Text

Description automatically generated

**Assessment and Feedback: Student Template**

**Student ID Number(s):** 2382208

**Programme:** M.Sc (Taught) economics

**Module:** Econometrics

**Name of Tutor:** Marco Barassi

**Assignment Title:** Computer project

**Date and Time of Submission:** 01 December 2021 12:00

**Actual Word Count:** 2000

**Extension:** Y / N \* **Extension Due Date:** N

I do / do not \* wish my assignment to be considered for including as an exemplar in the **School Bank of Assessed Work.** **\*** *delete as appropriate*

|  |
| --- |
| **The purpose of this template is to ensure you receive targeted feedback that will support your learning. It is a requirement to complete to complete all 3 sections, and to include this completed template as the first page of every assignment that is submitted for marking (your School will advise on exceptions).** |
|  |
| **Section One:** Reflecting on the feedback that I have received on previous assessments, the following issues/topics have been identified as areas for improvement: (add 3 bullet points). *NB – for first year students/PGTs in the first term, this refers to assessments in your previous institution*   * More coercive in my writing and more engaged in the topic of the assessment. * Use a wide variety of literature to make my point more valid * Proofread my work before submitting it. |
| **Section Two:** In this assignment, I have attempted to act on previous feedback in the following ways (3 bullet points)   * I made sure my writing was in line with assessment and I made sure I covered all the required by my lecturer. * I applied the knowledge gathered from the lecture notes and the microeconomic book to explain my results. * I proofread my work before submitting it |
| **Section Three:** Feedback on the following aspects of this assignment (i.e. content/style/approach) would be particularly helpful to me: (3 bullet points)   * Feedback on my content will be very helpful for me going forward as it will help me with my other assessment. * I would like an assessment on how my work style of work so I can be able to improve further on how deliver my assessment. * The last feedback I want is on my approach because is not something I’m confident in and I would like to improve it. |

**By submitting your work online you are confirming that your work is your own and that you understand and have read the University’s rules regarding plagiarism and the consequences that will arise should you submit plagiarised work.**

**Please ensure that you complete and attach this template to the front of all work that is submitted.**

In this assignment i will be interpretating the data lforce.dta to derive an empirical strategy to estimate the returns to education and I will also test whether there is any difference in returns to education for women with children.

Introduction

The following report I will estimate the return of education of the data lforce.dta. The data provided to us is a cross sectional data in which time is not a factor meaning there can’t be autocorrelation. Cross-sectional data, or a cross section of a study population, in statistics and [econometrics](https://www.bing.com/search?q=Econometrics&filters=sid%3a2e4a00be-b361-e498-3403-99ff7aa852a9&form=ENTLNK) is **a type of data collected by observing many subjects (such as individuals, firms, countries, or regions) at the one point or period of time.** I will perform a regression on the independent variable in respective to the dependent variable. I chose the independent variable education and experience due to the mincer earnings function which I will explain in the next paragraph but variable such as age were omitted due to collinearity. The variables are best variables to get the most consistent and unbiased result.

Literature review

The Mincer earnings function is a single equation model that explains wage income as a function of schooling and experience, named after Jacob Mincer. This equation helps the determine the returns to education in respect to certain independent variables. The equation makes sure to take into account various possibility such as experience, age and city ecc… The result provided will follow this literature review and will provide accurate results.

The methodology I used to interpret the data lforce.dta was through the mincer earning function. The mincer earning function is a single equation model that explains earnings as a function of schooling and experience. (Polachek, 2007) I will regress the dependent variable wages in respect to the independent variable’s education, experience and experience squared subsequently I will formulate the estimate to return to education by utilising a dummy variable. The following equation shows my regression equation:

lnWage =lny0 + β1Educi + β2Xi +3Xi2 + β4Zi ui

Figure 1 -

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Ln wage** | **Coefficient** | **Standard error** | **Prob>F** | **Prob>T** | **Standard deviation** |
| Educ | .1363014 | .024764 | 0.0000 | 0.000 | 2.139828 |
| Exp | .0984937 | .035355 | 0.0000 | 0.006 | 5.377247 |
| Exp squared | -.0033068 | .0016835 | 0.0000 | 0.051 | 104.4187 |

