FEM 11087 - Applied Microeconometrics

Assignment 2: Panel Data Analysis

Empirical Application

Group 33

Alejandra Betancourt 778686
Yinli Hu 744678
Teun Mulder 565586
Andrés Piñón 775387

30 September 2025

Question 1 [0.7 points]

A central question in labor economics is: ***How much more do individuals earn with higher levels of education?*** *Economists often estimate the* returns to education—*that is, the increase in earnings associated with completing high school, college, or additional years of schooling.*

Using the panel data provided, begin by constructing a ***bar chart*** *showing* ***mean income by education group****. Group individuals based on their* ***highest level of educational attainment*** *(e.g., less than high school, high school graduate, some college, college degree or more), and plot the* ***average income*** *for each category.*

```
gen edyears_cat = .
replace edyears_cat = 1 if edyears <= 11 & !missing(edyears)
replace edyears_cat = 2 if edyears == 12 & !missing(edyears)
replace edyears_cat = 3 if edyears >= 13 & ///
        edyears <= 15 & !missing(edyears)
replace edyears_cat = 4 if edyears >= 16 & !missing(edyears)
```

```
graph bar (mean) income, over(edyears_cat)
graph bar (mean) income, over(edyears_cat) by(male)
```

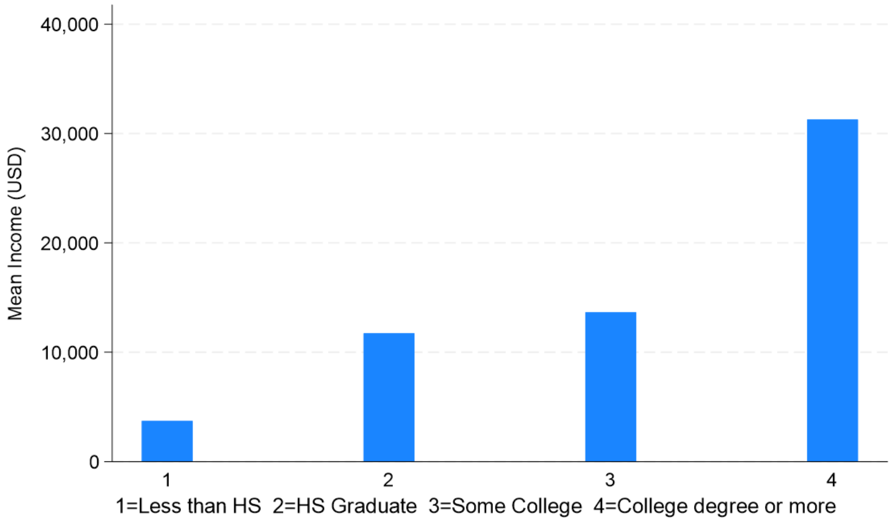

Figure 1.1: Mean income by education level

Recent debates around student debt and the value of higher education often assume that education "pays off" equally for everyone. ***Does your analysis support that assumption?*** *To explore this, create* ***separate plots by gender*** *to highlight any differences in the relationship between education and earnings. Discuss your findings.*

Note: *For this question, create and use a categorical education variable based on each individual's highest level of education completed across the panel. Construct four categories:*

- *Less than high school (11 or fewer years)*
- *High school graduate (exactly 12 years)*
- *Some college (13 to 15 years)*
- *College degree or more (16 or more years)*

