07 March 2021

Finance, gender, and entrepreneurship

India's informal sector firms

Ira N. Gang
Economics Department, Rutgers University
New Brunswick, USA
Email: gang@rutgers.edu

Rajesh Raj Natarajan
Department of Economics
Sikkim University, Sikkim
Email: rajeshraj.natarajan@gmail.com

Kunal Sen
UNU-WIDER, Helsinki, Finland and
University of Manchester, Manchester, UK
Email: sen@wider.unu.edu

Abstract: How does informal economic activity respond to increased financial inclusion? Does it become more entrepreneurial? Does access to new financing options change the gender configuration of informal economic activity and, if so, in what ways and what directions? We take advantage of nationwide data collected in 2010/11 and 2015/16 by India's National Sample Survey Office on unorganized (informal) enterprises. This period was one of rapid expansion of banking availability aimed particularly at the unbanked, under-banked, and women. We find strong empirical evidence supporting the crucial role of financial access in promoting entrepreneurship among informal sector firms in India. Our results are robust to alternative specifications and alternative measures of financial constraints using an approach combining propensity score matching and difference-in-differences. However, we do not find conclusive evidence that increased financial inclusion leads to a higher likelihood of women becoming entrepreneurs than men in the informal sector.

Key words: entrepreneurship, financial constraints, gender, informal sector, difference-in-differences, India

JEL classification: O12, G28, L26

Finance, gender, and entrepreneurship

India's informal sector firms

Introduction

One of the salient features of under-development is the existence of a large informal economy characterized by a large mass of non-entrepreneurial firms (De Vreyer and Roubaud 2013). These firms are typically household units, which are survivalist in nature, with limited prospects for growth (Grimm et al. 2012).[1] Many of these household units are run by women, who seem to face greater hurdles than men in making the transition to entrepreneurial firms (for a review, see Jennings and Brush 2013). A large literature has attempted to understand why so few firms expand beyond the household unit, consequently becoming more productive, and why we see gender differences in entrepreneurship in the informal sector (see Chen 2012 and Fields 2019). Credit constraints, a consequence of the widespread failure of credit markets in developing countries, are widely regarded as a key constraint to entrepreneurship development (Kerr and Nanda 2011). Yet systematic evidence on the causal role of finance in determining entrepreneurship in general, and female entrepreneurship in particular, is lacking. This paper focuses on the role that finance/credit/banking access plays in expanding informal sector enterprises beyond family and household units.[2] We ask: (1) Is there a causal link between the availability of finance and the likelihood of informal firms making the transition from strictly family organizations to entrepreneurial ones, i.e. from those employing only family members to those hiring outside labour? (2) Does this differ between female- and male-owned enterprises?

We look at India's unorganized sector, including manufacturing firms and unincorporated non-agricultural enterprises; we use the terms 'unorganized' and 'informal' interchangeably.[3] Approximately 75 per cent of manufacturing employment and 17 per cent of manufacturing output are in the sector; about 86 per cent of these firms are family owned (Raj and Sen 2016). Our strategy lies in taking advantage of nationwide data collected in 2010/11 and 2015/16 by India's National Sample Survey Office (NSSO) on unorganized (informal) enterprises. These are repeated cross-sections of unincorporated non-agricultural enterprises. Informal firms are quite heterogeneous. In line with the NSSO classification, Raj and Sen (2016) categorize firms in this sector into three types: very small pure household enterprises (PHEs), mixed household enterprises (MHEs) that are somewhat larger, using both family and at least one non-family labourer, and larger non-household enterprises (NHEs) employing mostly non-family labour. We define a firm as entrepreneurial if it employs at least one hired worker, in addition to members of

[1] Among the poor, informal firms are important in job creation (Naudé 2010; World Bank 2013). Moreover, Ivlevs et al. (2020) highlight entrepreneurship's contributions to bettering welfare through growth innovation, job creation (van Praag and Versloot 2007), and improving wellbeing and health (Nikolova 2019). Most notably, Landes (1969) placed informal entrepreneurial activity at the heart of early industrialization (though he did not use the word 'informal').

[2] For example, the development literature has spent decades addressing the many aspects of relieving financing constraints facing firms and families (Mead and Liedholm 1998; Rijkers et al. 2010).

[3] Factories registered under the Factories Act of 1948 are generally referred to as the organized or formal sector in India's data collections. Factories not using power during the preceding year and employing 20 or more workers, or those using power and employing ten or more workers, must register under this Act (Gang 1992; Gang and Pandey 2007). Other firms fall into the unorganized sector and may be registered under other legislation and government bodies.

the proprietor's family, i.e. we merge the MHEs and NHEs (both employ hired workers) into one, calling them entrepreneurial firms.

We find that lack of access to finance constrains the transition of firms in the informal sector in India. This effectively means that as the constraint weakens, firms are willing and able to expand hiring beyond family workers. However, we do not find clear gender differences in transition with the easing of finance constraints: female entrepreneurs are no more likely to join the entrepreneurial side of the sector than male entrepreneurs. Our findings are upheld when controlling for the endogeneity of the finance constraint using instrumental variables. We also supplement our main identification strategy with an alternate approach, combining propensity score matching (PSM) and difference-in-differences (DID), which also reinforces our main finding of the positive role of finance in promoting entrepreneurship among informal firms in India.

The paper is in six sections. Next, we provide background, including brief discussions of the literature on the roles of gender and finance in informal sector firm transition, and the relevant policy environment in India. In Section 3, we discuss the data, including descriptive statistics, and outline our empirical strategy and approach to identification. Section 4 takes up our estimation results while Section 5 outlines a critical robustness check. Section 6 concludes.

2 Related literature, background, and policy environment

In this section, we first examine the lessons that we draw about firm transition from the literature on gender roles and financial constraints in development. We then discuss the background to this study and the relevant policy environment in India in subsequent subsections.

2.1 Related literature

Financial constraints are a factor limiting firm's growth (Oliveira and Fortunato 2006). Investment in fixed assets is less likely for firms with financial difficulties (Ojah et al. 2010; Winker 1999). Smaller firms face greater financial limitations than larger ones – for example, Beck et al. (2008) find that financial constraints are associated with a 10 per cent decrease in small businesses growth.

Although studies show the clear role financial constraints play when starting a business, there is less consensus on the role of credit access for the growth and subsequent performance of small businesses. Some studies argue that access to credit helps small firms grow faster (Brown et al. 2005) and is important for the development of micro enterprises (Woodruff and Zenteno 2007). Other studies point to a lesser role played in enterprise growth by financial constraints, finding little convincing evidence that access to formal credit is an important influencing factor (Daniels and Mead 1998; Johnson et al. 2002). It seems that for enterprise growth, access to finance is needed but by itself is not enough (Nichter and Goldmark 2009).

Gender, financing, and their interaction seem to play roles in the development of entrepreneurship in the informal sector. The literature on gender and credit access has expanded rapidly, mostly after the appearance of the World Bank's Enterprise Survey (WBES) of 2006. It is difficult to use these data for studying gender issues, but a series of papers have rather inventively developed ways to use its information on gender to analyse linkages between gender and credit. Unfortunately, the evidence that gender differences in access to finance make a difference in firm development is not conclusive (Presbitero et al. 2014).[4] As we will soon see, our data overcome many problems faced by users of the WBES. This line of research also deals with firms larger than ours and, generally, not in the informal or unorganized economy.

2.2 Background

The incidence of entrepreneurship in India's unorganized economy has many elements: owner's gender, firm financial constraints faced, owner's caste, etc. In India, the unorganized sector has played an honoured role in development strategy. Government policy has encouraged small firms both in the unorganized and organized sectors. As part of the effort to promote industrial decentralization and increase employment, support for small firms was an important element of the Industrial Policy Resolution of 1956, reiterated in December 1977 in the Industrial Policy Statement. Such enterprises have generally been exempt from excise and other taxes, enjoyed protection from larger firms which were often restricted from producing competing products, given preferential pricing (for example, in sales to public sector firms), and so on (Gang 1992).

2.3 Policy environment

India's central bank, the Reserve Bank of India (RBI), follows specific policies aimed at expanding access to banking services.[5] In 2005 the RBI began classifying districts (state subdivisions) as under-banked if their population per bank branch was greater than the national average. Various policies then encouraged bank branch expansion in these districts. In 2011, banks were instructed to open at least 25 per cent of their total branches in a year in unbanked rural centres—a 4:1 norm as against the previous 1:4 norm (Chavan 2020). They were requested to plan for financial inclusion and to set targets for opening branches, small-sized savings deposit accounts, and debit cards, and for providing small-sized overdrafts. The period between 2010/11 and 2015/16 was one of rapid expansion of banking availability in India aimed particularly at the unbanked, under-banked, and women (see Young 2019). This is evident from the decline in number of under-banked districts in India, from 355 in 2010/11 to 344 in 2015/16.The population covered per branch, which was 13,027 in 2010/11, dropped significantly to 8,683 in 2015/16 (Table 1). Two main Indian government programmes were the Bharatiya Mahaila Bank (Indian Women's Bank) started on 19 November 2013, and the Pradhan Mantri Jan Dhan Yojana (Prime Minister's People's Wealth Scheme) launched on 28 August 2014. The Bharatiya Mahila Bank was a public sector bank mandated to cater to the banking needs of women. Under the Pradhan Mantri Jan Dhan Yojana programme, the number of small deposit accounts, debit cards, and banking agents engaged by banks have grown significantly.

[Table 1 near here]

In 2013, women were included in the priority sector, which prior to this year comprised small and marginal farmers, agricultural labourers, and Scheduled Castes (SCs) and Scheduled Tribes (STs). Under priority sector lending requirements, which form 40 per cent of adjusted net bank credit (ANBC), banks must lend 10 per cent of the ANBC to groups that were economically and socially disadvantaged and also included women. Though they began late in our time period, the programmes rapidly expanded bank account ownership and are representative of numerous policies undertaken by the RBI to expand financial inclusion over the twenty-first century. With

[4] These studies cover a wide geography; for example, no gender effect is found by Aterido et al. (2013) on Sub-Saharan Africa, Bruhn (2009) on Latin America, or Storey (2004) on Trinidad and Tobago. See Presbitero et al. (2014) for a succinct and informative summary of this aspect of the literature.

