2016

Effects of Parental Education on Return to Education

Zhirou Shi
Skidmore College

Effects of Parental Education on Return to Education

By

Zhirou Shi

A Thesis Submitted to

Department of Economics

Skidmore College

In Partial Fulfillment of the Requirement for the B.A Degree

Thesis Advisor: Qi Ge

May 3, 2016

Abstract

Return to education is the rate of return to income with each additional year of schooling. This paper utilizes data from the Panel Study of Income Dynamics (PSID) in 2012 to examine return to education in the U.S. I use the Mincer earnings function (Mincer, 1974) as a basic model to evaluate the relationship between relationship between school and earnings. I augment this wage function with control variables including gender, region, family economic background, and job industry to decrease the influence from other factors on the relationship between income and schooling. Then, in order to eliminate ability bias in the model, I use the method of instrumental variable (IV), with the father's and mother's education as instruments for schooling, to re-examine return to education. The results from using the augmented wage function indicate that the ordinary least square (OLS) estimate (5.87%) is downward-biased compared to the IV estimate (10.2%). These findings are aligned with findings from previous research. This study contributes to the field by updating the rate of return to education for year 2012 and showing that parental education is closely related with children's rate of return to education, which provide possible future policy directions on how to get better results from spending on education.

I. Introduction

As the most important and primary human capital investment, education has always been highly valued by citizens, welfare institutions, and governments because education can stimulate economic growth by improving employee's working skills and alleviating poverty by increasing income. Previous research suggests that education is one of the most efficient tools to improve economic well-being (Garay, Zereyesus, & Thompson, 2014). Getting educated not only benefits the learner, but also the society by decreasing the crime rate and insecurity. Due to these merits, many countries have spent an extraordinary amount of money on providing public education.

However, besides this, we know very little about the quantified benefit of education. Return to education can be thought as a quantified benefit from education if we agree that the major goal of education is to earn higher income in future. Return to education refer to the percent increase in income with each additional years of schooling. For example, a 10% return to education means that earning increases 10% with each additional year of education. Knowing the accurate return to education is important because governments can adjust the amount of spending on public education accordingly and make the investments reach their maximum profitability.

In this paper, I examine the rate of return to education in the United States for year 2012 by analyzing 2,583 observations from Panel Study of Income Dynamics (PSID). I will also explore the relationship between parental education and children's return to education. There are two major econometric issues that can bias the estimate for return to education: ability bias and measurement error. Ability bias refers to the phenomenon that education is not randomly assigned. People who are more capable are tend to have more years of education, and therefore are more likely to be paid higher not only because of their education, but also their valued abilities. Measurement error is the mis-measurement of schooling due to inaccurate self-reported schooling. People tend to over report their schooling when they are not satisfied with their education or uncomfortable with the survey question. To deal with the ability bias, I will use the method of instrumental variable (IV) and use parental education as the instrument for education. Unfortunately, measurement error cannot be corrected in this study because the dataset is very limited.

My study contributes to the field by both updating the rate of return to education for year 2012 and showing factors that may related with children's schooling and return to education. As explained before, knowing accurate return to education helps government invest more wisely on education, and knowing factors that can affect return to education, such as parental education, can help governments to increase the efficiency of using the expenditures.

The results suggest that ordinary least square (OLS) estimates and IV estimates are significantly different. Return to education under using the method of OLS is 5.87%, whereas the method of IV increases the return to 10.2%. OLS estimates are biased downward. Another finding is that parental education is positively correlated with children's schooling. Children of fathers with bachelor's degrees will have 0.59 more years of schooling compared to children whose fathers do not have college degrees. Children with college-graduate mothers will have 0.703 more years of schooling compared to children whose mothers have not attended any college.

I will first go over some previous essential literature that discuss rate of return to education and the econometric approaches they use to deal with ability bias and measurement error. Then, I will introduce my theoretical models. Last but not least, the ordinary least square

(OLS) and instrumental variable (IV) estimates will be presented and compared in the result section. Some limitations of this study will then be discussed.

II. Literature Review

Before starting the discussion of estimating the accurate return to education, I would like to first build a general understanding of the assumptions and the logic behind the economics of education by introducing a schooling model. Most of recent studies investigating returns to education base their concepts and ideas on the model introduced in Becker (1964). He defines a school as “an institution specializing in the production of training.” Since students usually do not get employed until they graduate from schools, the major benefit of getting education is to earn higher incomes in future. At the same time, they have to afford the cost of schooling to enjoy that benefit. The cost including direct costs, such as transportations, tuition, and study supplies, and indirect costs including foregone earnings. The model also incorporates interest rates that allow people to borrow or lend to cover the costs and discount rate for future gain. The major idea of this model is that individuals need to balance the benefits of getting higher education against the costs to reach an optimal schooling decision.

Becker (1964) gives a cost function and a benefit function according to the terms above in utility term. By integrating the cost function by the years of education, we compute the total amount of investment one would make in education. Similarly, we can calculate the schooling benefit by integrating the benefit utility function from the time we finished school to the last day of our life. By adding these two values together, we can calculate the benefit we could earn for the entire life from education. To maximize the lifecycle benefit, we can set the derivative of this equation equal to 0 and get the years of schooling that can optimize our benefit.

Mincer (1974) adopts the equation that implied a model for log earnings in Becker (1964) and introduces his earning function from an econometrics perspective to depict the relationship between observed schooling and earnings outcomes:

$$\mathrm{lny} = \mathrm{ln}y_0 + rS + \beta_1 exp + \beta_2 exp^2.$$

In this equation, *y* stands for wage rate; y_0 is a constant term; *S* is schooling in years; *exp* is the experience one has in years. It is a straightforward mode that uses ordinary least squares (OLS) to estimate for the return to schooling. What is special with this equation is that he added terms that depict one's experience into the function. Mincer argues in his paper that a simple regression of earnings on years of schooling could not capture the true relationship because schooling is not the only type of capital investment. Since earning is a return on cumulated net investment, to better predict human capital earnings function, we need to find out the function of the cumulated investment. After studying numbers of individual profiles, Mincer claims that wage function is a concave function of experience. Therefore, the linear schooling term is augmented by a nonlinear, squared experience term to complete the earnings function. This earning equation also fits the prediction of the concave lifecycle earning curve, which shows that income diminishes with age. The Mincer earnings equation is generally used in economic studies investigating relationship between schooling and income because it shows a direct relationship between income can represent the most important return from education.

