Regional differences in returns to higher education within the United Kingdom

Abstract

This study examines regional differences in the returns to higher education within the United Kingdom using data from the 2011 Quarterly Labour Force Survey (QLFS). The analysis employs Ordinary Least Squares (OLS) and Instrumental Variables (IV) regression to estimate the economic returns to schooling while addressing potential biases such as endogeneity and omitted variables. Key findings highlight substantial variation in returns across regions, with evidence suggesting that OLS underestimates the true effect of education on wages. These results contribute to understanding geographic disparities in educational outcomes, providing insights for policymakers focused on regional economic inequality and education policy.

Student ID: 210006045

Date: 12 December 2024

Introduction

Education is one of the most significant investments individuals make in their lifetimes, with profound implications for lifetime earnings, career prospects, and overall economic mobility. However, decisions regarding educational attainment are rarely made in isolation. They are influenced by expectations about the returns to education, which often vary depending on regional labor market conditions. In the United Kingdom, a commonly held belief among students and prospective workers is that London provides substantially higher wages compared to other regions, prompting many to pursue higher education with the expectation of accessing the capital's lucrative labor market. This project investigates whether such expectations are substantiated by empirical evidence, focusing specifically on the regional differences in returns to higher education within the UK.

The primary research question addressed in this study is: **Do returns to higher education differ significantly between London and other regions in the UK?** The hypothesis underpinning this research is that the returns to secondary and higher education are higher in London due to its concentration of high-skilled industries, professional job opportunities, and higher living costs, which may inflate wages. Understanding these regional disparities is critical for informing educational and career decisions, as well as for shaping policies that aim to reduce regional inequality.

Previous research has extensively analysed the returns to education using frameworks such as the human capital model developed by Becker (1964) and Mincer (1974). These studies emphasize that education enhances productivity, leading to higher wages over time. However, estimating the causal impact of education on wages is challenging due to biases in conventional methods, such as Ordinary Least Squares (OLS) regression. Factors such as ability bias and measurement error can lead to biased and inconsistent estimates. For instance, Harmon and Walker (1995) note that ability bias can inflate OLS estimates, while measurement error typically introduces downward bias. Additionally, David Card (1993) highlights the importance of geographic variation in education access and labor market conditions when analysing returns to schooling.

Building on this literature, this project uses the Mincer earnings equation as the baseline empirical framework. The analysis incorporates years of schooling, educational attainment milestones (secondary and higher education), work experience, gender, and regional indicators as key explanatory variables. To address endogeneity concerns in the relationship between education and earnings, an instrumental variable (IV) approach is employed, using the mother's occupation (ses_rank) as a proxy for socioeconomic status. This IV strategy aims to isolate the causal effect of education on wages, while also capturing the role of regional labor market dynamics through interaction terms for London and education levels.

The main findings of this study reveal significant regional disparities in returns to education. While OLS estimates suggest a 4.1% return for each additional year of schooling, the premiums for secondary and higher education are notably higher in London compared to other regions. These results are consistent with the hypothesis that London's labor market offers greater rewards for educational qualifications. However, the instrumental variable analysis yields imprecise and unreliable estimates, largely due to the weak instrument problem, with an F-statistic of 2.43. The limitations of the IV approach underscore the challenges of identifying robust causal effects in this context.

In comparison to the existing literature, the findings align with prior studies emphasizing regional differences in labor market returns to education. However, they also highlight the critical need for stronger instruments and richer data to accurately estimate the causal impact of education. This study contributes to the ongoing debate on regional inequalities in the UK by providing empirical insights into the differential valuation of education across regions. These findings have important implications for both students making educational decisions and policymakers designing strategies to reduce regional disparities.

Literature Review and Model

The human capital model, extensively developed by Becker (1964) and Mincer (1974), provides a theoretical framework for understanding the relationship between education, work experience, and earnings, building on earlier contributions by economists such as Adam Smith and Alfred Marshall. Central to this model is the premise that individuals invest in education and training to enhance their productivity, thereby leading to higher earnings over their lifetime. This model underpins the analysis of returns to education and serves as the foundation for the empirical framework used in this study. A key concept within the human capital model is the rate of return to schooling, which plays a crucial role in understanding educational attainment, labour market participation, wages, and income levels.

