Giresun Üniversitesi

İktisadi ve İdari Bilimler Dergisi

Dergi Sayfası: http://dergipark.gov.tr/guiibd

THE RETURNS TO FORMAL SCHOOLING IN TURKEY USING PSEUDO-PANEL DATA

Hacı Mevlüt KARATAŞ[1]

MAKALE BİLGİSİ

Makale Geliş Tarihi: 01.11.2018
Makale Düzenleme: 05.12.2018
Makale Kabul Tarihi: 28.12.2018

Anahtar Kelimeler:
Returns to education,
Pseudo panel,
Fixed effects,
Endogeneity,
Turkey

JEL Sınıflandırması: I20, I26, C23

ABSTRACT

The returns to education is of great interest in public policies and labor economics. It has been widely studied by many experts for decades. It has been acknowledged that in a Mincerian wage equation, ordinary least squares estimates are biased due to the endogeneity of education. One way to deal with this endogeneity could be removing individual fixed effects using panel data. However, the education of an individual is fixed once their wage is observed. Thus, the fixed effects model in panel data would wipe out the information on education. On the other hand, using a pseudo-panel fixed effects approach, the returns to education can be estimated making use of repeated cross-section data. There are limited studies on the returns to education in Turkey despite the fact that its crucial importance in the public policy. In this paper, we estimate the returns to education for Turkey using a pseudo-panel data approach. We make use of Turkish Household Labor Force data for the years 2009-2014 in order to construct pseudo-panel data. We find that one additional year of education increases individual wages by around 8.5 percent using ordinary least squares. However, using pseudo-panel fixed effect estimation leads a 9.3-percent rate of returns to education which show that there is a downward bias in ordinary least squares estimates of returns to education in Turkey.

ABSTRACT

The returns to education is of great interest in public policies and labor economics. It has been widely studied by many experts for decades. It has been acknowledged that in a Mincerian wage equation, ordinary least squares estimates are biased due to the endogeneity of education. One way to deal with this endogeneity could be removing individual fixed effects using panel data. However, the education of an individual is fixed once their wage is observed. Thus, the fixed effects model in panel data would wipe out the information on education. On the other hand, using a pseudo-panel fixed effects approach, the returns to education can be estimated making use of repeated cross-section data. There are limited studies on the returns to education in Turkey despite the fact that its crucial importance in the public policy. In this paper, we estimate the returns to education for Turkey using a pseudo-panel data approach. We make use of Turkish Household Labor Force data for the years 2009-2014 in order to construct pseudo-panel data. We find that one additional year of education increases individual wages by around 8.5 percent using ordinary least squares. However, using pseudo-panel fixed effect estimation leads a 9.3-percent rate of returns to education which show that there is a downward bias in ordinary least squares estimates of returns to education in Turkey.

Key Words: returns to education, pseudo panel, fixed effects, endogeneity, Turkey

JEL Classifications: I20, I26, C23

[1] Arş. Gör. Dr., Giresun Üniversitesi, İİBF, Ekonometri Bölümü, e-mail: hmkarata@syr.edu

INTRODUCTION

The relationship between education and earnings is in the heart of labor economics since the establishment of Human Capital model based on seminal works of Gary Becker in the early 1960s. The effect of education on individual earnings is extensively studied employing Mincer's (1974) earning function. The rate of return to education has been provided in hundreds of studies for different samples and different time periods in the last half-century. These estimates suggest that formal education is one of the most important components determining individual earnings. Psacharopoulos (1985, 1994), and Psacharopoulos and Patrinos (2004) provide an extensive review on the return to education for tens of countries and find that the average return to education is around 10%.

Although the popularity of the basic Mincerian earning function, the ordinary least squares (OLS) coefficient estimate of the rate of return suffers from the bias caused by the endogeneity of the years of schooling in the equation. The literature on the returns to education has been growing towards to the methods that are trying to deal with this bias. For example, the instrumental variables method, twins fixed effect studies, panel data methods are among the most popular methods.

There is mixed evidence on the source, direction, and size of the bias in OLS estimates and the true rate of return to education. The size and the direction of the bias may reveal itself in an opposite manner depending on the source of the bias, such that two possible biases induced by different sources may cancel each other. Griliches (1977) calls attention on this issue by introducing an ability measure explicitly into the Mincer (1974)'s specification. He points out that however omitted ability variable would have resulted in an upward bias in OLS estimates, it is also reasonable to work in the opposite way and induce a downward bias. Moreover, Lang (1993) introduces the "discount rate bias" arisen from the heterogeneous discount rates of individuals. Furthermore, Card (2001) points out that a measurement error bias would account for a 10-percent downward bias in the schooling coefficient estimated by OLS.

In this study, we use a fixed effect method in pseudo-panel data to provide a new piece of evidence on the estimates of the return to education in Turkey controlling for possible bias. We make use of the repeated cross-section data of the Turkish Household Labor Force Survey (THLFS) for the years 2009 through 2014. We restrict our dataset for individuals aged 25-54 in the first year of survey, 2009. We find that the returns to education is 9.3 percent controlling for the endogeneity. We also show that "discount rate bias" is the dominant source of bias drives the OLS results downward in lower levels of education, but in high school levels ability bias is the dominant source of the bias and drives OLS estimates upward.

The rest of the paper is organized as follow. Section 2 briefly summarizes the previous literature, while section 3 describes the data. Section 4 defines the model. Section 5 provides the details of construction of pseudo-panel data. The results are given and discussed in section 6. Section 7 concludes the paper.