However, this equation ignores the influence of inaccurate measurement of schooling and endogeneity in explanatory variables. A measurement error is the difference between a measured value of quantity and its true value. Years of schooling are easy to be misreported because people may consider it as privacy and holding negative attitudes when giving out that information. To explore the existence of this error, Griliches (1977) creates two terms: observed schooling and true schooling, and substitutes the true by the observed schooling into the regression model. He finds that the existence of measurement errors from the regression results and claims that measurement errors in schooling will create downward biased OLS estimations on return to schooling.

OLS estimate of return to schooling through the Mincer earnings function is also subject to endogeneity. Endogeneity is the phenomenon that explanatory variable is correlated with the error term. In my study, endogeneity is refer as the nonzero covariance between education and ability, which can be considered as ability bias in schooling problems. As motioned before, ability bias is the phenomenon that more capable people tend to have more years of schooling and are more likely to have higher income because of their capabilities. The OLS estimate results partially reflect the return to schooling, but they are usually assumed to be biased upward because people who tend to have higher education are usually smarter and are therefore more easily to be paid higher than others.

To estimate the rate of return more accurately, previous researchers use the method of instrumental variables (IV). IV is the observable covariate that only affects an independent variable but is uncorrelated with the error term in the equation. IV procedure is a tool that helps to estimate causal relationships when the explanatory variables (covariates) are correlated with the error terms in a regression. In my paper, using the method of IV allows us to observe the true relationship between education and earnings because there are other factors that affect one's schooling choices that are not included in the log wage equation. This method provides consistent estimates by replacing the original observations by predicted values that are formed by one or a set of instrumental variables. Two properties are required when selecting IV: being related to explanatory variables and uncorrelated with the error term. Card (2001) proves that when the chosen instrument satisfies the properties stated above, return to schooling can be consistently estimated by IV. However, when the assumptions are not satisfied, the IV estimates can be inconsistent and biased. Card gives an example of a less-satisfied IV. “Exposure to different institutional structures” is the instrument uses in the example to estimate schooling and eliminate individual ability difference. This selection of instrumental variable can violate our assumptions because ability and schooling are likely to be affected by a change in educational institution at the same time. For example, if reformation of a specific set of schools lowered their tuition, the marginal cost of education would decrease, and therefore people's schooling choice will be affected.

There are several alternatives to the IV procedure. A close-related alternative suggested by Card (2001) is a control function approach, first proposed by Garen (1984) when discussing the schooling issues. This model makes assumptions about the nature of covariance between the unobserved correlated components (ability) and the observable variables. The study includes additional control functions and interaction terms into the equation to capture the relationships. Card also suggests another more radical alternative to IV is maximum likelihood estimation of a structural model for earnings and schooling. This method is based on a complete specification of the unobservable components in the earning function. It allows the earnings function to be made very general, such as allowing the return to different years of schooling to vary flexibly with individual ability. However, it can be very challenging to completely write out the entire function because we cannot know what factors are related with the dependent variable. In this study, since we are unable to write out how ability can affect education and need to deal with endogeneity and estimation error, IV method is a desirable tool to be used. As long as the chosen instrument is appropriate, IV procedure will a give consistent estimate of return to schooling.

To sum up, adding instrumental variables can help estimate the return to schooling more consistently by reducing the ability difference among subjects. A suitable choice of instrument leads to high credibility of IV estimates. A lot of researchers use IV as a standard solution to the schooling problem. Previous researchers have chosen compulsory school laws, twins, high-education institution proximity, parental education, and etc. as instruments to estimate unbiased return to schooling. I will summarize and compare their OLS estimates and IV estimates in the following section.

Instrumental Variables Measuring Return to Education

Features of the schooling system are popular instruments picked by economists. Angrist and Krueger (1990)'s landmark study of examining relationship between compulsory schooling and education uses individual's quarter of birth as an instrument for schooling. They believe that age policy and compulsory school attendance laws can affect return to schooling because these policies must force some students to stay in the education system longer then they desire. These more years of schooling may then lead to higher education return. Time of birth can be a factor that influences the number of years that one stays in school. Therefore, Angrist and Krueger choose to use the quarter of birth as an instrument for schooling. They argue that it is a desirable instrument because it is unlikely to be correlated with personal attributes but generates exogenous variation in education. They expect that people born early in a year are more likely to drop out and therefore have less schooling then people born later in a year because people born in the beginning of the year start school at an older age and are able to drop out early compare to individuals born near the end of year. Angrist and Krueger use 5 percent Public Use Sample from the Census to collect data for men born from 1930 to 1959 in the United States and observe what they expected as explained before. However, they did not observe significantly difference in estimates from the instrument variables and the OLS on the return to schooling, which indicated that the conventional estimates are not significantly biased.

Angrist and Krueger (1990)'s findings attracted a lot of interest as well as some criticism. Bound and Jaeger (1996) criticizes that quarter of birth is not a satisfactory instrument because it may correlate with unobserved ability difference. They examine schooling outcomes of men who are born before the existence of compulsory school laws and found some evidence of seasonal patterns. They suggest that this pattern may correlate with family background. Children born earlier in the year tend to have poorer family backgrounds, and this may influence them to have shorter schooling and lower income. Since the instrument Angrist and Krueger choose also correlated with ability, it is not a perfect IV. Inspired by the criticism of Bound and Jaeger, I realize that family background has a strong influence on one's schooling choices. In my study, I would like to use family background indicators as instruments for schooling.

Ashenfelter and Krueger (1992) examines returns to schooling among monozygotic twins (identical twins from the same fertilized egg) by contrasting wage rates of them with different years of schooling. Monozygotic twins are considered genetically identical and share similar family background. By controlling for exogenous factors and individual's ability, using twin data can help discover the true relationship between education and earnings and eliminate the correlation between schooling and workers' ability or other characteristics. Another error that Ashenfelter and Krueger explore and fix is measurement error. They use multiple measurements of schooling levels to make years of schooling accurate. One major innovation way of reducing measurement error in this study is to ask the twins to report both their own and their twin's schooling. Ashenfelter and Krueger use the average of self-reported and twin-reported schooling to run the regression to reduce measurement error. To acquire desirable twin data, they conduct survey of identical twins during twin's festivals in 1991. They interview over 495 separate individuals over the age of 18 using their own questions and questions from Bureau of the Census for the Current Population Survey (CPS).

