Week 2: Multiple Regression, Training Exercises

Coursera/Erasmus U., Econometric Methods and Applications

Anthony Nguyen

Training Exercise 2.1

Notes:

- This exercise uses the datafile `TrainExer21`and requires a computer.
- The dataset `TrainExer21` is available on the website.

Questions

(a) Use dataset `TrainExer21` to regress log-wage on a constant and the gender dummy `Female`, and check the result presented in Lecture 2.1 that $log(Wage) = 4.73 - 0.25Female + e$.

(b) Let e be the series of residuals of the regression in part (a). Perform two regressions:

(i) e on a constant and education;
(ii) e on a constant and the part-time job dummy.

(c) Comment on the outcomes of regressions (i) and (ii) of part (b).

Answers

(a) Use dataset `TrainExer21` to regress log-wage on a constant and the gender dummy `Female`, and check the result presented in Lecture 2.1 that $log(Wage) = 4.73 - 0.25Female + e$.

We can verify the results from the lecture by quickly running a binary regression on `LogWage` by `Female`:

Table 1:

	Dependent variable:
	LogWage
Female	-0.251^{***}
	(0.040)
Constant	4.734^{***}
	(0.024)
Note:	*p<0.1; **p<0.05; ***p<0.01

We can see that the results are the same, with the intercept β_0 (or α if you prefer) term = 4.734 with a Standard Error = 0.024.

Likewise, we can see that the β_1 value = -0.251, with a Standard Error = .040.

(b) Let e be the series of residuals of the regression in part (a). Perform two regressions:

Using the residuals from this regression, we can now run two separate regressions to see if the error terms correlate with other variables in our dataset.

- **(i) e on a constant and education;**

Table 2:

	Dependent variable:
	residuals(LogWage_Female)
Educ	0.218^{***}
	(0.016)
Constant	-0.453^{***}
	(0.036)
Note:	*p<0.1; **p<0.05; ***p<0.01

- **(ii) e on a constant and the part-time job dummy.**

Table 3:

	Dependent variable:
	residuals(LogWage_Female)
Parttime	0.099**
	(0.043)
Constant	−0.028
	(0.023)
Note:	*p<0.1; **p<0.05; ***p<0.01

(c) Comment on the outcomes of regressions (i) and (ii) of part (b).

If the residual error from our first regression in part *a* can be thought of as wage difference between men and women that cannot be explained by our original regressor variable, `Female`, then the regressions we performed on these error terms in part *b* hint at what portion of this unexplained wage difference can be explained by other variables.

Specifically, what the first regression we performed in part *b* says is that an extra level of Education has a 0.218 (22%) effect on the unexplained wage difference between men and women. In other words, higher education levels leads to higher wages.

Similarly, in the second regression, we see that having a Part-timejob has an effect of 0.099 (or 10%) on the unexplained wage difference between men and women. This is unexpected, as we would have expected lower-wages for part-time work.

The results of both of these regressions on our residual error from part *a* suggest that these additional variables should be added to our model for predicting differences in `LogWage`.