

Workshop 4: Stationarity and ARIMA

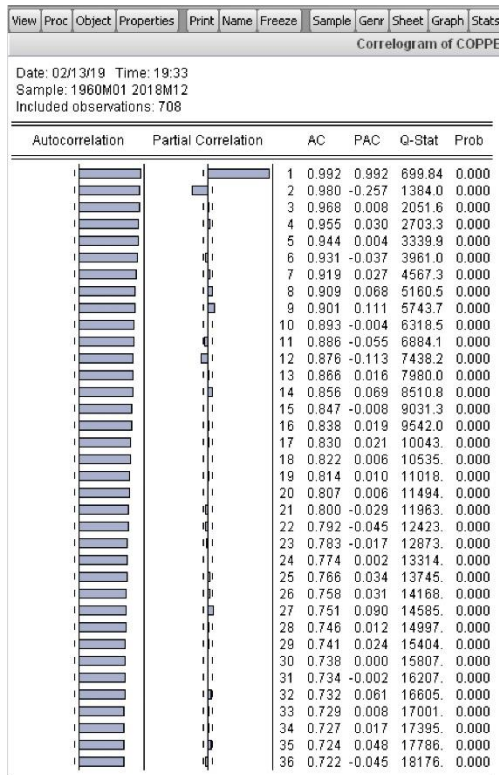
A0176595L ZHANG AO

A0186103M WANG XIN RUI

A0186040M LI LIPING

(a) Dickey-Fuller Tests

Does the series look stationary?



Based on the autocorrelation, the values range from 0.722 to 0.992, thus the series is not stationary

(ii) Carry out a basic Dickey-Fuller test with no constant or trend term:

View	Proc	Object	Properties	Print	Name	Freeze	Sample	Genr	Sheet	Graph	Stats	Ident
Augmented Dickey-Fuller Unit Root Test on COPPER												

Null Hypothesis: COPPER has a unit root
Exogenous: None
Lag Length: 1 (Automatic - based on SIC, maxlag=19)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.733304	0.3988
Test critical values:		
1% level	-2.568242	
5% level	-1.941272	
10% level	-1.616398	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(COPPER)
Method: Least Squares
Date: 02/13/19 Time: 19:35
Sample (adjusted): 1960M03 2018M12
Included observations: 706 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
COPPER(-1)	-0.001762	0.002403	-0.733304	0.4636
D(COPPER(-1))	0.335927	0.035588	9.439353	0.0000
R-squared	0.111507	Mean dependent var	7.573839	
Adjusted R-squared	0.110245	S.D. dependent var	243.9988	
S.E. of regression	230.1563	Akaike info criterion	13.71822	
Sum squared resid	37292239	Schwarz criterion	13.73114	
Log likelihood	-4840.533	Hannan-Quinn criter.	13.72321	
Durbin-Watson stat	1.964526			

The null hypothesis is that there exists unit root which means it is not stationary. From the ADF test above, the probability is above 0.05, therefore we cannot reject null hypothesis, thus it is not stationary.

(iii) Include a constant and a trend in the basic Dickey-Fuller test

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

Augmented Dickey-Fuller Unit Root Test on COPPER	
Null Hypothesis: COPPER has a unit root	
Exogenous: Constant, Linear Trend	
Lag Length: 1 (Automatic - based on SIC, maxlag=19)	

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-3.292650	0.0683
Test critical values:		
1% level	-3.971104	
5% level	-3.416195	
10% level	-3.130392	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(COPPER)
Method: Least Squares
Date: 02/13/19 Time: 19:37
Sample (adjusted): 1960M03 2018M12
Included observations: 706 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
COPPER(-1)	-0.020058	0.006092	-3.292650	0.0010
D(COPPER(-1))	0.343349	0.035445	9.686863	0.0000
C	-2.992700	17.33630	-0.172626	0.8630
@TREND("1960M01")	0.181759	0.067408	2.696395	0.0072

R-squared	0.124816	Mean dependent var	7.573839
Adjusted R-squared	0.121076	S.D. dependent var	243.9988
S.E. of regression	228.7512	Akaike info criterion	13.70880
Sum squared resid	36733624	Schwarz criterion	13.73463
Log likelihood	-4835.205	Hannan-Quinn criter.	13.71878
F-statistic	33.37245	Durbin-Watson stat	1.971804
Prob(F-statistic)	0.000000		

From the ADF test above, the probability is still above 0.05, therefore we cannot reject null hypothesis, thus it is not stationary.