From the table above we can infer that all independent variables are significant has their p value for the f-test stands, as their below the 0.05 threshold however in the case for experince squared it seems that p value for the test is slightly above eth significance level of 0.05 this means it is not significant. The coefficient for education is positive this means that there’s a positive correlation between educ and Ln wage as demonstrated by (Rodríguez-Pose and Tselios, 2009) which states there’s a positive correlation between education and wage similarly the same concept can be applied to exp squared. The variable defined as exp squared is negative because experience has a diminishing marginal returns which means that has experiences increases it increases at a diminishing rate.(Polachek, 2007).

The standard error measures the standard deviation of the mean and tells you how the mean is distributed. Figure 1 shows the standard error for the different dependent variable. The standard error of educ has a value of 0.024, this can be interpreted that given the fact that the value is small, it gives us an indication that the mean is relatively close to the true mean of our overall population.

When you’re regressing for different samples there’s a possibility that heteroskedasticity could arises because the standard deviations of a variable, monitored over a specific amount of time, are nonconstant. Heteroskedasticity does not cause bias or inconsistency in the ols estimators however what occurs is that since, standard errors are based on the estimator variances there are no longer valid for constructing confidence intervals and t-statistics making the p values obsolete. Therefore, to tackle this issue, we use the robust standard error which helps because it still allows to report new statistics that work regardless of if heteroskedasticity is present. Figure 2 depicts the robust standard error, the coefficient is still the same as it doesn’t change when it’s calculated, the standard error is slightly different but there isn’t a significant different between the two.

Robust standard error Figure 2

|  |  |  |  |
| --- | --- | --- | --- |
| **Ln wage** | **Coefficient** | **Standard error** | **95% confidence overall** |
| educ | .1363014 | .0216186 | .0936567 .1789461 |
| exp | .0984937 | .0389588 | .0216437 .1753436 |
| expsq | -.0033068 | .0017186 | -.0066969 .0000832 |

To answer the second part of the question which iswhether there is any difference in the returns to education for women with children I will have to create a dummy variable. In order to make sure we don’t fall for a dummy variable trap which arises when to many dummy variables describe a given number of groups, using two dummy variables will lead to perfect collinearity. I created the dummy variable for women who have kids under the age of 6- nkidslt6 and over the age of 6 nkindgt6 and it was named kids7. This helps gives us an indication of the return to education for women who have kids and even more specifically depending on their age. The result I collected is depicted in figure 3 and it shows that there’s a negative relationship between having kids and wages and also the value is statistically significant at the 0.05 level.

Dummy variable Figure 3

|  |  |  |
| --- | --- | --- |
| **Ln wage** | **Coefficient** | **P>|t|** |
| educ | .1341552 | 0.000 |
| exp | .0941265 | 0.009 |
| expsq | -.0031644 | 0.063 |
| Kids7 | -.1346316 | 0.410 |

Conclusions

In conclusion experience and education positively affect wages in positive correlated interaction as I explained and by considering that experience has diminishing return, we make our result more accurate. The estimate to returns for education for women with kids is in line with many literatures that provide the consensus that there’s a negative relationship between wages and having kids.

Part 2

In this assignment i will be interpretating the data NSLY2000.dta to derive an empirical strategy to estimate the returns to education and I will also test whether there is any difference in returns to education for women with children. I will test whether there is any difference in the returns to education with reference to gender and race.

Introduction

The NSLY2000.dta is a panel data which are data sets where the same cross-sectional units are followed over time. Panel data sets are most useful when controlling for time-constant unobserved features for example people, firms, cities ecc… which could be correlated with the explanatory variables. (Wooldridge, 2013)

The methodology I will use to estimate my result from the data NSLY2000.dta is a pooled OLS. The pooled OLS will be used in other to estimate the return for education for the individuals in my sample. Therefore, I will run a pooled ols regression on the dependent variable learning in respective to my independent variable exp-experience, expsq = experience squared, male, ethblack- dummy variable=1 if black, ethhisp-dummy variable =1 if Hispanic. Earnings are logged because using the log of earnings rules out negative predicted values. Since the distribution of earnings tends to be lognormal, taking logs   
reduces the problem of heteroscedasticity. Experience is squared because it has diminishing returns. Subsequently I have generated three interaction variables between schooling, male, black and Hispanic which are defined as the following variable sch\_male, sch\_black,sch\_hisp. The equation below illustrates my pooled regression and variables considering the error term.

