```
* MSBA 635 - Data Analytics II;

* print data;
*Time Series Example 1 - ARDL Models, Exponential Smoothing, and Forecasting
```

*__Autoregressive distributed lag__ (ARDL) models are heavily used in forecasting where lagged values of the dependent variable (AR or autoregressive terms) and lagged values of the independent variables (DL or distributed lag terms) constitute the right hand side of the regression model

*__Exponential smoothing__ (ES) models are not regression models, but are essentially weighted averages of past values of a time series variable used to predict future values of it

*Time series are either **stationary** (meanreverting) hence average looks flat over time, or **nonstationary** (not mean-reverting) hence average looks upward-sloping or downward sloping over time – How to test for **nonstationarity** (how to diagnose a nonstationary series) – How to convert a nonstationary series into a stationary series, then build ARDL and other time series models on that stationary series

*Explain variability in change in unemployment rate as a function of lagged change in unemployment rate and lagged growth rate of GDP

```
proc print data=tmp1.okun (obs=10);
run;
```

The SAS System   17:07 Tuesday, January 22, 2019   1

Obs	g	u
1	1.4	7.3
2	2.0	7.2
3	1.4	7.0
4	1.5	7.0
5	0.9	7.2
6	1.5	7.0
7	1.2	6.8
8	1.5	6.6
9	1.6	6.3
10	1.7	6.0

```
* display data attributes;

proc contents data=tmp1.okun;
run;
```

```
The SAS System         17:07 Tuesday, January 22, 2019   2

                                        The CONTENTS Procedure

             Data Set Name        TMP1.OKUN                      Observations          98
             Member Type          DATA                           Variables             2
             Engine               V9                             Indexes               0
             Created              07/02/2010 10:38:47            Observation Length    16
             Last Modified        07/02/2010 10:38:47            Deleted Observations  0
             Protection                                          Compressed            NO
             Data Set Type                                       Sorted                NO
             Label
             Data Representation  WINDOWS_32
             Encoding             wlatin1  Western (Windows)


                                  Engine/Host Dependent Information

                Data Set Page Size          4096
                Number of Data Set Pages    1
                First Data Page             1
                Max Obs per Page            252
                Obs in First Data Page      98
                Number of Data Set Repairs  0
                Filename                    C:\Users\nxnguy01\Downloads\okun.sas7bdat
                Release Created             9.0202M0
                Host Created                XP_PRO


                             Alphabetic List of Variables and Attributes

#    Variable    Type    Len    Label

1    g           Num       8    percentage change in U.S. Gross Domestic Product, seasonally adjusted
2    u           Num       8    U.S. Civilian Unemployment Rate (Seasonally adjusted)
```

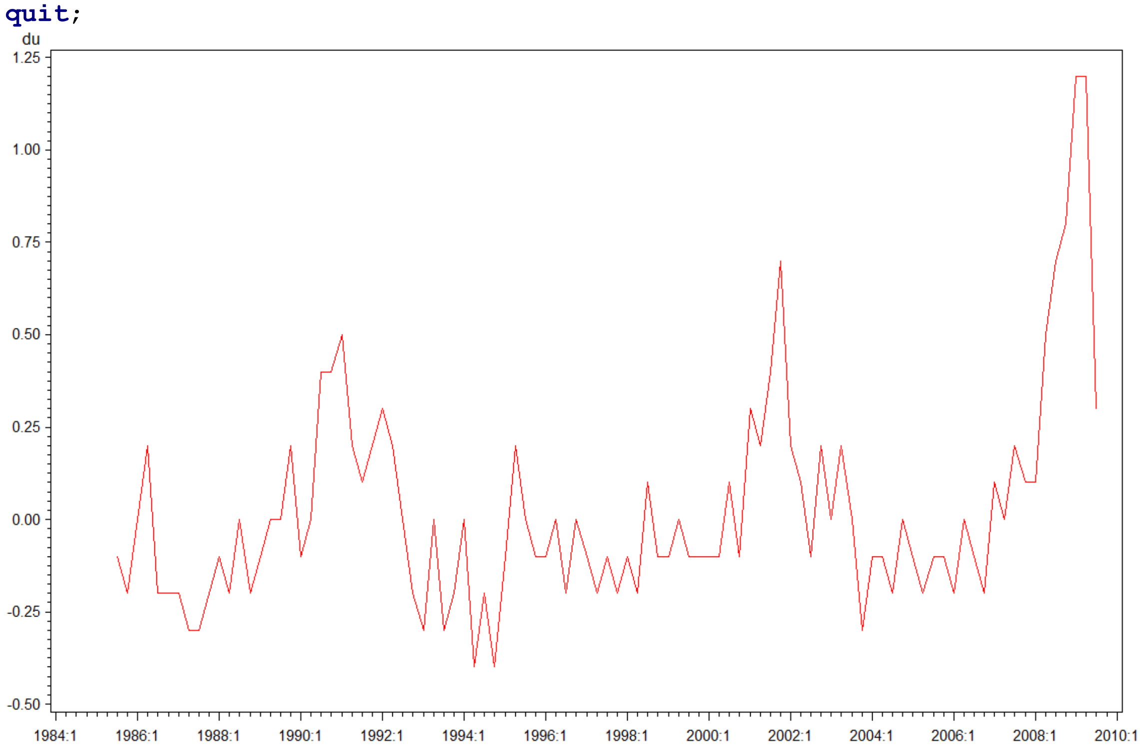
```
* construct new variables;
*you set up variables, do log transforms, do lags, do dates, create indicator variables
here!!!! Super important

data okundata;
set tmp1.okun;
du = dif(u);
du1 = lag(du);
g1 = lag(g);
g2 = lag2(g);
g3 = lag3(g);
retain date '1jan85'd;
date = intnx('qtr',date,1);
format date yyqc.;
year = 1985 + int(_n_/4);
qtr = mod(_n_,4) + 1;
run;
```

