

Gang, Ira N.; Natarajan, Rajesh Raj; Sen, Kunal

Working Paper

Finance, Gender, and Entrepreneurship: India's Informal Sector Firms

IZA Discussion Papers, No. 13854

Provided in Cooperation with:

IZA – Institute of Labor Economics

Suggested Citation: Gang, Ira N.; Natarajan, Rajesh Raj; Sen, Kunal (2020) : Finance, Gender, and Entrepreneurship: India's Informal Sector Firms, IZA Discussion Papers, No. 13854, Institute of Labor Economics (IZA), Bonn

This Version is available at:

<https://hdl.handle.net/10419/227381>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

DISCUSSION PAPER SERIES

IZA DP No. 13854

**Finance, Gender, and Entrepreneurship:
India's Informal Sector Firms**

Ira N. Gang
Rajesh Raj Natarajan
Kunal Sen

NOVEMBER 2020

DISCUSSION PAPER SERIES

IZA DP No. 13854

Finance, Gender, and Entrepreneurship: India's Informal Sector Firms

Ira N. Gang

Rutgers University and IZA

Rajesh Raj Natarajan

Sikkim University

Kunal Sen

*UNU-WIDER, University of Manchester and
IZA*

NOVEMBER 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

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

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

www.iza.org

ABSTRACT

Finance, Gender, and Entrepreneurship: India's Informal Sector Firms*

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.

JEL Classification: O12, G28, L26

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

Corresponding author:

Sikkim University
6th mile, Samdur
Tadong-737102 Gangtok
Sikkim
India

E-mail: rajeshraj.natarajan@gmail.com

* The authors gratefully acknowledge the support of UNU-WIDER. We are also thankful to the National Sample Survey Office for providing us with access to the data. We also thank two anonymous reviewers for their helpful and extensive comments on an earlier version of the paper. The usual disclaimers apply. This study is reproduced here with full acknowledgement of UNU-WIDER, Helsinki. The study was originally commissioned under the UNU-WIDER project Transforming informal work and livelihoods.

1 Introduction

Both women and men participate in the informal economy as they seek to escape poverty and improve their living standards. However, their economic activity typically falls into very different parts of this heterogeneous sector, leading to significant gender disparity in outcomes. The informal firm may be entrepreneurial or not, with higher performance and better outcomes generally seen in enterprises run by entrepreneurs.¹ If female-run enterprises are disproportionately located within the non-entrepreneurial part of the distribution, then it is likely that those women will be less successful in escaping poverty and in improving their living standards than the men in the sector. How to address this disparity in outcomes is a question faced by society and policy-makers.

The literature highlights beneficial influences of female ownership, with increased female firm ownership a stand-alone goal of policy. However, empowering female ownership without acceptable access to complementary resources such as market access and capital may lack the desired positive impact. In other words, in developing economies, female firm ownership may not be sufficient for female empowerment and for its complementary benefits.²

In this paper, our focus is on the role that finance/credit/banking access plays in expanding informal sector enterprises beyond the family and household unit. As we have just indicated, such access has long been viewed as critical for the development of informal firms—whether female- or male-owned. We examine the gender differences in access and in the impact that access has on firm expansion. This leads to our research questions: (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 that employ only family members to those that hire 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 2016b). 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 drawn from the NSSO surveys on unincorporated non-agricultural enterprises carried out for two years, 2010/11 and 2015/16. Informal firms are quite heterogeneous. In line with the NSSO classification, Raj and Sen (2016b) categorize firms in this sector into three types: very small pure household enterprises

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 (Kritikos 2014; 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 the financing constraint that firms and families face (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, 1995; Gang and Pandey 2007). Other firms fall into the unorganized sector and may be registered under other legislation and government bodies. We are analysing this sector.

(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 the proprietor's family, i.e. we merge the MHEs and NHEs (both employ hired workers) into one and call these entrepreneurial firms.⁴

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 we control for the endogeneity of the finance constraint using the instrumental variable method. We also supplement our main identification strategy with an alternate approach, combining propensity score matching (PSM) and difference-in-differences (DID), which also reinforces the main finding of our study and unequivocally highlights 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 provides various robustness checks. Section 6 concludes.

