MODULE THREE, PART FOUR: PANEL DATA ANALYSIS IN ECONOMIC EDUCATION RESEARCH USING SAS

Part Four of Module Three provides a cookbook-type demonstration of the steps required to use SAS in panel data analysis. Users of this model need to have completed Module One, Parts One and Four, and Module Three, Part One. That is, from Module One users are assumed to know how to get data into SAS, recode and create variables within SAS, and run and interpret regression results. They are also expected to know how to test linear restrictions on sets of coefficients as done in Module One, Parts One and Two. Module Three, Parts Two and Three demonstrate in LIMDEP and STATA what is done here in SAS.

THE CASE

As described in Module Three, Part One, Becker, Greene and Siegfried (2009) examine the extent to which undergraduate degrees (BA and BS) in economics or Ph.D. degrees (PhD) in economics drive faculty size at those U.S. institutions that offer only a bachelor degree and those that offer both bachelor degrees and PhDs. Here we retrace their analysis for the institutions that offer only the bachelor degree. We provide and demonstrate the SAS code necessary to duplicate their results.

DATA FILE

The following panel data are provided in the **comma separated values** (CSV) text file "bachelors.csv", which will automatically open in EXCEL by simply double clicking on it after it has been downloaded to your hard drive. Your EXCEL spreadsheet should look like this:

"College" identifies the bachelor degree-granting institution by a number 1 through 18.

"Year" runs from 1996 through 2006.

"Degrees" is the number of BS or BA degrees awarded in each year by each college.

"DegreBar" is the average number of degrees awarded by each college for the 16-year period.

"Public" equals 1 if the institution is a public college and 2 if it is a private college.

"Faculty" is the number of tenured or tenure-track economics department faculty members.

"Bschol" equals 1 if the college has a business program and 0 if not.

"T" is the time trend running from -7 to 8, corresponding to years from 1996 through 2006.

"MA_Deg" is a three-year moving average of degrees (unknown for the first two years).

College	Year	Degrees	DegreBa	r Public	Faculty	Bschol	Т	MA_Deg
1	1991	50	47.375	2	11	1	-7	0
1	1992	32	47.375	2	8	1	-6	0
1	1993	31	47.375	2	10	1	-5	37.667
1	1994	35	47.375	2	9	1	-4	32.667
\checkmark		\checkmark						
1	2003	57	47.375	2	7	1	5	56
1	2004	57	47.375	2	10	1	6	55.667
1	2005	57	47.375	2	10	1	7	57
1	2006	51	47.375	2	10	1	8	55
2	1991	16	8.125	2	3	1	-7	0
2	1992	14	8.125	2	3	1	-6	0
2	1993	10	8.125	2	3	1	-5	13.333
\checkmark	\checkmark	\checkmark	\downarrow	\checkmark	\checkmark	\checkmark		\checkmark
2	2004	10	8.125	2	3	1	6	12.667
2	2005	7	8.125	2	3	1	7	11.333
2	2006	6	8.125	2	3	1	8	7.667
3	1991	40	35.5	2	8	1	-7	0
3	1992	31	37.125	2	8	1	-6	0
\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\downarrow		\checkmark
17	2004	64	39.3125	2	5	0	6	54.667
17	2005	37	39.3125	2	4	0	7	51.333
17	2006	53	39.3125	2	4	0	8	51.333
18	1991	14	8.4375	2	4	0	-7	0
18	1992	10	8.4375	2	4	0	-6	0
18	1993	10	8.4375	2	4	0	-5	11.333
18	1994	7	8.4375	2	3.5	0	-4	9
\downarrow	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark	\checkmark		\checkmark
18	2005	4	8.4375	2	2.5	0	7	7.333
18	2006	7	8.4375	2	3	0	8	6

If you opened this CSV file in a word processor or text editing program, it would show that each of the 289 lines (including the headers) corresponds to a row in the EXCEL table, but variable values would be separated by commas and not appear neatly one on top of the other as in EXCEL.

As discussed in Module One, Part Two, SAS has a data matrix default restriction. This data set is sufficiently small, so there is no need to adjust the size of the matrix. We could write a "READ" command to bring this text data file into SAS similar to Module 1, Part 4, but like EXCEL, it can be imported into SAS directly by using the import wizard.

To import the data into SAS, click on 'File' at the top left corner of your screen in SAS, and then click 'Import Data'.

<u> </u>	SAS																		
File	Edit	View	Tools	Solutions	Window H	lelp													
Ľ.	<u>N</u> ew				Ctrl+N	102	H	6	D	X	e <mark>ð</mark>	K)	1	顒	<u> </u>	×	3	۲	
1	<u>O</u> pen				Ctrl+O												-		
	<u>C</u> lose						og - (l	Jntitle	d)										
	Open ()bject																	
₩	N <u>e</u> w Pr	rogram																	
2	Open P	rogram																	
	<u>S</u> ave				Ctrl+S	1													
	S <u>a</u> ve As	s																	
	Save As	s O <u>b</u> ject																	
	Import	Data																	
	Export	Data				1													
						1													
	Page Si	et <u>u</u> p																	
В	Print D	etup																	
u E	Print Pr	leview			Ctrl+ D	F													
	<u>r</u>				Cultr	Edito	r Un	titlod1											
	Sen <u>d</u> M	1ail					r - Un	uueu				 							
	1 C:\U	sers\\l	Indicator	rs\indicator.s	as														
	2 C:\U	sers\\	Code\inj	put_code.sas															
	<u>3</u> C:\U	sers\\	Code\an	alysis_code.	as														
	<u>4</u> C:\	\Contst	ructing /	All Northland	ł														
	E <u>x</u> it																		
						[

This will initialize the Import Wizard pop-up screen. Since the data is comma separated values, scroll down under the 'Select data source below.' tab and click on 'Comma Separated Values (*.csv)' as shown below.



