Caste and Credit: A woeful tale?

Sunil Mitra Kumar   Ragupathy Venkatachalam[*]

August 2016

Abstract

This paper examines caste-based differences in farmers' access to bank loans in rural India. We investigate whether banks practice taste-based discrimination on the basis of caste. In order to identify potential discrimination, we consider loan applications and approval decisions separately. We find significant inter-caste differences in application rates, and evidence of discrimination against Scheduled Tribe borrowers at the approval stage. To rule out the role of statistical discrimination, we simulate unobserved credit histories with various distributions. Evidence for taste-based discrimination persists despite accounting for unobservables. However, we find that this discrimination does not affect small farmers.

1 Introduction

Access to resources and opportunities can be a critical factor in improving outcomes for disadvantaged groups. Barriers to accessing social, political and economic opportunities may not automatically reduce over time, and intervention through legislation and public policy, for instance through affirmative action, are often necessary. In particular, improving access to financial resources is widely acknowledged to facilitate upward economic and social mobility.[1] Conversely, lack of access to resources for certain groups based on race, gender or ethno-social identities can perpetuate inequalities.

In India, caste as a form of social identity remains an enduring predictor of economic status. It is correlated with occupation and employment (Prakash, 2015; Ito, 2009; Thorat and Attewell, 2007), income and expenditure (Deshpande, 2000), and capital more generally (Kijima, 2006). Agriculture is the largest employer in India, and due to the gap between sowing and harvest, bank credit is usually needed to be able to purchase various inputs (Conning and Udry, 2007). One important avenue through which caste can influence income and productivity in the rural sector, where agriculture dominates, is by shaping access to bank credit. The question of whether caste influences access to credit is thus important, and while there is a general agreement that it does (Burgess et al., 2005; Kumar, 2013; Pal, 2002; Government of India, 2007), the reasons for these differences are less clear. One particular concern is whether lenders discriminate on the basis of caste (Kumar, 2013; Dréze et al., 1997).[2]

In this paper, we focus on the role of caste in shaping access to financial capital, in particular bank loans, in rural India. Is there differential access for different caste-groups, and if so, why? While varying levels of access might signal the presence of discrimination, they may also arise from the failure to seek resources and capitalise on opportunities. To make sense of potential caste-based discrimination in accessing credit, it is useful to note the parallels with discrimination in labor markets (Thorat and Attewell, 2007; Altonji and Blank, 1999) and the role of racial and gender-based discrimination in accessing bank credit in the USA (e.g. Blanchflower et al., 2003; Blanchard et al., 2008; Asiedu et al., 2010). Castes in India can be categorised into four major groups, viz. Scheduled Castes (SCs; the most disadvantaged), Other Backward Classes(OBCs; of middling disadvantage), Brahmins, the traditionally privileged higher castes, and Scheduled Tribes (STs;

---

[1]Equality in access to economic resources has been traditionally viewed as a vehicle to promote broader social equality. For example, this is the rationale behind passing the Equal Credit Opportunity Act in 1974 (and the subsequent amendment in 1976) in the United States.

[2]Akerlof (1976) provides one of the early theoretical explanations regarding the persistence of caste-based discrimination given informational costs in an Arrow-Debreu general equilibrium framework.

also very disadvantaged).[3,4,5]

There are at least two ways to explain discrimination within a rational choice framework. First, the *taste-based* discrimination proposed by Gary Becker (1971). According this definition, discrimination occurs if the differences in some outcome (loan approvals, wages) based on race, caste, or a similar attribute, are not 'objective'. That is, if residual differences remain even after taking into account all possible borrower or worker characteristics that are relevant to the outcome, and are attributed to a 'taste' for discrimination on the part of the bank or employer. Second, discrimination may occur due to information problems relating to unobservable 'objective' characteristics – creditworthiness, ability – leading bank managers or employers to focus on other observable but correlated proxies such as race to extract the *expected* value of the relevant characteristic. This is referred to as *statistical* discrimination (Arrow, 1973, 1998). There is yet another way in which outcomes might differ systematically by caste or gender or race: if demand for pay, loans, or jobs varies based on group-belonging (e.g. Chevalier, 2007). In this paper we attempt to distinguish between these three phenomena.

