DISCUSSION PAPER SERIES

IZA DP No. 15825

Human Capital and Self-Employment in India: An Empirical Analysis for Different Cohorts

Nabamita Dutta
Saibal Kar
Shaswata Ray

DECEMBER 2022

DISCUSSION PAPER SERIES

IZA DP No. 15825

Human Capital and Self-Employment in India: An Empirical Analysis for Different Cohorts

Nabamita Dutta
University of Wisconsin

Saibal Kar
Centre for Studies in Social Sciences and IZA

Shaswata Ray
Center for Studies in Social Sciences

DECEMBER 2022

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Human Capital and Self-Employment in India: An Empirical Analysis for Different Cohorts*

The ambiguity in the relationship between self-employment and educational attainment is well documented in the literature. Using an extensive individual level dataset from Periodic Labor Force Survey, we estimate the probability of being self-employed in India based on educational attainments. Our results suggest that the probability of being self-employed rises for an individual with education but not monotonically so. Indeed, the impact of education on likelihood of self-employment does not convey much information without considering how the effect varies across gender, caste, age, household size, religion, and industry as various cohorts chosen for this study using 418,297 observations. The probability to be self-employed varies considerably based on gender, caste and age when the level of education rises. A cohort based analysis for determination of self-employment is novel for India along with the findings where college educated women show higher probability of self-employment than men, for example. The importance of considering the non-linearity in the relationship between self-employment and education, usually part of analytical frameworks but inadequately addressed empirically, should be useful for better policies on the interaction between human capital and occupational choice. Robustness analysis considering further cohort effects in terms of household size and religion, buttresses our benchmark results.

JEL Classification: J24, N3, N35

Keywords: Labor Force Survey, education, occupation, self-employment, gender, India

Corresponding author:
Saibal Kar
Centre for Studies in Social Sciences
R-1, B. P. Township
Kolkata 700 094
India
E-mail: saibal@cssscal.org

* The research has been facilitated by financial support from the RBI Endowment at CSSSC. However, it does not in any manner implicate RBI. The paper has benefited from seminar presentation at the BITS-Pilani, Goa. The usual disclaimer applies.

1. Introduction and Related Literature

A voluminous literature has long explored the role of educational attainment on choice of self-employment and entrepreneurship in different countries. Over the last two decades, studies by Cassar (2006), van der Sluis et al. (2005), and Bosma et al. (2004) stress that the human capital is key to entrepreneurial success. Koellinger (2008)further states that higher educational attainment can lead to innovative entrepreneurs. Indeed, these and several such studies are elegantly reviewed and classified in Marvel, Davis and Sproul (2016). The compendium suggests that human capital is not only critical for identifying entrepreneurial opportunities (also see Marvel, 2013 for high-tech entrepreneurship), but also for acquiring financial and innovative advantages over others. Of course, in addition to human capital, business environment and institutions generate varying outcomes associated with quality and distribution of entrepreneurs across developing and developed countries (Chowdhury, Audretsch and Belitski, 2019).Individual-level or even macroeconomic spread of human capital continues to influence choice of entrepreneurship overwhelmingly (see, van der Sluis and van Praag, 2007 for a meta-analysis and how entrepreneurship allows unrestricted use of human capital), while the opposite that entrepreneurship has promoted human capital remains far less convincing. Notwithstanding, country-levelevidence in favor of the relationship varies, as we show in this paper with regard to rounds of data from India. It would also make the first attempt at systematically analyzing the influence of human capital on entrepreneurship for the country.

Note, the above findings do not offer a final verdict that higher educational attainmentsare conducive to self-employment, necessarily.For example, Baum and Silverman (2004) suggest that the relationship between the two variables might be overemphasized. The mixed findings have also been reflected for country specific studies such as those based on students taking courses in entrepreneurship and business planning (Martin, McNally, & Kay, 2013). In fact, university students who took a course in entrepreneurship were less likely to start a business (Oosterbeek, van Praag, and Ysselstein, 2010). Other studies have found similar results for South Africa (Mentoor and Friedrich, 2007) and for Germany (von Graevenitz, Harhoff, and Weber, 2010).

We add to this literature by emphasizing on a non-linear association between educational attainment and entrepreneurship, andto a limited extent, self-employment. To this end, we use firm level data from India. The purpose and contribution are worthy on several accounts. As already mentioned, while a few prior studies (viz. Monsen, Mahagaonkar and Dienes, 2012 about transition between employment and self-employment; Bhagabatula, Mudambi and Murmann, 2019 on innovation and entrepreneurship) discuss variant of this relationship,none focus exclusively on measures of human capital and entrepreneurial attainments in India. Second, the considerably high unemployment among educated youthsin India (see Dasgupta and Kar, 2018) is a concern for public policy, wherein opportunities for 'formal' entrepreneurship could reverse the impending crises for economic growth and income distribution (in view of human capital to growth trajectories in Galor and Zeira, 1993; and Banerjee and Newmann, 1993 where capital market imperfections restrict individuals in developing countries from being self-employed). Notably, a large section of workers outside the formal regime find self-employment (own account enterprises, or OAE according to categorization of the National Sample Survey of India) for sustenance in the unorganized sector of India (see, Marjit and Kar, 2011). This does not resolve the formal unemployment problem in India, however.

RecentlyDutta and Sobel (2018) show that the impact of human capital on entrepreneurship is contingent on levels of financial development of countries. While for low levels of financial development, human capital has a positive and significant effect on entrepreneurship, the impact at higher levels is moderate at best. Building on such literature, we estimate the probability of self-employment based on educational attainment of cohorts of labor force in India. We hypothesize that the impact of educational impact on likelihood of self-employment does not convey much information without considering how the effect varies across gender, caste, age, household size, religion, and industry. As mentioned, educational attainment can result in new knowledge creation, specific skill sets and development of cognition abilities all of which can be beneficial for an entrepreneur. But does it function efficiently in a predominantly non-formal business environment, where 5.4 million micro enterprises constitute 95% of MSMEs and large companies taken together and that only 4% of all such firms can access formal credit?[1]

Section 2 presents an analytical formulation. Section 3 describes the data source and features and explains the empirical methods used via different sub-sections. Section 4 presents empirical results and section 5 offers robustness analysis. Section 6 concludes with summary of results and shortcomings.

2. Analytical Formulations

Consider an individual who chooses between tertiary and secondary education as the highest level completed. The choice is driven by several factors and constraints that the earlier literature has discussed and established at length. It is well-known due to Spence (1973) that the return from signaling by an individual to the prospective employer regarding one's ability/skill is a prime determinant of the level of education chosen. It is complemented by a host of other determinants that have subsequently been added to the list of explanatory variables (see Hartog

[1]Ministry of Statistics and Programme Implementation, India, various years and Enterprise Survey, World Bank provide information on size distribution of firms and access to credit, respectively.

and van den Brink, 2007).Beyond individual and household characteristics, the choice also depends on country level features including wealth distribution, general access to credit, unemployment levels and even, income inequality (see Hu, 2021; Galor, 2012; Galor, Moav and Vollrath, 2009; etc.). Once the chosen level of education is completed, the individualhas to choose further between wage work and self-employment. Once again, a considerably large body of literature discusses the determinants of wage work and self-employment as nodes of a decision tree (Simoes, *et al*. 2016; Parker, 2004; Blanchflower, 2000; Fairlie and Meyer, 1996; Evans and Leighton, 1989). Presently, we contemplate that accumulation of human capital supports taking up entrepreneurship, especially when it improves the functioning of the business model (Unger, *et al*., 2011). The moot question is how?

