UNITED NATIONS UNIVERSITY
UNU-WIDER

WIDER Working Paper 2020/5

Returns to education in self-employment in India

A comparison across different selection models

Indrajit Bairagya*

January 2020

United Nations University World Institute for Development Economics Research
wider.unu.edu

Abstract: This study focuses on estimating the returns to education in non-farm self-employed businesses in the Indian context, using nationwide individual- and household-level data provided by the India Human Development Survey for the year 2011/12. Given that different studies have used different types of regression models for estimating returns to education in self-employment for different economies, the present paper further examines the sensitivity of the estimation of returns to education to the choice of different types of selection model—namely, ordinary least squares, instrumental variable technique, Heckman selection model, and double-selection models, based on the familiar Mincerian earnings equation. The results indicate that, although the trend in the *movement* of the rate of returns across different educational levels is similar for the different selection models, the estimated *value* of the rate of returns with respect to the different levels of education is very sensitive to the choice of model. This points to the need to exercise a fair degree of care in the choice of an appropriate model(s) for estimating the rate of returns to education and for drawing policy implications from it.

Key words: double-selection model, endogeneity, returns to education, self-employment

Acknowledgements: The initial version of the paper was presented at the UNU-WIDER Development Conference on 'Transforming economies – for better jobs' organized in partnership with UNESCAP on 11–13 September 2019 in Bangkok, Thailand. The author thanks the participants of the above conference for their comments and suggestions on the paper at the time of the conference.

* Centre for Human Resource Development, Institute for Social and Economic Change, Nagarabhavi, Bengaluru, India: indrajitisec@gmail.com

This study has been prepared within the UNU-WIDER project Transforming economies – for better jobs.

Copyright © UNU-WIDER 2020

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-762-0

https://doi.org/10.35188/UNU-WIDER/2020/762-0

Typescript prepared by Joseph Laredo.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

Human capital theory posits the importance of education in increasing the productivity of an individual and thereby earnings, as it helps them acquire the necessary knowledge and skills (Becker 1964). Studies have established a relationship between education and earnings and identified that better education results in higher earnings (Card 1999). Given the importance of education, an estimation of returns to education adds to our understanding of various development and policy issues, such as the determinants of economic growth and income distribution (Colclough et al. 2010; Psacharopoulos 1985; Söderbom et al. 2006).

Several empirical studies also show that basic education increases the productivity and wages of workers (Bigsten 1984; Fan et al. 2002; Lanjouw and Shariff 2004). Others argue that lack of education (human capital accumulation through skill formation) often acts as a barrier to the movement of workers from one sector to the other. Yet some studies have observed that the absence of education among a large number of individuals in rural India has stunted the growth of the rural non-farm sector (Mukherjee and Zhang 2005) and that lack of education results either in the stagnation of agricultural labour or its movement to casual work occupations in the non-farm sector, rather than to salaried employment with higher wages and benefits (Planning Commission 2000). Thus, less educated households tend to rely more on low-paying farm wage employment or the very low-productive non-farm sector, as evidenced by Corral and Reardon (2001) for Nicaragua, Hossain (2004) for Bangladesh, and Lanjouw and Shariff (2002) for India.

Education's payoff also differs across different types of employment. For instance, Taylor and Yunez-Naude (2000) showed that it was highest for rural non-farm wage labour, but relatively low for cash cropping and rural self-employment, implying that there were significant interactive differences in skills between these two types of employment. The literature has further suggested that in the context of developing nations, returns to education for the self-employed are lower than for wage employees. A study using models both with and without corrections for sample selection conducted by Rankin et al. (2010) for Ghana and Tanzania showed convex returns to education for the self-employed, while another study, by Sandefur et al. (2006), for Ghana, Tanzania, and Ethiopia observed convex returns to education for both the self-employed and wage employed in Ghana and Tanzania, but not in Ethiopia. Further, Kavuma et al. (2015), using the Mincerian framework with a pooled regression model based on the data for Uganda's labour market, observed a concave relationship between earnings and education for self-employment and a convex relationship for wage employment. Kavuma et al. (2015) found that the convex and concave education qualification earnings curve for wage employees and the self-employed, respectively, was primarily reflective of the difference in the demand for skills in the formal and informal sectors, demand being lower in the informal sector.

Teal (2001), using a model with correction for sample selection bias, noted a similar finding for Ghana, returns to education being far higher for wage earners than for the non-agricultural self-employed. Nyakundi et al. (2017), in determining the returns to education at various levels of education for the self-employed in the motor spare parts industry in Kenya using a descriptive design and ordinary least squares (OLS) estimation, observed a significant and positive association between the returns to education and increasing levels of education. A study conducted for Ghana, using the Ghana Living Standards Survey of 1998/99 based on an OLS model also with corrections for endogeneity and selectivity biases separately, found that education had raised earnings only modestly in respect of wage employment but not the earnings of the majority of workers in Ghana, as the returns to education for self-employment were significantly lower (Kingdon and Söderbom 2007a).

The returns to education for households engaged in self-employment activities in rural Peru, considering individuals' level of education as well as the average educational level of the households, were estimated using an OLS model by Laszlo (2005). The results reveal that an increase in individuals' education significantly reduces the likelihood of their participation in farm wage employment, whereas it has no significant impact for farm and non-farm self-employment. Further, average formal schooling of household members engaged in non-farm self-employment activities has a negative impact on the earnings from farm wage employment, indicating a shift from farm wage employment towards non-farm self-employment activities with the increase of household members' education (Laszlo 2005). The education earnings curve for Pakistan, which was estimated using an OLS model and also with corrections for endogeneity and selectivity biases separately, was found convex in nature for both the self-employed and wage employees (Kingdon and Söderbom 2007b).

A few studies have produced similar results for advanced economies as for the developing economies discussed above. A comparison of the returns to education for self-employed with respect to Portugal and Spain (both characterized by a high rate of self-employment), using an Efficient Generalized Instrumental Variables (EGIV) technique, revealed that for Spain, the returns were higher for wage earners than for the self-employed, whilst for Portugal, the returns were higher for wage earners with higher education, but also higher for the self-employed with a secondary-level education (Garcia-Mainar and Montuenga-Gomez 2005). Further, Williams (2002), using OLS log earning equations with adjustment for sample selection bias and comparing wage-employment and self-employment, found that additional schooling had a lesser effect on the earnings of the self-employed than on those of the wage employed. Another study, by Iversen et al. (2010), instrumenting education within the selection model, found that the returns to different levels of education were low and diverse for self-employment as compared with wage employment in the Danish labour market.

In this context it is important to note that India accounts for the highest percentage of non-agricultural informal employment among countries in South and Southeast Asia (ILO 2011). In fact, about 93 per cent of the total workforce is informally employed with no access to any social security benefits (NCEUS 2008). Moreover, among informal sector workers, the proportion of self-employed is about 57 per cent in rural areas and about 58 per cent in urban areas (NSSO 2014). Although a large number of studies have focused on estimating the returns to education in respect of wage employment (both regular and casual) in the Indian context (for example, Duraisamy 2002; Singhari and Madheswaran 2016; Tilak 1987; Vasudeva Dutta 2006), very few have focused on estimating the returns to education for non-farm self-employment at the national level. Considering that a significant proportion of the workforce is self-employed in India, focusing solely on wage earners results in only a partial picture of the economic benefits of education, especially when education is considered an essential determinant of self-employment. Moreover, existing studies have used only one-stage selection, considering sample selection bias (mostly Heckman selection bias correction or multinomial selection), based on Lee (1983) corrections. This study, therefore, focuses on an analysis of the returns to education for non-farm self-employment businesses in India, using nationwide individual- and household-level data, i.e. India Human Development Survey (IHDS) data for the Indian economy for the year 2011/12.

In addition, given the fact that different studies have used different types of regression models, such as OLS, the Heckman selection model, a multinomial selection model, and an instrumental variables model, for estimating the returns to education for self-employment for different economies, the paper has been extended to examining the sensitivity of the estimation of returns to education to the type of selection model used. In order to accomplish the objective of the study, we have used OLS, instrumental variable (IV) techniques, and the Heckman selection

model. Double-selection models have also been used, with one selection for participation in the labour force and another for choice of self-employment.

The rest of the paper is organized as follows. Section 2 describes the details of the different regression techniques used in the study. A description of the database is given in Section 3, and descriptive results related to education and self-employment are presented in Section 4. Sections 5 and 6 present the results of different regression models and the estimated rates of return to education for self-employment, respectively. Section 7 concludes.

2 Methodology

We begin with an estimation of the impact of educational levels on the earnings of self-employment businesses, based on an earnings equation similar to the Mincerian wage equation (Mincer 1970).

$$\text{LogYi} = \text{Xi}\beta\text{i} + \text{ui} \quad (1)$$

where Y_i represents the earnings of the i^{th} self-employed individual, dependent on the vector of covariates X_i, such as age, educational level, gender, caste, religion, region, and business- and social capital-related factors.

