Returns to Education in Self-Employment in India:

An Application of Double-Selection Model with Endogeneity

Indrajit Bairagya

Assistant Professor,

Centre for Human Resource Development,

Institute for Social and Economic Change,
Bangalore, India.

Email: indrajit@isec.ac.in; indrajitisec@gmail.com

Introduction

- Education (human capital accumulation through skill formation) often acts as barrier for the workers to move from one sector to another.

- Basic education increases the productivity and wages of the workers [Bigsten, 1984; Fan et al., 2002; Lanjouw and Sariff, 2004].

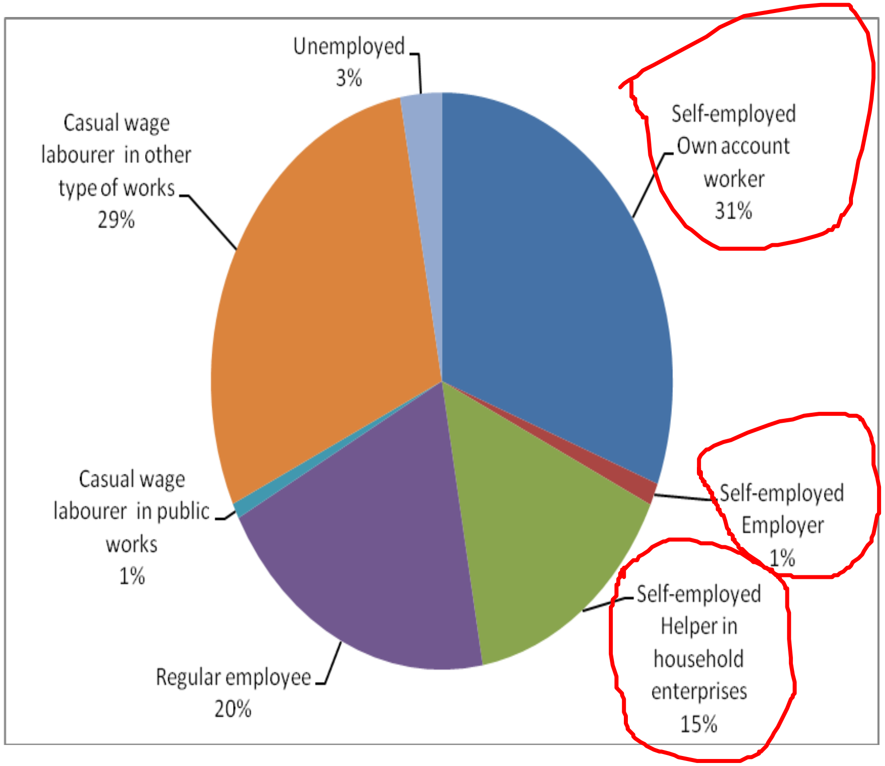
- Absence of education among a large number of individuals in rural India have held back the growth of the rural nonfarm sector [Mukherjee and Zhang, 2008].

- **Less educated households rely** on low-paying farm wage employment **or very low productive non-farm sector** rather than salaried employment–
  - evidence is given for India by Lannjouw and Shariff (2002); Planning Commission (2000),
  - for Bangladesh by Hossain (2004) and
  - for Nicaragua by Corral and Reardon (2001).

- **Education's pay off also differ across different types of employment.** An additional schooling has a lesser effect on earnings for the self-employed compared to the wage-employed [Taylor and Yunez-Naude (2000) ; Hamilton 2000; Williams (2002); Iversen et. al. (2010); Kavuma et. al. (2015)] .

Distribution of LFPR by different types of activity in India

	Activity status	Percentage
Labour force	self-employed own account workers	11.89
	self-employed employer	0.54
	self-employed helper in household enterprises	5.38
	regular employee	6.83
	Casual labourer in public works	0.31
	casual labourer in other works	10.48
	unemployed	1.00
Not in labour force	attended educational institution	27.82
	attended domestic duties only	11.55
	attended domestic duties and was also engaged in free collection of goods	9.95
	rentiers, pensioners , remittance recipients, etc.	1.48
	not able to work due to disability	1.18
	begging, prostitution, etc.	3.21
	Others	8.38
Total		100



Source: Authors' estimation based on NSSO data for 2011-12.