One way to look at it is that human capital allows individuals to cultivate new ideas, concepts and in many cases scientific inventions and innovations. In case of employees, these skill-driven benefits are utilized by the organization for no extra credit to the worker. In some caseshowever, merit pay, incentives and promotions are offered by the firms (Manso, 2011; Andersson *et al*. 2009, etc.). Since tolerance for failure, long-term contracts, compensation and performance pay are not generic to all firms, many high-ability individuals do not find returns to be adequate in case of employment. Entrepreneurship by high-skilled individuals is rather common in many countries. Second, self-employment can be pervasive for relatively low-skilled individuals as well (see Beladi and Kar, 2015) where retail trade, micro enterprises for manufacturing, home-based production and service, assembly, etc. constitute main business types. Themicro processes are often not part of the activities for large firms, such that smaller enterprises operate via sub-contracting received from large producers and service providers. Such units benefit from skilled entrepreneurs and workers to adapt to newer technologies, as perhaps is the case with the smaller enterprises (96% of all businesses) in Asia (Yoshino and Taghizadeh-Hesary, 2018) and of course, China (99.8% businesses are SMEs according to National Bureau of Statistics, 2018), some of which has also been able to expand to a much bigger size of operation over time. The level of education of the entrepreneur often helps to exercise industry connections with research and development. In developed countries, for example the network of businesses in the United Kingdom benefits from deep clusters associating the government, the research laboratories and the industrial centres (Govt. of UK, 2017). The industrial clusters in India have poor links with research generated in universities or technical schools and have weak mutual feedback. This is usually the reason behind technological dependence on rest of the world beyond basic uptake of scientific processes. Entrepreneurs with better human capital may be able to improve upon this pattern and generate higher returns from such activities. However, for India the relation between human capital and entrepreneurship could get substantially more complicated based on distribution of caste, religion and gender, somewhat in line with patterns of minority entrepreneurship in USA and Europe but with many differences.

3. Description of the Data

3.1 Data Source

We use the data from the 2019-20 edition of the Periodic Labour Force Survey (PLFS) which is conducted annually by the National Statistical Office (NSO, Govt. of India) since 2017.[2] The primary aim of the PLFS has been to estimate the employment and unemployment indicators

[2]The data was accessed from the Ministry of Statistics and Programme Implementation website http://164.100.161.63/unit-level-data-periodic-labour-force-survey-plfs-july-2019-june-2020

in rural and urban areas in India.[3] The survey uses a rotational panel sampling in urban areas where each selected household in urban areas was visited four times – once in the first visit schedule and thrice periodically with the revisit schedule. However, there were no revisits in the rural areas.[4]To make our analysis compatible and minimize biases, we focus on the data from the first visit samples when data from both urban and rural areas were collected. In 2019-20, the survey covered418, 297 persons (240,231 in rural areas and 178,066 in urban areas) belonging to 100,480 households (55,291 in rural areas and 45,189 in urban areas) in total. Below, we describe the variables used in our study.

3.2 Dependent and Independent Variables

Employing the recent wave of demographic data from PLFS, the main focus of our study is toempirically assess the probability of an individual to be self-employed in India based on the education level. The second part of the analysis estimates the probability for different cohorts – gender, caste, age, religion, industry, etc. The dependent variable of interest is a binary choice variable which assumes the value of 1 if an individual is self-employed and 0 otherwise. PLFS provides the information ontheeconomic/non-economic activities an individual has been engaged in during the reference period. Based on the activity status, individuals can be classified as employed, unemployed or not in labor force. Among the employed individuals, those who "operate their own farm or non-farm enterprises or are engaged independently in a profession or trade on own-account or with one or a few partners"are considered to be self-employed in

[3]See https://mospi.gov.in/web/plfs for further details regarding the survey
[4] See http://164.100.161.63/sites/default/files/reports_and_publication/PLFS_2019_2020/Instructions%20to%20Field%20Staff_PLFS_Vol%20I.pdffor further details regarding the sampling design

household enterprises.[5] We consider those individuals who either worked in household enterprise as an own-account worker or as an employer to be "self-employed" for the purpose of our study.

A little over 14 percent of the respondents in our sample were found to be self-employed(see Table 1). Among the respondents who were self-employed, there is a stark contrast between men and women with the former accounting for almost 84 percent of the self-employed individuals. Looking at the breakdown of self-employed individuals across the different caste groups, we find that individuals belonging to Scheduled Caste (SC) are lagging behind the other three groups with only 11 percent of SC individuals being self-employed.Additionally, the Other Backward Classes (OBC) make up close to 42% of the total self-employed individuals. Table 1 presents the summary statistics of our main variable of interest.

[Insert Table 1 about here]

The primary independent variable of interest is the level of education of the respondent. We make use of an education dummy variable (coded from 0 to 4) comprising five ordered categories based on the highest educational attainment of the respondent. These categories correspond to: 0 for individuals who are not literate; 1 for those who either gained literacy without formal schooling or gained literacy through formal schooling but have not completed primary education; 2 for those who have completed either primary, middle or secondary level of education; 3 for individuals who have completed higher secondary education and finally, 4 for respondents who have received a college degree or above. Almost one fourth of the respondents in our sample arenot literate, while on the other hand, only 10 percent have obtained a college or

[5]See http://164.100.161.63/sites/default/files/reports_and_publication/PLFS_2019_2020/Instructions%20to%20Field%20Staff_PLFS_Vol%20I.pdffor further details regarding definitions

higher degree. Around 43 percent of the respondents are revealed to have moderate levels of education completing primary education at the minimum or secondary education at the maximum. Furthermore, there are noticeable differences in levels of educational attainment between the two genders. About 29 percent of the female respondents are found to be not literate, while it is only 18 percent for males. This disparity continues to be present even at higher education levels with almost 13 percent of male respondents having completed at least a college degree, with the corresponding percentage for female respondents being only 9 percent.

Nearly 40% of the people in each caste group have either completed primary, middle or secondary level of education. Additionally, one-fourth of the individuals belonging to the religions of Hinduism, Islam and Sikhism are not literate. Regardless of their religious affiliation, the majority of people appear to have finished either primary, or middle, or secondary schooling. A little over 33 percent of *Jains*(a religious category) appear to have at least a college degree, making them the group with the highest average level of educational attainment.

3.3 Other Independent and Control Variables

As mentioned, our secondary hypothesis aims at assessing the probability of self-employment for different cohorts in India based on gender, caste,age, household size, industry and religion. As explained in the motivation, an in-depth empirical study for India is lacking in terms of how self-employment varies across different cohorts. Most of the variables have been shown to be important determinants of entrepreneurship and, thus, belonging to specific cohorts should matter in terms of self-employment. While the mentioned variables are interacted with education level, we also use most of them as controls. We follow the micro and macro literature on entrepreneurship while choosing controls.

The first variable we consider is gender. The significance of femaleentrepreneurship for creation of new ventures, generating independent sources of income, innovation, wealth creation and building ownership and growth have been emphasized in the literature, all of which can potentially lead to their empowerment (Dutta and Mallick, 2018; Estrin and Mickiewicz, 2011;Fontana, 2009; De Bruin et al, 2006). Caste is undoubtedly important for entrepreneurship in the Indian context. Iyer, Khanna and Varshney (2013) have shown that marginalized castes like Scheduled Castes (SC) and Scheduled Tribes (ST) are underrepresented in ownership of enterprises relative to other castes. Audretsch, Boente and Tamvada (2013) have also found caste to be important in the Indian context. Other than gender and caste, we also consider age as another control as well as a variable interacted with education. The impact of age on entrepreneurship is inconclusive in the literature. Studies like Lévesque and Minniti (2006) and Parker (2009) haveshown that the relationship is negative since entrepreneurial intention and willingness are likely to go down with age. But since experience and entrepreneurial opportunities enhance with age, the relationship can be positive too (Lee and Vouchilas, 2016; Fairlie et al. 2016). Zhang and Acs (2018) have shown the relationship to be non-linear.Factors like household size have been shown to be important as well(Blanchflower, 2000; Hundley, 2000; Krasniqi, 2009). As part of robustness analysis, we consider other variables. We talk about the variables and results in subsequent sections.

As we can see in Table 1, the respondents are evenly split by gender, with women making up roughly 49% of the total. The average age of the respondents is around30 years. In terms of caste, other backward classes (OBC) represent 39.6% of our sample. General caste (referenced here as ‘others’) represents about 28.8% of individuals. For scheduled caste (SC), the percentage is about 17.6% and finally for scheduled tribe (ST) the figure is 13.7%.

