Does education actually lead to a richer life or is it clickbait?

Introduction:

Graetz G. (2023) suggests that better educated workers may earn higher wages for two possible reasons: because they acquire more skills, or because more education illustrates intellect and enthusiasm. It has also been implied by the NCES that individuals with a college degree continue to enjoy both better employment prospects and greater annual earnings than those with lesser levels of education.

Furthermore, studies have shown that there are other variables which influence an individual's wage, such as family background, ethnicity, or siblinghood. Card, D. (1999) has suggested that education plays a central role in modern labour markets. Therefore, it's crucial to analyse whether education is the only important determinant to higher income.

We have a prominent focus on educational achievement levels across observations and observe potential wage inequalities by estimating regressions, conducting statistical tests and evaluating wage distributions between the different levels and gender.

During this investigation, we seek to determine "Whether education actually leads to higher income for individuals?" by utilising cross-sectional data from Understanding Society Survey (UKHLS) - specifically from Wave 13. This question doesn't merely carry a theoretical interest; but practical implications for individuals who contemplate whether an investment in education is worth their hard-earned income, as well as for policymakers to devise strategies and enhance human capital development for economic growth & dynamic efficiency.

Data and Methodology:

This study uses data from the Understanding Society Survey (UKHLS). It's a nationally represented survey of respondents based in the UK. Data relating to Wave 13 is analysed and represents a cross-sectional investigation.

The focal point of our investigation is education and how different levels of qualifications may impact wages. In the UKHLS, respondents are asked questions about what their highest education qualification was. Following a similar structure to Walker, I et al (2003), who focus on the relationship between education, earnings and productivity, we segregate our focal explanatory variable into 4 categories: secondary, undergraduate, postgraduate and other. The dependent variable is annual income. This is based on total monthly labour income gross and we estimated annual income by taking the observations from this variable and multiplying them by 12 (months). We estimated annual labour income by calculating the following: total monthly labour income x 12.

Initially, we've imposed two models: one is in linear-linear form, and the other in log-linear form (shown below) and regressions are performed by using the Ordinary-Least Squares (OLS) Estimator.

The following show the models that are estimated:

$$\begin{aligned} Wage_{i} &= \beta_{0} + \beta_{1}educ_{i} + \gamma\alpha_{i} + u_{i} \ (1) \\ ln(Wage_{i}) &= \beta_{0} + \beta_{1}educ_{i} + \gamma\alpha_{i} + u_{i} \ (2) \end{aligned}$$

Where the 'ln' (in equation 2) represents the natural logarithm and *wage* (in eq 1 and 2) is our estimated annual labour income. Our explanatory variable *educ* represents whether an individual's highest qualification is secondary school, undergraduate, postgraduate or another type of qualification. α represents the other explanatory variables which are typically included in wage determination models. The controls that are included with our analysis are similar to the determinants used in Woodhall, M. (1987). Variables included are: age, gender, ethnicity, postal code, job satisfaction, marital status, if they have children, socioeconomic status, whether they work for a private company & amount of sleep time. A full list of how the variables are coded and mean wage for people with the respective levels of education are shown in the appendix (Tables A1 and A2).

We have decided to move forward with equation (2) which follows Card, D. (1999) and Mincer's (1974) "human capital earnings" function as we believe that using log for income will allow us to mitigate the occurrence of MLR4 violation.

Results:

As mentioned in the Data and Methodology section, all results from the regressions (unless stated otherwise) are estimated using the natural logarithm of the estimated annual wage as the dependent variable. Using the natural logarithm for wage is nothing new, and is simply used as a mathematical tool to decrease the range of outliers and 'space' between observed values and has been used in historical papers such as Gruetter, M. et al (2009) and Bloch, F.E. et al (1978) just to name a few.

A Regression Specification Test (RESET) for the full sample was performed, where the dependent variable in equation (1) was linear, whilst being logarithmic in (2). Results from the respective tests can be found on Table B2. Our findings suggest that mis-specification appears in our linear model, which implies that there may be some variables that aren't transformed correctly to explain causation in our model. However, the same outcome is found with our logarithmic model, which suggests that there may be explanatory variables which are contained in the error term, hence creating a bias in our results. Although our results are a good indication, Wooldridge (2019) and Shukur, G. et al (2004) have questioned how reliable this test is by commenting that it tests for functional form and that the power of the test is low (resp).