Click 'Next', and then provide the location from which the file bachelor.cvs can be located wherever it is stored (in our case in "e:\bachelor.csv").

🖳 Import Wizard - Select file		
SAS Import Wizard Select file	Where is the file located? E:\bachelors.csv Options	Browse
	Help Cancel < Back	Next > Finish

To finish importing the data, click 'Next', and then name the dataset, known as a member in SAS, to be stored in the temporary library called 'WORK'. Recall that a library is simply a folder to store datasets and output. I named the file 'BACHELORS' as seen below. Hitting the Finish button will bring the data set into SAS.

📴 Import Wizard - Select library	and member	_ 0 🔀
SAS Impot/Export Facility SAS Destination	Choose the SAS destination: Library: WORK Member: BACHELORS	•
	Help Cancel < Back Next >	<u>F</u> inish Exit the v

To verify that the wizard imported the data correct, review the Log file and physically inspect the dataset. When SAS is opened, the default panels are the 'Log' window at the top right, the 'Editor' window in the bottom right and the 'Explorer/Results' window on the left. Scrolling through the Log reveals that the dataset was successfully imported. The details of the data step procedure are provided along with a few summary statistics of how many observations and variables were imported.

00	·,		
NOTE :	The infile 'E:\bache File Name=E:\bachelo RECFM=V,LRECL=32767	lors.csv' is: rs.csv,	
NOTE :	288 records were rea The minimum record 1 The maximum record 1	d from the infile 'E:\bachelors.csv ength was 22. ength was 51.	•••
NOTE:	The data set WORK.BA	CHELORS has 288 observations and 9	variables.
NOTE :	DATA statement used	(Total process time):	
	real time	0.14 seconds	
	cpu time	0.09 seconds	
288 r	ows created in WORK.B	ACHELORS	from E:\bachelors.csv.

To view the dataset, click on the "Libraries" folder, which is in the top left of the 'Explorer' panel, and then click on the 'Work' library. This reveals all of the members in the 'Work' library. In this case, the only member is the dataset 'Bachelors'. To view the dataset, click on the dataset icon 'Bachelors'.

10									
JAN VIEW	TABLE Work Ba	chelors							
	COLLEGE	YEAR	DEGREES	DEGREBAR	PUBLIC	FACULTY	BSCHOOL	т	MA DEG
1	1	1991	50	47.375	2	11	1	-7	0
2	1	1992	32	47.375	2	8	1	-6	0
3	1	1993	31	47.375	2	10	1	-5	37.6666667
4	1	1994	35	47.375	2	9	1	-4	32 6666667
5	1	1995	32	47.375	2	11	1	-3	32.6666667
6	1	1996	28	47.375	2	11	1	-2	31.66666657
7	1	1997	49	47.375	2	9.5	1	-1	36 3333333
8	1	1998	60	47.375	2	8	1	0	45.6666667
9	1	1999	49	47.375	2	9	1	1	52.6666667
10	1	2000	59	47.375	2	9	1	2	56
11	1	2001	58	47.375	2	10	1	3	55 3333333
12	1	2002	53	47.375	2	10	1	4	56.6666667
13	1	2003	57	47.375	2	7	1	5	56
14	1	2004	57	47.375	2	10	1	6	55.6666667
15	1	2005	57	47.375	2	10	1	7	57
16	1	2006	51	47.375	2	10	1	8	55
17	2	1991	16	8.125	2	3	1	-7	0
18	2	1992	14	8.125	2	3	1	-6	0
19	2	1993	10	8.125	2	3	1	-5	13.3333333
20	2	1994	5	8.125	2	3	1	-4	9.6666667
21	2	1995	7	8.125	2	3	1	-3	7.3333333
22	2	1996	3	8.125	2	3	1	-2	5
23	2	1997	3	8.125	2	3	1	-1	4.3333333
24	2	1998	5	8.125	2	3	1	0	3.6666667
25	2	1999	5	8.125	2	3	1	1	4.3333333
- 26	2	2000	5	8.125	2	3	1	2	5
27	2	2001	6	8.125	2	3	1	3	5.3333333
20	2	2002	11	8.125	2	3	1	4	7.3333333
29	2	2003	17	8.125	2	3	1	5	11.3333333
30	2	2004	10	8.125	2	3	1	6	12.6666667
31	2	2005	7	8.125	2	3	1	7	11.3333333
32	2	2006	6	8.125	2	3	1	8	7.6666667
33	3	1991	40	37.125	2	8	1	-7	0
34	3	1992	31	37.125	2	8	1	-6	0
35	3	1993	33	37.125	2	7	1	-0	34.6666667
	3	1994	24	37.125	2	9	1	4	29.3333333
	3	1995	27	37.125	2	7	1	-3	28
34 34 35 36 37 4	3333	1991 1992 1993 1994 1995	40 31 33 24 27	37.125 37.125 37.125 37.125 37.125 37.125	2 2 2 2 2 2 2	8 7 9 7	1	-7 -6 -5 -4 -3	34.666666 29.333333 2