3.4 Benchmark specifications

Our main hypothesis studies how education levels of individuals in India impacts their probability of being self-employed. The following probit specification is empirically tested:

$$SE_{irs} = \beta_0 + \beta_1 Edu_{irs} + \eta_i Controls_{irs} + \beta_2\gamma_r + \beta_3\tau_i + \varepsilon_{irs} \quad (1)$$

where, SE_{irs} is the binary choice variable indicating if person I is self-employed or not in location r for state s. Edu_{irs} denotes the education level of person i in location r for state s. The ordered dummy variable, as mentioned, ranges from 0 to 4 with 0 as the base group. We once again remind our reader that 0 refers to individuals who are not literate. To take into account location effect, we consider γ_r which represents location fixed effects. Specifically, we consider whether an individual belongs to the rural area in which the dummy takes the value 1 while a value of 0 would otherwise indicate the individual being in an urban area. Being in an urban or rural area may matter for access and quality of education (Sahoo and Klasen, 2021). Local labor market conditions should also be captured in location fixed effect. τ_i indicates state fixed effect that should take into account state characteristics like societal norms, regional variations in labor market conditions, and specific state policies with regard to labor market.

We start with binary probit models without any control.[6] The idea is to see how much ofthe probability to be self-employed is explained by educational attainment. We control for state fixed effects and location fixed effects (rural or urban). Subsequently, we add relevant controls to our model. The controls considered in our benchmark regression are gender, caste, age, and household size. The other controls are added as part of robustness analysis.

[6] In related studies (Audretsch, Bonte and Tamvada, 2013) multinomial probit models have been used because a variety of occupational types was the dependent variable. In our case, employment vis-à-vis self-employment is the main dependent variable leading to a binary choice problem.

Our secondary hypothesis is tested via the following specification

$$SE_{irs} = \beta_0 + \beta_1 Edu_{irs} + \beta_2 Cohort_{irs} + \beta_3 (Edu*Cohort)_{irs} + \eta_i Controls_{irs} + \gamma_r + \tau_i + \varepsilon_{irs} \quad (2)$$

We are interested in the coefficients β_1 and β_3 in specification (2).The probability to be employed based on specification (1) is given by ($\partial SE_{irs}/\partial Edu_{irs}$). Given that we can only interpret the sign of probit regressions but not their economic significance meaningfully, ($\partial SE_{irs}/\partial Edu_{irs}$)gives us the marginal estimates or the probability of an individual to be employed for different education levels with respect to the base group of not being literate.

($\partial SE_{irs}/\partial Edu_{irs}$) becomes conditional on the coefficient of the cohorts and the cohort value itself in the presence of the interaction terms. ($\partial SE_{irs}/\partial Edu_{irs}$) is given by $\beta_1 + \beta_3 Cohort_{irs}$. The sign and significance of the marginal effects will vary based on the two coefficients and the cohort value.

3.5 Empirical Methods

We consider limited dependent variable (LDV) models as has been used in many empirical studies employing binary dependent variables (Dutta, Beladi and Kar, 2022; Webster and Piesse, 2018; Swamy et al., 2001). Ordinary least square (OLS), under these circumstances, suffer from challenges like predicted probabilities lying outside the unit interval. Specifically, we employ probit specifications that, similar to logit, use Maximum Likelihood Estimation (MLE) but with a normal distribution function of the error terms.

The initial specification can be written as

$$\Pr(SF = 1) = F(\widehat{X}\,\Omega) \qquad (3)$$

$\Pr(SF = 1)$ denotes the probability of an individual being self-employed or not. F is the cumulative standard normal distribution. X is the vector of explanatory variables and Ω is the vector of coefficients to be estimated. Being self-employed will be considered as a success event. $y^*$is the latent continuous metric that underlies the observed responses by the analyst. An individual's probability of being self-employed depends on an unobservable latent (utility) index I_i which, in turn, is determined by an array of explanatory variables. The models we estimate, as formulated in specifications (1) and (2) can be written as $\Pr(SF_{irs} = 1|X_{irs}) = \Phi(\beta X_{irs})$.

As stated, meaningful interpretations of probit coefficients need marginal estimates. We present the marginal estimates of our main variable of interest, education level, while explaining the results in section 4. Marginal estimates are reported sinceaverage coefficients have the potential to be biased (Webster and Piesse, 2018; Fernández-Val, 2009). Additionally, for interacted variables the joint significance of the terms matter. It is quite possible that the combination of effects is statistically significant or insignificant at different values of the considered cohorts – gender, caste, household size or others - regardless of the individual significance of either coefficient in the regression.Thus, to meaningfully conclude the probability of being self-employed for different education levels, we need to estimate the marginal impacts for different values of each cohort(Dutta and Sobel, 2021; Berry, Golder, and Milton, 2012; Brambor, Clark and Golder, 2006; Braumoeller 2004). For example, in the case of binary variables like gender, we have to estimate ($\partial SE_{irs}/\partial Edu_{irs}$)for both 0 and 1 values of the variable for each education level. We present all these marginal estimates through graphical analysis, subsequently.

4. Benchmark Results

The benchmark results are presented in Table 2. We start with binary regression considering state fixed effects and location fixed effects (whether the individual lives in a rural area or not) in column (1). Our reference level of education in all of the models is the respondents who are *not literate*. The column (2) includes the control variables like gender, caste, ageand household size.

In column (1), apart from the below primary education level, the coefficients of the other higher education levels are positive and significant. Thus, an increase in the level of education is associated with an increase in the probability of self-employment if the individual completed at least primary education. For education levels below primary education, the probability to be self-employed is actually negative and significant.

[Insert Table 2 about here]

For individuals who have primary, middle or secondary education, the probability to be self-employed is about 5.8 percent more as compared to the no formal education group. For upper secondary, the probability is about 2.6 percent. In the case of the highest education category, the probability is about 3.4 percent. The marginal estimates from column (1) are visually presented in Figure 1A. Respondents with minimal level of literacy (below primary) are about five percent less likely to be self-employed than the not literate respondents.

For the column (2) specification when all controls are included, we find that for individuals below primary education, the probability is not significant any longer. For individuals with primary, middle or secondary education, the probability to be self-employed is about 6 percent in terms of marginal effect relative to the base group which is the same as column (1) results. The results are qualitatively similar for the other groups. In Figure 1B, the marginal estimates from column (6) are presented. As evident from both figures, the probability to be self-employed goes up with education beyond primary level but then it goes down or flattens out for higher levels of education.

[Insert Figures 1A and 1B here]

In Table 3, we start presenting results based on specification (2). In column (1), we consider the specification and interact education variable with gender. The coefficients of all the education groups are significant, but vary in sign. Conversely, the coefficients of the interacted variables are all negative and statistically significant. We again present the figure for marginal effects to closely interpret the findings. In Figure 2A, ($\partial SE_{irs}/\partial Edu_{irs}$) is presented for each education group for male and female cohorts. Importantly, for individuals with below primary education, the probability of females being self-employed is higher than males. The opposite is true for those with primary, middleor secondary education. For the education groupscategorized by higher secondary and beyond college education, the probabilities for both cohorts are relatively similar – to be self-employed is between 2 to 4% higher for both males and females for both these education groups relative to their respective cohorts of not literate individuals.

[Insert Table 3 about here]

In Columns (2), (3) and (4), we present the results interacted with caste dummies. Keeping space constraint in mind and to organize the results better, we present these in three columns for a single specification. Figure 2B helps us to understand the results better. Moving from below primary education to primary, middle and secondary level,OBC and SCbenefit substantially relative to ST. Interestingly, the probability of being self-employed is the highest for all groups with primary, middle and secondary education. Also, we find that general caste fares marginally better in terms of being self-employed for all education groups relative to all other castes.

[Insert Figures 2 A and 2B about here]

In Table 4, we continue to explore specification (2). We report the impact of education on self-employment for cohorts by age and by household size in columns (1) and (2) respectively. In both columns(1) and (2), the coefficients of all the education groups are significant.

[Insert Table 4 about here]

The positive and significant coefficient of age in column (1) indicates that as an individual becomes older, the likelihood of being self-employed also increases. In Figures 3A and 3B, the marginal effects are shown for each education group at different ages and household sizes.

[Insert Figures 3A and 3B about here]

The three education groups, except those with at least a college degree, show a similar trend with the probability of self-employment increasing with age. However, for individuals with college education, the likelihood actually decreases as one becomes older. It becomes close to zero at the age of 60 years. In the case of household size, one member households, for the two highest education levels, seem to have a detrimental effect on the likelihood of being self-employed compared to the reference group. Thus, single individual households are likely to stay away from self-employment even with higher educational attainments. The opposite is true for households with two members – for such households, individuals possessing a higher secondary degree have the highest probability of being self-employed.