Research Gap

Although a large number of studies have focussed on estimating the returns to education in wage employment (both regular and casual) in India (for instance, Tilak, 1987; Duraisamy, 2002; Vasudeva Dutta, 2007; Singhari and Madheswaran, 2016 etc.), studies are limited which have focussed on estimating the returns to education in non-farm self-employment at the national level in India.

Objective

- This study focuses on estimating the returns to education of self-employment businesses in India.
- In addition, given the fact that different studies have used different types of regression models [OLS, Heckman-selection model, multinomial selection model (Lee, 1983; Dubin and McFadden, 1984; Bourguignon et al. 2007) and 2SLS] to estimate the returns to education, the paper has been extended to assess the sensitivity of the estimation of returns to education across the selection of different types of regression models.

Methodology

- Our starting point is an earning equation similar to **Mincerian wage equation** (Mincer, 1970).

$$\text{Log}Y_i = X_i\beta_i + u_i \qquad \text{(i)}$$

- However, it is well-established in the literature that the OLS based estimation of the earning equation suffers with selection bias.

- Labour force participation **selection bias correction using Heckman (1976; 1979)** procedure has become increasingly popular among researchers with a wide body of research developed.
- In the first stage, $\text{Prob}\ (P = 1\ |Z) = Z_i\gamma + u_i$ (ii-a)

- Using equation (ii-a) one can estimate the predicted probability of the individuals to join in labour force. The second stage involves the estimation of the earnings equation by correcting the sample selection bias by way of including the above predicted probabilities as an added explanatory variable (Inverse mills ratio).
- In the second stage, the earnings equation can be written as $Y^* = X\beta + u$ (ii-b)

- This variable can not be observed for those who are not in the labour force. The conditional expected earnings for the employed individuals can be written as

$$E[Y \mid X, P=1] = X\beta + E[u \mid X, P=1] \qquad \text{(ii-c)}$$

- Error terms of the equations (ii-a) and (ii-b) follow joint normal distribution .

Methodology –contd…

- In our case, **we certainly have a selection bias for participation to job market.**

- In addition, **we have another selection bias for the choice of self-employment**; given the other options for casual and regular wage employment for those have participated in the job market.

- However, it **has not gained much appeal for selection bias correction for more than one stage**, even if existing in the data, sometimes.

- In fact, **it may lead to 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.

- Tunali (1986) has suggested a double selection model which can be used for this case.

Methodology- Double Selection Model

- 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**.

$$P^{*} = Z'_{i}\,\gamma + u_{i} \qquad \text{(iii-a)}$$

$$q^{*} = T'_{i\,\delta} + v_{i} \qquad \text{(iii-b)}$$

- Here, P^{*} and q^{*} are the latent variables. P and q 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.

Methodology- Double Selection Model

- **Another important issue arises regarding the independency of the two selections** i.e., whether the decision of choice for self-employment is independent from the choice of joining the labour market or these two are interdependent.

- To stay away from this issue of independency, **we have estimated the earnings equation considering both independency and interdependency** between two selection decisions in two separate models.

- In the first model, following Heitmueller (2004), we have first estimated two correction terms (inverse mills ratio) from two separate probit models and then using these correction terms, estimated the earnings equation.

- In the second model, considering the fact of interdependency between two selection decisions and following Tunali (1986) and Ham (1982), we have estimated a correction term (inverse mills ratio) based on a bivariate probit estimation of the two selection equations and then including the correction terms in the Mincerian earnings equation, we have estimated the earnings equation.

Database

- We have used the nation-wide individual and household-level India Human Development Survey (IHDS) Data for the Indian economy for the year 2011-12.

- IHDS data provides information on the earnings from the self-employed businesses

- Self-employment is a household-based business and also that earnings from it constitute household earnings.

- It is difficult to identify the actual decision maker when it comes to self-employment small businesses.

