Course Name	**Master of Science in Business Analytics**
Module Name	**Applied Econometrics For Business(EC6062)**
Module Leader	**Dr. Dionysios Karavidas**
Batch	**2019-2020**
Semester	**Spring**

Essay Topic:

Estimating the return to education (schooling) based on the Mincer equation

Submitted by:
Aishwarya C Inamdar
19084404

Table of Contents

Estimating the return to education (schooling) based on the Mincer equation

Introduction

Earnings functions are the most commonly used analytical equations in job and education economics.[1] In 1958, Jacob Mincer pioneered an important method for understanding the distribution of earnings through the population. He, along with his students and colleagues, expanded the original human capital model in the years following Mincer's ground-breaking work, drawing significant conclusions regarding a whole range of findings relating to human well-being. This line of research explained why schooling increases earnings; why earnings rise at a decreasing rate over one's life; why earnings growth is lower for those who expect irregular labor participation; why males earn more than females; why whites earn more than blacks; why gender disparities in occupational distribution; why geographic and work mobility predominate among young people; and why certain other anomalies on the labor market do occur.[2]
Human capital accrued on the job is considered to be one of the key determinants of a wage rate for an employee. According to the wage equation set up by Mincer (1974), a linear and a quadratic concept is used to include work experience or human capital. One stylized fact is that the resulting profile of wage-experience is hump-shaped. The explanatory variable of human capital is also determined by the number of years spent in jobs and the education.[3]

Data Description

The main source of data for the analysis of this model is the WAGE2.RAW. The data frame consists of 935 observations on 17 variables namely wage (monthly earnings), hours (average weekly hours), IQ (IQ score), KWW (knowledge of world work score), educ (years of education), exper (years of work experience), tenure (years with current employer), age (age in years), married (=1 if married), black (=1 if black), south (=1 if live in south), urban (=1 if live in SMSA), sibs (number of siblings), brthord (birth order), meduc (mother's education), feduc (father's education) and lwage (natural log of wage). The following table gives an overview of the data being used for the analysis. Table 1 represents a summary of the data frame obtained from Stata. It is observed that the data is categorized based on the family background like number of siblings, birth order and education of the mother and father, marital status and the working region along with the demographic factors like the age, number of years spent in education, the IQ level, the work experience and the number of years completed with the current employer.

Model and variable description

Model description

The empirical model,

$$\boxed{lwage = \beta_0 + \beta_1 educ + \beta_2 X + \beta_3 X^2 + u} \quad (1)$$

uses the dependent variables educ and X where X denotes the potential experience.

Following the empirical model, we model the effect of education on earnings with a standard Mincerian wage equation by adding a few more dummy variables in order to eliminate the violation of the zero conditional mean assumption, i.e., u being uncorrelated with educ and also to increase the value of R^2.

The following econometric model is used for the estimation of the

$$\boxed{\begin{aligned} lwage = \beta_0 + \beta_1 educ + \beta_2 exper1 + \beta_3 exper1\_sq + \beta_4 tenure + \beta_5 black + \beta_6 south + \beta_7 urban + \\ \beta_8 married + u \end{aligned}} \quad (2)$$

Variable description

In the suggested model in equation 2, wage denotes the logarithmic value of monthly earnings, educ denotes the years of education completed and exper1 and exper1_sq denote the potential experience and the rest of the dummy variables are used as explained in the data description part. As is common in the literature, we proxy potential experience exper by a quadratic age term. Here, exper1 and exper1_sq are correlated, but the quadratic in experience provides a better fit and hence we use it. Adding the explanatory variables to the regression model reduces the error variance. Hence, we have included some independent variables which are uncorrelated to the existing independent variables. The coefficient of interest is β_0 which describes the percent change in the monthly earnings due to a one-year marginal change in attained education, educ. [5]

Methodology

Firstly, we suppose to begin with what we are interested in the effect of education, β_1, measured in years of schooling, on wages:

$$wage = \beta_0 + \beta_1 educ + u \quad (3)$$

The u in the above equation denotes the dummy variables which are under the zero conditional mean assumption i.e., $E(u|educ) = 0$, the regression function for the above model describes the wages conditional on a level of schooling as a linear function of the parameters.