(iv) Carry out Augmented Dickey-Fuller tests, including lagged values of the dependent variable

Series: RESID Workfile: SESSION 4 WORKSHOP:Untitled\

View

Proc

Object

Properties

Print

Name

Freeze

Sample

Genr

Sheet

Graph

Stats

Ident

Correlogram of RESID

Date: 02/13/19 Time: 19:39

Sample: 1960M01 2018M12

Included observations: 706

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.014	0.014	0.1372	0.711	
2	-0.030	-0.031	0.7918	0.673	
3	-0.021	-0.020	1.0975	0.778	
4	-0.028	-0.028	1.6539	0.799	
5	0.070	0.070	5.1854	0.394	
6	-0.052	-0.056	7.0883	0.313	
7	0.002	0.007	7.0918	0.419	
8	-0.169	-0.173	27.567	0.001	
9	-0.030	-0.022	29.218	0.001	
10	0.014	-0.006	29.349	0.002	
11	0.180	0.189	51.674	0.000	
12	0.071	0.053	55.268	0.000	
13	-0.104	-0.074	63.049	0.000	
14	-0.003	-0.016	63.054	0.000	
15	0.013	0.017	63.169	0.000	
16	0.005	-0.047	63.185	0.000	
17	0.052	0.059	65.147	0.000	
18	-0.021	-0.006	65.470	0.000	
19	-0.054	0.000	67.624	0.000	
20	-0.003	0.022	67.632	0.000	
21	0.078	0.062	72.074	0.000	
22	0.059	0.008	74.594	0.000	
23	0.054	0.051	76.734	0.000	
24	-0.050	-0.031	78.556	0.000	
25	-0.009	0.024	78.621	0.000	
26	-0.037	-0.068	79.653	0.000	
27	-0.024	-0.019	80.064	0.000	
28	-0.029	-0.055	80.699	0.000	
29	-0.041	-0.009	81.912	0.000	
30	0.036	0.065	82.863	0.000	
31	-0.081	-0.062	87.749	0.000	
32	0.016	-0.031	87.946	0.000	
33	0.020	0.002	88.244	0.000	
34	-0.015	-0.048	88.404	0.000	
35	0.045	0.043	89.927	0.000	
36	0.061	0.082	92.721	0.000	

File Series: RESID01 Workfile: SESSION 4 WORKSHOP:Untitled\

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

Augmented Dickey-Fuller Unit Root Test on RESID01

Null Hypothesis: RESID01 has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=19)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-26.12485	0.0000
Test critical values:	1% level	-3.971123
	5% level	-3.416204
	10% level	-3.130397

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

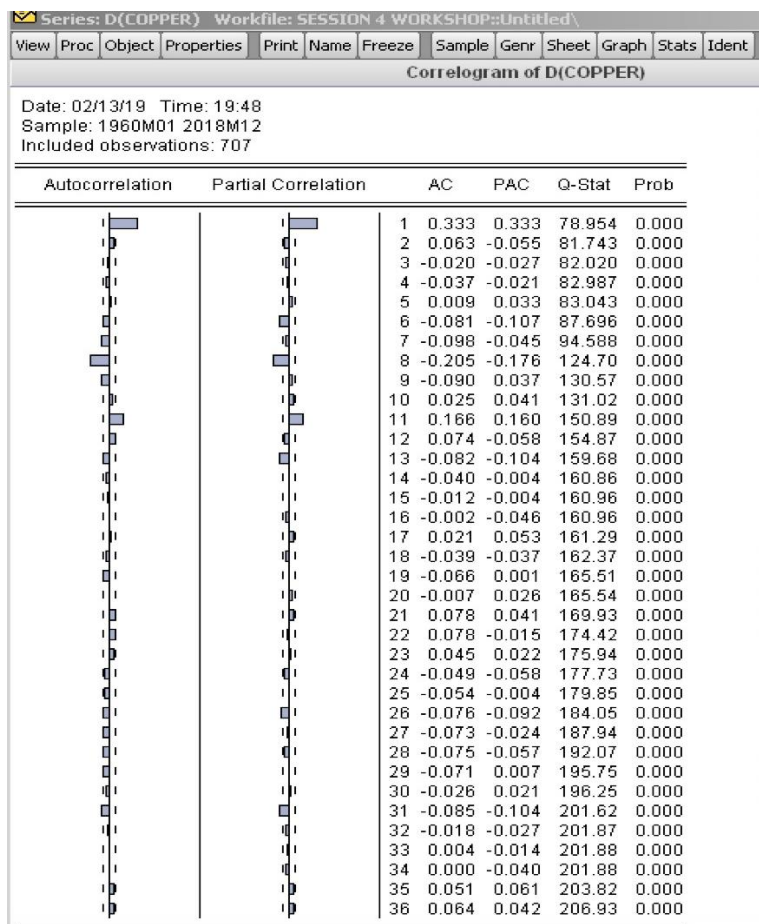
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(RESID01)
Method: Least Squares
Date: 02/13/19 Time: 19:45
Sample (adjusted): 1960M04 2018M12
Included observations: 705 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	-0.986085	0.037745	-26.12485	0.0000
C	0.178168	17.32039	0.010287	0.9918
@TREND("1960M01")	-0.000387	0.042328	-0.009136	0.9927

