```
* MSBA 635 - Data Analytics II;
```

*Vector Autoregressive (VAR) models are heavily used in forecasting where lagged values of the dependent variables (AR or autoregressive terms) from two or more stationary time series constitute the right hand side of each regression model

*Use the Lag values (AR term or regressive term) it's a regression. Lag value on the right hand side

*utilize VAR on time series to help make better forecasts

*Autoregressive Conditional Heteroscedasticity (ARCH) models are heavily used in forecasting where the time series variance of the errors is not constant due to periods of volatility in the underlying time series

*ARCH this is for time-series data. Like energy commodities

* Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models are ARCH models that also capture the effects of lagged, non-constant time series variance of the errors

*GARCH used to predict future volatility

```
* print data;
*This is the VAR example
*this is from the federal reserve banking system
proc print data=tmp1.fred (obs=10);
run;
```

```
The SAS System         17:24 Tuesday, January 22, 2019   1
```

Obs	c	y
1	7.47902	7.57840
2	7.49159	7.58381
3	7.48762	7.58467
4	7.48897	7.58406
5	7.48869	7.59312
6	7.50346	7.60808
7	7.50829	7.62164
8	7.52812	7.64108
9	7.53871	7.65050
10	7.55092	7.66162

```
* display data attributes;

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

```
The SAS System         17:24 Tuesday, January 22, 2019   2
```

The CONTENTS Procedure

Data Set Name	TMP1.FRED	Observations	200
Member Type	DATA	Variables	2
Engine	V9	Indexes	0
Created	01/24/2011 15:05:43	Observation Length	16
Last Modified	01/24/2011 15:05:43	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	2
First Data Page	1
Max Obs per Page	252
Obs in First Data Page	174
Number of Data Set Repairs	0
Filename	C:\Users\nxnguy01\Desktop\fred.sas7bdat

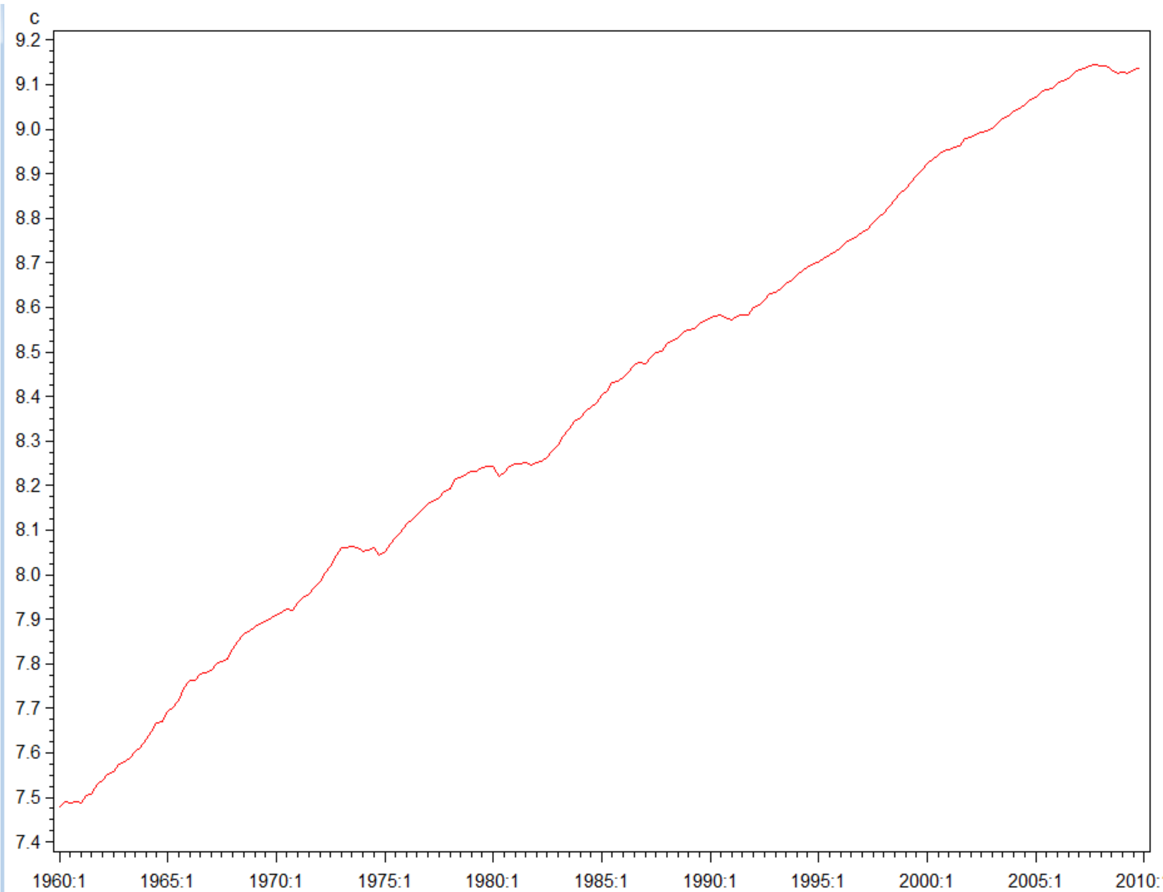
```
Release Created            9.0202M3
Host Created               W32_VSPRO