Using the data, we construct a scatter plot of the wages and education level and a histogram of residuals of wage and the log values of wage (lwage). The scatter plot in Fig. 1 shows that wages and education share a positive correlation. Comparing the figures 1 and 2 of the histogram plots, it is evident that the residuals from the lwage are more normally distributed as compared to the wage plot. Certainly, the histogram in Fig. 3 fits under its comparable normal density better than in Fig. 2. Moreover, the histogram for the wage residuals is notably skewed to the left, hence proving that the residuals are not normally distributed. Hence, we use the logarithmic value of wages to reduce the problem of heteroscedasticity.

Firstly, as per Mincer's equation, we create 2 independent variables for experience potential namely exper1 where exper1 = (age – educ – 6) and exper1_sq is the squared value of exper1 and then we find the correlation between the dependent and the independent variables used in equation 2 (refer figure 4). Using the model mentioned in equation 2, estimation of the same is done using multivariate OLS regression method in order to identify the association of the dependent variables educ, exper1, exper1_sq, tenure, black, south and urban on the wages. OLS regression method is used since it minimizes the sum of the squared residuals. The analysis begins by estimating the Mincerian wage equation as mentioned in equation 3 in order to estimate the overall returns to education for the potential experience (refer figure 5 for the results). Additionally, we estimate the returns to education by subsamples: potential experience, tenure, black and non-black, south and urban and marital status as mentioned in equation 2. Regional estimates provide a description of the geographic distribution of the returns within the rural and urban areas (refer figure 6 for the results).[6] Inferences on the results are made by observing the R-square values and the F and t statistics and to test the significance of the proposed model, i.e., the hypothesis that the selected variables are jointly significant.

H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = \beta_7 = 0$ (vs.) H_1: H_0 is not true

Observations and Results

Figure 4 represents the correlation of wage (dependent variable) with the independent variables. We can conclude that expect for the black and south variables, all the other variables are positively correlated with the wage. By estimating the equation 3, the usual for of the model after applying OLS regression is as follows (refer figure 5 for the coefficient values):

lwage = 5.973062 + .0598392 educ + u

The coefficient of education has a percentage interpretation, i.e., predicted wages increase by 5.9832% for every additional year of education. The ease in the interpretation of the rate of return of schooling is one of the reasons for using the log wage specification for estimation. In the human capital interpretation of the wage equation, it means that the rate of return of one year of schooling is almost 6%, which is not that bad! However, the value of R-square is very low for this model (0.0974). This means that 9.74% of the variation in wages is explained by their educational level. This means that 90.26% of the variation in wages remains unexplained. This is reason we add more variables as explained earlier. The t-statistic of 10.03 in this case confirms that education is definitely a significant factor in explaining the variation in wages. Furthermore, the confidence interval lies between .0481366 and .0715418, that is there is 95% chance that the true coefficient of schooling lies within this range. The value of 5.973062 for β_0 states that a person with zero years of schooling has 5.973062 predicted wage. We get a positive value for this constant as we have used the logarithmic values of wages which cannot take negative values. In this case, we reject the H_0 since none of the β values are zero.