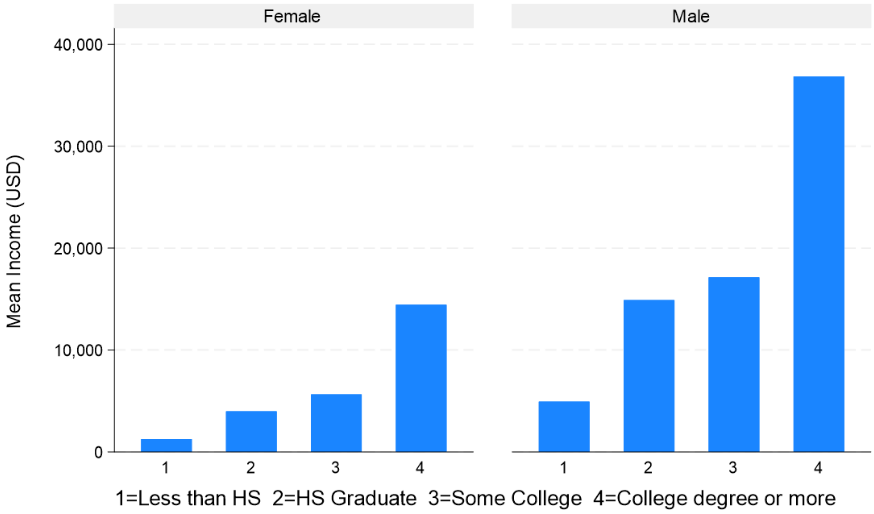

Figure 1.2: Mean income by education level, by gender

Question 2 [1 point]

*Now, we turn to formally estimating the effect of years of education on income. Use **pooled OLS** to examine the impact of years of education (edyears) on **log(income)**, controlling for age, gender (male), marital status categories, ethnicity categories, and childbirth.*

```
sum income, detail
count if income <= 0
```

```
gen log_income = log(income)
reg log_income edyears age i.male ib0.mstatus ib4.ethnicity ///
    i.child_birth
```

Table 2.1: Pooled OLS model

	Coefficient	Robust std. err.	P>|t|	Conf. int.
Treatment variable				
Education years	0.1458	(0.0027)	0.000	[0.140,0.151]
Control variables				
Age	0.1263	(0.0014)	0.000	[0.123,0.129]
Male	0.6000	(0.0126)	0.000	[0.575,0.625]
Childbirth	-0.0795	(0.0184)	0.000	[-0.116,-0.043]
Married	0.6189	(0.0185)	0.000	[0.583,0.655]
Separated or divorced	0.0767	(0.0375)	0.041	[0.003,0.150]
Widowed	0.0647	(0.2531)	0.798	[-0.431,0.561]
Black	-0.4286	(0.0139)	0.000	[-0.456,-0.401]
Hispanic	-0.0825	(0.0146)	0.000	[-0.111,-0.054]
Mixed Race (Non-Hispanic)	-0.3744	(0.0583)	0.000	[-0.489,-0.260]
Constant	3.0156	(0.0319)	0.000	[2.953,3.078]
Number of obs	55874			
F-statistic	3864.88			
Prob > F	0.0000			
R-squared	0.4089			
Adj. R-squared	0.4088			

Dependent variable: log(income)

*a) What is the estimated return to <u>an additional year of education</u>? Interpret the coefficient on years of education in terms of its **sign**, **magnitude**, and **statistical significance**.*

*b) Differences in returns to schooling by gender are sometimes interpreted as potential evidence of **labor market discrimination**. Test whether the effect of years of education using the categorical variable created in Question 1 on log(income) is the **same for men and women**. Based on your results, do you find any evidence consistent with discrimination?*

```
reg log_income i.edyears_cat##i.male age ib0.mstatus ib4.ethnicity ///
    i.child_birth
```

Table 2.2: Pooled OLS model with interaction effects

	Coefficient	Robust std. err.	P>|t|	Conf. int.
Treatment variable				
HS graduate	0.3802	(0.0256)	0.000	[0.330,0.430]
Some college	0.6488	(0.0277)	0.000	[0.594,0.703]
College degree or more	1.0239	(0.0388)	0.000	[0.948,1.100]
HS graduate × Male	0.3836	(0.0302)	0.000	[0.324,0.443]
Some college × Male	0.1664	(0.0326)	0.000	[0.102,0.230]
College degree or more × Male	0.2206	(0.0432)	0.000	[0.136,0.305]
Control variables				
Age	0.1225	(0.0014)	0.000	[0.120,0.125]
Male	0.4472	(0.0188)	0.000	[0.410,0.484]
Childbirth	-0.0759	(0.0182)	0.000	[-0.112,-0.040]
Married	0.6013	(0.0183)	0.000	[0.565,0.637]
Separated or divorced	0.0631	(0.0372)	0.090	[-0.010,0.136]
Widowed	0.0932	(0.2509)	0.710	[-0.399,0.585]
Black	-0.4143	(0.0138)	0.000	[-0.441,-0.387]
Hispanic	-0.0732	(0.0145)	0.000	[-0.102,-0.045]
Mixed Race (Non-Hispanic)	-0.3752	(0.0578)	0.000	[-0.488,-0.262]
Constant	4.4713	(0.0289)	0.000	[4.415,4.528]
Number of obs	55874			
F-statistic	2690.54			
Prob > F	0.0000			
R-squared	0.4195			
Adj. R-squared	0.4193			

