Econometrics II: Assignment 2

Walter Verwer & Bas Machielsen

January 12, 2021

Question 1

First use pooled OLS to check the impact of including and excluding ASVABC on the estimate of α_1. Present and explain the result.

	Dependent variable:	
	(1)	(2)
ASVABC	0.123***	
	(0.005)	
AGE	0.436***	0.389***
	(0.061)	(0.061)
AGESQ	-0.001	-0.000
	(0.001)	(0.001)
S	0.789***	1.042***
	(0.020)	(0.017)
ETHBLACK	-1.218***	-2.299***
	(0.124)	(0.115)
URBAN	1.301***	1.354***
	(0.086)	(0.087)
REGNE	-2.301***	-1.499***
	(0.189)	(0.187)
REGNC	-3.810***	-3.062***
	(0.184)	(0.182)
REGW	-2.956***	-2.223***
	(0.191)	(0.189)
REGS	-3.889***	-3.308***
	(0.182)	(0.181)
Observations	40,043	40,043
R^2	0.225	0.215
Adjusted R^2	0.225	0.215
Residual Std. Error	7.127(df = 40033)	7.173(df = 40034)
F Statistic	1293.484*** (df = 9.0; 40033.0)	1372.448*** (df = 8.0; 40034.0)
Note:		*p<0.1; **p<0.05; ***p<0.01

The inclusion of the proxy ability decreases the estimate for the coefficient of schooling. Hence, given all other standard assumptions, ability and schooling are positively correlated, and the omission of a proxy for ability overestimates the impact of schooling.

Question 2

Perform a pooled OLS analysis to obtain insight in the heterogeneity of returns to schooling by ethnicity. Present the results and comment on the outcomes: what are the conclusions based on this?

	Dependent variable:		
	Interaction	Not Black	Black
	(1)	(2)	(3)
BLACKxS	-0.076		
	(0.053)		
ASVABC	0.123***	0.117***	0.165***
	(0.005)	(0.006)	(0.011)
AGE	0.434***	0.457***	0.140
	(0.061)	(0.067)	(0.117)
AGESQ	-0.001	-0.001	0.001
	(0.001)	(0.001)	(0.002)
S	0.796***	0.791***	0.732***
	(0.021)	(0.022)	(0.041)
ETHBLACK	-0.242	0.000***	-8.262***
	(0.688)	(0.000)	(1.358)
URBAN	1.298***	1.431***	-0.079
	(0.086)	(0.093)	(0.210)
REGNE	-2.318***	-2.402***	-1.232***
	(0.190)	(0.208)	(0.372)
REGNC	-3.825***	-3.921***	-2.391***
	(0.184)	(0.202)	(0.372)
REGW	-2.970***	-3.085***	-1.494***
	(0.191)	(0.209)	(0.408)
REGS	-3.904***	-3.942***	-3.144***
	(0.182)	(0.201)	(0.352)
Observations	40,043	35,223	4,820
R^2	0.225	0.213	0.304
Adjusted R^2	0.225	0.213	0.303
Note:			*p<0.1; **p<0.05; ***p<0.01

We can see that the interaction effect is insignificant: that is to say, there is no significant different between blacks and non-black in the influence of schooling on earnings. When we split up the sample into blacks and non-black, we get a slightly different view: the point estimate for the effect of schooling seems to be slightly lower for black people than for non-black people. However, as seen in the pooled regression with interaction effect, the differential impact is not statistically significant.

Question 3

Perform the analysis for heterogenous schooling effects using the random effects model. Present the results and compare the outcomes with the pooled OLS results obtained before. Interpret the outcomes.

Table 1:

	Dependent variable:
	EARNINGS
ASVABC	0.124***
	(0.011)
AGE	0.433***
	(0.051)
AGESQ	−0.001
	(0.001)
S	0.826***
	(0.036)
ETHBLACK	1.908
	(1.236)
URBAN	0.559***
	(0.103)
REGNE	−15.959***
	(0.865)
REGNC	−17.416***
	(0.855)
REGW	−16.295***
	(0.862)
REGS	−17.305***
	(0.851)
BLACKxS	−0.231**
	(0.095)
Observations	40,043
R^2	0.163
Adjusted R^2	0.163
F Statistic	35,260.370***
Note:	*p<0.1; **p<0.05; ***p<0.01

The point estimate for schooling is now close to the point estimate for schooling in the pooled OLS regression including the proxy for ability. Hence, the random effects estimator looks a lot like the pooled estimator, indicating that the contribution from the within group estimator is marginal. This can also be observed when looking at the decomposition of the explained variance: the between R-squared is larger than the within R-squared, indicating the model does a better job explaining the changes between individuals rather than individuals over time.

(Still need to explain: black ethnicity · schooling is significant)

Question 4

A priori, would you plead for using fixed effects estimation or random effects estimation? Explain your answer.

A priori, it would make more sense to use fixed-effects rather than random effects, because it is very likely that the unobservable individual components η_i are correlated to the predictor variables X rather than being random. For example, η_i can be interpreted as being some measure of ability or innate willingness to exert effort, and that is likely related to age, schooling and test score. A possible correlation would violate the randomness of η_i required by random effects, and hence, fixed effects would be preferred.

Question 5

Apply the fixed effects estimator to analyze the heterogenous schooling effect. Interpret the outcomes.

Table 2:

	Dependent variable:
	EARNINGS
AGE	0.428***
	(0.052)
AGESQ	−0.0004
	(0.001)
S	0.852***
	(0.071)
URBAN	0.204*
	(0.118)
REGNE	0.672**
	(0.300)
REGNC	−0.447*
	(0.262)
REGW	1.060***
	(0.300)
BLACKxS	−1.054***
	(0.231)
Observations	40,043
R^2	0.134
Adjusted R^2	0.017
F Statistic	682.917*** (df = 8; 35270)
Note:	*p<0.1; **p<0.05; ***p<0.01

Question 6

Fixed effects estimation may not be as efficient as random effects estimation, but is robust to correlation between regressors and the random effect. Can we perform a Hausman test in this context? Perform the test you propose.

The test tests the null hypothesis that the unique errors are not correlated with the regressors.

```r
phtest(fixed_effects, random_effects)
```

```
##
##  Hausman Test
##
## data:  formula
## chisq = 54.565, df = 8, p-value = 5.364e-09
## alternative hypothesis: one model is inconsistent
```

The null hypothesis is rejected, implying that the unique parts are correlated with the regressors, and hence, random effects is an inconsistent estimator.

Question 7