Figure 6 represents the estimated values by applying OLS regression on the predicted model with the introduction of more independent variables. The concluding equation is as follows (refer figure 6 for the coefficient values):

lwage = 5.052158 + .0680631 educ + .0562756 exper1 - .0015017 exper1_sq + .0122592 tenure - .1867148 black - .0893135 south + .1830286 urban + .200146 married + u

From the above equation, we can say that the predicted wages increase by 6.806% for every additional year of schooling. Since the values of exper1 (.0562756>0) and exper1_sq (-.0015017<0) being contrast in signs, there is a concave relationship between log wages and experience. The predicted wage increases by 1.1867% for every 1-year rise with the current employer. The coefficient of black determines the difference between the monthly salary of blacks and non-blacks. From the values obtained, it can be inferred that the monthly salary of blacks is 18.67148% lower than that of the non-blacks. Similarly, the monthly salary of people living in the south is 9.37151% less than those living in the SMSA. However, married people have their monthly wage 20.0146% more than unmarried ones. The value of R-square has increased to 0.2496, i.e., 24.96% of the variation in the wages can be explained with the addition of these variables. This thus infers that these variables do have a positive effect on the monthly wages and this model with the introduction of dummy variables is better than the one without them. The p-value of all the coefficients of the independent variables expect the experience potential are less than the critical p-value of 0.05 at 5% level of significance indicating that the difference in monthly salary due to these variables is statistically significant at 5% level of significance. With the experience variable, the ceteris paribus assumption does not work directly, when we increase exper1, exper1_sq will increase as well, which needs the computation of the partial effects. This means that exper is uncorrelated with the lwage and thus educ and exper are also uncorrelated. Since none of the β values are zero, we reject the H_0.

Conclusion

Overall, education has a positive effect on the monthly wage of the individual. The dummy variables inserted to detect variation in differences in earnings have shown an overall positive effect on the monthly wages. As expected, the negative effects are predominant on the earnings of individuals who reside in the southern region. The gaps between the black's and the non-black's income shows a significant difference. Furthermore, the tenure of the individual with the current employer also has a positive effect on the monthly wage along with the marital status.

Limitations and Gaps

1) Introduction of omitted variable bias is possible if we omit the variables have an effect on our parameter estimate.
2) In case of the presence of heteroskedasticity, the standard errors of the estimates are biases even if the OLS estimates are unbiased and consistent. Thus, we cannot draw inferences using the t-statistics or F-statistics.
3) The schooling-earnings relationship does not necessarily mean causality.
4) Earnings functions include private (i.e. individual) returns to education, while government / public costs and other benefits are necessary to estimate the social return rates.