            Alphabetic List of Variables and Attributes

#    Variable    Type    Len    Label

1    c           Num       8    log of real consumption expenditure
2    y           Num       8    log of real disposable income
```

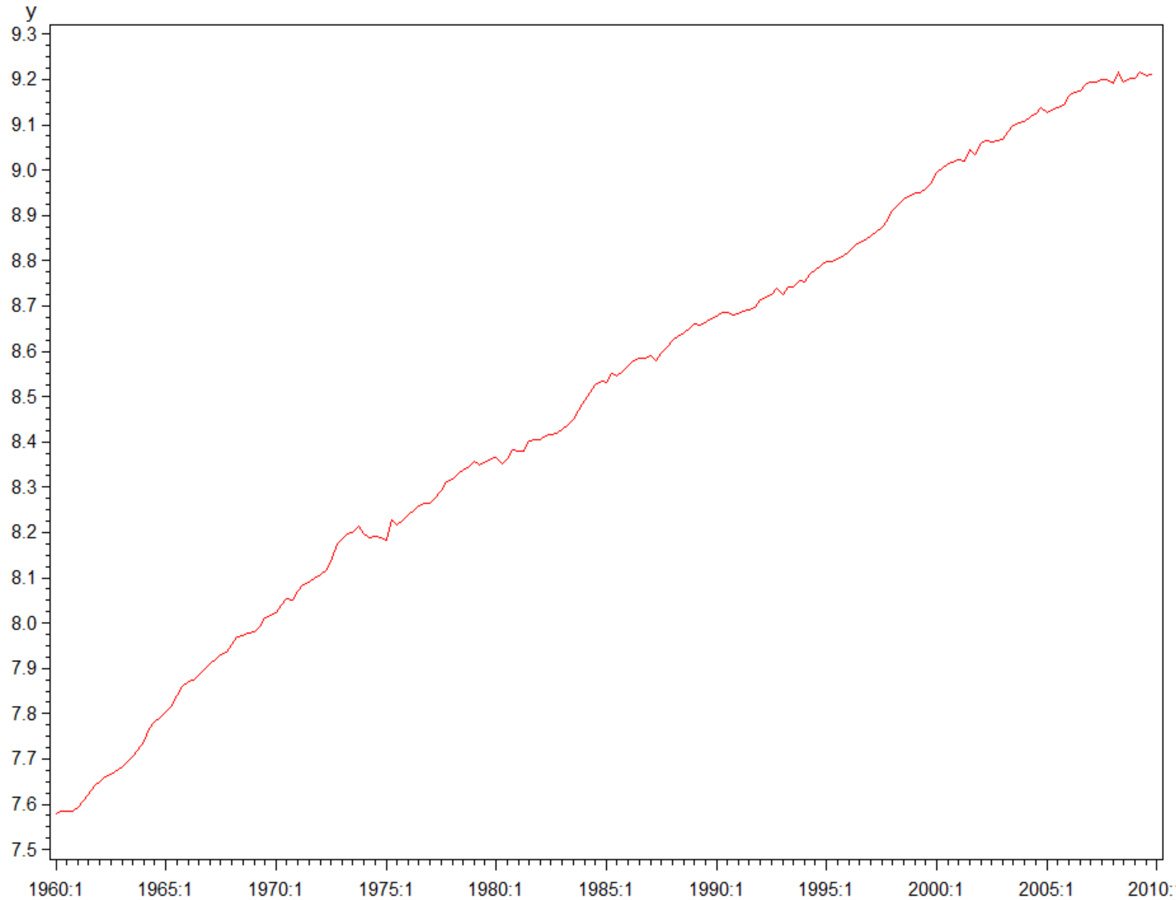
```
* construct new variables;

data freddata;
set tmp1.fred;
dc = dif(c);
dc1 = lag(dc);
dy = dif(y);
dy1 = lag(dy);
retain date '1oct59'd;
date = intnx('qtr',date,1);
format date yyqc.;
year = 1960 + int((_n_-1)/4);
qtr = mod((_n_-1),4) + 1;
run;

