Yi Ma

Question 1

1. Calculate the mean log wage for each year in the data set

1980 1981 1982 1983 1984 1985 1986 1987

1.393477 1.512867 1.571667 1.619263 1.690295 1.739410 1.799719 1.866479

1. Regress log(wage) on a constant and a dummy variable for each year

Call:

lm(formula = lwage ~ d81 + d82 + d83 + d84 + d85 + d86 + d87,

data = wagepan)

Residuals:

Min 1Q Median 3Q Max

-5.2694 -0.2820 0.0266 0.3231 2.3124

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.39348 0.02196 63.462 < 2e-16 \*\*\*

d81 0.11939 0.03105 3.845 0.000122 \*\*\*

d82 0.17819 0.03105 5.738 1.02e-08 \*\*\*

d83 0.22579 0.03105 7.271 4.21e-13 \*\*\*

d84 0.29682 0.03105 9.558 < 2e-16 \*\*\*

d85 0.34593 0.03105 11.140 < 2e-16 \*\*\*

d86 0.40624 0.03105 13.082 < 2e-16 \*\*\*

d87 0.47300 0.03105 15.232 < 2e-16 \*\*\*

1. Regress log(wage) on a full set of dummy variables for each year

Call:

lm(formula = lwage ~ d80 + d81 + d82 + d83 + d84 + d85 + d86 +

d87 + 0, data = wagepan)

Residuals:

Min 1Q Median 3Q Max

-5.2694 -0.2820 0.0266 0.3231 2.3124

Coefficients:

Estimate Std. Error t value Pr(>|t|)

d80 1.39348 0.02196 63.46 <2e-16 \*\*\*

d81 1.51287 0.02196 68.90 <2e-16 \*\*\*

d82 1.57167 0.02196 71.58 <2e-16 \*\*\*

d83 1.61926 0.02196 73.75 <2e-16 \*\*\*

d84 1.69029 0.02196 76.98 <2e-16 \*\*\*

d85 1.73941 0.02196 79.22 <2e-16 \*\*\*

d86 1.79972 0.02196 81.96 <2e-16 \*\*\*

d87 1.86648 0.02196 85.00 <2e-16 \*\*\*

1. Comment on the relationship among your estimates from the first three parts

The result for part a and part c are identical because what OLS does essentially is the conditional mean of Y given X. Result for part b shows the marginal effect and the intercept represents the mean for year 80. If we add up the intercept and the corresponding coefficient, we would get the same mean value as in part a and part c.

1. Estimate this equation by pooled OLS

Pooling Model

Call:

plm(formula = lwage ~ educ + black + hisp + exper + expersq +

married + union, data = wagepan, model = "pooling",

index = c("nr", "year"))

Balanced Panel: n = 545, T = 8, N = 4360

Residuals:

Min. 1st Qu. Median 3rd Qu. Max.

-5.268937 -0.248691 0.033205 0.296163 2.560777

Coefficients:

Estimate Std. Error t-value Pr(>|t|)

(Intercept) -0.03470569 0.06456900 -0.5375 0.5910

educ 0.09938779 0.00467760 21.2476 < 2.2e-16 \*\*\*

black -0.14384171 0.02355950 -6.1055 1.114e-09 \*\*\*

hisp 0.01569798 0.02081119 0.7543 0.4507

exper 0.08917907 0.01011105 8.8200 < 2.2e-16 \*\*\*

expersq -0.00284866 0.00070736 -4.0272 5.742e-05 \*\*\*

married 0.10766558 0.01569647 6.8592 7.897e-12 \*\*\*

union 0.18007257 0.01712053 10.5179 < 2.2e-16 \*\*\*

It is not reliable because we don’t know if Ci is correlated with the observed covariates. Even if we assume that Ci is uncorrelated with the covariates, errors will still be correlated with through the individual-specific unobserved heterogeneity Ci. Therefore, use of robust standard errors and use the robust variance-covariance matrix in testing hypothesis is compulsory.

POLS with robust standard errors.

t test of coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.03470569 0.12000704 -0.2892 0.772444

educ 0.09938779 0.00920092 10.8019 < 2.2e-16 \*\*\*

black -0.14384171 0.05007130 -2.8727 0.004089 \*\*

hisp 0.01569798 0.03916655 0.4008 0.688587

exper 0.08917907 0.01243303 7.1728 8.597e-13 \*\*\*

expersq -0.00284866 0.00086989 -3.2747 0.001066 \*\*

married 0.10766558 0.02606010 4.1314 3.673e-05 \*\*\*

union 0.18007257 0.02755815 6.5343 7.127e-11 \*\*\*

1. Estimate the equation by FE.

Call:

plm(formula = lwage ~ educ + black + hisp + exper + expersq +

married + union, data = wagepan, model = "within",

index = c("nr", "year"))

Balanced Panel: n = 545, T = 8, N = 4360

Residuals:

Min. 1st Qu. Median 3rd Qu. Max.

-4.1726214 -0.1257010 0.0092527 0.1595770 1.4701690

Coefficients:

Estimate Std. Error t-value Pr(>|t|)

exper 0.11684669 0.00841968 13.8778 < 2.2e-16 \*\*\*

expersq -0.00430089 0.00060527 -7.1057 1.422e-12 \*\*\*

married 0.04530332 0.01830968 2.4743 0.01339 \*

union 0.08208713 0.01929073 4.2553 2.138e-05 \*\*\*

FE can’t estimate variable that doesn’t change overtime. In FE we allow for correlation between the x and unobserved time fixed variable c and every individual in the sample work continuously for the period of 7 years which makes exper undistinguishable from c. however it is still statistically significant.

