



# The Rate of Return to Education in Iran

Ali Reza Oryoe<sup>1</sup> · Atiyeh Vahidmanesh<sup>1,2</sup>

Received: 4 April 2019 / Accepted: 19 January 2021 / Published online: 16 February 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

There is no doubt that education is the key to development. An important step in evaluating the condition of education in a society is to estimate the rate of return to education in that society. This paper estimates the private rate of return to schooling for men born in 1951–1983 in Iran using a pseudo-panel fixed-effect model. This is the first study that measures unbiasedly the rate of return to schooling in Iran. We construct the pseudo-panel data using six repeated cross sections of Household Expenditures and Income Surveys between 2008 and 2013. Our estimates show that the overall rate of return is around 9% based on the pseudo-panel fixed-effect approach. The estimates reveal a downward bias in ordinary least squares with individual data. This study also evaluates the rate of return to education in rural and urban areas. The rate of return in urban and rural areas is about 7.8 and 10%, respectively.

**Keywords** Pseudo-panel · Return to schooling · Economics of education · Iran

**JEL Classification** I26 · J01 · I21 · I25 · E24

## Introduction

Education plays a significant role in economic growth and development. One of the central topics in the economics of education is the estimation of the rate of private return to education. The rate of return has decreased in North Africa and the Middle East in recent decades, and this reduction has caused youth dissatisfaction and political instability (Arampatzi et al., 2018; Pellicer, 2018). We face a high investment in education in the MENA while we see high youth unemployment. It

---

✉ Ali Reza Oryoe  
oryoe1alireza@gmail.com

Atiyeh Vahidmanesh  
a.vahidmanesh@ut.ac.ir

<sup>1</sup> Faculty of Economics, University of Tehran, North Kargar St., P.O. Box 14155-6445, 1411734115 Tehran, Iran

<sup>2</sup> Sharif Institute for Public Policy, Sharif University of Technology, PO Box: 1459993599, No 7, Golestan Alley, Azadi Avenue, Tehran, Iran

seems that this educational investment has a very small effect on the labor market in this region. Like other countries in the region, Iran is also faced with this paradox in its labor market. There are extensive investments in education after the 1979 revolution while the country is struggling with a high unemployment rate for decades. The estimations of returns to schooling shed light on what is going on in the labor market and education system in Iran. For this purpose, we used Household Expenditure and Income Surveys of Iran to estimate the rates of returns to schooling and investigate the spillover of education in the labor market.

Mincer (1974) provides for the first time a piece of empirical evidence for the labor market effects of education. The Mincerian equation explores the causal relationship between schooling and earning and tries to measure the effect of education on people's well-being. In the original Mincer equation, the logarithm of income is the dependent variable, and the independent variables are the years of schooling and experience. Mincer estimates this equation using the ordinary least squares (OLS) method. In this framework, the coefficient of the years of schooling is the private rate of return to schooling. However, due to the endogeneity problem in the schooling variable, there are concerns about the true value of the return to education (Card, 2001). An individual with a higher level of ability and motivation may invest more in her education and find a better job with higher wages. This is the root of the endogeneity problem in the Mincer equation that leads to an upward bias in the OLS estimates.

To solve this problem, different solutions are proposed in the literature such as the instrumental variables (IV) and fixed-effects models. Card (1995) uses geographical proximity as an instrument to correct for the potential ability bias; Angrist and Krueger (1991) suggest the quarter of birth as an instrument. Many other studies use the instrumental variable method to study the causal relationship between schooling and earning (e.g., Butcher & Case, 1994; Duflo, 2001). In most of these studies, the IV coefficients are larger in magnitude than the OLS coefficients, while the OLS coefficients are expected to be smaller than the IV coefficients due to the ability bias. Some of these studies argue that the ability bias might be quite small and the downward bias caused by the measurement errors in schooling may crowd out the effect of ability bias. However, Card (2001) believes that it is hard to accept this argument. He discusses that the IV estimates might be biased upward in comparison to the OLS estimates. This is because some unobserved differences between the control and treatment groups, which are framed in the IV model implicitly, can cause the instruments to be invalid or weak.

If there is a genuine panel data set, then there is no need to use the instrumental variables method, and hence, there is no worry about the weakness or invalidity of the instruments. A fixed-effect model can remove all of the unobserved individual fixed effects; hence, the estimates should be unbiased. However, such a good panel data are not available in most countries. To solve this problem, Deaton (1985) proposes a new method for constructing a pseudo-panel dataset using repeated cross-sectional data sets. In this approach, some cohorts are constructed using some common time-invariant characteristics (e.g., birth-year), and these cohorts are considered as observations of the pseudo-panel (PP). This method is being progressively used in different fields of study (see, for example, Guarini et al., 2018).

This study estimates the rate of return to schooling in Iran, a country whose economy and educational system are under pressure, using the pseudo-panel approach (Oryoie & Abbasinejad, 2017). Iran has been investing extensively in its education system during the past decades. As a result of these investments, the average years of schooling has increased from just less than 4 years to more than 10 years, and the literacy rate has improved from 42% to more than 85%.<sup>1</sup> However, this huge improvement in educational attainment has not still translated to labor productivity. In the labor market, Iran faces a chronic high unemployment rate, especially among the youth population. The average unemployment rate is around 11.66% throughout 2000–2019.<sup>2</sup> The total labor force participation for ages 15–24 is also quite low. It is around 30.4% throughout 2000–2019, and it is falling during recent years.<sup>3</sup> This fall of labor force participation suggests that people are becoming unemployed and/or hopeless to find a job. Therefore, the level of educational attainment has increased a lot, while the labor force participation rate and the employment rate is very low in Iran. These two phenomena encourage us to find the rate of return to schooling in Iran. The policy implications of this exercise are first to learn better the relationship between education and labor market and second to help the policymakers to design better policies in the education system and labor market. Based on our results, it seems that Iran faces over-investment in education while the labor market and economy itself were not prepared to create high-skills jobs.