In addition to a visual inspection of the data, we use the "means" command to check the descriptive statistics. Since we don't list any variables in the command, by default, SAS runs the 'means' command on all variables in the dataset. First, however, we need to remove the two years (1991 and 1992) for which no data are available for the degree moving average measure. Since we may need the full dataset later, it is good practice to delete the observations off of a copy of the dataset (called bachelors2). This is done in a data step using an 'if then' command.

```
data bachelors2;
    set bachelors;
if year = 1991 then delete;
if year = 1992 then delete;
run;
PROC MEANS DATA=bachelors2;
RUN;
```

Typing the following commands into the 'Editor' window and then clicking the run bottom (recall this is the running man at the top) yields the following screen.

		The	MEANS Procedure			
Variable	N	Mean	Std Dev	Minimum	Max i mum	
COLLEGE	252	9.5000000	5.1984521	1.0000000	18.000000	
Year	252	1999.50	4.0391510	1993.00	2006.00	
DEGREES	252	23.1111111	19.2263606	0	81.0000000	
Degrebar	252	23.6527778	18.0142715	2.0000000	62.4375000	
PUBL IC	252	1.7777778	0.4165671	1.0000000	2.0000000	
FACULTY	252	6.5178571	3.1367692	2.0000000	14.0000000	
BSCHOOL	252	0.3888889	0.4884682	0	1.0000000	
т	252	1.5000000	4.0391510	-5.0000000	8.000000	
Ma_deg	252	23.1931217	18.5539832	1.3333333	80.0000000	

CONSTANT COEFFICIENT REGRESSION

The constant coefficient panel data model for the faculty size data-generating process for bachelor degree-granting undergraduate departments is given by

Faculty size_{it} = $\beta_1 + \beta_2 T_t + \beta_3 BA \& S_{it} + \beta_4 MEANBA \& S_i + \beta_5 PUBLIC_i + \beta_6 Bschl + \beta_7 MA_Deg_{it} + \varepsilon_{it}$

where the error term ε_{it} is independent and identically distributed (*iid*) across institutions and over time and $E(\varepsilon_{it}^{2}|\mathbf{x}_{it}) = \sigma^{2}$, for I = 18 colleges and T = 14 years (-5 through 8) for 252 complete records. To take into account clustering, include the cluster option with the cluster being on the colleges. The SAS OLS regression command that needs to be entered into the editor, including the standard error adjustment for clustering is

```
proc surveyreg data=bachelors2;
    cluster college;
    model faculty = t degrees degrebar public bschool ma_deg;
run;
```

Upon highlighting and hitting the "run" button, the Output panel shows the following results

Regression Analysis	s for De	ependent V	ariable FACUL	TY.
	Data Su	Immary		
Number of Ol Mean of FACI Sum of FACU	bservati ULTY LTY	ions	252 6.51786 1642.5	
	Design 8	Gummary		
Number of	Cluster	rs	18	
	Fit Stat	tistics		
R-squa Root M Denomin	re BE nator DF	0.6 1.8	6484 6827 17	
Test	s of Mod	del Effect	s	
Effect I	Num DF	F Value	Pr → F	
Model Intercept T DEGREES DEGREBAR PUBLIC BSCHOOL MA_DEG	6 1 1 1 1 1	27.70 123.99 1.59 0.77 10.28 46.00 0.38 4.36	<pre><.0001 <.0001 0.2242 0.3928 0.0052 <.0001 <.0001 0.5457 0.0521</pre>	
The denominator deg	rees of	freedom f	or the F test	s is 17.

Estimated Regression Coefficients

Parameter	Estimate	Standard Error	t Value	Pr > t
Intercept	10.1397401	0.91062638	11.13	<.0001
т	-0.0280875	0.02226545	-1.26	0.2242
DEGREES	-0.0163611	0.01865794	-0.88	0.3928
DEGREBAR	0.1083201	0.03378213	3.21	0.0052
PUBL IC	-3.8623935	0.56949614	-6.78	<.0001
BSCHOOL	0.5811154	0.94252689	0.62	0.5457
MA_DEG	0.0378038	0.01809658	2.09	0.0521

NOTE: The denominator degrees of freedom for the t tests is 17.

Contemporaneous degrees have little to do with current faculty size but both overall number of degrees awarded (the school means) and the moving average of degrees (MA_DEG) have significant effects. It takes an increase of 26 or 27 bachelor degrees in the moving average to expect just one more faculty position. Whether it is a public or a private college is highly significant. Moving from a public to a private college lowers predicted faculty size by nearly four members for otherwise comparable institutions. There is an insignificant erosion of tenured and tenure-track faculty size over time. Finally, while economics departments in colleges with a business school tend to have a larger permanent faculty, ceteris paribus, the effect is small and insignificant.

NOTE:

FIXED-EFFECTS REGRESSION

The fixed-effects model requires either the insertion of 17(0,1) covariates to capture the unique effect of each of the 18 colleges (where each of the 17 dummy coefficients are measured relative to the constant term) or the insertion of 18 dummy variables with no constant term in the OLS regression. In addition, no time invariant variables can be included because they would be perfectly correlated with the respective college dummies. Thus, the overall mean number of degrees, the public or private dummy, and business school dummy cannot be included as regressors.

