The Casual Effect of Education on Earnings in the United States

Bayesian Inference

I. Introduction

The education has long been commanded as a premier investment on human capitals, with an established economic importance and conspicuous implications on lifetime earnings (Becker 1964, Mincer 1974, and Card 1999). In spite of the rapid accumulation of empirical observations in many different countries and time periods confirming that more-educated individuals earn higher wages than their less-educated counterparts, an education premium is gradually losing its lustre amid the rising college tuition in the United States married with mix of industrial shifts. The trend has culminated in a new social phenomenon "*Toolbelt Generation"*, where more youngsters in the US are turning down traditional college paths in a disenchantment – choosing tools over textbooks (Chen 2024).

The aim of this paper is to scrutinize the heterogeneity in returns to education, probing beyond the conventional averages to uncover the stratified impacts of an education premium. By wielding a dataset rich in demographic granularity, encompassing variables such as location, gender, and ethnicity, the analysis is tailored to discern the contextual nuances that shape the economics of education. The approach seeks to illuminate how educational returns fluctuate in response to varying socio-economic landscapes, this providing a nuanced understanding that resonates with the 'Toolbelt Generation's' diverging individual-specific path. In doing so, I aim to furnish policymakers, educators and individuals concerned with the investment decisions of profound importance with insights that could better align educational strategies with the shifting contours of economic demand.

II. Data

This section describes the data used in the empirical analysis. I used IPUMS USA individual census and survey data in 2022 that reports various socioeconomic information about an individual at the period of investigation. The main variable of interest refers to salary-based income, years of education and other potential confounding variables such as age, working hours, race, and state-level location.

TABLE 1

Means and standard deviations

	Overall			Male			Female		
	(1) Observed	(2) Mean	(3) S.D.	(4) Observed	(5) Mean	(6) S.D.	(7) Observed	(8) Mean	(9) S.D.
Salary-based Income (US $)	1346155	71079	79332	704582	82339	91298	641573	58713	61310
Age	1346155	44.22	14.57	704582	44.24	14.66	641573	44.21	14.47
Average Working Hours in a week (hours)	1346155	39.63	11.66	704582	41.57	11.56	641573	37.49	11.39
Education (years)	1346155	14.46	2.54	704582	14.27	2.55	641573	14.66	2.51

Table 1 reports descriptive statistics for variables I used to establish the empirical results in the paper. The heterogeneity of gender-variant descriptive statistics reported in Table 1. Particularly, female respondent has a nearly more half year education than their male counterparts on average in 2022 survey data, but they earned less than three-fourths of what man earned. The presence of heterogeneous returns to schooling associated with gender status paves a way for a multi-faceted empirical analysis into the data. Whereas Figure 1 displays a consistently positive returns on education for both genders, as illustrated by upward trending box plots.

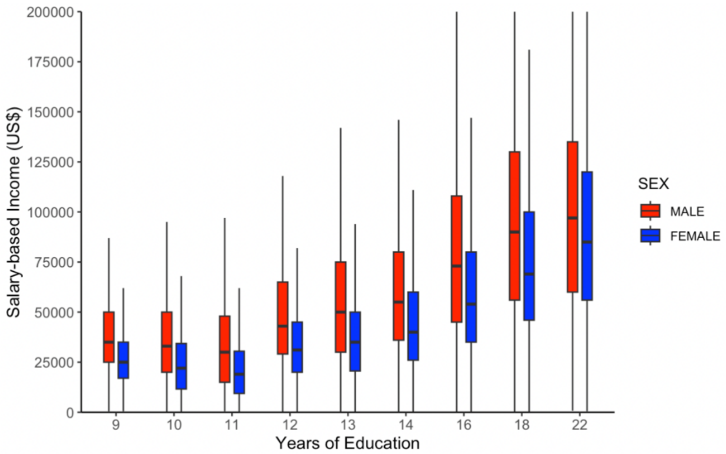

Figure 1. Salary-based Income by Years of Education

It is worth noting that there is a considerable heterogeneity on the state-level, suggesting an additional layer of hierarchy within the data structure and thereby rationalising the use of hierarchical models to better survey the contour of the issue. Figure 2 and 3 present kernel probability distributions that illustrate variations in salary-based income by state and the relationship between salary-based income and years of education across states, respectively. These visual representations corroborate the existence of a state-level hierarchy, underpinning the significant variability observed on a regional basis. Employing hierarchical models will facilitate a delineation of the nuanced stratifications within the data and thereby provide a more granular analysis of the issue. Given the constraints imposed by the word limit, this analysis has been streamlined to focus on the most critical aspects: variations by gender and state. Hence, in-depth empirical works to explore more complex dimensions is not covered in this paper. In order to account for racial variation within the data, fixed effects for each race were incorporated into the regression analysis. This approach effectively controls for unobserved heterogeneity by the racial differences, isolating the impact of the independent variables by comparing within-race variations, thereby enhancing the robustness of the estimates.