1. Previous Literature

Right after the first generation of studies on the estimation of the returns to education that use cross-section data employing basic Mincer (1974) equation, the endogeneity of the individual choice of the years of schooling attracted attention in the literature. The unobservable "ability" variable had been of great interest among the labor economists, because of its likely positive effect on both earnings and the educational attainment. That is, a more able person could get more years of schooling and earn higher wages. Since "ability" is unobservable, an estimate of the coefficient of schooling in the least squares estimation would capture a combined effect of both formal years of schooling and ability. Hence, an upward bias is expected in OLS estimate of the rate of returns to education. Griliches (1977) introduces a measure of IQ as a proxy to the unobserved ability in earning function and finds that the bias induced by the omitted ability variable is small. In fact, in some cases, it would even result in a downward. Later, Lang (1993) points out that heterogenous individual discount rates would be another source of bias. That is, individuals discount future in different rates such that some people would rather lower earnings in the present than higher future earnings as a result of the higher education and drop out the school in early life. Hence, failing to take this different discount rates into account would introduce an upward bias in OLS estimates. This bias is called as " discount rate bias" by Lang (1993). Moreover, the schooling variable is likely to be measured with an error which renders OLS estimates bias. Card (2001) points out that the measurement error introduces a 10-percent bias, on average, in the least squares estimates.

The literature has been devoted a great deal of attention on these bias in the least squares estimates of the schooling coefficient and searched ways to obtain the true rate of returns to education. Those include use of instrumental variables taking advantages of natural experiments (e.g., Angrist and Krueger, 1991; Harmon and Walker, 1995; Card, 1993; Kane and Rouse, 1993; Uutsitalo and Conneely, 1998; Duflo, 2001), exploiting the differences between siblings or twins in their level of education and earnings (e.g., Ashenfelter and Krueger, 1994; Ashenfelter and Rouse, 1998), using genuine panel data methods (e.g., Angrist and Newey, 1991). However, each of them has some drawbacks. First, instrumental variables based on a natural experiment would provide an estimate only on those who are affected by the treatment. This phenomenon is known as the local average treatment effect (LATE) (Imbens and Angrist, 1994). Second, siblings/twins studies are not convincing enough on explaining the differences between twins which shares similar ability and family background. Third, panel data methods would have been a good candidate for removing individual heterogeneity, or fixed effects only if the individual wages are not observed after completing the education. That is, individual wages can only be observed once the individual finishes the schooling and start to work. Once the wages are observed the years of schooling is fixed. Hence, panel data fixed effects cannot provide an estimate for the coefficient of schooling. However, some studies use panel data in which some individuals working while they are in school and provide estimates of the returns to schooling. But these only can be valid for those who are part-time workers and/ or at the lower end of the wage distribution.

Having mentioned the traditional methods and their drawbacks, a relatively new approach, pseudo-panel methods would provide us with a piece better evidence on the true rate of return to education. The pseudo-panels has been using in estimating various economic relationships, however, it is only in the last decade started to be used in estimating the relationship between education and earnings.

Warunsiri and McNown (2010) estimate the rate of return to education in Thailand using 20-year of repeated cross-sections of Thailand's Labor Force Survey. They construct pseudo-panel data by grouping individuals based on the birth cohorts which becomes the cross-section unit of the synthetic panel. The averages of variables of interest for each group of individuals are obtained and followed for the survey years. Employing a basic Mincer (1974) equation, they find that there is a downward bias in OLS estimate of the return to education based on cross-section data.

Following a similar methodology of Warunsiri and McNown (2010), Himaz and Aturupane (2016) estimate the rate of return to education for Sri Lanka. They make use of 9 years of repeated-cross-sections of Sri Lanka's Household Labor Surveys for years between 1997 and 2008. They use birth cohorts as cross-section unit of a synthetic panel data. It is found that there is an upward bias in OLS estimate of the return to education in Sri Lanka.

Moreover, in a developed country context Kirby and Riley (2008) use 10 years of UK's repeated-cross-sectional surveys to estimate the external returns of schooling defining cohorts in industry level. Furthermore, Fulford (2014) uses India's five waves of cross section surveys which span for 20 years. He constructs the pseudo-panel by grouping individuals in five-year birth cohorts by the state of residence. He finds that one extra year of education increases per capita consumption by 3-4%.

The estimates of schooling returns in Turkey are mostly obtained by employing OLS method. Some studies deal with selectivity problem while most of them provide OLS and MLE estimates for different levels of education for several disaggregated samples. There are some comparisons of the returns to education in different points of time. Moreover, some studies offer evidence on the effects of education on wage inequality. Endogeneity of the years of schooling is acknowledged but there is no published study dealing with this problem.

Anne Krueger (1972) has done the very first study on the rate of returns to schooling in Turkey using two surveys which are conducted in 1968 by Turkish Association of Metal Manufacturers and by US Military. She finds 20% and 25% of private returns for secondary and university education, respectively. On the other hand, Tansel (1994) is the first study that uses a Mincerian wage equation to estimate the returns to education in Turkey. She uses 1987 Household Income and Consumption Expenditure Survey (HICES). Her main finding is that the returns to education increase with the level of education. She also finds that the return to education for vocational high school is higher than those for general-curricula high schools which contrasts the findings in the developed countries. Furthermore, she finds that the rate of returns for males is higher than for females, except in primary and middle school levels. Tansel (2008) and Tansel and Bodur (2012) re-examine the returns to schooling and compares for different years by using HICES data for 1994 and Household Labor Force Surveys for years 2002-2005. They use Heckman two-step estimator and OLS estimator in estimating the rate of returns to education. They find no significant difference between the two estimates obtained from those methods. They conclude that the rate of return is higher for women than for man by a magnitude of 2-5%. Furthermore, they conclude that the rate of returns is declining between 1994 and 2004 for men and women.

