Education, social identity and self-employment over time: evidence from a developing country

Jagannadha Pawan Tamvada (ID) ·
Mili Shrivastava · Tapas Kumar Mishra

Abstract While the extant literature suggests that an individual's occupational choice depends on the level of economic development, little is known about how the dynamics of education and social identity impact self-employment choice over time in a developing country setting. The pseudo-panel regression results presented here suggest that higher education decreases the likelihood of individuals' choice of self-employment over time in India, but increases it in most developed cohorts. Such transitional dynamics are more prominent amongst individuals who belong to historically determined backward classes. Furthermore, the results suggest sectoral differences in self-employment choice—individuals with higher education and wealth have a greater likelihood of transiting into self-employment over time in agriculture while similar individuals in non-agriculture exit self-employment. In both sectors, however, education moderates the impact of social identity on self-employment choice. The paper underscores the need for policy shifts towards entrepreneurship along the path of economic development.

Plain English Summary University education helps overcome social barriers to self-employment in India. While it reduces the transition into self-employment over time, educated individuals are more likely to move back into self-employment in the most developed cohorts. We examine the occupational dynamics over ten years using large-scale databases by tracking self-employment in cohorts of individuals across India. The results suggest that university education enables individuals from socially backward classes to overcome barriers to self-employment in both non-agricultural and agricultural sectors. Although university education decreases the likelihood of self-employment over time, it increases self-employment in the most developed cohorts. The paper suggests that public policy should focus on facilitating industrial development in the initial stages of development as individuals are motivated to invest in professional human capital during these periods while encouraging entrepreneurial activities when a more developed economy can provide opportunities for greater returns in self-employment.

Keywords Social class · Self-employment choice · Entrepreneurship · India · Caste · Pseudo-panels

JEL Classification J24 · L26 · Z13

J. P. Tamvada (✉) · T. K. Mishra
Southampton Business School, University of Southampton, Southampton, SO17 1BJ, UK
e-mail: jp.tamvada@soton.ac.uk

T. K. Mishra
e-mail: T.K.Mishra@soton.ac.uk

M. Shrivastava
Bournemouth University Business School, Bournemouth University, BH12 5BB, Bournemouth, UK
e-mail: mshrivastava@bournemouth.ac.uk

1 Introduction

Although an emerging body of literature examines the determinants of self-employment and the push-pull aspects of self-employment choice in developing countries (Tamvada 2021; Audretsch et al., 2013; Kim & Cho, 2009; Earle & Sakova, 2000), little is known about the dynamics of self-employment in such contexts. Beginning with Kuznets (1966), a compelling body of theoretical literature suggests that self-employment and stages of economic development are related inversely. Lucas (1978) predicts that entrepreneurship decreases with economic development. Iyigun and Owen (1999) claim that as an economy develops, individuals invest time in accumulating professional skills through education than accumulating entrepreneurial human capital. Several empirical studies based on aggregated databases and cross-national studies confirm a negative association between the rate of self-employment and the stage of economic development (Pietrobelli et al., 2004; Folster 2002; Acs et al., 1994).[1]

However, analyses based on aggregate data provide a partial picture of the dynamics and distributional changes in self-employment choice. They do not capture the inherent dynamics which, in Indian context for instance, are driven by both the momentum of growth and a historically driven caste divide. There is a visible interplay of both *social and economic dynamics* in the entrepreneurial choice functions of individuals through self-employment, but micro analyses of self-employment dynamics in developing countries are sparse in the extant literature mainly because of the non-availability of panel data.[2] For overcoming this challenge, the paper offers a novel empirical approach to examining self-employment dynamics by constructing pseudo panels of cohorts of individuals and tracking them over time. These pseudo panels are constructed using repeated cross-sectional individual-level microdata (Deaton, 1985; Verbeek, 2006), and provide new evidence on the dynamics of self-employment choice using individual-level microdata from India.[3]

The paper aims to examine the following: (1) The role of university education for self-employment over time in a developing country, (2) the moderating role of development on the impact of education on self-employment over time, (3) the role of social class for self-employment over time[4] and (4) the moderating role of education on the impact of social class on self-employment over time. In addition to the above, the paper examines how these effects vary across (a) non-agriculture and agriculture sectors and (b) rural and urban areas.

This paper makes several novel contributions to literature on occupational choice. Firstly, it offers novel insights on the role of social class for self-employment choice along different stages of economic development, a topic that that has rarely received attention in the literature. In particular, it advances the seminal contribution of Audretsch et al. (2013) by demonstrating that higher education moderates the negative impact of social class on self-employment choice over time. Secondly, it empirically tests the theoretical claim of Iyigun and Owen (1999) that in the early stages of economic development individuals invest in professional human capital and not in entrepreneurial abilities. Thirdly, it contributes to an emerging body of literature on self-employment in developing countries (Tamvada 2021; Tamvada 2015; Audretsch et al., 2013; Mohapatra et al., 2007) by examining the dynamics of self-employment activity over time. It offers a unique approach to examine the dynamics of self-employment using pseudo panels to overcome limitations posed by the non-availability of panel data.

The pseudo-panels are constructed from three cross-sectional databases collected by India's National

---

[1]Pietrobelli et al., (2004) also present evidence of exceptions suggesting that in some cases self-employment is related to high value-added manufactured exports, representing a dynamic and emerging form of entrepreneurship.

[2]Although developing countries collect large-scale cross-sectional data through household surveys, the non-availability of panel data poses a significant barrier to examining the dynamics of self-employment choice over time. There are a few exceptions, however. See, for instance, Mohapatra et al. (2007) for an analysis of the changing nature of self-employment in China.

[3]India stands as one of the countries with the largest youth population with nearly 62% belonging to the working age group (15–59 years). With almost 54% of its total population below 25 years of age, it is slated to become the world's youngest nation by 2022. This demographic dividend not only provides a huge reservoir of manpower, but also draws focus towards making this talent pool employable. As one of the countries with currently the highest youth population in the world, India faces its greatest opportunity and challenge.

[4]We refer to social class interchangeably with belonging to backward social classes of the caste system in India.

Sample Survey Organisation (NSSO) during 1994–1995, 1999–2000, and 2004. The paper provides first insights into the role of social class for self-employment over time while providing empirical support to a compelling theoretical literature on the evolution of self-employment. The results partially support the claim of Iyigun and Owen (1999) that as an economy develops individuals prefer to invest in professional human capital instead of entrepreneurial human capital. The results suggest that individuals who acquire education and wealth are less likely to transition into self-employment over time in non-agriculture while similar individuals in agriculture are more likely to transition into self-employment over time. Furthermore, backward class individuals are less likely to transition into self-employment in both the sectors but education moderates this relationship. Although University education decreases the likelihood of self-employment over time, it increases self-employment in the most developed cohorts.

The remainder of the article is structured as follows. Section 2 presents the theoretical background for the study. In Section 3, the dataset and descriptive statistics are presented. In the fourth section, the empirical results relating to self-employment choice and the role of social class over time are discussed. The limitations and conclusions are presented in the final section.

2 Human capital, social identities, and self-employment over economic development stages

Amongst early researchers, Lucas (1978) predicts that entrepreneurship decreases with economic development. Iyigun and Owen (1999, pp. 213–215) suggest that "entrepreneurial human capital plays an important role in intermediate income countries, whereas professional human capital is relatively more important in richer economies." Under the assumption that entrepreneurship is riskier than providing professional services they show that individuals begin to invest more time in accumulating professional skills by way of education than in accumulating entrepreneurial human capital as an economy develops. In their words, "[a]s per capita income grows and the payoff to being a professional increases, individuals are less willing to gamble on entrepreneurial ventures . . . as the return to the safe activity increases and the payoffs to the risky

activity becomes more variable, human capital accumulators devote more time to schooling and less time to gaining entrepreneurial experience." This claim is tested here by tracking the occupational behaviour of individuals with greater levels of human capital over time.

Furthermore, irrespective of the social apparatus an individual faces, it is rational to assume that inherent in human nature is a strong desire to grow and prosper—preferably through less risk and greater stability. As an illustration, consider two states of economic environment, S_1 (a sustained period of economic growth volatility) and S_2 (a prospective period of stability and high growth momentum). Leaving aside any complex interplay of social dynamics, individuals may wish to experiment with a new business idea because of persistent unemployment and laggard growth during the initial periods of development. Self-employment is risky as the probability of a failure in the new venture is greater than the probability of success. However, in the absence of any certain and constant economic opportunity, individuals are more inclined to be self-employed rather than loosing time by a constant employment search if the probability of success in employment search is assumedly less than the probability of success in self-employment. As the economy progresses towards stability and transitions into S_2, they are more likely to be inclined towards securing a stable job, for which they will be ready to devote time and resources for professional human capital development. Thus, investment in entrepreneurial human capital, a conventionally a risky investment strategy for individuals, is likely to be less preferred when compared to undertaking training for professional human capital and subsequent stable income generation as economy transitions into S_2.