* print data;
*first two columns are from the original data set, we created the rest of the columns

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

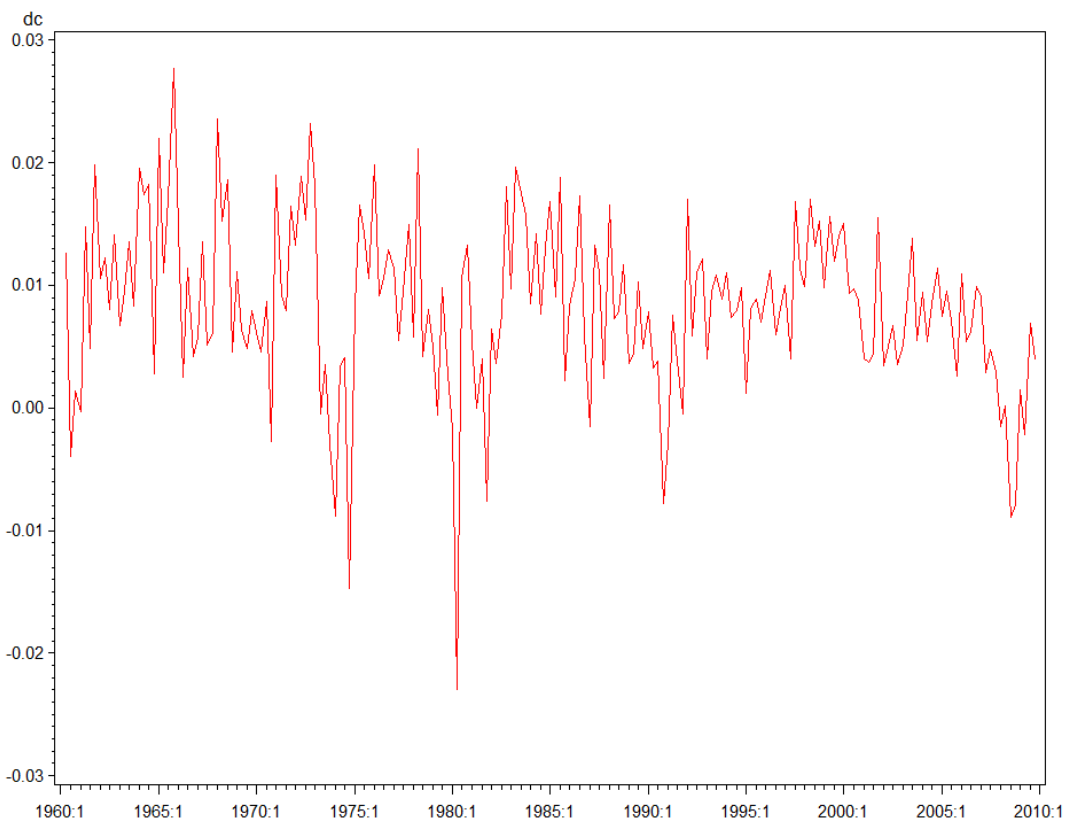
The SAS System 17:24 Tuesday, January 22, 2019 3