Dependent variable: log(income)

```
test 2.edyears_cat#1.male 3.edyears_cat#1.male 4.edyears_cat#1.male
```

Table 2.3: Joint F-test

F-statistic	54.85
Prob > F	0.0000

(1) HS graduate × Male = 0
(2) Some college × Male = 0
(3) College degree or more × Male = 0

c) Under what conditions is the pooled OLS estimate of the effect of years of education ***unbiased and efficient****? Do you believe these conditions are likely to hold in this context?*

Question 3 [0.5 points]

So far, the panel structure of the data has been largely unexploited. Random effects (RE) estimation can improve the efficiency of the estimates compared to pooled OLS.

a) Estimate the effect of years of education (edyears) on ***log(income)*** *using the* ***random effects*** *(RE) model, controlling for age, gender (male), marital status categories, ethnicity categories, and childbirth. Interpret the estimated coefficient for years of education in terms of its* ***sign***, ***magnitude***, *and* ***statistical significance***. *Then, compare the RE estimate and standard error of the education coefficient with those obtained from the* ***pooled OLS model***.

```
xtset pid wave
```

```
xtreg log_income edyears age i.male ib0.mstatus ib4.ethnicity ///
    i.child_birth, re
estimates store random
```

Table 3.1: Random effects (RE) model

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Education years	0.1509	(0.0031)	0.000	[0.145,0.157]
Control variables				
Age	0.1273	(0.0015)	0.000	[0.124,0.130]
Male	0.5698	(0.0167)	0.000	[0.537,0.603]
Childbirth	-0.0599	(0.0174)	0.001	[-0.094,-0.026]
Married	0.5107	(0.0197)	0.000	[0.472,0.549]
Separated or divorced	0.0454	(0.0392)	0.246	[-0.031,0.122]
Widowed	0.0855	(0.2619)	0.744	[-0.428,0.599]
Black	-0.3833	(0.0192)	0.000	[-0.421,-0.346]
Hispanic	-0.0655	(0.0203)	0.001	[-0.105,-0.026]
Mixed Race (Non-Hispanic)	-0.3244	(0.0813)	0.000	[-0.484,-0.165]
Constant	2.9423	(0.0341)	0.000	[2.875,3.009]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3380			
R-squared between	0.5546			
R-squared overall	0.4084			
Wald χ^2	34633.21			
Prob > χ^2	0.0000			
σ_u	0.4412			
σ_e	1.1766			
ρ	0.1233			

Dependent variable: log(income)

Table 3.2: POLS, RE comparison

	POLS	RE
Treatment variable		
Education years	0.1458***	0.1509***
	(0.0027)	(0.0031)
Control variables		
Age	0.1263***	0.1273***
	(0.0014)	(0.0015)
Male	0.6000***	0.5698***
	(0.0126)	(0.0167)
Childbirth	-0.0795***	-0.0599***
	(0.0184)	(0.0174)
Black	-0.4286***	-0.3833***
	(0.0139)	(0.0192)
Hispanic	-0.0825***	-0.0655**
	(0.0146)	(0.0203)
Mixed Race (Non-Hispanic)	-0.3744***	-0.3244***
	(0.0583)	(0.0813)
Married	0.6189***	0.5107***
	(0.0185)	(0.0197)
Separated or divorced	0.0767*	0.0454
	(0.0375)	(0.0392)
Widowed	0.0647	0.0855
	(0.2531)	(0.2619)
Constant	3.0156***	2.9423***
	(0.0319)	(0.0341)
Number of obs	55874	55874
Number of groups		7126
F-statistic	3864.88	
Wald χ^2		34633.21
P-value	0.0000	0.0000
R-squared	0.4089	
Adj. R-squared	0.4088	
R-squared within		0.3380
R-squared between		0.5546
R-squared overall		0.4084
σ_u		0.4412
σ_e		1.1766
ρ		0.1233