5. Robustness Analysis

5.1 Exploring additional cohort effects

We continue exploring our results in the context of the probability to be self-employed for additional cohorts. Presently, we consider different industries and how the likelihood of being self-employed varies by education levels. The PLFS survey provides 5-digit industry codespertaining to different industries which we use in order toaggregate the data into main industrial sectors based on National Industrial Classification 2008 (NIC-2008).[7]The main industry groups considered are agriculture, manufacturing, construction, and transportation. The remaining sectors areconsidered as the baseline group which we term as 'others'. We interact education dummy with the constructed industry dummies. The results are available in Table 5. Along with all benchmark controls, we also control for religion, discussed subsequently.

[Insert Table 5 about here]

In Figure 4A, we present the marginal estimates. Relative to no formal education, we find that being literate or having below primary education enhances likelihood of self-employment in agriculture, construction or transportation by 2 to 5 percent. This is true for 'others' as well but by a smaller magnitude. In the context of manufacturing, the likelihood of being self-employedfor higher secondary and college educated individuals goes down relative to individuals with no education. For transportation, the impact almost remains similar relative to the base group except in the case of the highest level of education where the probability falls. The likelihood of being self-employed has the greatest benefit for individuals who have at least a college degree relative to base group in the case of construction – the probability is almost 32

[7] See https://www.ncs.gov.in/Documents/NIC_Sector.pdf for more details on NIC-2008.

percent more.

Religion has been found to be important for entrepreneurship in the Indian context. Employing micro data, Audretsch, Boente and Tamvada (2013) find that while religions like Islam and Christianity are encouraging of entrepreneurship, Hinduism deters entrepreneurship. We explore if religion affects the probability of employment through the likelihood of education. We create dummies for all religions as stated in the database – Hinduism, Islam, Christianity, Buddhism, Jainism, Sikhismand others. As seen from Table 1, Hinduism constitutes about 75 percent of our sample. In comparison, Islam constitutes 13.6 percent, Christianity constitutes of about 7 percent and Sikhism constitutes about 2 percent. This is approximately the population distribution by religion in India as well, but Hindus are under-represented and Christians over-represented in the sample. The remaining 3 percent is divided among the other religions.

[Insert Table 6 about here]

The results are presented in Table 6 and the corresponding figure offers the marginal estimates as in 4B. The likelihood of self-employment remains positive for Hindus across all education groups with the magnitude being the smallest for the highest education group. Muslims are also likely to be self-employed except when they have the highest education level. The probability for Muslims to be self-employed is less than that of Hindus. The probability to be self-employed is the highest for Sikhs – individuals declaring their religion as Sikhism are almost 20 percent likely to be self-employed for all education groups relative to those with no formal education.

[Insert Figures 4A and 4B about here]

6. Summary of Results and Conclusion

The empirical analysis for India contributes to the literature by exploring the association between human capital and entrepreneurship. As mentioned above, findings of the literature are far from conclusive (Haber and Reichel, 2007; Cassar, 2006; van der Sluis et al., 2005; Bosma et al., 2004; Baum and Silverman, 2004, etc). Subsequent studies have tried to resolve the ambiguity in the findings by suggesting the relationship to be non-linear. For example, Dutta and Sobel (2018) find that human capital's effect on entrepreneurship at the macro level is contingent on access to financial capital. This paper not only re-qualifies such findings in the context of India as a novel exercise explaining self-employment on the basis of different cohorts, but clearly points out that the impact of education in non-monotonic on the choice of self-employment in India. Indeed, the effect varies across gender, caste, age, household size, religion, and industry and these add value to possible interactions between education and self-employment. To provide a structured analysis, we summarize our findings in Table 7. To be precise, we estimated the probability of being self-employed based on education levels but conditional on the cohort type. The first two rows present the likelihood to be self-employed for different education levels without considering any cohort effect. Next we present our findings for the different cohorts – gender, age, and caste. The table shows that the probability to be self-employed based on educational attainment is not only conditional on level of educational attainment but varies a lot by caste, age and gender of individuals.

For a country like India with strong group and gender identities, and often fragmented by religion and caste, our results strongly suggest that the effect of educational attainment is based on the cohort considered. Thus, the standard 'one size fits all' type entrepreneurial support schemes proposed by governments are unlikely to generate strong marginal effects. The policies shall need moderation according to the cohort effects as obtained here. The multiple schemes for micro, small, and medium enterprises (MSMEs) in India may find the above responses useful for choosing specific tax-subsidy patterns, for example. Our results clearly point out the importance of re-considering the effectiveness of support schemes to be based on cohorts (gender, age, caste etc.) when interactions with education yields precise probabilities of being self-employed.

References

Audretsch, D., Bonte, W., &Tamvada, J. 2013. Religion, social class, and entrepreneurial choice. *Journal of Business Venturing*. 28 (6), 774-789

Banerjee, A. V., & Newman, A. F. 1993. Occupational Choice and the Process of Development, *Journal of Political Economy*, 101(2), 274–298.

Baum, J. A., & Silverman, B. S. 2004. Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups. *Journal of Business Venturing*, 19(3), 411-436.

Beladi, H., & Kar, S. 2015. Skilled and Unskilled Immigrants and Entrepreneurship in a Developed Country. *Review of Development Economics*, 19 (3), 666-682

Berry, D.W., Golder, M. & Milton, D. 2012. Improving Tests of Theories Positing Interaction *The Journal of Politics*,74 (3), 1–19

Bhagavatula, S., Mudambi, R., &Murmann, J.P. 2019, Innovation and Entrepreneurship in India: An Overview, *Management and Organization Review*, 15(3), 467–493.

Blanchflower, D. 2000. Self-employment in OECD countries. *Labour Economics.* 7(5), 471-505.

Bosma, N., Van Praag, M., Thurik, R., & De Wit, G. 2004. The value of human and social capital investments for the business performance of startups. *Small Business Economics*, 23(3), 227-236.

Brambor, T., Clark, W. & Golder, M. 2006. Understanding Interaction Models: Improving Empirical Analyses. *Political Analysis*. 14 (1), 63–82.

Braumoeller, B. 2004. Hypothesis Testing and Multiplicative Interaction Terms*. International Organization*. 58 (4), 807–20.

Cassar, G. 2006. Entrepreneur opportunity costs and intended venture growth. *Journal of Business Venturing*, 21(5), 610-632.

Chowdhury, F., Audretsch, D. B., &Belitski, M. 2019. Institutions and Entrepreneurship Quality. *Entrepreneurship Theory and Practice*, 43(1), 438–456.

Dasgupta, I and Kar, S. 2018. The Labor Market in India since 1991, *IZA World of Labor: Bonn*.

De Bruin, A., Brush, C. G., & Welter, F. 2006. Introduction to the special issue: Towards building cumulative knowledge on women's entrepreneurship. *Entrepreneurship Theory and Practice*, 30(5), 585-593.

Dutta, N., Kar, S., &Beladi, H.2022. Innovation and Perceived Corruption: A Firm-Level Analysis for India. *Business and Politics,* 24(2), 151-170. doi:10.1017/bap.2021.20

Dutta, N., & Mallick, S. 2018. Enabling women entrepreneurs: Exploring factors that mitigate the negative impact of fertility rates on female entrepreneurship. *Kyklos*, 71(3), 402-432.

Dutta, N., & Sobel, R. S. 2021. Entrepreneurship, fear of failure, and economic policy. *European Journal of Political Economy*, 66, 101954.

Dutta, N., & Sobel, R. S.2018. Entrepreneurship and human capital: The role of financial development. *International Review of Economics & Finance*, 57, 319-332.

Estrin, S., & Mickiewicz, T. 2011. Institutions and female entrepreneurship. *Small Business Economics*, 37(4), 397-415.

Evans, D.S., &Leighton, L.S.1989. Some Empirical Aspects of Entrepreneurship, *American Economic Review*, 79(3), 519-535.

Fairlie, R.W., &B.D. Meyer, B.D.1996. Ethnic and Racial Self-Employment Differences and Possible Explanations, *Journal of Human Resources*, 31,757-793.