However, a serious econometric issue arising here is the problem of endogeneity because a non-zero correlation between education and earnings may lead to a serious estimation problem if we ignore the potential endogeneity of our key independent variable 'education' with the dependent variable 'earnings from self-employment businesses'. The reason for considering 'education' as an endogenous variable is that factors related to unobserved individual ability may determine both 'education' and 'earnings', as indicated by several existing studies (Card 2001; Griliches 1977; Iversen et al. 2011; Kingdon and Söderbom 2007a; Kolstad and Wiig 2015; Schwiebert 2012; and many others). In order to deal with the above endogeneity problem, we have adopted a two-stage least squares-based instrumental variable technique (IV 2SLS) with 'father's education' as an instrument for 'education of the self-employed individuals'.

The problem of endogeneity due to unobserved ability apart, it is well established in the literature that an OLS-based estimation of the earnings equation suffers from a selection bias due to non-randomness associated with the sample selection. Therefore, a model with a joint determination of the labour force participation and earnings has been widely used in the existing literature for measuring the returns to education, while controlling for other variables. Labour force participation selection bias correction, using the Heckman (1976, 1979) procedure, has become increasingly popular among researchers, and a wide body of research has developed. This procedure involves two stages in correcting the selection bias. At the first stage, it specifies employment selection with a probit model:

$$\text{Prob}\,(\text{P} = 1\,|\text{Z}) = \text{Z}_\text{i}\gamma + \text{u}_\text{i} \quad (2a)$$

where P represents the participation in employment; Z and γ represent the vector of explanatory variables that determine the employment participation and the vector of parameters, respectively. Using equation (2a), one can estimate the predicted probability of individuals engaging in employment.

The second stage involves the correction of the sample selection bias by way of including the above predicted probabilities as an added explanatory variable (Inverse Mills Ratio). The earnings equation can now be written as:

$$Y^* = X\beta + u \tag{2b}$$

where Y represents the earnings of employed individuals and this variable cannot be observed for those who are not employed. The conditional expected earnings of employed individuals with the error terms of equations (2a) and (2b) following a joint normal distribution can be written as:

$$E[Y \mid X, P=1] = X\beta + \rho\sigma_u \lambda(Z_i\gamma) \tag{2c}$$

where ϱ represents the correlation between unobserved factors that determines the employment participation and unobserved factors related to earnings; σu represents the standard deviation of error term and λ the Inverse Mills Ratio (IMR), computed on the basis of the vector of explanatory variables that determines the employment participation (Z) and the vector of parameters (γ) from equation (2a). Although the above model deals with different types of employment together, we are specifically interested in capturing only self-employment. Therefore, the issue with self-selection that arises here is: Should we make corrections for self-employment only or for entire employment participation? One other option is to divide the selection into two stages—one for choice of employment in general and the other for choice of self-employment—before estimating the earnings equation.

However, whilst selection bias correction for one selection equation is common in the literature, it has not gained much appeal for more than one stage. The objective of our paper is to identity the determinants of earnings from self-employment businesses with a special emphasis on education. In our case, we certainly have a selection bias for participation in the labour force, and we have another selection bias for the choice of self-employment, given the other options of casual and regular wage employment for those participating in the labour force. In fact, we might obtain a biased estimation if we completely ignored the issue of second selection (Co et al. 1999), i.e. selection of only the types of self-employment. A double selection model suggested by Tunali (1986) can be used in this case. Thus, in this paper, the regression equation of the determinants involves double sample selections. The first stage of sample selection captures participation in the labour force, while the second stage of selection includes the choice of self-employment types.

$$P^* = Z'_i\gamma + u_i \tag{3a}$$

$$S^* = T'_i\delta + v_i \tag{3b}$$

Here, P^* and S^* are the latent variables. P and S represent the selection for employment participation and the choice of self-employment, respectively. Z and T are the covariates that determine the selection for employment participation and the choice of self-employment, respectively. Further, u_i and v_i are the error terms for employment participation and the choice of self-employment, respectively.

In this context, another important issue arises regarding the independency of the two selections, i.e. whether the choice for self-employment is independent of the choice for joining the labour market or whether these two are interdependent. To stay away from this issue of independency, we have estimated the earnings equation considering both independency and interdependency of the two selection decisions, using two separate models. In the first model, following Heitmueller

(2004), we have first estimated two correction terms (IMR) based on two separate probit models (one for participation in the labour force and the other for choice of self-employment) and then, using these correction terms, estimated the earnings equation. In the second model, considering the interdependency of the two selection decisions and following Ham (1982) and Tunali (1986), we have estimated a correction term (IMR) based on a bivariate probit estimation of the two selection equations and then, including the correction terms in the Mincerian earnings equation, we have identified the factors that determine the earnings of self-employment businesses for India.

3 Database for the study

In order to accomplish our objective, we used individual- and household-level data from the nationwide India Human Development Survey (IHDS) for the Indian economy for the year 2011/12, as provided by Desai and Vanneman (2011–2012). Although National Sample Survey Office (NSSO) data for employment and unemployment in India have been extensively used in research for examining labour market issues, we have used IHDS data mainly because they provide information on the earnings of self-employed people, which is not present in the NSSO data. It has been well argued in the literature that self-employment is a household-based business and that earnings from it constitute household earnings. Often, it is difficult to identify the actual decision maker when it comes to small self-employment businesses. Interestingly, IHDS data include an answer to the question: 'Who is the decision maker of business activities from among the household members?'. Further, the data provide detailed accounts of gross receipts as well as of expenditure incurred on different inputs such as raw materials, labour, electricity, water, transport, loan repayments, and taxes. In this paper, we have considered the difference between gross receipts and payments as business earnings, which is quite similar to the measure of 'net profit' from self-employment businesses used by Hamilton (2000) and Iversen et al. (2011). In addition to earnings, it provides information on a number of variables related to the socio-economic features of households and individuals.

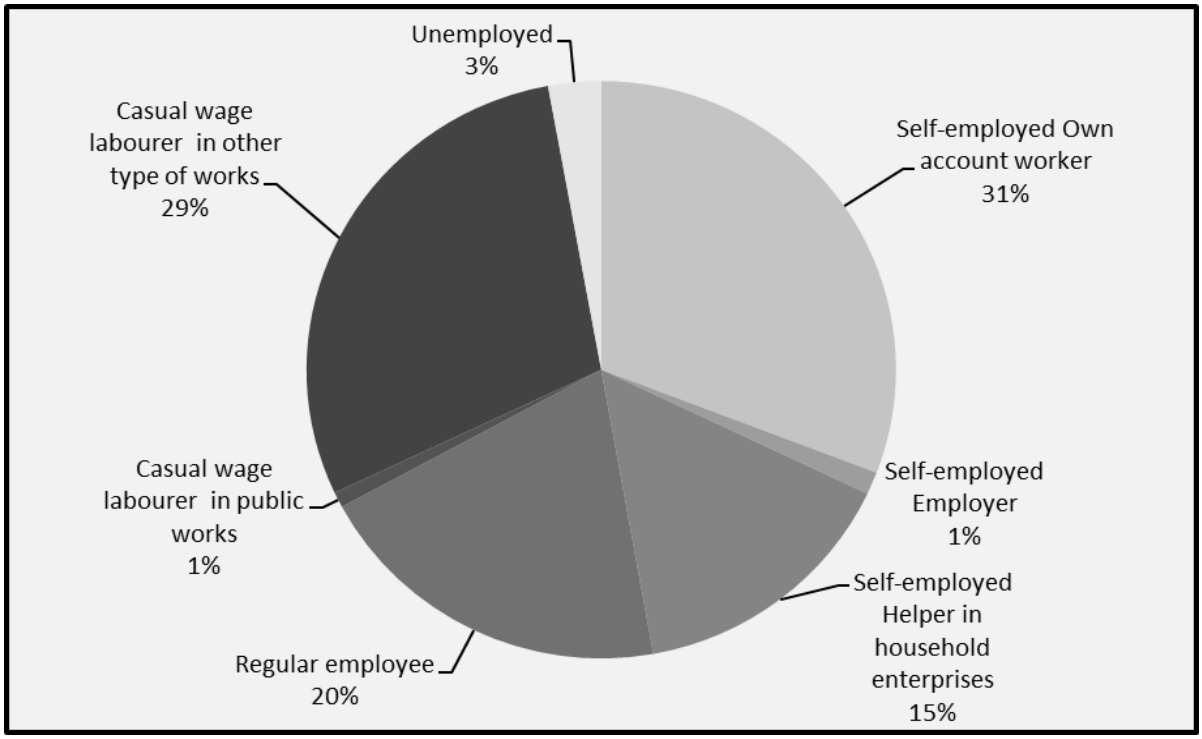
4 Descriptive statistics related to earnings from self-employment

How significant are self-employment businesses for the Indian economy? To what extent are self-employment businesses dependent on level of education in India? To answer the above questions, we start with a descriptive analysis of the labour force distribution in India by different types of activity, i.e. self-employment, casual wage employment, and regular wage employment.