Vural and Gulcan (2008) use the Turkish Household Budget Survey (HBS) for the years 1994 and 2004 to compare the rate of return to schooling over time. They find that the rate of returns to education is slightly higher for men than for women. Moreover, they show that the rate of returns has been increased for both men and women in this period. Bakis (2012) supports the findings that the returns to education has been increased by 2.5 percentage points from 1988 to 2008 and it is around 10% in 2008. Moreover, Guris and Caglayan (2012) provide some evidence that schooling returns are higher for women than those for men using 2003 and 2006 HLFS and employing OLS, and robust and resistant regression techniques.

Besides, there are some recent working papers that are trying to deal with the endogeneity of the years of schooling in a Mincerian equation. Those studies provide some evidence on the returns to education in Turkey using instrumental variables method based on natural experiments (e.g, Mocan, 2014; Aydemir and Kirdar, 2015).

2.Data and Samples

The data used in this paper is the Turkish Household Labor Force Survey (THLFS) for years between 2009 and 2014 which is obtained from the Turkish Statistical Institute (TURKSTAT). The THLFS is a cross-sectional individual-level survey which provides rich individual characteristics on a sample of around 500,000 observations in each year. The THLFS covers the whole population except non-residents, those who are living in institutions and conscripts in the territory of Turkey.

The THLFS provides labor-related information only for those who are 15 years old and above as well as demographic information for all individuals in the sample. Our main estimates use individuals who were born between 1955 and 1984. Moreover, wage information is available only for regular and casual workers, who are employed, therefore; employer, self-employed and unpaid family workers are excluded from the sample. The reported wage is monthly take-home-pay, that is, wage after deduction of taxes, compulsory social security, and other life insurance premiums. Working hours are reported as the hours worked in a week usually. Those who report working hours more than 84 hours in a week are excluded. Furthermore, individuals who started to work in the survey month or didn’t report their wage information are also excluded from the sample.

Table1: Summary Statistics for the Individual Sample 2009-2014

VARIABLES	# of Obs.	Mean	Std. Dev.
Age	388,352	38.64	7.663
Female*	388,352	0.234	0.423
Married*	388,352	0.849	0.358
Years of Education	388,352	9.505	4.196
Levels of Education			

Illiterate*	388,352	0.012	0.108
No Diploma Holder*	388,352	0.027	0.163
Primary School*	388,352	0.346	0.476
Middle School*	388,352	0.112	0.316
General High School*	388,352	0.116	0.320
Vocational High School*	388,352	0.123	0.329
College*	388,352	0.275	0.447
Hourly Wage	388,352	3.195	2.238
Hours Worked in a Week	388,352	49.69	13.02

+ Variables with * are dummy variables.

The dependent variable in our estimations is the natural logarithm of hourly wages calculated as the monthly wages divided by the usual working hours in a month. All wages are deflated to 2003 liras using consumer price index (CPI) for each NUTS2 regions provided by TURKSTAT. Upper and lower 1% of the distribution of the wages for each year are trimmed in order to exclude the influential observations. The variable of interest is either the years of education or levels of education. The education is recorded as the completed level of education in surveys, we use dummy variables for each level of schooling given in the surveys.

The college education may consist of 2-6 years of education based on the field and the university, although four year of college education is the most common one, and on top of that graduate studies may differ between 2-7 years. However, the THLFS doesn't distinguish between two-year college, four-year college and post-college education. Thus, we assume four additional years of education on top of high school education for all individuals went to college. Summary statistics on the variables of interest is presented for the individual sample in Table 1.

3. The Model

We use pseudo-panel data fixed effects method as an alternative way of estimating the returns to education in Turkey. We first construct a "synthetic" panel data defining cohorts as the year of birth and also the year of birth interacted with regions. Then using within estimation method we estimate the rate of return to education. The underlying model can be expressed in a genuine panel data setting as;

$$\ln w_{it} = \beta_1 S_i + \beta_2 X_{it} + \mu_i + u_{it}$$

(1)

where the dependent variable is the natural logarithm of hourly wages of individual $i,$ at time t. Schooling variable, S_i, is time-invariant. Schooling variable is likely to be correlated with unobservable individual characteristics, so the pooled OLS estimation ignoring the time component in the Eq (1) would lead bias estimates of the coefficient of schooling. Moreover, within estimator wipes out the unobservable fixed effects and bias can be removed, but this method wipes out schooling variable, as well. In this case, the genuine panel data fixed effects method does not provide an estimate for the schooling coefficient. However, following Deaton (1985) and Warunsiri and McNown (2010), we can construct pseudo-panel (synthetic panels) data making use of the rich repeated cross-sectional data. In this case, Eq (1) can be re-written as;

$$\ln w_{i(t)t} = \beta_1 S_{i(t)t} + \beta_2 X_{i(t)t} + \mu_{i(t)t} + u_{i(t)t}$$

(2)