Fairlie, R. W., Morelix, A., Reedy, E. J., & Russell, J. 2016. The Kauffman index of startup activity: national trends. Kansas City: *The Ewing Marion Kauffman Foundation*

Fernández-Val, I. 2009. Fixed effects estimation of structural parameters and marginal effects in panel probit models. *Journal of Econometrics*.150(1), 71-85.

Fontana, M. 2009. The gender effects of trade liberalization in developing countries: A review of the literature. *M. Bussolo& RE De Hoyos. Gender Aspects of the Trade and Poverty Nexus. A Macro-Micro Approach*, 25-50.

Galor, O. 2012. Inequality, Human Capital Formation and the Process of Development, *IZA DP No. 6328*, IZA: Bonn.

Galor, O., &Zeira, J.1993. Income Distribution and Macroeconomics, *Review of Economic Studies*, 60 (1), 35-52.

Galor, O., Moav, O., &Vollrath, D.2009. Inequality in Landownership, the Emergence of Human-Capital Promoting Institutions, and the Great Divergence, *Review of Economic Studies*, 76(1), 143-179.

Government of United Kingdom. 2017. *Industrial clusters in England*, Department for Business, Energy & Industrial Strategy, London.

Haber, S. & Reichel, A.2007. The cumulative nature of the entrepreneurial process: The contribution of human capital, planning and environment resources to small venture performance. *Journal of Business Venturing*, 22(1), 119-145.

Hu, Z. 2021.The effect of income inequality on human capital inequality: Evidence from China, *Structural Change and Economic Dynamics*, 58, 471-489.

Hundley G. 2000. Male/Female Earnings Differences in Self-Employment: The Effects of Marriage, Children, and the Household Division of Labor. *ILR Review*. 54(1), 95-114.

Iyer, L., Khanna, T., & Varshney, A. 2013. Caste and entrepreneurship in India. *Economic and Political Weekly*, 52-60.

Koellinger, P.D. 2008. Why are some entrepreneurs more innovative than others? *Small Business Economics*, 31(1), 21-37

Krasniqi, A. B.2009. Personal, Household and Business Environmental Determinants of Entrepreneurship. *Journal of Small Business and Enterprise Development*, 16 (1), 146-166.

Lee, M. O., &Vouchilas, G. 2016. Preparing to age in place: attitudes, approaches, and actions. *Housing and Society*. 43(6), 69–81. https://doi.org/10.1080/08882746.2016.1221039

Lévesque, M., &Minniti, M. 2006. The effect of aging on entrepreneurial behavior. *Journal of Business Venturing*, 21, 177–194. https://doi.org/10.1016/j.jbusvent.2005.04.003.

Manso, G. Motivating Innovation. 2011. *The Journal of Finance*. 66(5), 1823-1860

Marjit, S and Kar, S. 2011. *The Outsiders: Economic Reform and Informal Labor in a Developing Economy*. Oxford University Press: New Delhi, London.

Martin, B.C., McNally, J.J., & Kay, M.J. 2013. Examining the formation of human capital in entrepreneurship: A meta-analysis of entrepreneurship education outcomes. *Journal of Business Venturing*, 28(2), 211-224

Marvel, M.R. 2013. Human capital and search-based discovery: A study of high-tech entrepreneurship. *Entrepreneurship Theory and Practice*, 37(2), 403–419.

Marvel, M. R., Davis, J. L., & Sproul, C. R. 2016. Human Capital and Entrepreneurship Research: A Critical Review and Future Directions. *Entrepreneurship Theory and Practice*, 40(3), 599–626.

Mentoor, E. R., & Friedrich, C. 2007. Is Entrepreneurial Education at South African Universities Successful?: An Empirical Example. *Industry and Higher Education*, 21(3), 221–232.

Monsen, E., Mahagaonkar, P., & Dienes, C.2012. Entrepreneurship in India: The Question of Occupational Transition, *Small Business Economics*, 39(2), 359–382.

Osterbeek, H., Praag, M., &Ijsselstein, A. 2010. The impact of entrepreneurship education on entrepreneurship skills and motivation. *European Economic Review*, 54(3), 442-454

Parker, S. C. 2004. *The economics of entrepreneurship*. Cambridge: Cambridge University Press. https://doi. org/10.1017/CBO9780511817441.

Sahoo, S., & Klasen, S. 2021. Gender Segregation in Education: Evidence from Higher Secondary Stream Choice in India. *Demography*, 58(3), 987–1010

Simoes, N, Crespo, N. &Moreira, S. 2016. Individual Determinants of Self-Employment Entry: What Do We Really Know? *Journal of Economic Surveys*, 30, 4, 783-806.

Spence, M. 1973. Job market signaling, *Quarterly Journal of Economics*, 87, 355-374.

Swamy, A., Knack, S., Lee, Y. & Azfa, O. 2001. Gender and corruption. *Journal of Development Economics*. 64(1), 25-55.

Unger, J.M., Rauch, A., Frese, M., & Rosenbusch, N.2011. Human capital and entrepreneurial success: A meta-analytical review, *Journal of Business Venturing*, 26(3), 341–358.

van der Sluis, J., &van Praag, M. 2007. *Human capital and entrepreneurs*, in J. Hartog and H. Van den Brink (eds.) Human Capital, Cambridge: Cambridge University Press.

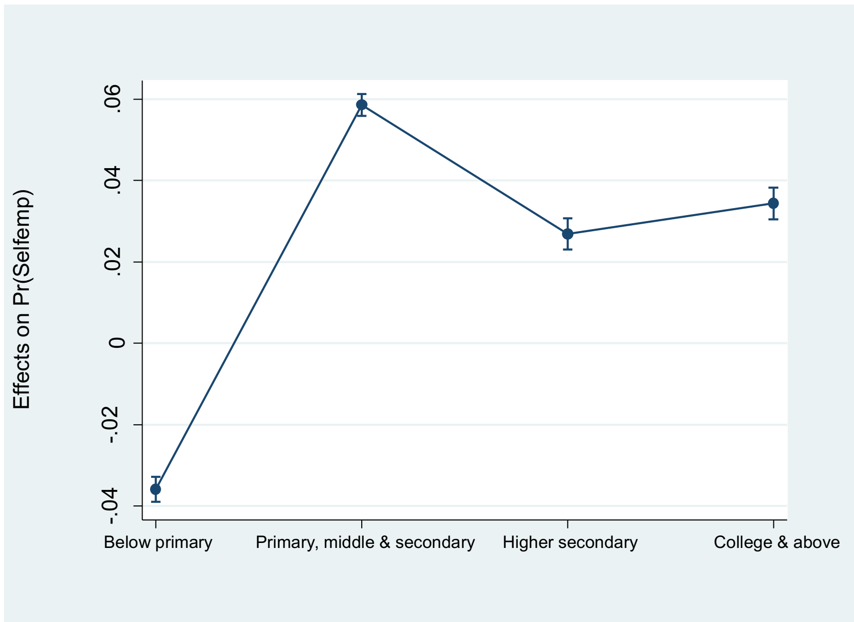
van der Sluis, J., Van Praag, M., &Vijverberg, W. 2005. Entrepreneurship selection and performance: A meta-analysis of the impact of education in developing economies. *The World Bank Economic Review*, 19(2), 225-261.

Webster, A. & Piesse. J. 2018. Are Foreign-Owned Firms More Likely to Pay Bribes than Domestic Ones? Evidence from Emerging Markets. *World Development*. 102, 142-161

Yoshino, N., &Taghizadeh-Hesary, F. 2018. The Role of SMEs in Asia and Their Difficulties in Accessing Finance. *ADBI Working Paper 911*. Tokyo: Asian Development Bank Institute.

