Econ 141: Solutions to Homework 5

Jon Schellenberg

Problem 1

(a) See regression output in do file. Population Regression Function is as follows:

$$\ln(salary_i) = \beta_0 + \beta_1 \ln(sales_i) + \beta_2 ros_i + \beta_3 roe_i + u_i$$

Normal form given below:

$$\underset{}{\ln(\widehat{salary_i})} = \underset{(0.315)}{4.311} + \underset{(0.035)}{0.280}\ \ln(sales_i) + \underset{(0.0041)}{0.0174}\ roe_i + \underset{(0.00054)}{0.00024}\ ros_i \qquad n = 209,\quad R^2 = 0.283$$

Also, this is not part of the normal form regression, but note from our output that $RSS_{ur} = 47.86$

(b) Here, we are doing a two-tailed test to see if the coefficients on ros_i and roe_i are equal. Formally, this is the test:

(i) $H_0 : \beta_2 = \beta_3$, or $\underbrace{\begin{bmatrix} 0 & 0 & 1 & -1 \end{bmatrix}}_{\boldsymbol{R}} \boldsymbol{\beta} = \underbrace{\begin{bmatrix} 0 \end{bmatrix}}_{\boldsymbol{r}}$

(ii) $H_1 : \beta_2 \neq \beta_3$

(iii) Assume H_0 is true. Under this assumption, we know that $F \sim F_{1,205}$. (You can also do a t-test here, we'll only show the F-test answer because Stata only reports the F-test).

(iv) F-stat: There are two ways to calculate this:

I. Use the formula $F = (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})'(\boldsymbol{R}\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}')^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})/q$
Stata does this computation with the "test" command - using this, you get $F = 16.26$

II. Run restricted and unrestricted regressions to collect RSS for the formula $F = \frac{(RSS_r - RSS_{ur})/q}{RSS_{ur}/(n-k-1)}$.

Restricted Regression:

$$\ln(salary)_i = \beta_0 + \beta_1 \ln(sales)_i + \beta_2(ros_i + roe_i) + u_i$$

After running this in Stata, you get $RSS_r = 51.66$

$$F = \frac{(51.66 - 47.86)/1}{47.86/(209 - 3 - 1)} = 16.26$$

(v) Rejection rule: we will reject iff $p < 0.05$. (Alternatively, we will reject iff $F > Fcrit$. Here, the critical value at the 5% significance level $F_{0.05,1,205} = 3.887$.)

(vi) Here, we see that $p = 0.0001 < 0.05$ (Alternatively, $F = 16.26 > Fcrit$).

(vii) Therefore, we reject our null hypothesis and conclude that $\beta_2 \neq \beta_3$, or in other words, return on equity and return on stocks do not equally impact CEO salaries (when holding sales constant).

(c) Here, we are doing a two-tailed test to see if the coefficients on $\ln(sales)_i$, ros_i and roe_i are jointly zero. Formally, this is the test:

(i) $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$, or $\underbrace{\begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\boldsymbol{R}} \boldsymbol{\beta} = \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}}_{\boldsymbol{r}}$

(ii) H_1 : at least one of $\beta_1, \beta_2, \beta_3$ are not zero.

(iii) Assume H_0 is true. Under this assumption, we know that $F \sim F_{3,205}$.

(iv) F stat: There are two ways to calculate this.

I. Use the formula $F = (\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})'(\boldsymbol{R}\widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}')^{-1}(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r})/q$
Stata does this computation with the "test" command - using this, you get $F = 26.93$

II. Run restricted and unrestricted regressions to collect RSS for the formula $F = \frac{(RSS_r - RSS_{ur})/q}{RSS_{ur}/(n-k-1)}$.