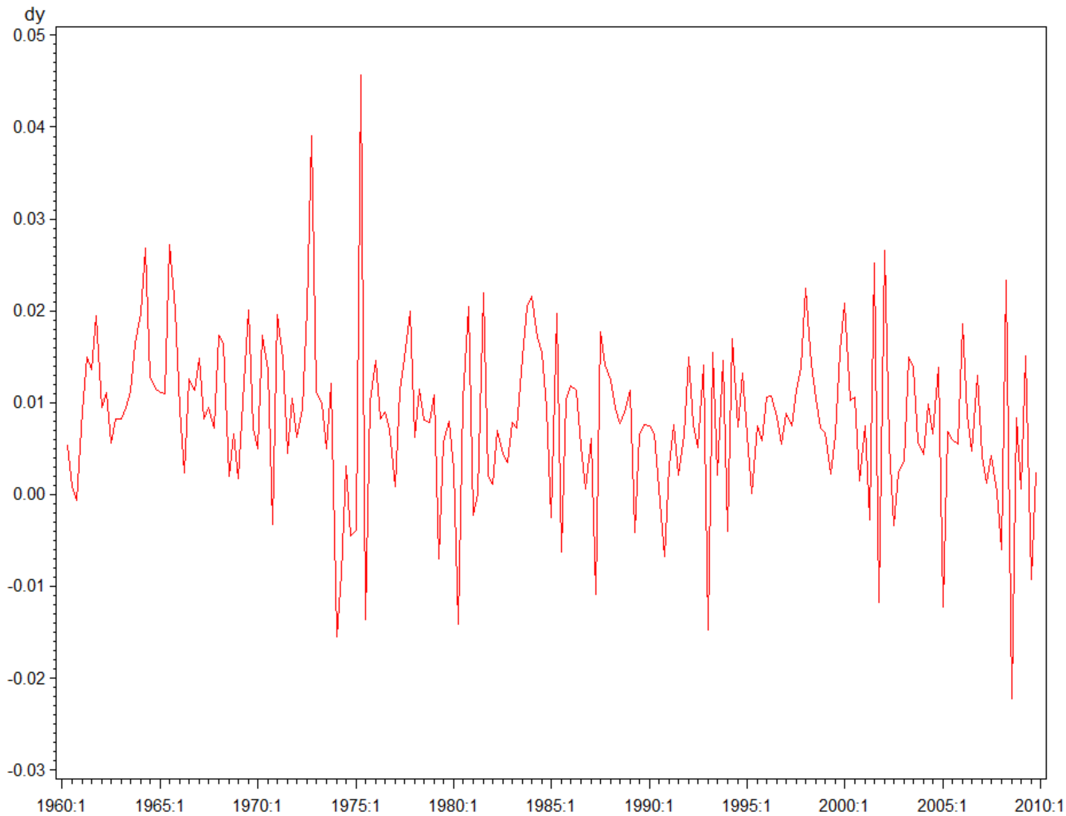
Obs	c	y	dc	dc1	dy	dy1	date	year	qtr
1	7.47902	7.57840	.	.	.	.	1960:1	1960	1
2	7.49159	7.58381	0.012573	.	0.005406	.	1960:2	1960	2
3	7.48762	7.58467	-0.003968	0.012573	0.000864	0.005406	1960:3	1960	3
4	7.48897	7.58406	0.001343	-0.003968	-0.000610	0.000864	1960:4	1960	4
5	7.48869	7.59312	-0.000280	0.001343	0.009061	-0.000610	1961:1	1961	1
6	7.50346	7.60808	0.014770	-0.000280	0.014955	0.009061	1961:2	1961	2
7	7.50829	7.62164	0.004839	0.014770	0.013559	0.014955	1961:3	1961	3
8	7.52812	7.64108	0.019823	0.004839	0.019448	0.013559	1961:4	1961	4
9	7.53871	7.65050	0.010591	0.019823	0.009418	0.019448	1962:1	1962	1
10	7.55092	7.66162	0.012216	0.010591	0.011119	0.009418	1962:2	1962	2

```
* plot data;
*vertical=consumption expenditures
*this is nonstationary
*second one is nonstationary
*third one is stationary
*4th one is stationary
options nolabel;
symbol1 value=none interpol=join color=red;
proc gplot data=work.freddata;
plot c*date;
plot y*date;
plot dc*date;
plot dy*date;
run;
quit;
```




y
9.3
9.2
9.1
9.0
8.9
8.8
8.7
8.6
8.5
8.4
8.3
8.2
8.1
8.0
7.9
7.8
7.7
7.6
7.5
1960:1
1965:1
1970:1
1975:1
1980:1
1985:1
1990:1
1995:1
2000:1
2005:1
2010:1


dc
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
1960:1
1965:1
1970:1
1975:1
1980:1
1985:1
1990:1
1995:1
2000:1
2005:1
2010:1


dy
0.05
0.04
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
1960:1
1965:1
1970:1
1975:1
1980:1
1985:1
1990:1
1995:1
2000:1
2005:1
2010:1

```
* estimate model using proc autoreg;
*my null hypothesis is nonstationary
*you are failing to reject the null meaning that this underlying data series is
nonstationary
*look at sero mean, single mean and trend. For Pr  Rho
options nolabel;
proc autoreg data=work.freddata;
model c = / stationarity=(phillips);
run;
quit;
```