Estimating the returns to education, however, presents significant challenges due to potential biases in Ordinary Least Squares (OLS) estimates. A positive correlation between schooling and its return can introduce upward bias, as does ability bias, which arises when unobservable factors like innate ability are positively correlated with both schooling and wages. Conversely, measurement error in schooling data typically results in downward bias, which may offset or compound these effects. These complexities necessitate careful model specification and robustness checks to accurately estimate the true returns to education.

The concept of discount rates, integral to the human capital model, provides further insight into the relationship between education and wage outcomes. Harmon and Walker (1995) emphasize that higher discount rates—reflecting a stronger preference for present consumption—lead individuals to forgo further schooling. This behaviour compresses the wage-schooling locus, reducing the returns to education observed in regions where individuals have higher discount rates. These dynamics are directly applicable to this project's focus on regional disparities in returns to higher education within the UK. By extending the Mincer equation to include regional indicators, this study explores how differences in discount rates and educational investments across regions may influence earnings. For example, regions with lower average discount rates may exhibit higher educational attainment and returns to schooling, while regions with higher discount rates may experience lower participation in higher education and compressed returns.

Incorporating these perspectives from the literature enhances the robustness of the empirical strategy in this project. For instance, the addition of dummy variables for secondary and higher education allows this study to go beyond cumulative years of schooling and investigate the distinct returns to completing specific educational levels. By focusing on the difference in returns between secondary and higher education, this project offers a more granular understanding of the value of educational qualifications in the UK labour market. Moreover, linking the human capital model to empirical tools such as the IV approach and regional heterogeneity analysis ensures that the findings provide meaningful insights into the relationship between education, wages, and regional disparities.

Finally, David Card's insights on interpreting the "rate of return to schooling" within the Mincer framework are particularly relevant. The coefficient on the schooling variable can only be strictly interpreted as the rate of return under the conditions outlined in Mincer (1974). This project adopts this terminology to reflect the theoretical underpinnings of the human capital model, emphasizing that the schooling coefficient in the Mincer equation represents the marginal return to an additional year of education. By incorporating insights from these foundational studies, this project seeks to provide a rigorous and nuanced analysis of the returns to education by comparing outcomes between London and other regions of the UK.

Estimation Methods

For my baseline estimation method, I will use the Mincer earnings equation, where the dependent variable is the log wage of individuals. The independent variables include years of schooling, secondary education, higher education, work experience, experience squared, region, and the sex of the individual.

Equation 1

$$LogWage = \beta0 + \beta1schooling + \beta2SecondaryEducation + \beta HigherEducation + \beta3Experience + \beta4ExperienceSQR + \beta5Region + \beta6MALE + \varepsilon$$

A key element of the model is the inclusion of education variables. I explore two distinct approaches to defining years of schooling, while also incorporating dummies for secondary and higher education attainment. The secondary education dummy identifies whether an individual completed A levels or an equivalent qualification, while the higher education dummy captures whether an individual attended university at the undergraduate level or its equivalent. These variables are critical for understanding how different levels of education contribute to wage determination.

The first approach incorporates years of schooling as a measure of total formal education completed. For simplification, I assume that individuals who ended their formal education at the minimum compulsory schooling age completed 11 years, those with A levels or an equivalent qualification completed 13 years, and those who attended university or equivalent completed 16 years. In this framework, the coefficient on years of schooling represents the marginal return to each additional year of education, consistent with the logic of the Mincer equation, where total years of schooling are used to quantify cumulative human capital. For instance, an individual who completed A levels but did not attend university would have years of schooling equal to 13, secondary education equal to 1, and higher education equal to 0. In this setup, the dummies for secondary and higher education capture the additional returns to completing these specific levels of education, beyond the marginal effect of additional years of schooling. This approach disentangles the qualitative benefits of completing particular educational milestones from the quantitative impact of additional years of education.

In the second approach, years of schooling is excluded, and the focus is on capturing the total effects of secondary and higher education. Here, the coefficients for secondary and higher education reflect a combined effect, encompassing both the additional years of schooling required to achieve these levels and the qualitative benefits of completing these milestones. This approach is useful for examining the total returns to education without isolating the incremental effects of individual years of schooling.