Ashenfelter and Krueger (1992) finds significant evidence of measurement error in schooling levels. They calculate the reliability ratio of their measures, which is the fraction of the variance in the reported measures of schooling that is due to true variation in schooling. The ratio indicates that between 8 and 12 percent of the measured variance is due to error. They also find that measurement error may lead to considerable underestimation of the returns to schooling in the conventional method. However, since they collect years of schooling reported by one's own and his or her sibling and generated estimations based on averages of that data, measurement error caused a smaller bias in their model than in the fixed-effect estimator.

Another econometric issue that this study deal with is ability bias. They first use data from identical twins to control ability difference. The nature of identical twins allows them to have similar levels of ability since they have the same genetic endowment and a similar growing environment. Secondly, Ashenfelter and Krueger (1992) uses the years of education reported by one's sibling as an instrument variable for one's education level. Estimates from this procedure are 0.17, which is much larger than the least-squares estimate (0.09). They suggest that if the IV is valid, the conventional methods significantly underestimated economic returns of schooling. From their estimates, Ashenfelter and Krueger conclude that they find no evidence to show that conventional estimates of return to schooling are biased upward due to imperfect control of individual and family-related characteristics. Also, measurement errors in self-reported schooling substantially downward biased conventional estimates. Their results suggest that every additional year of completed schooling increased average wage rate by about 12-16 percent using the described procedure.

Twins are the ideal experimental subjects to reduce ability bias because they have the similar levels of capability through sharing the same genetic endowment and family characteristics. If twin's data is available, I believe it would be helpful to use it to decrease ability bias for my study. However, if it is not available, family characteristics, as one of the key variables Ashenfelter and Krueger's (1992) controls in their study, can be used as instrument for schooling to reduce ability bias in this study. Since endogeneity is significantly reduced in Ashenfelter and Krueger's study by using twin's data. They are more focused on examining the effect of measurement errors on conventional regressions and finding results with measurement errors fixed. The innovative methods they use to collect accurate years of schooling and compute measurement error empirically are attractive. However, measurement error is not the major econometric issue that my paper will discuss about.

Subsequently, Card (1993) examines difference in return to schooling associated with variation in exogenous source. In the study, he explores how four-year college proximity related with schooling and earnings differentials. Card argues that the presence of a nearby college can be an instrument because it can be excluded from the earnings equation but it correlates with schooling. Students who grow up in an area with no high education institutions face a higher cost of schooling because they need to pay higher transportation fees and other expenses. Therefore, the benefit that can be gained from schooling reduces and students may decrease their years of education. The data is drawn from the National Longitudinal Survey of Young Men from 1966 with 5,525 men aged 14-24 and continued with follow-up surveys through 1981. Results from estimating a wage equation on 1976 wages indicated that men who grew up in an area with a nearby college have significantly higher education and earnings than other men. Also, when college proximity is used for the instrumental variable, the estimates are 25-60 percent higher than conventional OLS estimates, which means that college proximity matters when individuals make schooling decisions. This effect is even more prone to children who had less-educated parents. Since schooling is also affected by parental education level, Card uses interaction of college proximity with family background variables as instruments for schooling. The estimates of return to education are lower than solely using college proximity as the instrumental variable, but are still about 30% higher than the OLS estimates.

The substantial influence of family background is implicitly mentioned in Card's (1993) study. He claims that parental education influenced the schooling choices and therefore incorporated it into his instrument sets. This confirms my belief that family background is significant to one's schooling. Furthermore, parental education might be a good indicator to capture and quantify one's family background. I will then discuss some literature that reveal the importance of parental education and use parental education as an instrument.

Parental Education on Schooling

Family environment is crucial to children’s growth and development. Parents can have a significant impact on their children by accompanying them growing up. Financial ability, willingness to purchase education, and promoted values are all factors that can significantly affect the effectiveness of human capital investment, such as education. Altonji and Dunn (1996) examined the effects of family characteristics on the return to education by estimating the influence of parental education on wages. They augment the Mincer earnings function with interaction terms between individuals’ schooling and their father’s education and mother’s education. The equation is estimated by using data from the Panel Study of Income Dynamics and National Longitudinal Surveys of Labor Market Experience of Young Men and Young Women from 1966 to 1981. The NLS results show that the interaction between the father's education and the son's education is 0.691 and the interaction between mother's education and the son's education is 1.1. The coefficient on the interaction of daughter's and the father's education is -0.041 and the corresponding interaction with mother is 0.79. The PSID results indicated that the son's return to education is positively related with both the father's and mother's education. The daughter's return to schooling is positively related with the mother's education but negatively related with the father's. These results indicated that parental education has substantial effects on children's wage level and return to education.

Although Altonji and Dunn (1996) has not use the method of IV for their empirical analysis, they provide statistical evidence of the significant influence parental education has on children's return to schooling. Since parental education is closely related to children's schooling, I think using parental education as an instrument for schooling can be a good choice. However, the validity of using parental education as an instrument needs to be discussed.

Also recognizing the substantial influence of parental education on efficacy of return to education, Maluccio (1998) uses parental education as one of the instrumental variables to examine the return to education in the Bicol region of the Philippines. Data is drawn from the Bicol Multipurpose Surveys in panel design containing three observations (1978, 1983, and 1994). He selects type of schools (private or public), distance to the nearest high school, household wealth, the mother's and father's completed education, and short-term health measures to make up the instrument sets. These variables had a relatively strong effect on education in this sample, since Maluccio adapts logarithmic wage function to the local situation of Bicol. He argues that cooperating parental education in instrumental variables is important because they serve as proxy for permanent income, reflect parental preferences, and may affect the education production process. Results from conventional OLS models are in accord with the usual findings in other countries: male has higher income than female and wages are lower in rural areas. Results from the first-stage regressions instrumenting for education show that the parental education and household wealth instruments are all significant at 5 percent level and raise completed education. This indicates that parental education can be served as good instrumental variables. The method of IV increases returns to education more than 60 percent compare to using OLS estimation. Malcuccio (1998) proves empirically that using parental education as instrument is feasible in his study, which builds a solid base for me to use parental education as instrumental variable in this study.