The SAS code to be run from the editor window, including the commands to create the dummy variables is:

```
data bachelors2;
         set bachelors2;
  col1 = 0; col2 = 0; col3 = 0; col4 = 0; col5=0;
                                                                               col6=0;
 col7 = 0; col8 = 0; col9 = 0; col10 = 0; col11 = 0; col12 = 0;
col13 = 0; col14 =0; col15 = 0; col16 = 0; col17 = 0; col18 = 0;
                                                 if college = 2 then col2=1;
if college = 4 then col4=1;
if college = 1 then coll=1;
if college = 3 then col3=1;
if college = 3 then col3=1;
if college = 5 then col5=1;
if college = 7 then col7=1;
if college = 7 then col7=1;
if college = 9 then col9=1;
if college = 11 then col11=1;
if college = 13 then col13=1;
if college = 15 then col15=1;
if college = 17 then col17=1;
if college = 18 then col18=1;
run;
proc surveyreg data=bachelors2;
         cluster college;
         model faculty = t degrees ma_deg col1 col2 col3 col4 col5
                                            col6 col7 col8 col9 col10 col11 col12
                                            coll3 coll4 coll5 coll6 coll7;
```

quit;

The resulting regression information appearing in the output window is

The SURVEYREG Procedure

Regression Analysis for Dependent Variable FACULTY

Data Summary

Number of Observations	252
Mean of FACULTY	6.51786
Sum of FACULTY	1642.5

Design Summary

Number of Clusters 18

Fit Statistics

R-square	0.9406
Root MSE	0.7967
Denominator DF	17

Estimated Regression Coefficients

		Standard		
Parameter	Estimate	Error	t Value	$\Pr > t $
Intercept	2.6963636	0.15108692	17.85	<.0001
T	-0.0285342	0.02245298	-1.27	0.2209
DEGREES	-0.0160847	0.01520712	-1.06	0.3050
Ma_deg	0.0398470	0.01485281	2.68	0.0157
coll	5.7774674	0.76815649	7.52	<.0001
col2	0.1529889	0.01342928	11.39	<.0001
col3	4.2975911	0.55419559	7.75	<.0001
col4	6.2897280	0.65533467	9.60	<.0001
co15	4.9109414	0.56987008	8.62	<.0001
co16	5.0201570	0.02560770	196.04	<.0001
col7	1.2138416	0.01321172	91.88	<.0001
co18	0.7779701	0.06784745	11.47	<.0001
co19	3.1647365	0.06269579	50.48	<.0001
coll0	2.8634525	0.15539858	18.43	<.0001
coll1	5.1518149	0.02403066	214.39	<.0001
coll2	-0.0680152	0.02152566	-3.16	0.0057
co113	3.9889465	1.01414776	3.93	0.0011
coll4	-0.6319560	0.11986346	-5.27	<.0001
co115	8.2585866	0.47255240	17.48	<.0001
col16	8.0096959	0.55460921	14.44	<.0001
col17	0.4354377	0.59258369	0.73	0.4725

Once again, contemporaneous degrees is not a driving force in faculty size. An F test is not needed to assess if at least one of the 17 colleges differ from college 18. With the exception of college 17, each of the other colleges are significantly different. The moving average of degrees is again significant.

RANDOM-EFFECTS REGRESSION

Finally, consider the random-effects model in which we employ Mundlak's (1978) approach to estimating panel data. The Mundlak model posits that the fixed effects in the equation, β_{1i} , can be projected upon the group means of the time-varying variables, so that

$$\beta_{1i} = \beta_1 + \delta' \overline{x}_i + w_i$$

where \bar{x}_i is the set of group (school) means of the time-varying variables and w_i is a (now) random effect that is uncorrelated with the variables and disturbances in the model. Logically, adding the means to the equations picks up the correlation between the school effects and the other variables. We could not incorporate the mean number of degrees awarded in the fixedeffects model (because it was time invariant) but this variable plays a critical role in the Mundlak approach to panel data modeling and estimation.

The random effects model for BA and BS degree-granting undergraduate departments is

$$FACULTY \ size_{it} = \beta_1 + \beta_2 YEAR_t + \beta_3 BA \& S_{it} + \beta_4 MEANBA \& S_i + \beta_5 MOVAVBA \& BS + \beta_6 PUBLIC_i + \beta_7 Bschl + \varepsilon_{it} + u_i$$

where error term ε is *iid* over time, $E(\varepsilon_{it}^2 | \mathbf{x}_{it}) = \sigma^2$ for I = 18 and $T_i = 14$ and $E[u_i^2] = \theta^2$ for I = 18.

In SAS 9.1, there are no straightforward procedures to estimate this model. In the appendix, I do provide a lengthy procedure that estimates the random effects model by OLS regression on a transformed model. This is quite complex and is not recommended for beginners. See Cameron and Trivedi (2005) for further details. SAS 9.2 has a new command called the PANEL procedure to estimate panel data. For our model, we need to attach the / RANONE option to specify that a one-way random-effects model be estimated. We also need to correct for the clustering of the data. Unlike simple commands in LIMPDEP and STATA, SAS does not have an option for one-way random effects with clustered errors.