In addition to education, work experience is a crucial determinant of earnings, and its inclusion significantly enhances the explanatory power of the model. I account for the non-linear relationship between experience and earnings by including both experience and its squared term. This captures the well-documented life-cycle pattern of wages: earnings typically rise rapidly early in a career as individuals acquire skills and expertise, before the rate of growth diminishes and may eventually plateau or decline due to factors such as skill depreciation or reduced mobility. Including these variables also prevents omitted variable bias, ensuring that the estimated returns to education are not conflated with the effects of work tenure.

Other control variables, such as a region dummy and a dummy for the sex of the individual, are also included in the model. The region dummy captures whether an individual resides in London or not, reflecting regional wage disparities that are important for understanding the geographic heterogeneity of labour markets. London, as a major economic hub, is associated with higher wages due to factors such as greater job opportunities and a higher cost of living. The inclusion of this variable allows for a clearer estimation of the effects of education and experience across different geographic contexts. Similarly, the dummy variable for sex (MALE) accounts for the potential wage gap between men and women, ensuring that the coefficients on education and experience reflect their effects net of gender differences.

This model, therefore, provides a robust framework for analysing the determinants of wages. By accounting for both years of schooling and educational milestones, as well as non-linear experience effects, regional disparities, and gender differences, the specification enables a nuanced examination of wage dynamics. The dual approach to defining education variables further allows for a deeper understanding of the distinct roles played by quantitative measures of schooling and qualitative effects of educational attainment.

Data

I use data from the Northern Ireland Statistics and Research Agency, Central Survey Unit, & Office for National Statistics, 2015 to be precise, the QLFS 2011 January-March sample, which initially consists of 101,645 observations and 774 variables. After accounting for missing data and exclusions, the final dataset used in the estimation includes 10,518 observations. The dependent variable in my regression is the log wage rate, which I derive from the variable HOURPAY, the recorded gross hourly pay of individuals. To address the skewness commonly observed in wage distributions, I log-transform this variable to create the variable LogWage. However, a significant portion of missing values in HOURPAY—90,598 cases—were labelled as "Does not apply," likely corresponding to individuals outside the labor force, such as the unemployed, students, retirees, or others not earning wages. These cases were excluded from the analysis, leaving a sample of individuals with valid wage data.

Education is a central focus of the analysis. The secondary and higher education variables are constructed using HIQUL11D, which provides detailed information on the highest qualification attained by an individual. The HigherEducation dummy is coded as 1 if an individual has attained a degree or higher education qualification and 0 otherwise. Similarly, the SecondaryEducation dummy is coded as 1 if an individual has attained GCE A-levels or equivalent qualifications.

For years of schooling, I make assumptions to align with the UK's educational system. In the UK, formal education begins at age 5 and ends at the minimum compulsory schooling age of 16. I derive years of schooling by subtracting 5 (the starting age) from the age of completing formal education. This measure reflects the total years of formal education completed. For example, individuals who ended their education at 15 are assumed to have completed 10 years of schooling, reflecting the UK rule that school-leaving age includes students who turn 16 during the summer following their final term.

To construct the experience variable, I use AGE, which records an individual's age, and EDAGE, the age at which an individual completed their formal education. I restrict the sample to individuals aged between 16 and 65 using a new variable AGE1, which reflects working-age individuals. Experience is calculated as the difference between AGE and EDAGE, representing potential years in the labor market since completing education. During this process, I encountered negative values for experience due to invalid responses in the EDAGE variable, such as "Does not apply," "No answer," "Still in education," and "Never had education." To address this, I created a cleaned version of EDAGE, labelled EDAGE1, which excludes these responses, ensuring that experience values are accurate.