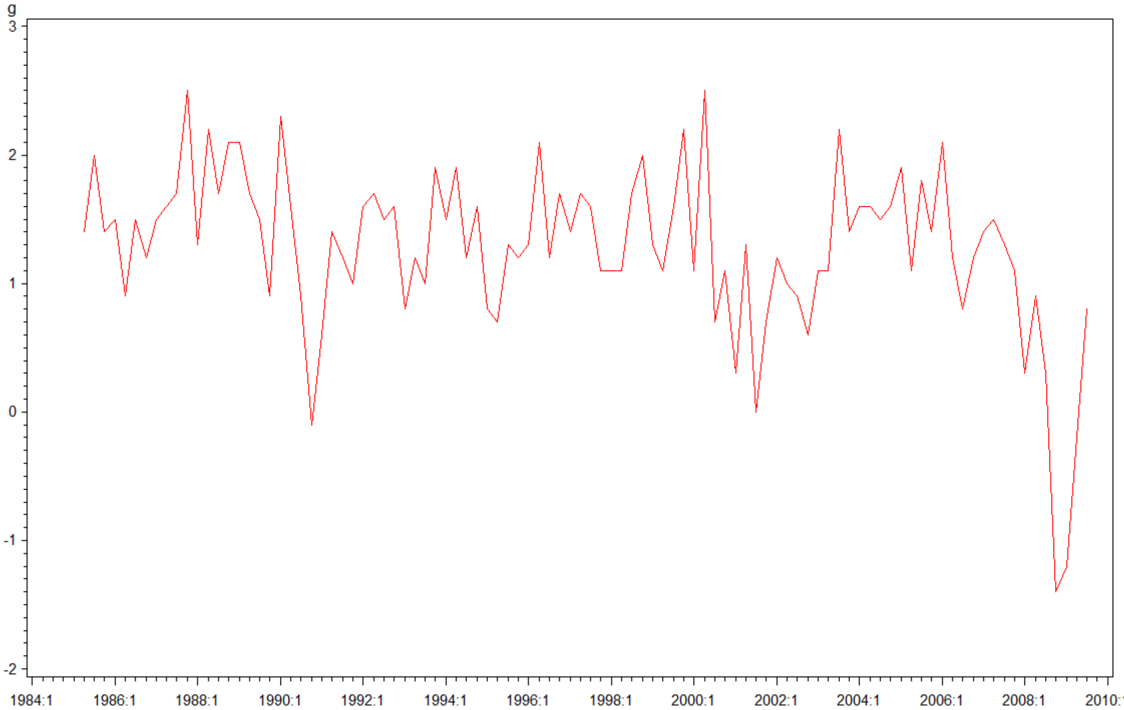
```
* print data;
*see what happened to g and u here. See what differencing does, what lagging does for the
date, year and qtr

proc print data=work.okundata (obs=10);
run;
```