Educated individuals have superior opportunities in salaried employment yet they are more likely to be able to identify entrepreneurial opportunities. The empirical studies examining the role of education for self-employment offer support for both views and are inconclusive (Van der Sluis et al., 2005; Van der Sluis et al., 2008). It is likely that the U-shaped relationship between education and self-employment (Åstebro et al., 2011; Blanchflower, 2000) can explain these diverse effects. Both low levels of education and high levels of education can increase self-employment. Individuals with low levels of education are more likely to be pushed into

self-employment while individuals with high levels of education are likely to be pulled into it. Education and human capital are linked to entrepreneurial success (Kolstad and Wiig, 2015; Unger et al., 2011) and returns to entrepreneurial activity (Van Praag et al., 2013; Hamilton, 2000). While pull factors such as high returns to capital can motivate individuals to enter self-employment and entrepreneurship, factors such as barriers to desirable salaried jobs can push individuals into self-employment as an occupation of last resort (Earle & Sakova, 2000). There is significant heterogeneity in informal enterprises in developing countries with the informal sector consisting of both upper-tier and lower-tier enterprises (Cunningham & Maloney, 2001). The upper-tier consists of competitive enterprises while the lower-tier consists of enterprises created by individuals rationed out of formal labour market (Fields, 2004). While in some other countries like Cote-d'Ivoire the informal sector consists of both voluntary and involuntary segments (Günther & Launov, 2012), there is increasing evidence that in some developing countries such as Ghana pull factors are attracting skilled workers into self-employment (Falco & Haywood, 2016). In more developed countries, self-employed individuals are likely to be more educated than self-employed individuals in less developed countries (van Stel & van der Zwan, 2020).

Some studies highlight that own-account self-employed individuals have characteristics that are consistent with the dualistic labour market hypothesis while self-employed individuals with employees resemble opportunity entrepreneurs of industrialized countries (Mandelman & Montes-Rojas, 2009). Tamvada (2010) examines the heterogeneity in the labour market in the Indian context and finds a hierarchy in returns to occupations with employers having greater returns than salaried individuals who in turn have greater returns than self-employed individuals across the welfare distribution. Using income data, Gindling and Newhouse (2014) examine the heterogeneity in self-employment in 74 developing countries, and find that occupations have a pecking order similar to Tamvada (2010). They further find that as the economy develops individuals move from agriculture into unsuccessful non-agricultural self-employment, and from there into wage employment in non-agriculture or successful self-employment. Furthermore, a significant proportion unsuccessful self-employed

individuals have characteristics similar to successful self-employed individuals suggesting that they experience barriers to their growth.

Margolis (2014) finds that social protection systems allow individuals to avoid subsistence self-employment while labour market frictions increase self-employment rates. He posits that about one-third self-employed individuals in the developing world are opportunity entrepreneurs who take initiative and risk. He suggests that self-employment rates decrease as countries develop and their institutions evolve, and wage employment becomes the main source of jobs. While self-employment is known to decline with economic development (Kuznets, 1966), it may increase in later stages of development, as in the case of OECD countries (Acs et al., 1994). Pietrobelli et al. (2004) examine data from 64 developing countries and 19 developed countries over 30 years to confirm the Kuznets hypothesis that self-employment tends to decline with economic development. However, they find that it can represent emerging opportunity entrepreneurship that boosts the development process in some contexts. Consistent with this, Mohapatra et al. (2007) find that self-employment in rural China has characteristics that are similar with productive small business sector. Wennekers et al. (2002) examine self-employment data between 1899–1997 and find that self-employment rates continuously declined until about early 1980s in Netherlands before experiencing a revival in the following period. Fritsch et al. (2015) find that self-employment has increased in Germany following the German reunification, and demographic developments, shifts to service sector employment, and increased university education have contributed to the increases in self-employment.

With respect to the role of education on self-employment, there is little consensus in the extant empirical literature (Premand et al. 2016; Van der Sluis et al. 2005, 2008).[5] While education expands an individual's knowledge base and increases exposure

---

[5] A vast literature has examined the role of education for self-employment. Some of the earliest studies on self-employment have originated this debate. These include, for example, Rees and Shaw (1986), Taylor (1996), Blanchflower and Meyer (1994) and Blanchflower (2000) find positive effects of education on self-employment; Evans and Leighton (1989b) and Evans and Jovanovic (1989) find no significant effects; and Blanchflower et al. (2001) find negative effects of education on the probability of selecting self-employment.

to new opportunities, it also increases the opportunity cost of being self-employed. This suggests that returns to salaried employment increase faster than returns to entrepreneurship as per capita income grows with the result being that individuals have "more to lose" by engaging in entrepreneurship (Lucas, 1978) but may eventually come back to self-employment when they are likely to get higher returns from self-employment during later stages of development. Thus, there are compelling reasons to posit that individuals who are more educated will opt for salaried employment instead of self-employment over time. These lead to the following hypotheses:

> **Hypothesis 1a:** University education decreases the likelihood of transitioning into self-employment over time.
> **Hypothesis 1b:** Economic development positively moderates the role of university education for self-employment over time.

The potential role of social identity for self-employment choice dynamics is an open-ended question. Concomitantly, entrepreneurship, for instance, a start-up type enterprise in a small scale, requires a definite clarity of success in the venture, especially when the economy is at a taking-off stage. Yet, questions may arise on the role of social gradients that may limit entrepreneurial ambition and the propagation of self-employment. Entrepreneurship literature has noted that occupational choice is influenced by individuals' social identities (Audretsch et al., 2013).

An individual's self-worth is influenced by their location in the social hierarchy (Foucault, 1982; Bourdieu, 1987; Goel and Deshpande, 2016). In line with the social dominance theory, individuals lower in the social hierarchy are likely to be dominated by those in the upper in the hierarchy (Sidanius et al., 1992; Sidanius et al., 1994). In such cases, symbolic interactionism theory suggests that the social interactions between individuals will influence their behavioural choices (Mead, 1934; Blumer, 1986). Consistent with this theory, individuals from socially backward classes may find their social identity to be a limiting factor for their entrepreneurial choice. For these reasons, the caste system in India, as a historically persistent phenomenon, has multi-layered implications for employment choice along with regional and educational opportunities. It is a rigid form of the social class structure that does not depend on the colour of the skin (Deshpande, 2005) but shares many qualities of social stratification found in other parts of the world (Donoghue, 1957; Berreman, 1960).

The caste divide in India is a cultural and social phenomenon that has been a part of the traditions of the Indian society for thousands of years. In the sociological discourse on the Indian caste system, some groups are referred to as lower castes or backward classes (Revankar, 1971; Srinivas et al. 1962). In line with the social dominance theory (Sidanius et al., 1992; Sidanius et al., 1994; Sidanius & Pratto, 2001), if backward class individuals believe that it is socially unacceptable for them to start businesses, they may not enter self-employment leading to caste-based differences in business ownership.

Furthermore, this may lead to strong social boundaries (Kim & Aldrich, 2005) that limit access to resources. The institutional profile associated with the caste system may not support individuals from backward classes becoming self-employed (Audretsch et al., 2013), and may limit access to finance and information (Bönte & Filipiak, 2012). Gupta et al. (2018), amongst others, study the extent caste systems impacts individuals' social identity and their consumption levels. Many authors also note that it is a major obstacle to achieving development goals because affected populations can get excluded from the development process. For instance, using NSS data for the periods 1983–1999, Kijima (2006) has reported that the disparities of living standards amongst marginalized backward social classes when compared to the non-marginalied backward social classes still remain very high. Despite significant governmental efforts since independence, these marginalized groups have low occupational mobility and are engaged mostly as agricultural labourers or in self-employed agricultural work (Gang et al., 2017) suggesting significant social inequality between high and low castes. As a consequence, individuals from backward social classes have lower mobility in the labour market and report higher poverty level and income inequality (Desai and Dubey, 2012).

Using social distance theory, Gupta et al. (2018) argue that social identity plays a role in individual alienation. However, it is well-known that the years after liberalization in early nineties have unleashed many economic opportunities in the Indian economy, and there are compelling reasons to assume that the dynamic economic environment is influencing the

occupational behaviour of people. Individuals from the deprived social class may seek professional human capital in times of rapid economic growth, rather than devote time to building entrepreneurial human capital that needs stronger social networks, amongst others, to be in place. However, they are likely to be less entrenched by the caste system over time as the tacit restrictions on occupational choice may become less binding as they gain more education. These lead to the following hypotheses:

Hypothesis 2a: As an economy develops, individuals belonging to backward social classes are more likely to become wage employees.

Hypothesis 2b: As an economy develops, education moderates the effect of social class, and educated individuals belonging to backward social classes are less likely to be inhibited from entering into self-employment.

3 Data construction, characteristics and estimation issues

3.1 Pseudo-panel data construction

As the objective is to characterise self-employment choice along the development path of an economy, conditional on existing social dynamics such as prevalence of caste hierarchy, creating a panel data by tracking individuals over a period of time is needed. However, the existing employment-unemployment data of the National Sample Survey of India (NSSO) does not report such information based on which an individual's employment status can be tracked. Panel data, in general, have limits in terms of availability over time and attrition. In this situation, a "pseudo-panel" based on cohorts that are stable groups of individuals, rather than individuals over time, can shed light on these dynamics. Here, individual variables are replaced by their intra-cohort means. As the database is in the form of independent repeated cross sections, pseudo-panels are constructed as an alternative to panel data for estimating the empirical models.