From the distribution of labour force participation by different types of activity (Figure 1) it can be seen that about 47 per cent of the total labour force in India are engaged in self-employment activities; 31 per cent of the total are engaged as self-employed own account workers, 1 per cent as self-employed employers,[1] and 15 per cent as self-employed helpers in household-based enterprises. If we exclude the number of unemployed people from the labour force, the size of self-employment goes up to 48.45 per cent of total employed people, the rest being engaged as either regular employees or casual employees.

[1] By the term 'self-employed own account workers', we refer to the self-employed enterprises that do not employ any hired workers, whereas 'self-employed employers' engage hired workers in their businesses.

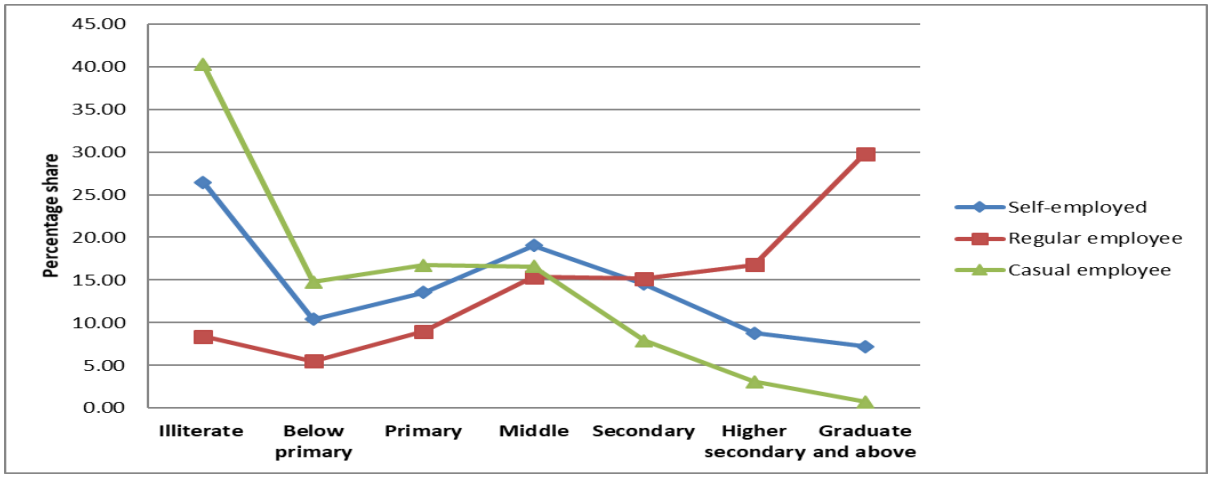
Figure 1: Percentage distribution of labour force participation by types of activity



Source: author's estimation based on NSSO data for 2011/12.

However, the level of participation in different activities may differ across different educational levels. Therefore, a percentage distribution of labour force participation rate in terms of regular employment, casual employment, and self-employment across different levels of education is presented in Figure 2.

Figure 2: Labour force participation rate of regular employment, casual employment, and self-employment across different levels of education



Source: author's estimation based on NSSO data for 2011/12.

Figure 2 shows that the labour force participation rate in regular employment increases with an increase in the educational level—and at a much faster rate after Higher secondary education than across the lower educational levels. Trends of self-employment and casual employment across educational levels are similar, but completely opposite to the trends of regular employment. Moreover, illiterate people account for the highest participation rate in self-employment, which falls significantly at the Below primary educational level, then shows a

slightly increasing trend to Middle school educational level and a continuous decrease after Middle school.

However, the participation rate in different activities may differ across different age groups as well. Hence, a percentage distribution of labour force participation across different types of activities by age group is presented in Table 1.

Table 1: Percentage distribution of labour force participation across different types of activities, by age group

UPSS status	Age group		
	15–29	30–44	45–59
Self-employed own account worker	15.22	35.11	43.65
Self-employed employer	0.42	1.44	2.00
Self-employed helper in household enterprises	23.76	12.66	8.09
Regular employee	21.02	19.88	18.75
Casual wage labourer in public works	0.78	1.00	0.75
Casual wage labourer in other type of works	31.24	29.08	26.44
Unemployed	7.57	0.83	0.32
Total	100	100	100

Source: author's estimation based on NSSO data for 2011/12.

Table 1 shows that the participation of those in the younger age groups (i.e. 15–29 and 30–44) in self-employed businesses is much lower than those in the older age group. Although a substantial number of youths are engaged as helpers in household-based enterprises, they move towards permanent settlement as their age grows. Thus, those belonging to the older age group account for a higher participation as self-employed own account workers and employers than those in the younger age groups. For instance, 43.65 per cent of people are engaged as self-employed own account workers in the age group 45–59 years. If we add together self-employed own account workers, employers, and helpers in household-based enterprises, we find about 54 per cent of people involved in self-employment in the age group of 45–59 years. Unemployment among the youth is much higher than among those in the older age group; perhaps most of the unemployed youth move towards self-employment as they grow older.

Moreover, in order to see whether there is greater participation in self-employment among those in the older group between educated and uneducated, we have estimated the labour force participation rate across different activities by different age groups for educated and uneducated separately, as presented in Table 2. The category 'educated' includes people with an educational level of Secondary and above, while the 'uneducated' category includes people with a below Secondary level of education.

Table 2 shows that the percentage of people engaged in self-employment in the young age group (15–29 years) is much lower than in the older age groups (30–44 and 45–59 years) for both the educated and uneducated. Both educated and uneducated youth are largely engaged as helpers in household-based enterprises without pay and as casual wage workers in other types of work (not public works). However, as they grow older, both educated and uneducated youth move towards permanent settlement either by finding a regular job or by working as own account workers or employers in household-based enterprises—i.e. as self-employed (Bairagya 2018).

Table 2: A comparative picture of different age groups of educated and uneducated people by different types of occupation (%)

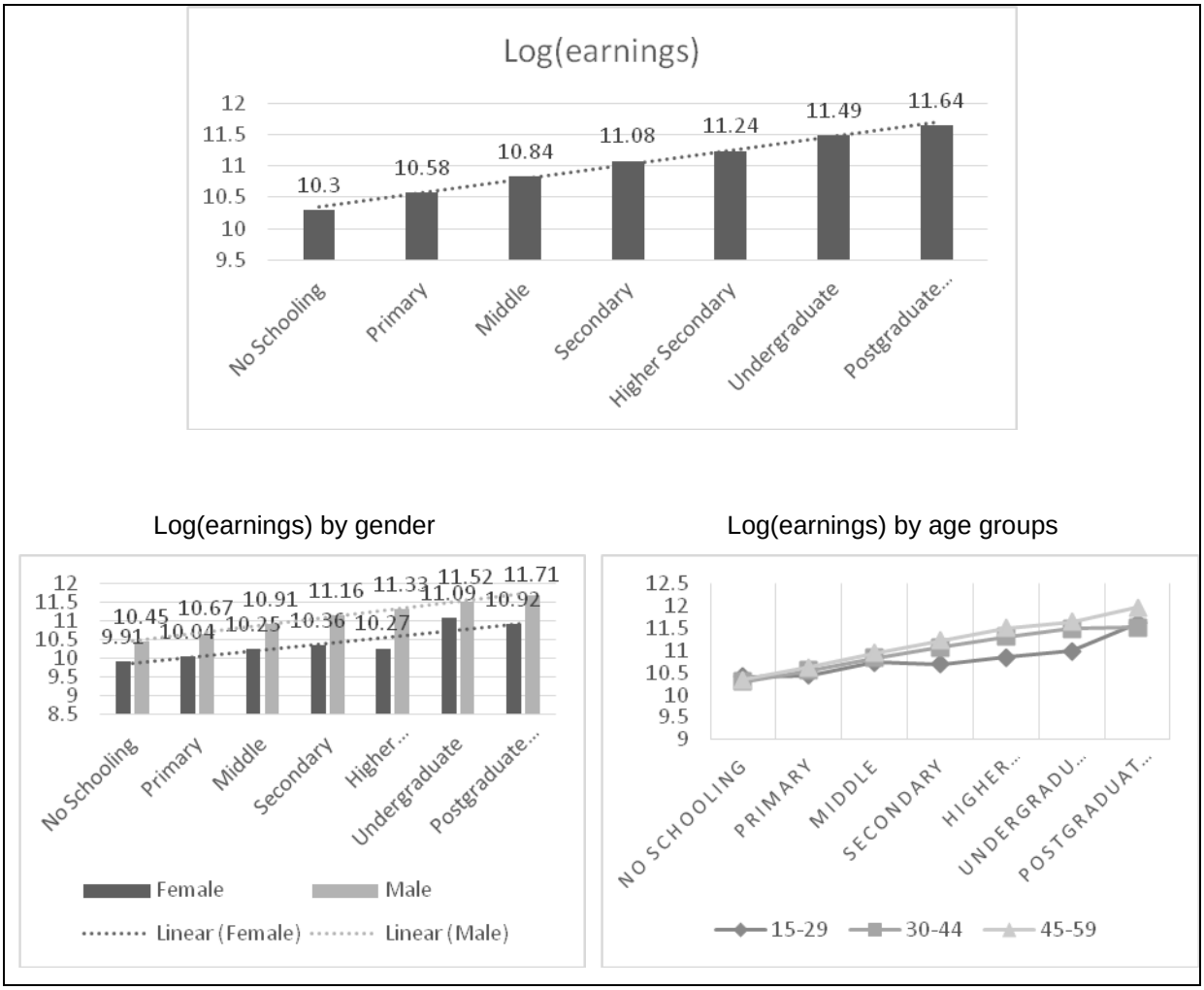
Activities	Educated (age groups)			Uneducated (age groups)		
	15–29	30–44	45–59	15–29	30–44	45–59
Self-employed own account worker	14.8	36.2	40.8	15.5	34.6	44.6
Self-employed employer	0.6	2.9	4.1	0.3	0.7	1.3
Self-employed helper in household enterprises	22.5	10.5	2.7	24.5	13.8	9.8
Regular employee	33.3	38.6	47.3	13.8	10.6	9.5
Casual wage labourer in public works	0.4	0.4	0.2	1	1.3	0.9
Casual wage labourer in other type of works	14.5	9.9	4.5	41.1	38.6	33.6
Unemployed	13.8	1.6	0.5	3.9	0.4	0.3
Total	100	100	100	100	100	100

