```
* Dr. Steven S. Vickner;
* MSBA 635 - Data Analytics II;

* print data;
```
*Seemingly unrelated regressions (SUR) models also utilize panel data, but a SUR model has a separate equation for each cross-sectional unit
*SUR models often exhibit contemporaneous correlation of the errors across equations
*The use of cross equation parameter hypothesis testing makes the SUR approach appealing for certain applications
*Explain variability in gross investment in plant and equipment as a function of capital stock and equity market value for two firms over 20 years
*data is horizontally concatenated
*multivariate regression – means you are trying to estimate more than one equation
*SUR will NOT BE ON the FINAL
*a betting pool will be an equation in the SUR
*more than one right hand side explanatory variable, now you have more than one equation
*you do this because these equations have correlation across the errors, it gives you more efficient estimates out of your regression
*F-test for parameters across different equations, gets to deeper insights into the data
*data is stacked on top of each other. Firm 2 is below firm 1.

```
proc print data=tmp1.grunfeld2 (obs=10);
run;
```
```
The SAS System         17:29 Tuesday, January 29, 2019   1

                        Obs     inv        v        k      firm    year

                          1     33.1    1170.6     97.8      1     1935
                          2     45.0    2015.8    104.4      1     1936
                          3     77.2    2803.3    118.0      1     1937
                          4     44.6    2039.7    156.2      1     1938
                          5     48.1    2256.2    172.6      1     1939
                          6     74.4    2132.2    186.6      1     1940
                          7    113.0    1834.1    220.9      1     1941
                          8     91.9    1588.0    287.8      1     1942
                          9     61.3    1749.4    319.9      1     1943
                         10     56.8    1687.2    321.3      1     1944
```

```
* display data attributes;
*
proc contents data=tmp1.grunfeld2;
run;
```
```
The SAS System         17:29 Tuesday, January 29, 2019   2

                                The CONTENTS Procedure

            Data Set Name        TMP1.GRUNFELD2              Observations          40
            Member Type          DATA                        Variables             5
            Engine               V9                          Indexes               0
            Created              11/13/2010 12:13:31         Observation Length    40
            Last Modified        11/13/2010 12:13:31         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            101
  Obs in First Data Page      40
  Number of Data Set Repairs  0
  Filename                    C:\Users\nxnguy01\Desktop\grunfeld2.sas7bdat
  Release Created             9.0202M3
  Host Created                W32_VSPRO


                        Alphabetic List of Variables and Attributes

     #    Variable    Type    Len    Label

     4    firm        Num       8    firm id
     1    inv         Num       8    gross investment in plant and equipment, millions of $1947
     3    k           Num       8    stock of capital, millions of $1947
     2    v           Num       8    value of common and preferred stock, millions of $1947
     5    year        Num       8    year
```

```
* construct new variables;
* 2 dummy variables with multicolinearlity in the model
*FIX THIS OUTPUT AND CODE
*

data grundata;
set tmp1.grunfeld2;
d = 0;
if firm = 2 then d = 1;
d2 = 0;
if firm = 2 then d2 = 1;
run;
```

```
The SAS System         17:29 Tuesday, January 29, 2019   3

                Obs     inv        v        k      firm    year    d     dv     dk

                  1     33.1    1170.6     97.8      1     1935    0      0      0
                  2     45.0    2015.8    104.4      1     1936    0      0      0
                  3     77.2    2803.3    118.0      1     1937    0      0      0
                  4     44.6    2039.7    156.2      1     1938    0      0      0
                  5     48.1    2256.2    172.6      1     1939    0      0      0
                  6     74.4    2132.2    186.6      1     1940    0      0      0
                  7    113.0    1834.1    220.9      1     1941    0      0      0
                  8     91.9    1588.0    287.8      1     1942    0      0      0
                  9     61.3    1749.4    319.9      1     1943    0      0      0
                 10     56.8    1687.2    321.3      1     1944    0      0      0
```

```
* print data;
```

*FIX THIS OUTPUT

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

The SAS System 17:29 Tuesday, January 29, 2019 4

Obs	inv	v	k	firm	year	d	dv	dk
1	33.1	1170.6	97.8	1	1935	0	0	0
2	45.0	2015.8	104.4	1	1936	0	0	0
3	77.2	2803.3	118.0	1	1937	0	0	0
4	44.6	2039.7	156.2	1	1938	0	0	0
5	48.1	2256.2	172.6	1	1939	0	0	0
6	74.4	2132.2	186.6	1	1940	0	0	0
7	113.0	1834.1	220.9	1	1941	0	0	0
8	91.9	1588.0	287.8	1	1942	0	0	0
9	61.3	1749.4	319.9	1	1943	0	0	0
10	56.8	1687.2	321.3	1	1944	0	0	0