R-squared 0.492961 Mean dependent var -0.118159
Adjusted R-squared 0.491517 S.D. dependent var 320.7579
S.E. of regression 228.7261 Akaike info criterion 13.70717
Sum squared resid 36725585 Schwarz criterion 13.72657
Log likelihood -4828.779 Hannan-Quinn critier. 13.71467
F-statistic 341.2546 Durbin-Watson stat 1.998558
Prob(F-statistic) 0.000000

From the ADF test above, the probability is 0, therefore we can reject null hypothesis, the residual series is stationary.

(v) If the series is non-stationary, examine whether the first differences are stationary. Graph the differenced series, plot its correlogram and then perform unit root tests. (The difference operator is D(copper) in EViews.)



Series: D(COPPER) Workfile: SESSION 4 WORKSHOP:Untitled\

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

Augmented Dickey-Fuller Unit Root Test on D(COPPER)

Null Hypothesis: D(COPPER) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=19)

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-18.74024	0.0000
Test critical values:		
1% level	-3.971104	
5% level	-3.416195	
10% level	-3.130392	

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

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(COPPER,2)
Method: Least Squares
Date: 02/13/19 Time: 19:49
Sample (adjusted): 1960M03 2018M12
Included observations: 706 after adjustments

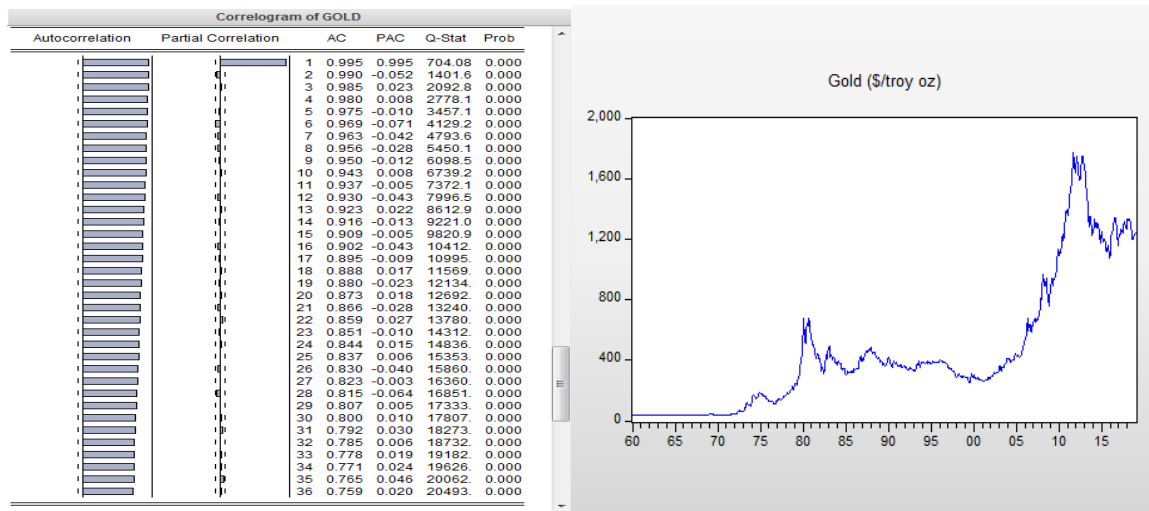
Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(COPPER(-1))	-0.666494	0.035565	-18.74024	0.0000
C	1.863961	17.39392	0.107162	0.9147
@TREND("1960M01")	0.008804	0.042540	0.206951	0.8361