2 Literature, background, and policy environment

In this section, we first examine the lessons that we can 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 Literature

Financial constraints are a factor limiting a firm's growth (Binks and Ennew 1996; Rajan and Zingales 1998; 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 (Beck et al. 2008; Beck 2007; Kuntchev et al. 2013). 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 that 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; McPherson, and Rous 2010). It seems that for enterprise growth, access to finance is needed but by itself is not enough (Nichter and Goldmark 2009).

⁴ In the section on robustness, we examine the sensitivity of our results to alternative definitions of entrepreneurial firms.

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 the information on gender in the data to analyse the linkages between gender and credit. Unfortunately, as clearly stated by Presbitero et al. (2014), the evidence that gender differences in access to finance make a difference in firm development is not conclusive.⁵ Many of these studies use the WBES. As we will soon see, our data overcome many of the 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: gender of the owner, financial constraints faced by the firm, caste of the owner, 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.⁶ 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; and the population covered per branch, which was 13,027 in 2010/11, dropped significantly to 8,683 in 2015/16 (Table 1). Two of the main programmes of the Indian government were the Bharatiya Mahila 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.

⁵ 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.

⁶ A superior source for understanding RBI policy in this period is Young (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).

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 had to 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 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, where one part is 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.⁷

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 variables of interest to us and discuss the basic descriptive statistics.

Table 2: Variables and their construction

Variables	Description	Source
<i>Dependent variable</i>		
Entrepreneurial firm (E)	A binary variable that takes the value 1 for firms that employ at least one hired worker, besides family workers	NSSO Survey data
<i>Independent variables</i>		
<i>Access to finance</i>		
FIN1	A binary variable for financial constraint; 1 if the firm faced any borrowing constraint in the last year, 0 otherwise	NSSO Survey data
FIN2	Ratio of loans from formal sources to total loans	NSSO Survey data
FIN3 DUM1	Dummy variable for bank loan; 1 for firms with bank loan	NSSO Survey data
FIN3 DUM2	Dummy variable for non-government loan; 1 for firms with non-government loan	NSSO Survey data
FIN3 DUM3	Dummy variable for government loan; 1 for firms with government loan	NSSO Survey data
FIN3	Dummy variable for loan; 1 for firms with any type of loan	NSSO Survey data
<i>Gender and interaction terms</i>		

⁷ Issues do exist with the consistency of districts over time, e.g., new districts formed from parts of several older districts (Pradhan 2016).

Female	Binary variable for female-run firms; 1 if the firm is a female-run firm	NSSO Survey data
FIN1*Female	Variable capturing the interaction between FIN1 and Female	NSSO Survey data
FIN2*Female	Variable capturing the interaction between FIN2 and Female	NSSO Survey data
FIN3DUM1 *Female	Variable capturing the interaction between FIN3DUM1 and Female	NSSO Survey data
FIN3DUM2 *Female	Variable capturing the interaction between FIN3DUM2 and Female	NSSO Survey data
FIN3DUM3 *Female	Variable capturing the interaction between FIN3DUM3 and Female	NSSO Survey data
<i>Control variables: firm characteristics</i>		
Location	Dummy variable for urban firms; 1 if the firm is located in an urban area	NSSO Survey data
ST	Dummy variable for ST-owned firms; 1 if the firm is owned by an individual belonging to Scheduled Tribe category	NSSO Survey data
SC	Dummy variable for SC-owned firms; 1 if the firm is owned by an individual belonging to Scheduled Caste category	NSSO Survey data
OBC	Dummy variable for OBC-owned firms; 1 if the firm is owned by an individual belonging to Other Backward Communities (OBC) category	NSSO Survey data
Age3–9	Dummy variable that takes the value 1 if the firm is aged between 3 and 9 years	NSSO Survey data
Age>9	Dummy variable that takes the value 1 if the firm has completed more than 9 years since its inception	NSSO Survey data
Asst	Dummy variable for government assistance; 1 if the enterprise has received any assistance from the government during last three years	NSSO Survey data
Regis	Dummy variable that takes the value 1 if the firm has 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 Survey data
lnLP	Log of labour productivity, where labour productivity is defined as the ratio of gross value added to employment	NSSO Survey data
<i>District-level variables</i>		
SHSCPOP	Proportion of Scheduled Caste population in total population	Census 2001
SHSTPOP	Proportion of Scheduled Tribe 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
ELECIVILLG	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) and NSSO (2013, 2017).