*estimate model using proc reg;
*one of the models will not be estimable because you will have perfect multicolinearity
*the parameter will not be estimable
*you cannot include d1 and d2 because they are your overall intercept, you need to drop one of the dummy variables
*don't do model 3 or model 4 because the first 2 give you the information you want
* there is no control for heterogeneity for the first model output
*this shows how we take this data set with unrelated regression and see that it's a panel data set
*second model there is a shifter for General electric. How is investment different for GE versus the investment case?
*model 3 how is investment for westing house different?
*Control for cross sectional heterogeneity by putting in a sector?
*-15.96560 is your base case (Intecept)
*10-26=-16 (the intercept)
*model 4 is a NO, SAS will give us an error
*it says it can't get a parameter estimate on d2- this is the **dummy variable track**
*model 5 has a note "**R-square is redefined**", the / noint is what did it. You can run model 1 and model 2 to get the results that you want
*ADD THIS OUTPUT

```
options nolabel;
model inv = v k;
model inv = d1 v k;
model inv = d2 v k;
model inv = d1 d2 v k;
model inv = d1 d2 v k / noint;
run;
quit;
```

*THERE IS A CODE HERE THAT NEEDS TO BE ADDED

```
* estimate model using proc reg;
*
options nolabel;
proc reg data=work.grundata;
model inv = d v dv k dk;
test d = dv = dk = 0;
run;
quit;
```


The REG Procedure
Model: MODEL1
Dependent Variable: inv

Number of Observations Read	40
Number of Observations Used	40

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	72079	14416	32.70	<.0001
Error	34	14990	440.87711		
Corrected Total	39	87069			

Root MSE	20.99707	R-Square	0.8278
Dependent Mean	72.59075	Adj R-Sq	0.8025
Coeff Var	28.92527		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-9.95631	23.62636	-0.42	0.6761
d	1	9.44692	28.80535	0.33	0.7450
v	1	0.02655	0.01172	2.27	0.0300
dv	1	0.02634	0.03435	0.77	0.4485
k	1	0.15169	0.01936	7.84	<.0001
dk	1	-0.05929	0.11695	-0.51	0.6155

The REG Procedure
Model: MODEL1

Test 1 Results for Dependent Variable inv

Source	DF	Mean Square	F Value	Pr > F
Numerator	3	524.39389	1.19	0.3284

```
                    Denominator        34      440.87711

* estimate model using proc reg;
*The by statement lets us run things by some sort of classification
*separates the firms that you selected: firm 1 and firm 2

options nolabel;
proc reg data=work.grundata;
by firm;
model inv = v k;
run;
quit;
```


---------- firm=1 ----------

The REG Procedure
Model: MODEL1
Dependent Variable: inv

Number of Observations Read	20
Number of Observations Used	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	31632	15816	20.34	<.0001
Error	17	13217	777.44634		
Corrected Total	19	44849			

Root MSE	27.88272	R-Square	0.7053
Dependent Mean	102.29000	Adj R-Sq	0.6706
Coeff Var	27.25850		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-9.95631	31.37425	-0.32	0.7548
v	1	0.02655	0.01557	1.71	0.1063
k	1	0.15169	0.02570	5.90	<.0001

---------- firm=2 ----------

The REG Procedure
Model: MODEL1
Dependent Variable: inv

Number of Observations Read	20
Number of Observations Used	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5165.55292	2582.77646	24.76	<.0001
Error	17	1773.23393	104.30788		
Corrected Total	19	6938.78685			