$i(t)$ indicates that we observe a different sample in each survey year, and we cannot follow the same individuals over time. Also note that the unobserved individual characteristics vary over time because of the different samples in each survey year. Moreover, it is likely that unobservable individual characteristics are correlated with the schooling variable such that the OLS estimator with pooled data would be biased. To deal with this bias we define cohorts based on some individual characteristics that do not change over years and can be observed in each survey year as described in Deaton (1985). Then, taking averages of the variables of interest over all individuals within each cohort-year cell, we construct pseudo-panel data. Hence, the Eq (2) can be expressed as;

$$\overline{lnw}_{ct} = \beta_2 \bar{S}_{ct} + \beta_2 \bar{X}_{ct} + \bar{\mu}_{ct} + \bar{u}_{ct}$$

(3)

Where the dependent variable is the average of the natural logarithm of hourly real wages of cohort c at time t and the variable of interest is the average of years of schooling (or completed level of school) of cohort c at time t. The rest of the control variables are defined in a similar way. Note that the fixed effects, $\bar{\mu}_{ct}$, is now time-varying and the conventional within estimation model will not be able to wipe them out. Deaton (1985) suggest that if we have large enough samples, $\bar{\mu}_{ct} \rightarrow \mu_c$. Therefore we can re-write Eq (3) as;

$$\overline{lnw}_{ct} = \beta_2 \bar{S}_{ct} + \beta_2 \bar{X}_{ct} + \mu_c + \bar{u}_{ct}$$

(4)

where μ_c is the cohort fixed effects. Here we can think of the cohort means, obtained from the sample, as estimates of their counterparts in the population cohort with some sampling error. Verbeek and Nijman (1993) show that if the cohorts have at least 100 observation, sampling error can be negligible.

4. The Construction of Pseudo-Panel (Synthetic Panel)

In generating the pseudo-panel data out of repeated cross sections, we follow a strategy close to Warunsiri and McNown (2010), Deaton (1985) and Fulford (2014). We start by defining the grouping variable that is not changing over time and is observed in each survey year. The natural candidate is the birth year of individuals. However, because we only have 30 single birth year, the sample size is small. In order to increase the sample size of pseudo-panel, we also use the region of residence as our second grouping variable. The region of residence is given as 12 NUTS1 regions or 26 NUTS2 regions in the THLFS. Verbeek and Nijman (1993) point out the trade-off between the number of cohorts and the size of cohorts. That is, the bigger the size of the cohort, the smaller the number of cohorts. Following the pseudo panel literature first, we defined 30 birth cohorts for those who were born between 1955 and 1984. Note that the birth cohorts become the cross-section units. Then, following these cohorts over 6 years, 2009-2014, we have a panel data on cohorts, which is called pseudo panel, or synthetic panel. Antman and McKenzie (2007) argue that although the pseudo panel estimator is consistent given large cell sizes, the standard errors would be higher than those obtained using a genuine panel data since the speed of convergence is based on the cell size rather than the sample size. Also, the small number of cohorts introduce higher standard errors. Considering this problem, we can utilize the large size of the sample to define a large number of cohorts. Following Propper, Rees, and Green (2001) we construct another set of cohorts as year-birth cohort interacted with 12 NUTS1 regions. For example, those who were born in 1960 and live in region TR2 consists of a group, those born in 1961 in the same region TR2 consists another group, those born in 1961 in region TRA consist another group and so on. In this way, we can obtain a relatively large sample. Moreover, it is expected to have small variation within each group cell but a higher variation between cells. After defining our cross-section units for a pseudo-panel setting, we take averages of each variable over individuals in each cell for each survey year. All cohort averages are weighted by the sampling weights provided in the

THLFS data sets. We are in favor of constructing the pseudo panel based on single birth year by 12 NUTS1 regions, in which 360 cross-section unit is followed for 6 years, however, we provide estimates based on pseudo-panels that is constructed by using single birth year, birth-year by 26 NUTS2 regions, two-year-birth cohorts interacted with 12 NUTS1 and 26 NUTS2 regions, in order to check robustness of our result.

5. Results

We estimate the return to education using OLS estimator on individual observations as well as pseudo-panel within estimator defining cohorts as single-birth-year, single-birth-year-NUTS1 regions, single-birth-year-NUTS2 regions, two-birth-year-NUTS1 regions, and two-birth-year-NUTS2 regions. We use two different measure of schooling; the years of schooling and dummy variables for completing each level of education. Each regression controls for tenure in the firm and its square as a measure of experience. We could not use age as a proxy for the experience because our pseudo panel construction uses the years of birth and including it would introduce multicollinearity.

Moreover, we control for gender by including a dummy variable for females, as well as for marital status by a dummy variable for married individuals. Furthermore, we control for the time fixed effects in each model, as well as region fixed effects in OLS estimation. The sampling weights are used in obtaining the averages of variables for each cohort-year cell. Lastly, within estimations are weighted by the square root of cell sizes to control for the heteroskedasticity arisen from the substantially different sizes of cohorts as in Dargay (2007).