Table 1 summarizes the OLS estimates and the IV estimates of basic wage function from studies that are discussed in literature review section. It is more clearly to see the comparisons between the conventional method and the IV method this way. There are two major findings from this table. First, family background is significant when considering education because it is considered or controlled in every study. Second, comparing OLS estimates and IV estimates, we can see that OLS estimates are tend to lower-estimate return to education. This return will increase and significantly different with OLS estimates by using the method of IV with valid instruments.

Table 1 Summary of comparisons of ordinary least square estimates and instrumental variable estimates from articles in literature review

Article	OLS estimate of basic model	IV estimate with listed instruments	Instruments	Is the difference between OLS and IV significant?
Angrist and Krueger's (1990)	0.0802 (0.0004)	0.0769 (0.0150)	season of birth dummies	No
	0.0711 (0.0003)	0.0891 (0.0161)		
	0.0573 (0.0003)	0.0553 (0.0138)		
Ashenfelter and Krueger (1992)	0.084 (0.014)	0.116 (0.030)	Sibling reported schooling	Yes
Card (1993)	0.073 (0.004)	0.194 (0.059)	College proximity	Yes
Maluccio (1998)	0.073 (p-test result=6.41)	0.1233 (p-test result=5.84)	Distance to secondary school, private secondary school in village, parental education, logarithm of value of house, and own agricultural land	Yes

In this paper, I will use the Mincer earnings equation to evaluate the relationship between schooling and economic return of schooling in the U.S. I will use the method of IV to explore how ability bias can affect return to education by using parental education as an instrument. There are two econometric issues that need to be discussed with this paper according to previous research: ability bias and measurement error. To explore how ability bias can affect return to education and obtain corrected results by eliminating ability bias, I decide to use the father's and mother's education levels as an instrumental variable for schooling. Previous research has shown that parental characteristics can substantially affect the effectiveness of children's investment in education from previous research. Ashenfelter and Krueger (1992) keeps family environment controlled in their study to avoid the effect of parental characteristics on children's wage rate. Card (1993) included parental education into instrumental variables because he believed that parental education would influence the decision they make for their children. Malcuccio (1998) proved that parental education can be use as an effective instrument for schooling. Parental education should be correlated with children's education, which satisfies the assumption that the correlation between instruments and endogenous variable is not zero. However, as for the second assumption that instruments should not be correlated with the error term, parental education may not be fully satisfied because parents may influence children's ability since they always closely accompany children's growth. The validity of instruments will be discussed further in discussion section.

The second issue that needs to be discussed is the bias generated by measurement error. Ashenfelter and Krueger (1992) brought up a great method of examining and fixing measurement error in the data, but this problem is actually relatively hard to be solved in this paper due to the limited resources of author. First, public twin's data with parental education information is barely available online. Second, the data used by Ashenfelter and Krueger is available online but may not be appropriate to use for this paper, because the data is collected in 1991. It could be outdated and the results may not be applicable to the current social environment. My paper will use the newest data, which is data of year 2014, from Panel Study of Income Dynamics (PSID) to see if there are any major changes in the return to education in today's society compared to the results from previous research.

III. Methodology

Data

Data used in this paper is from Panel Study of Income Dynamics (PSID) for year 2013. Observations are the head of family. Due to the characteristic of PSID, 2013 data actually records what happened in 2012. Therefore, the regression results in my paper describe the relationship between return to education and schooling for the United States in year 2012.

I include *age* (in years), *education* (in years), *hourly wage rate* (in dollar), *gender* (female = 0, male = 1), *father's education* (dummy variable), *mother's education* (dummy variable), economic *family background* the head growing up (dummy variable), *industry* of the head's occupation (dummy variable), and current *region* (dummy variable) in this study. The original data includes 9,107 observations. Several types of data are dropped. First, missing data is dropped. Second, wage that is 0 is dropped because it needs to be log transformed. Furthermore, observations of wage that belongs to the lowest 5% and highest 5% are dropped in order to decrease the influence of outliers on the regression results. After this process, the lowest hourly wage in the dataset becomes $7.5 per hour and the highest wage is $35 per hour, which is a reasonable range for hourly wage rate. Forth, observations of heads who are younger then 19 years old not including 19 years old are dropped because federal law regulates that only individuals who are 19 or older are qualified as heads of households. Finally, negative observations of *experience* are dropped from the dataset. Experience in this study is calculated by $age - edu - 6$, which is possible years of being employed and cannot be zero if age and education are recorded correctly. After a general cleaning of the data, there are 2,583 observations left in the data set. A general description of the continuous variables is displayed in Table 2. From the statistics, we can see that subjects in this study are mainly in their middle age and well educated because the mean of education is 13 which stands for high school graduates. The mean wage is $15.6 per hour which is significantly higher than the federal minimum wage $7.25 per hour. The average experience in the dataset is 20.2 years which indicates that most subjects in the dataset have some working experience. Mother's education is a little higher than the father's education level, but there is not significant different. Parental education is around the level of high school graduates. Method of categorizing levels of education will be showed in methodology section.

Table 2 Sample Means, Maximum, and Minimum

Variable	Mean	Smallest	Largest
age	39.2	19	85
	(13.1)		
education	13	0	17
	(2.2)		
wage	15.6	7.5	36
	(6.5)		
experience	20.2	0	69
	(13.4)		
father's education	3.9	0	8
(categorical)	(2.03)		
mother's education	4.1	0	8
(categorical)	(2.03)		
Number of observation	2,583		

Note: Standard deviations in parentheses.

Models

The regression model used in this paper is based on the Mincer earnings function:

$$\text{lny} = y_0 + \beta_1 edu + \beta_2 exp + \beta_3 exp^2 + \varepsilon$$

Where *y* is hourly wage rate; *edu* is the schooling of family heads; *exp* is working experience; ε is the error term. This a straightforward regression function showing that the relationship between income and education is also related with working experience. However, it is very easily influenced by unobserved factors, such as gender, industry, and region. According to the assumptions when setting up this model, I predict that the OLS estimates for coefficients of education and experience are positive, whereas the coefficient of experience squared is negative to give a concave shape of life-time earning curve.