Root MSE	10.21312	R-Square	0.7444
Dependent Mean	42.89150	Adj R-Sq	0.7144
Coeff Var	23.81153		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.50939	8.01529	-0.06	0.9501
v	1	0.05289	0.01571	3.37	0.0037
k	1	0.09241	0.05610	1.65	0.1179

```
* construct new variables;
*

data grun1;
set tmp1.grunfeld2;
if firm = 1;
inv_ge = inv;
v_ge = v;
k_ge = k;
keep year inv_ge v_ge k_ge;
run;

* construct new variables;

data grun2;
set tmp1.grunfeld2;
if firm = 2;
inv_we = inv;
v_we = v;
```

```
k_we = k;
keep year inv_we v_we k_we;
run;

* merge data;
*you concatenate the data once you have constructed the new variables
*merge them by year

data grunfeld;
merge work.grun1 work.grun2;
by year;
run;

* estimate model using proc reg;
* individual regressions
*one for GE
options nolabel;
proc reg data=work.grunfeld;
model inv_ge = v_ge k_ge;
output out=geout r=ehat_ge;
run;
quit;
```


The REG Procedure
Model: MODEL1
Dependent Variable: inv_ge

Number of Observations Read	20
Number of Observations Used	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	31632	15816	20.34	<.0001
Error	17	13217	777.44634		
Corrected Total	19	44849			

Root MSE	27.88272	R-Square	0.7053
Dependent Mean	102.29000	Adj R-Sq	0.6706
Coeff Var	27.25850		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-9.95631	31.37425	-0.32	0.7548
v_ge	1	0.02655	0.01557	1.71	0.1063
k_ge	1	0.15169	0.02570	5.90	<.0001

```
* estimate model using proc reg;
*regression for westing house

options nolabel;
proc reg data=work.grunfeld;
model inv_we = v_we k_we;
output out=weout r=ehat_we;
run;
quit;
```


The REG Procedure
Model: MODEL1
Dependent Variable: inv_we

Number of Observations Read	20
Number of Observations Used	20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5165.55292	2582.77646	24.76	<.0001
Error	17	1773.23393	104.30788		
Corrected Total	19	6938.78685			

Root MSE	10.21312	R-Square	0.7444
Dependent Mean	42.89150	Adj R-Sq	0.7144
Coeff Var	23.81153		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.50939	8.01529	-0.06	0.9501
v_we	1	0.05289	0.01571	3.37	0.0037
k_we	1	0.09241	0.05610	1.65	0.1179

```
* merge data;
*adds on the residual columns to the end, appends data by year

data all;
merge work.geout work.weout;
run;

* produce correlation matrix;
*the errors out of each equation are highly coordinated (73.896%)
*parameter estimates are more efficient when run as a system vs. individual
equations
proc corr data=work.all;
var ehat_ge ehat_we;
run;
```

```
quit;
```


The CORR Procedure

2 Variables: ehat_ge ehat_we

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
ehat_ge	20	0	26.37442	0	-37.51100	58.73724
ehat_we	20	0	9.66065	0	-13.50522	17.26484

Pearson Correlation Coefficients, N = 20
Prob > |r| under H0: Rho=0

	ehat_ge	ehat_we
ehat_ge	1.00000	0.72896 0.0003
ehat_we	0.72896	1.00000

```
* estimate model using proc syslin;
*equations in systems are typically coorelated via their error term
*you have a label for ge and for westing house
*you get better forecast in models like this then treating them separately

options nolabel;
proc syslin data=work.grunfeld sur;
ge: model inv_ge = v_ge k_ge;
we: model inv_we = v_we k_we;
run;
quit;
```


The SYSLIN Procedure
Ordinary Least Squares Estimation

Model	GE
Dependent Variable	inv_ge

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	31632.03	15816.02	20.34	<.0001
Error	17	13216.59	777.4463		
Corrected Total	19	44848.62			

Root MSE	27.88272	R-Square	0.70531

Dependent Mean	102.29000	Adj R-Sq	0.67064
Coeff Var	27.25850		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-9.95631	31.37425	-0.32	0.7548
v_ge	1	0.026551	0.015566	1.71	0.1063
k_ge	1	0.151694	0.025704	5.90	<.0001

The SYSLIN Procedure
Ordinary Least Squares Estimation

Model	WE
Dependent Variable	inv_we

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5165.553	2582.776	24.76	<.0001
Error	17	1773.234	104.3079		
Corrected Total	19	6938.787			

Root MSE	10.21312	R-Square	0.74445
Dependent Mean	42.89150	Adj R-Sq	0.71438
Coeff Var	23.81153		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.50939	8.015289	-0.06	0.9501
v_we	1	0.052894	0.015707	3.37	0.0037
k_we	1	0.092406	0.056099	1.65	0.1179

