000467	0.007923		

The graph between DLINDEX and DLINDEXF1



The graph is magnified to show the forecast clearly DLINDEXF1



Discussion:

Thus, we are able to perform ARMA / ARIMA modelling and forecasting to find the best fitting model for the last 22 observations of the data.