[5] A superior source for understanding RBI policy in this period is Young (2019).

these policies, the share of adults with a bank account more than doubled from 2011 to 2017, to 80 per cent; among women, account ownership increased more than 30 per cent between 2014 and 2017.

Banking the unbanked and under-banked is a policy pushed by governments and international organizations as a multi-goaled win. Has it really delivered? The number of financial accounts opened and by whom is often regarded as a measure of success in bringing the unbanked into the formal financial system. Does it translate into gains for the poor, to more productive and efficient firms, and to greater gender equity in these outcomes and turn some of the beneficiaries into owners of entrepreneurial firms in the informal sector? In brief, how does informal economic activity respond to the extension of financial inclusion? Does it help women enter sectors from which they previously were absent?

3 Data sources, variables, empirical strategy, and identification

In this section, we discuss our data, including descriptive statistics, and outline our empirical strategy and approach to identification. Our intention is to establish a structure that allows an analysis of the allocation of unorganized sector firms across two parts of the unorganized sector, one part entrepreneurial, the other not. To some extent, this is a distinction between a sector that is dynamic and one is residual, allowing us to characterize an informal sector with both entrepreneurial and subsistence aspects, reflecting conflicting characterizations that we find in the literature (see, for example, Fields 1990; Maloney 1999).

3.1 Data sources

Our analysis heavily relies on repeated cross-section, unit-level data drawn from the 67th (2010/11) and the 73rd (2015/16) rounds of the Government of India's National Sample Survey (NSS; see NSSO 2013, 2017), focusing on unincorporated non-agricultural enterprises. These firms are typically characterized as India's large unorganized sector. Both are India-wide enterprise-level surveys, stratified by district. Districts are subunits of Indian states for which many Indian agencies—e.g., Census, Reserve Bank, etc.—make data available. The NSSO uses a block enumeration approach in each district. We have about 620,000 firms in this pooled dataset, across 562 districts of 35 Indian states.

Although there is some variation in their survey questions across years, these two highly compatible random samples allow comparable estimates across years at the enterprise level. From these data, we get information on the firm unit, such as the gender of the owner and employees, family labour, hired/outside labour, labour productivity, various financial availability variables, outstanding loans, share of loans from institutional sources, and banking, among other firm-specific attributes, as well as regional, state, and district information. We use the wholesale price index for capital goods to deflate financial variables.

We draw information on district-level banking from the RBI publication *Basic Statistical Returns of Scheduled Commercial Banks in India*. These district-level banking variables include number of branches, number of deposits and amount deposited, and outstanding credit of scheduled commercial banks. India's 2001 and 2011 Censuses provide the relevant population figures in order to calculate district-level bank branches per capita (Census of India 2001, 2011). The NSSO surveys include the names of the districts in which firms are located; we merged the NSSO, the RBI, and the Census datasets using a one-to-one mapping of 562 districts for the three datasets.

Newly created districts during the period under study are merged with their parent districts to facilitate district-level comparisons over time.[6]

3.2 Variables and descriptive statistics

Our primary objective is to analyse the role of financing in explaining whether firms are entrepreneurial in the Indian informal sector, and whether female-owned firms are more likely to be entrepreneurial with greater access to finance. Critical, therefore, are what we mean by an entrepreneurial firm, how we capture the gendered role of financing, and our measures of the availability of financing to the firm. In Table 2 we outline the key variables of interest to us and discuss the basic descriptive statistics.

[Table 2 near here]

Dependent variable

We classify an entrepreneurial firm as one employing at least one hired worker on a regular basis (besides family workers) and the variable 'entrepreneurial firm' takes the value 1 for such firms, as in, for example, Earle and Sakova (2000). Firms that are not 'entrepreneurial' are own-account enterprises that exclusively make use of family labour; these are mainly found on household premises.

The decision to employ a hired worker transforms an informal firm from an own-account enterprise to an employer, and is seen as an indicator of entrepreneurial success (Gindling and Newhouse 2014). The hiring of non-family workers for a household enterprise involves an implicit barrier to entry, as these employers typically need to finance the wages of hired workers by borrowing from credit markets or through the profits of the enterprise (Banerji et al. 2016). Therefore, firms that have managed to make the transition from an own-account enterprise to becoming an employer have managed to overcome this barrier and can be classified as entrepreneurial firms.[7]

Main independent variables:

Finance constraint: We construct two measures to represent the availability of finance to the firm (or the firm's financial constraint).

1. We construct a direct measure (FIN1). The surveys ask firms if they encountered any borrowing constraints during the last year. This measure takes the form of a dummy variable with the value 1 for firms whose owners reported non-availability or high credit costs as a major problem that they faced over the last year.[8]
2. We also construct 0–1 categorical variables for firms receiving bank loans (FIN2DUM1), non-government loans (FIN2DUM2), and government loans (FIN2DUM3), while keeping firms not taking loans as the benchmark (reference) category.

[6] Issues do exist with the consistency of districts over time, e.g., new districts formed from parts of several older districts (Pradhan 2016).

[7] The existence of dynamic entrepreneurial firms along with subsistence firms in the informal sector is well documented in the literature (Grimm et al. 2012).

[8] Below we discuss a possible selection issue here, in addition to an endogeneity issue for the generic FIN variable.

Each of these two measures of financial constraints is imperfect in itself.[9] Using the two measures mitigates against the problem of measurement errors in any one measure. Specifically, a score of 1 may reflect either that the firm did not attempt to obtain credit or that it faced real difficulty in obtaining credit.

Female: This is a binary variable, which is equal to 1 if the firm is owned and managed by a woman and 0 otherwise. We consider only sole-proprietorship firms, that is, firms with sole owners. We do not consider partnership firms, that is, firms owned by more than one person who share the firm's proceeds. Conceptually, the reason for looking only at proprietorships and not partnerships is that partnerships involve joint decision-making, usually with the partners dividing responsibilities. As such, we do not know who the 'face' of the firm is and who may be running the firm; hence, if the partners are of different genders we do not have a clear indication of the role of gender in our question of interest. As we have such a large sample, limiting our sample to sole proprietorships does not cost in terms of losing observations and provides unambiguity.[10]

Year-district fixed effects: Year and district fixed effects are included to help capture otherwise unobserved year- and district-specific external finance constraints. Time effects include macro shocks with possible firm productivity effects. Unchanging district-specific effects can wield an independent impact on the use of non-family labour besides that wielded by the firm's financing constraints.

The summary statistics for the variables used in the analysis are presented in Table A1 in the online appendix with further description of firm characteristics as control variables. Table 3 presents a simple and fascinating picture of entrepreneurship in our sample. Panel A shows the cross-tabulation of entrepreneurship with the owner's gender; Panel B displays entrepreneurship versus whether the firm faces a financial constraint. Female-run firms are much less likely to be entrepreneurial than male-run firms. Firms not facing a financial constraint are more likely to be entrepreneurial than those with such constraints. We also see striking differences in the characteristics of entrepreneurial and non-entrepreneurial firms (Panel C in Table 3). Entrepreneurial firms are more productive than non-entrepreneurial firms, are more capital intensive, and are more likely to maintain accounts, have used a computer or internet, and be registered under any Act or Authority (the t-statistics on differences in characteristics between entrepreneurial and non-entrepreneurial firms are statistically significant in all cases). This suggests that entrepreneurial firms are significantly better performing than non-entrepreneurial firms and underscores the importance for household units to expand into non-household units for their further growth.

[Table 3 near here]

3.3 Estimation strategy

We estimate variations of

$$E_{idt} = \alpha_0 + \alpha_1 FIN_{idt} + \alpha_2 Female_{idt} + \sum_{i>0} \gamma_i X_{idt} + \mu_i + \delta_d + \varepsilon_j' \qquad (1)$$

---

[9] We also examine a 'source of financing' constraint, the proportion of outside financing firms receive from 'formal' sources. However, this measure faces a severe selection problem, as about 90 per cent of firms did not receive any financing from outside sources.

[10] We lose 0.5% of the 619,701 observations in our dataset when we follow this condition.

where the dependent variable E_{idt} is a dummy variable for an entrepreneurial firm *i* in district *d* at time *t*. We classify a firm as an entrepreneurial firm if it employs at least one hired worker (besides family workers) and the variable takes the value 1 for such firms.[11] FIN measures a firm's financial constraint, and we use two alternative measures of financial constraint.[12] X_{idt} is a vector of firm-specific control variables. In particular, we control for differences across firms in terms of age, nature of registration, location, assistance towards training and marketing, social group of the owner, and log of labour productivity.[13] The variables μ_i and δ_d control, respectively, for time- and district-specific fixed effects.

3.4 Identification and estimation

Our empirical strategy aims to identify the effect of the finance constraint on entrepreneurship, and whether it differs for men and women. The validity of our analysis rests on the exogeneity of the finance variables. We employ an instrumental variable (IV) approach to allay endogeneity issues and identify causal effects. This requires one or more variables — instruments — that are strongly correlated with the endogenous regressor (financial constraint) and influence the outcome variable (entrepreneurship) through only the endogenous regressor.

We face several possible sources of omitted variable bias. The enterprise owner's decision not to hire outside workers may be due to family environment, motivation, ability, and other unobserved characteristics. For example, the innate ability of the entrepreneur differs across firms. McKenzie and Woodruff (2014) have found this to be a significant positive factor in a firm's success and its ability to employ hired labour (i.e., become entrepreneurial). Another source of selectivity lies in the construction of the null category in our FIN variables, as discussed above. While our control variables help ease some of the endogeneity, they are incomplete. Hence, we need to account for omitted variable bias.