Standard errors in parentheses.
* p < 0.05, ** p < 0.01, *** p < 0.001

b) Under which conditions and why can the random effects estimator be ***more efficient*** *than pooled OLS?*

Question 4 [1.55 points]

*Alternatively, the panel structure of the data can be used to perform **fixed effects (FE)** estimation.*

*a) Based on theoretical considerations, would you **prefer** fixed effects or random effects estimation? Justify your answer.*

*b) Use a **fixed effects estimator** to examine the impact of years of education (edyears) on **log(income)**, controlling for age, gender (male), marital status categories, ethnicity categories, and childbirth. Interpret the coefficient on years of education in terms of its **sign**, **magnitude**, and **statistical significance**. Compare your results with those from the **pooled OLS** and **random effects** models.*

```
xtreg log_income edyears age i.male ib0.mstatus ib4.    ethnicity ///
    i.child_birth, fe
estimates store fixed
```

Table 4.1: Fixed effects (FE) model

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Education years	0.1610	(0.0038)	0.000	[0.153,0.168]
Control variables				
Age	0.1259	(0.0017)	0.000	[0.123,0.129]
Male	0.0000	(.)	.	[0.000,0.000]
Childbirth	-0.0545	(0.0180)	0.002	[-0.090,-0.019]
Married	0.4178	(0.0223)	0.000	[0.374,0.461]
Separated or divorced	-0.0112	(0.0432)	0.796	[-0.096,0.073]
Widowed	0.1575	(0.2861)	0.582	[-0.403,0.718]
Black	0.0000	(.)	.	[0.000,0.000]
Hispanic	0.0000	(.)	.	[0.000,0.000]
Mixed Race (Non-Hispanic)	0.0000	(.)	.	[0.000,0.000]
Constant	3.1536	(0.0378)	0.000	[3.080,3.228]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3383			
R-squared between	0.4945			
R-squared overall	0.3713			
Wald χ^2				
Prob > χ^2	0.0000			
σ_u	0.8174			
σ_e	1.1766			
ρ	0.3255			

Dependent variable: log(income)

Table 4.2: POLS, RE, FE comparison

	POLS	RE	FE
Treatment variable			
Education years	0.1458***	0.1509***	0.1610***
	(0.0027)	(0.0031)	(0.0038)
Control variables			
Age	0.1263***	0.1273***	0.1259***
	(0.0014)	(0.0015)	(0.0017)
Male	0.6000***	0.5698***	0.0000
	(0.0126)	(0.0167)	(.)
Childbirth	-0.0795***	-0.0599***	-0.0545**
	(0.0184)	(0.0174)	(0.0180)
Black	-0.4286***	-0.3833***	0.0000
	(0.0139)	(0.0192)	(.)
Hispanic	-0.0825***	-0.0655**	0.0000
	(0.0146)	(0.0203)	(.)
Mixed Race (Non-Hispanic)	-0.3744***	-0.3244***	0.0000
	(0.0583)	(0.0813)	(.)
Married	0.6189***	0.5107***	0.4178***
	(0.0185)	(0.0197)	(0.0223)
Separated or divorced	0.0767*	0.0454	-0.0112
	(0.0375)	(0.0392)	(0.0432)
Widowed	0.0647	0.0855	0.1575
	(0.2531)	(0.2619)	(0.2861)
Constant	3.0156***	2.9423***	3.1536***
	(0.0319)	(0.0341)	(0.0378)
Number of obs	55874	55874	55874
Number of groups		7126	7126
F-statistic	3864.88		4154.09
Wald χ^2		34633.21	
P-value	0.0000	0.0000	0.0000
R-squared	0.4089		0.3383
Adj. R-squared	0.4088		0.2415
R-squared within		0.3380	0.3383
R-squared between		0.5546	0.4945
R-squared overall		0.4084	0.3713
σ_u		0.4412	0.8174
σ_e		1.1766	1.1766
ρ		0.1233	0.3255