Previous research controlled factors that may significantly influence the results. For example, Card (1993) includes gender, current living region, race, and family structure into the basic wage function to examine the return to education more accurately. To reduce this undesirable influence on return to education, I refer to the control variables that are used in previous research and add them into the basic wage function. These control variables are *gender*, *gender-education interaction*, current living *region*, *family economic background* that head growing up with, and the *industry* that head is currently working in. After adding these control variables, the regression function becomes:

$$\begin{aligned}\ln \text{wage} = {} & \beta_0 + \beta_1 \text{edu} + \beta_2 \text{exp} + \beta_3 \text{exp}^2 + \beta_4 \text{gender} + \beta_5 \text{gender\_edu} + \beta_6 region^* \\ & + \beta_7 family^* + \beta_8 industry^* + \varepsilon\end{aligned}$$

Control variables are generated as dummy variables. Variables in the function with a star indicate that a set of dummies is included instead of one dummy variable.

Getting control of these variables can help us find a more accurate OLS estimates. Since men are always paid higher than women, gender can substantially influence the relationship between schooling and return to education. In this dataset, female is represented as 0 and male is represented as 1. I also add an interaction of education and gender into the equation in order to allow me to estimate the difference in return to schooling between genders. Regions can affect income significantly because income may be adjusted to compensate different levels of living expenses in different regions. For example, the living expense in New York state can dramatically higher than the cost of living in Mississippi. Wage in New York state, therefore, needs to be higher than the wage in Mississippi to ensure that employees are living the same quality of live. Thus, regions are needed to be controlled when examining return to schooling. Current region that household heads are living in is represented as a categorical variable in PSID. Number 1-5 indicate northeast, north central, south west, and Alaska/Hawaii, respectively, which are the same with census regions. Another control variable I choose to add is the economic family background which the heads growing up with. In PSID, this is asked as if the heads' parents were poor when they were growing up. The answers are categorized into three groups: 1indicates "poor", 3 indicates "average", and 5 indicates "pretty well off". As discussed in previous literatures, family background is an important factor that needs to be considered when examining return to education, since it can significantly affect the schooling decisions of children. Card (1993) has proved that economic background that one growing up with can affect schooling and Card has controlled this variable in his study. The last control variable I add is the head's current working industry. Industries need to be controlled because hourly wage rates various dramatically from one industry to another. For example, hourly wage rate in financial industry and hourly wage rate in agriculture industry cannot be comparable. There are 19 industry categories from PSID from agriculture to public administration.

With the exogenous variables controlled, the coefficient of education is expected to increase compared to the non-controlled model because return to education should be measured more accurately with the augmented model and therefore the results should increase to a high rate as showed in Table 1.

I will then use the method of IV to examine the relationship between schooling and income. The method of IV leaves the unobserved variable in the error term and uses an estimation method that recognizes the presence of the omitted variable. Assume we are using a simple regression model:

$$y = \beta_0 + \beta_1 x + u$$

where y is the dependent variable; x is endogenous variable that correlated with u; u is the error term. Suppose we have an instrumental variable *z* for *x*. The method of IV can consistently estimate return to education (β_1) when the following two assumptions are satisfied: instrumental variable is related to endogenous variable *education* ($\mathrm{Cov(z, x)} \neq 0$) and is unrelated with exogenous variable *ability* ($\mathrm{Cov(z, u)} = 0$). IV method identifies return to education, which is the coefficient of endogenous variable, as following:

$$Cov(z, y) = \beta_1 Cov(z, x) + Cov(z, u).$$

Since $\mathrm{Cov(z, u)} = 0$, the equation can be rewrite as:

$$Cov(z, y) = \beta_1 Cov(z, x)$$

Therefore,

$$\beta_1 = \frac{Cov(z, y)}{Cov(z, x)}$$

$$\text{and} \quad \widehat{\beta_1} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}$$

The IV estimator of β_0 is $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\bar{x}$, which is derived by the similar method using to calculate the OLS intercept estimator except that $\widehat{\beta_1}$ is now the IV estimator. When x is exogenous IV estimator will be identical to the OLS estimator (Wooldridge, 2003).

In this study, I choose to use parental education as instruments for schooling. As discussed previously in the literature review section, family background is an important factor that may influence one's schooling. Card (1993) proves that children with lower educated parents are more likely to have less education if college proximity is low in the current living area. Therefore, to investigate the effect of parental education has on children's education, I will use parental education as instruments to estimate schooling to treat the endogeneity between education and ability. The father's and mother's education in PSID are categorical variables with 1 through 8 indicate 0-5 grades, 6-8 grades, 9-11 grades, 12 grades, 12 grades plus nonacademic training, some college, college BA, and advanced or professional degree, respectively. A high number of mother's education indicates a highly educated mother, which is the same with the father's education.

If parental education is a valid instrument, an IV estimate that is higher than previous OLS estimates is expected. Furthermore, this new estimate of return to education should around 10% according to previous research (Table 1). However, it is possible that the IV cannot provide a desirable estimator because the instrument, parental education, may violate the assumption that instrument should not be correlated with the error term. In this case, parental education may affect children's ability from both nature and nurture.

Results

Table 3 shows the OLS estimates from the basic and augmented OLS estimates. The second column contains the regression results from basic wage function and the third column shows the estimates from augmented wage function. OLS estimates of the basic wage function are all significantly different from zero and indicate that each additional year in education significantly increases hourly wage by 5%. The negative coefficient of squared experience term and the positive coefficient of experience generate a concave regression curve between experience and wage rate, which is aligned with the previous predictions.

OLS estimates from the augmented wage function indicate that return to schooling increases to 5.87%, which is 17.4% larger compare to the estimation from basic wage function. The results also indicate that male's wage is 38% higher than female's wage in this dataset, even though women's return on education is 2% higher than men's. These findings are similar to the findings in other literatures.