R-squared	0.333142	Mean dependent var	-0.188938
Adjusted R-squared	0.331245	S.D. dependent var	281.6753
S.E. of regression	230.3468	Akaike info criterion	13.72129
Sum squared resid	37300931	Schwarz criterion	13.74066
Log likelihood	-4840.615	Hannan-Quinn criter.	13.72878
F-statistic	175.5890	Durbin-Watson stat	1.963010
Prob(F-statistic)	0.000000		

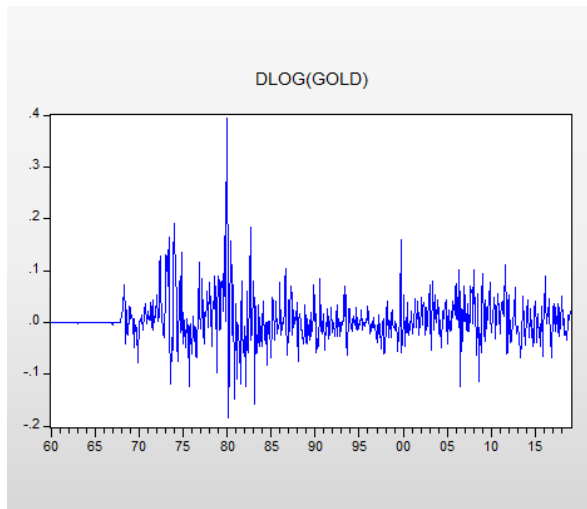
From the correlogram, the autocorrelation pattern indicates that it is stationary, similar result can be observed based on the probability of the ADF test statistics which is 0, shows that we can reject the null hypothesis, and series is stationary.

(b) Arima Modelling

(i) Establishing Stationarity



After Set the new workfile structure type to start from 1985M01:



(ii) Identification

Plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the series DLOG (gold).

Correlogram of DLOG(GOLD)					
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 0.136	0.136	7.6014	0.006
		2 -0.085	-0.106	10.580	0.005
		3 -0.000	0.028	10.580	0.014
		4 0.047	0.035	11.506	0.021
		5 0.018	0.009	11.647	0.040
		6 0.005	0.009	11.657	0.070
		7 0.072	0.074	13.829	0.054
		8 0.024	0.001	14.060	0.080
		9 -0.013	-0.004	14.130	0.118
		10 0.009	0.013	14.163	0.166
		11 0.154	0.148	24.152	0.012
		12 0.030	-0.017	24.534	0.017
		13 -0.001	0.027	24.534	0.027
		14 0.035	0.027	25.056	0.034
		15 0.045	0.027	25.923	0.039
		16 0.042	0.034	26.664	0.045
		17 0.019	0.015	26.820	0.061
		18 0.057	0.038	28.209	0.059
		19 0.021	0.002	28.390	0.076
		20 0.081	0.089	31.181	0.053
		21 0.027	-0.002	31.496	0.066
		22 0.036	0.020	32.060	0.076
		23 0.037	0.025	32.656	0.087
		24 0.047	0.037	33.608	0.092
		25 0.009	-0.016	33.647	0.116
		26 0.062	0.064	35.321	0.105
		27 -0.010	-0.056	35.367	0.130
		28 0.001	0.013	35.368	0.159
		29 -0.001	-0.030	35.368	0.193
		30 -0.024	-0.031	35.615	0.221
		31 0.052	0.025	36.817	0.218
		32 -0.005	-0.033	36.829	0.255
		33 -0.043	-0.052	37.634	0.265
		34 -0.047	-0.056	38.641	0.268
		35 0.051	0.038	39.790	0.265
		36 -0.032	-0.073	40.258	0.287

(iii) Estimation

Dependent Variable: DLOG(GOLD)

Method: ARMA Maximum Likelihood (OPG - BHHH)