Additionally, we found the VIF (Variation Inflation Factor), which informed us of the amount of multicollinearity that occurred in our regression. We obtained the mean VIF value for the models, which were both 6.09 (shown on Table B1). This is expected since both equations use the same independent variables in their regression. Since our VIF is less than the 0.10 (silver-bullet) benchmark, Daoud J.I (2017) suggests that this indicates collinearity and we can assume that there's no issue of multicollinearity in the model.

Another important test we decided to perform is the Breusch-Pagan test, which tests for evidence of heteroskedasticity in the model. In the linear-linear model, the f-statistic is equal to 25.47, while the p-value is lower than the 1% significance level (shown on Table B2), thus there is strong evidence of heteroskedasticity. Same conclusion is drawn for the logarithmic-linear model, where the f-statistic is 4.24 and the p-value is also lower than the 1% significance level (Table B2), making it possible to reject the null hypothesis of constant variance. As a result, MLR5, also known as the Gauss-Markov Theorem, is violated, indicating that the OLS estimator isn't the best unbiased estimator and Robust Standard Errors are used for all our estimated results. Therefore, we will be using the log-linear model due to this and aforementioned reasons.

The regression results for our base regression is shown on Table 1 (full set of results from the regression in Table B3). With reference to the results in this section of the appendix, 'undergraduates' earn 22.0% more on average, whilst 'postgraduates earn 30.4% (when being compared to secondary graduates, holding all else constant). These estimated coefficients

reinforce the investigation, in that, higher education is a crucial determinant for higher wages across graduates. Each of the coefficients noted have p-values < 0.01, indicating strong statistical significance across those observations.

Furthermore, with respect to location, (m_gordv) , there's a significant influence on graduate salary, and wage differentials with regards to geographical inequality. As seen in our estimated regression analysis and relative findings from Rienzo (2016), all major regions of the UK portray a lower estimated annual wage in comparison to our base category 'London' (comparison of wages shown on Table B4).

Although not a primary focus of our investigation, the quadratic form of variable 'age' plays an influential role in reflecting a concave relationship between itself and wage. We found that the turning point was around 45 years old (which can be seen on Graph B5). This is where wages start diminishing. While wages tend to increase with experience (overtime), there's typically a point where wages start to diminish, due to plentiful reasons such as retirement, skill depreciation, or age-related discrimination. Our graph, B6, and data from UKHLS Wave 13 may also illustrate the early career wage growth, for those who obtained a postgraduate degree, and the potential earlier retirement that corresponds to this.

We've decided to create a subgroup of our population to see how wages are affected if an employee works at a private company or not. From Table 1 (full regression can be found in Table B3), we see undergraduates (relative to secondary graduates) earn similar wages whether they work for a private/public based company (where UG graduates at public companies earn approximately 21.9% on average more than secondary graduates, and private company employees earn approximately 21.7% more on average). The most prominent finding here is between post-graduates' salaries. We see that there's a much wider gap between graduates who work in a public company and those who don't (compared to our base, holding everything else constant, post-graduates who are public employees earn approximately 31.4% more on average compared to private employees who earn 29.8% more on average). This is due to some obvious reasons, such as having larger economies of scale, ability to raise larger pools of investment from stakeholders or stronger collective bargaining in the public sector, suggested by Ehrenberg R.G et al (1986). Therefore, our results suggest, even though a higher level of qualification does have a high 'degree' of impact on wages, another consideration would be in what type of organisation the individual works at.

Table 1: Relationship Between Wage, Company Type and Level of Education

Regression	(a)	(d)	(e)
Type of Model	Full (log-linear)	Public Company	Private Company
age	0.116***	0.109***	0.126***
	(0.000)	(0.000)	(0.000)
c.age##c.age	-0.00129***	-0.00123***	-0.00139***
	(0.000)	(0.000)	(0.000)
female	-0.401***	-0.325***	-0.444***
	(0.0130)	(0.0207)	(0.0170)
ug	0.220***	0.219***	0.217***
	(0.0152)	(0.0249)	(0.0191)
pg	0.304***	0.314***	0.298***
	(0.0184)	(0.0288)	(0.0240)
other	-0.160***	-0.0958	-0.166***
	(0.0540)	(0.125)	(0.0603)
R^2	0.392	0.306	0.451
N	9309	3810	5499

Robust standard errors in parentheses p < 0.1, p < 0.05, p < 0.05. Other variables included in the model, but not shown here: age, age-squared, white, asian, black, mixed/other, NEng, MidEng, SEng, W+NI, Scot, LDN, dissatisfied, neutral, satisfied, not legally binded, legally binded, has no child, has child, unemployed, professional, intermediate, routine, worker in public company, worker in private company, sleeptime. Full regression results on Table B3.