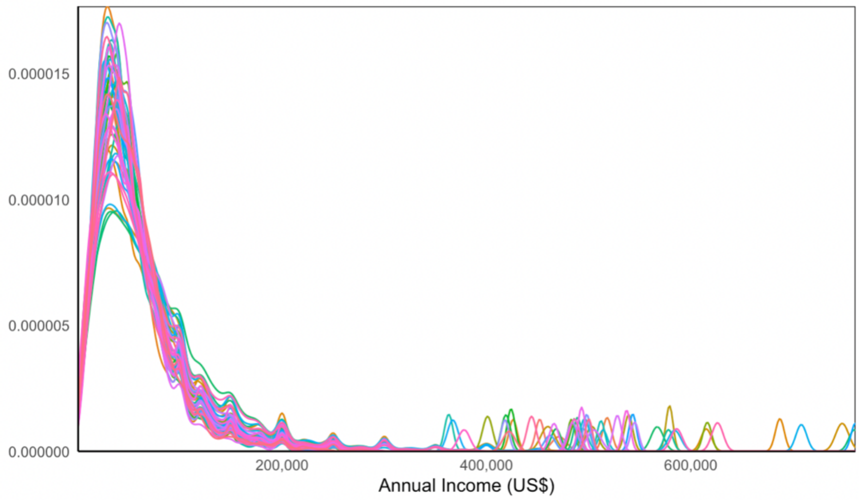

Figure 2. Kernel Density of State-Variant Salary-based Income

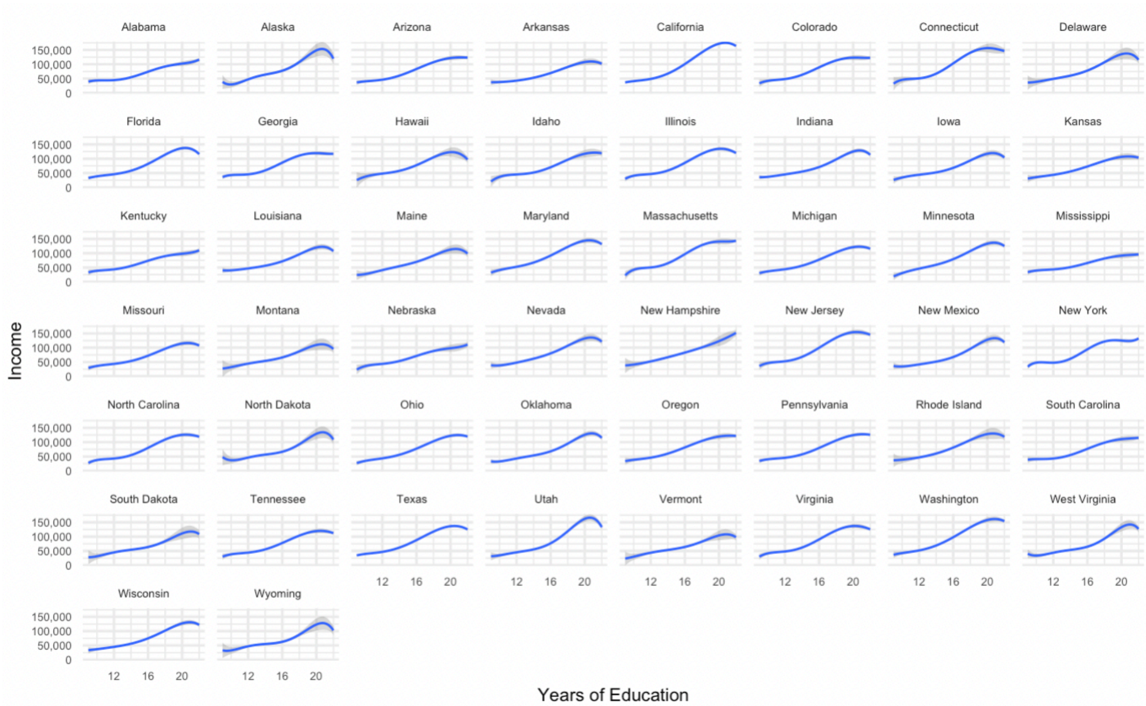

Figure 3. State-Variant Salary-based Income by Years of Education

III. Results

A. *Frequentist Linear Regression*

Table 2 summarises the regression results on the effect of education on income with frequentist approaches from a pooled data. For columns (1) and (2), I present OLS regression results, illustrating the implications of addition year of education on salary-based income. For regression results in column (2), I controlled for age and average working hours in a week as well as using state and race fixed effects to account for state-variant and race-variant nature of the income. For columns (3) and (4), I used a Ridge and Lasso regression with the same model specification from the OLS reported in column (2).