This new SAS 9.2 procedure has more options for specific error term structures in panel data. Although SAS does not allow the CLUSTER option, there is a VCOMP option that specifies the type of variance component estimate to use. For balanced data, the default is VCOMP=FB. However, the FB method does not always obtain nonnegative estimates for the cross section (or group) variance. In the case of a negative estimate, a warning is printed and the estimate is set to zero. Because we have to address clustering, WK option is specified, which is close to groupwise heteroscedastic regression. The SAS code to be run from the Editor panel (with 1991 and 1992 data suppressed) is

```
PROC SORT DATA=bachelors2;
BY college year;
PROC panel DATA=bachelors2;
ID college year;
MODEL faculty = t degrees degrebar public bschool MA_deg /RANONE VCOMP=WK;
RUN;
```

The resulting regression information appearing in the output window is

Output-(Untitled) The PANEL Procedure Hansbeek and Kapteyn Variance Components (RanOne) Dependent Variable: FACULTY Hodel Description Estimation Method Number of Cross Sections RanOne Number of Cross Sections SE 150.6509 MSE 0.5149 R-Square 0.1154 Variance Component Estimates Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Housen Test for Randon Effects Pr > n DF Nalue Pr > n 0 . . Variable DF Estimates Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Housen Test for Randon Effects DF DF Value Pr > n 0 . . 1 10.14197 2.9145 3.48 0.0000 T 1 0.1266 -2.26 0.0241 DEGREEM 1 0.106078 0.03937 2.67 0.0004 DEGREEM 1 0.106078 0.03937 <t< th=""><th></th><th></th><th></th><th></th><th></th><th></th></t<>							
The PANEL Procedure Hansbeek and Kapteyn Variance Components (RanOne)Dependent Variable: FACULTYModel DescriptionEstimation MethodRanOne Mumber of Cross SectionsRanOne 18 Time Series LengthFit StatisticsSSE150.6509DFE245 0.7842NSE0.65149Root MSE0.7842 0.7842Nor ince Component EstimatesVariance Component for Cross Sections8.109092 0.534793Variance Component for Error0.634793Housan Test for Randon EffectsDFNaluePr > n0.VariableDFEstimateStandard ErrorValuePr > 1 tIntercept110.141972.91453.480.0000T-0.028530.0126-2.260.0241DEGREESA1-0.016090.03372.670.0241DEGREENA10.1060780.03372.670.0261DEGREENA10.1060780.03372.670.0260DEGREENA10.1060780.03372.670.0260DEGREENA10.1060780.03372.670.0260DEGREENA10.1060780.03372.670.0260DEGREENA10.1060780.03372.670.0260DEGREENA10.0160780.0337<	🕐 Output - (Untitled)						
Nansbeek and Kapteyn Variance Components (Handne) Dependent Variable: FACULTY Model Description Estimation Method Mumber of Cross Sections 18 Time Series Length 14 Fit Statistics SEE 150.6509 DFE 245 MSE 0.6149 Root MSE 0.7842 Not MSE 0.6149 Root MSE 0.7842 Variance Component Estimates Variance Component for Cross Sections 8.109092 Variance Component for Cross Sections 0.634793 Hausnan Test for Randon Effects DF n Value Pr > n 0 Variable DF Estimates Variable DF Estimates Variable DF Estimate DEGREES 1 DEGREES 1 DEGREES 1 Standard Error t Value Pr > 1 I 10.14197 2.9145 3.48 0.0000 T 0.02853 0.0126 -2.26 0.0241 DEGREES 1 DEGREES 1 -0.01609 DEGREES 1 -0.01609 DEGREES 1 -0.01609 DEGREES 1 -0.0		- 18 Mar 19	The PANEL	Procedure		1.0	
Nependent Variable: FACULTY Hodel Description Estination Method RanOne Number of Cross Sections 18 Time Series Length 14 Fit Statistics SSE 150.6509 DFE 245 MSE 0.6149 Root MSE 0.7842 R-Square 0.1154 Variance Component Estinates Variance Component for Cross Sections 8.103092 Variance Component for Cross Sections 0.634793 Hausman Test for Randon Effects DF n Value Pr > n 0 Parameter Estinates Variable DF Estinate Variable DF Estinate T 1 0.14197 2.9145 3.48 0.0000 T 1 0.02853 0.0126 -2.26 0.0241 DEGREES 1 -0.01609 0.00983 -1.64 0.1033 PUBLIC 1 -3.85366 1.6551 -2.33 0.0270	На	nsbeek and	i Kapteyn Var	iance Compor	ents (BanOr	ie J	
Nodel DescriptionEst ination Method Number of Cross SectionsRanOne 18 14Fit StatisticsSSE150.6509 0.6149DFE 0.7842245 0.7842MSE0.6149 0.1154Root MSE 0.78420.7842Variance Component EstimatesVariance Component for Cross Sections 0.634793Housman Test for Randon EffectsDF n Value Pr > n 0OOVariableDF n Value Pr > n 0OOVariableDF EstimatesVariableDF EstimateStandard Errort VariableDF EstimateStandard ErrorLogREES 1 -0.016090.0126-2.260.0241DEGREES 1 -0.016090.03972.91453.480.0000TotologDIGUI1 0.1411972.91453.480.0000Totolog0.0000Totolog0.0000Totolog0.0000Totolog0.00001 </td <td>Dependent Variable: FACUL</td> <td>TY</td> <td></td> <td></td> <td></td> <td></td>	Dependent Variable: FACUL	TY					