Source: Bairagya (2018).

As the core focus of the paper is to estimate the returns to education in respect of self-employment businesses, we have presented the mean earnings of self-employment businesses across different levels of education for India in Figure 3. As NSSO data for 2011/12 do not provide the information related to the earnings of self-employment, for a subsequent analysis we have used IHDS data for the year 2011/12, wherein the self-employed businesses typically include self-employed own account workers and self-employed employers.

Figure 3 reveals a consistent increase in mean earnings with an increase in level of education. However, the extent of this increase differs across different levels of education. Moreover, the mean earnings of females are much lower than those of males across all the educational levels. In terms of earnings across the different age groups, there is not much difference between people belonging to the age groups of 15–29, 30–44, and 45–59 at the lower educational levels, but the gap between age groups increases considerably at the Middle educational level before coming down at the Postgraduate level.

Figure 3: Mean Log(earnings) of self-employment businesses across different levels of education in India



Source: author's estimation based on IHDS data for 2011/12.

5 Determinants of earnings from self-employment

5.1 Endogeneity problem

As discussed in Section 2, we adopted an IV 2SLS technique with 'father's education' as an instrument for 'education of the self-employed individuals' in order to deal with the problem of potential endogeneity of our key independent variable 'education' with the dependent variable 'earnings from self-employment businesses'. In this context, it is important to mention that we are directly able to match the father with his son/daughter using the variable 'father's ID' in the IHDS data (where the father is still alive and living with his son/daughter in the same household). In addition to father's ID, we have used the information available in the IHDS data related to the 'years of education of the household head's father'. Using this, we could identify the years of education of the father for the self-employed individuals who are also the heads of family. Using the above two results, we are able to identify the years of education of the father for 7,962 self-employed individuals, out of 9,457. The results of the OLS and IV models are presented in Table 3. The descriptions of the variables used in different regression equations and their descriptive statistics are presented in the Appendix.

Table 3: Determinants of earnings from self-employment businesses in India based on OLS and IV models

	OLS model	IV model	
	[dependent variable: Log(earnings)]	First Stage [dependent variable: Education-years]	Second Stage [dependent variable: Log(earnings)]
Education years	0.056***		0.094***
	(0.003)		(0.006)
Father's years of education		0.446***	
		(0.01)	
Age	0.02***	0.003	0.019***
	(0.005)	(0.021)	(0.006)
Age-square	-0.0002***	-0.001***	-0.0001**
	(0.0001)	(0)	(0.0001)
Female	-0.508***	-2.692***	-0.353***
	(0.035)	(0.228)	(0.07)
SC and ST	-0.276***	-1.045***	-0.198***
	(0.029)	(0.121)	(0.032)
Minority	0.133***	-1.432***	0.229***
	(0.027)	(0.11)	(0.032)
Rural	-0.488***	-0.97***	-0.453***
	(0.024)	(0.095)	(0.028)
Married	0.081**	-0.059	0.059
	(0.034)	(0.161)	(0.044)
Household's principal activity is business	0.387***	0.078	0.39***
	(0.025)	(0.1)	(0.028)
Business in home or in a fixed place	-0.058**	0.994***	-0.081***
	(0.026)	(0.103)	(0.028)
Member of business community	0.211***	0.63***	0.159***
	(0.039)	(0.147)	(0.043)
Member of religious community	0.146***	0.968***	0.092**
	(0.037)	(0.14)	(0.041)
Member of caste group	-0.034	0.311*	-0.069
	(0.043)	(0.17)	(0.048)
Member of political party	0.181***	0.927***	0.095
	(0.052)	(0.201)	(0.06)
Attend panchayat/ ward meeting regularly	-0.046*	0.41***	-0.064**
	(0.026)	(0.101)	(0.028)
Constant	9.959***	7.168***	9.618***
	(0.115)	(0.459)	(0.149)
Number of observations =	9457		7962
F =	F(15, 9441)= 196.72		F(15, 7946)= 133.82
Prob > F =	0		0
			Centred R-squared= 0.238
R-squared =	0.238		Uncentred R-squared= 0.238
Root MSE =	1.056		1.037

Source: author's estimation based on IHDS data for 2011/12.

Table 3 shows that years of education is positive and statistically significantly related to the earnings of self-employed businesses based on both the OLS and IV estimates, implying that earnings increase with an increase in the years of education of the self-employed individuals. In fact, the first-stage regression results of the IV model show that father's years of education is positive and statistically significantly related to the years of education of the self-employed individuals. Moreover, the value of the Cragg-Donald Wald F statistic and Kleibergen rk Wald F statistic of weak identification test are 1855.9 and 2101.5, respectively, which are way above the Stock-Yogo weak ID test critical value at 10 per cent level, implying that the instrument we have used (father's years of education) is not a weak instrument. Moreover, the Hansen J statistic shows that the equation has been exactly identified. However, it is important to note that the

magnitude of coefficient related to education (i.e. average rate of return) is higher in the IV-based estimation than in the OLS-based estimation. In fact, the above result is not unusual: Psacharopoulos and Patrinos (2004), based on an extensive review of the studies related to returns to education, observe that IV-based estimates are often higher than the OLS-based estimates of the returns to education.

Although differences are found in terms of the magnitude of coefficients related to control variables between OLS and IV models, no significant difference is found in terms of their signs and statistical significance levels.

The coefficient associated with age is positive and significant for both models, implying that self-employment earnings increase with an increase in the owner's age, as age may reflect their experience, too. However, the coefficient associated with age-square is negative and significant, indicating that self-employment earnings increase with an increase in the owner's age up to a certain age, after which they start to decrease.

The coefficient associated with females is negative and statistically significant, implying that females account for lesser earnings than males in self-employed businesses.

More importantly, the coefficient associated with SC and ST people is negative and significant for both models, implying that SC and ST people account for lesser earnings than non-SC/ST people.

Surprisingly, the coefficient for minority religious community is positive and statistically significant for both models (OLS and IV).

Further, the coefficient associated with people living in rural areas is negative and significant, implying that rural self-employed people account for lesser earnings than urban self-employed people.

Interestingly, the coefficient for self-employed people belonging to families with self-employed businesses as their principal activity is positive and significantly related to earnings. This may be because when self-employment is the main source of income for a household, they put more effort into it in order to ensure their survival.

Finally, having social capital—such as membership of a business community, religious community, or political party—has a positive and significant impact on earnings from self-employment businesses across all the models.

5.2 Double-selection models

Apart from the aforementioned issue of potential endogeneity, another serious issue that arises here is sample selection bias, as selection of only self-employed individuals may not generate a random sample. In order to correct the sample selection bias of the earnings equation, we adopted a double-selection model, considering the selection of employment participation at the first stage of selection and choice of self-employment at the second stage of selection. We first estimated these two selection equations separately and then used a bivariate model considering their interdependency. The self-selection models identify the factors that determine both employment participation and choice of self-employment. The results of both the univariate and bivariate self-selection models are presented in Table 4.