Furthermore, within splitting the sample by private and public employees, we should notice that gender wage discrimination is much smaller in public companies compared to private, with there being a 10% noticeable difference between female pay in public and private companies: it's estimated that females earn 32.5% less in public companies, compared to females earning 44.4% less in private companies. Our results are consistent with findings from Ehrenberg R.G et al (1986), where public sectors result in less gender discrimination than of [what] occurs in the private sector. Consequently, our results also suggest that gender is a key determinant of one's wages. However, we have also found that the gender wage gap can be mitigated through females pursuing higher levels of education, see Table 2 (full results on Table B7): We find, with strong evidence, that wage gap between secondary graduates and postgraduates, within our subsamples of females, are much larger compared to males; more education produces greater increases in the wage rate for females than for males, Huang, T. (1999).

Table 2: Relationship Between Wage, Gender, and Level of Education

Regression	(a)	(b)	(c)
Type of Model	Full (log-linear)	By Female	By Male
age	0.116***	0.109***	0.126***
	(0.000)	(0.000)	(0.000)
c.age##c.age	-0.00129***	-0.00123***	-0.00139***
	(0.0130)	(0.000)	(0.000)
female	-0.401***		
	(0.0152)		
ug	0.220***	0.242***	0.185***
	(0.0152)	(0.000)	(0.000)
pg	0.304***	0.335***	0.283***
	(0.0184)	(0.000)	(0.000)
other	-0.160***	-0.148**	-0.157*
	(0.0540)	(0.034)	(0.062)
R^2	0.392	0.349	0.399
N	9309	5155	4154

Robust standard errors in parentheses p < 0.1, p < 0.05, p < 0.01. Full regression results on Table B7.

Conclusion:

To conclude, this study investigated the relationship between wage and education, whilst looking at some of the importance of other explanatory variables. Our results imply that higher education is strongly correlated with higher income, and females benefit more from higher education, which is consistent with previous findings. We found through statistical tests and regressions that, despite the strong relationship between education and wages, other factors such as geographical location and age also have a widening impact on how much an individual earns.

A limitation of our investigation is that the number of observations in our analysis reduced drastically due to invalidation and so our model may not provide a holistic opinion on all individuals who filled out the survey.

Lastly, further work could investigate the cross-relationships between education with other explanatory variables, i.e. interaction terms, and fully determine whether higher education leads to more beneficial outcomes for females.

Bibliography:

Bloch, F.E. and Kuskin, M.S. (1978). Wage Determination in the Union and Nonunion Sectors. *Industrial and Labor Relations Review*, 31(2), p.183. doi:https://doi.org/10.2307/2522386.

Card, D. (1999). *Chapter 30 - The Causal Effect of Education on Earnings*. [online] ScienceDirect. Available at:

https://www.sciencedirect.com/science/article/pii/S1573446399030114.

Daoud, J.I. (2017). Multicollinearity and Regression Analysis. *Journal of Physics: Conference Series*, 949(1). doi:<u>https://doi.org/10.1088/1742-6596/949/1/012009</u>.

Ehrenberg, R.G. and Schwarz, J.L. (1986). *Chapter 22 Public-sector labor markets*. [online] ScienceDirect. Available at:

https://www.sciencedirect.com/science/article/pii/S1573446386020126.

Graetz, G. (2023). *More education signals ability to employers, but does it really improve skills?* [online] LSE Business Review. Available at:

https://blogs.lse.ac.uk/businessreview/2023/07/31/more-education-signals-ability-to-employers-but-does-it-really-improve-skills/#:~:text=Better%20educated%20workers%20may%20earn.

Gruetter, M. and Lalive, R. (2009). The importance of firms in wage determination. *Labour Economics*, 16(2), pp.149–160. doi: https://doi.org/10.1016/j.labeco.2008.09.001.