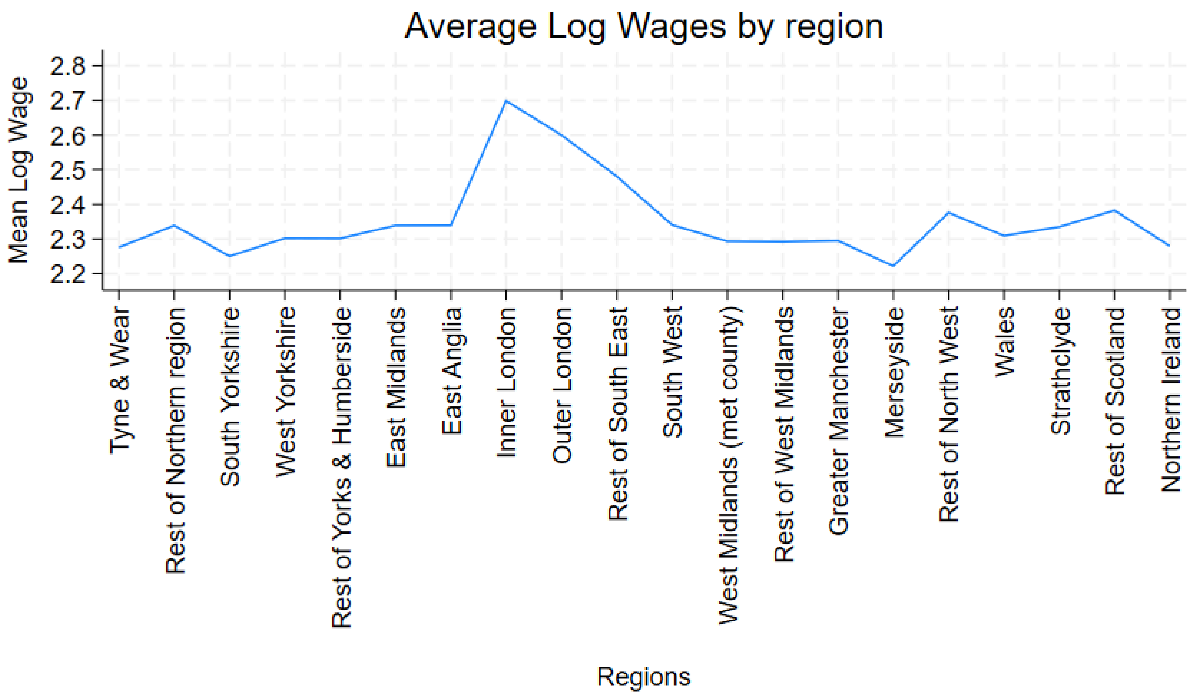
The region variable is a dummy that identifies whether an individual resides in London or not, capturing regional wage disparities and allowing for an analysis of geographic heterogeneity in labor markets. The sex variable is represented by a dummy (MALE) that identifies male individuals, controlling for potential gender-based differences in wages.

This dataset provides a comprehensive basis for analysing the relationship between education, experience, and wages while accounting for demographic and regional factors. The constructed variables and cleaning processes ensure the model captures meaningful variation in wages and minimizes biases due to invalid or missing data. Below are some summary statistics in Table 1 and Graph 1.

Table 1 Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
LogWage	11047	2.385	.586	-2.04	6.491
EDAGE1	65958	17.705	2.839	5	46
AGE1	66567	41.438	14.15	16	65
schooling	65958	12.705	2.839	0	41
SecondaryEducation	101645	.155	.362	0	1
HigherEducation	101645	.204	.403	0	1
Experience	60899	25.522	13.834	0	57
MALE	101645	.485	.5	0	1
Region	101645	.103	.303	0	1

Graph 1



Graph 1 shows us the mean log wages by region. We can see that the mean log wages in Inner and Outer London are significantly higher than in the rest of the UK which is consistent with our expectations of what these means should look like.

Results

In table 2 we can see the results from the OLS regressions that I ran with slightly different specifications. The baseline OLS regression can be seen in column 1 with the variables outlined previously in this paper. The coefficient on the schooling variable in Models 1 and 3 (column 1 and 3 in table 2) indicates a statistically significant and positive return of 4.1% per additional year of education. This return is slightly lower than the averages often reported in the literature, possibly reflecting regional differences or characteristics of the dataset. Another thing to consider is measurement error in our "schooling" variable. We made some simplifications for calculating the years of schooling which could lead to measurement error and as a result this might be why we have these lower OLS estimates for our $\beta1$ coefficient from Equation 1.