There are few studies on Iran's return to schooling. Salehi-Isfahani et al. (2009) estimate the rate of return to schooling using the OLS regression in Iran, Egypt, and Turkey. But as we discussed earlier, there is a notorious endogeneity problem in the OLS method. This paper is the first study that uses a powerful unbiased method for measuring the rate in Iran. We use six Household Expenditures and Income Surveys (HEIS) conducted by the Statistical Center of Iran (SCI) between 2008 and 2013 to construct our pseudo panel data set. Our results suggest that the rate of return is about 9% and 7% based on the PP model and the OLS, respectively. In urban and rural areas, the rate of return is, respectively, 7.8% (resp. 6.4%) and 10% (resp. 7.1%) based on the PP (resp. OLS) approach.

Estimation of the Mincer equation using the PP approach has two benefits: first, the time-invariant unobserved characteristics (such as ability) are removed, and hence we can correct for the potential endogeneity. Second, as Glenn (2005) explains, one of the weaknesses of the cross-sectional data is that it does not control for age-group effects. The people of the same age groups face different conditions at different times. For example, the conditions of the labor market may change over time. Therefore, differences by age may not be completely related to age effects if we do not control for age-group effects. Pseudo panel analysis provides an opportunity to overcome this limitation of the cross-sectional data by construing cohorts based on birth year. By the same logic, we can construct cohorts based on different areas if there is considerable heterogeneity in different areas of a country.

<sup>1</sup> <http://www.barrolee.com/>

<sup>2</sup> <https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS>

<sup>3</sup> <https://data.worldbank.org/indicator/SL.TLF.ACTI.1524.ZS>

Iran is a pretty diverse country in terms of culture, religion, ethnicity, weather, geography, and socioeconomic status. Therefore, people in different areas face pretty different conditions. By constructing the cohorts based on area, we can control the area effects as well.

Methodology is described in “[Methodology](#),” the data and the empirical results are discussed in “[Data](#)” and “[Results](#),” respectively, a comparative analysis is provided in “[Comparative Analysis](#),” suggestions for future studies are presented in “[Suggestions for Future Studies](#),” and the conclusions are stated in “[Conclusion](#).”

## Methodology

This paper uses the Mincer equation to estimate the return to education as follows:

$$\ln W_{it} = c + \beta_1 E_{it} + \beta_2 X_{it} + \beta_3 X_{it}^2 + \theta_i + \varepsilon_{it}; i = 1, \dots, N_t; t = 1, \dots, T \quad (1)$$

where  $\ln W_{it}$ ,  $E_{it}$  and  $X_{it}$  are, respectively, the natural logarithm of wage, the number of years of education, and the number of years of experience for individual  $i$  at time  $t$ .  $T$  is the number of periods,  $\theta_i$  is the unobserved individual effects, and  $N_t$  is the total number of individuals at time  $t$  (Consider that the number of individuals varies from one survey to another in repeated independent cross-sections.). The coefficient  $\beta_1$  measures the private return to education, so the purpose of the paper is to estimate  $\beta_1$  consistently. For the sake of simplicity, assume  $E\{E_{it}\varepsilon_{it}\} = 0, \forall t = 1, \dots, T$ .

Consider that  $\theta_i$  is unobservable, and hence in practice, the error term is  $\theta_i + \varepsilon_{it}$ . If  $\theta_i$  is not correlated with the independent variables, then the composite error term  $\theta_i + \varepsilon_{it}$  is not correlated with the independent variables, and Eq. (1) can be estimated consistently using the OLS after pooling all cross sections. However, the problem is that  $\theta_i$  might be correlated with the years of schooling.<sup>4</sup> Therefore, the composite error term is correlated with  $E_{it}$  and, as a result, the estimated  $\hat{\beta}_1$  is biased.

If there is a genuine panel data,  $\theta_i$  can be eliminated using a fixed-effect model, but such a data set is not often available. If some repeated cross sections are available, this problem can be solved by transforming the cross sections to pseudo-panel data using some cohorts and applying a fixed-effect model (Deaton, 1985; Verbeek & Nijman, 1992; Verbeek & Nijman, 1993; Verbeek, 2008). Each cohort is a set of individuals sharing some common characteristics that do not change over time (from one survey to another) and is observable for all individuals in the sample (e.g., the birth year). As an example, a cohort might be made of all individuals born in the period 1984–1987.

Suppose there are  $T$  periods (or independent cross sections). Construct  $C$  cohorts based on birth year and the place of residence in each period. After constructing the cohorts, the mean of all variables in each cohort over the cohort members should be calculated, and the averages are used as observations. In this way, we make a

<sup>4</sup> The higher levels of motivation/ability might be positively associated with more years of schooling.

pseudo-panel with repeated cohorts over  $T$  periods. This dataset can be used for estimating the following equation:

$$\overline{LnW}_{ct} = k + \beta_1 \overline{E}_{ct} + \beta_2 \overline{X}_{ct} + \beta_3 \overline{X}_{ct}^2 + \overline{\theta}_{ct} + \overline{\varepsilon}_{ct}; c = 1, \dots, C \quad (2)$$

where  $\overline{LnW}_{ct} = \sum_{i \in S_c} LnW_{it}/n_c$ ,  $\overline{X}_{ct} = \sum_{i \in S_c} X_{it}/n_c$ ,  $\overline{\theta}_{ct} = \sum_{i \in S_c} \theta_{it}/n_c$ ,  $\overline{\varepsilon}_{ct} = \sum_{i \in S_c} \varepsilon_{it}/n_c$ ,  $n_c$  and  $S_c$  are, respectively, the number and the set of individuals in cohort  $c$ .