The argument for our IV is strong and straightforward. We rely on RBI (central bank) policy discussed above. The policy the RBI followed was to increase the number of bank branches (or open accounts automatically, or increase the rupees available, etc.) in under-banked districts, where 'under-banked' was defined as applying to districts with a population per branch greater than India's nationwide mean. Since 2001, the RBI has maintained a list of under-banked districts where banks are required to open half of their new branches and are provided with incentives to do so. The policy affects entrepreneurship in the unorganized sector only via its effect on the enterprises' financial constraints. Our instrument is *average population per bank branch* (APPB). Our argument is that geographical access to banking provides better firm access to finance within the same geographic area, the district. We expect firms that are in districts with easier bank access (fewer people per bank branch, for example) to have better access to financing and, therefore, expansion (instrument relevance).[14] Moreover, we believe our instrument meets the necessary exclusion criterion for an IV, as it is only through the firm's financial constraint that it should influence the

---

[11] We have also tried an alternate measure of entrepreneurship where the dependent variable takes the value 1 if the number of hired workers employed by a firm is 2 or more and 0 otherwise. Our results remain unaffected even with this alternate measure of entrepreneurship.

[12] Their construction is discussed in detail in the section on data and variables.

[13] These variables are defined in Table A1 in the online appendix.

[14] Regional (here, district-level) data are useful and often used in constructing IVs to address reverse causality and selection for agents within the region (Dustmann and Preston 2001). This gives us instrument exogeneity, i.e. increased banking availability (etc.) is uncorrelated with the error term. It is better if there is a lag; for this, we rely on the earlier collection dates for the RBI data in comparison with the Survey data.

enterprise's decision to hire outside workers. We test for the suitability of the instrument in our estimations.

We attempt to correct endogeneity by employing the two-stage residual inclusion (2SRI) method (Terza et al. 2008). The conventional two-stage least squares (2SLS) method is the standard approach followed in these circumstances to address endogeneity when employing IV. However, as we have a categorical variable as the dependent variable, employing 2SLS will be susceptible to bias; 2SRI performs better than 2SLS and delivers consistent estimates (Wooldridge 2010). Following Ivlevs et al. (2020) in their analysis of entrepreneurship in former Soviet economies, we first estimate a standard first-stage auxiliary regression in which our instruments and all the control variables are used to explain our potentially endogenous regressor (i.e., FINCON). In the second-stage equation, we include the predicted first-stage residuals, in conjunction with the endogenous regressor.

The unbiased effect of the finance variable on entrepreneurial activity is given by the estimated coefficient of the endogenous regressor in the second stage, while endogeneity bias is captured by the coefficient estimate on the predicted residuals (Ivlevs et al. 2020). With 2SRI, we first estimate regression for the endogenous finance variable using the exogenous regressors (as used in earlier estimations) and the instrument, APPB, as explanatory variables. Our first-stage regression takes the following form:

$$First\ stage: FIN_{idt} = \alpha_0 + \alpha_1 Instruments + \sum_{i>0} \gamma_i X_{idt} + \mu_i + \delta_d + \theta_j \qquad (2)$$

We then retrieve the residuals of first-stage regression and include these residuals as a control variable in the second stage. We can think of these residuals as capturing the part of FIN that is potentially endogenous. Our second-stage regression takes the following form:

$$Second\ stage: E_{idt} = \beta_0 + \beta_1 FIN_{idt} + \gamma\theta_i^{est} + \sum_{i>0} \gamma_i X_{idt} + \mu_i + \delta_d + \varepsilon_j \qquad (3)$$

where variables are defined as above, θ is the first-stage regression's error term and θ_i^{est} its predicted residual, and ε is the second-stage regression's error term.

Ivlevs et al. (2020) point out that a direct test for the regressor of interest's exogeneity is given by the estimated coefficient of the predicted residuals, γ (Bollen et al. 1995). The null hypothesis that the regressor is exogenous is not rejected if γ is not statistically different from 0. If this is the case, non-linear regression (in our case, logit) is preferred.

In this section we have established the core of our approach. In the next sections we discuss the results of bringing data to the equations and considerations, as well as variations in the modelling and a robustness check for our story.

4 Results

This section presents our results, including those from logit and IV estimation methods and other implementations, as discussed in the previous section. Unless otherwise noted, to account for possible non-independence of the error term across districts our estimations employ robust standard errors clustered at the district level. The data comprise the two repeated cross-sections for 2010/11 and 2015/16, discussed earlier.

4.1 Baseline results

Table 4 presents the marginal effects from logit estimations of our measures of finance on the probability that a firm hiring non-family labour.[15] Columns 1–4 are for the 2010/11 wave; Columns 5–8 for the 2015/16 wave; and Columns 9–12 pool both waves, with year controls in the even columns. In all estimations, we include firm-level variables as controls. These controls include the critical categorical variable gender of the owner, as well as location (urban or not), social group of the owner, age of the firm, whether the firm has received government assistance towards training and marketing, whether the firm is registered with some government body, and firm productivity. Even-numbered columns in the entire table include fixed effects (FE) for districts.

[Table 4 near here]

FIN1 in the tables is the direct measure of financial constraints (Columns 1, 2, 5, 6, 9, 10) as discussed earlier, which is similar to the way papers using the WBES capture financial constraints. Notice that the marginal effects for FIN1 carry the expected sign and statistical significance at the 1 percent level. This implies that the financial constraint hinders firm's transitioning from non-entrepreneurial to entrepreneurial. To state it differently, as the constraint weakens, we are more likely to find firms in the part of the informal sector we have labelled 'entrepreneurial'. The marginal effect of FIN1 from the full model for pooled data (column 10) suggests a reduction in the likelihood of firm transition, on average, by 2.5 percent in the presence of financial constraints.

FIN2 also captures the sources of external funding, but here we use dummy variables to capture the types of loans the firm received, with FIN2DUM1, FIN2DUM2, and FIN2DUM3 as categorical variables for firms receiving bank loans, government loans, and other non-government loans, respectively. Firms not receiving any external financing are the reference group. Hence, these estimates are made with our complete sample. Again, in Table 4, relative to the base category of firms with no external financing, the marginal effects are all positive and significant at the 1 per cent level endorsing our earlier finding: more external finding, more entrepreneurial behaviour.

In all model specifications—all 12 columns—an informal firm headed by a woman is more likely to be in the family part of the informal economy: the marginal effects are negative and significant at the 1 per cent level in all specifications. We observe that the probability of a firm being non-entrepreneurial increases to 15 to 20 percent when a woman heads the firm. Our basic logit estimations in Table 4 focus on the role of financial constraints and gender in the distribution of unorganized sector firms across the entrepreneurial–non-entrepreneurial spectrum. The evidence that this offers to us is that alleviating the financial constraint promotes entrepreneurship—firms are willing and able to expand hiring beyond family workers.

We also augment Equation 1 by adding the interaction term $\alpha_3 FIN * Female_{idt}$ to the right hand side, estimating several variants as in Table 4. The results, both the coefficient values and the marginal effects, are available in the online appendix (Tables A3 and A4). The marginal effects on the variables for financing (FIN1and FIN2) and female firm heads continue to tell the same story as in Table 4 in terms of sign and statistical significance. If we use FIN1 as our preferred measure of finance constraints, we find that the interaction term of the owner's gender with FIN1 is not statistically significant. On the other hand, the interaction terms of owner's gender with the two FIN2 dummies are all significant at the 1 per cent level and positive (FIN2DUM1 and FIN2DUM3). As the results on the effect of the finance constraint on male versus female

[15] The coefficient values of these estimations are presented in the appendix Table A2.

entrepreneurship depends on our choice of the measure to capture the finance constraint, we do not have conclusive evidence that women are at a disadvantage in joining the entrepreneurial side.[16]

4.2 Instrumental variable results

Above we discussed endogeneity concerns between entrepreneurship and financial constraints; the constraints may enter in as entrepreneurial firms seek investments to expand, innovate, and stay in business. Because of this, we approach the endogeneity issue using 2SRI, as discussed above. In Table 5, we present the marginal effects for the second-stage reduced-form estimates for one specification with two versions of our financial constraints using our instrumental variables approach.[17] The results mimic, in sign and significance, the logit results discussed in the previous subsection.

However, again we see that this varies over the two alternative ways in which we capture the constraints that firms face regarding financial liquidity. Using FIN1, and referring to Equation 4, θ_i^{est} is the predicted residual from the first-stage equation, and its coefficient, γ, is a direct test for the exogeneity of the regressor of interest. γ is estimated as 0.130 with standard error 0.013 (see Appendix Table A5), suggesting that the regressor is endogenous (coefficient estimates are in the online appendix; in the paper we present marginal estimates. The marginal estimate on FIN1 is -0.117 with standard error 0.010, indicating significance at the 1 per cent level. In comparison, in Table 4, Column 9, the marginal estimate on FIN1 is -0.022 with standard error 0.010.

[Table 5 near here]

For the two-stage procedure, we collapse FIN2DUM1, FIN2DUM2, and FIN2DUM3 into a single dummy variable, FIN2. FIN2 equals 1 if the firm has taken out any sort of formal loan. It is 0 otherwise. Using FIN2, γ is estimated as 0.073 with standard error 0.010, suggesting that the regressor is endogenous. The marginal estimate on FIN2 is 0.036 with standard error 0.006, indicating significance at the 1 per cent level. Because in Table 5 we used three dummy variables to capture the loan sources of the firm and here we collapsed them to one, we cannot directly compare the coefficients. However, at a glance they look similar.

The two-stage procedure we employed shows that it is necessary to account for the endogeneity of enterprise financial constraints. However, doing so does not change the sign or significance of our estimates. In other words, the 2SRI estimates reinforce the main findings arrived at using the logit estimations. The coefficients of FIN1and FIN2 retain the same sign and significance, suggesting that lack of access to finance is a serious impediment to the transition of firms from household firms to non-household firms employing hired labour.

---

[16] We also estimated regressions using separate samples of male and female owners for each wave and for the waves pooled with one another. Our results endorse what we have observed for the pooled sample of male and female owners. One possible concern here is that the gender of the owner of the firm may be endogenous with regard to entrepreneurship, if the male members of the more successful firms take over the management or ownership of the firm when it switches from being a non-entrepreneurial firm to an entrepreneurial firm. However, we find that the industries where female owned firms are largely found are very different from the industries where male owed firms dominate – out of the 70 industries in our sample, 40 per cent of female owned firms are in only four of these industries. This suggests that there are specific attributes of these industries which make it more likely for female owned firms to be present, and the gender of the firm owner is unlikely to be endogenously determined.