Table 3 OLS Estimators of Relationship Between Wage Rate and Years of Schooling

	basic wage function	augmented wage function
Variables	lnwage	lnwage
education	0.0500***	0.0587***
	(0.00364)	(0.00628)
experience	0.0292***	0.0218***
	(0.00191)	(0.00175)
experience^2	-0.000454***	-0.000323***
	(3.96e-05)	(3.57e-05)
gender	-	0.383***
(female = 0, male = 1)		(0.0942)
gender × education	-	-0.0193***
		(0.00720)
constant	1.694***	1.520***
	(0.0529)	(0.117)
observations	2,583	2,583
R-squared	0.154	0.358

Then, I use education of both parents as instrumental variables to estimate return to education. Before showing the IV estimators, I will first present the results of first-stage regression to examine whether the instruments fulfil the assumption that parental education is correlated with children's education by showing the estimations of relationship between instrumental variables (*mother's education* and *father's education*) and the endogenous explanatory variable (*education)*. In Table 3, the second column displays estimates instrumenting for education and the third column displays regression results instrument for interaction of gender and education. The table also includes an F-test on the instruments to show the validity of using parental education as instruments. In the second column, where education is the treatment variable, both the father's education and mother's education instruments have positively significant effect on children's schooling. Therefore, children with highly educated parents are more likely to have more years of schooling. Furthermore, estimator of mother's education (16.9%) has a larger effect on schooling than the father's education (12.8%). The second column presents the results instrumenting for the interaction between education and gender. The instruments for interaction between parental education and gender are both significantly positively correlated with the interaction term at 1% level. F-test results of both OLS estimates are large enough to reject the null hypothesis that the instruments are not correlated with education.

Table 4 First-stage OLS Estimates of Education and Gender Multiply Education

Wage equation variables	education	gender x edu
experience	-0.260**	-0.0038
	(0.01048)	(0.0089)
expenrience^2	0.0000363	-0.00029*
	(0.0002071)	(0.0001755)

gender	-0.0175649	11.552***
(female = 0, male = 1)	0.2004323	(0.170)
Instrumental variables		
Father's education	0.1280537**	-0.0416
	(0.0529)	(0.0349)
Mother's education	0.169303***	-0.0388
	(0.041254)	(0.0350)
Father's education x gender	0.0316582	0.217***
	(0.0514)	(0.044)
Mother's education x gender	-0.0300053	0.1812***
	(0.0515)	(0.044)
constant	13.6911***	1.666**
	(0.903)	(0.765)
observations	2,583	2,583
F-test exclude instruments	42.06	52.13

Table 5 presents the IV estimators using the augmented wage equation as regression function. IV estimate of return to education indicates that each additional year of education increases wage by 10.2%. This estimate is 104% larger than the OLS estimates of the basic wage function and 73% larger than the OLS estimates of the augmented wage function. Estimates of gender difference in wage and return to education are not significant.

Table 5 Estimations of Return to Education Using the Method of IV

	IV estimators
Variables	lnwage
education	0.102***
	(0.0179)
experience	0.0234***
	(0.00186)
experience^2	-0.000324***
	(3.57e-05)
gender	0.378
(female = 0, male = 1)	(0.251)
gender × education	-0.0190
	(0.0190)

constant	1.403***
	(0.296)
observations	2,583
R-squared	0.311

Table 6 shows the comparison between OLS estimates and IV estimates from the augmented wage function. It is obvious to see that return to schooling increases substantially by using method of IV, from 5.87% to 10.2%.

Table 6 OLS and IV estimates of Return to Education

	OLS estimators	IV estimators
Variables	lnwage	lnwage
education	0.0587***	0.102***
	(0.00628)	(0.0179)
experience	0.0218***	0.0234***
	(0.00175)	(0.00186)
experience^2	-0.000323***	-0.000324***
	(3.57e-05)	(3.57e-05)
gender	0.383***	0.378
(female = 0, male = 1)	(0.0942)	(0.251)
gender × education	-0.0193***	-0.0190
	(0.00720)	(0.0190)
constant	1.520***	1.403***
	(0.117)	(0.296)
observations	2,583	2,583
R-squared	0.358	0.311

IV. Discussion

In discussion section, I will first compare my result to previous studies in general. Then, I will analyze two major comparisons: the comparison between OLS estimates of basic and augmented wage function and the comparison between OLS estimates and IV estimates. Later, the validity of instruments and limitations of this study will be discussed.

Comparing to previous literature, my OLS estimate of return to education on the basic wage function (5%) is lower than the OLS estimates on the basic wage function from previous literature. Ashenfelter and Krueger (1992) claims that return on education is 8.4% with the basic wage function, Card (1993) claims the return is 7.3%, and Maluccio (1998) shows that the return is 7.3%. This difference may due to using different dataset and measurement error in schooling. Low return to education may due to low education level of the participants. However, the mean of schooling in this dataset 13 years is not low. 13 years of education equals to the level of having a year of college education. This mean of education is similar to the average of education in Ahenfelter and Krueger's paper (14.11 years) and Card's paper (13.2 years) and even higher than the mean education in Maluccio's paper (8.9 years). Since the low OLS estimate is not due to less educated participants, schooling must be recorded with error in this dataset, which indicate the existence of measurement error.

Augmented Wage Function and IV Regression

In this paper, OLS estimate of return to education for the augmented wage function is 5.87%, which is 17.4% higher then the OLS estimate of the basic wage function. This change in estimator indicates that the control variables, including gender, interaction of gender and education, industry, region, and economic background can significantly influence return to education. Adding these variables into the wage function can effectively control influence from factors other than schooling on income. Previous research has also confirmed the significant influence of these factors on return to education. Therefore, adding these variables into the wage function can and has effectively controlled the external effects, which allows us to examine the return to schooling more accurately. For the comparison between OLS estimates and IV estimates, I will use the augmented wage function as the regression function.

The IV estimate and OLS estimate are significantly different. Return to education under using the method of IV is 104% larger compare to using OLS estimation, from 5.87% to 10.2%. This result is similar to the findings in previous research. Ashenfelter and Krueger (1992), Card (1993), and Maluccio (1998) also found that IV estimates are significantly larger than the OLS estimates. Ashenfelter and Krueger examines measurement error in schooling by using sibling reported education as an instrument. They explain that between 8 and 12 percent of this difference in estimation is due to the existence of measurement error. Maluccio uses distance to secondary school as an instrument to estimate return to schooling. He explains this significant difference as a result of binary instrumental variables. Binary instruments are dummy instruments, such as the instrumental variable he uses *distance to schooling*. This variable may pick up those individuals who stopped schooling earlier and generate a higher average marginal return. This intuition is similar to the reason why IV estimates are higher than the OLS estimates in my paper.