Table 2 Comparison of Regression Models

Variable	OLS1	OLS2	OLS3
schooling	0.041***		0.041***
	(0.002)		(0.002)
SecondaryEducation	0.160***	0.189***	0.157***
	(0.013)	(0.013)	(0.013)
HigherEducation	0.429***	0.578***	0.426***
	(0.014)	(0.011)	(0.014)
Region	0.178***	0.217***	0.153***
	(0.016)	(0.016)	(0.028)
MALE	0.201***	0.207***	0.201***
	(0.009)	(0.010)	(0.009)
Experience	0.041***	0.038***	0.041***
	(0.001)	(0.001)	(0.001)
Experience Squared	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)
londonhigheredu			0.038
			(0.035)
londonsecondaryedu			0.028
			(0.049)
Constant	1.043***	1.528***	1.045***
	(0.035)	(0.018)	(0.035)
R Squared	0.312	0.291	0.312
Observations	10518	10518	10518

Standard errors in parenthesis
* p<0.10, ** p<0.05, *** p<0.01

The dummy variables for SecondaryEducation and HigherEducation provide insights into the wage premiums associated with achieving specific education levels. In column 1 of table 2, which includes schooling as a covariate, the coefficients for SecondaryEducation (0.160) and HigherEducation (0.429) highlight the additional wage premiums conferred by these qualifications beyond the effect of years of schooling alone. We can see that the returns to higher education are significantly higher than those to secondary education indicating that a university degree or equivalent increases wages significantly more than secondary education. These findings reflect the "sheepskin effect," where completing an educational milestone, such as a secondary diploma or a university degree, yields additional labour market rewards beyond incremental years of schooling.

In column 2 of table 2, where schooling is excluded, the coefficients for SecondaryEducation (0.189) and HigherEducation (0.578) increase, capturing the combined effects of years of education and the attainment of educational milestones. This comparison underscores the distinct role of education milestones, suggesting that higher education confers a larger qualitative premium compared to secondary education.

The regional dummy variable is significant across all models, with a coefficient of 0.178 in Model 1 and 0.217 in Model 2. This indicates that residing in a region such as London is associated with an 18–22% increase in wages, controlling for other factors. Model 3 introduces interaction terms (londonhigheredu and londonsecondaryedu) to explore whether the returns to secondary and higher education differ in London compared to other regions. Although these coefficients are not statistically significant, their direction suggests a slightly higher premium for higher education in London (3.8%) compared to secondary education (2.8%). This finding, while inconclusive, hints at the need for more granular analysis to explore how regional labour markets influence the valuation of education.

The coefficients on gender (MALE) and experience are consistent and significant across all models. Being male is associated with a wage premium of approximately 20%, reflecting persistent gender wage disparities, which are well-documented in the literature. The quadratic specification of experience reveals the expected concave relationship with wages, where returns to experience diminish at higher levels.

For the R squared values we have a pretty consistent value across the 3 models which ranges between 0.29 and 0.32. These values are consistent with the literature. In the 3 models we have a high F statistic of 681.33, 718.27 and 530.01 in OLS1, OLS2 and OLS3 respectively.

Robustness checks and extensions

When estimating the returns to education using the baseline Mincer equation, the absence of a variable for ability introduces bias in the estimator $\beta1$. This bias arises because the covariance between education and ability is unlikely to be zero. By omitting ability, it becomes part of the error term, and since ability is correlated with education, the error term is also correlated with education. This violates the zero conditional mean assumption, leading to biased and inconsistent estimates.

To address this ability bias, I employed an instrumental variable (IV) approach. Upon an initial extensive review of the QLFS dataset, I discovered that it lacked suitable variables that could serve as valid instruments for education. To expand the scope of my analysis and identify a potential IV, I merged the QLFS dataset with the

Understanding Society, Wave 3, 2011-2012: Teaching Dataset from the UK Data Service. This merge provided access to a broader range of variables but came at the cost of reducing the sample size significantly, leaving me with 3,773 observations and 1,475 variables in the merged dataset.

For an instrument to be valid, it must influence educational attainment without directly affecting earnings, aside from its impact through education. After reviewing the literature and conducting a detailed analysis of the merged dataset, I identified the mother's occupation as a suitable instrumental variable. Chevalier et al. (2013) emphasize the strong influence of parental background on children's educational attainment. Using a mother's occupation aligns with this framework, as it reflects parental socioeconomic status, which is known to shape schooling decisions.

To operationalize this, I took the detailed occupation variable for mothers and categorized it into broader socioeconomic subgroups, creating a new variable called ses_rank. This variable represents the socioeconomic status of the mother's occupation. Using ses_rank as an IV, I conducted a two-stage least squares (2SLS) regression to estimate the returns to education, addressing the ability bias inherent in the baseline regression model.