[17] Coefficient values are presented in Table A5.

[Table 6 near here]

We also perform the 2SRI estimation using separate samples of male and female owners. The results with the marginal effects are reported in Table 6.[18] They show the same pattern observed for the pooled sample. As for our logit results, we obtain somewhat ambiguous results on whether finance constraints matter more for women than for men. For both FIN1 and FIN2 the finance constraint is binding for male and female owners alike, implying that as the finance constraint weakens, both male-owned and female-owned firms are willing and able to hire workers from outside the family. However, for FIN1 the marginal effect on the finance constraint is smaller (more negative) for women than for men; while for FIN2 the marginal effect on the finance constraint is smaller (less positive) for women than for men, indicating that the constraint is more binding for male owners than for female owners.

5 Robustness test: PSM-DID results

In this section, we outline a critical robustness test. One possible concern with our IV approach is that the instrument—APPB—may not meet the excludability condition requiring that the decision of banks to open branches in a specific district affects the likelihood of entrepreneurship only through this channel. Under the 2005 reforms, banks could choose an extensive and intensive level of entry in under-banked districts as well as the total expansion of their branch network (Young 2019). This implies the decision to open more branches (per capita) in some districts than in others could have been influenced by (unobserved) district characteristics, which may also influence business performance and the likelihood of entrepreneurship.

In order to address the possibility that the roll-out of banking services in under-banked districts was not random, we supplement our main identification strategy with an alternate approach, combining PSM and DID. We take advantage of the fact that while banks could choose which under-banked district to open new branches in, they were constrained by the RBI's policy to open branches in under-banked districts in order to receive licences for entry in the rich markets (Young 2019). Over the period of our analysis, the RBI vigorously pursued its expansion policy by increasing the number of branches, especially in areas that it deemed under-banked. Further details of our PSM-DID strategy are provided in the online appendix. Our PSM-DID estimates confirm the finding from our IV strategy. Our results clearly point to the positive effect of policy change on entrepreneurship.

[18] Coefficient values are presented in Table A6.

6 Conclusions

We began our paper by laying out two questions: (1) Does informal economic activity become more entrepreneurial in response to increased financial inclusion? (2) Does access to new financing options change the gender configuration of informal economic activity and, if so, in what ways and what directions? We picture the informal sector as composed of two firm types: family firms and entrepreneurial firms. The distinction is that entrepreneurial firms employ outside non-family labourers—that is, hired workers. We examine the impact of financial inclusion, partly captured by banking access, on proprietorship in entrepreneurial informal firms and its consequences, focusing on the gender differences in these impacts. The context of our study is India during the 2010s, a period during which banking policies greatly expanded banking access to women and to the unbanked.

We capture financial constraints using two core explanatory variables: (1) information self-reported by firms facing finance constraints (similar to the variable employed by studies making use of the WBES); (2) a set of categorical variables on whether the firm obtained bank loans, government loans, non-government loans, or no loans (the omitted variable). While one could dispute the degree to which each of these core explanatory variables face the problem, there is an element of them that is clearly potentially endogenous. This means that our logit estimation may not produce the true causal relationship between an enterprise's financial constraints and its entrepreneurship.

We address the endogeneity potential by implementing an IV approach, relying on the idea that an increase in the number of bank branches is correlated with the difficulty of obtaining financing: briefly, more bank branches in the district where you are living means more financial access. Under-banked districts have less access; during the 2010s, policy strove to increase the number of branches. Our principal banking access measure, i.e. our IV, is whether the district had a population per branch below (banked) or above (under-banked) the national average. This is the indicator the RBI used in implementing its policy to increase branches in under-banked districts. When the IV affects the likelihood of entrepreneurship only through our financial constraint measures, then we say that we can make causal inferences. We also supplement our IV analysis with a robustness check using PSM-DID and obtain similar results as we did with our IV estimation strategy.

Our results show a strong potential role for increased liquidity. Whichever measure of finance constraint we use, we find clear evidence that finance constraints matter for the likelihood of becoming an entrepreneur in the informal sector in India. The results are robust to concerns about reverse causality. When we use IV estimation using 2SRI or we use PSM-DID, we obtain similar results to when we use a logit estimation method. However, we find less-conclusive evidence that effective financing encourages female entrepreneurship more than male entrepreneurship. Our results here are sensitive to the choice of the finance constraint that we use as our explanatory variable, in both the logit and the 2SRI results. This does not mean that finance constraints do not matter for women entrepreneurs; instead, our results imply that financial inclusion matters for *both* women and men entrepreneurs.

Our findings have strong implications for policy. In the 2000s, the Indian government initiated an ambitious set of reforms with the objective of ensuring that the areas of the country which historically had not had much access to banking services would be able to able to get access to formal financial institutions. At the same time, India has an endemic problem of a large informal sector, mostly populated by micro household (own-account) enterprises that remain largely unproductive (Raj and Sen 2016). Our findings suggest that the policy actions of the Indian government to increase access to finance in India's under-banked districts succeeded in one important dimension—it contributed to the growth of entrepreneurship in India's informal sector, enabling many of the self-employed to become employers, hiring outside workers.

References

Aterido, R., T. Beck, and L. Iacovone (2013). 'Access to Finance in Sub-Saharan Africa: Is There a Gender Gap?' *World Development*, 47: 102–20. https://doi.org/10.1016/j.worlddev.2013.02.013

Banerji, S., R.S. Raj, and K. Sen (2016). 'Monitoring Costs, Credit Constraints and Entrepreneurship'. *The Manchester School*, *84*(5), 573–99. https://doi.org/10.1111/manc.12122

Beck, T., A. Demirgüç-Kunt, L. Laeven, and R. Levine (2008). 'Finance, Firm Size, and Growth'. *Journal of Money, Credit and Banking*, 40(7):1379–405. https://doi.org/10.1111/j.1538-4616.2008.00164.x

Bollen K.A., D.K. Guilkey, and T.A. Mroz (1995). 'Binary Outcomes and Endogenous Explanatory Variables: Tests and Solutions with an Application to the Demand for Contraceptive Use in Tunisia'. *Demography*, 32(1): 111–31. https://doi.org/10.2307/2061900

Brown, J.D., J.S. Earle, and D. Lup (2005). 'What Makes Small Firms Grow? Finance, Human Capital, Technical Assistance, and the Business Environment in Romania'. *Economic Development and Cultural Change*, 54(1): 33–70. https://doi.org/10.1086/431264

Bruhn, M. (2009). 'Female-Owned Firms in Latin America: Characteristics, Performance, and Obstacles to Growth'. Policy Research Working Paper 5122. Washington, DC: The World Bank. https://doi.org/10.1596/1813-9450-5122

Census of India (2001). *Primary Census Abstract*. New Delhi: Registrar General of India.

Census of India (2011). *Primary Census Abstract*. New Delhi: Registrar General of India.

Chavan, P. (2020). 'Women's Access to Banking in India: Policy Context, Trends, and Predictors'. *Review of Agrarian Studies*, 10(1): 7–36.

Chen, M.A. (2012). 'The Informal Economy: Definitions, Theories and Policies'. WIEGO Working Paper 1. Cambridge, MA: Women in Informal Employment Globalizing and Organizing.

Daniels, L., and D.C. Mead (1998). 'The Contribution of Small Enterprises to Household and National Income in Kenya'. *Economic Development and Cultural Change*, 47(1): 45–71. https://doi.org/10.1086/452386

De Vreyer, P., and F. Roubaud (eds) (2013). *Urban Labor Markets in sub-Saharan Africa*. Washington, DC: World Bank.

Dustmann, C., and I. Preston (2001). 'Attitudes to Ethnic Minorities, Ethnic Context and Location Decisions'. *The Economic Journal*, 111(470): 353–73. https://doi.org/10.1111/1468-0297.00611

Earle, J.S., and Z. Sakova (2000). 'Business Start-ups or Disguised Unemployment? Evidence on the Character of Self-Employment from Transition Economies'. *Labour economics*, 7(5): 575–601. https://doi.org/10.1016/S0927-5371(00)00014-2

Fields, G.S. (1990). 'Labour Market Modelling and the Urban Informal Sector: Theory and Evidence'. In D. Turnham, B. Salomé, and A. Schwarz (eds), *The Informal Sector Revisited*. Paris: Organisation for Economic Co-operation and Development.

Fields, G.S. (2019). *Employment and Development: How Work Can Lead from and into Poverty*. Oxford: Oxford University Press. https://doi.org/10.1093/oso/9780198815501.001.0001

Gang, I.N., and M. Pandey (2007). 'Small Scale Industry'. In K. Basu (ed.), *Oxford Companion to Economics in India*. Delhi: Oxford University Press.

Gang, I.N. (1992). 'Small Firm "Presence" in Indian Manufacturing'. *World Development*, 20(9): 1377–89. https://doi.org/10.1016/0305-750X(92)90085-A

Gindling, T.H., and D. Newhouse (2014). 'Self-Employment in the Developing World'. *World Development*, 56: 313–31. https://doi.org/10.1016/j.worlddev.2013.03.003

Grimm, M., P. Knorringa, and J. Lay (2012). 'Constrained Gazelles: High Potentials in West Africa's Informal Economy'. *World Development*, 40(7): 1352–68. https://doi.org/10.1016/j.worlddev.2012.03.009

Ivlevs, A., M. Nikolova, and O. Popova (2020). 'Former Communist Party Membership and Present-Day Entrepreneurship'. *Small Business Economics*, 1–18. https://doi.org/10.1007/s11187-020-00364-6

Jennings, J.E., and C.G. Brush (2013). 'Research on Women Entrepreneurs: Challenges to (and from) the Broader Entrepreneurship Literature?', *Academy of Management Annals*, 7(1): 661–713. https://doi.org/10.1080/19416520.2013.782190

Johnson, S., J. McMillan, and C. Woodruff (2002). 'Property Rights and Finance'. *American Economic Review*, 92(5): 1335–56. https://doi.org/10.1257/000282802762024539

Kerr, W.R., and R. Nanda (2011). 'Financing Constraints and Entrepreneurship'. In D. Audretsch, O. Falck, S. Heblich, and A. Lederer (eds), *Handbook of Research on Innovation and Entrepreneurship*, S. 88–103. Cheltenham: Edward Elgar Publishing. https://doi.org/10.4337/9781849807760.00015

Landes, D.S. (1969). *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge: Cambridge University Press.