There are three problems with the estimation of Eq. (2): first,  $\overline{\theta}_{ct}$  is unobservable, so it cannot be controlled. Second, it might be correlated with  $\overline{E}_{ct}$ , so the estimated  $\hat{\beta}_1$  might be biased. Third, it varies over time, so it cannot be eliminated using a fixed-effect model. However, Verbeek and Nijman (1992, 1993) show that if the size of each cohort is large enough (beyond 100 observations per cohort), then  $\overline{\theta}_{ct}$  is approximately equal to  $\overline{\theta}_c$ , and the coefficients of Eq. (2) can be estimated consistently using a within estimator (fixed-effect model). The within estimator is given by

$$\hat{\beta} = \left( \sum_{c=1}^C \sum_{t=1}^T (\overline{Z}_{ct} - \overline{Z}_c)(\overline{Z}_{ct} - \overline{Z}_c)' \right)^{-1} \sum_{c=1}^C \sum_{t=1}^T (\overline{Z}_{ct} - \overline{Z}_c)(\overline{LnW}_{ct} - \overline{LnW}_c) \quad (3)$$

where  $\hat{\beta}$  denotes the vector of coefficients and  $\overline{Z}_{ct}$  shows the vector of independent variables for cohort  $c$  at time  $t$ ,  $\overline{Z}_c = \sum_{t=1}^T \overline{Z}_{ct}/T$ , and  $\overline{LnW}_c = \sum_{t=1}^T \overline{LnW}_{ct}/T$ . Consider that if cohort sizes are not equal, then the error terms  $\overline{\varepsilon}_{ct}$  in Eq. (2) will be heteroskedastic. For solving this problem, Deaton (1985) suggests correcting the heteroskedasticity by weighting each cohort with the square root of the cohort size. Since cohort sizes vary considerably in this study, each cohort is weighted by the square root of the cohort size.

## Data

We construct the pseudo panel data using six rounds of the HEIS conducted by the SCI between 2008 and 2013. The HEIS has been collecting annually since 1963. The surveys are nationally representative in both urban and rural areas. These surveys contain information on consumption expenditures, income, household demographics, schooling, employment, and asset ownership. The number of households in the sample increases gradually every year (e.g., there are 5689, 21,950, and 38,252 households, respectively, in 1985, 1997, and 2015). Those surveys that are conducted over the period 2014–2017 are not included in this analysis, because the definition of education has been changed in those surveys by the Statistical Center of Iran. From 2014 onward, we can only observe the level of education of the respondents (e.g., elementary, secondary), and hence, the number of years of education is not computable.

The cross-sectional data sets are employed to build pseudo-panel data based on cohort means for the individuals who are born in 1951–1983. Therefore, this analysis considers only the individuals who are 25–62 years old. The reason that

this age range is chosen is that it leads to some extent to a balanced panel dataset. Women are excluded from the sample to avoid the selection bias problem, because women's labor force participation rate is quite low in Iran. Also, the sample is restricted to the wage income earners, that is, those who work either in public or private sectors.

The cohorts are constructed by the place of residence (see Appendix 1) and the year of birth (4-year and 3-year birth cohorts). Our sample is limited to the cohorts with more than 100 observations<sup>5</sup>, because we want to employ the asymptotic theory explained in “Methodology” that is based on the large number of observations per cohort. In this way, 4-year cohorts generate 222 cohorts with an average cohort size of 364 individuals, and 3-year cohorts generate 246 cohorts with an average cohort size of 309 individuals. We also build two distinct pseudo panel data sets for rural and urban areas based on the same criteria as above for further investigations.

The main variables of interest are annual gross wage earnings, age, and the number of years of schooling. Non-monetary incomes<sup>6</sup> are also added to the gross wage. The wages are deflated by the Consumer Price Index.<sup>7</sup> The number of years of schooling is made from a question that asks about the level of education. This variable varies from zero (for illiterate) to 24 (for Ph.D. degree). The descriptive statistics of the variables are reported in Appendix 5.

The individuals that come from high-income family groups might be more successful in finding high-salary jobs than those from low-income family groups. Therefore, the economic class of the household may be very important, and it is added to the regressions. We define this variable based on the relative income lines. The total consumption expenditures are usually used as a proxy for total income, because consumption expenditures are usually considered as the best measure of living standards (Deaton, 1997). Each cohort is classified into three classes which are poor, middle, and rich, respectively, if the mean of the total expenditures in the cohort is below 75% of the median, between 75 and 125% of the median, and above 125% of the median.

## Results

The results are presented in this section. In “Pooled Cross Section,” the Mincer equation is estimated using a simple OLS method when all cross-sectional data are pooled. The estimated return from this Pooled OLS (POLS) is considered as a benchmark. In “Pseudo Panel Data,” the return is estimated using the PP model, and finally, the return is estimated for urban and rural areas in “Disaggregation by Location.”

<sup>5</sup> Approximately 10% of the cohorts have fewer than 100 observations.

<sup>6</sup> Benefits-in-kind divided by the number of wage income earners in each household.