Huang, T. (1999). The impact of education and seniority on the male-female wage gap: is more education the answer? *International Journal of Manpower*, 20(6), pp.361–374. doi:https://doi.org/10.1108/01437729910289710.

Rienzo, C., 2016. Real wages, wage inequality and the regional cost-of-living in the UK. Empirical Economics, Volume 52, pp. 1309-1335.

doi: https://link.springer.com/article/10.1007/s00181-016-1122-4

Shukur, G. and Mantalos, P. (2004). Size and Power of the RESET Test as Applied to Systems of Equations: A Bootstrap Approach. *Journal of Modern Applied Statistical Methods*, 3(2), pp.370–385. doi:https://doi.org/10.22237/jmasm/1099267800.

Walker, I. and Zhu, Y. (2003). Education, earnings and productivity: Recent UK evidence. [online] Available at:

https://www.researchgate.net/profile/Yu-Zhu-38/publication/228871438_Education_earnings_and_productivity_Recent_UK_evidence/links/00b49520b43777fda2000000/Education-earnings-and-productivity-Recent-UK-evidence.pdf.

Woodhall, M. (1987). Earnings and Education. [online] Science Direct. Available at: https://www.sciencedirect.com/science/article/pii/B9780080333793500413.

Wooldridge, J.M. (2019). *Introductory Econometrics: a modern approach*. S.L.: Cengage Learning.

Appendix:

Table A1 - Variable Definitions & Codes

Variable Name	Original Variable Names	Coding	Base Group	Expected Sign of Coefficient
wage (dependent)	m_fimnlabgrs_dv	multiplied by 12 to get the annual income gross	N/A	N/A
age	m_dvage	renamed to "age"	N/A	(+ve) until local max.
male	m_sex	female, male	male	(+ve)
educ	m_isced11_dv	secondary, UG, PG, other	secondary	(-ve)
race	m_racel_dv	white, asian, black, mixed & other	white	(+ve)
post	m_gor_dv	NEng, MidEng, SEng, Wales+NI, Scotland, London	london	(+ve)
jobsat	m_jbsat	Inapplicable, dissatisfied, neutral, satisfied	satisfied	(+ve)
marbi	m_mastat_dv	Not legally binded, legally binded	not legally binded	(+ve) especially with females because more career focussed
hvchild	m_nchild_dv	No child, has child	no child	(+ve) because same reason as above
sociog	m_jbnssec3_dv	Unemployed, professional , intermediate, routine	professional	(+ve)
privcom	m_jbsect	Don't work for private company, Work for private company	Don't work for private company	(+ve) non-private are larger companies
sleeptime	m_hrs_slph & m_hrs_slpm	Time variable representing the number of minutes an individual sleeps	N/A	(+ve)

Table A2 - Means of Wages by Highest Qualification and Gender

Education Type	Full Sample	Female Sample	Male Sample
Secondary	£11237.14	£8684.44	£14199.70
Undergraduate (UG)	£20102.65	£16820.43	£24868.81
Postgraduate (PG)	£27910.08	£23938.2	£32319.91
Other	£2682.275	£1823.448	£4010.774

Table B1 - VIF values for equation (1) and (2):

Variable	VIF	1/VIF
age	51.88	0.019276
c.age#c.age	51.71	0.019337
0.male	1.08	0.922197
educ		
2	1.36	0.733959
3	1.44	0.694555
4	1.06	0.939844
race		
2	1.10	0.910959
3	1.12	0.892451
4	1.04	0.962126
post		
1	2.21	0.452697
2	3.16	0.316060
3	2.77	0.361470
4	2.25	0.444005
5	1.91	0.524117
jobsat		
1	1.08	0.927007
2	1.08	0.926233
1.marbi	1.42	0.703438
<pre>1.hvchild</pre>	1.40	0.712774
sociog		
2	1.21	0.826177
3	1.47	0.681180
1.privcom	1.14	0.878274
sleeptime	1.06	0.947445
Mean VIF	6.09	

Table B2 - Diagnostic Tests:

Test	Model (1): Linear-Linear	Model (2)/(a): Log-Linear
RESET (Ramsey Test)	97.51 (0.000)	7.81 (0.000)
Bresuch-Pagan Test	25.47 (0.000)	4.24 (0.000)

Note - f-stats are shown and p-values are in parentheses

Table B3 - Full Regression Results for population split by working for private company or not (with columns 2 and 3 showing the samples split by working for not private/private company, resp.)