Maloney, W.F. (1999). 'Does Informality Imply Segmentation in Urban Labor Markets? Evidence from Sectoral Transitions in Mexico'. *World Bank Economic Review*, 13(2): 275–302. https://doi.org/10.1093/wber/13.2.275

McKenzie, D., and C. Woodruff (2014). 'What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?' *World Bank Research Observer*, 29(1): 48–82. https://doi.org/10.1093/wbro/lkt007

Mead, D.M., and C. Liedholm (1998). 'The Dynamics of Micro and Small Enterprises in Developing Countries'. *World Development*, 26(1): 61–74. https://doi.org/10.1016/S0305-750X(97)10010-9

NSSO (National Sample Survey Office) (2013). *Processed Data on Survey of Unincorporated Non-Agricultural Enterprises (Excluding Construction) in India, NSS 67th Round (July 2010 – June 2011).* New Delhi: NSSO, Ministry of Statistics and Programme Implementation, Government of India.

NSSO (2017). *Processed Data on Survey of Unincorporated Non-Agricultural Enterprises (Excluding Construction) in India, NSS 73rd Round (July 2015 – June 2016).* New Delhi: NSSO, Ministry of Statistics and Programme Implementation, Government of India.

Naudé, W. (2010). 'Entrepreneurship, Developing Countries, and Development Economics: New Approaches and Insights'. *Small Business Economics*, 34(1): 1–12. https://doi.org/10.1007/s11187-009-9198-2

Nichter, S., and L. Goldmark (2009). 'Small Firm Growth in Developing Countries'. *World Development*, 37(9): 1453–64. https://doi.org/10.1016/j.worlddev.2009.01.013

Nikolova, M. (2019). 'Switching to Self-Employment Can Be Good for Your Health'. *Journal of Business Venturing*, 34(4): 664–91. https://doi.org/10.1016/j.jbusvent.2018.09.001

Ojah, K., T. Gwatidzo, and S. Kaniki (2010). 'Legal Environment, Finance Channels and Investment: The East African Example'. *Journal of Development Studies*, 46(4): 724–44. https://doi.org/10.1080/00220380903012722

Oliveira, B., and A. Fortunato (2006). 'Firm Growth and Liquidity Constraints'. *Small Business Economics*, 27(2): 139–56. https:/doi.org/10.1007/s11187-006-0006-y

Pradhan, S. (2016). 'Why You Should Care About New Boundaries for Districts in India'. Humans of Data, 8 January. Available at: https://humansofdata.atlan.com/2016/01/boundary-changes-districts-in-india (accessed 17 June 2020).

Presbitero, A.F., P. Rabellotti, and C. Piras (2014). 'Barking Up the Wrong Tree? Measuring Gender Gaps in Firms' Access to Finance'. *Journal of Development Studies*, 50(10): 1430–44. https://doi.org/10.1080/00220388.2014.940914

Raj, R.S.N., and K. Sen (2016). *Out of the Shadows? The Informal Manufacturing in Post-Reform India*. Delhi: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199460847.001.0001

RBI (2011). *Basic Statistical Returns of Scheduled Commercial Banks in India*, volume 40. Mumbai: RBI. Available at: https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=Basic+Statistical+Returns (accessed 10 October 2019).

RBI (2016). *Basic Statistical Returns of Scheduled Commercial Banks in India*, volume 45. Mumbai: RBI. Available at: https://www.rbi.org.in/Scripts/AnnualPublications.aspx?head=Basic+Statistical+Returns (accessed 10 October 2019).

Rijkers, B., M. Söderbom, and J.L. Loening (2010). 'A Rural–Urban Comparison of Manufacturing Enterprise Performance in Ethiopia'. *World Development*, 38(9):1278–96. https://doi.org/10.1016/j.worlddev.2010.02.010

Storey, D.J. (2004). 'Racial and Gender Discrimination in the Micro Firms Credit Market?: Evidence from Trinidad and Tobago'. *Small Business Economics*, 23: 401–22. https://doi.org/10.1007/s11187-004-7259-0

Terza, J.V., A. Basu, and P.J. Rathouz (2008). 'Two-Stage Residual Inclusion Estimation: Addressing Endogeneity in Health Econometric Modeling'. *Journal of Health Economics*, 27(3): 531–43. https://doi.org/10.1177/1536867X1801700409

Van Praag, C.M., and P.H. Versloot (2007). 'What Is the Value of Entrepreneurship? A Review of Recent Research'. *Small Business Economics*, 29(4): 351–82. https://doi.org/10.1007/s11187-007-9074-x

Winker, P. (1999). 'Causes and Effects of Financing Constraints at the Firm Level'. *Small Business Economics*, 12(2):169–81. https://doi.org/10.1023/A:1008035826914

Woodruff, C., and R. Zenteno (2007). 'Migration Networks and Microenterprises in Mexico'. *Journal of Development Economics*, 82(2): 509–28. https://doi.org/10.1016/j.jdeveco.2006.03.006

Wooldridge, J.M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

World Bank (2013). *World Development Report 2013: Jobs*. Washington, DC: The World Bank. https://doi.org/10.1596/978-0-8213-9575-2

Young, N. (2019). 'Banking and Growth: Evidence from a Regression Discontinuity Analysis: Online Appendix'. Available at: https://nateyoungecon.com/wp-content/uploads/2019/01/NYoung_Banking_and_Growth.pdf (accessed 13 December 2019).

Table 1: Number of banked and under-banked districts in India, 2010/11–2015/16

Districts	2010/11	2015/16
Banked	207	218
Under-banked	355	344
National average of population per bank branch	13,027	8,683

Note: 562 districts in each year; if district population per branch > national average, the district is under-banked.

Source: Authors' construction based on RBI (2011, 2016).

Table 2: Key Explanatory Variables and their construction

Variables	Description
Dependent variable	
Entrepreneurial firm (E)	Binary variable for firms that employ at least one hired worker, besides family workers
Independent variables	
Access to finance	
FIN1	Binary variable for firms that faced any borrowing constraint in the last year
FIN2 DUM1	Dummy variable for firms with bank loan
FIN2 DUM2	Dummy variable for firms with non-government loan
FIN2 DUM3	Dummy variable for firms with government loan
FIN2	Dummy variable for firms with any type of loan
Gender	
Female	Binary variable for female-run firms

Source: Authors' construction based on Census of India (2001, 2011) and NSSO (2013, 2017).

Table 3: Cross-tabulations: entrepreneurship characteristics

Panel A: entrepreneurship vs gender of owner

Owner	Owner	Entrepreneurship		Total
		Non-entrepreneurial firm	Entrepreneurial firm	
Male-run firms	Frequency	295,536	213,130	508,666
	Row percentage	58.1	41.9	100.0
	Column percentage	82.5	94.5	87.1
	Cell percentage	50.6	36.5	87.1
Female-run firms	Frequency	62,890	12,499	75,389
	Row percentage	83.4	16.6	100.0
	Column percentage	17.6	5.5	12.9
	Cell percentage	10.8	2.1	12.9
Total	Frequency	358,426	225,629	584,055
	Row percentage	61.4	38.6	100.0
	Column percentage	100.0	100.0	100.0
	Cell percentage	61.4	38.6	100.0

Panel B: entrepreneurship vs financial constraint

Financial constraint	Owner	Entrepreneurship		Total
		Non-entrepreneurial firm	Entrepreneurial firm	
No	Frequency	331,458	212,215	543,673
	Row percentage	61.0	39.0	100.0
	Column percentage	92.5	94.1	93.1
	Cell percentage	56.8	36.3	93.1
Yes	Frequency	26,968	13,414	40,382
	Row percentage	66.8	33.2	100.0
	Column percentage	7.5	6.0	6.9
	Cell percentage	4.6	2.3	6.9
Total	Frequency	358,426	225,629	584,055
	Row percentage	61.4	38.6	100.0
	Column percentage	100.0	100.0	100.0
	Cell percentage	61.4	38.6	100.0

Panel C: Firm Characteristics by entrepreneurship

Characteristics	Entrepreneurship		t-test
	Non-entrepreneurial firm	Entrepreneurial firm	
lnLP	9.788	10.388	0.600*** (0.003)
Firms that maintained accounts	0.053	0.286	0.233*** (0.001)
Firms that used computer	0.021	0.145	0.124*** (0.001)
Firms that used internet	0.014	0.103	0.090*** (0.001)
Regis	0.259	0.626	0.367*** (0.001)
Mean Land and Building to Employment (in logs)	10.607	11.233	0.626*** (0.004)
Mean Plant and Machinery to Employment (in logs)	7.973	8.486	0.513*** (0.007)
Number of Firms	358,426	225,629	

Source: Authors' estimates.

Table 4: Access to finance and entrepreneurship: Marginal Effects (dependent variable: entrepreneurial firm or not)

Variables	**2010-2011**				**2015-2016**				**Pooled Data**			
	FIN1		FIN2		FIN1		FIN2		FIN1		FIN2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
FIN1	-0.035*** (0.004)	-0.031*** (0.004)			-0.017*** (0.004)	-0.022*** (0.004)			-0.022*** (0.003)	-0.025*** (0.003)		
FIN2 DUM1			0.113*** (0.004)	0.121*** (0.004)			0.128*** (0.005)	0.128*** (0.005)			0.121*** (0.003)	0.126*** (0.003)
FIN2 DUM2			0.066*** (0.004)	0.051*** (0.004)			0.060*** (0.004)	0.045*** (0.004)			0.068*** (0.003)	0.050*** (0.003)
FIN2 DUM3			0.107*** (0.013)	0.125*** (0.013)			0.155*** (0.018)	0.152*** (0.018)			0.125*** (0.011)	0.143*** (0.011)
Female	-0.173*** (0.003)	-0.172*** (0.003)	-0.173*** (0.003)	-0.171*** (0.003)	-0.197*** (0.004)	-0.204*** (0.004)	-0.197*** (0.004)	-0.203*** (0.004)	-0.179*** (0.002)	-0.190*** (0.002)	-0.180*** (0.002)	-0.189*** (0.002)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	N	N	N	N	N	N	N	N	N	Y	N	Y
District FE	No	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	313586	313586	313586	313586	270469	270448	270469	270448	584055	584055	584055	584055

Note: Robust standard errors are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' estimates.