<sup>7</sup> We use the CPIs reported by the Statistical Center of Iran. The indices are calculated separately for urban/rural areas and different provinces.

**Table 1** Regression of logarithm of annual wage on schooling, OLS model, pooled all cross sections

	(1)	(2)	(3)	(4)	(5)	(6)
Schooling	0.074*** (0.001)	0.058*** (0.004)	0.072*** (0.001)	0.053*** (0.004)	0.073*** (0.001)	0.054*** (0.004)
Schooling <sup>2</sup>		0.001*** (0.000)		0.001*** (0.000)		0.001*** (0.000)
Age	0.113*** (0.005)	0.113*** (0.005)	0.115*** (0.005)	0.114*** (0.005)	0.125*** (0.005)	0.124*** (0.005)
Age <sup>2</sup>	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Area FE	No	No	Yes	Yes	Yes	Yes
Time FE	No	No	No	No	Yes	Yes
ME of schooling	0.074	0.073	0.072	0.071	0.073	0.071
Observations	80,622	80,622	80,622	80,622	80,622	80,622
R <sup>2</sup>	0.198	0.198	0.213	0.214	0.221	0.221

Robust standard errors in parentheses. The sample includes men born in 1951–1983. Data: HEIS 2008–2013

ME marginal effect of schooling, FE fixed effects

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

## Pooled Cross Section

In this section, the rate of return to education is estimated using the POLS during the period 2008–2013. Table 1 reports the estimates for all men born between 1951 and 1983. Unobservable local and time fixed effects are controlled in some of the regressions. Under the odd columns, the coefficients of the simple standard (traditional) Mincer equation are reported. Under the even columns, we include squared years of education into the simple Mincer equation to capture nonlinear behaviors of education. The relationship between the logarithm of wage and the years of schooling is supposed to be linear in the simple Mincer equation; however, many studies show that the wage-schooling relationship might be nonlinear (e.g., Hungerford & Solon, 1987). It seems that the logarithm of wage is often an increasing convex function of schooling. This fact is generally explained by an increase in the relative demand for high-skilled labor. Mincer (1996) considers that higher-educated individuals generally work in high-skilled jobs, and lower-educated individuals are expected to work in low-skill jobs. Therefore, if relative demand for schooling increases more than the relative supply, the marginal return to education increases for higher-educated workers relative to lower-educated workers. Since the marginal return is the slope of the wage-education relationship, an increase in the marginal return leads to a convex wage-schooling relationship. In fact, the odd columns estimate (variables were introduced in “Methodology”):

$$\ln W_{it} = c + \beta_1 E_{it} + \beta_2 X_{it} + \beta_3 X_{it}^2 + \varepsilon_{it}; i = 1, \dots, N_t; t = 1, \dots, T \quad (4)$$

and the even columns estimate

$$\ln W_{it} = c + \beta_1 E_{it} + \lambda_1 E_{it}^2 + \beta_2 X_{it} + \beta_3 X_{it}^2 + \varepsilon_{it}; i = 1, \dots, N_t; t = 1, \dots, T \quad (5)$$

In Eq. (4), the marginal effect of education is measured by  $\beta_1$ ,<sup>8</sup> and in Eq. (5), the marginal effect (ME) is measured by  $\beta_1 + 2\lambda_1 E_{it}$ , because

$$\frac{d(\ln W_{it})}{d(E_{it})} = \frac{\frac{dW_{it}}{W_{it}}}{d(E_{it})} = \beta_1 + 2\lambda_1 E_{it} \quad (6)$$

We replace  $E_{it}$  with its mean over all individuals and years to calculate the mean of the ME. The ME multiplied by 100 shows how much, on average, an individual's wage increases (in percentage) when the number of years of education increases by one unit<sup>9</sup>; in other words, it shows the average return to education. The MEs are reported on the seventh row. As can be seen, the average return to schooling is around 7% during the period 2008–2013. This estimate is considered as a benchmark for future comparisons in the next section.

### Pseudo Panel Data

In this section, the return to education is estimated using a fixed-effect model based on the pseudo-panel data during the period 2006–2013. As we discussed in “Methodology,” the model eliminates all of the unobserved heterogeneities and therefore corrects the endogeneity problem of the Mincer equation. The results are shown in Table 2. Unobservable time effects are controlled in all regressions. A joint  $F$  test is done to check whether the dummies for all years are statistically equal to zero or not. If they are, then no time effects are needed. The null hypothesis is rejected at the 1% level in all regressions, so it is necessary to control the time effects. The first four columns estimate the Mincer equation based on 3-year cohorts. To check the sensitivity of our results to cohort size, we estimate the equation based on 4-year cohorts as well, where the relevant results can be seen under the last four columns. We also estimated the rate of return based on 2-year cohorts and got pretty similar results.

As is shown in Table 2, the return to schooling is about 9% in Iran, and the results are robust in all specifications. Compare them with the OLS estimates in “Pooled Cross Section.” As can be seen, the OLS estimates are smaller than the pseudo-panel fixed-effect (PPFE) estimates by around two percentage points. Interestingly, the estimated return increases after removing the unobserved heterogeneities. The magnitude of the change is significant. Therefore, we face a downward bias if we do not control for the unobserved individual effects. This result is contrary to the ability bias argument that predicts an upward bias in an OLS model. Many other papers have also found a similar pattern of change from an OLS model to an IV model. That is, their estimated return using an IV model is larger in magnitude than their estimated return using an OLS model. We already explained that the issue emerges

<sup>8</sup> More precisely, the marginal effect is  $(e^{\beta_1} - 1)$ .