This method is helpful because the linear model with individual fixed effect corresponds to its pseudo-panel data counterpart due to the linearity of this transformation when using intra-cohort means. The individual

fixed effect is replaced by a cohort effect and the model is particularly simple to estimate if the cohort effect can itself be considered as a fixed effect. For these reasons, the criteria for forming the cohorts must be taken into account like, for example, being observable for all the individuals, the ability to form a partition of the population (each individual is classified into exactly one cohort), and corresponding to a characteristic of the individuals that will usually not change over time like gender. Furthermore, the size of the cohorts depends on the trade-off between bias and variance. It must be large enough to limit the extent of measurement error on intra-cohort variable means that leads to bias and imprecise estimators of the model parameters.

3.1.1 The architecture of pseudo-panels

As Deaton (1985) argues, in the absence of genuine panel data, repeated cross-sectional data can be used to construct synthetic or pseudo-panels. A pseudo-panel based on, for instance, age cohorts, gender, or education levels can be used to control for at least cohort fixed level effects. Such methods are similar to instrumental variable methods where group dummies are used as instruments.[6] To explicate, consider the following linear model with individual effects,

$$y_{it} = x_{it}\beta + \alpha_i + e_{it}, \tag{1}$$

for $i = 1, \ldots, N$ and $t = 1, \ldots, T$.

For simplicity, we assume that observations on N individuals are available for all the time periods. When the individual fixed effects α_i are uncorrelated with x_{it}, it is possible to pool the cross sections to consistently estimate the regression coefficients β. In most situations, the correlation between the individual effects and some of the explanatory variables implies that the K moment condition given by $E\{(y_{it} - x_{it}\beta)x_{it}\} = 0$ is violated, in which case the cross sections are not poolable. In case the data is genuine panel data, the fixed effects approach can be used to treat α_is as unknown fixed parameters. However, if the data on the same individual are not available for each year, this cannot be used.

---

[6]See Verbeek (2006) for details.

Following Deaton ([1985](#)), the observations are aggregated to cohort levels, where cohorts represent people of similar characteristics. In this case, the model assumes the following form,

$$\overline{y}_{ct} = \overline{x}_{ct}\beta + \overline{\alpha}_c + \overline{e}_{ct}, \tag{2}$$

for $c = 1 \ldots C$ and $t = 1 \ldots T$, where the variables are aggregated to cohort level averages. This pseudo-panel, however, does not allow consistent estimation of β as $\overline{\alpha}_{ct}$ is likely to be correlated with the $\overline{x}_{ct}$. Under an assumption that $\overline{\alpha}_{ct}$ is a term fixed over time, the above equation can be consistently estimated. This is very likely to be the case when the average cohort size, $n_c \to \infty$. In such a case, the natural estimator for β is the within estimator given by,

$$\hat{\beta}_W = \left(\sum_{c=1}^{C} \sum_{t=1}^{T} (\overline{x}_{ct} - \overline{x}_c)(\overline{x}_{ct} - \overline{x}_c)' \right)^{-1}$$
$$\sum_{c=1}^{C} \sum_{t=1}^{T} (\overline{x}_{ct} - \overline{x}_c)(\overline{y}_{ct} - \overline{y}_c) \tag{3}$$

As described in Verbeek ([2006](#)), the asymptotic behaviour of pseudo-panel estimators can be derived for the following alternative asymptotic sequences. First, when $N \to \infty$, with C fixed, so that $n_c \to \infty$. Second, when $N \to \infty$ and $C \to \infty$, with n_c fixed. Third, $T \to \infty$, with N, C and n_c fixed. While Moffitt ([1993](#)) and Verbeek and Vella ([2005](#)) employ the asymptotics of the first type, Deaton ([1985](#)) and Verbeek and Nijman ([1993](#)) employ the second type.

In this paper, we also assume asymptotics of the first type. In this case, the fixed effects estimator is consistent estimator for β, when

$$plim \frac{1}{CT} \Sigma \Sigma (\overline{x}_{ct} - \overline{x}_c)(\overline{x}_{ct} - \overline{x}_c)' \tag{4}$$

is finite and invertible and

$$plim \frac{1}{CT} \Sigma \Sigma (\overline{x}_{ct} - \overline{x}_c)\alpha_{ct} = 0 \tag{5}$$

As $n_c \to \infty$ the above conditions are automatically satisfied as the cohort fixed effects converge to a constant over time, that is, $\alpha_{ct} \to \alpha_c$ (Moffitt, [1993](#)).[7] Deaton ([1985](#)) relies on asymptotics of the first type and does away with the necessity to have large numbers of observations in each cohort. This is

achieved by considering the cohort averages as error-ridden measurements of the population averages of the cohorts. By assuming that measurement errors are distributed with zero mean, the moment matrices of the within estimator are adjusted to correct for the measurement error.[8]

3.2 Estimation and selection issues

When the cohort selection criterion has the qualities required to consider the psedio panel estimation as a fixed effects model, the parameters are generally estimated based on standard panel data estimation techniques. A within transformation is applied as in standard panel data estimation in which, for each cohort, the various variables are centred on the mean of the observed values for the cohort, for all the observation dates (in this case, years 1994, 2000 and 2004). This allows the estimation of the cohort effect estimator.

In the estimations, there may be some endogeneity or self-selection issues, particularly with respect to wages and assets. The empirical strategy addresses this problem by introducing an extensive set of land controls in the estimation. Thus, the panel estimation results can also be viewed as conditional on these controls. Due to the problem of non-availability of panel data, we have constructed pseudo panels for tracking self-employment over time. Our sample has three time periods (three survey rounds), and therefore application of system GMM or Arellano-Bond type dynamic panel estimation to tackle potential endogeneity in the estimation is not achievable. These estimations require deeper lags of the difference of both predictors and dependent variable but we use the land ownership variables to closely approximate assets.

3.3 Data source and distributional characteristics

For the pseudo-panel analysis, the 50th round (collected during 1994–1995), the 55th round (collected during 1999–2000) and 61st round employment-unemployment surveys (collected during 2004–2005) of the National Sample Survey Organisation of India

---

[7]Verbeek and Nijman ([1992](#)) show that even when the cohort sizes are large, the bias may be present.

[8]When cohorts are based on age groups, the asymptotics of the second type seldom get satisfied, as the number of cohorts is fixed. For this reason, as the cohort sizes in the sample are very large, we assume asymptotics of the first type, to consistently estimate β.

(NSSO) are used. The sample is restricted to those who are older than 15 years but younger than 70 years. Family members who assist household enterprises, children and the elderly, and people classified into other miscellaneous occupational categories are excluded from the analyses. Individuals who have reported their principal economic activity to be self-employment (including own-account workers and employers), salaried employment, casual labor, or unemployment are included in the sample.[9]

Table 1 presents the summary statistics from the three cross-sectional databases. As given in this Table, for the years 1994–1995, the sample consists of 164,543 individuals. For the years 1999–2000, it consists of 171,361 individuals, and for the years 2004–2005, the data are available for 170,281 individuals. The table suggests that 40.4% of individuals are self-employed in 1994–1995. This rises up to 44.6% in 2000 and 53% in 2005.

Pseudo-panels are constructed using these cross-sectional databases from 1994–1995, 1999–2000, and 2004–2005 by constructing cohorts of individuals based on 5 year age bands interacted with the state of their residence. For the asymptotic reasons described earlier, only those cohorts that have at least 500 observations in each of the surveys are considered for the analysis.[10] Table 2 presents the summary statistics of the cohorts of the pseudo-panel that are constructed from the cross-sectional databases. The estimations use these variables reported in Table 2.

Figure 1 reports the adaptive Kernel density plots for the dependent variable, the proportion of self-employed individuals in cohorts, for all the three periods. The distributions do not appear to be strictly unimodal as both left and right hand side of the distributions contain significant number of observations implying the existence of some clusters. Given the

Table 1 Descriptive statistics (full sample)

	1994–1995	1999–2000	2004–2005
Self-employed	0.386	0.393	0.454
Age	36.668	36.823	37.971
(std.dev)	(12.629)	(12.510)	(12.396)
Female	0.197	0.194	0.200
Married	0.763	0.760	0.781
No education	0.362	0.327	0.286
Informal Education	0.005	0.004	0.015
Primary School	0.117	0.108	0.101
High School	0.259	0.270	0.310
University	0.167	0.191	0.192
No Tech. Education	0.956	0.956	0.957
Tech. Degree	0.021	0.007	0.006
Tech. Diploma	0.023	0.036	0.025
Urban	0.389	0.396	0.353
Agriculture	0.454	0.424	0.362
Hindu	0.795	0.783	0.767
Backward Class	0.275	0.300	0.313
Land < 0.2 Hectares	0.129	0.154	0.063
0.2 < Land <0.4 Hectares	0.408	0.531	0.576
0.4 < Land < 2 Hectares	0.239	0.219	0.241
Land > 2 Hectares	0.224	0.097	0.120
Sample Size	164543	171361	170281

large number of cross-sectional observations, it may look like the distributions have equal mean, median and mode central values.

Figure 2 plots bivariate Kernel density with contour surface plot for the years 1995, 2000 and 2005 of self-employment and university education in the cohorts. These offer a visual impression of the surface of the joint density thus providing a clearer impression of relative heights and the concentration of the observations in the database. Figure 3 illustrates the surface of the joint density of proportion of self-employed individuals and social class. The expanding contours show relative spreads, that the principal economic activities are distributed heterogeneously across the distribution of castes. They also show some fixed patterns of red contour areas with a significant degree of association between these two variables.

