Diane Nguyen

MSBA 635 Data Analytics 2

2/1/19

Homework 3

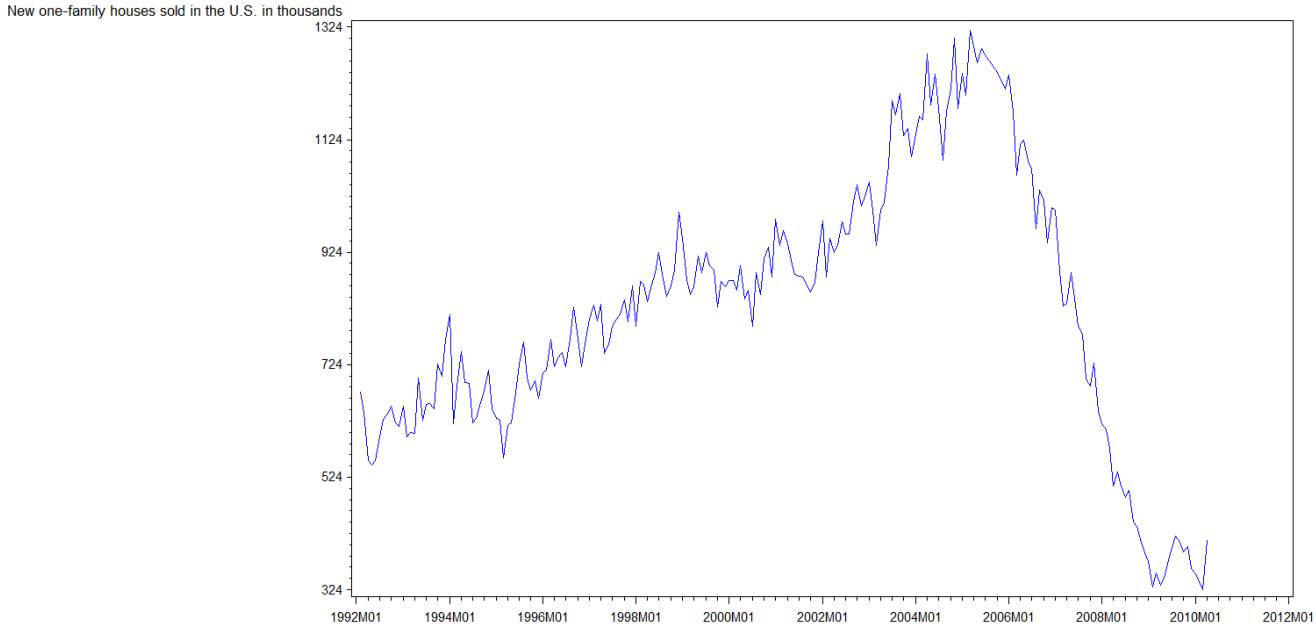
1) Use the homes data for the following:

a) Plot HOMES, DHOMES, IRATE, and DIRATE (where the letter D implies a first difference) with month and year on the horizontal axis using SAS 9.4 and proc gplot.