Though these notions of discrimination are intuitively straightforward to comprehend, the task of measurement is quite challenging. Empirical studies, for instance on mortgage lending in the USA, often use survey data to isolate the residual effects of race or gender. This approach suffers from the well known limitation that potential unobservables might lead to biased estimates given that such data are observational.[6] There is a small but growing literature that uses randomisation to study discrimination in labour markets (Bertrand and Mullainathan, 2004; Banerjee et al., 2009). However, it is difficult to study loan approvals by banks using randomisation given the practical informational requirements of applying for loans. Moreover, some aspects of access to credit are not readily amenable to study through randomisation. For instance, since caste cannot be randomly allocated, there are limitations in using this approach to study how caste-belonging influences the propensity to apply for a loan. In view of these limitations, the literature on discrimination in labour and credit markets has mainly used observational data, as does our paper. Yet because potential unobservables remain a concern, we employ counterfactual simulations to investigate how our results might change if we could control for certain key unobservables. We adapt an approach suggested by Ichino et al. (2008)

---

[3]There are substantial, enduring complexities and political agitation around the categorisation of castes as OBC (Ramaiah, 1992).

[4]STs are not technically part of the caste system. Yet tribals are one of the most disadvantaged peoples in India, and most analyses of caste therefore include the ST category, as does the current paper.

[5]It is generally agreed that the caste system applies not only to Hindus but to followers of other religions in the Indian subcontinent as well.

[6]See LaLonde (1986); Duflo et al. (2007)

and Rosenbaum and Rubin (1983) to simulate a binary 'credit history' variable, and study how the residual effect of caste-group belonging on loan approvals changes with different distributional assumptions.[7]

The paper asks the following questions: (a) How does caste influence farmers' access to bank loans? (b) Do banks discriminate on the basis of caste? By way of addressing these questions, the paper also, albeit indirectly, addresses a larger question concerning the effectiveness of affirmative action policies in the financial sector. While caste-based affirmative action in government jobs and admissions to educational institutions (Bertrand et al., 2010) are more prominent, the Reserve Bank of India (RBI) has long directed bank lending towards disadvantaged groups (Reserve Bank of India, 2014, 2004, 2008; Sriram, 2007). These policies are especially significant for bank lending in rural India, where expanding formal credit is thought to be key to increasing rural incomes.[8]

We use nationally-representative data from the 2011-12 round of the India Human Development Survey (Desai and Vanneman, 2015), which enables us to study the decision to apply for a loan separately from the subsequent approval of this application. We focus on farmer households in rural India and their borrowings from banks. Our main findings are as follows. First, caste-wise differences do exist in access to loans and a major part of these differences can be explained by corresponding differences in application rates. SCs and STs, the two most disadvantaged groups, are 16-20% less likely to apply for loans than are Brahmins and OBCs. Second, for those farmers who applied for bank loans, approval rates are almost uniform across Brahmins, OBCs, and SCs. ST borrowers are 5-7% less likely to have a loan approved. In other words, STs likely face taste-based discrimination. Third, we argue that this finding is largely robust through a sensitivity analysis that constructs counterfactual scenarios concerning a key unobservable, viz. credit histories. Fourth, we find positive supportive evidence regarding the success of the RBI's affirmative action initiatives for lending to small farmers. Dividing the sample according to land ownership, we find that caste-based differences in loan application and approval rates are largely muted for small farmers. Fifth, we find that the taste-based

---

[7]Although we refer to credit histories, the approach is more general, and can be interpreted in terms of any other key unobservable which might be considered relevant e.g. the quality of land, effort, or productivity.

[8]Though there is no equivalent legislation to the United States' Equal Credit Opportunity Act in India, the RBI has long encouraged banks to lend to 'priority-sectors', including SCs, STs and small farmers. For instance, a recent circular on priority lending targets and classification specifies a target of '18 percent of Adjusted Net Bank Credit (ANBC) or Credit Equivalent Amount of Off-Balance Sheet Exposure, whichever is higher' towards agricultural sector. The corresponding target for 'weaker sections', of which SC, ST, small and marginal farmers are a part, is '10 percent of ANBC or Credit Equivalent Amount of Off-Balance Sheet Exposure, whichever is higher'. See https://www.rbi.org.in/scripts/NotificationUser.aspx?Id=9688Mode=0

discrimination against STs is visible mainly in states with substantial ST populations, and not where they are a small minority.

Our study has certain key advantages over previous work in this area. Researchers so far have studied the determinants of access to credit including caste (Kumar, 2013; Pal, 2002) and whether farmers are credit-constrained (Kochar, 1997). These studies control for credit-worthiness characteristics including land-ownership, household demographics, and indicators of economic status, but do not control for the decision to *seek* credit. To our knowledge, this paper is the first attempt to study rural bank lending in India using data that include information about loan applications. This enables us to explicitly distinguish between lender-based differences in loan approval rates and borrower-based differences in the demand for credit.

