Changes in Wage Returns to Education: A Comparative Regression Analysis of the United States and Indonesia in the 1970s and 1990s

By Pranav Sundararajan

Abstract: This paper compares the wage returns to education between the United States of America and the Republic of Indonesia in two historical periods (the 1970s and 1990s) via a log-level regression of Mincer earnings functions to explore how the present relationship between education and income in both countries has been shaped by past trends in education and returns in the labor market. In essence, it asks the following research question: To what extent does an additional year of schooling affect wage returns in the United States compared to Indonesia in the 1970s and 1990s? The research finds that an additional year of schooling brings statistically more significant wage returns in Indonesia compared to the US in the 1970s and 1990s, by around 3 to 6% purely if a percentages-based relative comparison is made. Within each country, however, the effects of education on wage returns between the 1970s and 1990s vary. Returns to education for an additional year of schooling fall by 2.5% in Indonesia between 1976 and 1995. However, the returns rise in the US by 1.6% between 1970 and 1990. In consideration of other covariates, the effect of gender on wage returns is more pronounced compared to potential labor market experience and hours worked per week by the individual in question.

Introduction

Education plays an indisputably powerful role in today's modern societies and economies, particularly since the Second World War and the late 20th century as the number of years an individual would spend in school (i.e. a formal education system) increased across all low, low-middle, upper-middle-, and high-income countries. Economists disagree as to whether formal years of schooling have a causal effect on individuals' wages earned later in life as they participate in the labor market. Prevailing literature outlines theories surrounding education as human capital investments because education enhances one's ability and skills that cause them to be productive later when they enter the job market. The regressions conducted corroborate with the prevailing literature on education being
observed to be a key factor in economic growth (Barro, 2013). For lower-middle income countries in particular, access and investment into higher education attainment has shown to be positively correlated with national economic development, and individual ability and productivity (Kim & Mohtadi, 1992).

Understanding the relationship between education and earnings is vital to understanding a variety of social issues that both the United States and Indonesia face, namely gender, ethnic, and racial discrimination, and inequitable distributions of income. In this regard, it was important to demonstrate a comparison between two countries on different ends of the income and development spectrum, but with roughly similar population figures in the present day. According to the World Bank, the United States is presently and historically classified as high-income country, while Indonesia is presently and historically been classified as lower-middle income country. (World Bank Country and Lending Groups, 2022). Moreover, the large populations of the United States at 331,893,745 (Population, Total - United States | World Bank Data, 2021) and Indonesia’s at 276,361,788 (Population, Total - Indonesia | World Bank Data, 2021) is beneficial, as the results of this analysis will have high external validity as opposed to an analysis of wage returns to education for two economies with small populations.

**Literature Review**

Past research on wage returns to education in the United States arrive at the consistent conclusion that those who attain higher education qualifications significantly stand to gain more in the labor market returns than others (Hout, 2012). Greater worker productivity and heterogeneous wage returns that are unexplained by the attainment of tertiary education and university degrees (Brand & Xie, 2010) mean that there may be other factors that need to be investigated, namely the years of labor market experience one may have. Additionally, within the empirical analyses, issues surrounding the selection of meaningful regressors and the low external validity of many of the earnings functions often come up. These issues make quantification of the causal effect of education on wage returns a tedious undertaking (Hout, 2012).