From the results, it is obvious to see that return to education estimated with OLS is downward biased. This effect is mainly caused by two problems: the correlation between education and measurement error and measurement error in education. Consider the following model:

$$y^* = \beta x^* + u \tag{1}$$

Where y* and x* denote for the true values of dependent variable and explanatory variable; *u* represents the error term in the equation. First, when explanatory variable is correlated with the error term, the assumption that *Cov* (x*, u) = 0 has been violated. Therefore, OLS becomes biased and inconsistent. In my study, since education is covariate with ability, the assumption that *Cov* (x*, u) = 0 has been violated. Therefore, OLS estimate is no longer consistent and ability bias downward biased the results. Second, inaccurate measurement in education will biased OLS estimates downward. If there is measurement error in explanatory variable, we can write the biased explanatory variable as:

$$x = x^* + e$$

where *x* is the observation value and *e* is the inaccuracy in measurement. The equation can be rewritten as:

$$x^* = x - e$$

Substituting x* into equation (1):

$$y^* = \beta(x - e) + u$$

$$\text{or} \quad y^* = \beta x + (u - \beta e)$$

The OLS estimate is biased and inconsistent because *x* is correlated with the error term $u - \beta e$. When β is larger then 0, x and $u - \beta e$ are negatively correlated and OLS estimator will be biased downwards. Since the dataset used in my paper is prone to measurement error, my OLS estimates can be lower than OLS estimates from other literature.

After applying the method of IV to estimate return to education, the rate increases 107% from 5.87% to 10.2%. Since I use parental education as instruments, this difference in estimation is partially due to the effect of parental education. The increasing in estimation is due to the instrumental variable *parental education* picks up observations who have less schooling with less educated parents. Therefore, return to education increases compare to OLS estimates.

Validity of Instrument

However, can the method of IV correct the OLS estimates and accurately depict the relationship between schooling and income? To answer this question, I need to examine the robustness of the instrument, which is the validity of using parental education as instruments. I will examine this legitimacy from two aspects. First, I will check the satisfaction of the two assumptions by interpreting the first-stage regression results and explaining analytically. Then, I will categorize parental education in a different method than the previous way to prove that getting previous results is not due to the way how I categorize parental education but because parental education can truly affect children's education and return to education.

First, I will examine how parental education meets the two assumptions. As discussed in previous section, an instrumental variable needs to fulfill two criteria: be related to the endogenous explanatory variable and being exogenous in equation. Table 4 shows the results of first-stage OLS estimates of education and interaction between education and gender. First stage test examines the first assumption of instrument variable, which is be related to the endogenous explanatory variable. When doing first-stage test, the endogenous variable will be set as dependent variable and put on the left hand side of the regression function. Instrumental variables, exogenous regress or from the structure function, and the error term will be put on the right hand side. If the structural equation looks like:

$$y = \beta_1 x_1 + \beta_2 x_2 + u$$

Where y is the dependent variable; x_1 is the endogenous variable; x_2 is the exogenous variable; u is the error term. The first-stage equation will look as follow:

$$x_1{}' = \gamma_1 x_2 + \gamma_2 z + v$$

Where $x_1{}'$ estimates x_1; z is the instrument treating x_1, and v is the error term (Rose and Fjelstad). Estimators of instruments indicate that the parental education is significantly related with education, which means that the assumption that instrumental variable needs to relate to endogenous variable is qualified. Here, parental education has significantly positive influence on children's schooling. Furthermore, the results of F-test on instruments is large enough to prove that parental education is a valid instrumental variable.

The second assumption of instrument, not correlating with the error term, is hard to test since this relationship involves unobservable error. Therefore, we can only maintain $\mathrm{Cov}\,(z, u) = 0$ by appealing to economic behavior (Wooldridge, 2003). In this study, this assumption can be translated to as covariance between parental education and ability should be zero. In most cases, parental education should be irrelevant with children's ability, since some ability can be talent and parents do not have intervene in natural endowment. Moreover, it is possible for children with less educated parents to earn advanced degree. Therefore, parental education is not correlated with children's ability. However, for some ability that can be acquired and improved through practicing, parental education may play a role when helping their children to learn and advance skills. For example, higher educated parents may know the importance of education better then less educated parents and may strongly emphasize on training the valuable skills for the younger generation. Furthermore, highly educated parents are more likely to have higher income and therefore can provide sufficient financial resources for their children to improve their ability compare to less educated parents. Thus, parental education may correlate with ability and the second assumption about instrumental variable may be violated.

Furthermore, I re-categorize parental education into two groups: less educated group and higher educated group. Less educated group is represented by 0 and indicates the education level from having 0 years of education to high school graduates. To be categorized into higher educated group, subjects need to have at least some college or have advanced and professional degrees. The first-stage regression results are shown in Table 7. It is obvious to see that parental education is significantly related with children's education. Children with fathers having bachelor's degrees will have 0.59 more years of schooling compared to children whose fathers do not have college degrees. Children with college-graduated mother will have 0.703 more years of schooling compared to children whose mother have not attended any college. The fir-stage estimates suggest that the new defined parental education fulfils the assumption that it is related with children's education.

Table 7 First-stage OLS Estimates of Education and Interaction of Gender and Education

Wage equation variables	education	gender × education
experience	-0.0144	0.00442
	(0.0107)	(0.0091)
expenrience^2	-0.000238	-0.000468***
	(0.0002)	(0.000178)
gender	-.000258	12.673***
(female = 0, male = 1)	0.1079	(0.0915)
Instrumental variables		
Father's education	0.590***	-0.150
	(0.164)	(0.139)
Mother's education	0.703***	-0.176
	(0.155)	(0.131)
Father's education × gender	0.305	1.128***
	(0.206)	(0.174)
Mother's education × gender	-0.175	0.761***
	(0.193)	(0.163)
constant	14.897***	1.666**
	(0.891)	(0.765)
observations	2,583	2,583
F-test exclude instruments	36.72	46.3

Table 8 shows the comparison of IV estimates of using the old categorizing method and the new method. Column 1 represents the estimators using the old separating method, whereas column 2 shows the estimators using the new grouping method. IV estimate for return to education with using the old method is 10.2% and the estimate is also 10.2% with using the new categorizing method. The standard deviation is slightly different in these two cases. Estimators for return to education are extremely similar between using the two grouping methods. Therefore, we can conclude that the results we get are not due to the way how we categorize parental education but because of the true effect it has.