---

[9]The principal economic activity alone is considered in the analysis for two reasons. First, not all individuals are engaged in subsidiary activities. Second, as less than one-sixth of the entire sample is engaged in subsidiary activities, considering such activities would further complicate the analysis when individuals report being both self-employed and paid employees. Furthermore, the principal economic activity is the activity to which the individuals devote most of their time.

[10]The number of states and districts has increased in India in the decade following the collection of the 1994–1995 survey. Hence, the databases are harmonized based on the geographic boundaries in the year 1995 for the pseudo-panel analysis.

Table 2 Descriptive statistics (cohorts)

	1994–1995	1999–2000	2004–2005
Dependent Variable			
Prop. Self-employed	0.404	0.446	0.530
Independent Variables			
Avg. Age	34.887	39.754	44.752
(std.dev)	(11.90)	(11.97)	(12.02)
Prop. Female	0.200	0.189	0.183
Prop. Married	0.719	0.801	0.861
Prop. No education	0.331	0.306	0.300
Prop. Informal Education	0.004	0.004	0.018
Prop. Primary School	0.115	0.113	0.108
Prop. High School	0.280	0.269	0.289
Prop. University	0.180	0.203	0.187
Prop. No Tech. Education	0.957	0.957	0.962
Prop. Tech. Degree	0.020	0.008	0.005
Prop. Tech. Diploma	0.023	0.035	0.021
Prop. Urban	0.386	0.392	0.329
Prop. Agriculture	0.433	0.429	0.386
Prop. Hindu	0.745	0.714	0.711
Prop. Backward Class	0.299	0.331	0.341
Prop. Land <0.2 Hectares	0.117	0.131	0.056
Prop. 0.2< Land <0.4 Hectares	0.411	0.508	0.519
Prop. 0.4< Land < 2 Hectares	0.266	0.248	0.279
Prop. Land > 2 Hectares	0.206	0.114	0.146
Number of Cohorts	194	194	194

A number of other factors influence occupational choice, and these are introduced as controls in the estimations. In particular, wealthier individuals have more of a "safety net" when embarking on a new venture than their less wealthy counterparts. Wealth itself can make financing self-employment possible, but it also makes it easier to obtain credit. Households with very high levels of wealth have a higher propensity to take risk (Carroll, 2000). Blanchflower and Oswald (1998) find that inheritance increases the probability of self-employment. Banerjee and Neuman (1993) argue that wealth distribution determines occupational structure. For these reasons, wealthy individuals are more likely to enter self-employment over time, and

we control for these effects by introducing the land possessed variables. Furthermore, most empirical evidence suggests a positive relationship between age and entrepreneurship (Evans & Leighton, 1989a; Blanchflower & Meyer, 1994; Blanchflower, 2000), and married individuals are more likely to be self-employed than are their nonmarried counterparts (Borjas, 1986; Blanchflower & Oswald, 1998).[11] These variables are also introduced in the estimations as controls.

4 Empirical results

The empirical results are presented in the following order. Firstly, the results of the pseudo-panel regressions for all the cohorts are presented in Table 3. Following this the pseudo-panel regression results for the non-agricultural and agricultural sectors are presented in Table 4, and the results for rural and urban areas are presented in Table 5.

The first column of Table 3 presents the base specification of the empirical model for all cohorts.[12] The results suggest that increases in high school education in a cohort increase self-employment in the cohort but University education reduces it. As the negative and significant coefficient of the Backward Class variable suggests, an increase in the proportion of individuals belonging to socially backward classes in a cohort decreases the proportion of self-employed individuals in the cohort. Column 2 of Table 3 introduces the interaction between the Backward Class and University variables into the estimation but this interaction term is insignificant here.

In column 3, the results of the two staged estimation are presented. In the first stage, a measure of development (measured by log of mean per capita consumption) is regressed on the other variables. Following this, the predicted development measure is interacted with the University education variable and introduced in the estimation. This approach helps resolve a part of the correlation between the development indicator with other predictors, and allows us to examine how development moderates the effect education on self-employment. The results in column 3 suggest that the

---

[11]See Le (1999) and Parker (2018) for surveys of empirical studies on self-employment.

[12]The standard OLS model is likely to be biased as the F test in all the fixed effects regressions is significant, suggesting that cohort fixed effects are correlated with the exogenous variables.
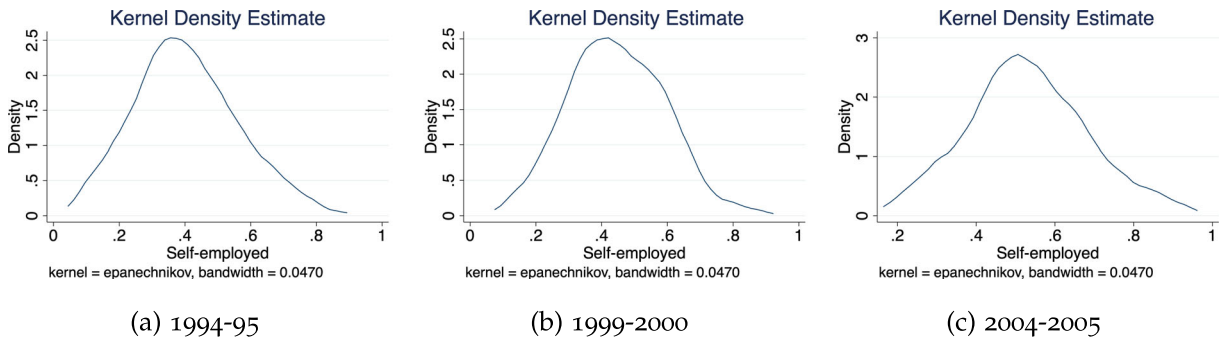
Fig. 1 Distribution of self-employment in cohorts. (a) 1994–1995. (b) 1999–2000. (c) 2004–2005
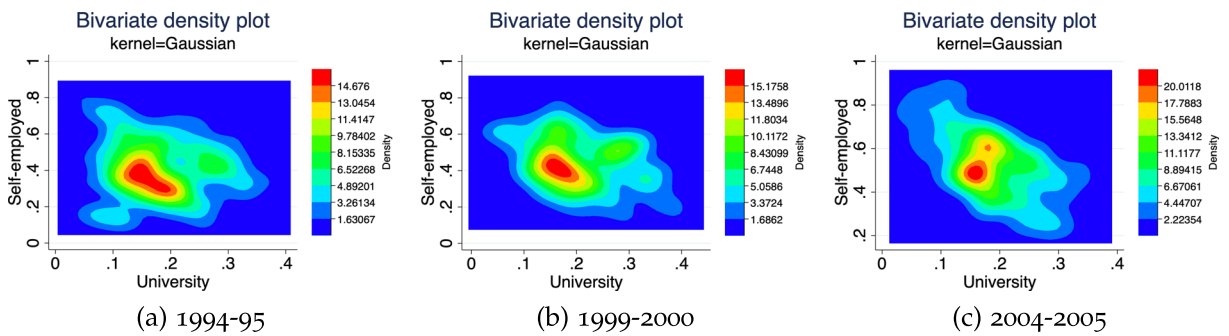

Fig. 2 Bivariate distribution of self-employment and university education. (a) 1994–1995. (b) 1999–2000. (c) 2004–2005
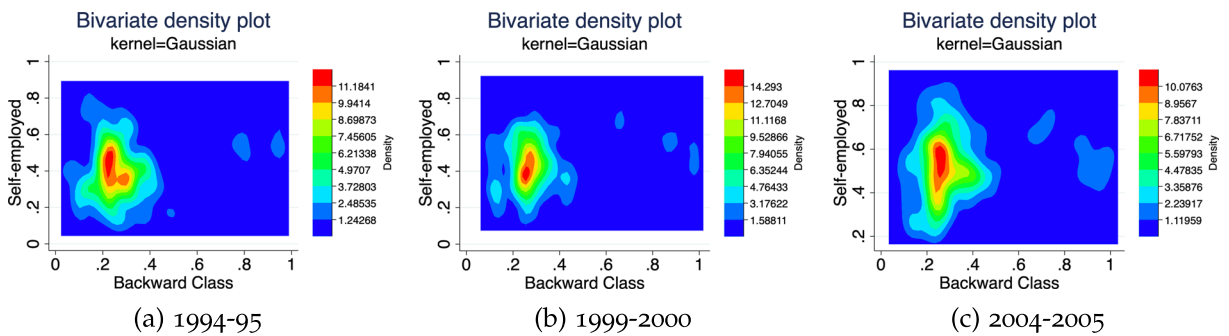

Fig. 3 Bivariate distribution of self-employment and social class. (a) 1994–1995. (b) 1999–2000. (c) 2004–2005

Table 3 Dynamics of self-employment, social class and education over development (pseudo-panel estimations)