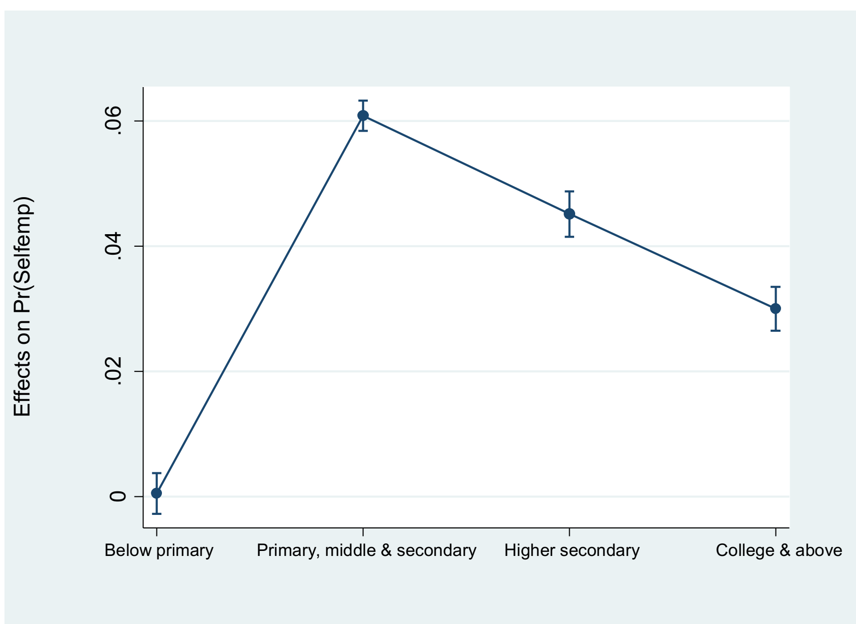
Zhang, T. &Acs, Z. 2018.Age and entrepreneurship: nuances from entrepreneur types and generation effects. *Small Business Economics.* 51, 773-809.