Regression	(a)	(d)	(e)
Type of Model	Full (log-linear)	Work for public company	Work for priv company
age	0.116***	0.111***	0.116***
	(0.00409)	(0.00732)	(0.00490)
c.age#c.age	-0.00129***	-0.00123***	-0.00139***
	(0.0000474)	(0.0000842)	(0.0000569)
female	-0.401***	-0.325***	-0.444***
	(0.0130)	(0.0207)	(0.0170)
male	0	0	0
	(.)	(.)	(.)
secondary	0	0	0
	(.)	(.)	(.)
ug	0.220***	0.219***	0.217***
	(0.0152)	(0.0249)	(0.0191)
pg	0.304***	0.314***	0.298***
	(0.0184)	(0.0288)	(0.0240)
other	-0.160***	-0.0958	-0.166***
	(0.0540)	(0.125)	(0.0603)
white	0	0	0
	(.)	(.)	(.)
asian	-0.113***	-0.0305*	-0.164***
	(0.0264)	(0.0400)	(0.0345)
black	-0.0104	0.0459	-0.0629
	(0.0368)	(0.0559)	(0.0476)
mixed or other	-0.0428	-0.00369	-0.0736
	(0.0389)	(0.0535)	(0.0540)
North England	-0.238***	-0.164***	-0.295***
	(0.0268)	(0.0416)	(0.0348)
Mid England	-0.204***	-0.131***	-0.265***
	(0.0241)	(0.0371)	(0.0316)

South England	-0.177***	-0.163***	-0.198***
	(0.0260)	(0.0416)	(0.0332)
Wales & Northern Ire	-0.235***	-0.164***	-0.287***
	(0.0275)	(0.0413)	(0.0365)
Scotland	-0.200***	-0.141***	-0.249***
	(0.0312)	(0.0497)	(0.0390)
London	0	0	0
	(.)	(.)	(.)
dissatisfied	-0.0336*	0.0143	-0.656***
	(0.0187)	(0.0300)	(0.0237)
neutral	-0.0152	-0.00447	-0.0208
	(0.0138)	(0.0223)	(0.0176)
satisfied	0	0	0
	(.)	(.)	(.)
not legally binded	0	0	0
	(.)	(.)	(.)
legally binded	0.00699	-0.0109	0.0183
	(0.0147)	(0.0225)	(0.0194)
have no child	0	0	0
	(.)	(.)	(.)
have children	-0.103***	-0.156***	0.0687***
	(0.0153)	(0.0243)	(0.0195)
professional	0	0	0
	(.)	(.)	(.)
intermediate	-0.271***	-0.254***	-0.282***
	(0.0170)	(0.0254)	(0.0228)
routine	-0.520***	-0.509***	-0.524***
	(0.0166)	(0.0295)	(0.0204)
work for non private	0		
	(.)		
work for private	0.0761***		
	(0.0136)		
sleeptime	-0.0000578	-0.00000395	-0.000139262
	(0.000105)	(0.000167)	(0.000135)
_cons	8.249***	8.307***	8.348***
	(0.0966)	(0.169)	(0.116)
R^2	0.392	0.306	0.451
N	9309	3810	5499

Robust standard errors in parentheses * p < 0.1, ** p < 0.05, *** p < 0.01

Table B4 - Difference in wages by location (with London as the base category):

Location	Wage difference (%)
Scotland	-20.0
Wales and Northern Ireland	-23.5
Middle of England	-20.4
North of England	-23.8
South of England	-17.7