Noter Description Estimation Method Number of Cross Sections RanOne 18 14 Fit Statistics SSE 150.6509 0.6509 DFE 0.7842 MSE 0.6149 0.6149 Root MSE 0.7842 Variance Component Estimates Variance Component for Cross Sections Variance Component for Error Variance Component for Cross Sections 0.634793 Hausman Test for Randon Effects DF n Value Pr > n 0 OF n Value Pr > n 0 Standard Error Volue Pr > it Intercept 1 10.14197 2.3145 3.48 0.0001 0.0033 T 1 -0.16078 0.03937 2.67 0.0041 DEGREES 1 -0.16078 0.03937 2.67 0.0041 DEGREENAN 1 0.106078 0.03937 2.67 0.0087 DEGREENAN 1 0.16078 0.03937 2.67 0.0087 DEGREENAN 1 0.16078 0.03937 2.67 0.0087 DEGREENAN 1 0.58177 1.4024 0.41 0.6708			Model De	secietion			
Estimation Method RanOne Number of Cross Sections 18 Time Series Length 14 Fit Statistics SSE 150.6509 DFE 245 MSE 0.6149 Root MSE 0.7842 R-Square 0.1154 Variance Component Estimates Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausman Test for Randon Effects DF n Value $Pr > n$ 0 Parameter Estimates Variable DF Estimate Error t Value $Pr > 1t$ Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 -0.02853 0.0126 -2.26 0.0244 DEGREES 1 -0.016078 0.03937 2.67 0.0060 PUBLIC 1 -3.85366 1.65511 -2.33 0.0200			HOUGT DE	seription			
Number of Cross Sections 18 Time Series Length 14 Fit Statistics SSE 150.6509 DFE 245 MSE 0.6149 Root MSE 0.7842 R-Square 0.1154 Variance Component Estimates 0.7842 Variance Component for Cross Sections 8.109092 0.634793 Variance Component for Error 0.634793 Hausman Test for Randon Effects 0 DF n Value Pr > n 0 . . Parameter Estimates Variable DF Estimate Standard Error t Value Pr > 1t Intercept 1 10.14197 2.9145 3.48 0.0001 T 1 -0.02853 0.0126 -2.26 0.0241 DEGREES 1 -0.16098 0.0397 2.67 0.0241 DEGREEMR 0.106078 0.0397 2.67 0.0241 DEGREENR 1 0.186078 0.0397 2.67 0.0241		Estin	ation Method	1.1	Ran0ne		
Fit Statistics Fit Statistics SSE 150.6509 DFE 245 MSE 0.6509 DFE 245 MSE 0.6149 Root MSE 0.7842 Variance Component for Cross Sections 8.109092 Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausman Test for Randon Effects DF m Value Pr > n 0 OF Standard Error t Value Pr > 1 Intercept 1 0.1154 DF R Value Pr > n 0 - Parameter Estimates Variable DF Estimate Standard DF Standard Pr > 1 DF Standard Pr		Numbe	er of Cross S	lections	18		
Fit Statistics SSE 150.6559 DFE 245 MSE 0.6149 Root MSE 0.7842 R-Square 0.1154 Variance Component Estimates Variance Component for Cross Sections 8.109092 Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausman Test for Randon Effects DF n Value Pr > n 0 OF stimates Variable DF Estimate Variable DF Estimate DF ror t Value Pr > n 0 0 To 2.9145 Variable DF Estimate Variable DF Estimate Variable DF Estimate DE Colspan= 2 DE Estimate DE Colspan= 2 DE Colspo 0.00983 <td co<="" td=""><td></td><td>line</td><td>Ser les Lengt</td><td>'n</td><td>14</td><td></td></td>	<td></td> <td>line</td> <td>Ser les Lengt</td> <td>'n</td> <td>14</td> <td></td>		line	Ser les Lengt	'n	14	
$\begin{array}{cccccccccccccccccccccccccccccccccccc$			Fit St	atistics			
MSE 0.6149 Rot MSE 0.7842 R-Square 0.1154 Variance Component Estimates 0.7842 Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausman Test for Randon Effects 0.634793 DF Nalue Pr > n 0 . . Parameter Estimates Standard Pr > lt Intercept 1 10.14197 2.9145 3.48 0.0001 T 1 -0.02853 0.0126 -2.26 0.0241 DEGREES 1 0.106078 0.0397 2.67 0.0081 PUBLIC 1 -3.86366 1.6551 -2.33 0.0200	88	F	150 6509	DEE		245	
R-Square 0.1154 Variance Component Estimates Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausnan Test for Randon Effects DF m Value Pr > n 0 . Parameter Estimates Variable DF Estimate Standard Error Pr > 1t Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 -0.02853 0.0126 -2.26 0.0241 DEGREES 1 -0.1609 0.00983 -1.64 0.1031 DEGREBAR 1 0.106078 0.0397 2.67 0.0084 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204	MS	Ē	0.6149	Root MSE	0.7	842	
Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausman Test for Randon Effects DF n Value Pr > n 0 . Parameter Estimates Variable DF Estimate Standard Error Pr > 1t Intercept 1 10.14197 2.9145 3.48 0.0001 T 1 -0.01609 0.00983 -1.64 0.1031 DEGREES 1 -0.106078 0.0397 2.67 0.0044 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204	B-	Square	0.1154		0.000	19.99 A	
Variance Component Estimates Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausnan Test for Randon Effects DF n Value Pr > n 0 . Variable DF Estimates Variable Variable DF Estimate Variable Variable DF Estimate Variable Variable Variable DF Estimate Standard Intercept 1 10.14197 2.9145 3.48 0.0001 T 1 0.0241 DEGREES 1 0.0126 -2.26 0.0241 DEGREES 1 -0.01609 0.00397 2.67 0.0081 PUBLIC 1 </td <td></td> <td></td> <td>37 529</td> <td>100000000 0</td> <td></td> <td></td>			37 529	100000000 0			