Standard errors in parentheses.
* p < 0.05, ** p < 0.01, *** p < 0.001

c) Perform the ***Hausman*** *test. What do the results indicate? Based on the test outcome,* ***which estimator*** *(RE or FE) is more appropriate in this context?*

```
hausman fixed random
```

Table 4.3: Hausman test

χ^2	141.94
Prob > χ^2	0.0000

H0: Difference in β not systematic

Question 5 [0.9 points]

*Next, estimate a **Correlated Random Effects (CRE)** model to examine the effect of years of education (edyears) on **log(income)**.*

```
by pid: egen age_mean = mean(age)
by pid: egen mstatus_mean = mean(mstatus)
```

```
xtreg log_income edyears age i.male ib0.mstatus ib4.ethnicity ///
    i.child_birth age_mean mstatus_mean, re
```

Table 5.1: Correlated Random Effects (CRE) model

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Education years	0.1505	(0.0031)	0.000	[0.144,0.156]
Control variables				
Age	0.1282	(0.0016)	0.000	[0.125,0.131]
Male	0.5600	(0.0178)	0.000	[0.525,0.595]
Childbirth	-0.0617	(0.0174)	0.000	[-0.096,-0.028]
Married	0.4463	(0.0221)	0.000	[0.403,0.490]
Separated or divorced	-0.0669	(0.0430)	0.119	[-0.151,0.017]
Widowed	-0.0927	(0.2632)	0.725	[-0.609,0.423]
Black	-0.3703	(0.0194)	0.000	[-0.408,-0.332]
Hispanic	-0.0684	(0.0203)	0.001	[-0.108,-0.029]
Mixed Race (Non-Hispanic)	-0.3117	(0.0813)	0.000	[-0.471,-0.152]
CRE variables				
age_mean	0.0013	(0.0032)	0.676	[-0.005,0.008]
mstatus_mean	0.2142	(0.0340)	0.000	[0.148,0.281]
Constant	2.8795	(0.0581)	0.000	[2.766,2.993]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3382			
R-squared between	0.5565			
R-squared overall	0.4093			
Wald χ^2	34714.61			
Prob > χ^2	0.0000			
σ_u	0.4412			
σ_e	1.1766			
ρ	0.1233			

Dependent variable: log(income)

*a) What is one advantage of the **CRE estimator** compared to the **random effects (RE)** estimator?*

*b) What is one advantage of the **CRE estimator** compared to the **fixed effects (FE)** estimator?*

*c) Compare the estimated coefficient for years of education from the **CRE model** with those from the **RE** and **FE** models. Are the coefficients similar or different? Explain why this is the case.*