Restricted Regression:

$$\ln(salary)_i = \beta_0 + u_i$$

After running this in Stata, you get $RSS_r = 66.72$

$$F = \frac{(66.72 - 47.86)/3}{47.86/(209 - 3 - 1)} = 26.93$$

(v) Rejection rule: we will reject iff $p < 0.05$. (Alternatively, we will reject iff $F > Fcrit$. Here, the critical value at the 5% significance level $F_{0.05,3,205} = 2.649$.)

(vi) Here, we see that $p < 0.0001 < 0.05$ (Alternatively, $F = 26.93 > Fcrit$).

(vii) Therefore, we reject our null hypothesis and conclude that our regressor coefficients are not jointly zero, or in other words, at least one of the following variables affects CEO salaries: sales, return on equity, and return on stock.

In the do file, F is calculated using both matrices and restricted/unrestricted models - either one works.

Problem 2

(a) There are two potential solutions for this part:

(i) **Math Solution:** Let $\boldsymbol{y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$, $\boldsymbol{x}_i = \begin{bmatrix} 1 \\ X_{1,i} \\ X_{2,i} \end{bmatrix}$ for all i, $\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} = \begin{bmatrix} 1 & X_{1,1} & X_{2,1} \\ 1 & X_{1,2} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{1,n} & X_{2,n} \end{bmatrix}$, $\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix}$,

and $\widehat{\boldsymbol{\alpha}} = \begin{bmatrix} \widehat{\alpha}_0 \\ \widehat{\alpha}_1 \\ \widehat{\alpha}_2 \end{bmatrix}$ (as usual). Also, (this is new) let $\boldsymbol{X}_2 = \begin{bmatrix} X_{2,1} \\ X_{2,2} \\ \vdots \\ X_{2,n} \end{bmatrix}$ be the $(n \times 1)$ column matrix of the observations of second regressors. Notice that $\boldsymbol{X}_2 = \boldsymbol{X} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$. We can rewrite our "stacked" versions of our regression equations as follows (with $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{u}$, and $\boldsymbol{v}$ defined as you would expect):

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} \tag{1}$$
$$\boldsymbol{y} - \boldsymbol{X}_2 = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{v} \tag{2}$$

Using the OLS formula for our coefficient estimates
$(Coefficients = (Regressors'Regressors)^{-1}Regressors'DependentVariable)$, we get the follow-

ing for our regressors:

$$
\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\
\widehat{\boldsymbol{\alpha}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X}_2) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\left(\boldsymbol{y} - \boldsymbol{X}\begin{bmatrix}0\\0\\1\end{bmatrix}\right) \\
&= \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}}_{=\widehat{\boldsymbol{\beta}}} - \underbrace{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}}_{=\boldsymbol{I}_3}\begin{bmatrix}0\\0\\1\end{bmatrix} \\
&= \widehat{\boldsymbol{\beta}} - \begin{bmatrix}0\\0\\1\end{bmatrix}
\end{aligned}
$$

From this, it's easy to see that

- $\widehat{\alpha}_0 = \widehat{\beta}_0$
- $\widehat{\alpha}_1 = \widehat{\beta}_1$
- $\widehat{\alpha}_2 = \widehat{\beta}_2 - 1$

(ii) **Less math solution:** Note that as long as your data are not perfectly collinear, the coefficients that solve the SSR minimization problem will be unique. Recall that our optimization problems that we're using to solve for our coefficients look like this:

- For the $\widehat{\beta}$s:

$$\min_{\widehat{\beta}_0,\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} \widehat{u}_i^2 = \min_{\widehat{\beta}_0,\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} \left(Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_{1,i} + \widehat{\beta}_2 X_{2,i})\right)^2$$