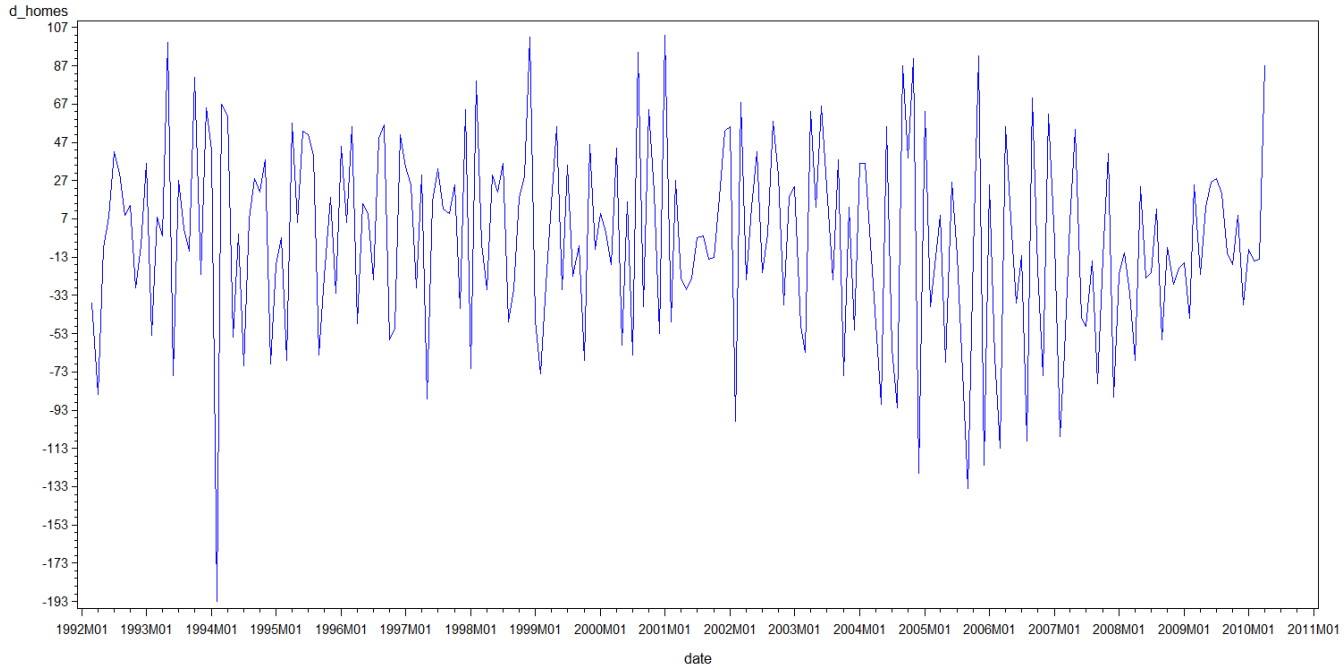
```
*print the data;
proc print data=tmp1.homes (obs=10);
run;

*display data attributes;
proc contents data=tmp1.homes;
run;

data homes_hw3;
set tmp1.homes;
d_homes=dif(homes);
d_homes_1=lag(d_homes);
d_homes_5=lag5(d_homes);
d_irate=dif(irate);
d_irate_1=lag(d_irate);
d_irate_3=lag3(d_irate);
retain date '31dec91'd;
date=intnx('month',date,1);
*fix the format;
format date yymm.;
year=1992+int(_n_/12);
month=mod(_n_/12)+1;
*date=_n_+1950;

proc print data=work.homes_hw3 (obs=20);
run;
quit;
*a) Plot HOMES, DHOMES, IRATE, and DIRATE (where the letter D
implies a first difference) with month and year on the
horizontal axis using SAS 9.4 and proc gplot.;
proc univariate data=work.homes_hw3;
run;
quit;

symbol1 value=none interpol=join color=blue;
axis1 order=(324 to 1389 by 200);
axis2 order=(-193 to 117 by 20);
axis3 order=(4.5 to 9.5 by .5);
axis4 order=(-1 to 1 by .5);

proc gplot data=work.homes_hw3;
plot homes*date/vaxis=axis1;
plot d_homes*date/vaxis=axis2;
plot irate*date/vaxis=axis3;
plot d_irate*date/vaxis=axis4;
run;
quit;
```

Below shows Homes, D_homes, Irate, and D_irate for the years 1992 up to 2010 for the data using proc gplot.