Table 2: Returns to Education (Years of Education)

		Fixed-Effects (FE) Estimates in Pseudo-Panel Data				
VARIABLES	OLS	Single-Birth Year	**Single-Birth Year by NUTS1**	Single-Birth Year by NUTS2	Two-Birth Year by NUTS1	Two-Birth Year by NUTS2
Years of Education	0.085***	0.095***	**0.093*****	0.092***	0.096***	0.094***
	(0.0002)	(0.011)	**(0.003)**	(0.002)	(0.004)	(0.003)
Tenure	0.033***	0.041***	**0.044*****	0.045***	0.047***	0.049***
	(0.000)	(0.006)	**(0.003)**	(0.003)	(0.004)	(0.003)

Tenure	-0.0003***	-0.0004	**-0.0007*** **	-0.0007***	-0.0008***	-0.0009***
Squared	(0.0000)	(0.0003)	**(0.0001)**	(0.0000)	(0.0002)	(0.0001)
Female	-0.013***	0.057	**-0.056* **	-0.050**	-0.053	-0.030
	(0.002)	(0.085)	**(0.034)**	(0.024)	(0.046)	(0.034)
Married	0.123***	0.449***	**0.276*** **	0.195***	0.309***	0.228***
	(0.003)	(0.064)	**(0.0342)**	(0.0256)	(0.0456)	(0.035)
Constant	-0.119***	-0.639***	**-0.468*** **	-0.404***	-0.527***	-0.461***
	(0.004)	(0.079)	**(0.037)**	(0.028)	(0.049)	(0.037)
Observations	388,352	180	**2,160**	4,680	1,080	2,340
R-squared	0.505	0.975	**0.887**	0.847	0.926	0.890

+ Robust Standard Errors in parentheses*** p<0.01, ** p<0.05, * p<0.1. + Each model controls for region and time fixed effects. + Sampling weights, provided in the THLFS, are used for obtaining the sample averages of the variables in the construction of the pseudo-panels. + Square root of the number of observations in each birth-year (birth-region-year) cell is used as weights in the FE estimations.

The first column of Table 2 shows the OLS estimates of the model utilizing individual observations. It is found that one additional year of schooling increases individual hourly wages by 8.5 percent which is in line with what is found in the literature. The third column of Table 2 shows the pseudo-panel fixed effect estimates of the return to education with the cross-section unit of the pseudo panel being single-birth-year-NUTS1 regions. It is found that once we take care of endogeneity of the years of schooling in the earning function, the returns to one year of schooling is 9.3 percent. Thus, the schooling coefficient estimate with the least squares is downward bias, even though this bias is not as big as in the studies using instrumental variables or in Warunsiri and McNown (2010). Furthermore, we checked our results for various ways of construction of the pseudo-panels. Column 2 of Table 2 shows the within estimates of the coefficients for the case single-birth-year is the cross-section unit. It shows that one additional year of schooling increases wages by 9.5 percent. Likewise, columns 4-6 of Table 2 provide very similar schooling coefficient estimates.

At this point, we need to conduct another robustness check for our results against our assumption on the region of residence. Throughout the paper, we assume that the region of residence is a fixed characteristic of individuals. Ideally, it would have been better to use the region of education or region of birth in the construction of the synthetic panels.

However, the THLFS has information on the region of residence only. Thus, we are forced to assume that the region of residence is fixed over years. But, because of the migration, this might not be true. Some individuals might attain their education in one region and move to another region in which they currently reside in the course of the sample period. This could jeopardize our results. Therefore, we conduct a robustness check based on the information in the THLFS on whether the individual lives in the same city since birth. That is, we can identify those individuals who have been living in the same city for their entire life. We replicate Table 2 restricting the sample only on those who have been living in their residence city for their entire life. Almost half of the sample does not live in the residence city for their entire life. The results are presented in the appendix in Table A2. It is found that the estimates of the returns to education is lower in both OLS and pseudo-panel fixed effects. This is what is expected, as it is likely that more educated people migrate to another city in a higher rate for work. As a result, we get lower estimates. However, the pattern in comparison of the least squares and within estimation models remains unaltered.

Table 3: Returns to Education for The Levels of Education

		Fixed-Effects (FE) Estimates in Pseudo-Panel Data				
VARIABLES	OLS	Single- Birth Year	**Single-Birth Year by NUTS1**	Single-Birth Year by NUTS2	Two-Birth Year by NUTS1	Two-Birth Year by NUTS2
Primary School	0.128***	0.268	**0.256*** **	0.278***	0.221**	0.254***
	(0.005)	(0.172)	**(0.056)**	(0.037)	(0.089)	(0.053)
Middle School	0.246***	0.361*	**0.457*** **	0.426***	0.420***	0.395***
	(0.005)	(0.198)	**(0.066)**	(0.045)	(0.097)	(0.064)
General-High	0.428***	0.533***	**0.610*** **	0.624***	0.571***	0.588***
	(0.005)	(0.201)	**(0.064)**	(0.043)	(0.096)	(0.063)
Vocational-High	0.464***	0.594***	**0.624*** **	0.649***	0.635***	0.627***
	(0.005)	(0.191)	**(0.065)**	(0.044)	(0.097)	(0.063)