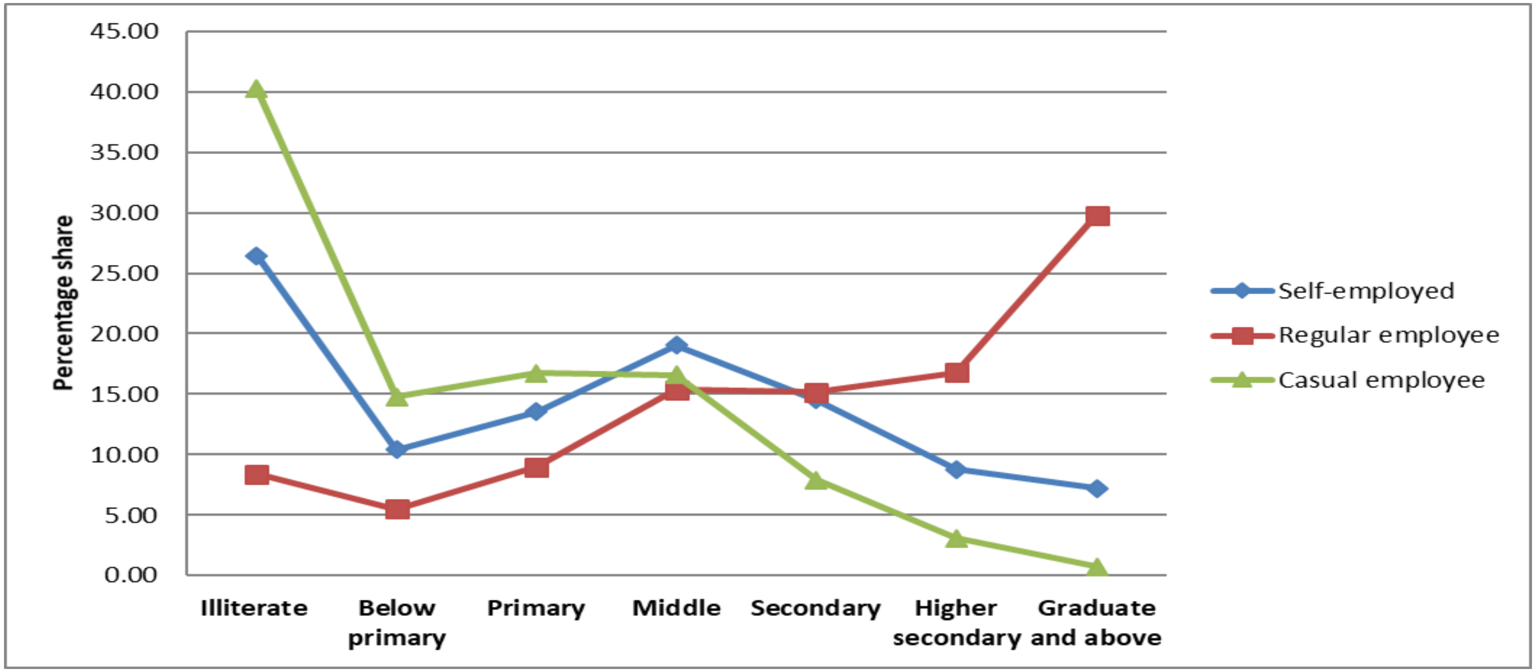
- Interestingly, IHDS 2011-12 data included **a question** on who is the decision maker of business activities from among the member of households.

- It provides detailed accounts of gross receipts and also of expenditure incurred on different inputs such as raw materials, labour, electricity, water, transport, repayment for loan and taxes.

- The difference between the gross receipts and payments is considered as earnings from the business for a given year.

- In addition to earnings, it provides information on a number of variables related to the socio-economic features of the households and individuals.