The SAS System 17:07 Tuesday, January 22, 2019 3

Obs	g	u	du	du1	g1	g2	g3	date	year	qtr
1	1.4	7.3	.	.	.	.	.	1985:2	1985	2
2	2.0	7.2	-0.1	.	1.4	.	.	1985:3	1985	3
3	1.4	7.0	-0.2	-0.1	2.0	1.4	.	1985:4	1985	4
4	1.5	7.0	0.0	-0.2	1.4	2.0	1.4	1986:1	1986	1
5	0.9	7.2	0.2	0.0	1.5	1.4	2.0	1986:2	1986	2
6	1.5	7.0	-0.2	0.2	0.9	1.5	1.4	1986:3	1986	3
7	1.2	6.8	-0.2	-0.2	1.5	0.9	1.5	1986:4	1986	4
8	1.5	6.6	-0.2	-0.2	1.2	1.5	0.9	1987:1	1987	1
9	1.6	6.3	-0.3	-0.2	1.5	1.2	1.5	1987:2	1987	2
10	1.7	6.0	-0.3	-0.3	1.6	1.5	1.2	1987:3	1987	3

```
* plot data;
*do a lot of plotting in time-series, we want to look at our data
*change in unemployment is the first graph
*change in gdp is the second graph

options nolabel;
symbol1 value=none interpol=join color=red;
axis1 order=(-0.50 to 1.25 by 0.25);
axis2 order=(-2 to 3 by 1);
proc gplot data=work.okundata;
plot du*date / vaxis=axis1;
plot g*date / vaxis=axis2;
run;
quit;
```





g
3
2
1
0
-1
-2
1984:1
1986:1
1988:1
1990:1
1992:1
1994:1
1996:1
1998:1
2000:1
2002:1
2004:1
2006:1
2008:1
2010:1

```
* estimate distributed lag model using proc autoreg;
*first time using proc autoreg that lets us do time series testing and things like it
*on the right hand side you have different measures of g. (g, g1, g2, g3)
*put a lot in and the one's that are insignificant you prune back
*slope on g1, and g2 are statistically significant
*g3 not statistically significant so you prune it back. The point is to fit the data over a period of time!
*do past values in predicted values predict current unemployment? This is what the lag means. You are lagging g1, g2, and g3. g is just there as an extraneous business cycle

options nolabel;
proc autoreg data=work.okundata;
model du = g g1 g2 g3;
run;
quit;
```

The SAS System 17:07 Tuesday, January 22, 2019 4

The AUTOREG Procedure

Dependent Variable du

Ordinary Least Squares Estimates

SSE	2.73516422	DFE	90
MSE	0.03039	Root MSE	0.17433
SBC	-44.662409	AIC	-57.431793
MAE	0.13705398	AICC	-56.757636
MAPE	73.46055	HQC	-52.272004
Durbin-Watson	1.2741	Total R-Square	0.6524

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.5810	0.0539	10.78	<.0001
g	1	-0.2021	0.0330	-6.12	<.0001
g1	1	-0.1645	0.0358	-4.59	<.0001
g2	1	-0.0716	0.0353	-2.03	**0.0456**
g3	1	0.003303	0.0363	0.09	**0.9276**

```
* estimate distributed lag model using proc autoreg;
*this model is the same, but you got rid of g3
*when you throw out g3 notice that it impacted g, g1 and g2
*65.39% of this model is explained

options nolabel;
proc autoreg data=work.okundata;
model du = g g1 g2;
run;
quit;
```


The AUTOREG Procedure

Dependent Variable du

Ordinary Least Squares Estimates

SSE	2.74074794	DFE	92
MSE	0.02979	Root MSE	0.17260
SBC	-50.693673	AIC	-60.951066
MAE	0.13649738	AICC	-60.511505
MAPE	73.0423003	HQC	-56.80486
Durbin-Watson	1.2829	Total R-Square	**0.6539**