The SAS System 17:24 Tuesday, January 22, 2019 4

The AUTOREG Procedure

Dependent Variable c

Ordinary Least Squares Estimates

SSE	47.8777699	DFE	199
MSE	0.24059	Root MSE	0.49050
SBC	286.940518	AIC	283.6422
MAE	0.42019737	AICC	283.662402
MAPE	5.04974767	HQC	284.976979
Durbin-Watson	0.0005	Total R-Square	0.0000

Phillips-Perron Unit Root Test

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	2	0.1952	0.7283	12.8335	1.0000
Single Mean	2	-0.5621	0.9220	-2.2815	0.1788
Trend	2	-4.5255	0.8514	-1.2864	0.8881

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	8.3836	0.0347	241.72	<.0001

```
* estimate model using proc autoreg;
*this is where you first difference it
*reject the null, conclude the first difference is stationary
*means its an integrated series of order 1 meaning you can run VAR models on it

options nolabel;
proc autoreg data=work.freddata;
model dc = / stationarity=(phillips);
run;
quit;
```


The AUTOREG Procedure

Dependent Variable dc

Ordinary Least Squares Estimates

SSE	0.00960413	DFE	198
MSE	0.0000485	Root MSE	0.00696
SBC	-1407.8036	AIC	-1411.0969
MAE	0.00519821	AICC	-1411.0766
MAPE	148.815172	HQC	-1409.764
Durbin-Watson	1.3964	Total R-Square	0.0000

Phillips-Perron Unit Root Test

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	2	-43.0680	<.0001	-5.1929	<.0001
Single Mean	2	-141.1549	0.0014	-10.3139	<.0001
Trend	2	-146.6295	0.0006	-10.5800	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.008330	0.000494	16.87	<.0001

```
* estimate model using proc autoreg;
*done on the graph of y above
*at alpha .05 the value of 0.5 you fail to reject. Reject at the 0.10 level
options nolabel;
proc autoreg data=work.freddata;
model y = / stationarity=(phillips);
run;
quit;
```


The AUTOREG Procedure

Dependent Variable y

Ordinary Least Squares Estimates

SSE	44.5980298	DFE	199
MSE	0.22411	Root MSE	0.47340
SBC	272.748194	AIC	269.449876
MAE	0.40295459	AICC	269.470078
MAPE	4.7911218	HQC	270.784655
Durbin-Watson	0.0007	Total R-Square	0.0000

Phillips-Perron Unit Root Test

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	2	0.1896	0.7269	12.2387	1.0000
Single Mean	2	-0.7268	0.9091	-2.8190	0.0578
Trend	2	-4.7492	0.8359	-1.6531	0.7685

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	8.4883	0.0335	253.57	<.0001

```
* estimate model using proc autoreg;
*for the first difference of the series graph
*reject the null and reject the null of non stationarity Pr < Rho which are the p-values.
This is how we test for non stationarity
The smaller the AIC the better, you are on a real number line. Since there is no model of
c or dc or y or of dy. the proc auto reg is to do the Phillips-Perron test
* Look at the metrics. Difference in homes at various times. T-1, t-5, T-1, t-3. Estimate
it by setting up differences, lags of those difference, then proc auto reg, LGM test, ACF
plot
options nolabel;
proc autoreg data=work.freddata;
model dy = / stationarity=(phillips);
run;
quit;
```

The SAS System 17:24 Tuesday, January 22, 2019 7

The AUTOREG Procedure

Dependent Variable dy

Ordinary Least Squares Estimates

SSE	0.01610075	DFE	198
MSE	0.0000813	Root MSE	0.00902
SBC	-1304.9858	AIC	-1308.2791
MAE	0.00644854	AICC	-1308.2588
MAPE	212.02648	HQC	-1306.9462
Durbin-Watson	2.1091	Total R-Square	0.0000

Phillips-Perron Unit Root Test

Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau
Zero Mean	2	-110.2268	<.0001	-8.8426	<.0001
Single Mean	2	-217.8586	0.0014	-14.7771	<.0001
Trend	2	-222.2950	0.0006	-15.2918	<.0001