Dependent variable

We classify an entrepreneurial firm as one that employs 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 usually seen as an indicator of entrepreneurial success (see 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.⁸

Main independent variables:

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

1. First, 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 had faced over the last year.⁹
2. Second, using the loan amount reported by the firms, we take the ratio of loans from formal sources to total loans as a proxy for finance availability (FIN2). We employ the following method to arrive at this variable: $FIN2 = (\text{loans from commercial banks} + \text{loans from cooperative banks} + \text{loans from microfinance institutions} + \text{loans from other institutional agencies}) / (\text{total loan} - \text{loans from central/state term lending institutions} - \text{loans from government})$. This ratio, however, effectively treats enterprises not taking loans as missing, resulting in a reduction of our sample size by several hundred thousand.
3. Third, as an alternative to the ratio variable, we also construct 0–1 categorical variables for firms receiving bank loans (FIN3DUM1), non-government loans (FIN3DUM2), and government loans (FIN3DUM3), while keeping firms not taking loans as the benchmark (reference) category.

Each of these three measures of financial constraints is imperfect in itself. Using the three measures mitigates against the problem of measurement errors in any one measure. Specifically, with the first and third measures, 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

⁸ The existence of dynamic entrepreneurial firms along with subsistence firms in the informal sector is well documented in the literature (Grimm et al. 2012).

⁹ Below we discuss a possible selection issue here, in addition to an endogeneity issue for the generic FIN variable.

¹⁰ We lose 4.5% of the 619,701 observations in our dataset when we follow this condition.

Firm characteristics as control variables

Urban: This variable stands for the location of the firm, assuming the value 1 for urban firms and 0 for rural firms. Urban firms generally grow faster than their rural counterparts do, as their market is larger and they have better access to infrastructure and inputs.

Social group: Under the possibility that the socioeconomic group to which owners belong will affect firms' ability to transition, we capture the firm owner's social group using three dummy variables, for Other Backward Caste (OBC), Scheduled Caste (SC), and Scheduled Tribe (ST) membership. 'Others' is the omitted group.

Age: The age of the firm is measured as the number of years elapsed since its establishment. We classify firms into three age categories, those from zero to two years old (Age0–2), those from three to nine years old (Age3–9), and those active for ten years or more (Age>9). We introduce two dummy variables, for Age3–9 and Age>9, with Age0–2 as the reference category.

Assistance: This variable takes the value 1 for firms that have received government assistance towards training and marketing. Assistance may affect firm hiring within the informal sector.

Registration: This variable takes the value 1 if firms have registered under any Act and 0 if they have not. Registration itself may form a type of collateral, providing the firm with access to otherwise unavailable financial and non-financial resources (Levenson and Maloney 1998; Sharma 2014). The unorganized sector firms we are considering are not registered under the Factories Act of 1948 (see footnote 3). However, they may be registered with state and municipal governments and co-operative authorities.

Labour productivity: Labour productivity is measured as the ratio of gross value added to employment. We follow the existing literature in including labour productivity as an additional control, as more productive firms are more likely to be entrepreneurial firms (Raj and Sen 2016a). We use the log of labour productivity in regression estimations.

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 appendix. Table 3 presents a simple and fascinating picture of entrepreneurship in our sample. The top panel shows the cross-tabulation of entrepreneurship with the owner's gender; the bottom panel 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.

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	299,021	216,074	515,095
	Row percentage	58.05	41.95	100
	Column percentage	82.45	94.45	87.09
	Cell percentage	50.56	36.53	87.09
Female-run firms	Frequency	63,666	12,693	76,359
	Row percentage	83.38	16.62	100
	Column percentage	17.55	5.55	12.91
	Cell percentage	10.76	2.15	12.91
Total	Frequency	362,687	228,767	591,454
	Row percentage	61.32	38.68	100
	Column percentage	100	100	100
	Cell percentage	61.32	38.68	100
Panel B: entrepreneurship vs financial constraint				
Financial constraint	Owner	Entrepreneurship		Total
		Non-entrepreneurial firm	Entrepreneurial firm	
No	Frequency	347,499	229,757	577,256
	Row percentage	60.2	39.8	100
	Column percentage	92.54	94.09	93.15
	Cell percentage	56.08	37.08	93.15
Yes	Frequency	28,025	14,420	42,445
	Row percentage	66.03	33.97	100
	Column percentage	7.46	5.91	6.85
	Cell percentage	4.52	2.33	6.85
Total	Frequency	375,524	244,177	619,701
	Row percentage	60.6	39.4	100
	Column percentage	100	100	100
	Cell percentage	60.6	39.4	100

Source: authors' construction based on own estimates.