Table 4: Determinants of selection of employment participation and self-employment in India

	Single probit estimations				Bi-variate probit estimations		
	Employment participation		Choice of self-employment		Coefficients		Marginal effects
	Coefficients	Marginal effects	Coefficients	Marginal effects	Employment participation	Choice of self-employment	
Age	0.017***	0.006***	0.011***	0.001***	0.016***	0.015***	0.001***
	(0)	(0)	(0)	(0)	(0)	(0)	(0)
Female	-1.369***	-0.435***	-0.838***	-0.043***	-1.368***	-0.758***	-0.037***
	(0.007)	(0.002)	(0.013)	(0.001)	(0.007)	(0.013)	(0.001)
SC and ST	0.255***	0.088***	-0.19***	-0.008***	0.251***	-0.17***	-0.004***
	(0.008)	(0.003)	(0.014)	(0.001)	(0.008)	(0.013)	(0)
Minority	0.004	0.001	0.103***	0.005***	0.005	0.107***	0.004***
	(0.009)	(0.003)	(0.014)	(0.001)	(0.009)	(0.014)	(0.001)
Education level - primary	0.008	0.003	0.128***	0.007***	0.008	0.14***	0.005***
	(0.01)	(0.003)	(0.018)	(0.001)	(0.01)	(0.018)	(0.001)
Education level - middle	0.253***	0.088***	0.286***	0.016***	0.253***	0.316***	0.015***
	(0.01)	(0.003)	(0.017)	(0.001)	(0.01)	(0.017)	(0.001)
Education level - secondary	0.19***	0.066***	0.361***	0.023***	0.19***	0.397***	0.021***
	(0.012)	(0.004)	(0.019)	(0.002)	(0.012)	(0.019)	(0.001)
Education level - higher secondary	0.266***	0.094***	0.39***	0.026***	0.265***	0.434***	0.024***
	(0.014)	(0.005)	(0.022)	(0.002)	(0.014)	(0.022)	(0.002)
Education level - undergraduate	0.478***	0.176***	0.421***	0.03***	0.48***	0.478***	0.031***
	(0.018)	(0.007)	(0.025)	(0.002)	(0.018)	(0.025)	(0.002)
Education level - post-graduate	0.673***	0.254***	0.316***	0.02***	0.667***	0.399***	0.027***
	(0.024)	(0.01)	(0.033)	(0.003)	(0.024)	(0.033)	(0.003)
Rural	0.172***	0.056***	-0.308***	-0.016***	0.166***	-0.293***	-0.01***
	(0.008)	(0.003)	(0.012)	(0.001)	(0.008)	(0.012)	(0.001)
Married	0.994***	0.33***	0.632***	0.033***	0.995***	0.604***	0.029***
	(0.008)	(0.003)	(0.014)	(0.001)	(0.008)	(0.014)	(0.001)
Member of religious community	-0.141***	-0.046***	-0.053***	-0.002***	-0.14***	-0.059***	-0.003***
	(0.012)	(0.004)	(0.019)	(0.001)	(0.012)	(0.018)	(0.001)
Member of caste group	0.084***	0.029***	-0.05**	-0.002**	0.084***	-0.041*	-0.001
	(0.014)	(0.005)	(0.022)	(0.001)	(0.013)	(0.021)	(0.001)
Member of political party	-0.094***	-0.031***	0.025	0.001	-0.094***	0.008	-0.0003
	(0.018)	(0.006)	(0.027)	(0.001)	(0.018)	(0.026)	(0.001)
Attend panchayat/ ward meeting regularly	-0.005	-0.002	-0.003	0	-0.005	-0.007	-0.0003
	(0.008)	(0.003)	(0.013)	(0.001)	(0.008)	(0.013)	(0.0004)
Constant	-1.167***		-2.197***		-1.159***	-2.41***	
	(0.012)		(0.022)		(0.012)	(0.025)	
athrho						0.639***	
						(0.01)	
rho						0.564	
						(0.007)	
Number of observations		204565		204565			204565
LR chi2(16)		86665.63		15615.26			
Wald chi2(32)							68359.61
Prob > chi2		0		0			0
Pseudo R2		0.33		0.20			
Log likelihood		-87473.58		-30837.69			-115739.67

Source: author's estimation based on IHDS data for 2011/12.

Table 4 shows that the results for determinants of both employment participation and choice of self-employment are almost similar, based on univariate and bivariate selection models. All educational levels are positive and significantly related to participation in employment. With an increase in the level of education, the coefficient also increases, i.e. the probability of participating in the labour force increases with an increase in the level of education. However, with an increase in the educational level, the probability of engaging in self-employment increases only until Undergraduate educational level before starting to decrease, an indication that highly educated individuals with an educational level of Postgraduate and above prefer not to engage in self-employment.

Moreover, age is positively and significantly related to both employment participation and engaging in self-employment.

Females display less likelihood of joining either the labour force or self-employment businesses than males.

People belonging to SC and ST communities show a greater likelihood of joining the labour force. Once they join the labour force, they are less likely to choose self-employment.

Similarly, the likelihood of people in rural areas joining the labour force is greater, but they are less likely to choose self-employment.

Married people are more likely both to join the labour force and to engage in self-employment.

Controlling for the above selection models, we subsequently estimated the earnings equations for self-employed individuals in India, which are presented in Table 5.

Table 5 presents the results of the earnings equations based on univariate corrections for selection biases and bivariate corrections for selection biases. The coefficient for years of education is positive and statistically significant across all the models, but with a varying magnitude. The results show a substantial difference in terms of the magnitude (value) of coefficients of control variables across different types of selection models, whereas there is no significant difference in the signs and statistical significance levels of the variables, except for coefficients related to the variables 'female' and 'married'. Moreover, the sign and statistical significance levels of the control variables in the selection models are also consistent with OLS and IV models, except for coefficients related to the variables 'female' and 'married'. The coefficient associated with females is negative across all models, but is not statistically significant for the single selection model for selection with employment participation or for the bivariate probit selection model. Most importantly, IMRs for both employment participation and choice of self-employment are statistically significant, showing the importance of correction for selection bias when estimating the returns to education for self-employment.

Table 5: Determinants of earnings from self-employment businesses in India based on different types of selection models

	Single selection		Double selection	
	Heckman with selection for employment participation	**Heckman with selection for self-employment**	**Univariate probit correction**	**Bivariate probit correction**
Education-years	0.048***	0.074***	0.065***	0.048***
	(0.004)	(0.006)	(0.007)	(0.004)
Age	0.014***	0.027***	0.021***	0.015***
	(0.005)	(0.006)	(0.006)	(0.005)
Age-square	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Female	-0.001	-1.075***	-0.561**	-0.005
	(0.162)	(0.19)	(0.254)	(0.162)
SC and ST	-0.345***	-0.402***	-0.461***	-0.344***
	(0.036)	(0.051)	(0.054)	(0.036)
Minority	0.133***	0.197***	0.193***	0.132***
	(0.027)	(0.034)	(0.034)	(0.027)
Rural	-0.533***	-0.684***	-0.715***	-0.531***
	(0.028)	(0.069)	(0.069)	(0.027)
Married	-0.269**	0.502***	0.145	-0.267**
	(0.114)	(0.143)	(0.184)	(0.115)
Household's principal activity is business	0.389***	0.391***	0.393***	0.389***
	(0.025)	(0.025)	(0.025)	(0.025)
Business in home or in a fixed place	-0.055**	-0.057**	-0.055**	-0.055**
	(0.026)	(0.026)	(0.026)	(0.026)
Member of business community	0.209***	0.21***	0.209***	0.209***
	(0.039)	(0.039)	(0.039)	(0.039)
Member of religious community	0.185***	0.115***	0.153***	0.184***
	(0.039)	(0.038)	(0.04)	(0.039)
Member of caste group	-0.057	-0.065	-0.086*	-0.057
	(0.044)	(0.045)	(0.045)	(0.044)
Member of political party	0.207***	0.196***	0.22***	0.207***
	(0.053)	(0.052)	(0.053)	(0.053)
Attend panchayat/ ward meeting regularly	-0.046*	-0.046*	-0.045*	-0.046*
	(0.026)	(0.026)	(0.026)	(0.026)
Constant	10.815***	8.034***	8.956***	10.805***
	(0.29)	(0.644)	(0.711)	(0.29)
Inverse Mills Ratio for employment participation	-0.567***		-0.539***	
	(0.176)		(0.177)	
Inverse Mills Ratio for choice of self-employment		0.776***	0.732***	-0.563***
		(0.255)	(0.256)	(0.177)
Number of observations =	9457	9457	9457	9457
F =	F(15, 9440)= 185.25	F(15, 9440)= 185.16	F(15, 9439)= 174.97	F(15, 9440)= 185.23
Prob > F =	0	0	0	0
R-squared =	0.239	0.239	0.24	0.239
Root MSE =	1.055	1.055	1.055	1.055

Source: author's estimation based on IHDS data for 2011/12.