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Cross Model Covariance

	GE	WE
GE	777.446	207.587
WE	207.587	104.308

Cross Model Correlation

	GE	WE
GE	1.00000	0.72896
WE	0.72896	1.00000

Cross Model Inverse Correlation

	GE	WE
GE	2.13397	-1.55559
WE	-1.55559	2.13397

Cross Model Inverse Covariance

	GE	WE
GE	0.002745	-.005463
WE	-.005463	0.020458

System Weighted MSE	0.9719
Degrees of freedom	34
System Weighted R-Square	0.6284

Model	GE
Dependent Variable	inv_ge

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-27.7193	29.32122	-0.95	0.3577
v_ge	1	0.038310	0.014415	2.66	0.0166
k_ge	1	0.139036	0.024986	5.56	<.0001

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Model WE
Dependent Variable inv_we

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.25199	7.545217	-0.17	0.8702
v_we	1	0.057630	0.014546	3.96	0.0010
k_we	1	0.063978	0.053041	1.21	0.2443

*these results: look at the standard errors for the equations, these are much smaller

```
* estimate model using proc syslin;
*once we're in a system we can do hypothesis testing
*null hypothesis: they are identical.
*reject the null, they are not the same
*for test 2 fail to reject the null
*this is a system wide test
*2 different parameters in 2 different equations are tested
*reject the null

options nolabel;
proc syslin data=work.grunfeld sur;
ge: model inv_ge = v_ge k_ge;
test_1: test v_ge = k_ge;
we: model inv_we = v_we k_we;
test_2: test v_we=k_we;
stest_1: stest ge.v_ge = we.v_we;
stest_2: stest ge.intercept=we.intercept,
               ge.v_ge = we.v_we,
                     ge.k_ge = we.k_we;
run;
quit;
```

The SYSLIN Procedure
Ordinary Least Squares Estimation

Model GE
Dependent Variable inv_ge

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	31632.03	15816.02	20.34	<.0001
Error	17	13216.59	777.4463		
Corrected Total	19	44848.62			

Root MSE	27.88272	R-Square	0.70531
Dependent Mean	102.29000	Adj R-Sq	0.67064
Coeff Var	27.25850		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-9.95631	31.37425	-0.32	0.7548
v_ge	1	0.026551	0.015566	1.71	0.1063
k_ge	1	0.151694	0.025704	5.90	<.0001

The SYSLIN Procedure
Ordinary Least Squares Estimation

Model WE
Dependent Variable inv_we

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	5165.553	2582.776	24.76	<.0001
Error	17	1773.234	104.3079		
Corrected Total	19	6938.787			

Root MSE	10.21312	R-Square	0.74445
Dependent Mean	42.89150	Adj R-Sq	0.71438
Coeff Var	23.81153		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-0.50939	8.015289	-0.06	0.9501
v_we	1	0.052894	0.015707	3.37	0.0037
k_we	1	0.092406	0.056099	1.65	0.1179

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Cross Model Covariance

	GE	WE
GE	777.446	207.587
WE	207.587	104.308

Cross Model Correlation

	GE	WE
GE	1.00000	0.72896
WE	0.72896	1.00000

Cross Model Inverse Correlation

	GE	WE
GE	2.13397	-1.55559
WE	-1.55559	2.13397

Cross Model Inverse Covariance

	GE	WE
GE	0.002745	-.005463
WE	-.005463	0.020458

System Weighted MSE	0.9719
Degrees of freedom	34
System Weighted R-Square	0.6284

Model	GE
Dependent Variable	inv_ge

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-27.7193	29.32122	-0.95	0.3577
v_ge	1	0.038310	0.014415	2.66	0.0166
k_ge	1	0.139036	0.024986	5.56	<.0001

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Test Results

Num DF	Den DF	F Value	Pr > F	Label
1	34	11.44	0.0018	TEST_1

Model WE
Dependent Variable inv_we

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-1.25199	7.545217	-0.17	0.8702
v_we	1	0.057630	0.014546	3.96	0.0010
k_we	1	0.063978	0.053041	1.21	0.2443

The SYSLIN Procedure
Seemingly Unrelated Regression Estimation

Test Results

Num DF	Den DF	F Value	Pr > F	Label
1	34	0.01	0.9209	TEST_2

Test Results

Num DF	Den DF	F Value	Pr > F	Label

Num DF	Den DF	F Value	Pr > F	Label
1	34	2.80	0.1033	STEST_1

Test Results

Num DF	Den DF	F Value	Pr > F	Label
3	34	3.01	0.0437	STEST_2