Homes by date

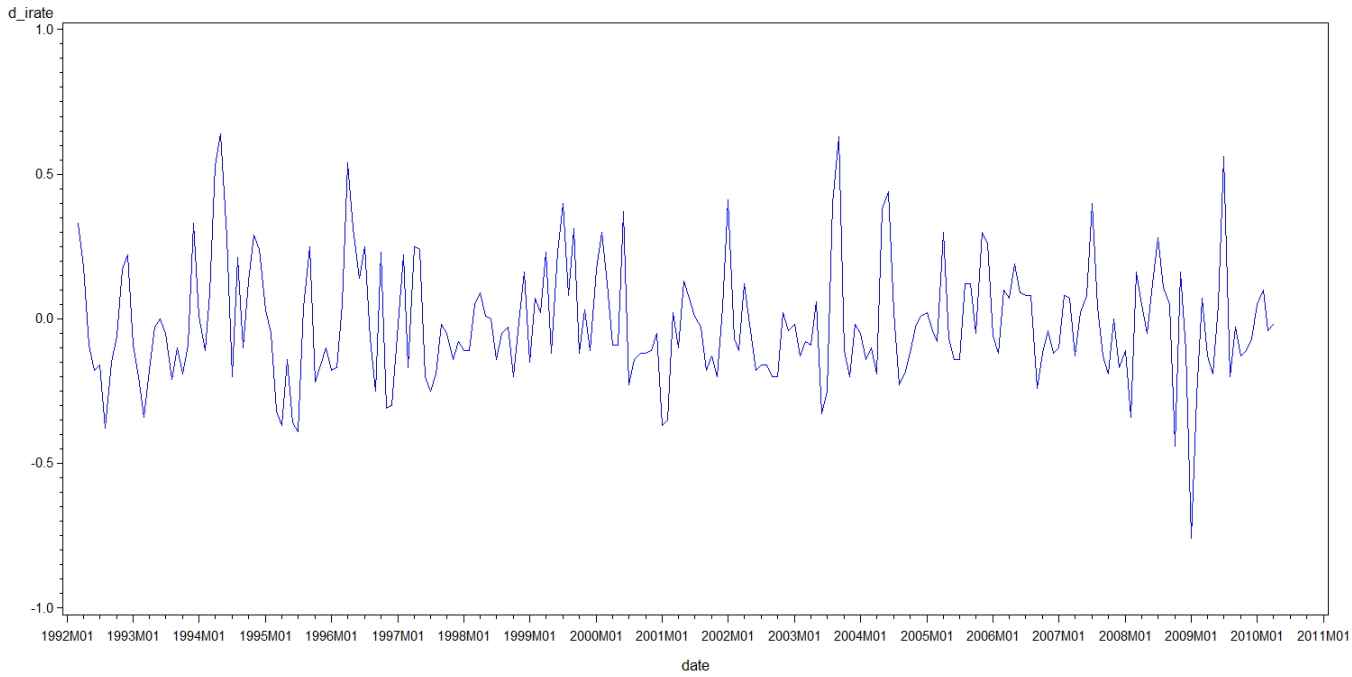


D_homes by date



I_rate by date

D_irate by date



b) Estimate the following ARDL model using SAS 9.4 and proc autoreg:

$$DHOMES_t = \beta_1 + \beta_2 HOMES_{t-1} + \beta_3 DHOMES_{t-5} + \beta_4 DIRATE_{t-1} + \beta_5 DIRATE_{t-3} + \varepsilon_t$$

```
proc autoreg data=work.homes_hw3;
model d_homes= d_homes_1 d_homes_5 d_irate_1 d_irate_3;
run;
quit;
```

The SAS System 07:22 Sunday, January 27, 2019 25

The AUTOREG Procedure

Dependent Variable d_homes

Ordinary Least Squares Estimates

SSE	469339.262	DFE	208
MSE	2256	Root MSE	47.50199
SBC	2270.90333	AIC	2254.09687
MAE	36.8414518	AICC	2254.38673
MAPE	131.702874	HQC	2260.88892
Durbin-Watson	2.0541	Total R-Square	0.2026

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-2.9215	3.2841	-0.89	0.3747
d_homes_1	1	-0.3073	0.0635	-4.84	<.0001
d_homes_5	1	0.2069	0.0633	3.27	0.0013
d_irate_1	1	-64.3245	15.9739	-4.03	<.0001
d_irate_3	1	-46.6306	16.0940	-2.90	0.0042

Ascertain the statistical significance of each slope coefficient (i.e., skip intercept).

The total R-square is .2026 meaning that 20.26% variance of this model can be explained.

- D_homes_1 is statistically significant at the .05 alpha level of significance since p=<.0001 < .05. For every one unit increase there is a .3073 decrease in d_home_1.
- D_homes_5 is statistically significant at the .05 alpha level of significance since p=0.0013 < .05. For every one unit increase there is a .2069 increase in d_home_5.
- D_irate_1 is statistically significant at the .05 alpha level of significance since p=<.0001 < .05. For every one unit increase there is a 64.32 decrease in d_irate_1.
- D_irate_3 is statistically significant at the .05 alpha level of significance since p=0.0042 <.05. For every one unit increase there is a 46.63 decrease in d_irate_3.