5.3 Estimation by different educational levels

Since the studies by Duraisamy (2002), Iversen et al. (2010, 2011), and Vasudeva Dutta (2006, show the importance of estimating returns to education at different levels of education, we have, in addition to estimating average returns to education years, introduced dummies for different levels of education, i.e. Primary, Middle, Secondary, Higher secondary, Undergraduate, and Postgraduate and above. Here, Illiterate/Below Primary is considered as the reference category. Determinants of self-employment earnings at different levels of education are presented in Table 6.

Table 6: Determinants of self-employment earnings at different levels of education

	No selection	Single selection		Double selection	
	OLS without selection	Heckman with selection for employment participation	Heckman with selection for self-employment	Univariate probit correction	Bivariate probit correction
Age	0.019***	0.013**	0.021**	0.025**	0.013**
	(0.005)	(0.006)	(0.01)	(0.01)	(0.006)
Age-square	-0.0002***	-0.0002***	-0.0002***	-0.0002***	-0.0002***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Female	-0.523***	0.032	-0.672	-0.919	0.03
	(0.041)	(0.23)	(0.708)	(0.705)	(0.231)
SC and ST	-0.28***	-0.356***	-0.313**	-0.613***	-0.355***
	(0.028)	(0.041)	(0.157)	(0.18)	(0.041)
Minority	0.13***	0.13***	0.147*	0.255***	0.129***
	(0.027)	(0.027)	(0.085)	(0.09)	(0.027)
Education level - Primary	0.116***	0.115***	0.139	0.279**	0.115***
	(0.038)	(0.038)	(0.112)	(0.118)	(0.038)
Education level - Middle	0.298***	0.224***	0.347	0.562**	0.225***
	(0.035)	(0.045)	(0.231)	(0.236)	(0.045)
Education level - Secondary	0.485***	0.428***	0.547*	0.862***	0.429***
	(0.04)	(0.045)	(0.29)	(0.301)	(0.045)
Education level - Higher secondary	0.644***	0.567***	0.711**	1.028***	0.567***
	(0.046)	(0.055)	(0.311)	(0.321)	(0.055)
Education level - Undergraduate	0.791***	0.654***	0.863***	1.13***	0.654***
	(0.052)	(0.073)	(0.335)	(0.34)	(0.073)
Education level - Postgraduate and above	0.968***	0.784***	1.022***	1.13***	0.786***
	(0.069)	(0.099)	(0.26)	(0.26)	(0.098)
Rural	-0.488***	-0.538***	-0.54**	-0.929***	-0.536***
	(0.024)	(0.03)	(0.246)	(0.27)	(0.03)
Married	0.089**	-0.295*	0.2	0.421	-0.295*
	(0.036)	(0.155)	(0.522)	(0.521)	(0.156)
Household's principal activity is business	0.393***	0.393***	0.393***	0.391***	0.393***
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Business in home or in a fixed place	-0.05**	-0.049**	-0.051**	-0.051**	-0.049**
	(0.025)	(0.025)	(0.025)	(0.025)	(0.025)
Member of business community	0.204***	0.202***	0.204***	0.202***	0.202***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Member of religious community	0.151***	0.192***	0.142**	0.136**	0.191***
	(0.038)	(0.041)	(0.057)	(0.057)	(0.041)
Member of caste group	-0.033	-0.059	-0.042	-0.125*	-0.059
	(0.045)	(0.045)	(0.059)	(0.064)	(0.045)
Member of political party	0.182***	0.21***	0.186***	0.247***	0.21***
	(0.056)	(0.057)	(0.059)	(0.063)	(0.057)
Attend panchayat/ ward meeting regularly	-0.042	-0.042	-0.043	-0.045*	-0.042
	(0.026)	(0.026)	(0.026)	(0.026)	(0.026)
Constant	10.066***	10.985***	9.544***	7.408***	10.979***
	(0.122)	(0.377)	(2.452)	(2.5)	(0.378)
Inverse Mills Ratio for employment participation		-0.619**		-0.767***	
		(0.247)		(0.264)	-0.617**
Inverse Mills Ratio for choice of self-employment			0.206	1.5	(0.248)
			(0.971)	(1.034)	
Number of observations =	9460	9460	9460	9460	9460
F =	F(20, 9439)= 136.51	F(21, 9438)= 130.68	F(21, 9438)= 130.04	F(22, 9437)= 125.18	F(21, 9438)= 130.65
Prob > F =	0	0	0	0	0
R-squared =	0.24	0.241	0.24	0.241	0.241
Root MSE =	1.055	1.054	1.055	1.054	1.054

Source: author's estimation based on IHDS data for 2011/12.

Table 6 presents the results of earnings equations based on no correction for selection bias, univariate correction for selection bias, and bivariate correction for selection bias. The results do

not show a significant difference in the sign or statistical significance levels of variables across different types of selection models, excepting two variables ('female' and 'married'). The coefficients for different educational levels are found to be positive and statistically significant. Moreover, the magnitude of the coefficient increases with an increase in the level of education for all the regression models. Moreover, the sign and statistical significance levels of the coefficient associated with the other control variables remain consistent even after introducing the different educational dummies in the regression equations. Surprisingly, the coefficients associated with Primary and Middle educational levels are not statistically significant in the case of the Heckman selection correction model with selection for self-employment.

6 Estimated rate of returns to education

Although a large number of studies during the last five decades have concentrated on estimating the rate of returns to education for wage employment among different socio-economic groups, comparatively few studies have concentrated on self-employment, especially for the Indian economy. Even fewer have considered endogeneity and selection issues together, while estimating the rate of returns to different levels of education. Estimates of the average rate of returns to education in respect of self-employment in India are presented in Table 7.

Table 7: Estimated average rate of returns to education in respect of self-employment in India

			Single selection		Double selection	
	OLS without selection	IV 2SLS	Heckman with selection for employment participation	Heckman with selection for self-employment	Univariate probit correction	Bivariate probit correction
Average rate of returns to education	5.6	9.4	4.8	7.4	6.5	4.8

Source: author's estimation based on IHDS data for 2011/12.

Table 7 shows that the average rate of returns to education in self-employment in India is positive across different selection models. However, the value differs across different models. The average rate of returns to education based on the IV model is much higher than all the other estimates (OLS as well as other selection models). In fact, the above result is not unusual: Psacharopoulos and Patrinos (2004), based on an extensive review of the studies related to returns to education, observe that IV-based estimates are often higher than the OLS-based estimates of returns to education.

In addition to estimating the average rate of returns, following the studies of Psacharopoulos (1989, 1994) and Duraisamy (2002), we have estimated the rate of returns of per year of education across different levels by:

$$r_k = (\beta_k - \beta_{k-1})/Y_k \quad (4)$$

where r_k is the rate of returns of per year of education for the *k*th level of education; β_k and β_{k-1} are the coefficients for the *k*th level of education and immediate previous level of education, respectively (depicted in Table 7); Y_k is the number of years of education within the *k*th level of education.

It is also important to mention that one needs to pursue 5 years, 3 years, 2 years, 2 years, 3 years, and 2 years of education to complete the Primary, Middle, Secondary, Higher secondary,

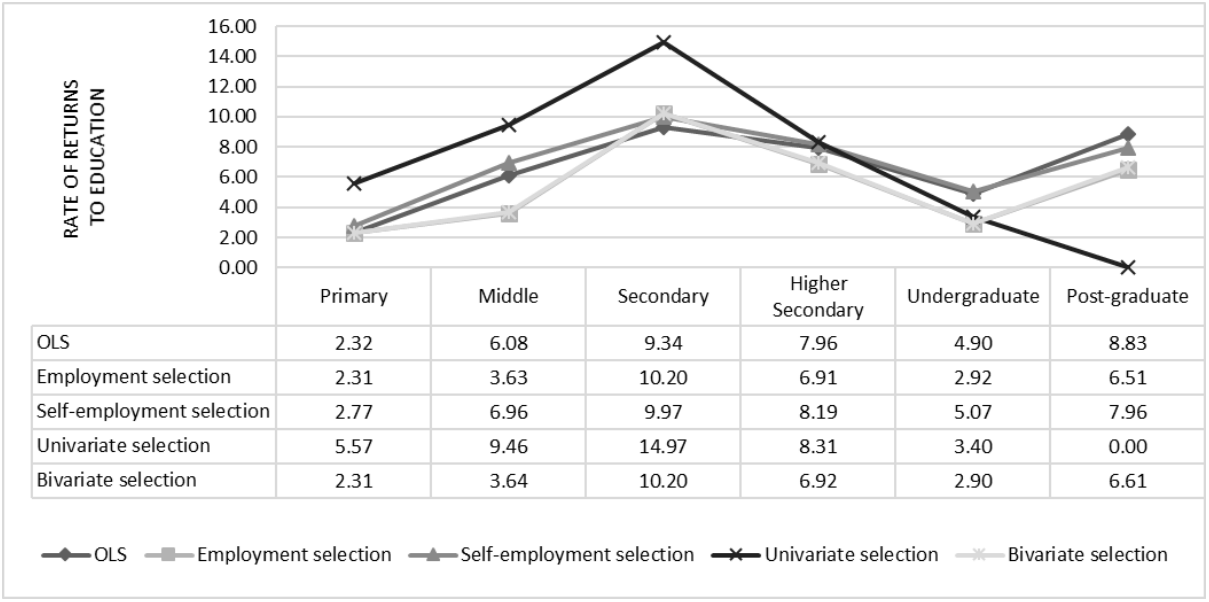
Undergraduate, and Postgraduate levels of education. Using eq. (4), the rate of returns to education from Middle level to Postgraduate level can be estimated. Therefore, following Vasudeva Dutta (2006), we have estimated the rate of returns to the Primary level of education by:

$$r_{primary} = (\beta_{primary}) / (Y_{primary}) \quad (5)$$

where $r_{primary}$, $\beta_{primary}$, and $Y_{primary}$ represent the rate of returns, coefficient, and years of schooling at the Primary level of education, respectively.