3.3 Estimating strategy

We estimate variations of

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

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. FIN measures a firm's financial constraint, and we use three alternative measures of financial constraint.¹¹ 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

¹¹ Their construction is discussed in detail in the section on data and variables.

owner, and log of labour productivity.¹² The variables μ_i and δ_d control, respectively, for time- and district-specific fixed effects.

We then augment Equation 1 with another indicator variable, *Female*, which equals 1 if the firm is a female-run firm, and its interaction with *FIN*. The generic augmented equation is of the following form:

$$E_{idt} = \alpha_0 + \alpha_1 FIN_{idt} + \alpha_2 Female_{idt} + \alpha_3 FIN * Female_{idt} + \sum_{i>0} \gamma_i X_{idt} + \mu_i + \delta_d + \varepsilon_j \quad (2)$$

where the variables are defined as above.

Equations 1 and 2 are estimated using logit. Parameter α_1 captures the relationship between the financial constraint and the probability of entrepreneurship. To ensure the robustness of our results, we estimate several variants of both equations.

3.4 Identification and estimation

Our empirical strategy aims to identify the effect of the finance constraint on entrepreneurship—in particular, 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).¹³ Moreover, we believe our instrument meets the necessary exclusion

¹² These variables are defined in Table 2.

¹³ 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.

criterion for an IV, as it is only through the firm’s financial constraint that it should influence the enterprise’s decision to hire outside workers. We test for the suitability of the instrument in our estimations.

Following Terza et al. (2008), we attempt to correct endogeneity by employing the two-stage residual inclusion (2SRI) method. 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:

$$\textit{First stage: } FIN_{idt} = \alpha_0 + \alpha_1 \textit{Instruments} + \sum_{i>0} \gamma_i X_{idt} + \mu_i + \delta_d + \theta_j \quad (3)$$

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:

$$\textit{Second stage: } E_{idt} = \beta_0 + \beta_1 FIN_{idt} + \gamma \theta_i^{est} + \sum_{i>0} \gamma_i X_{idt} + \mu_i + \delta_d + \varepsilon_j \quad (4)$$

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.

3.5 Robustness check: PSM-DID

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 that 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.

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.¹⁴ 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).

¹⁴ As a further robustness check, we also compare the outcome changes in districts that received the treatment prior to the study period, that is, those districts that remained 'banked', and districts that received the treatment during the study period.

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 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.¹⁵ 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 a 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.¹⁶ 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

¹⁵ These fitted values would lie between 0 and 1.

¹⁶ We tried this with different kernel-matching algorithms with different bandwidths and trimming levels to arrive at an ideal estimation model for this study.

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

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 robustness checks for our story.

4 Results

This section presents our baseline 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 shows our logit estimates of the impact of financial constraints on unorganized sector firms on the probability of hiring non-family labour. Columns 1–6 are for the 2010/11 wave; Columns 7–12 for the 2015/16 wave; and Columns 13–18 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: Access to finance and entrepreneurship: logistic regression estimates with control variables (dependent variable: entrepreneurial firm or not)

Variables	2010-2011						2015-2016						Pooled Data					
	FIN1 (1)	(2)	FIN2 (3)	(4)	FIN3 (5)	(6)	FIN1 (7)	(8)	FIN2 (9)	(10)	FIN3 (11)	(12)	FIN1 (13)	(14)	FIN2 (15)	(16)	FIN3 (17)	(18)
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			0.314*** (0.027)	0.440*** (0.032)					0.343*** (0.025)	0.429*** (0.029)					0.311*** (0.019)	0.428*** (0.021)		
FIN3					0.533*** (0.020)	0.579*** (0.021)					0.522*** (0.019)	0.524*** (0.020)					0.527*** (0.014)	0.553*** (0.014)
DUM1																		
FIN3					0.312*** (0.019)	0.244*** (0.020)					0.245*** (0.016)	0.185*** (0.017)					0.297*** (0.012)	0.217*** (0.013)
DUM2																		
FIN3					0.506*** (0.061)	0.596*** (0.062)					0.633*** (0.073)	0.623*** (0.074)					0.543*** (0.046)	0.627*** (0.047)
DUM3																		
Female	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	0.817*** (0.015)	0.825*** (0.016)	0.582*** (0.048)	0.578*** (0.051)	0.816*** (0.015)	0.820*** (0.016)	0.807*** (0.014)	0.837*** (0.015)	0.636*** (0.041)	0.633*** (0.044)	0.807*** (0.014)	0.831*** (0.015)	0.781*** (0.010)	0.833*** (0.011)	0.591*** (0.031)	0.603*** (0.033)	0.781*** (0.010)	0.826*** (0.011)
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Time FE	N	N	N	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	Y
District FE	No	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y
N	313589	313586	28416	28297	313589	313586	270471	270448	35591	35490	270471	270448	584060	584055	64407	63950	584060	584055
Pseudo R2	0.145	0.169	0.096	0.146	0.148	0.171	0.147	0.172	0.111	0.150	0.151	0.174	0.151	0.171	0.104	0.139	0.154	0.173