Wage returns to education in Indonesia, while investigated on later than in the United States, show similar results over much of the 21st century. With the ushering of market-oriented economic reforms in the late 1980s expecting to lead to “greater alignment of wages with productivity-related characteristics” (Purnastuti et al., 2015), estimates have been made on the returns to education (Behrman & Deolalikar, 1993; Duflo, 2001), which show that as a human capital investment, schooling is profitable. This complements the finding that the returns to schooling increase with the level of education (Purnastuti et al., 2013); however, these returns are not as high when compared to surrounding countries in the Southeast Asia region. This is because of significant variations in the rate of return to education (and subsequent wage gaps) between each income class, which have been found to contribute to rising wage inequality in Indonesia (Widyanti, 2018). While wage
returns on education may vary across gender, urban and rural geographical areas, industries of employment, and years of labor experience an individual may have, previous literature has shown that attaining higher education qualifications had directly increased productivity, which lead to higher earnings (Mincer, 1958). It is also hypothesized in the literature that higher education attainment indirectly increases productivity (and hence earnings) because of credential effects, stemming from the idea being that higher education merely certifies, not augments, one’s productivity (Card, 1999). This expounds two main theories of human capital that are at play here: the human capital hypothesis, and the screening hypothesis. The human capital hypothesis is the idea that schooling provides the skills for one to be productive. Earnings are brought about by higher productivity, which are themselves caused by the expenditures one makes in education. (Becker, 2002). While the screening hypothesis is the idea that employers select those who have pursued higher education purely because this minimizes their risk of potentially hiring someone who does not have high education backgrounds (Layard & Psacharopoulos, 1974). Here, higher wages may not be entirely attributed to high productivity, but rather to various heterogeneous factors, namely compensating differentials, or the perceived societal reputation of a certain occupation.

Data Sources & Methodology

Microdata on income/wages and educational attainment was collected for both countries in the 1970s and 1990s using the Integrated Public Use Microdata Series (IPUMS) International platform. A total of 4 datasets were collected for both countries and periods. The primary variables of interest selected with the help of IPUMS in each of the micro datasets were the observations’ monthly or annual wage and salary income, educational attainment in terms of the level of schooling completed, and their highest grade/level of schooling the respondent completed, in years, that is, years of study. In addition to wage and educational attainment variables, data on observations’ age, sex, the industry of employment, the status of their employment either as a wage/salary worker or self-employed person, and hours worked per week were also collected, to potentially examine the effects of other covariates.

For Indonesia in the 1970s, microdata was sourced using IPUMS International from the Intercensal Population Survey of 1976 (indon76), a sample of households in 770 clusters throughout the country that amounted to 281,170 total observations. Data from the 1990s for Indonesia was sourced from the Intercensus Population Survey of 1995 (indon95), a questionnaire of dwelling information including social and demographic characteristics that covered 718,837 individuals. Any observations with missing values, unknown, or denoted as “Not In Universe (NIU)”2, for the above-mentioned dependent variables were dropped. In addition, due to the presence of top codes for 16+ years of schooling in the indon76 dataset and 17+ years of schooling in the indon95 dataset, the years of schooling variable had to be adjusted. This was done by adding one year to the top-coded value.
In addition, the years of schooling had to be estimated for partial education attainment categories, based on Indonesia’s 6-3-3 system of years a pupil must spend in primary, middle secondary, and higher secondary education. For any observations coded in as 91 (indicating some primary education), 92 (indicating some technical education after primary education), 93 (indicating some secondary education), and 94 (indicating some tertiary education), years of schooling was estimated to be 3, 7.5, 9, and 14 years of schooling. Both indon76 and indon95 datasets were then filtered by age and employment status to include those above the age of 25 and below the age of 65 who were employed to focus the analysis solely on those who had passed the general age of matriculation from tertiary education and were working at the time when the survey was conducted. Finally, the monthly income of the observations was recorded in Indonesian Rupiah (IDR) for the indon76 and indon95 dates, while annual income was recorded in United States Dollars (USD) in the US datasets; therefore, to keep the income measurements consistent across all datasets, each observation’s income in the Indonesia datasets was multiplied by 12 to acquire annual income values.

For the United States in the 1970s, microdata was sourced using IPUMS International from the 1970 Census of Population and Housing, a 1-in-100 national random sample drawn by the US Census Bureau covering 2029666 observations (US70). Data for the United States in the 1990s was sourced from the 1990 Census of Population and Housing (US90), a 1-in-20 national random sample drawn by the US Census Bureau covering 12,501,046 individuals. A similar process to clean and process the data for the indon76 and indon95 datasets was applied to the US70 dataset. As the years of schooling variable was not top coded in both US70 and US90 datasets, it sufficed to drop all unknown observations, missing or NIU.