Table 5.2: POLS, RE, FE, CRE comparison

	POLS	RE	FE	CRE
Treatment variable				
Education years	0.1458***	0.1509***	0.1610***	0.1505***
	(0.0027)	(0.0031)	(0.0038)	(0.0031)
Control variables				
Age	0.1263***	0.1273***	0.1259***	0.1282***
	(0.0014)	(0.0015)	(0.0017)	(0.0016)
Male	0.6000***	0.5698***	0.0000	0.5600***
	(0.0126)	(0.0167)	(.)	(0.0178)
Childbirth	-0.0795***	-0.0599***	-0.0545**	-0.0617***
	(0.0184)	(0.0174)	(0.0180)	(0.0174)
Black	-0.4286***	-0.3833***	0.0000	-0.3703***
	(0.0139)	(0.0192)	(.)	(0.0194)
Hispanic	-0.0825***	-0.0655**	0.0000	-0.0684***
	(0.0146)	(0.0203)	(.)	(0.0203)
Mixed Race (Non-Hispanic)	-0.3744***	-0.3244***	0.0000	-0.3117***
	(0.0583)	(0.0813)	(.)	(0.0813)
Married	0.6189***	0.5107***	0.4178***	0.4463***
	(0.0185)	(0.0197)	(0.0223)	(0.0221)
Separated or divorced	0.0767*	0.0454	-0.0112	-0.0669
	(0.0375)	(0.0392)	(0.0432)	(0.0430)
Widowed	0.0647	0.0855	0.1575	-0.0927
	(0.2531)	(0.2619)	(0.2861)	(0.2632)
CRE variables				
age_mean				0.0013
				(0.0032)
mstatus_mean				0.2142***
				(0.0340)
Constant	3.0156***	2.9423***	3.1536***	2.8795***
	(0.0319)	(0.0341)	(0.0378)	(0.0581)
Number of obs	55874	55874	55874	55874
Number of groups		7126	7126	7126
F-statistic	3864.88		4154.09	
Wald χ^2		34633.21		34714.61
P-value	0.0000	0.0000	0.0000	0.0000
R-squared	0.4089		0.3383	
Adj. R-squared	0.4088		0.2415	
R-squared within		0.3380	0.3383	0.3382
R-squared between		0.5546	0.4945	0.5565
R-squared overall		0.4084	0.3713	0.4093
σ_u		0.4412	0.8174	0.4412
σ_e		1.1766	1.1766	1.1766
ρ		0.1233	0.3255	0.1233

Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

*d) Based on your CRE estimates, does the assumption of **exogeneity** appear to hold? Which estimator would you consider most appropriate in this context?*

Question 6 [0.9 points]

Recent research provides compelling evidence that after the birth of a first child, women's earnings decline sharply and remain persistently lower, while men's earnings remain largely unaffected.

*a) Estimate the effect of childbirth on **log(income)** using the **most appropriate model**. Control for age, gender (male), marital status categories, ethnicity categories, and years of education (edyears). Interpret the estimated coefficient for childbirth in terms of its **sign**, **magnitude**, and **statistical significance**.*

```
xtreg log_income i.child_birth age i.male ib0.mstatus ib4.ethnicity ///
    edyears age_mean mstatus_mean, re
```

Table 6.1: Correlated Random Effects (CRE)

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Childbirth	-0.0617	(0.0174)	0.000	[-0.096,-0.028]
Control variables				
Age	0.1282	(0.0016)	0.000	[0.125,0.131]
Male	0.5600	(0.0178)	0.000	[0.525,0.595]
Education years	0.1505	(0.0031)	0.000	[0.144,0.156]
Married	0.4463	(0.0221)	0.000	[0.403,0.490]
Separated or divorced	-0.0669	(0.0430)	0.119	[-0.151,0.017]
Widowed	-0.0927	(0.2632)	0.725	[-0.609,0.423]
Black	-0.3703	(0.0194)	0.000	[-0.408,-0.332]
Hispanic	-0.0684	(0.0203)	0.001	[-0.108,-0.029]
Mixed Race (Non-Hispanic)	-0.3117	(0.0813)	0.000	[-0.471,-0.152]
CRE variables				
age_mean	0.0013	(0.0032)	0.676	[-0.005,0.008]
mstatus_mean	0.2142	(0.0340)	0.000	[0.148,0.281]
Constant	2.8795	(0.0581)	0.000	[2.766,2.993]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3382			
R-squared between	0.5565			
R-squared overall	0.4093			
Wald χ^2	34714.61			
Prob > χ^2	0.0000			
σ_u	0.4412			
σ_e	1.1766			
ρ	0.1233			

Dependent variable: log(income)

b) Test whether the effect of childbirth on log(income) ***differs*** *between males and females. What conclusions can you draw from your results?*

```
xtreg log_income i.child_birth##i.male age ib0.mstatus ib4.ethnicity ///
    edyears age_mean mstatus_mean, re
```