TABLE 2

Frequentist Regression Estimates: Effects of Education on Salary-Based Income

	(1) OLS (simple)	(2) OLS (Pooled)	(3) Ridge	(4) Lasso
Intercept	-78687.80***	-167260.35***	-167811.13	-167821.60
	(373.42)	(925.75)		
Education (Years)	10358.55***	8307.16***	8353.89	8354.70
	(25.44)	(24.79)		
Controls				
Age		445.83***	442.21	442.26
		(4.95)		
Average Working Hours in a week (hours)		1868.44*** (5.24)	1870.62	1870.80
Lambda			9.3260	0.3765
Fixed Effects				
State FE	NO	YES	YES	YES
Race FE	NO	YES	YES	YES
Observations	1346155	1346155	1346155	1346155

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2 presents compelling evidence of statistically significant returns on investment in education, with robustness even at the 1% significance level. Upon controlling for variables such as age, race, and state-level variation, the effect size is determined to be $8307.16. This signifies that, on average, each additional year of education is associated with an $8307.16 increase in annual wage-based income. Supporting this finding, both Ridge and Lasso regression analyses concur, yielding coefficients of $8353.89 and $8354.70 respectively, thus reinforcing the original conclusion regarding the economic value of educational attainment.

Table 3 reports the regression results on the returns on education with frequentist approaches from a separated data by gender, thus offers a gender-specific examination of the economic returns to educational investments. Columns (1) and (2) delineate the OLS regression results for data corresponding to males, while columns (3) and (4) detail the OLS regression outcomes for data pertaining to females.

TABLE 3

Frequentist Regression Estimates: Effects of Education on Salary-Based Income (by Gender)

	Male		Female	
	(1) OLS (simple)	(2) OLS	(3) OLS (simple)	(4) OLS
Intercept	-78687.80***	-188934.27***	-63386.46***	-128987.12***
	(577.01)	(1472.73)	(427.06)	(1030.03)
Education (Years)	12941.50***	10615.43***	8327.24***	6507.24***
	(39.80)	(39.43)	(28.71)	(27.96)
Controls				
Age		445.39***		311.03***
		(7.83)		(5.55)
Average Working Hours in a week (hours)		1793.44***		1633.84***
		(8.43)		(5.99)
Fixed Effects				
State FE	NO	YES	NO	YES
Race FE	NO	YES	NO	YES
Observations	704582	704582	641573	641573

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3 offers convincing evidence of significantly positive returns on educational investment, which, however, exhibit considerable variation between genders. Upon controlling for variables such as age, race, and state-level variation, the effect size for male is determined to be \$10615.43, whereas \$6507.24 for female counterparts. That says, an additional year of education correlates with \$10615.43 increase in annual income for American men, whereas for American women, the increase is associated with \$6507.24.

B. *Bayesian Linear Regression*

While results from Bayesian regression largely agrees with a significantly positive returns on educational investment, the Bayesian interpretation of credible intervals offer a scope of calibrating the precise range within which the returns to education are situated. Columns (1) and (2) presents the Bayesian regression results from pooled data, illustrating the implications of addition year of education on salary-based income. Columns (3) and (4) delineate the Bayesian regression results for data corresponding to males and females, respectively.

TABLE 4

Bayesian Regression Estimates: Effects of Education on Salary-Based Income

	(1) Simple Model	(2) Pooled Model	(3) Separate Model (Male)	(4) Separate Model (Female)
Intercept	-78687.0	-231857.8	-188884.4	-133957.9
	(-79406.4, -77967.3)	(-240230.7, -165959.2)	(-270295.1, -141059.6)	(-189518.9, -127469.5)
Education (Years)	10358.6	8307.6	10615.5	6506.8
	(10310.6, 10407.7)	(8248.0, 8366.9)	(10533.5, 10698.4)	(6441.8, 6573.7)
Controls				
Age		445.8	445.9	311.1
		(433.5, 457.5)	(429.5, 461.4)	(297.9, 323.4)
Average Working Hours in a week (hours)		1868.4	1793.4	1633.8
		(1855.0, 1881.7)	(1776.3, 1810.3)	(1620.1, 1647.4)
Fixed Effects				
State FE	YES	YES	YES	YES
Race FE	YES	YES	YES	YES
Observations	1346155	1346155	1346155	1346155