The primary manipulation made to the US70 and US90 datasets were in the hours worked per week variable, which—unlike the hours worked per week variable in the Indonesia datasets—signifies an interval of hours worked per week. To convert those intervals into a singular number of hours worked per week, the mean of the interval was taken. So, observations were coded in as 1 (indicating that the individual works 1 to 14 hours weekly), 2 (15 to 29 hours), 3 (30 to 39 hours), and 4 (40 to 48 hours), the hours worked per week for US observations was estimated to be 7.5, 22, 34.5, and 44 hours, respectively. For observations coded as 5 (49 hours or more), hours worked was estimated to be 58.5. This was calculated by taking the mean of the interval between 49 and 67 hours, a set interval range similar in scope to its IPUMS data codes that also take labor-leisure considerations into account. US70 and US90 were then filtered in the same fashion as the Indonesia datasets for age and employment status.

Further manipulation was done on the US90 dataset, as it did not contain the years of schooling variable. Therefore, a years of schooling variable was created using the educational attainment data. With the education attainment variable representing the highest level of attainment of the
observation, years of schooling would represent the minimum years of schooling that the observation would have completed. For observations with less than primary education completed ("<primary"), primary education completed ("primary"), secondary education completed ("secondary"), and tertiary education completed ("university"), years of schooling would take values of 2, 6, 12, and 16 years, respectively.

Finally, for all four datasets (indon76, indon95, US70, and US90), a new variable, PEXPER (potential experience), representing the number of years an individual has in the labor force, was created. This was done by subtracting each observation’s age from their years of schooling.

**Model and Estimation**

The method of estimation for the wage returns to education is a multivariate log-level regression model of the following specification, inspired by the Mincer regression (Mincer, 1958):

\[
\log (INCWAGE_i) = \beta_0 + \beta_1 YRSCHOOL_i + \beta_2 PEXPER_i + \beta_3 PEXPER_i^2 + \beta_4 HRSWORK_i + \beta_5 SEX_i + \epsilon_i
\]

Where the subscripts \(i\) denote the observations, \(\beta_0\) is a constant term, \(PEXPER_i^2\) denotes the square of the years of potential experience an individual has in the labor market, \(SEX_i\) is a dummy variable that takes the value 1 if the observation is male, and \(\epsilon_i\) denotes the unobserved error term. The natural logarithm of observations’ annual income (\(\log (INCWAGE_i)\)) was taken instead of solely annual income (\(INCWAGE_i\)) to linearize the relationship between the above covariates (independent variables) and annual income.

Educational attainment was not chosen for the regression as adding it would lead to multicollinearity issues, as it is highly correlated with years of schooling, and can subsequently be linearly predicted from each other at a substantial level of statistical significance. The industry of employment and employment status were also ultimately left out to limit overestimation (upward bias) or underestimation (downward bias) of coefficients but can be incorporated in any possible extension analyses to exploit possible heterogeneity. It is impossible to account for all the variables in the data, even though they correlate with incomes. Therefore, robust standard errors were used in the regressions to address the effects of omitted variable bias and heteroscedasticity.

**Main Results**

Figures 1 and 2 are scatterplots with an estimated linear best fit line that illustrates how the relationship between years of education and the natural log of income for the United States and Indonesia has changed between the 1970s and 1990s, categorized by sex. Figure 1 illustrates Indonesia in 1976 and 1995, while Figure 2 illustrates the United States in 1970 and 1995.