Date: 02/13/19 Time: 19:56

Sample: 1985M02 2018M12

Included observations: 407

Convergence achieved after 16 iterations

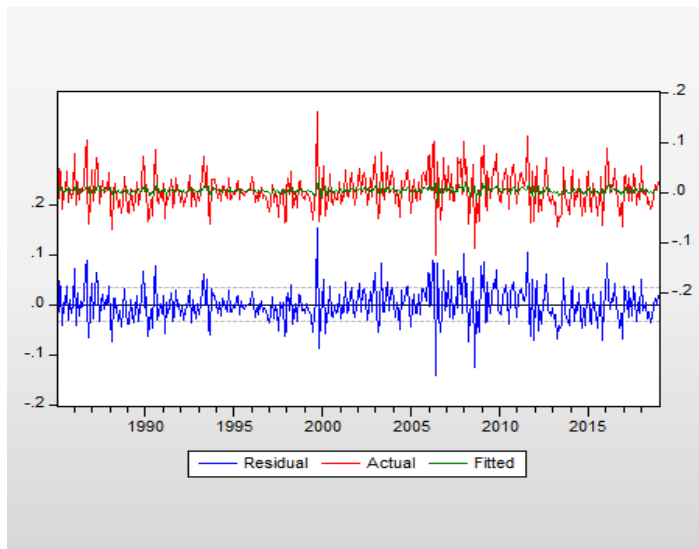
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003484	0.002086	1.670293	0.0956
MA(1)	0.167650	0.040449	4.144712	0.0000
SIGMASQ	0.001196	6.34E-05	18.84932	0.0000
R-squared	0.022957	Mean dependent var		0.003484
Adjusted R-squared	0.018120	S.D. dependent var		0.035029
S.E. of regression	0.034710	Akaike info criterion		-3.876179
Sum squared resid	0.486726	Schwarz criterion		-3.846630
Log likelihood	791.8023	Hannan-Quinn criter.		-3.864485
F-statistic	4.746237	Durbin-Watson stat		2.026107
Prob(F-statistic)	0.009174			
Inverted MA Roots	-.17			

From the above statistics, both of the coefficients are significant, the model looks quite reasonable.

(iv) Testing

Do the residuals look random?



From the graph above, the residuals look random. The mean of residuals is approximately zero, which indicates random walk.

Does the correlogram suggest random residuals?

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.014	-0.014	0.0774	
		2 -0.083	-0.084	2.9417	0.086
		3 0.006	0.004	2.9572	0.228
		4 0.044	0.037	3.7588	0.289
		5 0.012	0.015	3.8229	0.431
		6 -0.009	-0.002	3.8558	0.570
		7 0.071	0.073	5.9516	0.429
		8 0.014	0.013	6.0305	0.536
		9 -0.013	-0.002	6.0984	0.636
		10 -0.014	-0.013	6.1854	0.721
		11 0.155	0.150	16.262	0.092
		12 0.006	0.005	16.277	0.131
		13 -0.007	0.020	16.296	0.178
		14 0.030	0.028	16.689	0.214
		15 0.034	0.025	17.181	0.247
		16 0.035	0.037	17.715	0.278
		17 0.004	0.014	17.723	0.340
		18 0.056	0.042	19.070	0.325
		19 -0.001	-0.006	19.071	0.387
		20 0.079	0.087	21.741	0.297
		21 0.009	0.011	21.778	0.353
		22 0.030	0.016	22.175	0.389
		23 0.025	0.021	22.446	0.434
		24 0.044	0.045	23.295	0.444
		25 -0.009	-0.021	23.332	0.500
		26 0.067	0.069	25.267	0.447
		27 -0.022	-0.047	25.478	0.492
		28 0.003	0.007	25.483	0.547
		29 0.004	-0.022	25.491	0.601
		30 -0.034	-0.040	25.989	0.626
		31 0.059	0.024	27.526	0.596
		32 -0.009	-0.021	27.565	0.644
		33 -0.032	-0.046	28.030	0.668
		34 -0.052	-0.071	29.253	0.654
		35 0.065	0.038	31.144	0.608
		36 -0.033	-0.054	31.638	0.631

The correlogram suggest random residuals.

The Box-Pierce Q-statistic: no p value under 5%, thus there is no Q-statistics is significant. Thus we cannot reject null hypothesis, the model fits the data well.