Log (earnings)= β0+β1Schooling+ β2 exp+β3exp squared++β4male+ β5ethblack+β6ethhisp+ β7sch\_male +β8sch\_black+β9sch\_hisp+Ui

Figure 4 illustrates results of the pooled ols regression depicted in the equation above and it shows the coefficient, standard error and p value for the f-test and t-test. Instantly we can infer that p-value of the f-test are significant since their respective value is 0.00 which is below the significant value of 0.05 this also means that null hypothesis has been rejected and the variables are jointly significant.

The t-test follows a similar understanding since the respective p-value is below the significant level of 0.05 we would conclude that it is also jointly significant. The respective coefficient is showing the relationship between the dependent variable and the independent variable for example a positive number which the variables s, exp and male show that there’s a positive correlation with earnings, these variables increase the chances of having higher earnings for example has a value .0816156 and exp. has a value of 0772562. The result is not surprising however the one variable that stands out of the three mentioned is the variable male with a value of .2318133 because it shows that being a male increases the possibility of you acquiring higher earnings, that follows the writings of (Meara, Pastore and Webster, 2019) that states that earnings are higher for men compared female. Experience squared has a negative coefficient of -.0011505 meaning that has experience increased, earning increase at a diminishing rate. The other independent variables measuring the effect of earnings is ethblack and ethhisp. There’s an interesting difference it seems that if you are of black ethnicity, the earnings decrease and it is statistically significant, and this is shown by the coefficient -.191258 which shows a negative relationship between the independent and dependent variable. In contrast ethhisp shows a positive relationship of the value of .145264 meaning that there’s a positive relationship between earning and being of a Hispanic ethnicity.

Figure 4 OLS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Log earnings | Coef. | Std. Err. | Prob>F | P>|t| |
| s | .0816156 | .001429 | 0.00 | 0.000 |
| exp | .0772562 | .0023323 | 0.00 | 0.000 |
| expsq | -.0011505 | .0001101 | 0.00 | 0.000 |
| male | .2318133 | .0232589 | 0.00 | 0.000 |
| ethblack | -.191258 | .0387842 | 0.00 | 0.000 |
| ethhisp | .145264 | .0397038 | 0.00 | 0.000 |
| sch\_male | -.0027907 | .0017382 | 0.00 | 0.108 |
| sch\_black | .0059174 | .0029916 | 0.00 | 0.048 |
| sch\_hisp | -.0103841 | .0031436 | 0.00 | 0.001 |
|  |  |  |  |  |
| Time |  |  |  | P>[t] |
| 1 | -.0607664 | .0157276 | -3.86 | 0.00 |
| 2 | -.1168361 | .0148777 | -7.85 | 0.00 |
| 3 | -.1659544 | .0150999 | -10.99 | 0.00 |
| 4 | -.2105105 | .0151338 | -13.91 | 0.00 |
| 5 | -.2272039 | .0153691 | -14.78 | 0.00 |
| 6 | -.2298459 | .0157439 | -14.60 | 0.00 |
| 7 | -.2213574 | .0161005 | -13.75 | 0.00 |
| 8 | -.2348251 | .0168497 | -13.94 | 0.00 |
| 9 | -.2819101 | .0171239 | -16.46 | 0.00 |
| 10 | -.3068747 | .0175949 | -17.44 | 0.00 |
| 11 | -.3511175 | .0179381 | -19.57 | 0.00 |
| 12 | -.3842307 | .0182445 | -21.06 | 0.00 |
| 13 | -.4106337 | .0185386 | -22.15 | 0.00 |
| 14 | -.4339888 | .0197929 | -21.93 | 0.00 |
| 16 | -.4683938 | .0190652 | -24.57 | 0.00 |
| 18 | -.4840601 | .0194971 | -24.83 | 0.00 |
| 20 | -.5083602 | .0202552 | -25.10 | 0.00 |
| Cons | .9787472 | .0218561 | 44.78 | 0.00 |

The random effects model determines individual effects of unobserved independent variable as random variables overtime. (Wooldridge, 2013) Figure 6 illustrates the random effect model which as depicted is statistically significant as the p value is lover than 0.05. The reason we used the random effect is because we assumed that the unobserved error might be uncorrelated with the explanatory variable and using the fixed effect would result in inefficient estimators.