Table 8 Estimations of Return to Schooling with IV Method
(based on augmented wage function)

	(1)	(2)
Variables	lnwage	lnwage
education	0.102***	0.102***
	(0.0179)	(0.0193)
experience	0.0234***	.0230***
	(0.00186)	(0.00184)
$experience^2$	-0.000324***	-.000328***
	(3.57e-05)	(3.52e-05)
gender	0.378	.7002***
(female = 0, male = 1)	(0.251)	(.267)
gender_edu	-0.0190	-.0435**
	(0.0190)	(.0203)
constant	1.403***	1.441***
	(0.296)	(.3107)
observations	2,583	2,583
R-squared	0.311	0.336

Limitation

As discussed above, of the two assumptions of instrumental variable, only one of them is fully satisfied. Therefore, parental education is not a perfect instrument to estimate schooling and the IV estimates may be inaccurate this way.

The second limitation to this study is that the measurement error in schooling cannot be treated. Using parental education as instruments take care of ability bias when estimating return to education but ignore the influence of measurement error has on the return. The major reason that why measurement error cannot be treated is that PSID is a survey based database that only offers self-reported schooling. Since there is no other method of recording education in PSID, I cannot compare and examine the effect measurement error in this dataset. For example, Ashenfelter and Krueger (1992) uses sibling-reported education as instrument for schooling to find out the effect measurement error. Since I only have self-reported data, this method is not applicable for this study.

V. Conclusion

In this study, I use the method of IV to decrease the effect of ability bias on return to education. I choose to use the father's and mother's education as instrumental variables because family background can significantly affect children's education. The OLS estimator of return to education is 5.9%, whereas the IV estimator is 10% for the United States in 2012. The IV estimate is 10.2%, which is 107% larger than the OLS estimate, a statistically significant difference.

I then test the validity of using the education of both parents as instrumental variables. The first-stage OLS estimates show that the first assumption of parental education being correlated with education is satisfied. However, the second assumption that education is uncorrelated with ability is not fully satisfied because parental education is relating to children's level of ability, for reasons that were beyond the scope of the current study. Therefore, parental education is not a perfect IV.

Knowing the 10.2% return to education and the positive relationship between parental education and children's education is crucial for making more effective policies that leverage the influence of parents. I believe that the major reason why parental education and children's education are positively related is that higher-educated parents understand the importance of education better then less-educated parents and therefore may encourage their children to get more education. Therefore, broadcasting the significance of education to less-educated parents can increase the schooling of their children and increase the efficiency of government expenditures on education.

I have two suggestions for future studies. First, researchers could make improvement on the current study by finding instruments that are not correlated to ability but still have significant influence on education, although this would be challenging. Researchers could include more instruments into the instrument sets in order to estimate return to schooling more accurately. Another suggestion is to use twins' data in future studies to treat ability bias and measurement error at the same time. Ability bias in twins is very small because twins are considered genetically identical and grow up with the same family background. Using both self-reported schooling and sibling-reported schooling can detect and decrease the influence from measurement error (Krueger and Ashenfelter, 1992).

Second, future research can initiate a new research topic on exploring the correlation between parental education and children's ability and reasons why they are correlated from both nature and nurture aspects. This would help us better understanding the effect of education on us and future generations as well as the significance of getting education.

References

Angrist, J. D., & Krueger, A. B. (1990). "Does Compulsory School Attendance Affect Schooling and Earnings?" (No. w3572). *National Bureau of Economic Research*.

Altonji, J. G., & Dunn, T. A. (1996). "The Effects of Family Characteristics on the Return to Education". *The Review of Economics and Statistics*, 692-704.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* New York: Columbia University Press.

Bound, J., & Jaeger, D. A. (1996). "On the Validity of Season of Birth as an Instrument in Wage Equations: A Comment on Angrist & Krueger's Does Compulsory School Attendance Affect Schooling and Earning?". *National Bureau of Economic Research,* No. w5835.

Card, D. (1993). "Using Geographic Variation in College Proximity to Estimate the Return to Schooling". *National Bureau of Economic Research*, No. w4483.

Card, D. (2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems". *Econometrica*, *69*(5), 1127-1160.

Garay, P.V., Zereyesus, Y.A., & Thompson, A. (2004). "Making Every Dollar Count: Local Government Expenditures and Welfare". *Modern Economy*, Vol.5 No.1.

Garen, J. (1984). "The Returns to Schooling: A Selectivity Bias Approach with a Continuous Choice Variable". *Econometrica*, 52, 1199-1218.

Griliches, Z. (1977). "Estimating the Returns to Schooling: Some Econometric Problems". *Econometrica: Journal of the Econometric Society*, 1-22.

Maluccio, J. (1998). “Endogeneity of Schooling in the Wage Function: Evidence from the Rural Philippines”. *Food Consumption and Nutrition Division Discussion Paper*, 54.

Mincer, J. A. (1974). “Age and Experience Profiles of Earnings. Schooling, Experience, and Earnings”. *National Bureau of Economic Research*, December 1994, 1157-1173.

Krueger, A., & Ashenfelter, O. (1992). “Estimates of the Economic Return to Schooling from a New Sample of Twins”. *American Economic Review*, December 1994National Bureau of Economic Research, No. w4143.

Rose, E., & Fjelstad, J. *Cheat Sheet for Instrumental Variable*. *http://courses.washington.edu/pbafadv/student%20presentations/IVCheatSheet-FjelstadRose.pdf*

Wooldridge, J.M. (2003) *Introductory Econometrics: A Modern Approach*. Mason, Ohio: South-Western College Press.