	(1)	(2)	(3)
Prop. Primary School	0.0901	0.0959	0.171
	(0.145)	(0.147)	(0.139)
Prop. High School	0.250***	0.240***	0.269***
	(0.0907)	(0.0916)	(0.0804)
Prop. University	−0.192*	−0.113	−1.053***
	(0.116)	(0.138)	(0.181)
Backward Class	−0.226***	−0.177***	−0.205***
	(0.0435)	(0.0605)	(0.0619)
Backward Class* University		−0.280	−0.373
		(0.322)	(0.306)
$\widehat{Develop}$			−0.0180***
			(0.00370)
$\widehat{Develop}$ * Prop. University			0.0861***
			(0.0108)
Avg. Age	0.0116***	0.0115***	0.0100***
	(0.00125)	(0.00124)	(0.00168)
Prop. Female	−0.0805	−0.0851	−0.0889
	(0.114)	(0.112)	(0.0993)
Prop. Married	0.135***	0.136***	0.138***
	(0.0219)	(0.0224)	(0.0230)
Prop. Urban	−0.110	−0.109	−0.0857
	(0.0676)	(0.0678)	(0.0745)
Prop. Agriculture	0.129	0.116	0.0717
	(0.0836)	(0.0880)	(0.108)
Prop. Hindu	0.0339	0.0367	0.0345
	(0.0547)	(0.0553)	(0.0497)
Prop. 0.2< Land <0.4 Hectares	−0.0223	−0.0166	−0.0118
	(0.0497)	(0.0493)	(0.0364)
Prop. 0.4< Land < 2 Hectares	0.433***	0.442***	0.459***
	(0.0932)	(0.0922)	(0.0877)
Prop. Land > 2 Hectares	0.0361	0.0338	
	(0.0540)	(0.0550)	
Constant	−0.211	−0.221*	0.0316
	(0.131)	(0.128)	(0.168)
Observations	582	582	582
R-squared	0.696	0.697	0.743
Number of Cohorts	194	194	194

Dependent variable is the proportion of individuals who are self-employed in a cohort. * signifies $p < 0.05$; ** signifies $p < 0.01$; *** signifies $p < 0.001$. Standard errors are reported in parentheses. Average per capita consumption expenditure is used as a measure of Development (*Develop*). The predicted $\widehat{Develop}$ from the first staged equation is included here to estimate the effect of development on self-employment

predicted $\widehat{Develop}$ variable is significant and negative suggesting that the process of development gets individuals out of self-employment. However, the positive and significant interaction between $\widehat{Develop}$ and University suggests that in the most developed cohorts, University education has a positive impact on self-employment in the cohorts over time. Thus, the results support hypotheses 1(a), 1(b) and 2(a).

As these results consider agricultural and non-agricultural individuals together in the cohorts, the estimated effects are net effects of both the sectors. If the effects are distinct across these two sectors, the net effects are likely to cancel out these distinct effects. For this reason, in Table 4, the models are re-estimated separately for cohorts of individuals in non-agriculture in columns 1–3 and cohorts of individuals in agriculture in columns 4–6. The results strongly support the view that education, social class and development all have has a significantly negative impact on self-employment in non-agriculture. However, the interaction term of Backward Class and University is positive and significant in columns 2 and 3 suggesting that University education enables socially backward class individuals to overcome the barriers to self-employment. Similarly, the coefficient of the interaction between predicted $\widehat{Develop}$ variable and University is positive suggesting that in the most developed cohorts, individuals with University education are more likely to enter self-employment. Thus the empirical results presented in columns 1–3 of Table 4 strongly support hypotheses 1(a), 1(b), 2(a) and 2(b).

In contrast to the non-agriculture sector, University education has no significant effect in the agriculture sector although increases in high school education have a positive effect on self-employment in this sector. However, the effect of belonging to backward class and the moderating role of university education are similar to the effects in non-agriculture. Thus, the inhibiting effects of caste system continue to persist in the agriculture sector but University education allows backward class individuals to overcome the social barriers to self-employment. In contrast to the non-agriculture sector, the positive and significant coefficient of the predicted $\widehat{Develop}$ variable suggests that the process of development increases self-employment in agriculture. Thus, the empirical results presented in columns 4–6 of Table 4 for agriculture sector do not support hypotheses 1(a) and 1(b) but support 2(a) and 2(b).

Table 4 Dynamics of self-employment, social class and education over development in non-agriculture and agriculture (pseudo-panel estimations)

	(1) Nonagri	(2) Nonagri	(3) Nonagri 2-Staged	(4) Agri	(5) Agri	(6) Agri 2-Staged
Prop. Primary School	−0.581***	−0.565***	−0.0943	0.0742	0.0634	0.140
	(0.154)	(0.150)	(0.140)	(0.0983)	(0.0951)	(0.0984)
Prop. High School	−0.0960	−0.115	0.103	0.155*	0.132	0.123
	(0.116)	(0.110)	(0.107)	(0.0871)	(0.0863)	(0.0918)
Prop. University	−0.820***	−1.139***	−2.024***	0.175	−0.289	−0.227
	(0.126)	(0.152)	(0.215)	(0.128)	(0.187)	(0.378)
Backward Class	−0.318***	−0.699***	−0.607***	−0.148***	−0.233***	−0.206***
	(0.107)	(0.166)	(0.151)	(0.0318)	(0.0401)	(0.0437)
Backward Class*University		1.260***	1.132***		1.384***	1.669***
		(0.372)	(0.325)		(0.457)	(0.504)
$\widehat{Develop}$			−0.0412***			0.0164***
			(0.00623)			(0.00383)
$\widehat{Develop}$ * Prop. University			0.113***			−0.00998
			(0.0185)			(0.0274)
Avg. Age	0.00464**	0.00569***	0.00233	0.0110***	0.0117***	0.0154***
	(0.00179)	(0.00179)	(0.00179)	(0.000774)	(0.000785)	(0.00106)
Prop. Female	−0.0714	−0.0734	−0.0243	−0.119	−0.0581	−0.0371
	(0.106)	(0.101)	(0.0948)	(0.0994)	(0.0995)	(0.101)
Prop. Married	0.180***	0.176***	0.238***	0.00660	−0.00336	0.00102
	(0.0281)	(0.0278)	(0.0267)	(0.0260)	(0.0270)	(0.0273)
Prop. Urban	−0.168**	−0.193***	−0.163**	0.157	0.196	0.0206
	(0.0755)	(0.0733)	(0.0722)	(0.130)	(0.129)	(0.147)
Prop. Hindu	−0.0852	−0.0686	−0.0856	0.00337	0.0169	−0.00638
	(0.0714)	(0.0701)	(0.0707)	(0.118)	(0.0926)	(0.0922)
Prop. 0.2< Land <0.4 Hectares	0.000295	−0.0348	0.0262	0.130*	0.141*	0.102
	(0.0468)	(0.0484)	(0.0390)	(0.0744)	(0.0738)	(0.0666)
Prop. 0.4< Land < 2 Hectares	0.268**	0.200*	0.316***	0.544***	0.587***	0.426***
	(0.119)	(0.112)	(0.111)	(0.0920)	(0.0939)	(0.0772)
Prop. Land > 2 Hectares	−0.227***	−0.222***		0.333***	0.382***	
	(0.0568)	(0.0549)		(0.0782)	(0.0784)	
Constant	0.575***	0.668***	0.870***	−0.270**	−0.309**	−0.431***
	(0.141)	(0.140)	(0.151)	(0.129)	(0.120)	(0.136)
Observations	456	456	456	366	366	366
R-squared	0.690	0.700	0.745	0.823	0.829	0.829
Number of Cohorts	152	152	152	122	122	122

Dependent variable is the proportion of individuals who are self-employed in a cohort. * signifies $p < 0.05$; ** signifies $p < 0.01$; *** signifies $p < 0.001$. Standard errors are reported in parentheses. Average per capita consumption expenditure of the cohort is used as a measure of Development (*Develop*). The predicted $\widehat{Develop}$ from the first staged equation is included here to estimate the effect of development on self-employment

Table 5 Dynamics of self-employment in rural and urban areas (pseudo-panel estimations)

	(1) Nonagri Rural	(2) Nonagri Rural	(3) Nonagri Rural 2-Staged	(4) Nonagri Urban	(5) Nongri Urban	(6) Nongri Urban 2-Staged
Prop. Primary School	−0.0309	−0.0354	0.0262	−0.517**	−0.517**	−0.325
	(0.235)	(0.233)	(0.208)	(0.220)	(0.218)	(0.262)
Prop. High School	0.00916	0.00412	0.166	−0.118	−0.129	−0.0984
	(0.124)	(0.121)	(0.128)	(0.127)	(0.128)	(0.132)
Prop. University	−0.322*	−0.388	−0.824*	−0.546***	−0.655***	−1.521***
	(0.191)	(0.235)	(0.465)	(0.140)	(0.181)	(0.552)
Backward Class	−0.263**	−0.315	0.00167	−0.262*	−0.423	−0.603
	(0.108)	(0.216)	(0.260)	(0.145)	(0.267)	(0.371)
Backward Class*University		0.218	−0.698		0.544	1.354
		(0.693)	(0.877)		(0.629)	(1.008)
$\widehat{Develop}$			−0.0391***			−0.0270*
			(0.0105)			(0.0142)
$\widehat{Develop}$ * Prop. University			0.0659			0.0836*
			(0.0415)			(0.0503)
Avg. Age	0.00935***	0.00933***	0.00540**	0.00569***	0.00616***	0.00511**
	(0.00191)	(0.00190)	(0.00223)	(0.00187)	(0.00207)	(0.00203)
Prop. Female	0.151	0.150	0.361***	0.145	0.151	0.129
	(0.112)	(0.110)	(0.134)	(0.122)	(0.122)	(0.117)
Prop. Married	0.241***	0.238***	0.343***	0.148***	0.146***	0.156***
	(0.0351)	(0.0380)	(0.0446)	(0.0304)	(0.0311)	(0.0320)
Prop. Hindu	−0.515***	−0.512***	−0.393**	−0.289*	−0.297*	−0.290*
	(0.152)	(0.157)	(0.161)	(0.165)	(0.165)	(0.170)
Prop. 0.2< Land <0.4 Hectares	−0.388***	−0.386***	−0.243**	0.0515	0.0423	0.0674
	(0.130)	(0.129)	(0.109)	(0.0461)	(0.0481)	(0.0432)
Prop. 0.4< Land < 2 Hectares	−0.0904	−0.0899	0.401***	0.582**	0.496*	0.535*
	(0.124)	(0.125)	(0.127)	(0.293)	(0.287)	(0.275)
Prop. Land > 2 Hectares	−0.570***	−0.568***		−0.0905*	−0.0875*	
	(0.141)	(0.140)		(0.0500)	(0.0505)	
Constant	0.744***	0.761***	0.613***	0.432**	0.464**	0.712***
	(0.195)	(0.184)	(0.196)	(0.176)	(0.185)	(0.214)
Observations	246	246	246	312	312	312
R-squared	0.741	0.741	0.753	0.623	0.625	0.643
Number of Cohorts	82	82	82	104	104	104