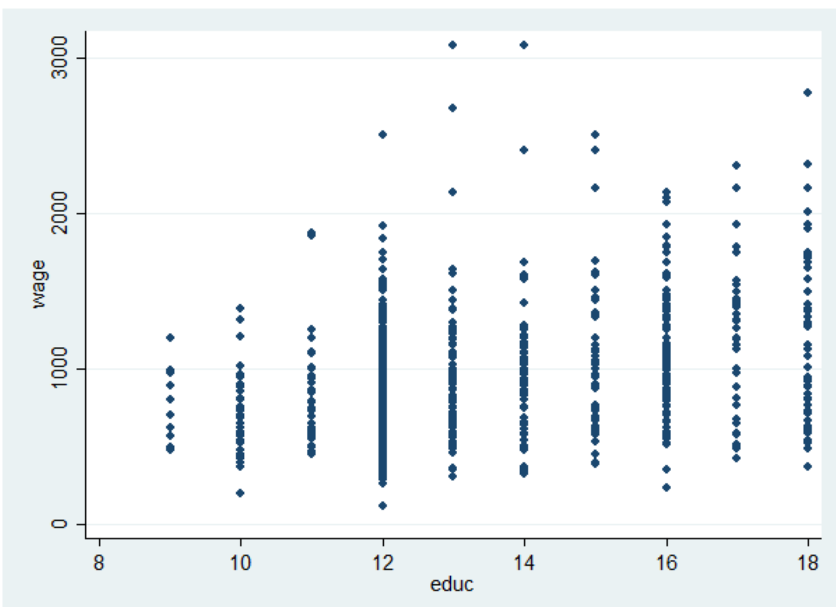
Figures and Tables

Table 1: Summary of the data used for analysis

wage	hours	educ	exper	tenure	age	married	black	south	urban	lwage	experl	expecl_sq
769	40	12	11	2	31	1	0	0	1	6.645091	13	169
808	50	18	11	16	37	1	0	0	1	6.694562	13	169
825	40	14	11	9	33	1	0	0	1	6.715384	13	169
650	40	12	13	7	32	1	0	0	1	6.476973	14	196
562	40	11	14	5	34	1	0	0	1	6.331502	17	289
1400	40	16	14	2	35	1	1	0	1	7.244227	13	169
600	40	10	13	0	30	0	0	0	1	6.39693	14	196
1081	40	18	8	14	38	1	0	0	1	6.985642	14	196
1154	45	15	13	1	36	1	0	0	0	7.05099	15	225
1000	40	12	16	16	36	1	0	0	1	6.907755	18	324
930	43	18	8	13	38	1	0	0	0	6.835185	14	196
921	38	14	9	11	33	1	0	0	0	6.82546	13	169
900	45	15	4	3	30	0	0	0	0	6.802395	9	81
1318	38	16	7	2	28	1	0	0	1	7.183871	6	36
1792	40	16	9	9	34	1	0	0	1	7.491087	12	144
958	50	10	17	2	35	1	0	0	1	6.864848	19	361

Source: https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041 (STATA data editor output)

Figure 1: Scatter plot of wage against education



Figures 2 and 3: Histogram plots of the residuals for variables wage and lwage respectively.

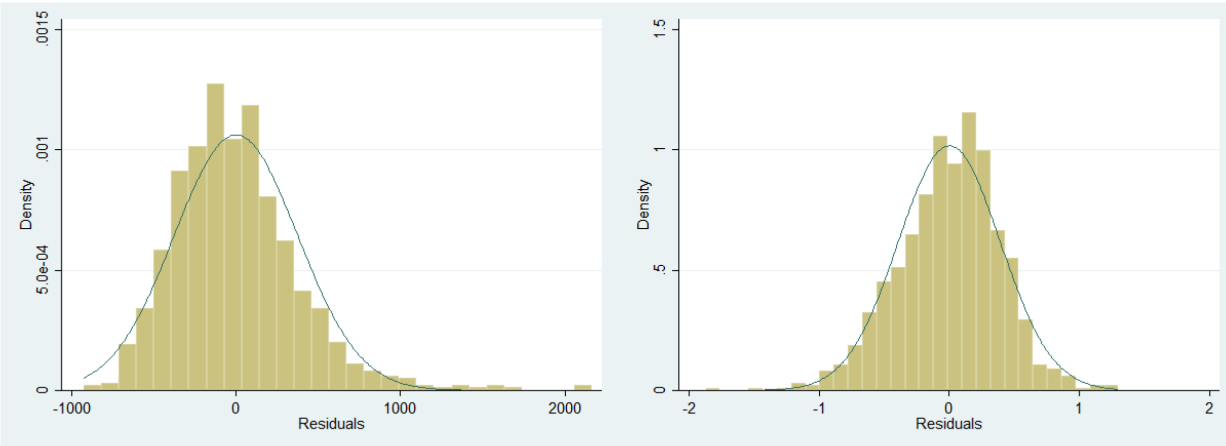


Figure 4: correlation between the dependent and independent variables (refer equation 3):

```
. correlate lwage educ exper1 exper1_sq tenure black south urban married
(obs=935)
```

	lwage	educ	exper1	exper1~q	tenure	black	south	urban	married
lwage	1.0000								
educ	0.3121	1.0000							
exper1	-0.0477	-0.5838	1.0000						
exper1_sq	-0.0513	-0.5681	0.9897	1.0000					
tenure	0.1859	-0.0362	0.2405	0.2410	1.0000				
black	-0.2321	-0.1795	0.0740	0.0736	-0.0782	1.0000			
south	-0.1948	-0.0970	0.0318	0.0354	-0.0617	0.2365	1.0000		
urban	0.2038	0.0722	-0.0469	-0.0451	-0.0385	0.0702	-0.1099	1.0000	
married	0.1500	-0.0586	0.1205	0.1200	0.0726	-0.0534	0.0228	-0.0402	1.0000