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



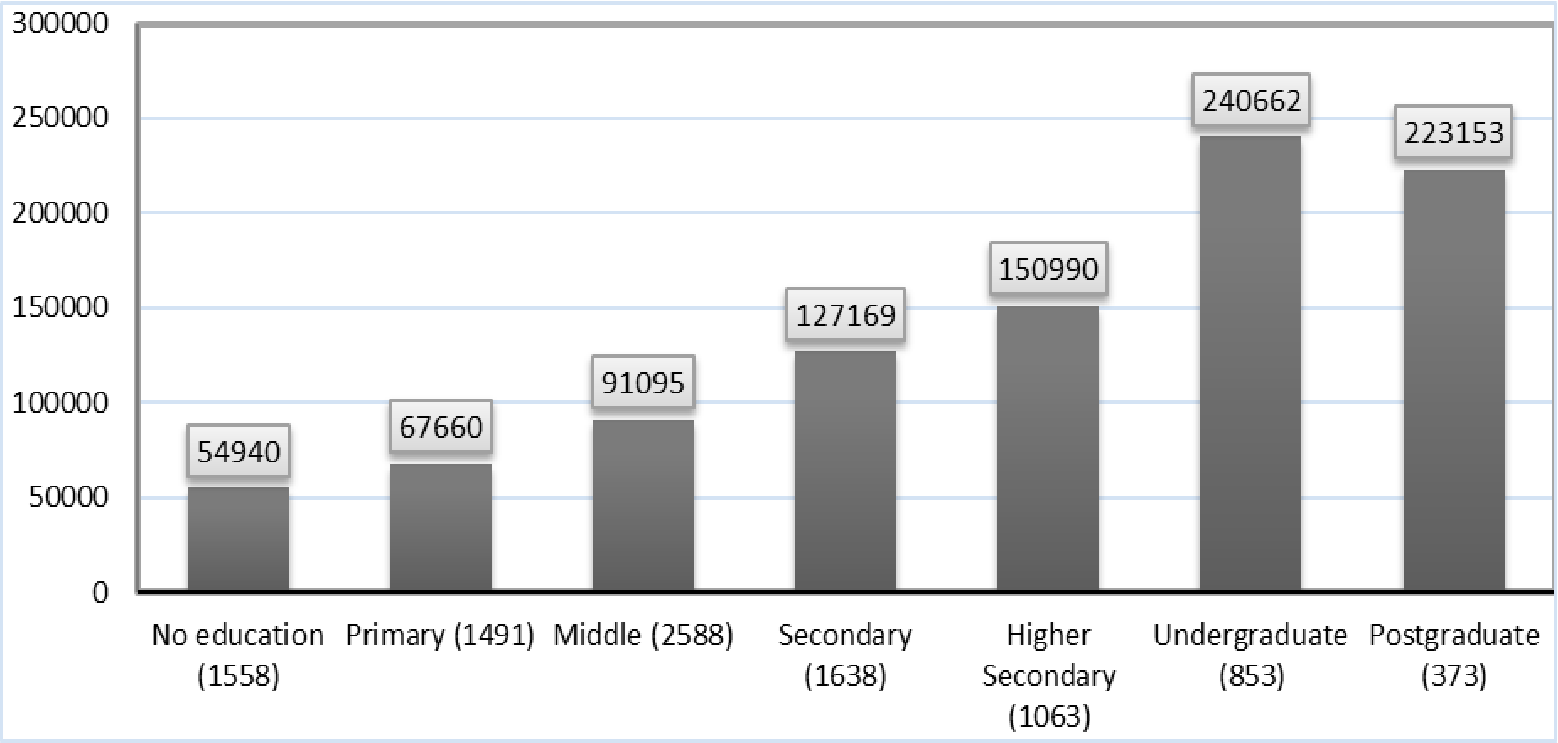
Source: Author’s estimation based on NSSO data for 2011-12.

Age group-wise percentage distribution of labour force across different types of activity

UPSS Status	Age groups		
	15-29	30-44	45-60
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: Authors' estimation based on NSSO data for 2011-12.

Mean earnings (in Indian rupees) in self-employment businesses across different levels of education in India



Source: Author’s estimation based on IHDS data for 2011-12.