- For the $\widehat{\alpha}$s:

$$
\begin{aligned}
\min_{\widehat{\alpha}_0,\widehat{\alpha}_1,\widehat{\alpha}_2} \sum_{i=1}^{n} \widehat{u}_i^2 &= \min_{\widehat{\alpha}_0,\widehat{\alpha}_1,\widehat{\alpha}_2} \sum_{i=1}^{n} (Y_i - X_{2,i} - (\widehat{\alpha}_0 + \widehat{\alpha}_1 X_{1,i} + \widehat{\alpha}_2 X_{2,i}))^2 \\
&= \min_{\widehat{\alpha}_0,\widehat{\alpha}_1,\widehat{\alpha}_2} \sum_{i=1}^{n} (Y_i - (\widehat{\alpha}_0 + \widehat{\alpha}_1 X_{1,i} + (\widehat{\alpha}_2 + 1) X_{2,i}))^2
\end{aligned}
$$

Notice that the minimizations of the two look...very similar - the only things that change are the coefficient names that are in front of our variables. However, these are what we are optimizing over to minimize SSR, and since the solutions must be unique, the multiplicative factors in each sum be the same. Thus,

- $\widehat{\alpha}_0 = \widehat{\beta}_0$
- $\widehat{\alpha}_1 = \widehat{\beta}_1$
- $\widehat{\alpha}_2 + 1 = \widehat{\beta}_2$

(b) Again, you can either solve this in scalar form or matrix form. The solution below writes it out in scalar form.

The residuals $\widehat{u}_i$ from regression 1 are as follows (by definition):

$$\widehat{u}_i = Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_{1,i} + \widehat{\beta}_2 X_{2,i})$$

The residuals $\widehat{v}_i$ from regression 2 are as follows (by definition):

$$\begin{aligned}
\widehat{v}_i &= (Y_i - X_{2,i}) - (\widehat{\alpha}_0 + \widehat{\alpha}_1 X_{1,i} + \widehat{\alpha}_2 X_{2,i}) \\
&= Y_i - (\widehat{\alpha}_0 + \widehat{\alpha}_1 X_{1,i} + (\widehat{\alpha}_2 + 1) X_{2,i}) \\
\text{(using what we found in part (a))} \quad &= Y_i - (\widehat{\beta}_0 + \widehat{\beta}_1 X_{1,i} + \widehat{\beta}_2 X_{2,i}) \\
&= \widehat{u}_i
\end{aligned}$$

Problem 3

(a) Table given below.

	(1) OLS	(2) OLS	(3) FE	(4) FE	(5) FE
fhpolr (t-1)	0.679***	0.701***	0.701***	0.361***	0.230***
	(0.0245)	(0.0253)	(0.0370)	(0.0480)	(0.0741)
lrgdppc (t-1)	0.0830***	0.0739***	0.0739***	-0.00518	-0.000585
	(0.00847)	(0.00866)	(0.0101)	(0.0333)	(0.0610)
laborshare (t-1)					0.206
					(0.137)
lpop (t-1)					-0.0741
					(0.108)
socialism					0
					(.)
State dummies?	NO	NO	NO	YES	YES
Year Dummies?	NO	YES	YES	YES	YES
R-squared	0.698	0.713	0.713	.233	.137
RSS	36.32	34.52	34.52	25.42	12.69
n	906	906	906	906	474
Standard Errors			cluster	cluster	cluster

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

(b) Coefficient on $flpolr_{t-1}$: This is "the effect of a unit increase on the past period's Freedom House Political Rights Index on the current periods' Freedom House Political Rights Index, while holding last period's GDP constant." That is a one unit increase in last period's democratic freedom is predicted to increase this year's democratic freedom by about 0.7 points *c.p.* This could be interpreted as considerable amount of persistence in democracy.

Coefficient on $lrgdppc_{t-1}$: This is "the effect of a percent increase in the past period's GDP on the current periods' Freedom House Political Rights Index, while holding last period's Freedom House Political Rights Index constant." That is if last year's measure of income per capita increases by one percent our measure of democratic freedom is predicted to increase by $0.083{*}10^{-2}$ (or 0.00073) points, *c.p.*

(c) With panel data, it's usually a good idea to add both. In this case, you *definitely* need to add state fixed effects, because different countries have different average levels of democracy over time (think the US vs North Korea - you have to account for the fact that US is much more democratic than North Korea, so including country fixed effects will help account for this difference). Differences in culture and historical factors would influence political developments. Additionally, it's probably a good idea to include time fixed effects as well, since countries are becoming more democratic over time, so the average level of democracy in the world differs across years.