Table 6.2: Correlated Random Effects (CRE) with interaction effects

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Childbirth	-0.3390	(0.0312)	0.000	[-0.400,-0.278]
Childbirth × Male	0.3959	(0.0370)	0.000	[0.323,0.468]
Control variables				
Age	0.1272	(0.0016)	0.000	[0.124,0.130]
Male	0.5165	(0.0183)	0.000	[0.481,0.552]
Education years	0.1512	(0.0031)	0.000	[0.145,0.157]
Married	0.4458	(0.0221)	0.000	[0.403,0.489]
Separated or divorced	-0.0692	(0.0429)	0.107	[-0.153,0.015]
Widowed	-0.0878	(0.2629)	0.739	[-0.603,0.428]
Black	-0.3674	(0.0194)	0.000	[-0.405,-0.329]
Hispanic	-0.0680	(0.0203)	0.001	[-0.108,-0.028]
Mixed Race (Non-Hispanic)	-0.3043	(0.0813)	0.000	[-0.464,-0.145]
CRE variables				
age_mean	0.0021	(0.0032)	0.515	[-0.004,0.008]
mstatus_mean	0.2211	(0.0340)	0.000	[0.154,0.288]
Constant	2.9032	(0.0582)	0.000	[2.789,3.017]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3396			
R-squared between	0.5563			
R-squared overall	0.4106			
Wald χ^2	34887.51			
Prob > χ^2	0.0000			
σ_u	0.4418			
σ_e	1.1753			
ρ	0.1238			

Dependent variable: log(income)

```
test 1.child_birth#1.male
```

Table 6.3: Single coeff. test

χ^2	114.41
Prob > χ^2	0.0000

(1) Childbirth × Male = 0

<u>Question 7 [1.2 points]</u>

Without conducting any empirical analysis:

*a) Compare the key assumptions underlying **pooled OLS**, **fixed effects (FE)**, and **random effects (RE)** estimators. Discuss theoretically in which scenarios you would prefer to use each method.*

*b) Within the practical context of this assignment (effect of education on earnings), provide an example situation for each estimator in the form of a **Directed Acyclic Graph (DAG)**. For each case (Pooled OLS, FE, and RE), explain why the assumptions required for the respective method hold in that example, and why that method would be preferred.*

Question 8 [0.75 points]

*Finally, revisit your data and evaluate whether **attrition** is present in your sample. Based on your preferred model, discuss the likelihood of **attrition bias**. What conclusions can you draw regarding its presence, and how might it affect the validity of your results?*

```
bysort pid (wave): gen n_waves = _N
gen all_waves = n_waves == 17
```

```
xtreg log_income i.child_birth##i.male age ib0.mstatus ib4.ethnicity ///
    edyears age_mean mstatus_mean all_waves, re
```

Table 8.1: Attrition bias: *all waves* indicator

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Childbirth	-0.3376	(0.0312)	0.000	[-0.399,-0.276]
Childbirth × Male	0.3971	(0.0370)	0.000	[0.325,0.470]
Control variables				
Age	0.1272	(0.0016)	0.000	[0.124,0.130]
Male	0.5146	(0.0183)	0.000	[0.479,0.550]
Education years	0.1511	(0.0031)	0.000	[0.145,0.157]
Married	0.4451	(0.0221)	0.000	[0.402,0.488]
Separated or divorced	-0.0687	(0.0429)	0.110	[-0.153,0.015]
Widowed	-0.0913	(0.2629)	0.728	[-0.607,0.424]
Black	-0.3660	(0.0194)	0.000	[-0.404,-0.328]
Hispanic	-0.0673	(0.0203)	0.001	[-0.107,-0.027]
Mixed Race (Non-Hispanic)	-0.3058	(0.0813)	0.000	[-0.465,-0.147]
CRE variables				
age_mean	-0.0001	(0.0033)	0.988	[-0.007,0.006]
mstatus_mean	0.2220	(0.0340)	0.000	[0.155,0.289]
Bias indicator				
all_waves	0.0586	(0.0252)	0.020	[0.009,0.108]
Constant	2.9418	(0.0605)	0.000	[2.823,3.060]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3396			
R-squared between	0.5565			
R-squared overall	0.4108			
Wald χ^2	34898.11			
Prob > χ^2	0.0000			
σ_u	0.4416			
σ_e	1.1753			
ρ	0.1237			