Figure 5: Regression of the Mincer equation without dummy variables (refer equation 3)

```
. regress lwage educ
```

Source	SS	df	MS
Model	16.1377074	1	16.1377074
Residual	149.518587	933	.16025572
Total	165.656294	934	.177362199

Number of obs	=	935
F(1, 933)	=	100.70
Prob > F	=	0.0000
R-squared	=	0.0974
Adj R-squared	=	0.0964
Root MSE	=	.40032

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.0598392	.0059631	10.03	0.000	.0481366	.0715418
_cons	5.973062	.0813737	73.40	0.000	5.813366	6.132759

Figure 6: Regression with the dummy variables (refer equation 2)

```
. regress lwage educ exper1 exper1_sq tenure black south urban married

      Source |       SS           df       MS      Number of obs   =       935
-------------+----------------------------------   F(8, 926)       =     38.50
       Model |  41.3481377         8  5.16851721   Prob > F        =    0.0000
    Residual |  124.308157       926  .13424207    R-squared       =    0.2496
-------------+----------------------------------   Adj R-squared   =    0.2431
       Total |  165.656294       934  .177362199   Root MSE        =    .36639

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .0680631    .006917     9.84   0.000     .0544883    .0816378
      exper1 |   .0562756   .0223079     2.52   0.012     .0124958    .1000555
   exper1_sq |  -.0015017    .000778    -1.93   0.054    -.0030285    .0000251
      tenure |   .0122592   .0024688     4.97   0.000      .007414    .0171044
       black |  -.1867148   .0377642    -4.94   0.000    -.2608281   -.1126015
       south |  -.0893135   .0263304    -3.39   0.001    -.1409876   -.0376393
       urban |   .1830286   .0270271     6.77   0.000     .1299871      .23607
     married |    .200146   .0392029     5.11   0.000     .1232092    .2770828
       _cons |   5.052158   .2038236    24.79   0.000     4.652148    5.452167
------------------------------------------------------------------------------
```

References

[1] Heckman, J. J., Lochner, L. J. and Todd, P. E. (2003) ‘Mincer_Heckman.Pdf’.

[2] Polachek, S. W. (2008) ‘Earnings over the life cycle: The mincer earnings function and its applications’, Foundations and Trends in Microeconomics, 4(3), pp. 165–272. doi: 10.1561/0700000018.

[3] Beblo, M. and Wolf, E., 2003. How Much Does A Year Off Cost? Estimating the Wage Effects of Employment Breaks and Part-Time Periods. SSRN Electronic Journal,.

[4] Heckman, J. J., L. J. Lochner and P. E. Todd (2005). "Earnings Functions, Rates of Return and Treatment Effects: The Mincer Equation and Beyond." National Bureau of Economic Research Working Paper Series No. 11544.

[5] Anderson, S. (2016) ‘Econ 495 -Econometric Review’.

[6] Peet, E., Fink, G. and Fawzi, W., 2015. Returns to education in developing countries: Evidence from the living standards and measurement study surveys. Economics of Education Review, 49, pp.69-90.

[7] Patrinos, H., 2016. Estimating the return to schooling using the Mincer equation. *IZA World of Labor*,.

[8] Arraes, R., Menezes, F. and Simonassi, A., 2014. Earning differentials by occupational categories: Gender, race and regions. *EconomiA*, 15(3), pp.363-386.

[9] UKEssays. November 2018. Earnings Function To Best Explain Wage Gaps Economics Essay. [online]. Available from: https://www.ukessays.com/essays/economics/earnings-function-to-best-explain-wage-gaps-economics-essay.php?vref=1 [Accessed 18 April 2020].