Determinants of selection for employment participation and self-employment in India

	Single Probit		Bivariate Probit	
	Employment participation	Choice of Self-employment	Employment participation	Choice of Self-employment
Age	0.016*** (0)	0.011*** (0)	0.016*** (0)	0.015*** (0)
Female	-1.37*** (0.007)	-0.839*** (0.013)	-1.368*** (0.007)	-0.761*** (0.013)
SC and ST	0.255*** (0.008)	-0.202*** (0.013)	0.251*** (0.008)	-0.182*** (0.013)
Education level -primary	0.008 (0.01)	0.124*** (0.018)	0.007 (0.01)	0.136*** (0.018)
Education level -middle	0.253*** (0.01)	0.277*** (0.017)	0.253*** (0.01)	0.307*** (0.017)
Education level -secondary	0.19*** (0.012)	0.35*** (0.019)	0.19*** (0.012)	0.384*** (0.019)
Education level -Higher secondary	0.266*** (0.014)	0.376*** (0.022)	0.265*** (0.014)	0.42*** (0.022)
Education level -undergraduate	0.478*** (0.018)	0.404*** (0.025)	0.48*** (0.018)	0.46*** (0.025)
Education level -post-graduate	0.672*** (0.024)	0.298*** (0.033)	0.666*** (0.024)	0.38*** (0.032)
Rural	0.172*** (0.008)	-0.315*** (0.012)	0.166*** (0.008)	-0.301*** (0.012)
Married	0.994*** (0.008)	0.632*** (0.014)	0.995*** (0.008)	0.603*** (0.014)
Member of religious community	-0.141*** (0.012)	-0.043** (0.019)	-0.139*** (0.012)	-0.047** (0.018)
Member of caste group	0.084*** (0.014)	-0.05** (0.022)	0.084*** (0.013)	-0.041* (0.021)
Member of political party	-0.094*** (0.018)	0.029 (0.027)	-0.094*** (0.018)	0.012 (0.026)
Attend panchayat/ ward meeting	-0.005 (0.008)	-0.006 (0.013)	-0.005 (0.008)	-0.009 (0.013)
Constant	-1.165*** (0.012)	-2.157*** (0.022)	-1.158*** (0.012)	-2.363*** (0.024)
athrho			0.638*** (0.01)	
rho			0.563 (0.007)	
Number of observations	204565	204565	204565	
LR chi2(16)	86665.42	15559.25		
Wald chi2(32)			68328.15	
Prob > chi2	0	0	0	
Pseudo R2	0.33	0.20		
Log likelihood	-87473.68	-30865.69	-115770.57	