College	1.056***	1.269***	**1.247*** **	1.259***	1.242***	1.277***
	(0.005)	(0.192)	**(0.060)**	(0.040)	(0.086)	(0.055)
Tenure	0.0309***	0.0466***	**0.0465*** **	0.0461***	0.0502***	0.0492***
	(0.0003)	(0.007)	**(0.003)**	(0.002)	(0.004)	(0.003)
Tenure Sq	-0.001***	-0.001**	**-0.001*** **	-0.001***	-0.001***	-0.001***
	(0.0001)	(0.0001)	**(0.0001)**	(0.0000)	(0.0002)	(0.0001)
Female	-0.053***	-0.0479	**-0.106*** **	-0.108***	-0.121**	-0.110***
	(0.002)	(0.089)	**(0.034)**	(0.024)	(0.050)	(0.035)
Married	0.096***	0.359***	**0.238*** **	0.177***	0.261***	0.204***
	(0.01)	(0.062)	**(0.033)**	(0.025)	(0.044)	(0.033)
Constant	-0.363***	-0.271*	**-0.176*** **	-0.144***	-0.184**	-0.156***
	(0.022)	(0.162)	**(0.059)**	(0.040)	(0.090)	(0.057)
Observations	388,352	180	**2,160**	4,680	1,080	2,340
R-squared	0.535	0.978	**0.892**	0.854	0.929	0.896

+ Robust Standard Errors in parentheses*** p<0.01, ** p<0.05, * p<0.1. + Each model controls for region and time fixed effects. + Sampling weights, provided in the THLFS, are used for obtaining the sample averages of the variables in the construction of the pseudo-panels. + Square root of the number of observations in each birth-year (birth-region-year) cell is used as weights in the FE estimations.

Having shown that there is a downward bias in the least squares estimates of schooling coefficient, we can check whether the returns to education differs by levels of education. Note that using the years of schooling we assume a linear average return to education for each year of education irrespective of the levels of education.

The bias seems to be small comparing to other studies. Thus, having to look at the rate of returns to education for levels of education would provide us with a better understanding of the true relationship between education and wages. Hence, we estimate the earning function by including a dummy variable for each completed level of education. Those who are illiterate or literate but did not finish a level of school is considered as the base level. The results are shown in Table 3.

Column 1 of Table 3 shows the OLS estimates of returns to each level of education with respect to the base level, which is illiterate or literate but not a degree holder using individual-level data. It is found that a primary school graduate has 12.8 percentage points higher wages compared to those do not graduated from primary school. It is also found that having completed three extra years after primary school, that is completing middle school education, would lead an 11.8-percentage-point increase in the individual wages. Moreover, having completed general-curricula high school result in an extra 18.2-percentage-point increase in individual wages on top of middle school education while having completed a vocational high school over middle school lead a 21.8-percentage-point increase in hourly wages. Lastly, being a college graduate increases wages by 62.8 percentage points compared to a general-curricula high school graduate. The general patterns in this estimation suggest that returns to a degree is increasing by the level of education. Note that individual choice of a completing a degree is endogenous in the model because of the likely correlation between the degree an individual attained and unobservable characteristics.

Thus, pseudo-panel fixed effect model in the third column of Table 3 presents estimates of the returns to levels of education by taking individual unobserved fixed effects into account. The FE estimates, in the third column, indicate that being a primary school graduate increases wages by 25.6 percentage points compared to those are not graduated from any school. Also, having finished middle school allows a 20.1-percentage-point extra returns in wages over primary school. The completion of general-curricula high school over middle school increases wages by 15.3 percentage points, whereas completion of vocational high increases wages by 16.7 percentage points. Furthermore, college graduation over general-high school education leads a 63.7-percentage-point increase in wages. Columns 2 and 4-6 lead similar estimates. Here we compare the returns to education levels, but each level consists of different years of schooling. Considering this, we can calculate the average rate of return to one-year education at each level from Table 3. We assume a linear rate of returns within each level of education. Some levels of education may consist of different years of education based on the type of school. However, we do not have any information in the data set to identify these differences. Thus, we are forced to assume the least year of education required to complete each level of education as the year of education in each level. That is, we assume 5-year primary school, 3-year middle school, 3-year general or vocational school and 4-year college. Table 4 shows the rate of return to one-year of education in the corresponding education levels.

Table 4: Rate of Returns to One-year schooling by Levels of Education (%)

	OLS	Pseudo- Panel FE
Primary School	2.6	5.1
Middle School	3.9	6.7

General-High	6.1	5.1
Vocational High	7.3	5.6
College	15.7	15.9

+ The rates are calculated using the estimates in columns of 1 and 3 of Table 3, respectively.

Table 4 depicts a better comparison among levels of education as well as between two estimators. OLS estimates suggest that the rate of returns to education is increasing with the levels of education. The highest rate of return to one-year education is experienced at college level. However, controlling for endogeneity by employing pseudo-panel fixed effects model we find that the rate of returns to one-year education is higher in middle school than in high schools. The highest rate of return to one year of education is still experienced by college graduates. Note that there is a downward bias in the estimates for primary and middle school level, while an upward bias in the estimates for general curricula or vocational high school. The estimate for college graduates is very close in both columns. These findings might be indicative for the source of bias in the OLS coefficients. It looks like the discount rate bias dominates the ability bias until high school, while ability bias dominates the discount rate bias at the high school level.

As we mentioned earlier, we assume fixed region of residence over years, which could jeopardize our estimates. We check the robustness of our results against this assumption considering only those individuals spent their entire life in the residence region. The results are presented in the appendix in Table A3. The conclusions remain unaffected except for the college. Using only non-mover sample the pseudo-panel fixed effects result in a lower rate of return to one-year education that OLS does. In fact, this backs up our point that the higher educated people tend to move more than lower educated people do.

CONCLUSION

This study estimates the returns to education in Turkey accounting for the endogeneity of schooling. The study takes advantage of rich cross-sectional data of the THLFS for the years between 2009 and 2014 and construct pseudo-panel data. The construction of the cross-section units of the pseudo panel is based on single-birth year by 12 NUTS2 regions. After generating the pseudo-panel data the within estimation method is used to estimate the true rate of returns to education.