In order for a pooled OLS to be consistent the unobserved effect most be uncorrelated with explanatory variable in order to fix that we use differencing which eliminates the unobserved factor prior to the estimation. Fixed effects is one way to remove the unobserved factor and in this particular data the unobserved factor were s, male, ethblack , ethhisp sch\_male, sch\_black, sch\_hisp these were all differenced. This can be shown by Figure 5 which shows the remaining variables experience and experience squared. These are also statistically significant as their p values have a value of zero.

The Hausman test is a test of endogeneity, by running the respective test the null hypothesis is that the covariance between Iv(s) and alpha is zero. If this is not the case the random effect is preferred to the fixed effect. The result of the Hausman test are depicted in Figure 6 and b = consistent under Ho and Ha; obtained from xtreg B = inconsistent under Ha, efficient under Ho; obtained from xtreg. This shows that both the FE and Re can be used and their both statistically significant as the reject null hypothesis.

Figure 5 (Fixed Effect)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Log earnings | Coefficient | Standard Error | P>t | Prob>F |
| exp | .102559 | .0029151 | 0.000 | 0.00 |
| expsq | -.0020036 | .0000926 | 0.000 | 0.00 |

Figure 6 (Random Effect)

|  |  |  |  |
| --- | --- | --- | --- |
| Log earnings | Coefficient | Standard error | P|z| |
| s | .0735495 | .0025247 | 0.000 |
| exp | .0918236 | .0022978 | 0.000 |
| expsq | -.0016901 | .000086 | 0.000 |
| male | .1450458 | .043355 | 0.001 |
| ethblack | .0654628 | .0741442 | 0.377 |
| ethhisp | .1708605 | .0742699 | 0.021 |
| sch\_male | .0034074 | .0032067 | 0.288 |
| sch\_black | -.0121622 | .0057144 | 0.033 |
| sch\_hisp | -.0135005 | .0058003 | 0.020 |

Hausman Test Figure 7

|  |  |  |
| --- | --- | --- |
|  | B  FE | B  RE |
| exp | .102559 | .0918236 |
| expsq | -.0020036 | -.0016901 |