The reminder of the paper is structured as follows. Section 2 presents the data, while section 3 lays out our empirical approach. Results are presented in section 4, and section 5 discusses their implications.

2 Data

We use data from the second round of the India Human Development Survey (IHDS), a nationally-representative household survey undertaken in 2011-12 (Desai and Vanneman, 2015). These data provide information on a rich set of social, economic and demographic characteristics which are well suited to analysing the determinants of households' borrowing. Crucially, households were asked whether they had applied for loans from various sources, and the success of those applications. This question is critical to exploring whether any caste-wise differences in loan patterns do indeed reflect discrimination on the part of lenders, and to our knowledge, no other large-scale household survey in India has asked this question.[9] As per the questionnaire, data on bank loan applications and approvals refer to those during the five years preceding the survey.

We focus only on formal sector, i.e. bank loans in the current paper. Such loans are widely considered superior than what the informal sector provides, owing to the usurious lending practices in the latter. Therefore, the question of caste-based differences is importance for the formal sector. It is also usually assumed that given a choice, farmers would opt for

---

[9]The National Sample Survey Organisation's All India Debt and Investment Survey is the other main source of information on household's access to credit. This is a decennial survey, conducted most recently in 2013. In keeping with previous rounds, it asks about the details of existing loans but does not ask about loan applications.

formal sector loans (e.g. Kochar, 1997). Therefore, excluding informal sector borrowings is also unlikely to compromise the identification of caste-based differences in application and approval rates for formal sector loans.

The second round of the IHDS data currently consist of individual and household-level information, but village-level information are covered only in the first round of the survey that took place in 2004-05. In particular, the presence of a bank in the village or distance to the nearest such is potentially an important covariate that helps control for loan supply. But, since the data on village covariates are from the earlier survey round, we present results with and without including them and do not find these to be significantly different.

We focus on farmer households in the rural sector, which we define as those households who report a) their main occupation as cultivation, and b) that they cultivate land. We focus on the 18 largest states (out of 28), and exclude Jammu and Kashmir, the North-Eastern states, Union Territories, Delhi and Goa, since these account for very few loans. Table 1 presents means and proportions for all variables by caste-group, and all statistics shown are population estimates that take survey weights into account, as do all our estimations.

Caste-wise means and proportions follow expected patterns in that land ownership, renting and leasing, incomes, and access to credit are all higher for Brahmins and OBCs, followed by SCs and then STs. It might seem surprising that average incomes are *lower* than average consumption expenditure, even though the former is an annual figure and the latter monthly. The survey documentation (Desai and Vanneman, 2015) offers two reasons for this: first, several incomes are negative, reflecting the precarious nature of cultivator livelihoods. Second, calculating incomes is notoriously difficult when several components are measured in kind.[10]

3 Empirical framework

Our analysis seeks to estimate the residual effects of caste belonging on the demand for loans, as well as their subsequent approval. We use logit specifications to model both the decision to apply for an agricultural loan, and subsequent loan approvals for those who applied. For household i, let Y_{1i} and Y_{2i} denote the binary outcomes of applying for an agricultural bank loan and, respectively, having this application approved and thus

---

[10]This pattern is also borne out by data from the National Sample Survey Organisation. See table 2 in Mishra (2008)