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

	No selection	Single selection		Double selectio	
	OLS without selection	Heckman with selection for employment participation	Heckman with selection for self-employment	Univariate Probit correction	Bivariate Prob correctio
Log(Amount of loan)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)	0.003 (0.002)
Age	0.019*** (0.005)	0.013** (0.006)	0.024** (0.01)	0.029*** (0.01)	0.013** (0.006)
Age-square	0*** (0)	0*** (0)	0*** (0)	0*** (0)	0*** (0)
Female	-0.518*** (0.041)	0.041 (0.229)	-0.905 (0.718)	-1.197 (0.715)	0.039 (0.231)
SC and ST	-0.279*** (0.028)	-0.355*** (0.041)	-0.369** (0.167)	-0.707*** (0.193)	-0.354*** (0.04)
Minority	0.131*** (0.027)	0.132*** (0.027)	0.131*** (0.027)	0.133*** (0.027)	0.132*** (0.027
Education level -primary	0.117*** (0.038)	0.116*** (0.038)	0.174 (0.11)	0.322*** (0.117)	0.116*** (0.038
Education level -middle	0.302*** (0.035)	0.227*** (0.045)	0.425* (0.227)	0.653*** (0.234)	0.227*** (0.045
Education level -secondary	0.491*** (0.04)	0.434*** (0.045)	0.645** (0.284)	0.979*** (0.297)	0.434*** (0.045
Education level -Higher secondary	0.65*** (0.046)	0.571*** (0.055)	0.815*** (0.305)	1.15*** (0.315)	0.572*** (0.055
Education level -undergraduate	0.796*** (0.052)	0.658*** (0.073)	0.973*** (0.326)	1.259*** (0.332)	0.658*** (0.074
Education level -post-graduate	0.971*** (0.069)	0.786*** (0.099)	1.103*** (0.25)	1.209*** (0.25)	0.788*** (0.098
Rural	-0.49*** (0.024)	-0.541*** (0.03)	-0.628** (0.255)	-1.059*** (0.281)	-0.539*** (0.03)
Married	0.09** (0.036)	-0.297* (0.155)	0.378 (0.528)	0.633 (0.528)	-0.296* (0.156)
Household's principal activity is business	0.393*** (0.026)	0.393*** (0.026)	0.392*** (0.026)	0.391*** (0.026)	0.393*** (0.026
Business in home or in a fixed place	-0.049* (0.025)	-0.048* (0.025)	-0.049** (0.025)	-0.05** (0.025)	-0.048* (0.025)
Member of business community	0.205*** (0.04)	0.204*** (0.04)	0.205*** (0.04)	0.203*** (0.04)	0.204*** (0.04)
Member of credit-saving group	-0.088** (0.037)	-0.088** (0.037)	-0.088** (0.037)	-0.088** (0.037)	-0.088** (0.037)
Member of religious community	0.153*** (0.038)	0.194*** (0.041)	0.134*** (0.052)	0.138*** (0.052)	0.194*** (0.041
Member of caste group	-0.025 (0.044)	-0.05 (0.045)	-0.046 (0.059)	-0.136** (0.064)	-0.05 (0.045)
Member of political party	0.187*** (0.056)	0.215*** (0.057)	0.2*** (0.061)	0.269*** (0.064)	0.215*** (0.057
Attend panchayat/ ward meeting regularly	-0.039 (0.026)	-0.038 (0.026)	-0.041 (0.027)	-0.046* (0.027)	-0.038 (0.026)
Constant	10.044*** (0.122)	10.968*** (0.377)	8.708*** (2.448)	6.385** (2.508)	10.962*** (0.378)
Inverse Mills Ratio for employment Participation		-0.623** (0.247)		-0.816*** (0.266)	-0.621** (0.248)
Inverse Mills Ratio for choice of self-employment			0.535 (0.982)	1.949* (1.054)	

Methodology- Double Selection with Endogeneity

- It may lead to a serious estimation problem, if we ignored the potential endogeneity of our key independent variable ‘amount of loan’ with the dependent variable ‘earnings from self-employment businesses’.

- The reason for considering ‘amount of loan’ as an endogenous variable is that there are some factors like ownership assets which may determine jointly both the ‘amount of loan’ and ‘earnings from self-employment businesses’.

- Moreover, more earnings from businesses may evident for more collaterals which enhances the chance for receiving greater loans.

- Therefore, addressing the issue of endogeneity of ‘amount of loan’ within the sphere of selection models is important.

- However, whilst the issue of selectivity and endogeneity has been separately dealt with enormously in the literature, studies dealing with both selectivity and endogeneity together within a regression model are relatively less.

Methodology- Double Selection with Endogeneity

- Among the existing studies,

  - Das, Newey and Vella (2003) have addressed the issue of sample selectivity together with endogeneity using a nonparametric estimation framework.

  - Chib et al. (2019) have addressed the issue using a Bayesian framework.

  - Wooldridge (2010) has advocated for augmenting the estimable equation by including the 'inverse mills ratio' estimated from the selection equation and then estimating the estimable equation using two stage least square (2SLS) method.

- The approach, we have followed in this study, is quite similar to Wooldridge (2010).

- We have estimated a correction term (inverse mills ratio) based on a bivariate probit estimation of the two selection equations and then performed an instrumental variable (IV) regression model by incorporating the correction term in the Mincerian earnings equation.

- Likewise, for the univariate double selection model, we have first estimated two correction terms (inverse mills ratio) based on two separate probit models and then using those correction terms estimated the earnings equation in a two-stage least square (2SLS) based IV-regression model.