Table 5: Access to finance and entrepreneurship: Marginal Effects (2SRI approach: no. of replications: 500)

Variables	FIN1	FIN2
FIN1	−0.117***	
	(0.010)	
FIN2		0.036***
		(0.006)
Female	−0.152***	−0.148***
	(0.002)	(0.002)
Location	0.041***	0.044***
	(0.001)	(0.001)
ST	−0.117***	−0.114***
	(0.003)	(0.003)
SC	−0.140***	−0.142***
	(0.002)	(0.002)
OBC	−0.044***	−0.045***
	(0.001)	(0.001)
Age3–9	−0.071***	−0.069***
	(0.002)	(0.002)
Age>9	−0.065***	−0.063***
	(0.002)	(0.002)
Asst	0.144***	0.123***
	(0.005)	(0.006)
Regis	0.225***	0.224***
	(0.001)	(0.001)
lnLP	0.089***	0.088***
	(0.001)	(0.001)
XUhat	0.025***	0.014***
	(0.002)	(0.002)
Observations	577,558	577,558

Notes: Figures in parentheses are bootstrapped standard errors; in this table we collapse FIN2DUM1, FIN2DUM2 and FIN2DUM3 into a single dummy variable, FIN2: FIN2 equals 1 if the firm has taken out any sort of formal loan, 0 otherwise; *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' estimates.

Table 6: Access to finance and entrepreneurship by gender: Marginal Effects (2SRI approach: no. of replications: 500)

Variables	**Male-run firms**		**Female-run firms**	
	FIN1	**FIN2**	**FIN1**	**FIN2**
FIN1	−0.136*** (0.011)		-0.172*** (0.030)	
FIN2		0.048*** (0.007)		0.026** (0.011)
XUhat	0.029*** (0.003)	0.012*** (0.002)	0.036** (0.006)	0.004 (0.003)
Controls	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Observations	503,560	503,560	73,998	73,998

Notes: Figures in parentheses are bootstrapped standard errors; in this table we collapse FIN2DUM1, FIN2DUM2 and FIN2DUM3 into a single dummy variable, FIN2: FIN2 equals 1 if the firm has taken out any sort of formal loan, 0 otherwise; *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' estimates.

Online Appendix

Table A1: Summary statistics, variables and their construction

Part A: Summary Statistics

Variable	Observations	Mean	SD	Min.	Max.
Entrepreneurial firm (E)	584,055	0.38631	0.48690	0	1
FIN1	584,055	0.06914	0.25369	0	1
FIN2 DUM1	584,055	0.05354	0.22511	0	1
FIN2 DUM2	584,055	0.06156	0.24035	0	1
FIN2 DUM3	584,055	0.00468	0.06828	0	1
Female	584,055	0.12908	0.33529	0	1
Location	584,055	0.51206	0.49986	0	1
ST	584,055	0.05574	0.22943	0	1
SC	584,055	0.10004	0.30005	0	1
OBC	584,055	0.46240	0.49858	0	1
Age3–9	584,055	0.44885	0.49738	0	1
Age>9	584,055	0.42633	0.49454	0	1
Asst	584,055	0.01196	0.10870	0	1
Regis	584,055	0.40060	0.49002	0	1
lnLP	584,055	10.01993	0.98134	2.10843	15.9002

Part B: Variables and their construction

Variables	Description	Source
Dependent variable		
Entrepreneurial firm (E)	Binary variable, firms employing at least one hired worker, besides family workers	NSSO data
Independent variables		
Access to finance		
FIN1	Binary variable for firms that faced any borrowing constraint in the last year	NSSO data
FIN2 DUM1	Dummy variable for firms with bank loan	NSSO data
FIN2 DUM2	Dummy variable for firms with non-government loan	NSSO data
FIN2 DUM3	Dummy variable for firms with government loan	NSSO data
FIN2	Dummy variable for firms with any type of loan	NSSO data
Gender and interaction terms		
Female	Binary variable for female-run firms	NSSO data
FIN1*Female	Interaction between FIN1 and Female	NSSO data
FIN2DUM1 *Female	Interaction between FIN2DUM1 and Female	NSSO data
FIN2DUM2 *Female	Interaction between FIN2DUM2 and Female	NSSO data
FIN2DUM3 *Female	Interaction between FIN2DUM3 and Female	NSSO data
Control variables: firm characteristics		
Location	Dummy variable for urban firms	NSSO data
ST	Dummy variable for Scheduled Tribe (ST)-owned firms	NSSO data
SC	Dummy variable for Scheduled Caste (SC)-owned firms	NSSO data
OBC	Dummy variable for Other Backward Communities (OBC)-owned firms	NSSO data
Age3–9	Dummy variable for firms aged between 3 and 9 years	NSSO data
Age>9	Dummy variable for firms that completed more than 9 years	NSSO data
Asst	Dummy variable for firms that received any government assistance during last three years	NSSO data
Regis	Dummy variable for firms that registered under any one of the Shops and Establishment Act, Municipal Corporation/Panchayats/Local Body, Vat/Sales Tax Act, Provident Fund Act, or Employees State Insurance Corporation Act, or with the SEBI/Stock Exchange or any other industry-specific Act/authority	NSSO data
lnLP	Log of labour productivity, where labour productivity is defined as the ratio of gross value added to employment	NSSO data
District-level variables		
SHSCPOP	Proportion of SC population in total population	Census 2001
SHSTPOP	Proportion of ST population in total population	Census 2001
MIDGRADEDU	Proportion of individuals educated at secondary level and above	Census 2001
ROADVILLG	Share of villages with paved approach road in total villages	Census 2001
ELECVILLG	Share of electrified villages in total villages	Census 2001
POSTVILLG	Share of villages with post office in total villages	Census 2001
BUSVILLG	Proportion of villages situated on a bus route in total villages	Census 2001
PRIMSCHVILLG	Proportion of villages with at least a primary school in total villages	Census 2001

Source: Authors' construction based on Census of India (2001, 2011), NSSO (2013, 2017).

Table A2: Access to finance and entrepreneurship: logistic regression coefficient estimates with control variables (dependent variable: entrepreneurial firm or not)

Variables	**2010-2011**				**2015-2016**				**Pooled Data**			
	FIN1		FIN2		FIN1		FIN2		FIN1		FIN2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
FIN1	-0.164*** (0.018)	-0.149*** (0.019)			-0.068*** (0.017)	-0.088*** (0.018)			-0.097*** (0.012)	-0.111*** (0.013)		
FIN2 DUM1			0.533*** (0.020)	0.579*** (0.021)			0.522*** (0.019)	0.524*** (0.020)			0.527*** (0.014)	0.553*** (0.014)
FIN2 DUM2			0.312*** (0.019)	0.244*** (0.020)			0.245*** (0.016)	0.185*** (0.017)			0.297*** (0.012)	0.217*** (0.013)
FIN2 DUM3			0.506*** (0.061)	0.596*** (0.062)			0.633*** (0.073)	0.623*** (0.074)			0.543*** (0.046)	0.627*** (0.047)
Female	-0.817*** (0.015)	-0.825*** (0.016)	-0.816*** (0.015)	-0.820*** (0.016)	-0.807*** (0.014)	-0.837*** (0.015)	-0.807*** (0.014)	-0.831*** (0.015)	-0.781*** (0.010)	-0.833*** (0.011)	-0.781*** (0.010)	-0.826*** (0.011)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	N	N	N	N	N	N	N	N	N	Y	N	Y
District FE	No	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	313586	313586	313586	313586	270469	270448	270471	270448	584055	584055	584055	584055
Pseudo R2	0.145	0.169	0.148	0.171	0.147	0.172	0.151	0.174	0.151	0.171	0.154	0.173

Note: Robust standard errors are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' estimates.

Table A3: Access to finance and entrepreneurship: logistic regression coefficient estimates with control variables and interaction terms (dependent variable: entrepreneurial firm or not)

Variable	**2010–11**				**2015–16**				**Pooled data**			
	FIN1		FIN2		FIN1		FIN2		FIN1		FIN2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
FIN1	−0.166*** (0.018)	−0.149*** (0.020)			−0.070*** (0.017)	−0.089*** (0.018)			−0.099*** (0.012)	−0.113*** (0.013)		
FIN2 DUM1			0.499*** (0.021)	0.547*** (0.021)			0.488*** (0.020)	0.491*** (0.021)			0.493*** (0.015)	0.520*** (0.015)
FIN2 DUM2			0.310*** (0.020)	0.243*** (0.021)			0.246*** (0.017)	0.187*** (0.018)			0.298*** (0.013)	0.219*** (0.013)
FIN2 DUM3			0.459*** (0.063)	0.557*** (0.064)			0.566*** (0.075)	0.556*** (0.076)			0.488*** (0.048)	0.576*** (0.048)
Female	−0.818*** (0.016)	−0.826*** (0.016)	−0.845*** (0.016)	−0.847*** (0.017)	−0.809*** (0.015)	−0.838*** (0.015)	−0.830*** (0.015)	−0.853*** (0.016)	−0.782*** (0.011)	−0.834*** (0.011)	−0.806*** (0.011)	−0.851*** (0.012)
FIN1*Female	0.032 (0.078)	0.011 (0.079)			0.029 (0.064)	0.019 (0.064)			0.036 (0.049)	0.025 (0.050)		
FIN2DUM1 *Female			0.522*** (0.069)	0.499*** (0.071)			0.409*** (0.064)	0.408*** (0.065)			0.455*** (0.047)	0.454*** (0.048)
FIN2DUM2 *Female			0.022 (0.075)	0.013 (0.077)			0.032 (0.060)	−0.047 (0.063)			−0.025 (0.047)	−0.043 (0.048)
FIN2DUM3 *Female			0.634*** (0.201)	0.538*** (0.211)			0.774*** (0.229)	0.760*** (0.233)			0.681*** (0.151)	0.639*** (0.154)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	N	N	N	N	N	N	N	N-	N	Y	N	Y
District FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	313,586	313,586	313,586	313,586	270,469	270,448	270,469	270,448	584,055	584,055	584,055	584,055
Pseudo R2	0.145	0.169	0.148	0.171	0.148	0.172	0.151	0.174	0.151	0.171	0.154	0.173

Note: controls include gender, location, dummies for social group (ST, SC, and OBC), age categories (Age3to9 and Age>9), assistance, registration status, and labour productivity; robust standard errors are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.
Source: Authors' estimates.