The figures confirm the primary hypothesis that income increases for each additional year of schooling (either because of the human capital or screening hypotheses), this result is consistent across both sexes (male & female). The addition of an approximate linear model line of fit for all the plots for each period does provide some elementary
insight into the direction of the relationship between education and wages in each of the two periods (the 1970s and 1990s), especially for the Indonesia data as seen in Figure 1. As per the summary statistics of the Indonesia dataset, since there are more observations in 1995 (78,074) than in 1976 (19,620), the line of best fit for Indonesia in 1995 might paint a more accurate picture of the relationship between years of schooling and wage than Indonesia in 1976. This does not consider the effect of the other covariates (potential years of experience in the labor market and hours worked per week). The results from Figure 1 are more tedious to synthesize for United States data. From Figure 2, it might be tempting first to conclude that there are more data points for the United States in 1970 than in 1990, and so the line of best fit best approximates for US 1970 data; however, if one consults the summary statistics of the US datasets, we see that there are more observations for 1990 (2,428,419) than in 1970 (393,565). As described in the Data Sources & Methodology section, the years of schooling variable for US 1990 data only takes four values (2, 6, 12, or 16). A broad majority of the data points are clustered to these four values of years of schooling by construction. This data, along with the fact that other covariates mentioned above have not been considered in Figure 2's best fit models, we must turn to the regression estimates to obtain the most holistic picture of the wage-education relationship.

Four regressions were conducted in total, as per the specification outlined above, for each country in each period. Figure 3 below details the results for each of the four regressions. Each estimation of the coefficient terms in the above regression equation is provided, along with an indication of their significance levels and standard errors. For each regression, the number of observations, the coefficient of determination ($R^2$), adjusted $R^2$, residual standard errors, and F-statistics. The results for all four regressions are statistically significant at the 99% confidence level. The robust standard errors were measured to be less than the regular standard errors, and their low magnitude indicates that the sample means of these coefficients ($\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$ and $\beta_5$) are closely distributed to the true coefficient that signifies the effect of each of the covariates on the (natural logarithm) of incomes. The constant terms, while having no real economic interpretation (as it represents the natural log of income given zero years of schooling), are high, indicating possible selection issues.

**Discussion**

For Indonesia in 1976, with an additional year of schooling, we would expect income to increase by 14.7%. In contrast, an additional year of schooling in 1995 would increase incomes by 12.2%, thereby complementing Behrman & Deolalikar and Duflo’s findings (Behrman & Deolalikar, 1993; Duflo, 2001); however, these analytical results show that in the intervening 19-year period, wage returns to education have dropped by 2.5% in Indonesia, which supports Purnastuti's findings that the returns to education in Indonesia have been decreasing, especially when compared within Southeast Asia in the same period (Purnastuti et al., 2013). The results also
show that incomes increased by 60.4% in Indonesia in 1976 and by 32.5% in 1995 relative to women. While it speaks to potential topics for future research such as wage gaps across the sexes (Widyanti, 2018), the scope of this result, particularly in the context of wage returns and educational attainment, is generally not substantiated in the literature review conducted for this analysis.

The effect of potential years of labor market experience on incomes is also a worthy finding that supports the secondary hypothesis: an additional year spent employed in Indonesia’s labor market led to a sizeable 8.1% increase in incomes in 1976 and a 5.1% income increase in 1995, though the effect decreases along with years of schooling. The coefficient representing the square of the potential years of labor market experience, while statistically significant, is very small and negative, indicating that the returns for an additional year of work should generally be positive. Hours worked were not as correlated with income compared to the other covariates; while the coefficient term is significant, it is the smallest of all the independent variables. While it was expected that an additional hour worked would be correlated with an increase in income, the result here does make sense since an additional hour worked per week would not factor much as we consider annual incomes here.

A similar synthesis of the regression results can be applied to the US in the 1970s and 1990s, where the difference in wage returns (which increased 8.0% in 1970 and 9.6% in 1990 for an additional year of schooling) increases only by 1.6% between the 20 years. Objectively compared to the coefficients in the Indonesia regressions, the returns are less, which is to be expected given that Indonesian market-oriented economic reforms in the 1980s led to high levels of economic growth. This was likely due to greater access to education which spurred an increase in the number of productive yet highly educated workers in 1990s (Purnastuti et al., 2015). The high-income, developed status that the US economy has maintained since the 1970s explains how education may not have as much as a pronounced effect on health, growth, and positive externalities.