Table 1: Means and proportions by caste-group

N=8,543	Brahmin	OBC	SC	ST
Loans				
Proportion currently have loan	0.422	0.406	0.285	0.206
Proportion applied for loan	0.445	0.428	0.306	0.243
Proportion loan approved if applied (N=3,619)	0.947	0.947	0.930	0.849
Land and income				
Land owned (acres)	15.635	12.097	7.138	10.129
Land rented in (acres)	2.106	2.567	2.280	1.568
Land rented out (acres)	2.003	1.515	1.261	1.182
Monthly consumption expenditure (Rs '000)	133.255	109.646	85.422	62.919
Annual income (Rs '000)	133.087	90.237	66.278	51.817
Education and household composition				
Age of household head	53.787	52.076	51.103	49.249
Proportion with male household head	0.946	0.944	0.948	0.920
Years of education highest male	8.587	7.325	6.162	5.461
Years of education highest female	6.025	4.203	3.388	2.686
Household size	5.303	5.379	5.242	5.261
Household proportion adult males	0.380	0.356	0.340	0.341
Household proportion adult females	0.362	0.354	0.345	0.343
Village characteristics				
Distance to nearest bank branch (km)	5.705	5.942	6.543	8.500
Distance to nearest town/city (km)	14.411	15.537	16.493	22.113
Percentage of households with electricity	62.248	58.413	55.701	45.788
State population proportions				
Himachal Pradesh	0.673	0.076	0.223	0.029
Punjab	0.740	0.167	0.093	0.000
Uttarakhand	0.230	0.237	0.533	0.000
Haryana	0.575	0.352	0.071	0.002
Rajasthan	0.180	0.531	0.168	0.121
Uttar Pradesh	0.311	0.562	0.127	0.000
Bihar	0.278	0.622	0.087	0.013
West Bengal	0.484	0.185	0.296	0.035
Jharkhand	0.151	0.270	0.096	0.484
Orissa	0.102	0.523	0.170	0.205
Chhattisgarh	0.026	0.491	0.077	0.405
Madhya Pradesh	0.230	0.481	0.091	0.198
Gujarat	0.325	0.443	0.051	0.182
Maharashtra	0.375	0.471	0.060	0.094
Andhra Pradesh	0.319	0.480	0.174	0.027
Karnataka	0.265	0.587	0.062	0.086
Kerala	0.677	0.275	0.048	0.000
Tamil Nadu	0.117	0.712	0.157	0.015

Notes:
This table presents means and proportions of all covariates used according to caste group, viz. Brahmin, OBC, SC and ST. All statistics are estimates for the population that take into account survey probability weights.

receiving a loan. We model the probability of each outcome using a logit specification:

$$\Pr[Y_{1i} = 1 | \mathbf{X}] = \frac{e^{\mathbf{X}'\beta}}{1 + e^{\mathbf{X}'\beta}} \tag{1}$$

$$\forall i \quad \text{s.t.} \quad Y_{1i} = 1, \Pr[Y_{2i} = 1 | \mathbf{X}] = \frac{e^{\mathbf{X}'\beta}}{1 + e^{\mathbf{X}'\beta}} \tag{2}$$

$\mathbf{X}$ is a vector of covariates that includes caste-group dummies and β is a vector of coefficients that we seek to estimate. We use the same set of covariates to model loan applications and approvals, since *a priori* there is no theoretical rationale to exclude any of them.

Unbiased estimation of the residual effects of caste-group on loan applications and bank approval decisions requires controlling for characteristics that influence application decisions and respectively, signal creditworthiness. To this end we include variables known to influence access to credit (see, for example, Pal, 2002; Swain, 2007), including household income and land owned – the latter the primary form of collateral demanded by banks.[11]

Second, we must account for the supply of loans. For loan applications, distance to the nearest bank (including if a bank is present in the village) is an important indicator of the access to credit. This distance might also function as a proxy of banks' knowledge about potential borrowers, thus influencing approvals. We include this distance as a village-level covariate, together with two indicators of development and connectedness, viz. the proportion of households with electricity, and the distance to the nearest town/city. Unfortunately the data on these covariates are not ideal, since they are from an earlier round of the IHDS survey conducted in 2004-05. However the results are essentially identical with or without these village-level covariates, thus suggesting that this data limitation is not a serious problem.

We now focus on loan approvals. Analyses of racial discrimination in access to mortgage or small business loans typically control for past repayment histories and default behaviour (e.g. Blanchflower et al., 2003). Since credit ratings and the bureaus that maintain such information are not prevalent in the rural Indian context, it is unlikely that banks systematically use such information.[12] Nonetheless, credit histories or ratings could influence loan applications or approvals, and to the extent that they might be correlated with caste-group, their exclusion could lead to biased estimates. In particular, informa-

---

[11]According to the Reserve Bank of India's guidelines, banks should not demand collateral for loans of up to Rs 50,000, and for larger loans, land is the main form of collateral (Reserve Bank of India, 2007).

[12]In the case of lending to small businesses, banks are also known to gather 'soft' information on creditworthiness in the absence of credit ratings (Berger and Udell, 2002).

tion on repayment rates or credit histories is important to distinguish between statistical discrimination (owing to caste-repayment correlation), and taste-based discrimination.[13] We can infer the latter should residual caste-wise differences persist even after controlling for repayment histories.