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.5836	0.0472	12.36	<.0001
g	1	-0.2020	0.0324	-6.24	<.0001
g1	1	-0.1653	0.0335	-4.93	<.0001
g2	1	-0.0700	0.0331	-2.12	**0.0371**

```
* estimate autoregressive distributed lag model using proc autoreg;
*try to get to your best model specification here
*acf=auto correlation function. You want your plots and you only want your acf
*dul is your autoregressive; Godfrey = 5 will do some testing for us
*ouput out=yhat out and you are saving your predicted variables
*you have good statistical significant for g1 at 0.0084
*whenever you run a time series model you have an a1 term model where the estimate is
between 0 and 1 (0.0992) and highly statistically significant. Total R-square is 0.6941
*Godfrey's serial correlation test- null hypothesis says we do not have Serial
correlation. We either reject or fail to reject. This one says at time t or time t-1.
Based on the p-values we fail to reject the null so we don't have first order serial
correlation. I specified a model such that when I check for serial correlation I won't
have it. Look at the Pr>LM for the Godfrey's serial correlation test, are they correlated
with the last time period? No. 2 time periods ago? No. you purged it from your model,
that's a good thing. Failed to reject the null going back 5 time periods.
*your errors right now are not correlated with the errors in the previous time periods
*they are not correlated over time, that's what we want in a time series model though. We
want to purge that as we did here
*HW 3, estimate the final functional form and verify that it has good functional
properties. Real estate data, part b, given homes and you need to get dhomes lagged 1
period, then lagged 5 periods, then different interest rate lag 1 period and different
interest rate lag 3 periods. This is the final functional form. Did you set up the
differences correctly? The lags correctly up to and including the lags. Do the /godfrey=5
in SAS as below in this code. If in the blue ban you don't have any singular
collinearity. You've got all your output, just interpret it.
*du is a function of du1 the lag term period
*try to make the SBC and AIC as small as possible

ods graphics on;
proc autoreg data=work.okundata plots(only)=(acf);
model du = du1 g g1 / godfrey=5;
output out=yhatout p=yhat;
run;
ods graphics off;
```

The SAS System 17:07 Tuesday, January 22, 2019 6

The AUTOREG Procedure

Dependent Variable du

Ordinary Least Squares Estimates

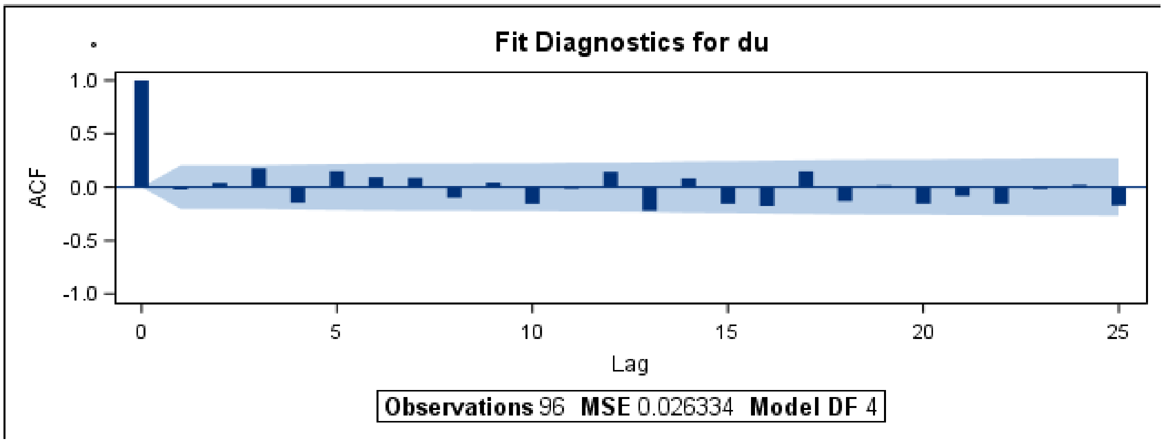
SSE	2.42272399	DFE	92
MSE	0.02633	Root MSE	0.16228
SBC	-62.534153	**AIC**	-72.791546
MAE	0.12863529	AICC	-72.351985
MAPE	69.340139	HQC	-68.64534
Durbin-Watson	1.9861	Total R-Square	**0.6941**