* signifies $p < 0.05$; ** signifies $p < 0.01$; *** signifies $p < 0.001$. Standard errors are reported in parentheses. Dependent variable is the proportion of individuals who are self-employed in a cohort

In summary, the results in Table 4 suggest that belonging to a socially backward class has, on an average, a negative effect on self-employment choice; however, these effects are moderated by University education. In both non-agriculture and agriculture sectors, the interaction of higher education with social class produces a positive effect on self-employment choice suggesting that higher education ameliorates the negative impact of social class on self-employment. Thus, education enables individuals to

overcome the limitations posed by the caste system, and the results offer significant support to hypotheses 2 (a) and 2 (b) in both sectors.

For the control variables in the non-agriculture equations in column 1–3 of Table 4, the coefficients are positive and significant in the midsized land ownership category and negative for the largest land size variable. In contrast to this, in the agriculture equations in columns 4–6 of Table 4, the largest land ownership variables are positive and significant. Thus, small amounts of land enable individuals to enter self-employment in non-agriculture, while individuals with large amounts of land choose self-employment in agriculture. Furthermore, the urban variable is significant and negative in nongariculture equation in columns 1–3 but insignificant in columns 4–6. This result is expected as one can expect an increase in the share of urban population in the cohort to have a positive influence on self-employment as rural-urban migration is likely to increase the share of people working as salaried employees in non-agriculture. The effects of the age and gender are consistent with extant literature.

For a more nuanced picture of these effects in rural and urban areas, the models are re-estimated for rural and urban areas separately in Table 5. While most individuals working in agriculture live in rural ares, individuals working in non-agriculture are found in both urban as well as rural areas. As these areas are likely to have distinct effects on self-employment, the estimation results on pseudo-panels of cohorts of individuals in rural areas and urban areas in the non-agriculture sector are presented in Table 5. The results for rural cohorts are presented in columns 1–3 and urban cohorts in columns 4–6.

In both rural and urban areas, the results suggest that increases in University education over time have a significant negative effect on self-employment in cohorts demonstrating that educated individuals become less likely to be self-employed.[13] Thus, in both rural as well as urban areas, individuals who acquire higher human capital are more likely to move out of self-employment over time. The coefficient of

the backward class variable is significantly negative in columns 1 and 4 of Table 5 suggesting that an increase in the proportion of the socially backward class people in a cohort reduces self-employment in it. The interaction variables of Backward Class with University are insignificant in columns 2 and 3 and 5 and 6. It is likely that these insignificant coefficients are because of the smaller sample number of cohorts in the estimated models and the smaller number of individuals within the cohorts.

The predicted $\widehat{Develop}$ variable has a significantly negative coefficient in columns 3 and 6 suggesting that in both rural as well as urban areas the process of development decreases self-employment. However, in column 6 the interaction between the predicted $\widehat{Develop}$ variable and Prop. University is positive and significant suggesting that in the most developed urban areas, highly educated individuals are more likely to enter self-employment over time.

In Appendix Table 6, we introduce more education variables to test for the robustness of the empirical results presented in Tables 3 and 4. The results are entirely consistent with the findings presented above. In Appendix Table, as endogeneity can result from a reverse causality in the relationships, it is examined by regressing land ownership on self-employment. In general, we find that self-employment does not statistically significantly impact land ownership in the majority of specifications. Hence, the proposed direction of causality—land ownership impacting self-employment choice can be considered valid.

5 Discussion and conclusion

Extant literature suggests that the evolution of an individual's choice function is conditional on development stages of an economy: from an unstable and regressive growth path to a path of progression and sustainability (Lucas, 1978; Iyigun and Owen, 1999). Focusing mainly on the industrialized nations and aggregate data, empirical literature has showed that there is a long-term decline in self-employment in the 1970s and an observed trend of revival since 1970s. Taking the case of OECD countries, for example, Bogenhold and Staber (1991) argue that rise in self-employment in these countries since 1970s may be "a response to deficiencies in labour markets rather than a sign of economic vitality". However, such an inference is

---

[13]There are many reasons why this might be the case. If the credit constraints are relatively more severe for educated individuals in rural areas than in urban areas, this result is plausible. Moreover, it should be noted that the estimations are based on two separate pseudo-panels constructed for the rural and urban cohorts. Thus the coefficients are not strictly comparable.

problematic for a fast-growing developing economy like India. This article is amongst the sparse body of literature in the developing country context examining the dynamics of self-employment choice over time using microdata.

Despite significant deficiencies in the labour market, self-employment and entrepreneurial choices in a developing country like India may be a sign of economic vitality. A certain sectoral bias while interacting with individuals' social identities may form a natural bound for self-employment growth and entrepreneurial choice. However, educational attainments increasingly offset negative effects of caste-based social identities. This article studies and establishes these predictions by using pseudo-panel techniques to analyse the dynamics of self-employment in India. Using three different cross-sectional databases collected over 1994–2004, the pseudo-panel analysis tracks the dynamics of self-employment in India. The results suggest that in both agricultural and non-agricultural sectors, individuals from socially backward classes are less likely to enter into self-employment over time. However, University education moderates this effect of social class on self-employment.

Furthermore, the results support the claim that as an economy develops individuals prefer to invest in professional human capital instead of entrepreneurial human capital. Consistent with the theoretical predictions, the results suggest that individuals in non-agriculture who acquire education and wealth are likely to move out of self-employment over time. However, individuals in agriculture who acquire education and wealth are more likely to enter self-employment over time.

By examining the dynamics of self-employment and the characteristics of individuals who are entering into it over time, we are able to indirectly examine the push-pull dynamics of self-employment. The inherent heterogeneity within the self-employment sector is evident as we find that University education and development processes decrease self-employment in the cohorts suggesting that necessity entrepreneurship reduces over the course of economic development. The positive and significant coefficient on the interaction term between the predicted development measure and University education suggests that individuals are able to pursue opportunity entrepreneurship in cohorts

that are most developed. The results demonstrate that educated individuals are less likely to enter self-employment as the economy develops suggesting that they are less likely to be pushed into self-employment but in the most developed cohorts they are pulled into self-employment. Thus, we are able to indirectly identify the push-pull aspects of self-employment choice over the economic development trajectory.

The article has important policy implications. It suggests that public policy should focus on facilitating industrial development in the initial stages of development as individuals are motivated to invest in professional human capital during these periods while encouraging entrepreneurial activities when a more developed economy can provide opportunities for greater returns in self-employment. However, at all stages of economic development, policy needs to ensure that social barriers do not inhibit individuals from transitioning into self-employment and entrepreneurship.

Overall, the paper makes several novel contributions to the extant literature. In particular, it sheds fresh light on the role of social identity for self-employment along stages of development. Empirically, it offers partial support to the claim of Iyigun and Owen (1999) that as an economy develops individuals prefer to invest in professional human capital instead of entrepreneurial human capital. The article sheds fresh light on the self-employment process in India, a rapidly growing country that is trying to break social barriers to entrepreneurship.

Institutional theory (North, 1991; Scott, 1995; DiMaggio, 1998) has underscored the role of institutions in shaping economic behaviour of individuals. The institutional context may have a significant influence on entrepreneurial decision-making (Tolbert et al., 2011; Welter & Smallbone, 2011). For instance, following Mead (1934), Klyver et al. (2013) suggest that the gender gap in self-employment is attributable to their socially constructed institutional context. However, we do not pursue institutional aspects here as we restrict the focus to the individual determinants of self-employment to avoid introducing further complexity to the construction of the pseudo-panels and the empirical models. Furthermore, the limitations of the data do not allow us to examine the hybrid nature of the self-employment activity. These can be fruitful areas for future research examining the

self-employment dynamics in developing countries. Following O'Connor (2013), future research can also examine the role of entrepreneurship education, rather than education per se, for self-employment choice.