However, the data do not provide information on repayment histories or a suitable proxy. Instead, we therefore examine their potential role through a Monte-Carlo simulation approach adapted from Ichino et al. (2008). This approach tests the sensitivity of our estimation results to the presence of an unobservable. Section 4.1 details the method, and the intuition is as follows. Excluding credit histories would lead to biased estimates if (unobserved) credit histories are correlated with caste-group as well as a bank's decision to lend. We can simulate a variable that is correlated with both a caste-group of interest and loan approvals and re-estimate our results. Using Monte-Carlo draws and varying the correlation between the simulated variable and, respectively, caste-group and loan approval, we can study the distribution of caste-group residuals by repeatedly re-estimating the model. The difference between these results and the estimates obtained without the simulated variable provide a handle on the sensitivity of results, and thus the potential importance of credit histories. It is important to note that credit history is only one potential unobservable. The analysis can be interpreted more generally, as accounting for the influence of *any* such potential unobservable that is correlated with caste-group and loan approvals.

Finally, it seems plausible to assume the absence of reverse causation. That is, in the short run at least, it is unlikely that loan applications, and more importantly the receipt of agricultural loans, would influence any of the right hand side variables.

4 Results

The first three rows of table 1 provide (unadjusted) caste-wise proportions in terms of access to credit. The first row confirms that there exist substantial inter-caste differences in loan access, and in particular, SC and ST borrowers have relatively low access to credit. These proportions are the type of statistic provided in official reports such as National Sample Survey Organisation (2005), and previous analyses essentially examine how these proportions change once borrower characteristics are accounted for, but do not differentiate between loan applications and their approval (e.g. Kumar, 2013). In con-

---

[13]Kumar (2013) uses data from the National Sample Survey Organisation which lists past repayments, and finds that these rates are not correlated with caste-group, but this analysis excludes STs.

Table 2: Logit estimation results

Model	Loan applications			Loan approvals		
	OBC	**SC**	**ST**	**OBC**	**SC**	**ST**
(1) State dummies	0.892 (0.081)	0.531 (0.062) p=0.000	0.411 (0.055)	1.062 (0.272)	0.823 (0.289) p=0.010	0.368 (0.130)
(2) Model 1 plus land and income[a]	1.041 (0.095)	0.744 (0.092) p=0.000	0.589 (0.085)	1.116 (0.278)	0.948 (0.329) p=0.034	0.435 (0.149)
(3) Model 2 plus education and household characteristics[b]	1.076 (0.099)	0.784 (0.099) p=0.000)	0.642 (0.094)	1.220 (0.297)	1.151 (0.381) p=0.053	0.484 (0.167)
(4) Model 3 plus village characteristics[c]	1.075 (0.099)	0.783 (0.099) p=0.000	0.658 (0.099)	1.205 (0.299)	1.155 (0.384) p=0.080	0.523 (0.177)

Notes:
This table presents estimates from logit models in terms of odds ratios. Only the coefficients for caste dummy are shown, with the full regression results given in appendix A. The base category is Brahmins. All estimates are survey-weighted, and standard errors in parentheses account for village-level clustering as per the survey design. The p-values shown are for Wald tests for the null that all three caste dummies are jointly zero. Sample size is 8,543 for loan application models and 3,619 for loan approval models.
[a] Land owned, land rented out, land rented in, household monthly consumption expenditure, and annual income.
[b] Age and sex of household head, years of education of highest-educated male and female, household size, household proportion of adult males and females.
[c] Distance to nearest bank branch, distance to nearest town/city, proportion households with electricity.

trast, this paper aims to investigate the role of caste in both applications and approvals. The second row shows that inter-caste differences in loan applications might indeed help explain patterns of overall access. Loan application rates follow the stereotypical caste hierarchy, with Brahmins the most likely to apply, followed by OBCs, SCs, and finally STs; a similar pattern to that of average consumption expenditures and incomes for instance. The magnitudes of these differences are substantive, with STs about half as likely to apply for a bank loan as Brahmins.

The third row of table 1 suggests that loan approvals are also not uniform across all caste-groups: while 93%-95% of loan applications by Brahmins, OBCs and SCs are approved, the proportion for STs substantially lower, at 85%. Our approach aims to analyse the extent of residual inter-caste differences that remain in loan application and approval statistics once we control for household characteristics that proxy ability, credit-worthiness, and access.

Table 2 presents the estimation results from a series of logit models. These are presented

in terms of odds ratios for respective caste dummies. In all cases the Brahmin caste group is the base category, and full regression results are presented in appendix A. The bottom part of each row reports p-values from Wald tests of joint equality of all three caste dummies.