<sup>9</sup> More precisely, the marginal effect is  $(e^{\beta_1 + 2\lambda_1 E_{it}} - 1)$ .



**Table 2** Regression of logarithm of annual wage on schooling, fixed-effects model, and pseudo panel data

	Three-year cohorts			Four-year cohorts				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Schooling	0.090*** (0.009)	0.069 (0.051)	0.088*** (0.011)	0.071 (0.049)	0.088*** (0.015)	0.112** (0.033)	0.090*** (0.015)	0.117*** (0.026)
Schooling <sup>2</sup>		0.001 (0.003)		0.001 (0.003)		-0.002 (0.002)		-0.002 (0.001)
Age	0.005 (0.021)	0.004 (0.021)	-0.002 (0.025)	-0.003 (0.024)	0.100*** (0.012)	0.102*** (0.012)	0.102*** (0.013)	0.105*** (0.013)
Age <sup>2</sup>	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)	-0.001*** (0.000)
Class of household								
Middle			-0.029* (0.012)	-0.029* (0.012)			-0.005 (0.025)	-0.006 (0.025)
Rich			-0.019 (0.016)	-0.019 (0.016)			-0.015 (0.019)	-0.017 (0.018)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Area FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
ME of schooling	0.090	0.090	0.088	0.089	0.088	0.088	0.090	0.088
Individual obs. per cohort								
Min	121	121	121	121	109	109	109	109
Mean	339	339	339	339	392	392	392	392
Max	766	766	766	766	932	932	932	932
Observations	246	246	246	246	222	222	222	222
R <sup>2</sup>	0.535	0.536	0.540	0.540	0.510	0.511	0.511	0.512

Robust standard errors in parentheses. The sample includes men born in 1951–1983. All regressions include time effects. The panels are balanced. Data: HEIS 2008–2013

ME marginal effect of schooling FE fixed effects

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

due to the measurement error problem and/or the weak/invalid instrument problem (Angrist & Keueger, 1991; Card, 1995; Butcher & Case, 1994). An advantage of the PPFE model is that it does not face the weak/invalid instrument problem. In addition, Deaton (1985) explains that when the cohort size is large enough, the measurement errors can be ignored. Since the cohort sizes are sufficiently large in this research, we are not worried about the measurement error too. As far as there is a good variation inside the cohorts (cohort sizes are large), the potential coefficient bias is minimized (Verbeek & Nijman, 1992). Therefore, we are confident that our estimates are unbiased.

In addition to estimating the return to education, we can also estimate the effect of the socioeconomic status of the individuals on their wage based on columns 3, 4, 7, and 8. For this purpose, we use columns 3 and 4 since they are estimated based on 3-year cohorts in which cohort sizes are larger than the other columns. Suppose  $\beta_M$  and  $\beta_R$  denote respectively the estimated coefficients of the middle and rich cohorts, and  $\bar{W}_P$ ,  $\bar{W}_M$ , and  $\bar{W}_R$  stand, respectively, for the mean of wage in the poor, middle, and rich cohorts, then it can be easily shown that

$$100 \times \frac{\bar{W}_M - \bar{W}_P}{\bar{W}_P} = 100 \times (e^{\beta_M} - 1) \quad (7)$$

and

$$100 \times \frac{\bar{W}_R - \bar{W}_M}{\bar{W}_M} = 100 \times (e^{\beta_R - \beta_M} - 1) \quad (8)$$

Based on Eqs. (7) and (8) and 3-year cohorts, if an individual is transferred from a poor family to a middle-class family, her wage rises by about 6%, and if she is transferred to a rich family, her wage increases by about 8%.

In summary, this section concludes that the rate of return to schooling for males aged 25–62 years old is around 9% and this result is more reliable than the simple OLS; because the pseudo panel fixed-effects model eliminates all of the unobserved time-invariant characteristics of the individuals, hence, we are confident at least that this approach corrects for the bias caused by the omitted variable problem.

### Disaggregation by Location

This section estimates the rate of return to schooling in urban and rural areas to check the robustness of the results that we got in the previous section and to see the differences between the urban and rural areas. The PPFE (based on 4-year cohorts) and POLS estimates are presented in Table 3. Unobservable time effects are controlled in all regressions. A joint  $F$  test is done to check whether the dummies for all years are equal to zero or not. The null hypothesis is rejected at the 1% level in all regressions, so we control the time effects in all regressions.

As can be seen, the estimated return in urban areas (resp. rural areas) is 7.8% (resp. 10%) based on the PPFE estimators, while the return is about 6.4% (resp.

**Table 3** Regression of logarithm of annual wage on schooling, OLS, and fixed-effect model, urban and rural areas

	Urban			Rural		
	(1) POLS	(2) Pseudo panel	(3) Pseudo panel	(4) POLS	(5) Pseudo panel	(6) Pseudo panel
Schooling	0.064*** (0.001)	0.078*** (0.011)	0.077*** (0.012)	0.071*** (0.001)	0.100*** (0.010)	0.100*** (0.011)
Age	0.135*** (0.006)	0.050** (0.018)	0.046* (0.018)	0.094*** (0.006)	0.167** (0.045)	0.162** (0.042)
Age <sup>2</sup>	−0.001*** (0.000)	−0.001** (0.000)	−0.001** (0.000)	−0.001*** (0.000)	−0.001*** (0.000)	−0.001*** (0.000)
Class of household						
Middle			0.015 (0.015)			0.046 (0.038)
Rich			0.024 (0.017)			0.043 (0.046)
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Area FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual obs. per cohort						
Min		103	103		100	100
Mean		244	244		233	233
Max		491	491		452	452
Observations	44,854	162	162	35,768	144	144
R <sup>2</sup>	0.196	0.585	0.586	0.136	0.497	0.501