Table A4: Access to finance and entrepreneurship: Marginal effects and interaction effects (dependent variable: entrepreneurial firm or not)

Variable	2010–11				2015–16				Pooled data			
	FIN1		FIN2		FIN1		FIN2		FIN1		FIN2	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
FIN1	−0.033*** (0.004)	−0.030*** (0.004)			−0.016*** (0.004)	−0.021*** (0.004)			−0.022*** (0.003)	−0.025*** (0.003)		
FIN2 DUM1			0.130*** (0.005)	0.140*** (0.005)			0.134*** (0.005)	0.135*** (0.005)			0.133*** (0.003)	0.140*** (0.004)
FIN2 DUM2			0.070*** (0.005)	0.053*** (0.005)			0.060*** (0.004)	0.045*** (0.004)			0.070*** (0.003)	0.050*** (0.003)
FIN2 DUM3			0.125*** (0.015)	0.144*** (0.015)			0.165*** (0.017)	0.162*** (0.018)			0.140*** (0.011)	0.160*** (0.012)
Female	−0.150*** (0.002)	−0.149*** (0.002)	−0.150*** (0.002)	−0.149*** (0.002)	−0.183*** (0.003)	−0.188*** (0.003)	−0.182*** (0.003)	−0.187*** (0.003)	−0.161*** (0.002)	−0.169*** (0.002)	−0.161*** (0.002)	−0.168*** (0.002)
FIN1 *Female	0.012 (0.011)	0.008 (0.011)			0.007 (0.011)	0.006 (0.011)			0.009 (0.008)	0.008 (0.007)		
FIN2DUM1 *Female			0.077*** (0.014)	0.071*** (0.014)			0.074*** (0.013)	0.071*** (0.013)			0.074*** (0.009)	0.071*** (0.009)
FIN2DUM2 *Female			-0.010 (0.012)	-0.008 (0.012)			-0.012 (0.011)	−0.013 (0.011)			−0.014* (0.008)	−0.015* (0.048)
FIN2DUM3 *Female			0.099*** (0.040)	0.078*** (0.040)			0.148*** (0.047)	0.140*** (0.046)			0.119*** (0.031)	0.107*** (0.030)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	N	N	N	N	N	N	N	N-	N	Y	N	Y
District FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	313,586	313,586	313,586	313,586	270,469	270,448	270,469	270,448	584,055	584,055	584,055	584,055
Pseudo R2	0.145	0.169	0.148	0.172	0.148	0.172	0.151	0.174	0.151	0.171	0.154	0.173

Note: controls include gender, location, dummies for social group (ST, SC, and OBC), age categories (Age3to9 and Age>9), assistance, registration status, and labour productivity; robust standard errors are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1. In the table, we report the mean of the interaction effects.
Source: Authors' estimates.

Table A5: Access to finance and entrepreneurship: Coefficient Values (2SRI approach: no. of replications: 500)

Variables	**FIN1**	**FIN2**
FIN1	−0.609*** (0.052)	
FIN2		0.187*** (0.032)
Female	−0.790*** (0.011)	−0.772*** (0.011)
Location	0.211*** (0.006)	0.228*** (0.007)
ST	−0.609*** (0.015)	−0.594*** (0.015)
SC	−0.727*** (0.011)	−0.741*** (0.012)
OBC	−0.230*** (0.006)	−0.237*** (0.007)
Age3–9	−0.367*** (0.010)	−0.361*** (0.010)
Age>9	−0.335*** (0.010)	−0.326*** (0.010)
Asst	0.747*** (0.028)	0.640*** (0.033)
Regis	1.167*** (0.006)	1.169*** (0.007)
lnLP	0.460*** (0.004)	0.458*** (0.004)
XUhat	0.130*** (0.013)	0.073*** (0.010)
Observations	577,558	577,558

Notes: Figures in parentheses are bootstrapped standard errors; in this table we collapse FIN2DUM1, FIN2DUM2, and FIN2DUM3 into a single dummy variable, FIN2: FIN2 equals 1 if the firm has taken out any sort of formal loan, 0 otherwise; *** p<0.01, ** p<0.05, * p<0.1.

Source: Authors' estimates.

Table A6: Access to finance and entrepreneurship by gender: Coefficient Values (2SRI approach: no. of replications: 500)

Variables	**Male-run firms**		**Female-run firms**	
	FIN1	FIN2	FIN1	FIN2
FIN1	−0.658*** (0.054)		-1.839*** (0.325)	
FIN2		0.231*** (0.036)		0.277** (0.122)
XUhat	0.141*** (0.014)	0.061*** (0.011)	0.382*** (0.066)	0.038 (0.033)
Controls	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Observations	503,560	503,560	73,998	73,998

Notes: Figures in parentheses are bootstrapped standard errors; in this table we collapse FIN2DUM1, FIN2DUM2, and FIN2DUM3 into a single dummy variable, FIN2: FIN2 equals 1 if the firm has taken out any sort of formal loan, 0 otherwise; *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Source: Authors' estimates.

Appendix to Section 5: Robustness check using: PSM-DID

For our DID strategy, we take advantage of the rule that regulators employed to select the set of under-banked districts under the 2005 reform of bank branch licensing in India. This rule compares the average number of persons per branch in a district against a statistic termed the 'national average' of population per branch for India (RBI 2009; Young 2019). This national average acts as a threshold, and the districts whose populations per branch exceeds this threshold receive treatment while others do not. The reform led to additional branch expansion over 2010–15 in some districts that were initially under-banked in 2010/11, so that they became banked by 2015/16. Our empirical strategy lies in examining whether informal firms show a greater propensity to become entrepreneurial in the districts that received the treatment, that is, the districts that changed status from under-banked to banked during the period of our analysis, relative to untreated districts, which remain under-banked throughout the period of our analysis.[19] In the period 2010–15, 22 districts changed from under-banked to banked, while 333 districts remained under-banked. Our unit of analysis is the firm, and we compare the entrepreneurship status of firms in the 22 treated districts relative to the 333 control districts.

A limitation of our DID strategy is that we do not have data on firm entrepreneurship status at the district level prior to 2010. Because of this we cannot explicitly test for parallel trends, which is a crucial assumption behind the validity of a DID estimation strategy. In other words, we cannot test for the assumption that the untreated districts provide the appropriate counterfactual of the trend that treated districts would have followed if they had not been treated. For example, control districts may differ significantly from treated districts in many important characteristics that are themselves correlated with why a particular district was treated. In order to guard against the possibility of the violation of the parallel trend assumption, we use PSM to construct a set of control districts that can be matched with treated districts in observable characteristics.

We therefore combine PSM and DID to estimate the causal impact of financial access on entrepreneurship. In the first stage, we employ PSM to construct matched control and treated districts, as the baseline. In the second stage, we apply the DID method in the matched data to estimate the impact of financial access on entrepreneurship. We follow the steps outlined in Unnikrishnan and Imai (2020).

First step: PSM

The PSM method matches the treated group of enterprises with the control group of enterprises based on observable characteristics. The intuition behind this method is to arrive at a control group of enterprises that were not exposed to the treatment whose observable characteristics are similar to those of the treated group. We match the enterprises based on the binary variable on the status of the districts as under-banked or not in the baseline period. In other words, we match the units of observation based on whether they are located in banked or under-banked districts. We then construct the propensity score based on the covariates that determine the treatment and also simultaneously affect the outcome (in our case, entrepreneurship).

Following Imbens and Rubin (2015), we estimate the propensity score by employing a logit regression. To do this, we first construct a binary variable for treatment, which takes only the values 0 (for control group) or 1 (for treatment group). We then estimate the propensity scores as

[19] As a further robustness check, we also compare the outcome changes in districts that received the treatment prior to the study period, which is, those districts that remained 'banked', and districts that received the treatment during the study period.

the fitted values that are derived from a logit estimation, with the binary treatment variable as the dependent variable and the covariates that are supposed to ensure balance between control and treatment groups as regressors.[20] As covariates, we include district-level variables that are likely to explain why a district is under-banked at the baseline and also to be correlated with entrepreneurship. These variables include proportion of villages that are on a bus route out of total inhabited villages, proportion of villages with electricity out of total inhabited villages, proportion of villages with a post and telegraph office out of total inhabited villages, proportion of villages with paved approach road out of total inhabited villages, proportion of villages with a primary school out of total inhabited villages, proportion of Scheduled Caste households out of total households, and proportion of Scheduled Tribe households out of total households.

The propensity scores are then used to match the control and treatment group enterprises. The key objective of the matching exercise is to find appropriate control group enterprises for treatment group enterprises. For matching, we use the kernel-matching algorithm, which employs weighted averages of all firms in the control group to build the counterfactual group to pair treatment with control firms.[21] As mentioned earlier, the matching is performed for two subsamples of firms: one subsample that includes firms in treated districts and firms in the districts that are under-banked throughout the study period, and another subsample that includes firms in treated districts and firms in the districts that were banked prior to the study period. Once the matching is done, we move to the second stage to disentangle the effect of bank branch penetration on entrepreneurship.