We estimate two sets of models, one for loan applications (columns 2-4) and one for their subsequent approval by a bank resulting in a loan (columns 5-7). For instance, the first row says that in a simple model that includes only state dummies, the odds ratio for an SC household applying for a loan (respectively, having that application approved) relative to Brahmins is 0.531 (respectively, 0.823). This simple model thus mirrors the unadjusted caste-wise proportions in table 1.

Subsequent rows in table 2 show how respective odds ratios change as additional variables are added to the model. Row 2 adds household-level characteristics including land owned, rented out and rented in, and income and consumption expenditure, while row 3 adds the age and sex of the household head, the years of education of the most educated male and female members, household size, and the proportion of adult males and females. Row 4 adds village-level information on distance to the nearest bank and to the nearest town/city, and the proportion of households with electricity. In general, as we move from simpler to more complex models, inter-caste differences in both applications and approvals are reduced as more covariates are controlled for. The odds ratios become closer to unity even though, as the p-values suggest, substantive inter-caste differences persist.

To examine how these estimation results relate to unadjusted the caste-wise proportions in table 1, we calculate the predicted probabilities of loan application and approval. We report the sample-averaged predicted probabilities corresponding to model 4 in table 3 (those for model 3 are very similar, as the odds ratios would suggest). These sample-averaged predicted probabilities have the useful interpretation of the treatment effect of caste-group belonging, provided that the estimation model has been correctly specified and that there are no omitted variables. For these predictions, all variables except caste are held at their sample values, and the caste variable is changed to Brahmin (say) to calculate the predicted probabilities for Brahmins, and likewise for other caste-groups. The decision to hold all non-caste variables at their sample values (rather then means, for instance) ensures that these calculations have a genuine interpretation as the actual probabilities that would have resulted were all households Brahmin (and so on). The bottom half of each row provides results from Wald tests for comparisons of these caste-wise predictions, and these show that inter-caste differences are also statistically significant.

Table 3: Predicted probabilities by caste group

Loan applications				Loan approvals			
Brahmin	**OBC**	**SC**	**ST**	**Brahmin**	**OBC**	**SC**	**ST**
0.405	0.419	0.357	0.324	0.937	0.947	0.945	0.892
(0.016)	(0.014)	(0.022)	(0.024)	(0.012)	(0.007)	(0.012)	(0.026)

All castes equal: p=0.000   All castes equal: p=0.218
                                         ST and non-ST equal: p=0.047

Notes:
This table shows sample-average probabilities by caste group predicted by model 4 in table 2.
Standard errors are given in parentheses. p-values correspond to Wald tests for respective null hypotheses.

Columns 2-5 of table 3 show that loan applications vary by caste, in a way that mirrors unadjusted proportions (table 1) but with attenuation. This pattern confirms that residual inter-caste differences in loan application rates are an important factor behind differences in overall access to credit. Next, columns 6-9 show that Brahmins, OBCs and SCs all have similar rates of loan approval, but that the corresponding rate for STs is about 5% lower.[14] This distinct residual difference despite controlling for a rich set of relevant covariates suggests that banks might discriminate against STs. So far we have focused on loan applications and approvals, but not loan amounts or interest rates. Banks can also discriminate by providing smaller amounts or lending at higher interest rates to certain caste-groups. We examine this possibility in appendix B and find that conditional on household and village characteristics, overall, there are no significant residual effects of caste-group belonging on loan amounts or interests rates.

Recall that while taste-based discrimination captures residual inter-caste differences after controlling for all relevant characteristics, one or more of these characteristics might be unobserved. In our case, while we have controlled for important indicators of credit-worthiness – in particular in the form of collateral (land) and incomes – the data do not have information about credit histories. Similarly, additional borrower characteristics might also influence a bank's decision to lend, such as the quality (and not just area) of land owned. If one or more of these unobserved characteristics are correlated with caste, then residual inter-caste differences in loan approvals might simply reflect statistical discrimination on the part of lenders. Since they do not have explicit information on, for instance, repayment abilities, but have found through past experience that certain

---

[14]The predicted approval rates STs in table 3 are lower than the average of approval rates for Brahmins, OBCs and SCs by 0.051 (S.E.= 0.026, p-value=0.047)

caste-groups have lower repayment rates, then a bank could simply use caste-group as a proxy for repayment abilities.

While the ethical appropriateness of this decision is questionable, it is not hard to see why lenders might resort to such decisions when driven by profit motives. In that case, the estimated residual differences using our data might in fact be a combination of statistical and taste-based discrimination. In order to examine whether the observed residual differences are solely due to taste, we need to control for potential sources of statistical discrimination.