Determinants of earnings from self-employment businesses in India based on corrections for selection bias within 2SLS-based instrumental variable (IV) model

	No selection	Single selection		Double selection	
	IV without selection bias correction	IV with a single selection for employment participation	IV with a single selection for self-employment	IV with double selection based on Univariate Probit correction	IV with double selection based on Bivariate Probit correction
Log(Amount of loan)	0.076*** (0.011)	0.076*** (0.011)	0.076*** (0.011)	0.076*** (0.011)	0.076*** (0.011)
Age	0.014** (0.006)	0.008 (0.006)	0.029*** (0.011)	0.035*** (0.011)	0.008 (0.006)
Age-square	0* (0)	0 (0)	0* (0)	0 (0)	0 (0)
Female	-0.51*** (0.043)	0.051 (0.241)	-1.716** (0.775)	-2.054*** (0.774)	0.047 (0.242)
SC and ST	-0.232*** (0.031)	-0.308*** (0.044)	-0.512*** (0.18)	-0.907*** (0.207)	-0.307*** (0.044)
Minority	0.165*** (0.03)	0.166*** (0.03)	0.165*** (0.03)	0.167*** (0.03)	0.166*** (0.03)
Education level -primary	0.103** (0.041)	0.102** (0.041)	0.278** (0.118)	0.451*** (0.126)	0.102** (0.041)
Education level –middle	0.307*** (0.038)	0.232*** (0.048)	0.691*** (0.246)	0.957*** (0.253)	0.233*** (0.048)
Education level -secondary	0.546*** (0.043)	0.489*** (0.048)	1.026*** (0.309)	1.414*** (0.323)	0.489*** (0.048)
Education level -Higher secondary	0.702*** (0.049)	0.624*** (0.059)	1.217*** (0.331)	1.607*** (0.344)	0.625*** (0.059)
Education level -undergraduate	0.859*** (0.056)	0.72*** (0.078)	1.409*** (0.354)	1.742*** (0.362)	0.721*** (0.079)
Education level -post-graduate	1.067*** (0.074)	0.882*** (0.105)	1.477*** (0.273)	1.6*** (0.273)	0.885*** (0.105)
Rural	-0.573*** (0.028)	-0.624*** (0.034)	-1.002*** (0.277)	-1.505*** (0.306)	-0.622*** (0.034)
Married	0.077** (0.038)	-0.311* (0.164)	0.973* (0.571)	1.269** (0.572)	-0.309* (0.165)
Household's principal activity is business	0.414*** (0.028)	0.413*** (0.028)	0.412*** (0.028)	0.41*** (0.028)	0.413*** (0.028)
Business in home or in a fixed place	-0.022 (0.027)	-0.021 (0.027)	-0.023 (0.027)	-0.024 (0.027)	-0.021 (0.027)
Member of business community	0.168*** (0.042)	0.166*** (0.042)	0.168*** (0.042)	0.166*** (0.042)	0.166*** (0.042)
Member of credit-saving group	-0.242*** (0.045)	-0.243*** (0.045)	-0.241*** (0.045)	-0.241*** (0.045)	-0.243*** (0.045)
Member of religious community	0.179*** (0.041)	0.22*** (0.045)	0.12** (0.055)	0.124** (0.055)	0.219*** (0.044)
Member of caste group	-0.069 (0.048)	-0.094* (0.049)	-0.135** (0.064)	-0.239*** (0.07)	-0.094* (0.049)
Member of political party	0.18*** (0.059)	0.208*** (0.06)	0.219*** (0.064)	0.3*** (0.068)	0.208*** (0.06)
Attend panchayat/ ward meeting regularly	-0.087*** (0.028)	-0.087*** (0.028)	-0.093*** (0.029)	-0.1*** (0.029)	-0.086*** (0.028)
Constant	9.679*** (0.141)	10.607*** (0.401)	5.521** (2.661)	2.818 (2.731)	10.598*** (0.402)
Inverse Mills Ratio for employment Participation		-0.626** (0.261)		-0.953*** (0.281)	
Inverse Mills Ratio for choice of self-employment			1.665 (1.063)	3.314*** (1.142)	-0.622** (0.262)
Number of observations =	9460	9460	9460	9460	9460
Wald chi2 =	Wald chi2(22)= 2518.40	Wald chi2(23)= 2524.87	Wald chi2(23)= 2527.39	Wald chi2(24)= 2545.91	Wald chi2(23)= 2520.50
Prob > chi2 =	0	0	0	0	0
R-squared =	0.137	0.138	0.138	0.1397	0.1379