To find an alternative model which has better or comparable fit, and which also has no residual autocorrelation, we tried the following models:

ARMA (2, 0):

Dependent Variable: DLOG(GOLD)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/13/19 Time: 20:01
Sample: 1985M02 2018M12
Included observations: 407
Convergence achieved after 15 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003473	0.001897	1.830899	0.0679
AR(1)	0.150628	0.039910	3.774163	0.0002
AR(2)	-0.105918	0.043571	-2.430929	0.0155
SIGMASQ	0.001188	6.36E-05	18.67461	0.0000
R-squared	0.029621	Mean dependent var		0.003484
Adjusted R-squared	0.022398	S.D. dependent var		0.035029
S.E. of regression	0.034634	Akaike info criterion		-3.878078
Sum squared resid	0.483406	Schwarz criterion		-3.838679
Log likelihood	793.1888	Hannan-Quinn criter.		-3.862486
F-statistic	4.100590	Durbin-Watson stat		1.992539
Prob(F-statistic)	0.006942			
Inverted AR Roots	.08+.32i	.08-.32i		

ARMA (0, 2):

Dependent Variable: DLOG(GOLD)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/13/19 Time: 20:01
Sample: 1985M02 2018M12
Included observations: 407
Convergence achieved after 16 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003479	0.001948	1.786143	0.0748
MA(1)	0.151899	0.040588	3.742491	0.0002
MA(2)	-0.079066	0.043340	-1.824306	0.0688
SIGMASQ	0.001188	6.36E-05	18.67115	0.0000
R-squared	0.029432	Mean dependent var		0.003484
Adjusted R-squared	0.022207	S.D. dependent var		0.035029
S.E. of regression	0.034637	Akaike info criterion		-3.877886
Sum squared resid	0.483500	Schwarz criterion		-3.838487
Log likelihood	793.1497	Hannan-Quinn criter.		-3.862294
F-statistic	4.073607	Durbin-Watson stat		1.996835
Prob(F-statistic)	0.007200			
Inverted MA Roots	.22	-.37		

ARMA (1, 0):

Dependent Variable: DLOG(GOLD)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/13/19 Time: 20:02
Sample: 1985M02 2018M12
Included observations: 407
Convergence achieved after 11 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003486	0.002101	1.659647	0.0978
AR(1)	0.136010	0.039860	3.412218	0.0007
SIGMASQ	0.001201	6.43E-05	18.67665	0.0000
R-squared	0.018565	Mean dependent var		0.003484
Adjusted R-squared	0.013706	S.D. dependent var		0.035029
S.E. of regression	0.034788	Akaike info criterion		-3.871717
Sum squared resid	0.488914	Schwarz criterion		-3.842168
Log likelihood	790.8945	Hannan-Quinn criter.		-3.860024
F-statistic	3.820989	Durbin-Watson stat		1.969347
Prob(F-statistic)	0.022702			
Inverted AR Roots	.14			

ARMA (1, 1):

Dependent Variable: DLOG(GOLD)
Method: ARMA Maximum Likelihood (OPG - BHHH)
Date: 02/13/19 Time: 20:02
Sample: 1985M02 2018M12
Included observations: 407
Convergence achieved after 14 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003484	0.002013	1.731338	0.0842
AR(1)	-0.330921	0.215195	-1.537772	0.1249
MA(1)	0.490279	0.203645	2.407521	0.0165
SIGMASQ	0.001190	6.33E-05	18.80114	0.0000
R-squared	0.027818	Mean dependent var		0.003484
Adjusted R-squared	0.020581	S.D. dependent var		0.035029
S.E. of regression	0.034666	Akaike info criterion		-3.876232
Sum squared resid	0.484304	Schwarz criterion		-3.836833
Log likelihood	792.8132	Hannan-Quinn criter.		-3.860640
F-statistic	3.843834	Durbin-Watson stat		2.011501
Prob(F-statistic)	0.009822			
Inverted AR Roots	-.33			
Inverted MA Roots	-.49			

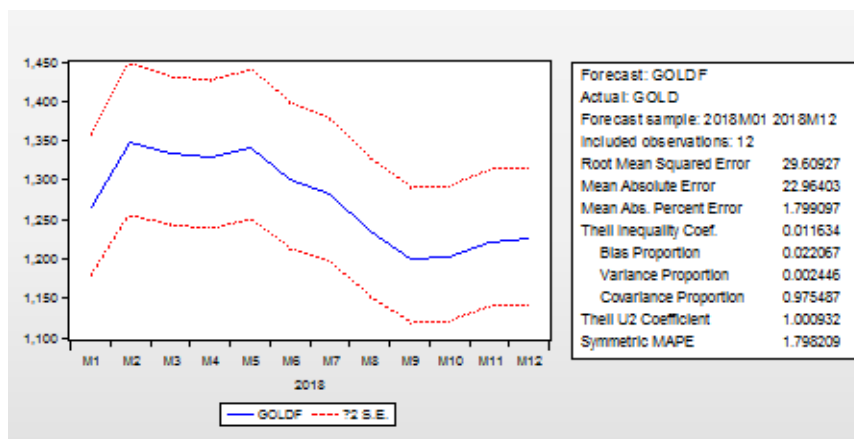
ARMA (2, 0) model gives the lowest AIC (-3.878), so we choose $p = 2$, $q = 0$ as the final parameters for our model. The coefficients of AR (2,0) are significant.