Robust standard errors in parentheses. The sample includes men born in 1951–1983. The pseudo panel data are constructed based on 4-year cohort means. All regressions include time effects. The panels are balanced. Data: HEIS 2008–2013

*FE* fixed effects

\* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

7.1%) based on the OLS estimators. We again see a gap between the OLS and the PPF estimates. The estimated return in rural areas is slightly larger, while one may expect a higher rate of return in urban areas because of more and better educational and job opportunities in urban areas in comparison with rural areas. This small gap might be due to this fact that the rate of return to education is decreasing at higher levels of education (the number of years of education is around 6.5 years in rural areas, and this number is around 9.8 years in urban areas), or it might be because measurement error problem.

In addition to urban/rural criteria, Warunsiri and McNown (2010) disaggregate their sample based on gender and marital status as well. However, we could not disaggregate our sample due to some data restrictions. As we explained in “Data,” women’s labor force participation rate is quite low in Iran which may cause the selection bias problem. The female labor force participation rate was around 17.3% in 2013 and 21.2% in 2008 for women between 20 and 55 years of age (Salehi-Isfahani et al., 2009). In addition, if

cohorts are constructed based on marital status, there will be many cohorts with very small sizes due to heterogeneity in the marital status of the individuals. In our sample (males born in 1951–1983), about 75% of the individuals are married, about 0.3 and 0.5% are widowed and divorced, and 24% are never married. Thus, the distribution of the marital status is very heterogeneous, and we face many small cohort sizes if we construct cohorts based on marital status.

## Comparative Analysis

This section compares the rate of return to education in Iran with other countries (for more comparisons, see, Psacharopoulos & Patrinos, 2004). Table 4 summarizes the comparisons. Most studies use either OLS or IV methods to estimate the private return to schooling. As can be seen, there are only two other studies that employ the pseudo-panel approach in addition to this study. The IV estimates are generally larger in magnitude than the OLS estimates. Our results indicate that the return is around 7% (9%) based on the POLS (the PPFE) model in Iran between 2008 and 2013.

Montenegro and Patrinos (2013) use about 545 harmonized household surveys from 98 countries to estimate the return to schooling in a comparable way around the world between 1970 and 2011. They estimate the traditional Mincer equation using the OLS method, which may suffer from the endogeneity problem. They show that the rate of return is around 10.4% on average, and the rate is on average around 10% in high-income countries. From a regional perspective, they show that the rate is around 5.6% in the Middle East and North Africa (MENA) between 2000 and 2011.

Salehi et al. (2009) estimate the rate of return to schooling in Iran in 2006 around 7.6% using the OLS method. Our estimates are 8.8% and 9% using, respectively, 3-year and 4-year cohorts suggesting a downward bias in Salehi's study. In comparison with other countries, as can be seen in Table 4, most of the studies use either OLS or IV methods to estimate the rate of return to schooling. There are only two papers that use the PP approach. Therefore, we compare our results only with these two studies. The results suggest that Iran's rate of return to schooling is larger than that of Sri Lanka and smaller than that of Thailand. The rate of return to education is determined by many economic, social, and political factors. To explore the causal relationship between the factors and the return to education, we need to employ the data sets of the countries and control the economic, social, and political factors in our regressions. However, we do not have access to such data at this point of time. Therefore, we cannot say why the rate of return in Iran is different from other countries. It is left for future studies.

## Suggestions for Future Studies

We have two suggestions for further research on this topic. First, let us denote the total number of individuals by  $N$ , the number of periods by  $T$ , the number of cohorts by  $C$ , and the number of observations in cohort  $c$  by  $n_c$ . We explained that if the

**Table 4** Return to education, cross-country comparison

Author	Data and sample	Schooling coefficients		
		OLS	IV	PP
Angrist and Krueger (1991)	1970 and 1980 Census Data, 1920–29 cohort in 1970, USA	0.070 (0.000)	0.101 (0.033)	
Kane and Rouse (1993)	NLS Class of 1972, Women Sample, USA	0.08 (0.005)	0.091 (0.033)	
Card (1995)	NLS, Young Men, 1966 cohort, USA	0.073 (0.004)	0.132 (0.049)	
Duflo (2001)	1995 Intercensal Survey, Men born 1950–72, Indonesia	0.078 (0.0001)	0.064 (0.025)	
Salehi-Isfahani and Assaad (2009)	2006 household Survey, Urban Men aged 20–54, Iran	0.076 (0.002)		
Butcher and Case (1994)	1985 Panel Study of Income Dynamics (PSID), White Women, 1920–1965 cohorts, USA	0.091 (0.007)	0.184 (0.113)	
Himaz and Aturupane (2016)	Labor Force Survey-1997–2008, 1953–74 Cohorts, Men, Sri Lanka	0.09 (0.00)		0.05 (0.02)
Waransiri and Menown (2010)	National Labor Force Surveys, 1946–67 cohorts, Thailand	0.115 (0.0002)	0.141 (0.103)	0.151 (0.01)
Harmon and Walker (1995)	1978–1986 Family Expenditure Survey, Men sample, Britain	0.061 (0.001)	0.153 (0.015)	
Psacharopoulos and Patrinos (2004)	Household Surveys and Labor Force Surveys- Middle East, North Africa, and Europe	0.071		
Psacharopoulos and Patrinos (2004)	Household Surveys and Labor Force Surveys-Latin America	0.12		
Lemieux and Card (2001)	1971 and 1981 Census, Only French-speaking men in Quebec and English-speaking men in Ontario, Canada	0.070 (0.002)	0.164 (0.053)	

Standard errors are in parenthesis

cohorts are constructed such that the number of observations in every cohort is above 100 ( $n_c \rightarrow \infty$ ), then we get unbiased estimates. In addition to this method, the cohorts can also be constructed in three other different ways (Verbeek, 2008):

1. Both  $N$  and  $C$  go to infinity, and  $n_c$  is remained constant.
2.  $T$  goes to infinity, with  $N$  and  $C$  kept constant.
3. Both  $T$  and  $n_c$  goes to infinity (McKenzie, 2004).