Figure 1A: Marginal Estimates: Self-employment and Education Levels



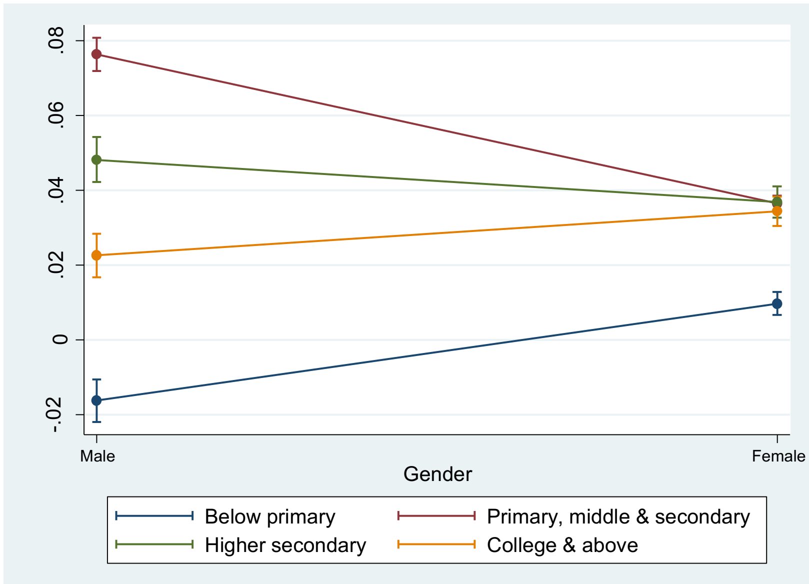
Source: Authors' calculations

Figure 1B: Marginal Estimates: Self-employment and Education Levels



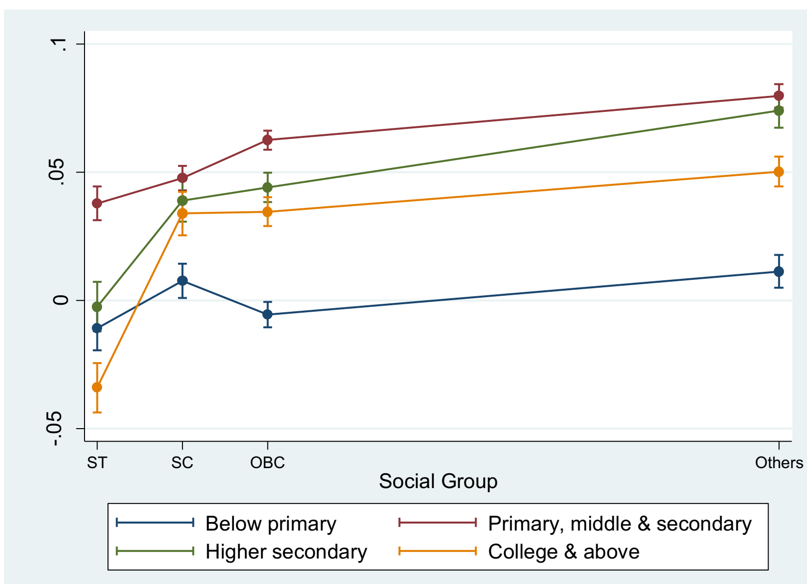
Source: Authors' calculations

Figure 2A: Marginal Estimates: Self-employment, Education Levels and Gender



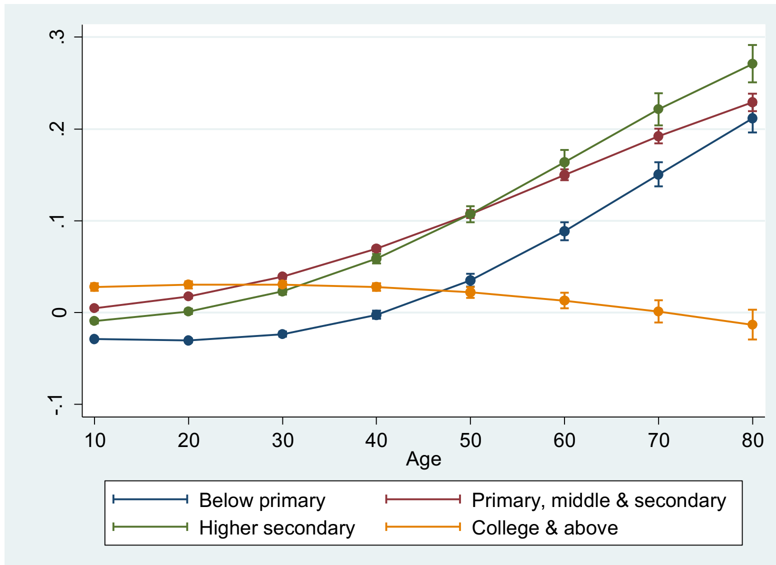
Source: Authors' calculations

Figure 2B: Marginal Estimates: Self-employment, Education Levels and Caste



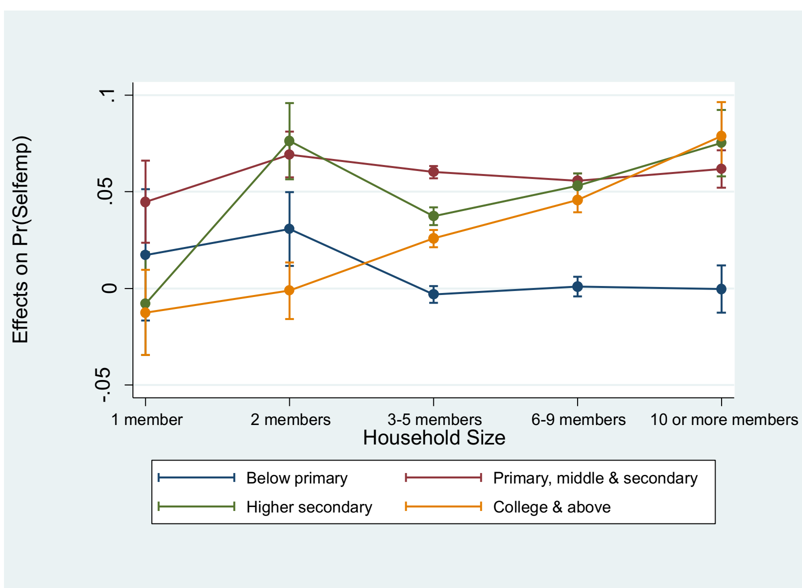
Source: Authors' calculations

Figure 3A: Marginal Estimates: Self-employment, Education Levels and Age



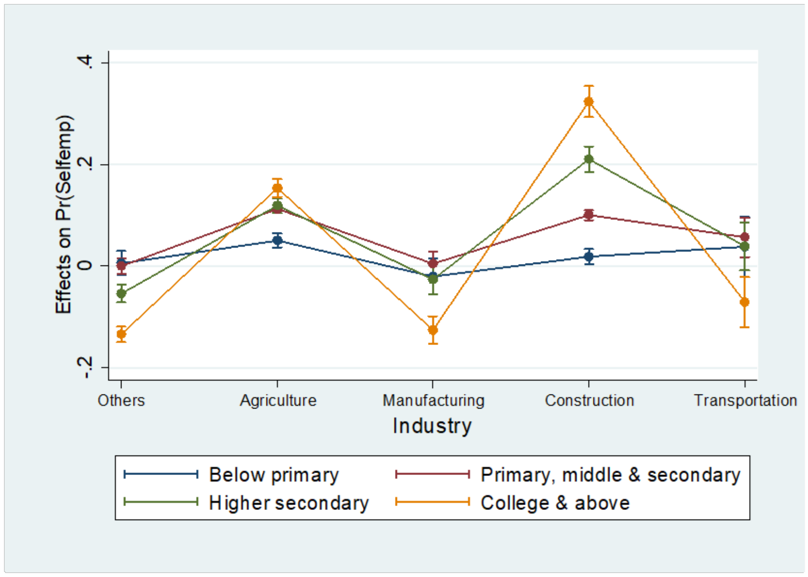
Source: Authors' calculations

Figure 3B: Marginal Estimates: Self-employment, Education Levels and Household Size



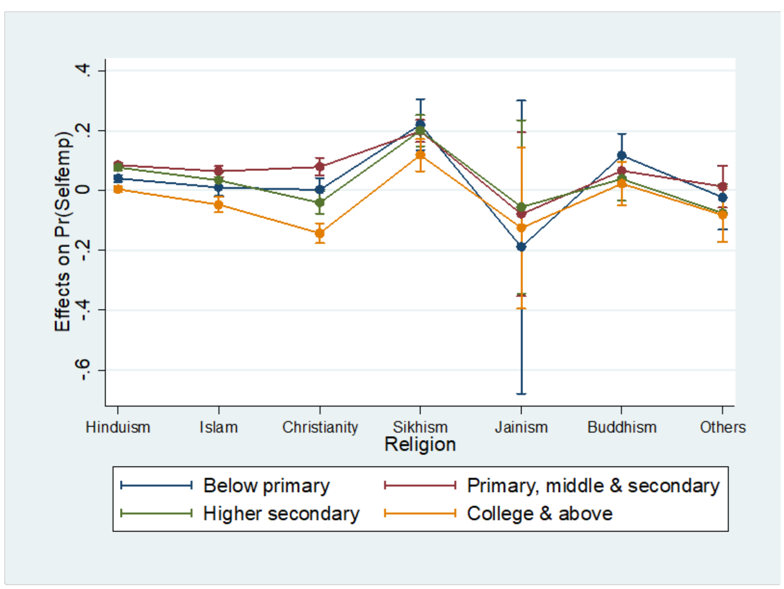
Source: Authors' calculations

Figure 4A: Marginal Estimates: Self-employment, Education Levels and Industry



Source: Authors' calculations

Figure 4B: Marginal Estimates: Self-employment, Education Levels and Religion



Source: Authors' calculations

Table 1: Summary statistics of variables

Variable	Observations	Mean	Std. Dev.	Min	Max
Self-Emp.	418297	0.142	0.349	0	1
Edu dummy	418297	1.732	1.231	0	4
Female	418297	0.49	0.5	0	1
Age	418297	30.994	19.568	0	110
Rural	418297	0.574	0.494	0	1
Household Size	418297	5.035	2.116	1	24

Social Group	Observations	Percent
ST	57,576	13.76
SC	73,980	17.69
OBC	1,65,903	39.66
Others	1,20,838	28.89
Total	4,18,297	100

Religion	Observations	Percent
Hinduism	3,14,615	75.21
Islam	57,221	13.68
Christianity	29,419	7.03
Sikhism	8,872	2.12
Jainism	890	0.21
Buddhism	4,019	0.96
Zoroastrianism	113	0.03
Others	3,148	0.75
Total	4,18,297	100

Education	Frequency	Percent
Not Literate	96,672	23.11
Below Primary	52,098	12.45
Primary, Middle & Secondary	1,81,746	43.45
Higher Secondary	42,077	10.06
College & above	45,704	10.93
Total	4,18,297	100

Source: Authors' calculations

Table 2: Probit Specifications: Self-employment and Education Levels

The dependent variable is a dummy indicating if an individual is self-employed or not. Not literate is the base group for education ordered dummy. *Below primary* indicates those individuals who are literate but have not attended primary schooling. *Prim, Midd, Sec* indicates educational attainment upto secondary. *Higher Sec* represents above secondary till higher secondary education. *College & abv* indicates undergraduate and beyond education. The controls are gender, caste, age, and household size. We control for location and state fixed effects. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

	(1)	(2)
Below Primary	-0.201***	0.003
	(0.009)	(0.011)
Prim, Midd, Sec	0.281***	0.357***
	(0.006)	(0.007)
Higher Sec	0.167***	0.275***
	(0.009)	(0.010)
College & abv	0.238***	0.190***
	(0.009)	(0.011)
Female	---	-1.009***
		(0.006)
SC	---	-0.137***
		(0.011)
OBC	---	0.072***
		(0.010)
Gen. caste	---	-0.002
		(0.010)
Age	---	0.027***
		(0.0001)
Household size	---	-0.014***
		(0.001)
Location F.E.	Yes	Yes
State F.E.	Yes	Yes
Constant	-1.338***	-1.977***
	(0.0132)	(0.0212)
Observations	418,297	418,297

Source: Authors' calculations

Table 3: Probit Specifications: Self-employment, Education Levels, Gender and, Caste

The dependent variable is a dummy indicating if an individual is self-employed or not. Not literate is the base group for education ordered dummy. *Bel. Prim.* indicates those individuals who are literate but have not attended primary schooling. *Pri, Mid, Sec* indicates educational attainment up to secondary. *High Sec* represents above secondary till higher secondary education. *Col & abv.* indicates undergraduate and beyond education. In Column (1), we interact female with education groups. In columns (2), (3) and (4), we interact caste dummies with education groups. We control for location and state fixed effects. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

	(1) Female	(2) SC	(3) OBC	(4) Others
Bel. Prim.	-0.073***	-0.065**	-0.065**	-0.065**
	(0.013)	(0.0266)	(0.026)	(0.026)
Pri, Mid, Sec	0.302***	0.203***	0.203***	0.203***
	(0.009)	(0.0187)	(0.0187)	(0.018)
High Sec	0.197***	-0.0142	-0.0142	-0.014
	(0.012)	(0.028)	(0.028)	(0.028)
Col & abv.	0.095***	-0.219***	-0.219***	-0.219***
	(0.012)	(0.033)	(0.033)	(0.033)
Bel.Prim*Cohort	-0.999***	0.125***	0.0279	0.148***
	(0.020)	(0.037)	(0.031)	(0.035)
Pri, Mid, Sec*Cohort	-0.712***	0.126***	0.151***	0.278***
	(0.0112)	(0.025)	(0.021)	(0.024)
High Sec*Cohort	-0.709***	0.289***	0.273***	0.466***
	(0.019)	(0.039)	(0.032)	(0.035)
Col &abv.*Cohort	-0.731***	0.463***	0.428***	0.544***
	(0.0193)	(0.044)	(0.037)	(0.038)
Fem	-1.142***	-1.010***	-1.010***	-1.010***
	(0.0134)	(0.006)	(0.006)	(0.006)
SC	-0.136***	-0.264***	-0.264***	-0.264***
	(0.011)	(0.021)	(0.021)	(0.021)
OBC	0.073***	-0.054***	-0.054***	-0.054***
	(0.0103)	(0.018)	(0.018)	(0.018)
Gen. Caste	-0.003	-0.248***	-0.248***	-0.248***
	(0.010)	(0.021)	(0.021)	(0.021)
Age	0.027***	0.027***	0.027***	0.027***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Household Size	-0.014***	-0.013***	-0.013***	-0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
Location F.E.	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes
Constant	-1.193***	-0.759***	-0.759***	-0.759***
	(0.0132)	(0.0164)	(0.0164)	(0.0164)
Observations	418,297	418,297	418,297	418,297