(v) Forecasting

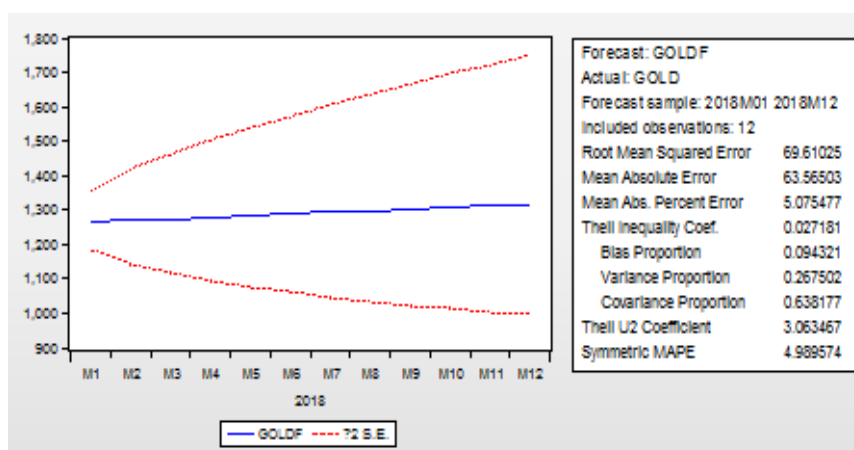
Dependent Variable: DLOG(GOLD)
Method: ARMA Generalized Least Squares (BFGS)
Date: 02/13/19 Time: 20:07
Sample: 1985M02 2017M12
Included observations: 395
Convergence achieved after 3 iterations
Coefficient covariance computed using outer product of gradients
d.f. adjustment for standard errors & covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.003602	0.002054	1.753633	0.0803
MA(1)	0.168036	0.049747	3.377795	0.0008
R-squared	0.022688	Mean dependent var		0.003619
Adjusted R-squared	0.020201	S.D. dependent var		0.035318
S.E. of regression	0.034959	Akaike info criterion		-3.864134
Sum squared resid	0.480310	Schwarz criterion		-3.843988
Log likelihood	765.1665	Hannan-Quinn criter.		-3.856152
F-statistic	9.123316	Durbin-Watson stat		2.028732
Prob(F-statistic)	0.002689			
Inverted MA Roots	-.17			