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

Source: authors' construction based on own estimates.

Table 5: Access to finance and entrepreneurship: logistic regression estimates with control variables and interaction terms (dependent variable: entrepreneurial firm or not)

Variable	2010–11						2015–16						Pooled data						
	FIN1		FIN2		FIN3		FIN1		FIN2		FIN3		FIN1		FIN2		FIN3		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	
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			0.284*** (0.028)	0.408*** (0.033)					0.311*** (0.026)	0.395*** (0.030)					0.279*** (0.019)	0.395*** (0.021)			
FIN3 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)	
FIN3 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)	
FIN3 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)	-0.826*** (0.073)	-0.828*** (0.077)	-0.845*** (0.016)	-0.847*** (0.017)	-0.809*** (0.015)	-0.838*** (0.015)	-0.836*** (0.059)	-0.838*** (0.063)	-0.830*** (0.015)	-0.853*** (0.016)	-1.782*** (0.011)	-0.834*** (0.011)	-0.816*** (0.046)	-0.836*** (0.049)	-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)					
FIN2*Female			0.474*** (0.099)	0.488*** (0.105)					0.429*** (0.086)	0.439*** (0.091)					0.462*** (0.065)	0.477*** (0.068)			
FIN3DUM1 *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)	
FIN3DUM2 *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)	
FIN3DUM3 *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	Y	Y	Y	Y	Y	Y	
Time FE	N	N	N	N-	N	N	N	N-	N	N	N	N	N-	N	Y	N	Y	N	Y
District FE	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	N	Y	Y
N	313,589	313,586	28,416	28,297	313,589	313,586	270,471	270,448	35,591	35,490	270,471	270,448	584,060	584,055	64,007	63,950	584,060	584,055	
Pseudo R2	0.145	0.169	0.096	0.147	0.148	0.171	0.148	0.172	0.112	0.151	0.151	0.174	0.151	0.171	0.105	0.140	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' construction based on own estimates.

FIN1 in the tables is the direct measure of financial constraints (Columns 1, 2, 7, 8, 13, 14) as discussed earlier, which is similar to the way papers using the WBES capture financial constraints. Notice that the coefficient estimate is negative and significant at the 1 per cent level in all the estimations in Table 4: as the constraint weakens, we are more likely to find firms in the part of the informal sector we have labelled as entrepreneurial.

FIN2 is the proportion of loans the firms receive from ‘formal sources’ (Columns, 3, 4, 9, 10, 15, 16). This measure is likely to face a severe selection problem, as about 90 per cent of firms did not receive any financing from outside sources. Still, this leaves us with many firms across the distribution of types of unorganized sector firms that did rely on various types of loans conditional on firms receiving outside funding. The coefficient estimate is positive and significant at the 1 per cent level in all the estimations in Table 4: as formal funding increases proportional to total external funding, we are more likely to find firms in the part of the informal sector we have labelled as entrepreneurial.

FIN3 again captures the sources of external funding, but here we use dummy variables to capture the types of loans the firm received, with FIN3DUM1, FIN3DUM2, and FIN3DUM3 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 do not suffer the selection issue as in FIN2; they are estimated with our complete sample. Again, in Table 4, relative to the base category of firms with no external financing, the coefficient estimates are all positive and significant at the 1 per cent level: more external finding, more entrepreneurial behaviour.

In all model specifications—all 18 columns—an informal firm headed by a woman is more likely to be in the family part of the informal economy: the coefficient estimates are negative and significant at the 1 per cent level in all specifications.