Table 3 Comparison of Regression Models

Variable	OLS1	OLS2	OLS3	IV Regression
schooling	0.041***		0.041***	0.761
	(0.002)		(0.002)	(0.669)
SecondaryEducation	0.160***	0.189***	0.157***	-0.354
	(0.013)	(0.013)	(0.013)	(0.511)
HigherEducation	0.429***	0.578***	0.426***	-2.286
	(0.014)	(0.011)	(0.014)	(2.524)
Region	0.178***	0.217***	0.153***	-0.965
	(0.016)	(0.016)	(0.028)	(1.007)
MALE	0.201***	0.207***	0.201***	-0.011
	(0.009)	(0.010)	(0.009)	(0.280)
Experience	0.041***	0.038***	0.041***	0.058*
	(0.001)	(0.001)	(0.001)	(0.034)
Experience Squared	-0.001***	-0.001***	-0.001***	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
londonhigheredu			0.038	0.292
			(0.035)	(0.545)
londonsecondaryedu			0.028	0.703
			(0.049)	(0.952)
Constant	1.043***	1.528***	1.045***	-7.939
	(0.035)	(0.018)	(0.035)	(8.452)
R Squared	0.312	0.291	0.312	--
Observations	10518	10518	10518	585

Standard errors in parenthesis
* $p<0.10$, ** $p<0.05$, *** $p<0.01$

Results from running the regression with an instrumental variable can be seen above in table 3 and need a very careful analysis. The coefficient for the returns to schooling is 0.761 with a standard error of 0.669. This is a drastic increase from our OLS estimates.

The fact that the IV coefficient on schooling is larger is consistent with literature and this implies that the OLS results understate the true return to schooling. As mentioned before this likely due to measurement error or ability bias. However this coefficient is not statistically significant and there is a very high standard error which limits confidence in interpreting this coefficient.

For the returns to higher and secondary education we have very strange results which indicate that upon completing secondary or higher education an individual would obtain a smaller wage. We run into the same unreliable coefficients with almost all of the variables in the 2sls regression with our instrumental variable.

Despite the many efforts to try and account for the endogeneity issue it has proven to be very difficult to deal with. The review of Angrist and Krueger (1991) led me to investigate into using birth month as a potential instrumental variable with no luck of success. As suggested by Harmon and Walker (1995), changes in school leaving age could be considered however, there haven't been any changes in such laws in recent years. With all the considerations and data available to me, the best I could find was the occupation of an individuals mother. This IV has proven to be weak with an F-statistic of 2.43. An F-statistic below 10 is widely regarded as indicative of a weak instrument, as outlined by Staiger and Stock (1997). After merging the two datasets, a significant number of observations were lost, leaving only 585 observations for the final 2SLS regression. This limited sample size is another reason why we were left with unreliable coefficient estimates for some variables in the IV regression, rendering them unsuitable for interpretation.

In the case of weak instruments, the first-stage regression produces a noisy and imprecise prediction of the endogenous variable, which propagates to the second stage, leading to unreliable estimates of the causal effect. With an F-statistic of 2.43, the instrument ses_rank in this analysis is far below the threshold, suggesting it lacks sufficient explanatory power. Weak instruments result in inflated standard errors, as seen in the 2SLS regression where the coefficient for schooling has a standard error of 0.669, rendering it statistically insignificant. These issues highlight the need for stronger instruments or alternative methodologies to address the bias effectively.

Discussion and Conclusion

The results of this study highlight important insights into the returns to education, particularly the roles of secondary and higher education in influencing earnings across different regions, with a specific focus on London versus other parts of the UK. While the instrumental variable (IV) approach using mother's occupation (ses_rank) attempted to address the endogeneity issue inherent in Ordinary Least Squares (OLS) estimation, the findings reveal limitations in both the strength of the instrument and the precision of the estimates. Nonetheless, the regional disparities uncovered in the analysis have significant economic and policy implications.

The findings of this study, particularly the higher returns to education observed in London, must be understood within the broader economic context of 2011. During this period, the UK was still recovering from the 2008 financial crisis, with unemployment reaching its highest level in over a decade. Despite these challenges, London demonstrated significant economic resilience, maintaining a robust labor market characterized by a concentration of high-skilled industries and relatively higher wages. This economic environment likely amplified the returns to secondary and higher education in the capital, as the demand for highly qualified workers remained strong.