Just like in the Indonesia results, hours worked per week also do not account for much for annual incomes in the United States in 1970 and 1990. However, the effect of potential labor market experience on wages is much less pronounced in the US (2.8% and 4% for 1970 & 1990, respectively) than in Indonesia, though it increases between 1970 and 1990, indicating the effects other variables aside from education are having on wage returns. The coefficient on $PEXPER_i \times (\beta_2)$ decreases by 3% in Indonesia from 1976 to 1995, while it rises in the US by 1.2% from 1970 to 1990. As is with the situation in Indonesia, the $\beta_3$ coefficient representing the effect of the square of potential labor market experience on wages is also negative but small in magnitude, so we can conclude that the effect of an additional year worked should generally be positive. It, along with the slight increase in the coefficient for an additional year of schooling, had a greater effect on incomes at the expense of hours worked per week. It is also
observed that the effect of the individual being male also had the greatest effect on wage returns relative to women in the US, and this is larger than the gender effect in Indonesia on a relative consideration.

**Conclusion & Policy Implications**

This study was successfully able to replicate the wage returns to education results in the literature. Without a doubt, however, the effect of non-education covariates in the above regressions corroborates the heterogeneous wage returns findings, where other factors, like years of labor market experience and gender, need to be considered (Brand & Xie, 2010). While gender wage gap is much beyond the scope of this research, gender discrimination in the labor market is a social issue that is still ever-present in both American and Indonesian societies today, and it would be a worthy topic of analysis to explore wage returns to education in present years and see whether the effect of gender has changed.

Given the data manipulations and processing outlined above and the data sources, it can also be argued that the above regression model has low external validity and hence is limited in its number of observations. Nevertheless, there are several useful policy implications both countries’ small and medium-sized enterprise (SME) owners for the present day. In Indonesia, SMEs have found to be essential to Indonesia’s future economic growth (Indonesia’s SMEs Are Key to Development. How Can They Grow?, 2021), and in the United States, have created almost 66% of net new jobs since the turn of the millennium (The State of Small Business in America, 2021). An 8 to 15% increase in historical wage returns for an additional year of schooling, is a notable result that can be used to guide possible policy in the United States and Indonesia to further incentivize and harmonize educational institutions and labor markets.
Figure 1. Trends in wages by gender and education in Indonesia in 1976 & 1995. This scatterplot displays the logarithm of income in Indonesian Rupiah (IDR) against the years of schooling for individuals in Indonesia, separated by gender. The data points are color-coded to distinguish between the years. Solid lines represent linear fits for each gender and year, illustrating changes in the relationship between education and income over the 19-year span. Notably, both men and women show an upward trend in income with more years of education, with the increase more pronounced by 1995.
Figure 2. Wage trends by gender and education in the United States in 1970 and 1990. This scatterplot depicts the relationship between years of schooling and the logarithm of wages (in United States Dollar; USD) for male and female workers in the United States. The data points are differentiated by year, with 1970 represented in red and 1990 in green. The data points indicate the range of wage values for each year of schooling within each gender group, providing a visual comparison of wage dispersion over time. They also highlight the evolution of wage outcomes across two decades, emphasizing the potential impact of education on earnings in the US labor market.
**Figure 3.** A table displaying the coefficients from Mincer ordinary least squares (OLS) regression models. The coefficients estimate the causal effects of years of schooling, potential years of labor market experience (and its square), and hours worked on the natural logarithm of incomes in Indonesia for the years 1976 and 1995, and in the United States for the years 1970 and 1990. The table includes the numerical coefficient estimates, standard errors (in parentheses), and significance levels (*p<0.1; **p<0.05; ***p<0.01). Additionally, for each model, the table reports the number of observations, the R-squared and adjusted R-squared values, the residual standard errors, and F-statistics.
References