Godfrey's Serial Correlation Test

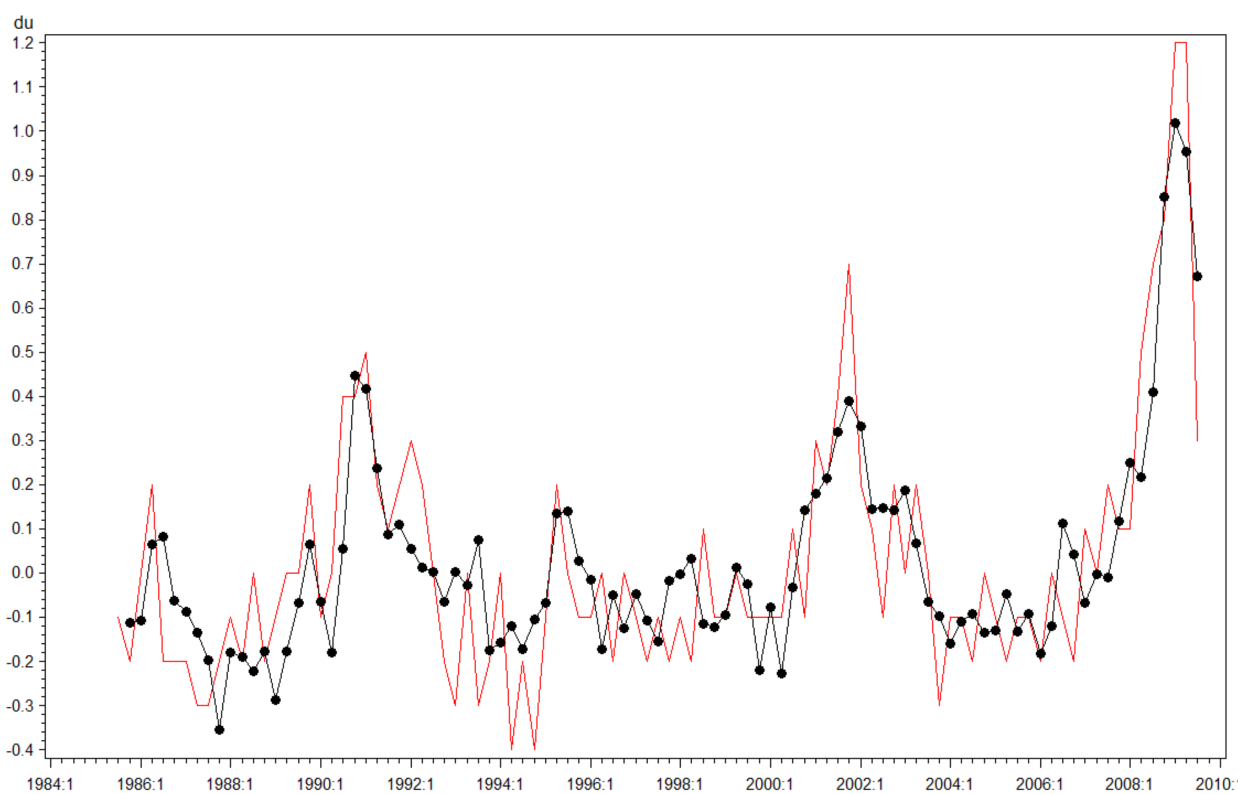
Alternative	LM	Pr > LM
AR(1)	0.1697	**0.6804**
AR(2)	0.2713	**0.8731**
AR(3)	3.8964	**0.2729**
AR(4)	6.1406	**0.1889**
AR(5)	8.2256	**0.1442**

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.3780	0.0578	6.54	<.0001
du1	1	0.3501	0.0846	4.14	<.0001
g	1	-0.1841	0.0307	-6.00	<.0001
g1	1	**-0.0992**	0.0368	-2.69	**0.0084**




```
* plot data;
*actual vs. predicted is plotted.
*the black line is predicted

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



```
* utilize exponential smoothing for a forecast;
*use the expo method for exponential smoothing. The higher the weight the more the
forecast will look like an echo of the actual value

proc forecast data=work.okundata interval=qtr method=expo weight=0.8 trend=1 lead=1
out=du_pred outest=est outfull;
var du;
id date;
run;
quit;

* plot data;
*actual=blue, forecast is the red.
*exponential smoothing says what happened in the past will be forecasted in the future.
*looks like something happens and then it is echoed by the forecast
*the higher the weight, the closer it will echo. The lower you give weight, it will look
like its kind of smooth through there

proc sgplot data=work.du_pred;
series x=date y=du / group=_type_;
yaxis values=(-0.50 to 1.25 by 0.25);
xaxis values=('1apr85'd to '1jan04'd by qtr);
run;
quit;
```