We propose a procedure to investigate this through a simulation approach.

4.1   A simulation approach to unobserved credit histories

Suppose that credit histories influence banks' decisions to approve agricultural loans, and that as a group, STs have inferior credit histories compared to non-STs. Then, our finding that ST group-belonging has a residual, negative effect on loan approvals, might get partially or fully explained away by controlling for loan histories in equation 2. That is, while our results suggest that banks discriminate against STs, the absence of this variable in the data means that we cannot infer whether the discrimination is statistical – if STs have worse credit histories on average – or taste-based.

We approach this problem using Monte-Carlo simulation, adapting an approach suggested by Ichino et al. (2008) and Rosenbaum and Rubin (1983). Both papers investigate how including a hitherto binary unobservable might change the estimate of a treatment effect. They assume that assignment to treatment is not unconfounded if the unobservable (e.g. creditworthiness) is excluded from the estimation, but becomes so once it is controlled for. Ichino et al. (2008) implement this idea for treatment effects estimation using propensity score matching, and propose a simple way of parameterising the distribution of the binary unobservable. Partitioning the set of observations into four quadrants according to (binary) treatment status and (also binary) outcome, four corresponding probabilities that the unobservable equals one are specified. Monte-Carlo draws are used to construct the unobservable such that it conforms, on average, to this set of probabilities. Throughout the discussion of this simulation approach, we focus only on households who have applied for loans.

The treatment effect is then re-estimated for each draw having included the unobservable as a covariate, and this exercise is repeated across multiple draws and with different combinations of the probabilities characterising the unobservable. Using Ichino et al.

(2008)'s notation, if the number of simulations undertaken is m and the estimated treatment effect for the kth simulation is $\hat{ATT}_k$, then the overall estimate $\hat{ATT}$ is obtained as $\frac{1}{m}\sum_{k=1}^{m}\hat{ATT}_k$. If se_k^2 denotes the variance of $\hat{ATT}_k$ then the variance of $\hat{ATT}$ is calculated as (eq.11, p.321):

$$T = \frac{1}{m}\sum_{k=1}^{m}\text{se}_k^2 + \frac{m+1}{m(m-1)}\sum_{k=1}^{m}(\hat{ATT}_k - \hat{ATT})^2 \qquad (3)$$

We adapt Ichino et al. (2008)'s method for characterising the distribution of the unobservable, even as our set-up is more similar to Rosenbaum and Rubin (1983) who use a logit model to estimate the effect of a binary treatment on a binary outcome. Note that 'treatment' in our case is the categorical variable of caste-group belonging, and outcome is the binary loan-approval. Caste-group cannot be manipulated in reality, but conceptually we are studying the consequences that would arise as if it could. And, caste-wise differences in predicted loan approval probabilities calculated by averaging over the full sample (as we did in table 3) are in effect estimates of this treatment effect.

Since our focus is on potential taste-based discrimination against STs, for this part of the analysis the four caste categories can be simplified to just two: ST and non-ST. This simplification is also intuitively appealing: our results thus far suggest that differences in loan approvals between Brahmins, OBCs and SCs are negligible, and it is the gap between these groups and STs that potentially suggest institutional discrimination.

4.1.1 Constructing the unobservable

We consider only those households who have applied for a bank loan, and our focus is on loan approvals. Let H denote hitherto unobserved credit histories, where for household k, H_k can either be 1 (good history) or 0 (poor history).[15] Let $ST_k = 1$ if household k belongs to a Scheduled Tribe, and $ST_k = 0$ for all other caste-groups. And, let Y denote loan approval outcomes, with $Y_k = 1$ if the loan is approved and $Y_k = 0$ if it is denied.

In the following, we omit the subscript k for clarity unless otherwise indicated. We can partition households into four mutually exclusive groups based on their caste-belonging and loan approval outcomes: $\{Y = i, ST = j\}, i, j \in \{0,1\}$. For instance, $\{Y = 0, ST = 1\}$ denotes ST households who have been denied loans. Let P_{ij} be the probability that

---

[15]The choice of a binary confounder is a simplification over a more generalised, continuous confounder. Ichino et al. (2008) show, through Monte-Carlo simulation, that this choice is conservative: if results are in fact vulnerable to the inclusion of a confounder, then a simulation exercise using a binary instead of continuous confounder is less likely to lead to the (mistaken) conclusion that results are robust.