Table 5 repeats the columnar structure and content of Table 4, now with the addition of rows for the interaction of owner’s gender with the financial constraints. In other words, the complete specification includes our core financial constraints, owner’s gender, interactions of owner’s gender with financial constraints, and our other control variables. The coefficient estimates on the variables for financing (FIN1, FIN2, and FIN3) 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 FIN2 and FIN3 across the bottom rows are all significant at the 1 per cent level and positive. For FIN2 and FIN3, this reinforces the intercept shifts we saw in Tables 4. For both FIN2 and FIN3, the intercept shifts upwards and the slope becomes more positive.

Our basic logit estimations shown in Tables 4 and 5 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. However, we do not provide conclusive evidence that women are at a disadvantage in joining the entrepreneurial side as compared with men, as the results on the effect of the finance

constraint on male versus female entrepreneurship depends on our choice of the measure to capture the finance constraint.¹⁷

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 6, we present the second-stage reduced-form estimates for one specification with three versions of our financial constraints using our instrumental variables approach. The results mimic, in sign and significance, the logit results discussed in the previous subsection.

Table 6: Access to finance and entrepreneurship: logit model using 2SRI approach (no. of replications: 500)

Variable	FIN1	FIN2	FIN3
FIN1	-0.079*** (0.012)		
FIN2		0.316*** (0.019)	
FIN3			0.428*** (0.010)
Female	-0.930*** (0.012)	-0.648*** (0.035)	-0.789*** (0.011)
Location	0.173*** (0.008)	0.144*** (0.021)	0.108*** (0.012)
ST	-0.676*** (0.017)	-0.726*** (0.055)	-0.693*** (0.018)
SC	-0.649*** (0.014)	-0.712*** (0.037)	-0.727*** (0.013)
OBC	-0.220*** (0.008)	-0.146*** (0.029)	-0.186*** (0.008)
Age3–9	-0.331*** (0.011)	-0.205*** (0.028)	-0.496*** (0.015)
Age>9	-0.304*** (0.011)	-0.089*** (0.029)	-0.488*** (0.017)
Asst	0.805*** (0.033)	-0.079 (0.128)	1.904*** (0.109)
Regis	1.073*** (0.009)	0.683*** (0.070)	1.352*** (0.018)
lnLP	0.422*** (0.005)	0.294*** (0.025)	0.549*** (0.009)
XUhat	-5.073*** (0.203)	1.016*** (0.296)	-2.503*** (0.192)
Observations	619,701	64,350	619,701

Note: figures in parentheses are bootstrapped standard errors; in this table we collapse FIN3DUM1, FIN3DUM2, and FIN3DUM3 into a single dummy variable, FIN3: FIN3 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' construction based on own estimates.

¹⁷ We 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.

However, again we see that this varies over the three 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 -5.073 with standard error 0.203 , suggesting that the regressor is endogenous. The coefficient estimate on FIN1 is -0.079 with standard error 0.012 , indicating significance at the 1 per cent level. In comparison, in Table 6, Column 13, the coefficient estimate on FIN1 was -0.097 with standard error 0.012 .

Using FIN2, γ is estimated as 1.016 with standard error 0.296 , suggesting that the regressor is endogenous. The coefficient estimate on FIN2 is 0.316 with standard error 0.019 , indicating significance at the 1 per cent level. In comparison, in Table 6, Column 15, the coefficient estimate on FIN2 was 0.311 with standard error 0.019 .

For the two-stage procedure, we collapse FIN3DUM1, FIN3DUM2, and FIN3DUM3 into a single dummy variable, FIN3. FIN3 equals 1 if the firm has taken out any sort of formal loan. It is 0 otherwise. Using FIN3, γ is estimated as -2.503 with standard error 0.192 , suggesting that the regressor is endogenous. The coefficient estimate on FIN3 is 0.428 with standard error 0.010 , indicating significance at the 1 per cent level. Because in Table 6 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, significance, or size of our estimates. In other words, the 2SRI estimates reinforce the main findings arrived at using the logit estimations. The coefficients of FIN1, FIN2, and FIN3 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.

We also perform the 2SRI estimation using separate samples of male and female owners. The results are reported in Table 7. They show the same pattern as observed for the pooled sample. As in the case of our logit results, we get ambiguous results on whether finance constraints matter more for women than for men. In the case of FIN1, we find that the coefficient is not statistically significant for women, though it is significant and of the right sign (negative) for men. In the case of FIN2 and FIN3, 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, in the case of FIN2 and FIN3, the coefficients on the finance constraint are lower in magnitude for women than for men, indicating that the constraint is more binding for male owners than for female owners.