95% Credible Interval in parentheses

Column (2) indicates that the coefficient for education stands at 8307.6, which aligns closely with the 8307.16 figure obtained from the frequentist OLS method in Table 2. Bayesian credible intervals offer a highly precise, narrow range (8248.0, 8366.9), within which there is a 95% probability that the true return on an additional year of education is contained. Moreover, column (3) reveals that the Bayesian coefficient for education is 10615.5 when considering data for males, closely mirroring the 10615.43 result from the frequentist OLS method reported in Table 3. Similarly, column (4) shows that the Bayesian coefficient for education is also 6506.8 for females, which is consistent with the frequentist finding in Table 3. The reported ranges with which the true return on an additional year of education is contained with 95% probability are (10533.5, 10698.4) and (6441.8, 6573.7) for American men and women, respectively.

Hence, the congruence between the frequentist and Bayesian regression results underscores the robustness of the estimations regarding gender-variant returns to education. Additionally, the Bayesian credible intervals enhance the interpretability of the results by providing a concrete range within which the true parameters are believed to reside with high probability.

C. *Hierarchical Model*

Table 5 presents the results from a hierarchical model, corroborating the claims initially posited regarding state-level variation in the returns to education. Especially, multilevel models with random slopes and both intercepts and slopes are provided in column (2) and (3).

TABLE 5

Multilevel Model: Effects of Education on Salary-Based Income by States

	(1) Random Intercepts Model	(2) Random Slopes Model	(3) Random Intercepts + Slopes Model
Education (Years)	10675.9		
	(293.5)		
State-Specific Effects			
California		15600.8	15605.0
		(745.8)	(723.8)
Texas		10269.4	10270.0
		(830.7)	(835.7)
Florida		10006.0	9972.3
		(957.5)	(963.1)
New York		11078.2	11091.1
		(941.2)	(925.1)
Pennsylvania		7979.2	7941.0
		(1105.9)	(1104.9)
Illinois		12308.2	12323.6
		(1222.1)	(1204.7)
Ohio		6635.9	6660.9
		(1303.8)	(1280.0)
Georgia		9566.9	9603.6
		(1292.2)	(1260.3)
North Carolina		9560.2	9585.9
		(1275.1)	(1290.6)
Michigan		7796.9	7787.4
		(1427.7)	(1455.0)
Control			
Sex	-29141.3	-29078.0	-29085.5
	(1569.9)	(1466.9)	(1475.8)
Observations	10,000	10,000	10,000

Standard deviation in parentheses

Table reports state-specific effects for the top 10 most populous states in the United States as of Dec 2023

As demonstrated in Table 5, the returns on educational investment are significantly positive but exhibit considerable variation from state to state. Specifically, Column (2), which details the estimations from a hierarchical model with random slopes, indicates that the returns on additional year of education extend from $6635.9 in Ohio to $15,600.8 in California. Column (3), which outlines the estimations form a hierarchical model incorporating both random intercepts and slopes, substantially aligns with the findings, demonstrating state-dependent variation in the returns to education ranging from $6660.9 in Ohio to $15,605.0 in California.

IV. Conclusion

In conclusion, the inquiry has revealed that the returns on educational investment are not monolithic but are instead influenced by variables such as gender and geographic location. This nuanced view is critical for individuals considering the economic value of furthering their education. This gender and location variant nature of educational returns illustrates the intricate interplay between education and the broader economic and industrial landscapes in the United States. It is evident that utilizing an averaged value to gauge the benefits of education does not capture the complex reality faced by individuals with diverging circumstances. Such findings not only speak to the evolving narrative of the 'Toolbelt Generation' but also serve as an important guidepost for policy formulation and educational guidance. Indeed, a granular approach that acknowledges the distinct pathways in earnings between genders and among states is imperative for accurate assessment and accurate policymaking.

Future extensions of this research should consider integrating additional variables such as racial factors, marital status, and individual choices in education, which may offer a fuller depiction of the economic implications of educational pursuits. This more comprehensive analysis promises to refine our understanding and help tailor educational strategies to meet the nuanced demands of our evolving economy.

Reference

Becker, Gary S. (1964), Human capital: a theoretical and empirical analysis, with special reference to education (Columbia University Press, New York).

Card, David (1999). The causal effect of education on earnings. *in Handbook of Labor* Economics (pp. 1801-1863)

Chen, Te-ping. (2024). “How gen Z is becoming the toolbelt generation.” The Wall Street Journal, April 1, 2024. https://www.wsj.com/lifestyle/careers/gen-z-trades-jobs-plumbing-welding-a76b5e43.

Mincer, Jacob. (1974). Schooling, experience, and earnings. New York: NBER.