The schooling variable in a Mincerian earning function is estimated by considering two kinds of variables; years of schooling and dummy variables for each level of education.

The least squares suggest that one additional year of schooling increases individual wages by 8.5 percent. However, the panel-data fixed effect model suggests a little higher rate of returns to education with a magnitude of 9.3 percent. Hence, we conclude that there is a small downward bias in OLS estimates of the returns to education. We further investigated the rate of returns to one-year education in different levels of education. OLS results indicate that the rate of returns to one-year schooling is increasing with levels of education. However, pseudo-panel fixed effect model suggests that the rate of returns to one-year education at high school level is less than that is at middle school education. Moreover, the comparison of these two sets of estimates suggests that there is a downward bias at primary and middle school levels while there is an upward bias at high school levels of education. This might indicate that the discount bias is more apparent at lower levels of education and ability bias is more apparent at high school level.

Overall, this paper provides a new piece of evidence on the returns to education in a developing country context controlling for the endogeneity of schooling and finds that the returns to education in Turkey is around 10% level which is close to the average returns to education for the whole world. The findings provide some helpful insights for policy makers in designing their education policies.

REFERENCES

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *Quarterly Journal of Economics, 106*(4), 979-1014.

Angrist, J. D., & Newey, W. K. (1991). Over-identification tests in earnings functions with fixed effects. *Journal of Business & Economic Statistics*, *9*(3), 317-323.

Antman, F., & McKenzie, D. J. (2007). Earnings mobility and measurement error: A Pseudo-Panel approach. *Economic Development and Cultural Change*, *56*(1), 125-161.

Ashenfelter, O., & Krueger, A. (1994). Estimates of the economic return to schooling from a new sample of twins. *The American Economic Review, 84*(5), 1157-1173.

Ashenfelter, O., & Rouse, C. (1998). Income, schooling, and ability: Evidence from A new sample of identical twins. *The Quarterly Journal of Economics, 113*(1), 253-284.

Aydemir, A., & Kirdar, M. G. (2015). *Low Wage Returns to Schooling in a Developing Country: Evidence from a Major Policy Reform in Turkey* (No. 9274). Institute for the Study of Labor (IZA).

Bakıs, O. (2012). Türkiye'de egitimin bireysel getirisi: 1988-2008. *TUSIAD-Sabanci University Competitiveness Forum working Paper No. 2012-1.*

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

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, *69*(5), 1127-1160.

Dargay, J. (2007). The effect of prices and income on car travel in the UK. *Transportation Research Part A: Policy and Practice, 41*(10), 949-960.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics, 30*(1), 109-126.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review*, *91*(4), 795- 813.

Fulford, S. (2014). Returns to education in India. *World Development, 59,* 434-450.

Griliches, Z. (1977). Estimating the returns to schooling: Some econometric problems. *Econometrica: Journal of the Econometric Society,* 1-22.

Güriş, S., & Çağlayan, E. (2012). Returns to education and wages in Turkey: Robust and resistant regression. *Quality & Quantity, 46*(5), 1407-1416.

Harmon, C., & Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review,* 1278-1286.

Himaz, R., & Aturupane, H. (2016). Returns to education in Sri Lanka: a pseudo-panel approach. *Education Economics*, *24*(3), 300-311.

Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467-475.

Kane, T. J., & Rouse, C. E. (1993). Labor market returns to two-and four-year colleges: Is a credit a credit and do degrees matter?, *NBER Working Paper No:4268.*

Kirby, S., & Riley, R. (2008). The external returns to education: UK evidence using repeated cross-sections. *Labour Economics*, *15*(4), 619-630.

Krueger, A. O. (1972). Rates of return to Turkish higher education. *Journal of Human Resources,* 482-499.

Lang, K. (1993). Ability bias, discount rate bias and the return to education. *Mimeo,Boston University*.

Mincer, J. (1974) *Schooling, Experience, and Earnings* .New York, National Bureau of Economic Research.

Mocan, L. (2014). The Impact of Education on Wages: Analysis of an Education Reform in Turkey. In *Koç University-TUSIAD Economic Research Forum Working Papers* (No. 1424). Koc University-TUSIAD Economic Research Forum.

Propper, C., Rees, H., & Green, K. (2001). The demand for private medical insurance in the UK: A cohort analysis. *The Economic Journal*, *111*(471), 180-200.

Psacharopoulos, G. (1985). Returns to education: A further international update and implications. *Journal of Human Resources,* , 583-604.

Psacharopoulos, G. (1994). Returns to investment in education: A global update. *World Development*, *22*(9), 1325-1343.

Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, *12*(2), 111-134.

Tansel, A. (1994). Wage employment, earnings and returns to schooling for men and women in Turkey. *Economics of Education Review, 13*(4), 305-320.

Tansel, A. (2008). Changing returns to education for men and women in a developing country: Turkey, 1994-2005. *ESPE Conference, London,* pp. 18-21.

Tansel, A., & Bodur, F. B. (2012). Wage inequality and returns to education in turkey: A quantile regression analysis. *Review of Development Economics*, *16*(1), 107-121.

Türkiye Istatistik Kurumu (Turkish Statistical Institute, TURKSTAT). *Household Labor Force Survey Data 2009-2014.*

Uusitalo, R. & Conneely, K., 1998. "Estimating Heterogeneous Treatment Effects in the Becker Schooling Model," University of Helsinki, Department of Economics 435.