Dependent variable: log(income)

```
bysort pid (wave): gen next_wave = (wave[_n+1] == wave + 1)
```

```
xtreg log_income i.child_birth##i.male age ib0.mstatus ib4.ethnicity ///
    edyears age_mean mstatus_mean next_wave, re
```

Table 8.2: Attrition bias: *next wave* indicator

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Childbirth	-0.3361	(0.0312)	0.000	[-0.397,-0.275]
Childbirth × Male	0.3941	(0.0370)	0.000	[0.322,0.467]
Control variables				
Age	0.1290	(0.0016)	0.000	[0.126,0.132]
Male	0.5165	(0.0183)	0.000	[0.481,0.552]
Education years	0.1509	(0.0031)	0.000	[0.145,0.157]
Married	0.4435	(0.0221)	0.000	[0.400,0.487]
Separated or divorced	-0.0704	(0.0429)	0.101	[-0.155,0.014]
Widowed	-0.0932	(0.2629)	0.723	[-0.608,0.422]
Black	-0.3668	(0.0194)	0.000	[-0.405,-0.329]
Hispanic	-0.0679	(0.0203)	0.001	[-0.108,-0.028]
Mixed Race (Non-Hispanic)	-0.3061	(0.0812)	0.000	[-0.465,-0.147]
CRE variables				
age_mean	-0.0014	(0.0033)	0.672	[-0.008,0.005]
mstatus_mean	0.2226	(0.0340)	0.000	[0.156,0.289]
Bias indicator				
next_wave	0.0641	(0.0139)	0.000	[0.037,0.091]
Constant	2.8931	(0.0582)	0.000	[2.779,3.007]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3398			
R-squared between	0.5567			
R-squared overall	0.4109			
Wald χ^2	34925.32			
Prob > χ^2	0.0000			
σ_u	0.4414			
σ_e	1.1752			
ρ	0.1236			

Dependent variable: log(income)

```
xtreg log_income i.child_birth##i.male age ib0.mstatus ib4.ethnicity ///
    edyears age_mean mstatus_mean n_waves, re
```

Table 8.3: Attrition bias: *number of waves* indicator

	Coefficient	Std. err.	P>|t|	Conf. int.
Treatment variable				
Childbirth	-0.3358	(0.0312)	0.000	[-0.397,-0.275]
Childbirth × Male	0.3979	(0.0370)	0.000	[0.325,0.470]
Control variables				
Age	0.1272	(0.0016)	0.000	[0.124,0.130]
Male	0.5103	(0.0185)	0.000	[0.474,0.546]
Education years	0.1510	(0.0031)	0.000	[0.145,0.157]
Married	0.4451	(0.0221)	0.000	[0.402,0.488]
Separated or divorced	-0.0699	(0.0429)	0.104	[-0.154,0.014]
Widowed	-0.0865	(0.2629)	0.742	[-0.602,0.429]
Black	-0.3666	(0.0194)	0.000	[-0.405,-0.329]
Hispanic	-0.0676	(0.0203)	0.001	[-0.107,-0.028]
Mixed Race (Non-Hispanic)	-0.3058	(0.0812)	0.000	[-0.465,-0.147]
CRE variables				
age_mean	-0.0058	(0.0045)	0.197	[-0.015,0.003]
mstatus_mean	0.2210	(0.0340)	0.000	[0.154,0.288]
Bias indicator				
n_waves	0.0065	(0.0027)	0.014	[0.001,0.012]
Constant	3.0107	(0.0728)	0.000	[2.868,3.153]
Number of obs	55874			
Number of groups	7126			
R-squared within	0.3396			
R-squared between	0.5567			
R-squared overall	0.4108			
Wald χ^2	34902.86			
Prob > χ^2	0.0000			
σ_u	0.4412			
σ_e	1.1753			
ρ	0.1235			

Dependent variable: log(income)