c) Conduct and interpret the Lagrange multiplier tests for serially correlated errors up to and including five lags

```
ods graphics on;
proc autoreg data=work.homes_hw3 plots(only)=(acf);
model d_homes= d_homes_1 d_homes_5 d_irate_1 d_irate_3/godfrey=5;
output out=yhatout p=yhat;
run;
ods graphics off;
quit;
```


The AUTOREG Procedure

Dependent Variable d_homes

Ordinary Least Squares Estimates

SSE	469339.262	DFE	208
MSE	2256	Root MSE	47.50199
SBC	2270.90333	AIC	2254.09687
MAE	36.8414518	AICC	2254.38673
MAPE	131.702874	HQC	2260.88892
Durbin-Watson	2.0541	Total R-Square	0.2026

Godfrey's Serial Correlation Test

Alternative	LM	Pr > LM
AR(1)	1.4826	0.2234
AR(2)	2.7602	0.2516
AR(3)	3.0685	0.3812
AR(4)	5.0241	0.2848
AR(5)	5.0262	0.4127

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	-2.9215	3.2841	-0.89	0.3747
d_homes_1	1	-0.3073	0.0635	-4.84	<.0001
d_homes_5	1	0.2069	0.0633	3.27	0.0013
d_irate_1	1	-64.3245	15.9739	-4.03	<.0001
d_irate_3	1	-46.6306	16.0940	-2.90	0.0042

The Total R-square for this model is 0.2026 meaning that only 20.26% variance of the model can be explained with our data. Here we are looking at Godfrey's serial correlation test. The null hypothesis here states that we do not have Serial correlation. Based on the p-values we fail to reject the null hypothesis. Therefore we don't have first order serial correlation.

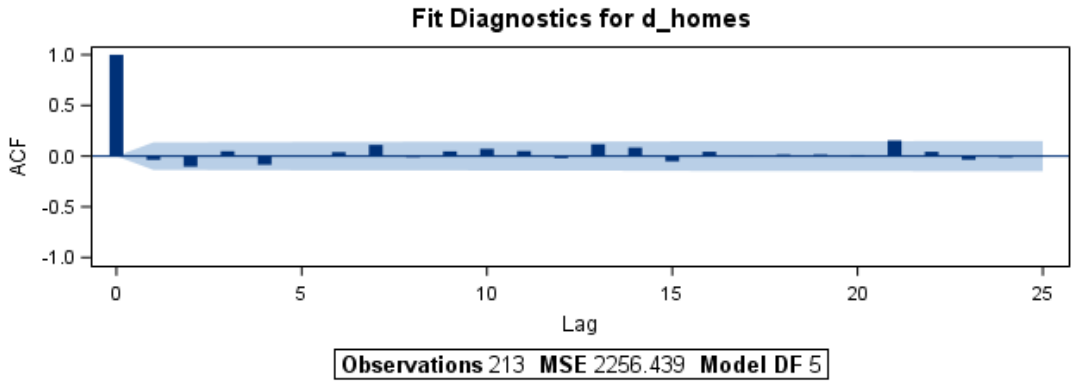
The Lagrange Multiplier test for serially correlated errors for the 5 lags examines the Pr > LM column of values. The values for each time period are not correlated so they are purged from our model. We failed to reject the null going back 5 time periods. None of the p-values are statistically significant.

- AR(1) is not statistically significant at the .05 level of significance since .2234 > .05
- AR(2) is not statistically significant at the .05 level of significance since .2516 > .05
- AR(3) is not statistically significant at the .05 level of significance since .3812 > .05

- AR(4) is not statistically significant at the .05 level of significance since .2848 > .05
- AR(5) is not statistically significant at the .05 level of significance since .4127 > .05

d) Obtain and interpret the error correlogram (ACF plot), and compare to your results in part (c).

According to the correlogram (ACF plot) below, if the data is within the light blue ban then we do not have any singular collinearity. According to the results we don't have first order serial correlation which corresponds to the results that we found in part (c).



e) Plot actual and predicted DHOMES using SAS 9.4 and proc gplot.

```
symbol1 value=none interpol=join color=blue;
symbol2 value=dot interpol=join color=black;
proc gplot data=work.yhatout;
plot d_homes*date=1 yhat*date=2/overlay;
run;
quit;
```

Below is Dhomes actual and predicted plotted on a single graph using proc gplot.