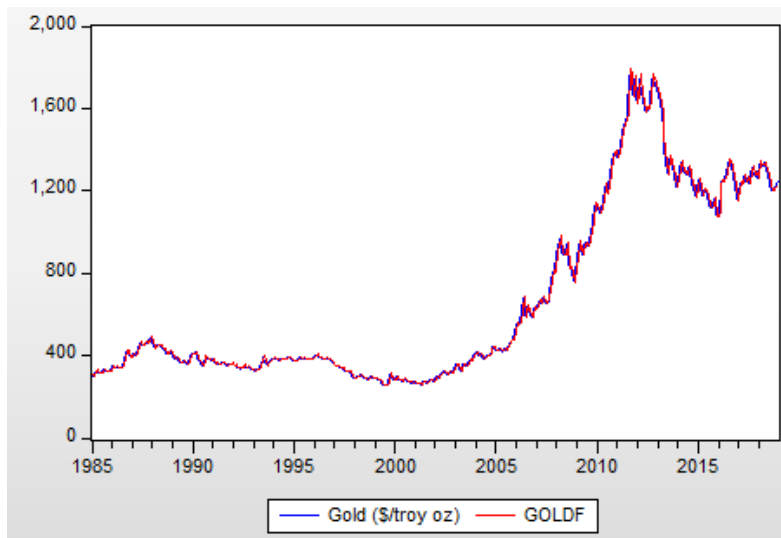
Static forecast



Dynamic forecast

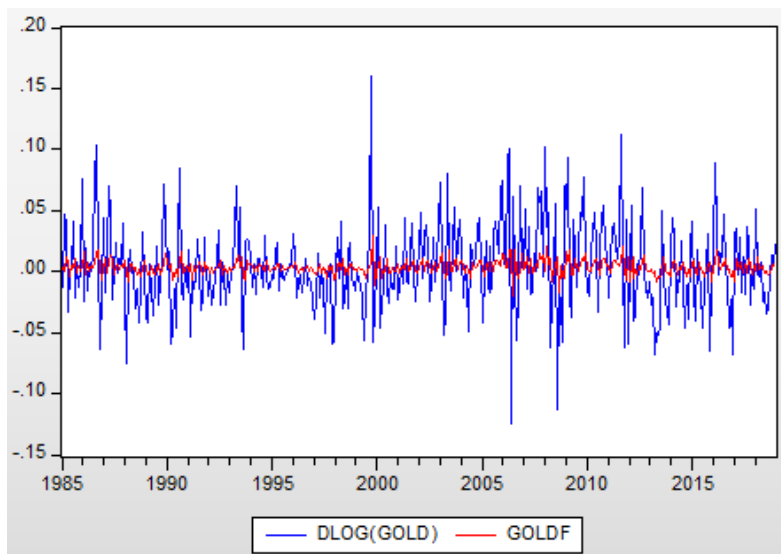


Produce Static forecasts for Gold and then plot these forecasts, goldF, and gold on the same graph by selecting Quick Graph. How good is the fit?



From the graph above, we can see the actual and forecast for gold fit quite well.

Produce Static forecasts for DLOG(gold) by specifying this at the top left of the forecast dialog box. Plot the forecasts and DLOG(gold) on the same graph. How good is the fit?



The graph above shows that the Dlog and GoldF are relatively stable at zero. Thus they are fitted quite good.

Do you expect gold prices to go up or down in the next 12 months? How about risk? What is the prediction interval?

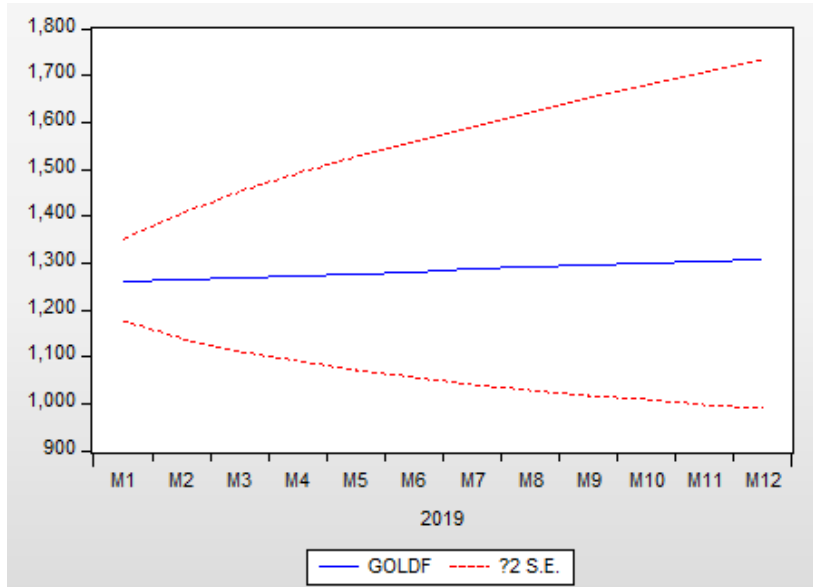
From the dynamic forecast, the gold price trends towards 1300 dollars.

The solid line in the graph represents the prediction, and two dashed lines provide the confidence interval with 2 standard deviations. With the increase of predictive period, the prediction moves towards the mean of the series. Next to the graph, we have a Theil inequality coefficient of 0.027, which means the model has a good predictive ability. The

decomposition implies that the proportion of bias is low and the proportion of covariance is high. So the actual series fluctuates more seriously than the prediction series.

Click the Forecast button which can be found in the menu at the top of the current equation window. [FORECAST observations 2019m01 to 2019m12]. In the top right of the dialog box, try a Dynamic forecasting Method and click OK.

Why can't you use the Static Forecasting method?



The curve is almost horizontal, one of the potential reason is that the forecasting period is too far away from the actual period.