Test for endogeneity of different IV models

Ho: Variables are exogeneous

	IV without selection bias correction	IV with a single selection for employment participation	IV with a single selection for self-employment	IV with double selection based on Univariate Probit correction	IV with double selection based on Bivariate Probit correction
Robust score Chi2(1)	49.516 (P = 0)	49.688 (P = 0)	49.376 (P = 0)	49.203 (P = 0)	49.694 (P = 0)
Robust regression F(1, 9436)	54.503 (P = 0)	54.707 (P = 0)	54.356 (P = 0)	54.157 (P = 0)	54.719 (P = 0)

First stage regression statistics of different IV models

Ho: Instruments are weak.

	R-square	Adjusted R-square	Partial R-square	Robust F(1, 9437)	Prob>F	Maximum eigenvalue statistics
IV without selection bias correction	0.0965	0.0943	0.0459	614.961	0	453.69
IV with a single selection for employment participation	0.0965	0.0942	0.0459	614.879	0	453.642
IV with a single selection for self-employment	0.0978	0.0956	0.0461	614.12	0	453.547
IV with double selection based on Univariate Probit correction	0.098	0.0957	0.0461	614.771	0	453.902
IV with double selection based on Bivariate Probit correction	0.0965	0.0942	0.0459	614.904	0	453.642

Critical Values

	10%	15%	20%	25%
2SLS size of nominal 5% Wald test	16.38	8.96	6.66	5.53
LIML size of nominal 5% Wald test	16.38	8.96	6.66	5.53

Estimated rate of returns to education

- Following the study of Psacharopoulos (1989; 1994) and Duraisamy (2002), we have estimated the rate of return of per year of education across different levels by

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

- Using eq(1), the rate of return to education from middle level to post-graduate level can be estimated. Therefore, following Dutta (2006), we have estimated the rate of returns for the primary level of education by

$$r_{primary} = (\beta_{primary)} /(Y_{primary}) \qquad (2)$$

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

Estimated rate of returns to education in self-employment

	OLS without selection	Heckman with selection for employment participation	Heckman with selection for self-employment	Univariate Probit correction	Bivariate Probit correction
Considering No endogeneity					
Education level -primary	2.3	2.3	3.5	6.4	2.3
Education level -middle	6.0	4.5	8.5	13.1	4.5
Education level -secondary	9.8	8.7	12.9	19.6	8.7
Education level -Higher secondary	13.0	11.4	16.3	23.0	11.4
Education level -undergraduate	15.9	13.2	19.5	25.2	13.2
Education level -post-graduate	19.4	15.7	22.1	24.2	15.8
Considering 2SLS model with endogeneity					
Education level -primary	2.1	2.0	5.6	9.0	2.0
Education level -middle	6.1	4.6	13.8	19.1	4.7
Education level -secondary	10.9	9.8	20.5	28.3	9.8
Education level -Higher secondary	14.0	12.5	24.3	32.1	12.5
Education level -undergraduate	17.2	14.4	28.2	34.8	14.4
Education level -post-graduate	21.3	17.6	29.5	32.0	17.7

Source: Author's estimation based on IHDS data for 2011-12.

PSACHAROPOULOS & PATRINOS (2004) based on a vast review of the empirical works argued that Instrumental variable (IV) estimates of the returns to education based on family background are higher than classic Ordinary Least Squares estimates.

Conclusion

- Results based on univariate and bivariate selection models show that the likelihood to participate in labour force increases with an increase in educational levels.

- However, with an increase in educational levels, the probability to join in self-employment increases initially, then it decreases.

- Moreover, the rate of returns to education increases with an increase in educational levels, but the estimated **magnitude** of rates of returns of different levels of education are very sensitive to the specification of the selection models.

- When we compare between with-endogeneity model and without-endogeneity model under a specific type of selection model, marginal differences are found in the rate of returns to education.

- Therefore, one needs to be careful about using an appropriate regression model for estimating the rate of return to education and interpreting the magnitude while suggesting policies based on it.

Thank You