The estimated rates of returns to education in self-employment across different educational levels for different models are presented in Figure 4.

Figure 4: Estimated rates of returns to education in self-employment across different educational levels



	Primary	Middle	Secondary	Higher Secondary	Undergraduate	Post-graduate
OLS	2.32	6.08	9.34	7.96	4.90	8.83
Employment selection	2.31	3.63	10.20	6.91	2.92	6.51
Self-employment selection	2.77	6.96	9.97	8.19	5.07	7.96
Univariate selection	5.57	9.46	14.97	8.31	3.40	0.00
Bivariate selection	2.31	3.64	10.20	6.92	2.90	6.61

Source: author's estimation based on IHDS data for 2011/12.

From Figure 4 it can be seen that the rate of returns to education in self-employment increases with an increase in the educational level up to Secondary education before starting to decrease. After Undergraduate level, it again shows an increasing trend. The trend of movement of the rate of returns across different educational levels is similar for the different selection models. However, the estimated *value* of the rate of returns for different levels of education is very sensitive to the specification of selection models. A comparison of the univariate selection model and bivariate selection model shows substantial differences in the value of the rate of returns to education. This points to the need to exercise a fair degree of care in the choice of an appropriate model for estimating the rates of returns to education at different levels and for drawing inferences therefrom.

7 Conclusion

Given the importance of education in influencing people's earnings in general, the focus of this study was to analyse the returns to education in respect of non-farm self-employed businesses in India, using nationwide individual- and household-level data provided by India Human

Development Survey (IHDS) for the Indian economy for the year 2011/12 with corrections for potential endogeneity and a double-selection model for the Mincerian earnings equation.

A descriptive analysis of returns to education shows a consistent increase in mean earnings with an increase in the educational level in India. Moreover, the results, based on univariate and bivariate selection models, point to the possibility of increased labour force participation with an increase in educational level. However, with an increase in educational level, the probability of engaging in self-employment increases up to the Undergraduate educational level but starts to decrease thereafter. The coefficient for years of education is positive and statistically significant for all the models, but with a varying magnitude. The results show a substantial difference in the magnitude of coefficients with respect to the control variables for different types of selection model, but it is important to note that there is no significant difference in the sign and statistical significance levels of the variables. Moreover, the coefficients for different educational levels are positive and statistically significant, with the magnitude of coefficients increasing with an increase in the level of education.

The average rate of returns to education in self-employment in India is positive for different selection models. However, the value differs across different models, implying that the estimation of returns to education is highly sensitive to the specifications of the models, considering the issues of selection bias and potential endogeneity. Further, the average rate of returns to education based on the IV model is much higher than all other estimates (OLS as well as other selection models), supporting the observation of Psacharopoulos and Patrinos (2004) (based on an extensive review of the studies related to returns to education) that IV-based estimates are often higher than OLS-based estimates of returns to education.

Further, the trend in the movement of the rate of returns across different educational levels is similar for the different selection models. However, the estimated *value* of the rate of returns across different levels of education varies with the type of selection model. A comparison of univariate selection and bivariate selection models reveals a substantial difference in the value of the rates of returns to education. An estimation of the rate of returns to education in self-employment is indeed essential if relevant educational policies are to be framed and educated youth encouraged to engage in self-employment in order to utilize their entrepreneurial talents; these, in turn, may even help address problems such as educated unemployment and underemployment in the economy. Therefore, care should be exercised when choosing an appropriate model(s) for estimating the rate of returns to education and drawing policy implications therefrom.

References

Bairagya, I. (2018). 'Why is unemployment higher among the educated?'. *Economic & Political Weekly*, 53(7): 43–51.

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

Bigsten, A. (1984). *Education and Income Determination in Kenya*. Farnham, UK: Gower Publishing.

Card, D. (1999). 'The causal effect of education on earnings'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics* (Vol. 3, pp. 1801–63). Amsterdam: North Holland.

Card, D. (2001). 'Estimating the return to schooling: Progress on some persistent econometric problems'. *Econometrica*, 69(5): 1127–60. https://doi.org/10.1111/1468-0262.00237

Co, C.Y., I.N. Gang, and M.S. Yun (1999). 'Switching models with self-selection: Self-employment in Hungary'. Unpublished paper. Rutgers University, New Jersey.

Colclough, C., G. Kingdon, and H. Patrinos (2010). 'The changing pattern of wage returns to education and its implications'. *Development Policy Review*, 28: 733–47. https://doi.org/10.1111/j.1467-7679.2010.00507.x

Corral, L., and T. Reardon (2001). 'Rural nonfarm incomes in Nicaragua'. *World Development*, 29(3): 427–42. DOI: https://doi.org/10.1016/S0305-750X(00)00109-1

Desai, S., and R. Vanneman (2011–2012). *India Human Development Survey-II (IHDS-II)*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2018-08-08. https://doi.org/10.3886/ICPSR36151.v6

Duraisamy, P. (2002). 'Changes in returns to education in India, 1983–94: By gender, age-cohort and location'. *Economics of Education Review*, 21(6): 609–22. https://doi.org/10.1016/S0272-7757(01)00047-4

Fan, S., L. Zhang, and X. Zhang (2002). 'Growth, inequality and poverty in rural China: The role of public investments'. Research Report 125. Washington, DC: International Food Policy Research Institute.

Garcia-Mainar, I., and V.M. Montuenga-Gomez (2005). 'Education returns of wage earners and self-employed workers: Portugal vs. Spain'. *Economics of Education Review*, 24(2): 161–70. https://doi.org/10.1016/j.econedurev.2004.07.004

Griliches, Z. (1977). 'Estimating the returns to schooling: Some econometric problems'. *Econometrica*, 45(1): 1–22. DOI: 10.2307/1913285

Ham, J.C. (1982). 'Estimation of a labour supply model with censoring due to unemployment and underemployment'. *The Review of Economic Studies*, 49(3): 335–54. https://doi.org/10.2307/2297360

Hamilton, B.H. (2000). 'Does entrepreneurship pay? An empirical analysis of the returns to self-employment'. *Journal of Political Economy*, 108(3): 604–31. https://doi.org/10.1086/262131

Heckman, J.J. (1976). 'The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models'. *Annals of Economic and Social Measurement*, 5(4): 475–92. Available at: https://www.nber.org/chapters/c10491.pdf (accessed 24 January 2020).

Heckman, J. (1979). 'Sample selection bias as a specification error'. *Econometrica*, 47: 153–61. DOI: 10.2307/1912352

Heitmueller, A. (2004). 'Public–private sector wage differentials in Scotland: An endogenous switching model'. Discussion Paper 992. Bonn: IZA Institute of Labor Economics.

Hossain (2004). 'Rural non-farm economy in Bangladesh: A view from household surveys'. Occasional Paper Series 40. Dhaka: Center for Policy Dialogue.

ILO (2011). 'Statistical update on employment in the informal economy'. Geneva: International Labour Organization, Department of Statistics. Available at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---stat/documents/presentation/wcms_157467.pdf (accessed 31 January 2020).

Iversen, J., N. Malchow-Møller, and A. Sørensen (2010). 'Returns to schooling in self-employment'. *Economics Letters*, 109(3): 179–82. https://doi.org/10.1016/j.econlet.2010.05.014

Iversen, J., N. Malchow-Møller, and A. Sørensen (2011). 'The returns to education in entrepreneurship: Heterogeneity and non-linearities'. *Entrepreneurship Research Journal*, 1(3): 1–36. https://doi.org/10.2202/2157-5665.1001

Kavuma, S. Namirembe, O. Morrissey, and R. Upward (2015). 'Private returns to education for wage-employees and the self-employed in Uganda'. WIDER Working Paper 2015/021. Helsinki: UNU-WIDER.

Kingdon, G., and M. Söderbom (2007a). 'Education, skills, and labor market outcomes: Evidence from Ghana'. Oxford: Social Science Open Access Repository. Available at: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-69486 (accessed 17 January 2020).