(d) Time fixed effects don't change the results by much (the confidence intervals for both coefficients between columns 1 and 3 overlap, so they aren't that noticeably different). This indicates that countries over time aren't really changing their level of "average level of democracy" by much. However, once we add country fixed effects, the estimates change by a LOT, which indicates that observed countries vary greatly in their average level of democracy.

(e) This is testing that the country fixed effects are not necessary for explaining level od democracy. Alternatively, if we run the regression by including an overall intercept and then all but one dummy for each possible country, this would be testing that the coefficients on those dummy variables are jointly 0. The test performed in Stata (hw5 final.do) has the following regression function:

$$fhpolr_{it} = \beta_0 + \beta_1 lrgdppc_{it-1} + \beta_2 lrgdppc_{it-1} + \beta_3 laborshare_{it-1} + \beta_3 lpop_{it-1} + \beta_4 socialism_{it} + \underbrace{\alpha_2 Country2_{it} + ... + CountryN_{it}}_{\text{country fixed effects}} + \underbrace{\gamma_2 Time2_{it} + ... + \gamma_T TimeTit}_{\text{time fixed effects}} + u_{it}$$

Our test has been written below:

(i) $H_0 : \alpha_2 = ... = \alpha_N = 0$

(ii) H_1 : Not H_0

(iii) Assume H_0 is true. Under this assumption, we know that $F \sim F_{10,96}$ (we get the dfs from Stata after doing the hypothesis test. It's not immediately apparent why there are only 10 restrictions, but basically since we're really restricting the data, it turns out that several of our restrictions are collinear (for example, it would be like if we had one restriction that was $x = 2$ and another that was $2x = 4$, which are really testing the same thing) - don't worry about this too much).

(iv) Rejection rule: we will reject iff $p < 0.05$. (Alternatively, we will reject iff $F > Fcrit$. Here, the critical value at the 5% significance level $F_{0.05,10,96} = 1.931$.)

(v) Here, we see that $p < 0.05$ (Alternatively, $F = 168.56 > Fcrit$).

(vi) Therefore, we reject our null hypothesis and conclude that the coefficients on our country dummies are not jointly 0 and it is appropriate to include them in the regression.

If you did the hypothesis test for regression 4, that would be OK too - you still get the same conclusion.

(f) The controls do a bit to the $fhpolr_{t-1}$ coefficient - the value drops to about 2/3 of what it was earlier, but it is still significant and positive. The coefficient on $lrgdppc_{t-1}$ doesn't change much.

However, note that the *socialism* variable was dropped from our regression (omitted due to multicollinearity). This is because in our data set, none of the countries change their socialism status over time - they either are always socialist or never socialist. Thus, this variable will be collinear with the country fixed effects, which also vary across individual and do not vary over time.

(g) No, that's why we included fixed effects. For more, see answers to parts c and d and more formal test from part e.

(h) From regressions (1)-(3), we see that the coefficient on lagged GDP per capita is positive and significantly different from 0 at the 5% level (in fact, if we did our 1-tailed hypothesis tests, we would see that we can conclude that the effect is positive)- this would suggest that increases in GDP lead to increases in future political freedoms. However, these regressions fail to account for the differences in average country freedom levels, which is undesirable. Inclusion of fixed effects (specifications (4) and (5)) makes this relationship to disappear. The coefficient on $lrgdppc_{t-1}$ is not statistically different from 0 at 5% significance.

Furthermore we can't claim any causality even for regressions where the coefficient on $lrgdppc_{t-1}$ is statistically significant. Those regressions are likely to suffer from the OVB and simultaneous causality. (more on that later in the course)