Verbeek, M., & Nijman, T. (1992). Can cohort data be treated as genuine panel data? *Empirical Economics*, *17*(1), 9-23.

Vural, B. M., & Gülcan, Y. (2008). Impact of education on individual earnings in Turkey. *International Journal of Economic Perspectives*, *2*(3)

Warunsiri, S., & McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development*, *38*(11), 1616-1625.

Appendix:

Table A2: Returns to Education (Years of Education) (Only non-mover Sample)

VARIABLES	OLS	Fixed-Effects (FE) Estimates in Pseudo-Panel Data				
		Single-Birth Year	**Single-Birth Year by NUTS1**	Single-Birth Year by NUTS2	Two-Birth Year by NUTS1	Two-Birth Year by NUTS2
Years of Education	0.068*** (0.000)	0.071*** (0.011)	**0.068*** (0.003)**	0.071*** (0.002)	0.068*** (0.005)	0.074*** (0.003)
Tenure	0.028*** (0.001)	0.036*** (0.01)	**0.038*** (0.004)**	0.037*** (0.003)	0.045*** (0.005)	0.042*** (0.004)
Tenure Sq.	-0.0002 (0.0002)	-0.0002 (0.0004)	**-0.0004*** (0.0001)**	-0.0004*** (0.0001)	-0.0006*** (0.0002)	-0.0006*** (0.0001)
Female	-0.027*** (0.003)	0.045 (0.088)	**-0.071** (0.032)**	-0.075*** (0.022)	-0.052 (0.050)	-0.037 (0.034)
Married	0.120*** (0.00)	0.502*** (0.08)	**0.218*** (0.03)**	0.150*** (0.02)	0.264*** (0.04)	0.197*** (0.03)
Constant	0.054*** (0.006)	-0.519*** (0.097)	**-0.253*** (0.037)**	-0.221*** (0.028)	-0.322*** (0.048)	-0.309*** (0.038)
Observations	193,658	180	**2,160**	4,677	1,080	2,340
R-squared	0.417	0.947	**0.866**	0.814	0.915	0.873

+ Robust Standard Errors in parentheses*** p<0.01, ** p<0.05, * p<0.1. + Each model controls for region and time fixed effects. + Sampling weights, provided in the THLFS, are used for obtaining the sample averages of the variables in the construction of the pseudo-panels. + Square root of the number of observations in each birth-year (birth-region-year) cell is used as weights in the FE estimations.

Table A3: Returns to Education for The Levels of Education(Only Non-mover Sample)

VARIABLES	OLS	Fixed-Effects (FE) Estimates in Pseudo-Panel Data				
		Single-Birth Year	**Single-Birth Year by NUTS1**	Single-Birth Year by NUTS2	Two-Birth Year by NUTS1	Two-Birth Year by NUTS2
Primary School	0.136***	0.170	**0.274*** **	0.269***	0.260***	0.249***
	(0.007)	(0.164)	**(0.059)**	(0.037)	(0.092)	(0.053)
Middle School	0.261***	0.403*	**0.407*** **	0.421***	0.391***	0.386***
	(0.008)	(0.21)	**(0.065)**	(0.042)	(0.096)	(0.059)
General-High	0.412***	0.563***	**0.547*** **	0.561***	0.541***	0.560***
	(0.008)	(0.188)	**(0.066)**	(0.041)	(0.098)	(0.058)
Vocational-High	0.443***	0.336	**0.542*** **	0.587***	0.556***	0.571***
	(0.008)	(0.222)	**(0.066)**	(0.043)	(0.097)	(0.059)
College	0.921***	0.916***	**1.009*** **	1.049***	0.988***	1.079***
	(0.008)	(0.202)	**(0.065)**	(0.042)	(0.097)	(0.059)
Tenure	0.027***	0.044***	**0.041*** **	0.039***	0.048***	0.043***
	(0.0005)	(0.011)	**(0.004)**	(0.003)	(0.005)	(0.003)
Tenure Sq.	-0.0007***	-0.0005	**-0.0005*** **	-0.0005***	-0.0007***	-0.0006***
	(0.00004)	(0.0004)	**(0.0001)**	(0.00005)	(0.0002)	(0.0001)
Female	-0.059***	0.015	**-0.092*** **	-0.099***	-0.073	-0.074**
	(0.003)	(0.090)	**(0.033)**	(0.023)	(0.053)	(0.035)
Married	0.0943***	0.436***	**0.202*** **	0.141***	0.245***	0.189***
	(0.003)	(0.086)	**(0.032)**	(0.023)	(0.042)	(0.031)
Constant	-0.406***	-0.271	**-0.131** **	-0.0947**	-0.189**	-0.142***
	(0.0307)	(0.173)	**(0.0609)**	(0.0389)	(0.0931)	(0.054)
Observations	193,658	180	**2,160**	4,677	1,080	2,340
R-squared	0.441	0.948	**0.868**	0.817	0.916	0.875

+ Robust Standard Errors in parentheses*** p<0.01, ** p<0.05, * p<0.1. + Each model controls for region and time fixed effects. + Sampling weights, provided in the THLFS, are used for obtaining the sample averages of the variables in the construction of the pseudo-panels. + Square root of the number of observations in each birth-year (birth-region-year) cell is used as weights in the FE estimations.