Variance Component for Cross Sections 8.109092 Variance Component for Error 0.634793 Hausnan Test for Randon Effects DF n Value Pr > n 0 . . Parameter Estimates Variable DF n Value Pr > n 0 . . Parameter Estimates Variable DF Estimate Error t Value Pr > it Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 -0.02853 0.0126 -2.26 0.0241 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0081 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH001 1 0.58177 1.4024 0.41 0.6781		Va	ariance Compo	nent Estinat	es		
Variance Component for Error 0.634793 Hausman Test for Randon Effects DF n Value Pr > n 0 . . Parameter Estimates Variable DF Estimate Variable DF Estimate Variable DF Estimate Variable DF Estimate Error t Value Pr > it Intercept 1 10.14197 2.9145 3.48 0.0004 T 1 -0.02853 0.0126 -2.26 0.0241 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0081 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH001 1 0.58177 1.4024 0.41 0.6781	Va	riance Com	ponent for C	ross Section	ns 8.1090	92	
Hausman Test for Randon Effects DF n Value Pr > n 0 Parameter Estimates Variable DF Estimate Variable DF Estimate Variable DF Estimate Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 0.0126 -2.26 0.0241 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0080 PUBLIC 1 -0.106078 0.0397 2.67 0.0080 PUBLIC 1 -0.58167 1 -0.02853 -0.02853 -0.0397 2.67 0.0080 PUBLIC 1 -3.86366 1.6551 <th colspa="</td"><td>Va</td><td>riance Cor</td><td>ponent for E</td><td>rror</td><td>0.6347</td><td>'93</td></th>	<td>Va</td> <td>riance Cor</td> <td>ponent for E</td> <td>rror</td> <td>0.6347</td> <td>'93</td>	Va	riance Cor	ponent for E	rror	0.6347	'93
Handon Effects DF n Value Pr > n 0 Parameter Estimates Variable DF Estimate Standard Variable DF Estimate Standard Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 -0.02853 0.0126 -2.26 0.0244 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0084 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH001 1 0.58177 1.4024 0.41 0.678			Hausnar	Test for			
DF n Value Pr > n 0 . . Parameter Estimates Standard Variable DF Estimate Error t Value Pr > it Intercept 1 10.14197 2.9145 3.48 0.0001 T 1 -0.02853 0.0126 -2.26 0.0243 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0064 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH001 1 0.58177 1.4024 0.41 0.6784			Handor	Effects			
O . . Parameter Estimates Variable DF Estimate Standard Error Value Pr > it Intercept 1 10.14197 2.9145 3.48 0.0004 T 1 -0.02853 0.0126 -2.26 0.0244 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0084 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH000 1 0.58177 1.4024 0.41 0.6784			DF m	Value Pr	> n		
Parameter Estimates Variable DF Estimate Standard Error Value Pr > it Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 -0.02853 0.0126 -2.26 0.0241 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0088 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH000 1 0.58177 1.4024 0.41 0.6788			0				
Variable DF Estinate Standard Error Value Pr > t Intercept 1 10.14197 2.9145 3.48 0.0001 T 1 -0.02853 0.0126 -2.26 0.0244 DEGREES 1 -0.01609 0.00983 -1.64 0.1033 DEGREBAR 1 0.106078 0.0397 2.67 0.0084 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCHODI 1 0.58177 1.4024 0.41 0.6784			Parameter	Estinates			
Variable DF Estinate Error t Value Pr > it Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 -0.02853 0.0126 -2.26 0.0241 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0086 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCHODI 1 0.58177 1.4024 0.41 0.678				Standard			
Intercept 1 10.14197 2.9145 3.48 0.0000 T 1 -0.02853 0.0126 -2.26 0.0244 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0084 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH001 1 0.58177 1.4024 0.41 0.6784	Var iab le	DF	Estinate	Error	t Value	Pr > [t]	
T 1 -0.02853 0.0126 -2.26 0.0240 DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.0080 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH001 1 0.58177 1.4024 0.41 0.678	Intercept	1	10.14197	2.9145	3.48	0.0006	
DEGREES 1 -0.01609 0.00983 -1.64 0.1030 DEGREBAR 1 0.106078 0.0397 2.67 0.008 PUBLIC 1 -3.86366 1.6551 -2.33 0.0204 BSCH001 1 0.58177 1.4024 0.41 0.678	T	1	-0.02853	0.0126	-2.26	0.0248	
DEGREBAR 1 0.106078 0.0397 2.67 0.008 PUBLIC 1 -3.86366 1.6551 -2.33 0.020 BSCHDDI 1 0.58177 1.4024 0.41 0.6781	DEGREES	1	-0.01609	0.00983	-1.64	0.1030	
PUBLIC: 1 -3.86366 1.6551 -2.33 0.020 BSCH001 1 0.58177 1.4024 0.41 0.6781	DEGREBAR	1	0.106078	0.0397	2.67	0.0080	
BSCHOOL 1 0.58177 1.4024 0.41 0.678	PUBL IC	1	-3.86366	1.6551	-2.33	0.0204	
	BSCHOOL	1	0.58177	1.4024	0.41	0.6786	