Table 7: Access to finance and entrepreneurship by gender: logit model using 2SRI approach (no. of replications: 500)

Variable	Male-run firms			Female-run firms		
	FIN1	FIN2	FIN3	FIN1	FIN2	FIN3
FIN1	-0.086*** (0.011)			0.065 (0.044)		
FIN2		0.256*** (0.019)			0.168*** (0.056)	
FIN3			0.402*** (0.010)			0.344*** (0.028)
Location	0.146*** (0.008)	0.111*** (0.019)	0.256*** (0.011)	-0.023 (0.022)	-0.068 (0.060)	0.014 (0.037)
ST	-0.817*** (0.019)	-0.695*** (0.058)	-0.673*** (0.018)	-0.856*** (0.044)	-0.318 (0.253)	-0.805*** (0.058)
SC	-0.623*** (0.015)	-0.629*** (0.041)	-0.743*** (0.012)	-0.940*** (0.048)	-0.213 (0.162)	-1.074*** (0.040)
OBC	-0.211*** (0.009)	-0.111*** (0.030)	-0.255*** (0.007)	-0.636*** (0.024)	-0.004 (0.120)	-0.619*** (0.021)
Age3-9	-0.336*** (0.013)	-0.317*** (0.026)	-0.350*** (0.014)	-0.593*** (0.031)	-0.886*** (0.079)	-0.608*** (0.033)
Age>9	-0.290*** (0.013)	-0.139*** (0.026)	-0.288*** (0.016)	-0.322*** (0.032)	-0.506*** (0.084)	-0.312*** (0.038)
Asst	0.495*** (0.029)	-0.124 (0.103)	0.044 (0.102)	0.180*** (0.057)	-0.862*** (0.178)	0.095 (0.266)
Regis	0.994*** (0.010)	0.745*** (0.058)	1.072*** (0.016)	2.177*** (0.024)	1.226*** (0.141)	2.146*** (0.058)
lnLP	0.387*** (0.005)	0.478*** (0.012)	0.414*** (0.007)	0.770*** (0.012)	0.985*** (0.029)	0.795*** (0.015)
XUhat	-6.451*** (0.240)	0.507** (0.234)	0.397** (0.172)	-7.011*** (1.128)	2.263*** (0.559)	-0.155 (0.470)
District FE	Y	Y	Y	Y	Y	Y
Observations	543,342	543,342	543,342	104,606	104,606	104,606

Note: figures in parentheses are bootstrapped standard errors; in this table we collapse FIN3DUM1, FIN3DUM2, and FIN3DUM3 into a single dummy variable, FIN3: FIN3 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' construction based on own estimates.

5 PSM-DID results

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 2 and include pre-intervention measures of infrastructure and human capital variables. There are eight such variables, namely SHSCPOP, SHSTPOP, MIDGRADEDU, ROADVILLG, ELECIVILLG, 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. ELECIVILLG, 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.¹⁸ The results of our DID estimations are presented in Table 8.¹⁹ 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 8: 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' construction based on own estimates.

¹⁸ 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 A3). We also carry out the estimation without weights, where we just use the matched districts from the first round for the second round (Table A4).

¹⁹ Means of the covariates used in the first-stage PSM estimation for the treated and untreated before and after matching are presented in Table A2.

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 three 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 measure of the amount of external firm borrowing; (3) 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 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 2016b). 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. (2007). 'Financing Constraints of SMEs in Developing Countries: Evidence, Determinants and Solutions'. Tilberg: TiSEM (Tilburg School of Economics and Management), Tilburg University.
- 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>
- Binks, M.R., and C.T. Ennew (1996). 'Growing Firms and the Credit Constraint'. *Small Business Economics*, 8(1): 17–25. <https://doi.org/10.1007/BF00391972>
- 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.
- 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>
- 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](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.
- 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](https://doi.org/10.1016/0305-750X(92)90085-A)