Source: Authors' calculations

Table 4: Probit Specifications: Self-employment, Education Levels, Age and,Household Size
The dependent variable is a dummy indicating if an individual is self-employed or not. Not literate is the base group for education ordered dummy. *Bel. Prim.* indicates those individuals who are literate but have not attended primary schooling. *Pri, Mid, Sec* indicates educational attainment up to secondary. *High Sec* represents above secondary till higher secondary education. *Col & abv.* indicates undergraduate and beyond education. In Column (1), we interact age with education groups. In columns (2), (3), (4) and (5), we interact household size dummies with education groups. The columns mention the size of the households. We control for location and state fixed effects. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

	Age	2 members	3-5 members	6-9 members	>= 10 members
Bel. Prim.	-0.649***	0.099	0.099	0.099	0.099
	(0.016)	(0.097)	(0.097)	(0.097)	(0.097)
Pri, Mid, Sec.	-0.038***	0.241***	0.241***	0.241***	0.241***
	(0.012)	(0.059)	(0.059)	(0.059)	(0.059)
High Sec.	-0.254***	-0.049	-0.049	-0.049	-0.049
	(0.022)	(0.085)	(0.085)	(0.085)	(0.085)
Col. & abv.	0.316***	-0.078	-0.078	-0.078	-0.078
	(-0.022)	(0.0708)	(0.0708)	(0.0708)	(0.0708)
Below Prim. * Coh.	0.016***	0.051	-0.120	-0.092	-0.102
	(0.000)	(0.108)	(0.098)	(0.099)	(0.115)
Pri, Mid, Sec. * Coh.	0.009***	0.080	0.099*	0.129**	0.225***
	(0.000)	(0.066)	(0.060)	(0.061)	(0.071)
High Sec. * Coh.	0.013***	0.399***	0.270***	0.404***	0.595***
	(0.001)	(0.096)	(0.086)	(0.087)	(0.102)
Col. & abv. * Coh.	-0.004***	0.0721	0.236***	0.391***	0.645***
	(0.001)	(0.080)	(0.072)	(0.074)	(0.091)
Age	0.022***	---	---	---	---
	(0.000)				
2 members	0.158***	---	---	---	---
	(0.026)				
3-5 members	0.325***	---	---	---	---
	(0.024)				
6-9 members	0.202***	---	---	---	---
	(0.024)				
10 or more members	0.170***	---	---	---	---
	(0.028)				
Controls	Yes	Yes	Yes	Yes	Yes
Location F.E.	Yes	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes	Yes
Constant	-2.044***	-2.227***	-2.227***	-2.227***	-2.227***
	(0.031)	(0.050)	(0.050)	(0.050)	(0.050)
Observations	418,297	418,297	418,297	418,297	418,297

Source: Authors' calculations

Table 5: Probit Specifications: Self-employment, Education Levels, and Industry

The dependent variable is a dummy indicating if an individual is self-employed or not. Not literate is the base group for education ordered dummy. *Bel. Prim.* indicates those individuals who are literate but have not attended primary schooling. *Pri, Mid, Sec* indicates educational attainment up to secondary. *High Sec* represents above secondary till higher secondary education. *Col & abv.* indicates undergraduate and beyond education. We interact education groups with industry dummies. We include all industry dummies in one specification – we present the same specification in different columns keeping space constraint in mind. We control for location and state fixed effects. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

	(1) Agriculture	(2) Manufacturing	(3) Construction	(4) Transportation
Bel. Prim.	0.017	0.017	0.017	0.017
	(0.036)	(0.036)	(0.036)	(0.036)
Pri, Mid, Sec	-0.002	-0.002	-0.002	-0.002
	(0.023)	(0.023)	(0.023)	(0.023)
High Sec	-0.165***	-0.165***	-0.165***	-0.165***
	(0.027)	(0.027)	(0.027)	(0.027)
Col & abv.	-0.429***	-0.429***	-0.429***	-0.429***
	(0.024)	(0.024)	(0.024)	(0.024)
Bel.Prim *Ind.	0.131***	-0.078	0.148**	0.090
	(0.041)	(0.065)	(0.073)	(0.092)
Pri, Mid, Sec * Ind.	0.332***	0.0152	0.647***	0.161***
	(0.0257)	(0.041)	(0.044)	(0.061)
High Sec * Ind.	0.514***	0.087*	1.245***	0.274***
	(0.034)	(0.052)	(0.060)	(0.073)
Col & abv. * Ind.	0.883***	0.036	1.871***	0.224***
	(0.036)	(0.049)	(0.061)	(0.077)
Controls	Yes	Yes	Yes	Yes
Location F.E.	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes
Constant	-1.193***	-0.759***	-0.759***	-0.759***
	-0.013	-0.016	-0.016	-0.016
Observations	418,297	418,297	418,297	418,297

Source: Authors' calculations

Table 6: Probit Specifications: Self-Employment, Education and Religion

The dependent variable is a dummy indicating if an individual is self-employed or not. Not literate is the base group for education ordered dummy. *Bel. Prim.* indicates those individuals who are literate but have not attended primary schooling. *Pri, Mid, Sec* indicates educational attainment up to secondary. *High Sec* represents above secondary till higher secondary education. *Col & abv.* indicates undergraduate and beyond education. We interact education with different religious groups. We control for gender, caste, age, household size, relation to the household held and industry. Keeping space constraint in mind, we do not report the controls but they are available on request. We control for location and state fixed effects. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

	(1) Islam	(2) Christianity	(3) Sikhism	(4) Jainism	(5) Buddhism	(6) Others
Bel. Prim.	0.128***	0.128***	0.128***	0.128***	0.128***	0.128***
	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)	(0.018)
Pri, Mid, Sec	0.268***	0.268***	0.268***	0.268***	0.268***	0.268***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
High Sec	0.247***	0.247***	0.247***	0.247***	0.247***	0.247***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Col & abv.	0.0124	0.0124	0.0124	0.0124	0.0124	0.0124
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Bel.Prim*Cohort	-0.097**	-0.121*	0.623***	-0.680	0.281**	-0.202
	(0.049)	(0.067)	(0.141)	(0.737)	(0.131)	(0.168)
Pri, Mid, Sec * Cohort	-0.075**	-0.0235	0.417***	-0.501	-0.0271	-0.227**
	(0.030)	(0.048)	(0.072)	(0.418)	(0.097)	(0.110)
High Sec * Cohort	-0.144***	-0.376***	0.442***	-0.411	-0.102	-0.485***
	(0.045)	(0.062)	(0.093)	(0.443)	(0.140)	(0.156)
Col &abv.* Cohort	-0.160***	-0.498***	0.414***	-0.377	0.0725	-0.270*
	(0.042)	(0.059)	(0.097)	(0.412)	(0.138)	(0.143)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Location F.E.	Yes	Yes	Yes	Yes	Yes	Yes
State F.E.	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-2.128***	-2.128***	-2.128***	-2.128***	-2.128***	-2.128***
	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)	(0.040)
Observations	151,934	151,934	151,934	151,934	151,934	151,934

Source: Authors' calculations

Table 7: Summarizing the findings

Table 7 summarizes the main findings in terms of probability of self-employment for educational attainment, and the different cohorts. The education groups are - Group 1 - Below Primary Education; Group 2 - Primary Middle Secondary; Group 3 - Above Secondary.

Cohort	Main Variable – Education Groups	Probability to be Self-employed (SE)
NA	Group 1	Falls
NA	Group 2	Rises
NA	Group 3	Rises but by a lesser magnitude than group 2
Gender	Group 1	Females more likely to be SE than males
	Group 2	Males more likely to be SE than females
	Group 3	Both equally likely
Caste	Group 1	General Caste fares marginally better than SC, ST, and OBC
	Group 2	General Caste benefits the most; SC and OBC benefits more than ST. Interestingly group 2 education benefits all groups the most.
	Groups 3	SC and OBC benefits substantially more compared to ST with group 3 education. General caste benefits marginally more relative to SC and OBC.
Age	Group 1	Rises
	Group 2	Rises
	Group 3	Rises for above college but falls for secondary education

Source: Authors' calculations