Second step: DID

We apply a version of the DID model to understand the effect of this policy change on entrepreneurship. We compare the firms in districts which remained under-banked (henceforth 'untreated') and those in districts which benefitted from the policy change over the period 2010/11–2015/16 (henceforth 'treated'). In the second stage, our estimation is confined to enterprises in the matched districts. Unlike in the typical DID settings, we lack the baseline data with untreated firms, as there were already both treated and untreated firms in our baseline year, 2010/11. To reduce any sample selection bias and attrition bias influencing our core findings, we follow the strategy employed by Unnikrishnan and Imai (2020). We use propensity scores (PS) as weights in regressions so that the regressions reflect the probability of firms being treated in 2010/11 and 2015/16 as different firms exhibit different probabilities of getting treated. This is certainly not a perfect strategy to eliminate selection bias given that the PS depends on the specification and the results of the probit model. However, we believe that the PS-weighted DID should yield a robust estimate given the data constraints. The generic model we estimate takes the following form:

$$E_{idt} = \beta_0 + \beta_1 U_{idt} + \beta_2 T_{idt} + \beta_3 U * T_{idt} + \beta_4 X_{idt} + \varepsilon_{idt} \quad (5)$$

where E is entrepreneurship and X is a vector of individual-level controls. The subscripts i, d, and t stand respectively for enterprise, district, and time. In our study, t equals 0 for pre-treatment and 1 for post-treatment. U is a dummy variable taking the value 0 for under-banked districts and 1 for other districts. T is a dummy variable that takes the value 1 if t equals 1 and 0 otherwise. The interaction term of U and T identifies the effect of policy change on E_{idt}. The coefficient of the

[20] These fitted values would lie between 0 and 1.

[21] We tried this with different kernel-matching algorithms with different bandwidths and trimming levels to arrive at an ideal estimation model for this study.

interaction term, β_3, therefore, yields a DID estimate that captures the effect of the programme change on the outcome variables.

We also confirm the robustness of our results by combining PSM with a DID strategy. The version of the DID model we use in this study helps us to address the bias resulting from self-selection and confounding. The first step of this method is to employ PSM to match the firms in the comparison group to similar firms in the treatment group. As discussed in the methodology, we used a kernel-matching algorithm to match each firm in the treatment group with firms in the control group. The estimates of DID are likely to be biased if outcomes in already banked districts or already under-banked districts are trending differently from outcomes in districts that witnessed a change in their status from under-banked to banked during the period under study.

The district-level variables that we use in the PSM are shown in Table A7 and include pre-intervention measures of infrastructure and human capital variables. There are eight such variables, namely SHSCPOP, SHSTPOP, MIDGRADEDU, ROADVILLG, ELECVILLG, POSTVILLG, BUSVILLG, and PRIMSCHVILLG. SHSCPOP and SHSTPOP represent the proportion of Scheduled Castes and of Scheduled Tribes in total population, respectively. MIDGRADEDU stands for the proportion of individuals educated to the secondary level and above and ROADVILLG represents the share of villages with paved approach roads in total villages. ELECVILLG, POSTVILLG, BUSVILLG, and PRIMSCHVILLG represent proportion of electrified villages, proportion of villages with post and telegraph offices, proportion of villages situated on a bus route, and proportion of villages with at least a primary school, respectively.

We then apply PS-weighted DID to the matched sample. Following Imbens (2000) and Hirano and Imbens (2001), we use the inverse of propensity scores as weights in the estimations.[22] The results of our DID estimations are presented in Table A8.[23] As mentioned earlier, we applied this method to two matched subsamples of firms. Our results clearly point to the positive effect of policy change on entrepreneurship. To be specific, we find that the probability of household firms becoming non-household firms is greater in districts that received the treatment during the period under study.

Table A7: Means of the covariates for the treated and untreated before and after matching

Variable			Before			After
	Treated	Untreated	StdDif	Treated	Untreated	StdDif
SHSCPOP	0.175616	0.168682	0.095617	0.16124	0.170671	−0.13005
SHSTPOP	0.088978	0.121001	−0.1945	0.143007	0.11887	0.146608
MIDGRADEDU	0.181736	0.116333	1.238938	0.126688	0.113365	0.252375
ROADVILLG	0.748775	0.553677	0.841523	0.606586	0.54467	0.267062
ELECVILLG	0.931218	0.755992	0.809186	0.874246	0.747923	0.583353
POSTVILLG	0.648715	0.427013	0.966474	0.505132	0.417673	0.381264
BUSVILLG	0.655319	0.387542	0.8585	0.529793	0.374197	0.498845
PRIMSCHVILLG	0.892691	0.830285	0.451128	0.850107	0.826545	0.170326

Note: StdDif stands for standardized difference between the treated and the untreated.
Source: Authors' estimates.

[22] Our results are also robust to alternative weighting schemes. For instance, we follow the weighting procedure proposed by Nichols (2007), where we weight the untreated subjects by $pi/(1-pi)$ and treated ones by 1 (Table A9). We also carry out the estimation without weights, where we just use the matched districts from the first round for the second round (Table A10).

[23] Means of the covariates used in the first-stage PSM estimation for the treated and untreated before and after matching are presented in Table A7.

Table A8: Branch expansion and entrepreneurship: PSM-DID using PS weights

Variable	Group 1: unbanked versus unbanked to banked			Group 2: banked versus unbanked to banked		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Banked	−0.040 (0.040)	−0.126*** (0.047)	−0.159 (0.132)	−0.138*** (0.030)	−0.050 (0.035)	0.967*** (0.133)
Time	0.444*** (0.008)	0.328*** (0.014)	0.354*** (0.014)	0.433*** (0.010)	0.219*** (0.013)	0.221*** (0.013)
Banked*Time	0.257*** (0.067)	0.141** (0.074)	0.009 (0.067)	0.098** (0.043)	0.166*** (0.049)	0.175*** (0.050)
Controls	N	Y	Y	N	Y	Y
District FE	N	N	Y	N	N	Y
Observations	290,642	278,456	278,456	195,743	183,278	183,278

Note: controls include gender, location, dummies for social group (ST, SC, and OBC), age categories (Age3to9 and Age>9), assistance, registration status, and labour productivity; we use the inverse of propensity scores as weights in the DID estimations.

Source: Authors' estimates.

Table A9: Branch expansion and entrepreneurship: PSM-DID using PS weights

Variable	Group 1: unbanked versus unbanked to banked			Group 2: banked versus unbanked to banked		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Banked	0.036 (0.029)	0.010 (0.034)	−0.226* (0.121)	−0.068** (0.029)	−0.082** (0.034)	1.005*** (0.135)
Time	0.392*** (0.015)	0.199*** (0.018)	0.290*** (0.019)	0.395*** (0.015)	0.164*** (0.019)	0.173*** (0.020)
Banked*Time	0.124*** (0.040)	0.119*** (0.047)	0.075 (0.049)	0.121*** (0.04)	0.159*** (0.047)	0.196*** (0.049)
Controls	N	Y	Y	N	Y	Y
District FE	N	N	Y	N	N	Y
Observations	290,642	278,456	278,456	195,743	183,278	183,278

Note: controls include gender, location, dummies for social group (ST, SC, and OBC), age categories (Age3to9 and Age>9), assistance, registration status, and labour productivity; we use the inverse of propensity scores as weights for untreated observations ($pi/(1 - pi)$) and 1 for treated observations.

Source: Authors' estimates.

Table A10: Branch expansion and entrepreneurship: PSM-DID

Variable	Group 1: unbanked versus unbanked to banked			Group 2: banked versus unbanked to banked		
	Model 1	**Model 2**	**Model 3**	**Model 1**	**Model 2**	**Model 3**
Banked	0.181*** (0.027)	0.128*** (0.030)	−0.992*** (0.113)	−0.162*** (0.027)	−0.107*** (0.030)	0.689*** (0.126)
Time	0.446*** (0.008)	0.287*** (0.009)	0.295*** (0.009)	0.435*** (0.009)	0.181*** (0.011)	0.201*** (0.011)
Banked*Time	0.070* (0.038)	0.085** (0.042)	0.131*** (0.044)	0.080** (0.039)	0.170*** (0.042)	0.215*** (0.044)
Controls	N	Y	Y	N	Y	Y
District FE	N	N	Y	N	N	Y
Observations	290,642	278,456	278,456	195,743	183,278	183,278

Note: controls include gender, location, dummies for social group (ST, SC, and OBC), age categories (Age3to9 and Age>9), assistance, registration status, and labour productivity; we use the matched districts obtained from the first year for the second year.

Source: Authors' estimates.

References cited in the online appendix

Census of India (2001). *Primary Census Abstract*. New Delhi: Registrar General of India.

Census of India (2011). *Primary Census Abstract*. New Delhi: Registrar General of India.

Hirano, K., and G. Imbens (2001). 'Estimation of Causal Effects using Propensity Score Weighting: An Application to Data on Right Heart Catheterization'. *Health Services and Outcomes Research Methodology*, 2(3): 259–78. https://doi.org/10.1023/A:1020371312283

Imbens, G. (2000). 'The Role of the Propensity Score in Estimating Dose-Response Functions'. *Biometrika* 87(3): 706–10. https://doi.org/10.1093/biomet/87.3.706

Imbens, G., and D. Rubin (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

NSSO (2017). *Processed Data on Survey of Unincorporated Non-Agricultural Enterprises (Excluding Construction) in India, NSS 73rd Round (July 2015 – June 2016)*. New Delhi: NSSO, Ministry of Statistics and Programme Implementation, Government of India.

NSSO (National Sample Survey Office) (2013). *Processed Data on Survey of Unincorporated Non-Agricultural Enterprises (Excluding Construction) in India, NSS 67th Round (July 2010 – June 2011)*. New Delhi: NSSO, Ministry of Statistics and Programme Implementation, Government of India.

RBI (Reserve Bank of India) (2009). *Report of the Group to Review Branch Authorisation Policy*. Mumbai: RBI.

Young, N. (2019). 'Banking and Growth: Evidence from a Regression Discontinuity Analysis: Online Appendix'. Available at: https://nateyoungecon.com/wp-content/uploads/2019/01/NYoung_Banking_and_Growth.pdf (accessed 13 December 2019).