The marginal effect of an additional economics major is again insignificant but slightly negative within the sample. Both the short-term moving average number and long-term average number of bachelor degrees are significant. A long-term increase of about 10 students earning degrees in economics is required to predict that one more tenured or tenure-track faculty member is in a department. Ceteris paribus, economics departments at private institutions are smaller than comparable departments at public schools by a large and significant number of four members. Whether there is a business school present is insignificant. There is no meaningful trend in faculty size.

It should be clear that this regression is NOT identical to similar one-way random effect models controlling for clustering in LIMDEP or STATA. The standard errors are adjusted for a general groupwise heteroscedastic error structure. The difference does not alter the significance and the standard errors are, for the most part, very comparable.

CONCLUDING REMARKS

The goal of this hands-on component of this third of four modules is to enable economic education researchers to make use of panel data for the estimation of constant coefficient, fixed-effects and random-effects panel data models in SAS. It was not intended to explain all of the statistical and econometric nuances associated with panel data analysis. For this an intermediate level econometrics textbook (such as Jeffrey Wooldridge, *Introductory Econometrics*) or advanced econometrics textbook (such as William Greene, *Econometric Analysis*) should be consulted.

APPENDIX: Alternative Means to Estimate Random-Effects Model with Clustered Data.

The following code provides a necessary code to estimate the random-effect models with clustering. The estimation procedure is two-step feasible GLS. In the first step, the variance matrix is estimated. In the second step, this variance matrix is used to transform the equation.

Because the variance matrix is *estimated* and not the true variance, this causes the standard errors to be slight different than the standard errors provided by LIMPDEP or STATA when estimating a random effects model with clustering.

The code to be run in the editor window is:

```
/* get SSE and SSU */
proc sort data= bachelors2;
by college year; quit;
proc tscsreg data=bachelors2 outest=covvc;
id college year;
model faculty = t degrees degrebar public bschool MA_deg / ranone;
quit;
/* find number of years */
data numobs (keep = year);
set bachelors2;
run;
proc sort nodupkey;
by year;
quit;
proc means data = numobs
      max;
output out = num;
      quit;
/* create lamda */
proc iml;
use covvc;
read all var {_VARERR_ _VARCS_} into x;
use num;
read var {_freq_} into y;
print y;
sesq = x[1,1];
susq = x[1,2];
lamda = 1 - sqrt( sesq / (y[1,1]*susq + sesq) );
print x y lamda;
cname = {"lamda"};
```

```
create out from lamda [ colname=cname];
append from lamda;
quit;
/* find averages of each variable grouped by college #*/
proc MEANS NOPRINT
data=bachelors2;
class college;
output out=stats
mean= avg_year avg_degrees avg_degrebar avg_public avg_faculty avg_bschool
avg_t avg_ma_deg;
run;
data bachelors3 (drop = _type_ _freq_);
      merge bachelors2 stats;
      by college;
      if _type_ = 0 then delete;
      one = 1;
      run;
DATA bachelors4;
    if _N_ = 1 then set out;
    SET bachelors3;
      l = one*lamda;
run;
/* transform data */
data clean (keep = college con nfaculty nt ndegrees ndegrebar npublic
nbschool nMA_deg year);
      set bachelors4;
nfaculty = faculty - lamda*avg_faculty;
nt = t - lamda*avg_t;
ndegrees = degrees - lamda*avg_degrees ;
ndegrebar = degrebar - lamda*avg_degrebar;
npublic = public - lamda*avg_public;
nbschool = bschool - lamda*avg_bschool;
nMA_deg = ma_deg - lamda*avg_ma_deg;
con = 1 - lamda*1;
run;
/* run regression on transformed equation assuming clustering */
/* Since intercept is included in transformed equation, use noint option*/
proc surveyreg data=clean;
      cluster college;
      model nfaculty = con nt ndegrees ndegrebar npublic nbschool nMA_deg /
      noint;
quit;
```

The output for this regression is:

	Regression Analysis for Dependent	Variable nfaculty	
	Data Summary		
	Number of Observations Mean of nfaculty Sum of nfaculty	252 0.72801 183.45775	
	Design Summary		
oin 8-30	Number of Clusters	18	
	Fit Statistics		

0.5134

R-square

Gil

15

The standard errors associated with this regression are much closer to the standard errors from LIMPDEP and STATA. However, this is a complex sequence of codes which should not be attempted by beginners.

REFERENCES

Becker, William, William Greene and John Siegfried (2009). "Does Teaching Load Affect Faculty Size?" Working Paper (July).

Cameron, Colin and Pravin Trivedi (2005). *Microeconometrics*. 1st Edition, New York, Cambridge University Press.

Mundlak, Yair (1978). "On the Pooling of Time Series and Cross Section Data," *Econometrica*. Vol. 46. No. 1 (January): 69-85.

Greene, William (2008). Econometric Analysis. 6th Edition, New Jersey: Prentice Hall.

Wooldridge, Jeffrey (2009). *Introductory Econometrics*. 4th Edition, Mason OH: South-Western.