1. Now add the interactions of the form

Oneway (individual) effect Within Model

Call:

plm(formula = lwage ~ educ + d81educ + d82educ + d83educ + d84educ +

d85educ + d86educ + d87educ + black + hisp + exper + expersq +

married + union, data = wagepan, model = "within",

index = c("nr", "year"))

Balanced Panel: n = 545, T = 8, N = 4360

Residuals:

Min. 1st Qu. Median 3rd Qu. Max.

-4.155670 -0.122980 0.010886 0.155469 1.490770

Coefficients:

Estimate Std. Error t-value Pr(>|t|)

d81educ -0.00104501 0.00260662 -0.4009 0.688512

d82educ -0.00624542 0.00410232 -1.5224 0.127989

d83educ -0.01136452 0.00568061 -2.0006 0.045509 \*

d84educ -0.01356809 0.00722194 -1.8787 0.060357 .

d85educ -0.01616558 0.00870137 -1.8578 0.063272 .

d86educ -0.01699123 0.01011152 -1.6804 0.092965 .

d87educ -0.01674212 0.01145253 -1.4619 0.143859

exper 0.17048411 0.02729410 6.2462 4.669e-10 \*\*\*

expersq -0.00596719 0.00085759 -6.9581 4.044e-12 \*\*\*

married 0.04747240 0.01831126 2.5925 0.009564 \*\*

union 0.07941945 0.01930557 4.1138 3.974e-05 \*\*\*

No. It decreased overtime.

Question 2

1. Explain why only one of D1 or D2 can be included in the population model

Because inclusion of D1 and D2 would lead to multicollinearity (e.g. D1 + D2 = const). This is a violation of the rank condition for identification. The rank condition for identification states that X has to be a full-rank matrix. In R, we calculated that the rank of X is 4 which implies that it is not a full-rank matrix, meaning that we have to drop either D1 and D2.

1. Calculate the within transformation.

Calculation done in R, the transformed matrix X is

[,1] [,2] [,3] [,4] [,5]

[1,] 0 0.5 -0.5 -0.5 -2.5

[2,] 0 -0.5 0.5 0.5 2.5

[3,] 0 0.5 -0.5 -0.5 -3.5

[4,] 0 -0.5 0.5 0.5 3.5

[5,] 0 0.5 -0.5 -0.5 -6.5

[6,] 0 -0.5 0.5 0.5 6.5

1. Which of these four variables (const, D2, exp, exp^2) can be included in a FE.

FE.2 states that  and according to the calculation in R, is

[const] [D1] [D2] [exp] [exp^2]

[1,] 0 0.5 -0.5 -0.5 -2.5

[2,] 0 -0.5 0.5 0.5 2.5

[3,] 0 0.5 -0.5 -0.5 -3.5

[4,] 0 -0.5 0.5 0.5 3.5

[5,] 0 0.5 -0.5 -0.5 -6.5

[6,] 0 -0.5 0.5 0.5 6.5

Because we have a population model, we can remove the expectation. According to the calculation done in R, the rank of the transformed X is 2, which means that we can only include 2 variables in our FE regression. Since the constant is canceled out during the within transformation process, and D2 and exp are linearly dependent (they are identical). Therefore, we only keep either D2 or exp and exp^2 in our FE regression.

1. Calculate the difference transformation.

After the difference transformation, dX becomes

[,1] [,2] [,3] [,4] [,5]

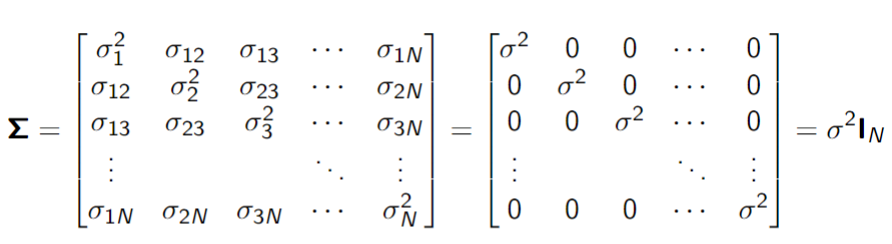
[1,] 0 -1 1 1 5

[2,] 0 -1 1 1 7

[3,] 0 -1 1 1 13

1. Write down the variance-covariance matrix (Vit uncorrelated with the covariates)

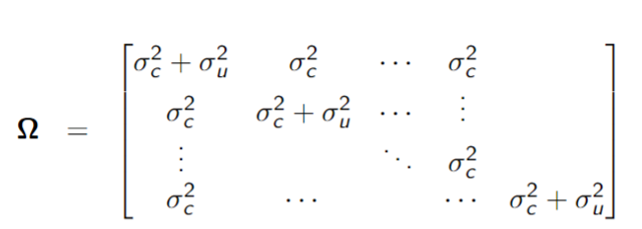
Vit = Ci + Uit and since it is assumed that the population model has Vit uncorrelated with the observed covariates, we know that POLS will be consistent. The variance-covariance matrix under OLS.3 is highly restrictive.



Where n = 6 in this case.

1. Write down the variance-covariance matrix (the RE assumptions hold)

For RE, we combine the assumptions of FE and POLS. Under these assumptions, we have



Where the variance-covariance matrix is a 6x6 matrix.