In contrast, regions outside London experienced higher unemployment and fewer opportunities in high-paying industries, which likely contributed to the lower returns to education observed in these areas. The stark differences between London and other regions highlight the importance of geographic location in determining the value of educational qualifications. Policymakers should consider how regional economic disparities influence the returns to education and explore strategies to foster high-skilled job creation and labor mobility across the UK. By doing so, the benefits of higher education can be more evenly distributed, reducing reliance on London as the primary economic hub.

Drawing from David Card's (1993) work, it is essential to consider the potential downward bias in OLS estimates due to measurement error in schooling. This aligns with the literature suggesting that IV estimates may be larger than OLS estimates, capturing the causal effect of education on earnings more accurately. However, in this study, the weak instrument problem, as reflected in the F-statistic of 2.43, limits confidence in the IV results. The large standard errors in the IV regression and counterintuitive coefficients for secondary and higher education suggest that the instrument does not adequately capture the exogenous variation required to isolate the true causal effect. This aligns with the challenges Card discusses when using less robust instruments, emphasizing the need for careful interpretation of IV estimates.

The regional focus of this study underscores the importance of considering geographic disparities in the returns to education. The findings suggest that secondary and higher education yield greater economic benefits in London compared to other regions. This is likely due to London's unique labor market dynamics, characterized by a high concentration of professional and high-skilled jobs, which amplify the returns to educational qualifications. However, the elevated cost of living in London offsets some of these benefits, raising important questions about the real returns to education for individuals residing in the capital.

For policy, these results have profound implications. Investments in secondary and higher education remain critical for enhancing individual productivity and earnings potential, particularly in regions outside London where the returns may be smaller due to limited access to high-paying jobs. Policymakers should prioritize regional economic development initiatives that create high-skilled employment opportunities beyond London, thereby narrowing the gap in returns to education between regions. In London, policies to address the high cost of living—such as affordable housing initiatives or transportation subsidies—could enhance the net benefits of higher education.

In conclusion, while this study faced significant challenges due to weak instrumentation and data limitations, the findings reinforce the critical role of secondary and higher education in shaping economic outcomes, particularly in London's labor market. Addressing regional disparities in returns to education and mitigating structural inequalities will require coordinated efforts across education, labor, and social policy sectors. Future research should focus on identifying stronger instruments and leveraging natural experiments, such as policy reforms, to provide more robust insights into the causal effects of education on earnings.

References

Angrist, J. D., & Krueger, A. B. (1991). Does Compulsory School Attendance Affect Schooling and Earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014. https://doi.org/10.2307/2937954.

Becker, G. S. (1964). *Human capital: A theoretical and empirical analysis, with special reference to education*. University of Chicago Press.

Card, D. (1993). *Using geographic variation in college proximity to estimate the return to schooling* (NBER Working Paper No. 4483). National Bureau of Economic Research. https://doi.org/10.3386/w4483.

Chevalier, A., Harmon, C., O'Sullivan, V., & Walker, I. (2013). The impact of parental income and education on the schooling of their children. *IZA Journal of Labor Economics, 2*(1), 1–22. https://doi.org/10.1186/2193-8997-2-8.

Harmon, C., & Walker, I. (1995). Estimates of the Economic Return to Schooling for the United Kingdom. *The American Economic Review*, *85*(5), 1278–1286. http://www.jstor.org/stable/2950988.

Mincer, J. (1974). *Schooling, experience, and earnings*. Columbia University Press.

Northern Ireland Statistics and Research Agency, Central Survey Unit, Office for National Statistics, Social Survey Division. (2015). *Quarterly Labour Force Survey, January – March, 2011*. [data collection]. *4th Edition.* UK Data Service. SN: 6782, DOI: http://doi.org/10.5255/UKDA-SN-6782-4.

Staiger, D., & Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments. *Econometrica*, *65*(3), 557–586. https://doi.org/10.2307/2171753.

University of Essex, Institute for Social and Economic Research, University of Manchester, Cathie Marsh Centre for Census and Survey Research, UK Data Service. (2014). *Understanding Society, Wave 3, 2011-2012: Teaching Dataset*. [data collection]. University of Essex, Institute for social and Economic Research, [original data producer(s)]. University of Essex, Institute for Social and Economic Research. SN: 7549, DOI: http://doi.org/10.5255/UKDA-SN-7549-1.