Kingdon, G., and M. Söderbom (2007b). 'Education, skills, and labour market outcomes: Evidence from Pakistan'. Oxford: Social Science Open Access Repository. Available at: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-69510 (accessed 17 January 2020).

Kolstad, I., and A. Wiig (2015). 'Education and entrepreneurial success'. *Small Business Economics*, 44(4): 783–96. https://doi.org/10.1007/s11187-014-9621-1

Lanjouw, P., and A. Shariff (2002). 'Rural nonfarm employment in India: Access, income and poverty impacts'. NCAER Working Paper 81. New Delhi: National Council of Applied Economics Research.

Lanjouw, P., and A. Shariff (2004). Rural non-farm employment in India: Access, incomes and poverty impact. *Economic and Political Weekly*, 39(40): 4429–46.

Laszlo, S. (2005). 'Self employment earnings and returns to education in rural Peru. *The Journal of Development Studies*, 41(7): 1247–87. https://doi.org/10.1080/00220380500170915

Lee, L. (1983). 'Generalised models with selectivity'. *Econometrica*, 51: 507–12. DOI: 10.2307/1912003

Mincer, J. (1970). 'The distribution of labor incomes: A survey with special reference to the human capital approach'. *Journal of Economic Literature*, 8(1): 1–26. Available at: www.jstor.org/stable/2720384 (accessed 24 January 2020)

Mukherjee, A., and X. Zhang (2005). 'Rural nonfarm development in China and India: The role of policies and institutions'. Discussion Paper 24. Washington, DC: International Food Policy Research Institute.

NCEUS (2008). 'Definitional and statistical issues'. Task Force Report. National Commission for Enterprises in the Unorganized Sector, Government of India.

NSSO (2014). 'Employment–unemployment situation in India 2011–2012, 68th round'. Report 554 (68/10/1). National Sample Survey Organisation, Government of India.

Nyakundi E., E.M.W. Simatwa, T.M. Ayodo, and P.O. Nyabera (2017). 'Analysis on returns to levels of education and challenges faced by the motor spare parts urban self-employed in Kisii County, Kenya'. *International Journal of Current Research*, 9(8): 55606–13. Available at: www.journalcra.com/sites/default/files/issue-pdf/24942.pdf (accessed 24 January 2020).

Planning Commission (2000). 'Report of the Task Force on Employment Opportunities'. New Delhi: Government of India.

Psacharopoulos, G. (1985). 'Returns to education: A further international update and implications'. *Journal of Human Resources*, 20: 583–604. DOI: 10.2307/145686

Psacharopoulos, G. (1989). 'Time trends of the returns to education: Cross-national evidence'. *Economics of Education Review*, 8(3): 225–31. https://doi.org/10.1016/0272-7757(82)90002-4

Psacharopoulos, G. (1994). 'Returns to investment in education: A global update'. *World Development*, 22(9): 1325–43. https://doi.org/10.1016/0305-750X(94)90007-8

Psacharopoulos, G., and H.A. Patrinos (2004). 'Returns to investment in education: A further update'. *Education Economics*, 12(2): 111–34. https://doi.org/10.1080/0964529042000239140

Rankin, N., J. Sandefur, and F. Teal (2010). 'Learning and earning in Africa: Where are the returns to education high?'. WPS/2010-02, CSAE Working Paper Series. Centre for the Study of African Economies, University of Oxford.

Sandefur, J., P. Serneels, and F. Teal (2006). 'African poverty through the lens of labor economics: Earnings and mobility in three countries'. Working Paper 060. University of Oxford, Global Poverty Research Group.

Schwiebert, J. (2012). 'Revisiting the composition of the female workforce – A Heckman selection model with endogeneity'. Discussion Paper 502. Wirtschaftswissenschaftlichen Fakultät, Universität Hannover.

Singhari, S., and S. Madheswaran (2016). 'The changing rates of return to education in India: Evidence from NSS data'. Working Paper 358. Karnataka, India: Institute for Social and Economic Change.

Söderbom, M., F. Teal, A. Wambugu, and G. Kahyarara (2006). 'The dynamics of returns to education in Kenyan and Tanzanian manufacturing'. *Oxford Bulletin of Economics and Statistics*, 68(3): 261–88. https://doi.org/10.1111/j.1468-0084.2006.00162.x

Taylor, J.E., and A. Yunez-Naude (2000). 'The returns from schooling in a diversified rural economy'. *American Journal of Agricultural Economics*, 82(May): 287–97. https://doi.org/10.1111/0002-9092.00025

Teal, F. (2001). *Education, Incomes, Poverty and Inequality in Ghana in the 1990s*. Oxford: Centre for the Study of African Economies, University of Oxford.

Tilak, J.B.G. (1987). *The Economics of Inequality in Education*. New Delhi: Sage Publications.

Tunali, I. (1986). 'A general structure for models of double selection and an application to a joint migration/earnings process with remigration'. In R. Ehrenberg (ed.), *Research in Labor Economics*, 8 (Part B, pp. 235–82). Greenwich, CT: JAI Press.

Vasudeva Dutta, P. (2006). 'Returns to education: New evidence for India, 1983–1999'. *Education Economics*, 14(4): 431–51. https://doi.org/10.1080/09645290600854128

Williams, D.R. (2002). 'Returns to education and experience in self-employment: Evidence from Germany'. IRISS Working Paper Series 2002-04. Luxembourg: Integrated Research Infrastructure in the Socio-economic Sciences at CEPS-INSTEAD (now LISER).

Appendix

Table A1: Descriptions of the variables used in different regression equations

Variables	Descriptions
Socio-economic and demographic characteristics	
Age	**Age of the self-employed individual**
Age-square	Square term of the age
Female	Female = 1 if the self-employed individual is female Ref: Male = 0
SC and ST	SC and ST = 1 if the self-employed individual belongs to the socially disadvantaged groups (schedule caste and schedule tribe) Ref: Non-SC and ST = 0
Minority	Minority = 1 if the self-employed individual belongs to a religious minority community (i.e. non-Hindu) Ref: Hindu = 0
Education level - Primary	Education level - Primary = 1 if the self-employed individual completed primary school education
Education level - Middle	Education level - Middle = 1 if the self-employed individual completed middle school education
Education level - Secondary	Education level - Secondary = 1 if the self-employed individual completed secondary education
Education level - Higher secondary	Education level - Higher secondary = 1 if the self-employed individual completed higher secondary education
Education level - Undergraduate	Education level - Undergraduate = 1 if the self-employed individual completed undergraduation
Education level - Postgraduate	Education level - Postgraduate = 1 if the self-employed individual completed post-graduation education Ref: Self-employed individual does not have any formal schooling = 0
Rural	Rural = 1 if the self-employed individual belongs to rural areas Ref: Urban = 0
Married	Married = 1 if the self-employed individual is married Ref: Single
Business-related factors	
Household's principal activity is business	Household's principal activity is business = 1 Ref: Other activities are the primary activity = 0
Business at home or in a fixed place	Business at home or in a fixed place = 1 Ref: business in moving place = 0
Social capital	
Member of business community	Member of business community = 1 Ref: Non-member of business community = 0
Member of religious community	Member of religious community = 1 Ref: Non-member of religious community = 0
Member of caste group	Member of caste group = 1 Ref: Non-member of caste group = 0
Member of political party	Member of political party = 1 Ref: Non-member of political party = 0
Attend panchayat/ward meetings regularly	Attend panchayat/ward meetings regularly = 1 Ref: Do not attend panchayat/ward meetings regularly = 0

Source: author's estimation based on IHDS data for 2011/12.

Table A2: Descriptive statistics of the variables used in the regression equations

Variables	No. of observations	Mean	Standard deviation	Minimum	Maximum
Log(earnings)	9,460	10.89	1.21	4.38	16.25
Age	9,569	42.99	12.87	5	98
Age-square	9,569	2014.28	1193.84	25	9604
Female	9,569	0.13	0.33	0	1
SC and ST	9,569	0.18	0.39	0	1
Minority	9,569	0.22	0.42	0	1
Education level - Primary	9,569	0.16	0.36	0	1
Education level - Middle	9,569	0.27	0.44	0	1
Education level - Secondary	9,569	0.17	0.38	0	1
Education level - Higher secondary	9,569	0.11	0.31	0	1
Education level - Undergraduate	9,569	0.09	0.28	0	1
Education level - Postgraduate	9,569	0.04	0.19	0	1
Rural	9,569	0.48	0.50	0	1
Married	9,569	0.85	0.36	0	1
Household's principal activity is business	9,569	0.68	0.47	0	1
Business in home or in a fixed place	9,569	0.76	0.42	0	1
Member of business community	9,569	0.09	0.29	0	1
Member of religious community	9,569	0.13	0.33	0	1
Member of caste group	9,569	0.09	0.28	0	1
Member of political party	9,569	0.05	0.21	0	1
Attend panchayat/ward meetings regularly	9,569	0.28	0.45	0	1

Source: author's estimation based on IHDS data for 2011/12.