- Gang, I.N. (1995). 'Small Firms in India: A Discussion of Some Issues'. In D. Mookherjee (ed.), *Indian Industry: Policies and Performance*. Delhi and New York: Oxford University Press.
- 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>
- 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>
- 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>
- 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>
- Kritikos, A.S. (2014). 'Entrepreneurs and Their Impact on Jobs and Economic Growth'. *IZA World of Labor*, 8: 1–10. <https://doi.org/10.15185/izawol.8>
- Kuntchev, V., R. Ramalho, J. Rodriguez-Meza, and J.S. Yang (2013). 'What Have We Learned from the Enterprise Surveys regarding Access to Finance by SMEs?' Policy Research Working Paper 6670. Washington, DC: The World Bank. <https://doi.org/10.1596/1813-9450-6670>
- Landes, D.S. (1969). *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present*. Cambridge: Cambridge University Press.
- Levenson, A.R., and W.F. Maloney (1998). *The Informal Sector, Firm Dynamics and Institutional Participation*. Washington, DC: The World Bank. <https://doi.org/10.1596/1813-9450-1988>
- 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>
- McPherson, M.A., and J.J. Rous (2010). 'Access to Finance and Small Enterprise Growth: Evidence from East Java'. *Journal of Developing Areas*, 43(2): 159–72. <https://doi.org/10.1353/jda.0.0066>
- 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](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>

- Nichols, A. (2007). 'Causal Inference with Observational Data'. *Stata Journal*, 7: 507–41. <https://doi.org/10.1177/1536867X0800700403>
- 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., and K. Sen (2016a). 'Moving Out of the Bottom of the Economy? Constraints to Firm Transition in the Indian Informal Manufacturing Sector'. *IZA Journal of Labor & Development*, 5(1). <https://doi.org/10.1186/s40175-016-0056-8>
- Raj, R.S.N., and K. Sen (2016b). *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>
- Rajan, G.R., and L. Zingales (1998). 'Financial Dependence and Growth'. *American Economic Review*, 88(3): 559–586.
- RBI (Reserve Bank of India) (2009). *Report of the Group to Review Branch Authorisation Policy*. Mumbai: RBI.
- 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>
- Sharma, S. (2014). 'Benefits of a Registration Policy for Microenterprise Performance in India'. *Small Business Economics*, 42: 153–64. <https://doi.org/10.1007/s11187-013-9475-y>
- 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>
- Unnikrishnan, V., and K.S. Imai (2020). 'Does the Old-Age Pension Scheme Improve Household Welfare? Evidence from India'. *World Development*, 134: 105017. <https://doi.org/10.1016/j.worlddev.2020.105017>
- 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).

Appendix

Table A1: Summary statistics

Variable	Observations	Mean	SD	Min.	Max.
Entrepreneurial firm (E)	619,701	0.394024	0.48864	0	1
FIN1	619,701	0.068493	0.25259	0	1
FIN2	72,224	0.49822	0.490601	0	1
FIN3 DUM1	619,701	0.060365	0.238161	0	1
FIN3 DUM2	619,701	0.062228	0.24157	0	1
FIN3 DUM3	619,701	0.005651	0.074961	0	1
Female	591,454	0.129104	0.335315	0	1
Location	619,701	0.509158	0.499917	0	1
ST	619,701	0.053986	0.22599	0	1
SC	619,701	0.096403	0.295143	0	1
OBC	619,701	0.448534	0.497345	0	1
Age3–9	619,701	0.44809	0.497299	0	1
Age>9	619,701	0.425087	0.494357	0	1
Asst	619,697	0.017473	0.131026	0	1
Regis	619,701	0.413553	0.492471	0	1
InLP	614,005	10.00492	1.025542	0.61619	19.52315

Source: authors' construction based on own estimates.

Table A2: Means of the covariates for the treated and untreated before and after matching

Variable	Before		After	
	Treated	Untreated	StdDif	StdDif
SHSCPOP	0.175616	0.168682	0.095617	-0.13005
SHSTPOP	0.088978	0.121001	-0.1945	0.146608
MIDGRADEDU	0.181736	0.116333	1.238938	0.252375
ROADVILLG	0.748775	0.553677	0.841523	0.267062
ELECVILLG	0.931218	0.755992	0.809186	0.583353
POSTVILLG	0.648715	0.427013	0.966474	0.381264
BUSVILLG	0.655319	0.387542	0.8585	0.498845
PRIMSCHVILLG	0.892691	0.830285	0.451128	0.170326

Note: StdDif stands for standardized difference between the treated and the untreated.

Source: authors' construction based on own estimates.

Table A3: 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' construction based on own estimates.

Table A4: 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' construction based on own estimates.