Appendix

Table 6 Dynamics of self-employment and development (pseudo-panel estimations)

	(1) All	(2) All	(3) All	(4) Nonagri	(5) Nonagri	(6) Nonagri	(7) Agri	(8) Agri	(9) Agri
Prop. Informal Education	0.694*	0.684*	0.981**	1.099	1.127	1.950***	−0.340	−0.281	−0.661*
	(0.402)	(0.402)	(0.442)	(0.717)	(0.694)	(0.745)	(0.345)	(0.334)	(0.363)
Prop. Primary School	0.100	0.102	0.167	−0.610***	−0.576***	−0.135	0.0766	0.0584	0.148
	(0.146)	(0.147)	(0.138)	(0.150)	(0.148)	(0.144)	(0.0978)	(0.0965)	(0.0991)
Prop. High School	0.282***	0.272***	0.313***	−0.180	−0.157	0.0163	0.158*	0.149*	0.135
	(0.0932)	(0.0945)	(0.0850)	(0.126)	(0.119)	(0.118)	(0.0841)	(0.0850)	(0.0891)
Prop. University	−0.210	−0.153	−1.226***	−0.742***	−1.051***	−2.208***	0.166	−0.341*	−0.331
	(0.127)	(0.142)	(0.189)	(0.122)	(0.162)	(0.242)	(0.127)	(0.198)	(0.375)
Prop. Tech. Degree	0.187	0.155	−0.460	−0.382	−0.171	−0.432	−0.788	−1.183	−0.851
	(0.488)	(0.477)	(0.494)	(0.417)	(0.437)	(0.542)	(1.190)	(1.144)	(1.126)
Prop. Tech. Diploma	0.252	0.219	0.574**	−0.400**	−0.232	−0.132	−0.211	0.0909	0.0888
	(0.240)	(0.238)	(0.256)	(0.193)	(0.195)	(0.239)	(0.712)	(0.730)	(0.746)
Backward Class	−0.252***	−0.221***	−0.280***	−0.377***	−0.710***	−0.640***	−0.142***	−0.235***	−0.195***
	(0.0477)	(0.0632)	(0.0657)	(0.112)	(0.179)	(0.161)	(0.0312)	(0.0390)	(0.0436)
Backward Class*University		−0.180	−0.102		1.117**	1.049***		1.473***	1.726***
		(0.301)	(0.288)		(0.442)	(0.369)		(0.459)	(0.511)
$\widehat{Develop}$			−0.0186***			−0.0499***			0.0172***
			(0.00491)			(0.00742)			(0.00400)
$\widehat{Develop}$ * Prop. University			0.0926***			0.145***			−0.00608
			(0.0113)			(0.0222)			(0.0244)
Avg. Age	0.0109***	0.0108***	0.00843***	0.00225	0.00368	−0.00115	0.0114***	0.0118***	0.0165***
	(0.00154)	(0.00152)	(0.00220)	(0.00219)	(0.00232)	(0.00201)	(0.000873)	(0.000895)	(0.00125)
Prop. Female	−0.0695	−0.0726	−0.0781	−0.0661	−0.0588	0.00988	−0.123	−0.0662	−0.0438
	(0.112)	(0.111)	(0.0991)	(0.107)	(0.104)	(0.0960)	(0.0985)	(0.0984)	(0.0992)
Prop. Married	0.132***	0.133***	0.137***	0.196***	0.185***	0.254***	0.00862	−0.00395	0.000696
	(0.0225)	(0.0231)	(0.0232)	(0.0286)	(0.0283)	(0.0256)	(0.0267)	(0.0273)	(0.0284)
Prop. Urban	−0.128*	−0.127*	−0.0948	−0.145*	−0.177**	−0.134*	0.173	0.220	0.0551
	(0.0676)	(0.0676)	(0.0821)	(0.0783)	(0.0772)	(0.0758)	(0.130)	(0.133)	(0.148)

Table 6 (continued)

	(1) All	(2) All	(3) All	(4) Nonagri	(5) Nonagri	(6) Nonagri	(7) Agri	(8) Agri	(9) Agri
Prop. Agriculture	0.155*	0.145*	0.112						
	(0.0826)	(0.0874)	(0.107)						
Prop. Hindu	0.0316	0.0333	0.0371	−0.0945	−0.0802	−0.113*	−0.0129	0.00723	−0.0168
	(0.0550)	(0.0554)	(0.0473)	(0.0678)	(0.0673)	(0.0678)	(0.115)	(0.0979)	(0.0982)
Prop. 0.2< Land <0.4 Hectares	−0.0125	−0.00850	0.00291	0.0160	−0.0179	0.0542	0.122	0.137*	0.0768
	(0.0502)	(0.0491)	(0.0398)	(0.0448)	(0.0486)	(0.0390)	(0.0746)	(0.0756)	(0.0658)
Prop. 0.4< Land < 2 Hectares	0.376***	0.381***	0.382***	0.237**	0.168	0.287**	0.561***	0.601***	0.440***
	(0.0995)	(0.0984)	(0.0832)	(0.117)	(0.114)	(0.116)	(0.0927)	(0.0952)	(0.0767)
Prop. Land > 2 Hectares	0.00694	0.00713		−0.233***	−0.241***		0.356***	0.401***	
	(0.0605)	(0.0606)		(0.0675)	(0.0670)		(0.0827)	(0.0837)	
Constant	−0.184	−0.188	0.0943	0.680***	0.739***	1.073***	−0.278**	−0.310**	−0.465***
	(0.134)	(0.133)	(0.194)	(0.157)	(0.152)	(0.150)	(0.126)	(0.122)	(0.142)
Observations	582	582	582	456	456	456	366	366	366
R-squared	0.700	0.700	0.753	0.701	0.707	0.756	0.824	0.830	0.831
Number of Cohorts	194	194	194	152	152	152	122	122	122

Dependent variable is the proportion of individuals who are self-employed in a cohort. * signifies $p < 0.05$; ** signifies $p < 0.01$; *** signifies $p < 0.001$. Standard errors are reported in parentheses

Table 7 Testing reverse causality of land ownership (pseudo-panel estimations)

Variables	(1) All Land 2	(2) All Land 3	(3) All Land 4	(4) Nonagri Land 2	(5) Nonagri Land 3	(6) Nonagri Land 4	(7) Agri Land 2	(8) Agri Land 3	(9) Agri Land 4
Self-Employed	−0.0564	0.153***	−0.0596	0.0848	0.0366	−0.220***	−0.389***	0.361***	0.133*
	(0.0651)	(0.0319)	(0.0591)	(0.0761)	(0.0293)	(0.0568)	(0.0872)	(0.0601)	(0.0708)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Constant	−0.361***	0.427***	0.667***	−0.456***	0.362***	0.730***	−0.0534	0.722***	0.280**
	(0.128)	(0.0626)	(0.116)	(0.156)	(0.0601)	(0.116)	(0.142)	(0.0978)	(0.115)
Observations	582	582	582	456	456	456	366	366	366
R-squared	0.558	0.554	0.419	0.664	0.653	0.638	0.410	0.474	0.292
Number of Cohorts	194	194	194	152	152	152	122	122	122

* signifies $p < 0.05$; ** signifies $p < 0.01$; *** signifies $p < 0.001$. Standard errors are reported in parentheses. Dependent variable is the size of average land ownership in the cohort. Land 1 is having no land ownership. Land 2 is for 0.2< Land Owned <0.4 Hectares; Land 3 is for 0.4< Land Owned < 2 Hectares; Land 4 is for Land Owned > 2 Hectares

Data statement

1. The datasets generated during and/or analysed during the current study are available with the Ministry of Statistics, Government of India at http://microdata.gov.in/nada43/index.php/catalog

2. The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

References

Acs, Z. J., Audretsch, D. B., Evans, D. S., & et al (1994). Why does the self-employment rate vary across countries and over time? Technical report. CEPR Discussion Papers.

Åstebro, T., Chen, J., & Thompson, P. (2011). Stars and misfits: self-employment and labor market frictions. *Management Science*, *57*(11), 1999–2017.

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

Banerjee, A. V., & Neuman, A. F. (1993). Occupational choice and the process of economic development. *Journal of Political Economy*, *101*, 274–298.

Berreman, G. D. (1960). Caste in India and the united states. *American Journal of Sociology*, *66*(2), 120–127.

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

Blanchflower, D. G., & Meyer, B. D. (1994). A longitudinal analysis of young entrepreneurs in Australia and the United States. *Small Business Economics*, *6*(1), 1–19.

Blanchflower, D. G., & Oswald, A. (1998). What makes an entrepreneur? *Journal of Labor Economics*, *16*(1), 26–60.

Blanchflower, D. G., Oswald, A., & Stutzer, A. (2001). Latent entrepreneurship across nations. *European Economic Review*, *45*(4-6), 680–691.

Blumer, H. (1986). *Symbolic interactionism: perspective and method*. Berkeley: University of California Press.

Bogenhold, D., & Staber, U. (1991). The decline and rise of self-employment. *Work, Employment & Society*, *5*(2), 223–239.

Bönte, W., & Filipiak, U. (2012). Financial literacy, information flows, and caste affiliation: empirical evidence from india. *Journal of Banking & Finance*, *36*(12), 3399–3414.

Borjas, G. J. (1986). The self-employment experience of immigrants. *The Journal of Human Resources*, *21*(4), 485–506.

Bourdieu, P. (1987). *Distinction: a social critique of the judgement of taste*. Cambridge: Harvard University Press.

Carroll, C. D. (2000). Portfolios of the rich. NBER Working Paper.