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.008205	0.000639	12.84	<.0001

```
* estimate model using proc autoreg;
*to build a model we use dc and dy
Change in c= lag in change in c and lag in change in y
*gives you first equation from VAR
options nolabel;
proc autoreg data=work.freddata;
model dc = dc1 dy1;
run;
quit;
```


The AUTOREG Procedure

Dependent Variable dc

Ordinary Least Squares Estimates

SSE	0.00843105	DFE	195
MSE	0.0000432	Root MSE	0.00658
SBC	-1414.9276	AIC	-1424.7924
MAE	0.0049531	AICC	-1424.6687
MAPE	124.332415	HQC	-1420.7995
Durbin-Watson	2.0857	Total R-Square	0.1205

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.005278	0.000757	6.97	<.0001
dc1	1	0.2156	0.0747	2.88	0.0044
dy1	1	0.1494	0.0577	2.59	0.0104

```
* estimate model using proc autoreg;
*then run this for lags of dc and dy
*gives you second equation from VAR
options nolabel;
proc autoreg data=work.freddata;
model dy = dc1 dy1;
run;
quit;
```


The AUTOREG Procedure

Dependent Variable dy

Ordinary Least Squares Estimates

SSE	0.01429345	DFE	195
MSE	0.0000733	Root MSE	0.00856
SBC	-1310.4072	AIC	-1320.272
MAE	0.0063767	AICC	-1320.1483
MAPE	191.931556	HQC	-1316.2791
Durbin-Watson	1.9935	Total R-Square	0.1118

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	0.006037	0.000986	6.12	<.0001
dc1	1	0.4754	0.0973	4.88	<.0001
dy1	1	-0.2172	0.0752	-2.89	0.0043

```
* estimate model using proc varmax;

ods graphics on;
proc varmax data=work.freddata plot=impulse;
model dc dy / p=1;
output lead=6;
run;
ods graphics off;
```


The VARMAX Procedure

Number of Observations	199
Number of Pairwise Missing	1

Simple Summary Statistics

Variable	Type	N	Mean	Standard Deviation	Min	Max
dc	Dependent	199	0.00833	0.00696	-0.02296	0.02773
dy	Dependent	199	0.00820	0.00902	-0.02218	0.045

*two equations estimated here, you'll see your same coefficient and statistical significance here
*see the graphs that were printed out on to the desktop
*what's a shock to dc and what filters through the dc equation
*second graph is shock in dy and trace its impact on dc

The VARMAX Procedure

Type of Model	VAR(1)
Estimation Method	Least Squares Estimation

Model Parameter Estimates

Equation	Parameter	Estimate	Standard Error	t Value	Pr > |t|	Variable
dc	CONST1	0.00528	0.00076	6.97	0.0001	1
	AR1_1_1	0.21561	0.07475	2.88	0.0044	dc(t-1)
	AR1_1_2	0.14938	0.05773	2.59	0.0104	dy(t-1)
dy	CONST2	0.00604	0.00099	6.12	0.0001	1
	AR1_2_1	0.47543	0.09733	4.88	0.0001	dc(t-1)
	AR1_2_2	-0.21717	0.07517	-2.89	0.0043	dy(t-1)

Covariances of Innovations

Variable	dc	dy
dc	0.00004	0.00003
dy	0.00003	0.00007

Log-likelihood 1764.344

Information Criteria

AICC	-3509.73
HQC	-3498.71
AIC	-3510.69
SBC	-3481.09
FPEC	2.617E-9

Forecasts

Variable	Obs	Forecast	Standard Error	95% Confidence Limits	
dc	200	0.00650	0.00658	-0.00639	0.01938
	201	0.00778	0.00696	-0.00586	0.02143
	202	0.00808	0.00701	-0.00565	0.02181
	203	0.00823	0.00701	-0.00551	0.02197
	204	0.00826	0.00701	-0.00548	0.02201
	205	0.00828	0.00701	-0.00546	0.02203
dy	200	0.00741	0.00856	-0.00938	0.02419
	201	0.00752	0.00902	-0.01016	0.02519
	202	0.00811	0.00908	-0.00968	0.02589

The VARMAX Procedure

Forecasts

Variable	Obs	Forecast	Standard Error	95% Confidence Limits	
dy	203	0.00812	0.00908	-0.00968	0.02592
	204	0.00819	0.00908	-0.00962	0.02599
	205	0.00819	0.00908	-0.00961	0.02599