It might be interesting to see how the results change in each approach mentioned above. But the problem is that there are no specific criteria for constructing cohorts based on these approaches. For example, we still do not know how large  $N$  must be when we say  $N$  goes to infinity.

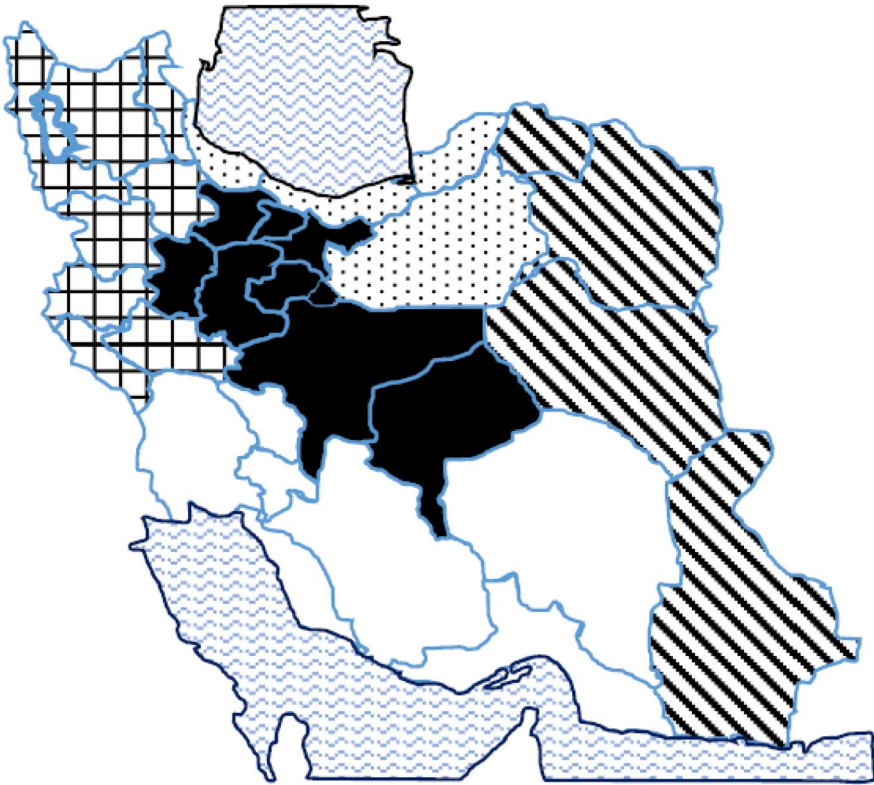
Second, we discussed that the best method for estimating the rate of return to education, in the absence of a genuine panel data set, is to use the pseudo-panel method. If there is a genuine panel data set, we can get better results. There are two panel data sets in Iran that are not publicly accessible. First, there is a comprehensive panel data set about income, job, family background, and education in the Ministry of Cooperatives, Labor, and Social Welfare. Second, there are two comprehensive panel data sets about educational attainment of Iranians in the ministries of education and higher education. If researchers are allowed to get access to these data sets, we can have much more precise estimates using panel data methods.

## Conclusion

This study was one of the few studies in the literature of the economics of education in Iran. It estimated the private return to schooling using a pseudo-panel fixed-effect model. The point of the model is that it removes all of the unobserved heterogeneities, and the estimated coefficients are unbiased if the cohort sizes are large enough. Since our cohort sizes were sufficiently large, we expected to derive unbiased estimates for the coefficients of the years of schooling. Our results showed that the rate of return was around 9% in Iran. That is, it is expected that each individual's wage increases on average by about 9% for each additional year of education. The rate of return in urban and rural areas was around 7.8 and 10%, respectively.

This study provided the correct estimates of the return to schooling for policymakers in Iran. We now know that each additional year of schooling increases an individual's earning by 9% on average. This is not a large number given the vast investment in Iran's educational system within the past four decades. It seems that education does not transform into human capital and job skills well. Most of the investment in education in Iran is in terms of providing schools for kids in remote areas. Iran needs more investment in the quality of education to improve human capital accumulation which later on shows itself in the labor market. Besides, we think that the Education Ministry needs to boost the job skills of students during primary/secondary education. This training can improve labor productivity and hence increase the rates of return to schooling.

## Appendix 1



**Fig. 1** Division of Iran into five different areas

The country is divided into five areas, which are center, south, north, east, and west, based on socio-economic characteristics and migration between different provinces as follow:

- Tehran, Qom, Qazvin, Alborz, Hamedan, Markazi, Yazd, and Esfahan.
- Khuzestan, Kerman, Fars, Bushehr, Chaharmahal, Hormozgan, and Kohkiluyeh.
- Kurdistan, Kermanshah, Ilam, Lorestan, East and West Azarbaijan, Ardebil, and Zanjan.
- Gilan, Mazandaran, Golestan, and Semnan.
- North and South Khorasan, Razavi Khorasan, and Sistan.