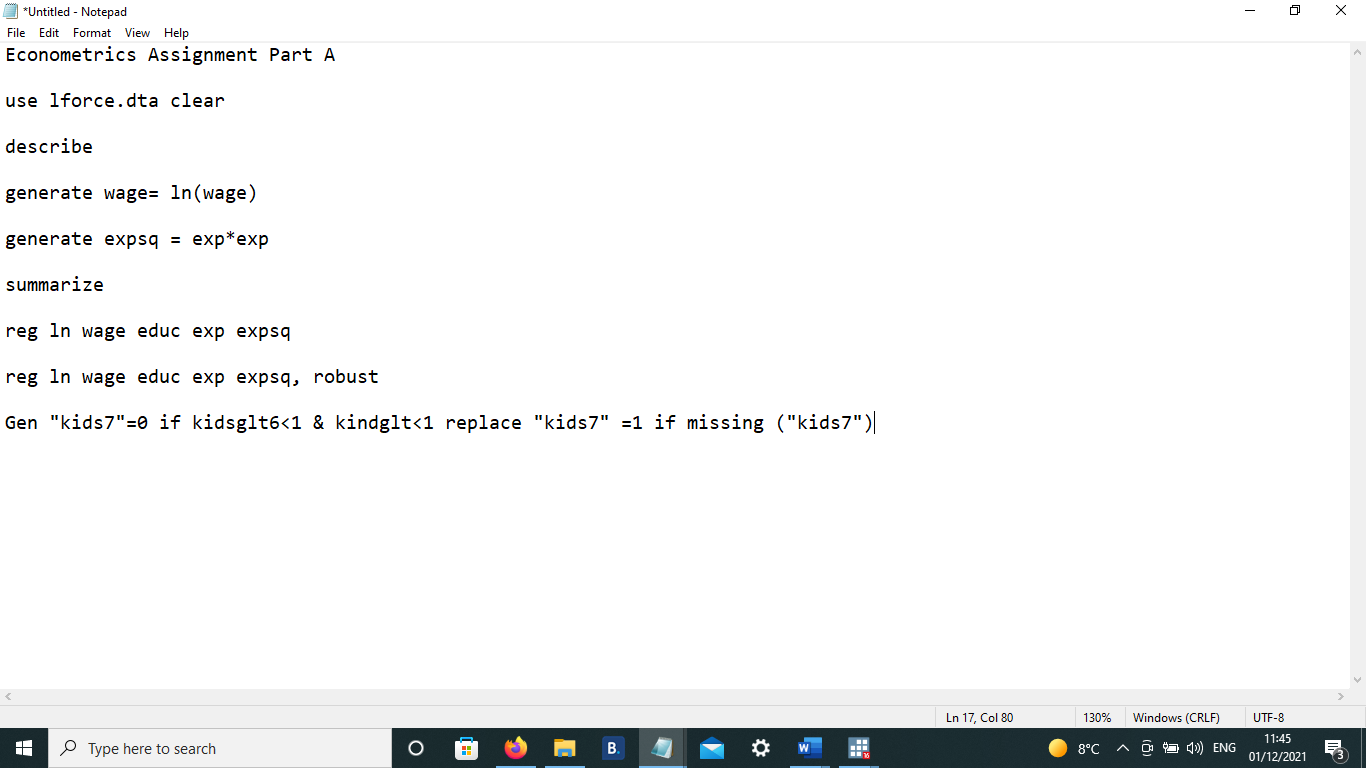
Figure 8 estimation ( OLS,RE, FE)

|  |  |  |  |
| --- | --- | --- | --- |
| Log Earnings | OLS | FE | RE |
| s | 0.0816\*\*\* |  | 0.0735\*\*\* |
| exp | 0.0773\*\*\* | 0.1026\*\*\* | 0.0918\*\*\* |
| expsq | -0.0012\*\*\* | -0.0020\*\*\* | -0.0017\*\*\* |
| male | 0.2318\*\*\* |  | 0.1450\*\*\* |
| ethblack | -0.1913\*\*\* |  | 0.0655 |
| ethhisp | 0.1453\*\*\* |  | 0.1709\*\* |
| sch\_male | -0.0028 |  | 0.0034 |
| sch\_black | 0.0059\*\* |  | -0.0122\*\* |
| sch\_hisp | -0.0104\*\*\* |  | -0.0135\*\* |

Conclusion

To conclude it seems that my result follows the literature review and the subsequent papers I have found that there’s positive relationship between experience, schooling and the gender male and the results are significant and follow the trend. Subsequently race seems to play an important in earnings especially if you are of a black ethnicity in contrast if you’re Hispanic it seems that there’s a positive relationship.

Appendix



Graphical user interface, text, application

Description automatically generated

Bibliography

Blau, F. and Kahn, L., 2017. The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3), pp.789-865.

Meara, K., Pastore, F. and Webster, A., 2019. The gender pay gap in the USA: a matching study. *Journal of Population Economics*, 33(1), pp.271-305.

Mincer, J., 1993. *Schooling, experience, and earnings*. Brookfield: Gregg Revivals.

Polachek, S., 2007. Earnings Over the Life Cycle: The Mincer Earnings Function and its Applications. *SSRN Electronic Journal*,.

Polachek, S., 2007. Earnings Over the Life Cycle: The Mincer Earnings Function and its Applications. *SSRN Electronic Journal*,.

Rodríguez-Pose, A. and Tselios, V., 2009. EDUCATION AND INCOME INEQUALITY IN THE REGIONS OF THE EUROPEAN UNION. *Journal of Regional Science*, 49(3), pp.411-437.

Wooldridge, J., 2013. *Introductory econometrics*. 5th ed. Australia: South-Western Cengage Learning.