Cunningham, W. V., & Maloney, W. F. (2001). Heterogeneity among Mexico's microenterprises: an application of factor and cluster analysis. *Economic Development and Cultural Change*, *50*(1), 131–156.

Deaton, A. S. (1985). Panel data from time series of cross-sections. *Journal of Econometrics*, *30*(1–2), 109–126.

Desai, S., & Dubey, A. (2012). Caste in 21st century india: competing narratives. *Economic and Political Weekly*, *46*(11), 40.

Deshpande, A. (2005). Affirmative action in India and the United States.

Dillon, E. W., & Stanton, C. T. (2017). Self-employment dynamics and the returns to entrepreneurship. Working Paper 23168, National Bureau of Economic Research.

DiMaggio, P. (1998). The new institutionalisms: avenues of collaboration. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, *154*(4), 696–705.

Donoghue, J. D. (1957). An eta community in Japan: the social persistence of outcaste groups. *American Anthropologist*, *59*(6), 1000–1017.

Earle, J. S., & Sakova, Z. (2000). Business start-ups or disguised unemployment? Evidence on the character of self-employment from transition economies. *Labour Economics*, *7*(5), 575–601.

Evans, D. S., & Jovanovic, B. (1989). An estimated model of entrepreneurial choice under liquidity constraints. *Journal of Political Economy*, *97*(4), 808–827.

Evans, D. S., & Leighton, L. S. (1989a). The determinants of changes in US self-employment, 1968–1987. *Small Business Economics*, *1*, 111–119.

Evans, D. S., & Leighton, L. S. (1989b). Some empirical aspects of entrepreneurship. *American Economic Review*, *79*(3), 519–535.

Falco, P., & Haywood, L. (2016). Entrepreneurship versus joblessness: explaining the rise in self-employment. *Journal of Development Economics*, *118*, 245–265.

Fields, G. S. (2004). A guide to multisector labor market models. http://www.ilr.cornell.edu/directory/gsf2/downloads/Multisector_mar05.pdf.

Fölster, S. (2002). Do lower taxes stimulate self-employment? *Small Business Economics*, *19*(2), 135–145.

Foucault, M. (1982). The subject and power. *Critical Inquiry*, *8*(4), 777–795.

Fritsch, M., Kritikos, S. A., & Sorgner, A. (2015). Why did self-employment increase so strongly in Germany? *Entrepreneurship & Regional Development*, *27*(5–6), 307–333.

Gang, I. N., Sen, K., & Yun, M.-S. (2017). Is caste destiny? Occupational diversification among dalits in rural India. *The European Journal of Development Research*, 29(2), 476–492.

Gindling, T. H., & Newhouse, D. (2014). Self-employment in the developing world. *World Development*, 56, 313–331.

Goel, D., & Deshpande, A. (2016). Identity, perceptions and institutions: Caste differences in earnings from self-employment in India.

Günther, I., & Launov, A. (2012). Informal employment in developing countries: opportunity or last resort? *Journal of Development Economics*, 97(1), 88–98.

Gupta, P., Mallick, S., & Mishra, T. (2018). Does social identity matter in individual alienation? Household-level evidence in post-reform India. *World Development*, 104, 154–172.

Hamilton, B. H. (2000). Does entrepreneurship pay? An empirical analysis of the returns to self-employment. *Journal of Political Economy*, 108(3), 604–631.

Iyigun, M. F., & Owen, A. L. (1999). Entrepreneurs, professionals, and growth. *Journal of Economic Growth*, 4, 213–232.

Kijima, Y. (2006). Why did wage inequality increase? Evidence from urban India 1983–99. *Journal of Development Economics*, 81(1), 97–117.

Kim, P., & Aldrich, H. (2005). Social capital and entrepreneurship.

Kim, G., & Cho, J. (2009). Entry dynamics of self-employment in South Korea. *Entrepreneurship & Regional Development*, 21(3), 303–323.

Klyver, K., Nielsen, S. L., & Evald, M.R. (2013). Women's self-employment: an act of institutional (dis) integration? A multilevel, cross-country study. *Journal of Business Venturing*, 28(4), 474–488.

Kolstad, I., & Wiig, A. (2015). Education and entrepreneurial success. *Small Business Economics*, 44(4), 783–796.

Kuznets, S. S. (1966). *Modern economic growth*. New Haven: Yale University Press.

Le, A. T. (1999). Empirical studies of self-employment. *Journal of Economic Surveys*, 13(4), 381–416.

Lucas, R. (1978). On the size distribution of business firms. *Bell Journal of Economics*, 9, 508–523.

Mandelman, F. S., & Montes-Rojas, G. V. (2009). Is self-employment and micro-entrepreneurship a desired outcome? *World Development*, 37(12), 1914–1925.

Margolis, D. N. (2014). By choice and by necessity: entrepreneurship and self-employment in the developing world. *The European Journal of Development Research*, 26(4), 419–436.

Mead, G. H. (1934). *Mind, self and society* Vol. 111. Chicago: University of Chicago Press.

Moffitt, R. (1993). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics*, 59(1), 99–124.

Mohapatra, S., Rozelle, S., & Goodhue, R. (2007). The rise of self-employment in rural China: development or distress? *World Development*, 35(1), 163–181.

North, D. C. (1991). Institutions. *Journal of Economic Perspectives*, 5(1), 97–112.

O'Connor, A. (2013). A conceptual framework for entrepreneurship education policy: meeting government and economic purposes. *Journal of Business Venturing*, 28(4), 546–563.

Parker, S. C. (2018). *The economics of entrepreneurship*, 2nd edn. Cambridge: Cambridge University Press.

Pietrobelli, C., Rabellotti, R., & Aquilina, M. (2004). An empirical study of the determinants of self-employment in developing countries.

Premand, P., Brodmann, S., Almeida, R., Grun, R., & Barouni, M. (2016). Entrepreneurship education and entry into self-employment among university graduates. *World Development*, 77, 311–327.

Rees, H., & Shaw, A. (1986). An empirical analysis of self-employment in UK. *Journal of Applied Econometrics*, 1, 95–108.

Revankar, R. G. (1971). *The Indian constitution–: a case study of backward classes*. Rutherford: Fairleigh Dickinson University Press.

Scott, W. R. (1995). *Institutions and organizations* (Vol. 2).

Sidanius, J., & Pratto, F. (2001). *Social dominance: an intergroup theory of social hierarchy and oppression*. Cambridge: Cambridge University Press.

Sidanius, J., Devereux, E., & Pratto, F. (1992). A comparison of symbolic racism theory and social dominance theory as explanations for racial policy attitudes. *The Journal of social psychology*, 132(3), 377–395.

Sidanius, J., Liu, J. H., Shaw, J. S., & Pratto, F. (1994). Social dominance orientation, hierarchy attenuators and hierarchy enhancers: social dominance theory and the criminal justice system. *Journal of Applied Social Psychology*, 24(4), 338–366.

Srinivas, M. N. et al. (1962). Caste in modern India and other essays. Caste in modern India and other essays.

Tamvada, J. P. (2010). Entrepreneurship and welfare. *Small Business Economics*, 34(1), 65–79.

Tamvada, J. P. (2015). The spatial distribution of self-employment in India: evidence from semiparametric geoadditive models. *Regional Studies*, 49(2), 300–322.

Tamvada, J. P. (2021). Microentrepreneurship in a Developing Country: Evidence for Public Policy. Palgrave Macmillan.

Taylor, M. P. (1996). Earnings, independence or unemployment: why become self-employed? *Oxford Bulletin of Economics and Statistics*, 58, 253–266.

Tolbert, P. S., David, R. J., & Sine, W.D. (2011). Studying choice and change: the intersection of institutional theory and entrepreneurship research. *Organization Science*, 22(5), 1332–1344.

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., & Vijverberg, W. (2005). Entrepreneurship selection and performance: a meta-analysis of the impact of educaton in developing economies. *World Bank Economic Review*, 2, 225–261.

Van der Sluis, J., Van Praag, M., & Vijverberg, W. (2008). Education and entrepreneurship selection and performance: a review of the empirical literature. *Journal of Economic Surveys*, 22(5), 795–841.

Van Praag, M., van Witteloostuijn, A., & van der Sluis, J. (2013). The higher returns to formal education for entrepreneurs versus employees. *Small Business Economics*, 40(2), 375–396.

van Stel, A., & van der Zwan, P. (2020). Analyzing the changing education distributions of solo self-employed workers and employer entrepreneurs in Europe. *Small Business Economics*, *55*(2), 429–445.

Verbeek, M. (2006). Pseudo panels and repeated cross sections. In L. Matyas, & P. Sevestre (Eds.) *The econometrics of panel data: handbook of the theory with applications*.

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

Verbeek, M., & Nijman, T. (1993). Minimum mse estimation of a regression model with fixed effects and a series of cross sections. *Journal of Econometrics*, *59*, 125–136.

Verbeek, M., & Vella, F. (2005). Estimating dynamic models from repeated cross-sections. *Journal of Econometrics*, *127*, 83–102.

Welter, F., & Smallbone, D. (2011). Institutional perspectives on entrepreneurial behavior in challenging environments. *Journal of Small Business Management*, *49*(1), 107–125.

Wennekers, S., Folkeringa, M., & et al (2002). The development of self-employment rate in the Netherlands 1899–1997. Technical report, EIM Business and Policy Research.