These areas are shown graphically in Fig. 1. Area 1 is shown by black color, area 2 is shown by white color, area 3 is shown by squares, area 4 is shown by dots, and area 5 is shown by backslash solid lines.

## Appendix 2

**Table 5** Summary statistics for wage earners

Year	Statistics	Wage	Age	Schooling years	Urban
2008	<i>N</i>	18,033	18,033	15,899	18,033
	Mean	1.49E+08	37.20	9.50	0.73
	sd	1.23E+08	8.62	4.33	0.44
2009	<i>N</i>	15,914	15,914	13,812	15,914
	Mean	1.56E+08	38.05	9.47	0.75
	sd	1.27E+08	8.55	4.32	0.43
2010	<i>N</i>	15,874	15,874	13,826	15,874
	Mean	1.55E+08	39.14	9.37	0.75
	sd	2.02E+08	8.52	4.36	0.43
2011	<i>N</i>	14,797	14,797	12,924	14,797
	Mean	1.47E+08	40.48	9.51	0.75
	sd	1.07E+08	8.47	4.40	0.43
2012	<i>N</i>	13,899	13,899	12,160	13,899
	Mean	1.39E+08	41.39	9.63	0.77
	sd	1.11E+08	8.39	4.47	0.42
2013	<i>N</i>	13,568	13,568	12,000	13,568
	Mean	1.36E+08	40.80	9.85	0.77
	sd	1.88E+08	8.04	4.44	0.42

The sample includes only the men born 1951–1983  
*sd* standard deviation

**Acknowledgements** The authors acknowledge and appreciate the useful comments received from the editor of the journal and the reviewers.

### Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Disclaimer** Any errors and omissions are wholly our responsibility.

### References

- Angrist, J. D., & Keueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.
- Arampatzi, E., Burger, M., Ianchovichina, E., Röhrich, T., & Veenhoven, R. (2018). Unhappy development: Dissatisfaction with life on the eve of the Arab Spring. *Review of Income and Wealth*, *64*, S80–S113.
- Butcher, K. F., & Case, A. (1994). The effect of sibling sex composition on women's education and earnings. *The Quarterly Journal of Economics*, *109*(3), 531–563.



- Card, D. (1995). Using geographical variation in college proximity to estimate the return to schooling. In L. N. Christofides, E. K. Grant, & R. Swidinsky (Eds.), *Aspects of labour market behaviour: Essays in honour of John Vanderkamp* (pp. 201–222). Toronto: University of Toronto Press.
- Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5), 1127–1160.
- Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, 30, 109–126.
- Deaton, A. (1997). *The analysis of household surveys: A microeconomic approach to development policy*. Washington, DC: Johns Hopkins University Press for the World Bank.
- 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.
- Glenn, N. (2005). *Cohort analysis* (Vol. 5). Newbury Park, California: Sage Publications.
- Guarini, G., Laureti, T., & Garofalo, G. (2018). Territorial and individual educational inequality: A capability approach analysis for Italy. *Economic Modelling*, 71, 247–262.
- Himaz, R., & Aturupane, H. (2016). Returns to education in Sri Lanka: A pseudo-panel approach. *Education Economics*, 24(3), 300–311.
- Hungerford, T., & Solon, G. (1987). Sheepskin effects in the returns to education. *The Review of Economics and Statistics*, 69(1), 175–177.
- 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?* (No. w4268). National Bureau of Economic Research.
- Lemieux, T., & Card, D. (2001). Education, earnings, and the ‘Canadian GI Bill.’ *Canadian Journal of Economics/Revue canadienne d’économique*, 34(2), 313–344.
- McKenzie, D. J. (2004). Asymptotic theory for heterogeneous dynamic pseudo-panels. *Journal of Econometrics*, 120, 235–262.
- Mincer, Jacob A., (1974), Schooling, experience, and earnings. National Bureau of Economic Research.
- Mincer, J. (1996). Economic development, growth of human capital, and the dynamics of the wage structure. *Journal of Economic Growth*, 1(1), 29–48.
- Montenegro, C. E., & Patrinos, H. A. (2013). *Returns to schooling around the world. Background paper for the World Development Report*. Washington, DC: World Bank.
- Oryoe, A., & Abbasnejad, H. (2017). The Rise and Fall of Top Incomes in Iran 1985–2015. *Iranian Economic Review*, 21(4), 941–968.
- Pellicer, M. (2018). The evolution of returns to education in the Middle East and North Africa: Evidence from comparable education policy changes in Tunisia. *Economics of Education Review*, 62, 183–191.
- Psacharopoulos, G., & Patrinos, H. A. (2004). Returns to investment in education: A further update. *Education Economics*, 12(2), 111–134.
- Salehi-Isfahani, D., Tunali, I., & Assaad, R. (2009). A comparative study of returns to education of urban men in Egypt, Iran, and Turkey. *Middle East Development Journal*, 1(02), 145–187.
- Verbeek, M. (2008). Pseudo-panels and repeated cross-sections. In *the econometrics of panel data* (pp. 369–383). Springer, Berlin, Heidelberg.
- Verbeek, M., & Nijman, T. (1992). Can cohort data be treated as genuine panel data? *Empirical Economics*, 17, 9–23.
- Verbeek, M., & Nijman, T. (1993). Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics*, 59, 125–136.
- Warunsiri, S., & McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development*, 38(11), 1616–1625.
- Harmon, C., & Walker, I. (1995). Estimates of the economic return to schooling for the United Kingdom. *The American Economic Review*, 85(5), 1278–1286.