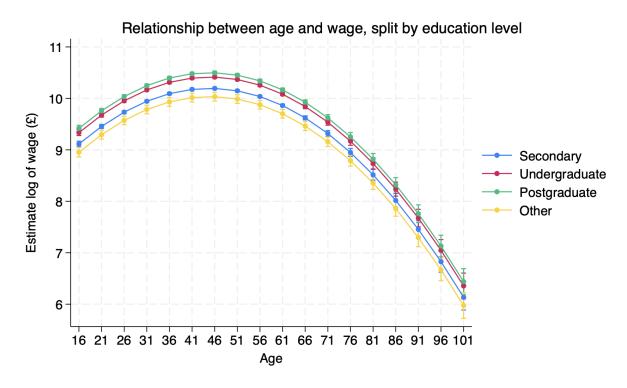
Graph B5 - Mapping showing the relationship between age and wage ($Y = Wage, X = Age; Age^2$) and deriving the maxima using FOCs:



$$\frac{\delta ln(wage)}{\delta age} = 0.116 - 0.00258age = 0$$

$$age = \frac{0.116}{0.00258} = 44.96 \approx 45 \text{ years old}$$

Graph B6 - Mapping showing the relationship between age and wage ($Y = Wage, X = Age; Age^2$, segregated by education level



We identify that postgraduate individuals always earn more compared to other levels of graduates

Table B7 - Full Regression Results for population split by gender (with columns 2 and 3 showing the samples split by gender)

Regression	(a)	(b)	(c)
Type of Model	Full (log-linear)	By Female	By Male
age	0.116***	0.109***	0.126***
	(0.00409)	(0.00533)	(0.00628)
c.age#c.age	-0.00129***	-0.00123***	-0.00139***
	(0.0000474)	(0.0000620)	(0.0000726)
female	-0.401***		
	(0.0130)		
male	0		
	(.)		
secondary	0	0	0
	(.)	(.)	(.)
	0.220***	0.242***	0.185***
ug	(0.0152)	(0.0212)	(0.0208)
	0.304***	0.335***	0.283***
pg	(0.0184)	(0.0262)	(0.0249)
	-0.160***	-0.148**	-0.157*
other	(0.0540)	(0.0697)	(0.0841)
	0	0	0
white	(.)	(.)	(.)
	-0.113***	-0.0662*	-0.164***
asian	(0.0264)	(0.0384)	(0.0361)
	-0.0104	0.0589	-0.155***
black	(0.0368)	(0.0481)	(0.0521)
	-0.0428	-0.00957	-0.0952*
mixed or other	(0.0389)	(0.0545)	(0.0555)
	-0.238***	-0.229***	-0.272***
North England	(0.0268)	(0.0358)	(0.0387)
1 torus Englusia	-0.204***	-0.228***	-0.188***
Mid England	(0.0241)	(0.0323)	(0.0350)
Titla Bilgiana	-0.177***	-0.203***	-0.151***
South England	(0.0260)	(0.0346)	(0.0379)
South England	-0.235***	-0.198***	-0.278***
Wales & Northern Ire	(0.0275)	(0.0366)	(0.0402)
wates & Northern ne	-0.200***	-0.194***	-0.218***
Caatland			
Scotland	(0.0312)	(0.0431)	(0.0434)
London	(.)	(.)	(.)
LUIIUUII	-0.0336*	-0.00982	-0.0902***
dissatisfied			
uissatisticu	(0.0187)	(0.0265)	(0.0245)
	-0.0152	-0.00996	
neutral	(0.0138)	(0.0194)	(0.0189)

satisfied	(.)	(.)	(.)
	0	0	0
not legally binded	(.)	(.)	(.)
	0.00699	-0.0484**	0.0795***
legally binded	(0.0147)	(0.0200)	(0.0207)
	0	0	0
have no child	(.)	(.)	(.)
	-0.103***	-0.210***	0.0164
have children	(0.0153)	(0.0219)	(0.0198)
	0	0	0
professional	(.)	(.)	(.)
	-0.271***	-0.319***	-0.187***
intermediate	(0.0170)	(0.0225)	(0.0252)
	-0.520***	-0.644***	-0.377***
routine	(0.0166)	(0.0239)	(0.0222)
	0	0	0
work for non private	(.)	(.)	(.)
	0.0761***	0.0378**	0.124***
work for private	(0.0136)	(0.0185)	(0.0196)
	-0.0000578	-0.00000752	-0.000139
sleeptime	(0.000105)	(0.000134)	(0.000163)
	8.249***	8.136***	7.893***
_cons	(0.0966)	(0.122)	(0.150)
R^2	0.392	0.349	0.399
N	9309	5155	4154

Robust standard errors in parentheses * p < 0.1, ** p < 0.05, *** p < 0.01