ECON613 Assignment 4

Promvarat Pradit

April 16, 2019

Exercise 1 Data

Load the data “Koop - Tobias”, which comes from Koop and Tobias (2004) Labor Market Experience Data. (See Koop, G. and J. Tobias, ”Learning About Heterogeneity in Returns to Schooling,” Journal of Applied Econometrics, 19, 2004, pp. 827-849.

The data file is in two parts. The first file contains the panel of 17,919 observations on the Person ID and 4 time-varying variables. The second file contains time invariant variables for the individual or the 2,178 households. Represent the panel dimension of wages for 5 randomly selected individuals.

Solution to Exercise 1

Figure 1 shows panel dimension of $log(wage)$ for 5 randomly selected individuals. Rows in the figure represent each individual ID while the columns show time period for the data. From the figure we can notice that there are many NA which indicate that our data is unbalanced, and we should be carefully investigate if the missing values are random or not. If so, we could proceed on studying this data.

Figure 1: Panel dimension of 5 randomly selected individuals (Row=ID , Column=Time)

	0	3	4	5	6	7	8	9	10	11	12	13	14
993	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.11	NA
1571	NA	NA	NA	NA	NA	1.27	3.72	0.99	0.83	1.09	1.20	0.20	0.83
1656	NA	NA	NA	NA	1.38	NA	0.84	1.99	NA	2.13	2.16	2.13	2.16
1907	2.29	2.71	1.57	NA	2.20	2.69	2.35	2.29	3.15	2.26	2.33	2.20	2.00
1928	NA	1.80	2.10	1.98	2.35	2.42	2.40	2.50	2.29	2.31	2.49	2.63	2.60

Exercise 2 Random Effects

We are interested in

$$LOGWAGE_{it} = \alpha_i + \beta_1 EDUC_{it} + \beta_2 POTEXPR_{it} + \epsilon_{it} \tag{1}$$

Where α_i is a random effect. Estimate the random effect model under the normality assumption of the disturbance terms.

Solution to Exercise 2

The estimates from random effect model is presented in Figure 3. The estimates suggest that 1 unit increase in education will increase wage by approximately $exp(0.09386374) - 1 \approx 9.84$ percent, and 1 unit increase in potential experience will increase wage by approximately $exp(0.0374053) - 1 \approx 3.81$ percent.

Figure 2: Estimated coefficients from Random Effect model

	(Intercept)	EDUC	POTEXPER
Estimated Coefficient	0.7941911	0.09386374	0.0374053

Exercise 3 Fixed Effects Model

We are interested in

$$LOGWAGE_{it} = \alpha_i + \beta_1 EDUC_{it} + \beta_2 POTEXPR_{it} + \epsilon_{it} \tag{2}$$

Where α_i is individual fixed effect. Estimate the following estimators

- Between Estimator
- Within Estimator
- First time difference Estimator

Compare the estimates of β_1 and β_2 under the different models.

Solution to Exercise 3

The estimates from all 3 fixed effect model is presented in Figure 3. We can see that β_1 and β_2 for all of the 3 estimators are positive as expected, which means that education and potential experience will improve wage outcome. However, the magnitude of the estimate for first time difference estimator is a lot lower than the other 2 estimators. This suggests that there might be some time effect such as time trends in the data.

Considering our variables, it is very likely that our data is having a time trend. Wage for an individual usually being raised every year (I assume wage here is nominal since the author did not mention about real wage in the article). Education is only stable or increasing over year as it is measured by year of schooling. Lastly, potential labor market experience is defined as (Age - Education - 5) so it is only stable or increasing over year as well. Hence, since our model did not include and time trend, I think the estimates from first time difference model is the most reliable in this case.

Figure 3: Estimated coefficients from all 3 Fixed Effect models

	(Intercept)	EDUC	POTEXPER
Between Estimates	8.455688e-01	0.09309987	0.025998745
Within Estimates	9.231826e-17	0.12366202	0.038561065
First Time Difference Estimates	4.946435e-02	0.03835231	0.003989071

Exercise 4 Understanding Fixed Effects

In the rest of the assignment, we consider only a random selected 100 individuals. We are interested in

$$LOGWAGE_{it} = \alpha_i + \beta_1 EDUC_{it} + \beta_2 POTEXPR_{it} + \epsilon_{it} \tag{3}$$

Where α_i is individual fixed effect.

1. Write and optimize the likelihood associated to the problem and estimate the individual fixed effect parameters

   Solution to Exercise 4.1)

   The model is written in the associated R script.

2. Run a regression of estimated individual fixed effects on the invariant variables.

   Solution to Exercise 4.2)

   I use this model

   $$\hat{\alpha}_i = c+\gamma_1 ABILITY+\gamma_2 MOTHERED+\gamma_3 FATHERED+\gamma_4 BRKNHOME+\gamma_5 SIBLINGS+\epsilon_i$$

   to regress individual fixed effects on the invariant variables. The regression result is shown in Figure 4. Of all 5 variables, only ability is significant (at 90% confident level). In addition, R^2 is only 0.06519 which indicates that these time invariant variables cannot explain much of the variation in individual fixed effects. Hence, fixed effect model is preferred to pooled OLS with time invariant variables.

Figure 4: Regression Result from the individual fixed effects ← invariant variables model

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.343805   0.225272  -1.526   0.1303
ABILITY     -0.134068   0.057695  -2.324   0.0223 *
MOTHERED     0.024175   0.019439   1.244   0.2167
FATHERED     0.009404   0.015691   0.599   0.5504
BRKNHOME    -0.123595   0.108555  -1.139   0.2578
SIBLINGS     0.003080   0.018586   0.166   0.8687
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4387 on 94 degrees of freedom
Multiple R-squared:  0.06519,   Adjusted R-squared:  0.01546
F-statistic: 1.311 on 5 and 94 DF,  p-value: 0.266
```

3. The standard errors in the previous may not be correctly estimated. Explain why, and propose an alternative method to compute standard errors.

<u>**Solution to Exercise 4.3)**</u>

The standard errors in previous question may not be correct because the dependent variable we used, $\hat{\alpha}_i$, is estimated from the question 4.1), so it has uncertainty ($S.E._{\hat{\alpha}_i}$) attached to it. In particular, using $\hat{\alpha}_i$ as the dependent variable introduces measurement error in our dependent variable. To fix this, we could use bootstrapping technique to generate many γ and find S.D. from the variation instead. The S.E. form bootstrapping process is reported in Figure 5.

Figure 5: S.E. for gamma from Bootstrapping process

	SE from Bootstrapping
(Intercept)	0.35017928
ABILITY	0.04191664
MOTHERED	0.01734622
FATHERED	0.01556473
BRKNHOME	0.11309096
SIBLINGS	0.02199740