households with loan approval status $Y = i$ and caste group $ST = j$ have a good credit history.[16] That is

$$P_{ij} = Pr(H = 1 | Y = i, ST = j) \qquad i, j \in \{0, 1\} \tag{4}$$

Then the set $\{P_{00}, P_{01}, P_{10}, P_{11}\}$ fully characterises the distribution of H. An intuitive way of interpreting these probabilities in terms of a bank's decision to approve loans is as follows: let the total sample population be N and θ_{ij} denote the proportion of this sample in each of the sets $\{Y = i, ST = j\}$ where $i, j \in \{0, 1\}$[17]:

$$\theta_{ij} = \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{ij}(k), \tag{5}$$

where $\mathbb{1}_{ij}$ is the indicator function:

$$\mathbb{1}_{ij}(k) = \begin{cases} 1 & \text{if } k \in \{Y = i, ST = j\} \\ 0 & \text{otherwise} \end{cases}$$

In the absence of banks practicing taste-based discrimination against STs, the probability that a loan is approved conditional on good credit history will be the same for both ST and non-ST households. Denote this b. Then

$$b = Pr(Y = 1 \mid H = 1, ST = 1) = \frac{P_{11}\theta_{11}}{P_{11}\theta_{11} + P_{01}\theta_{01}} \tag{6}$$

and

$$b = Pr(Y = 1 \mid H = 1, ST = 0) = \frac{P_{10}\theta_{10}}{P_{10}\theta_{10} + P_{00}\theta_{00}} \tag{7}$$

Now let μ_{ST} (μ_{non}) denote the proportion of ST (non-ST) households with good credit histories. Then

$$\mu_{ST} = \frac{P_{11}\theta_{11} + P_{01}\theta_{01}}{\theta_{11} + \theta_{01}} \tag{8}$$

$$\mu_{non} = \frac{P_{11}\theta_{11} + P_{01}\theta_{01}}{\theta_{11} + \theta_{01}} \tag{9}$$

Together, equations 6-9 contain four unknowns (viz. $P_{10}, P_{11}, P_{00}, P_{01}$) while the θ_{ij} terms

---

[16]These definitions are along the lines of Ichino et al. (2008).

[17]The θ_{ij} terms are all sample statistics. For STs, we have $\theta_{11} = 0.849$ and $\theta_{01} = 0.151$, while for non-STs we have $\theta_{10} = 0.945$ and $\theta_{00} = 0.055$.

are all sample estimates. Solving eqns. 6-9 we get:

$$P_{00} = (1 - b)\frac{\mu_{non}(\theta_{10} + \theta_{00})}{\theta_{00}}; \qquad P_{10} = b\frac{\mu_{non}(\theta_{10} + \theta_{00})}{\theta_{10}}$$

$$(10)$$

$$P_{01} = (1 - b)\frac{\mu_{ST}(\theta_{11} + \theta_{01})}{\theta_{01}}; \qquad P_{11} = b\frac{\mu_{ST}(\theta_{11} + \theta_{01})}{\theta_{11}}$$

The aim is to investigate whether poorer credit histories amongst STs could be an explanation for their lower rates of loan approval. We therefore assume that $\mu_{non} \geq \mu_{ST}$, and simulate unobservable credit histories H for different combinations of $\{b, \mu_{ST}, \mu_{non}\}$. Not all combinations are feasible, since each of $\mu_{ST}, \mu_{non}, b, P_{00}, P_{01}, P_{10}, P_{11}$ must all lie between 0 and 1, which can be verified using eq. 10.

4.1.2 Simulation results

With caste-group belonging simplified to ST and non-ST, the baseline difference in loan approval rates using all available covariates (i.e. model 4 in table 2) is 0.053 (s.e.= 0.026, p-value=0.042). Figure 1 summarises how this approval gap changes if we control for unobserved credit histories across a range of distributional assumptions. The figure consists of two parts: subfigure (a) provides the estimated loan approval gaps, while subfigure (b) provides the corresponding p-values using standard errors defined in eq. 3.

In this figure we have held μ_{ST}, the proportion of STs with good credit histories, at 0.1. To put this into perspective, we know that about one in four STs applied for a loan (table 1), so $\mu_{ST} = 0.1$ assumes that just under half this proportion have good credit histories.[18] With $\mu_{ST} = 0.1$, we assume that $\mu_{non} \geq 0.1$ since the aim is to examine whether higher loan approval rates for non-STs might in fact arise from better credit histories. In appendix C we provide further results holding $\mu_{ST} = 0.25$ and $\mu_{ST} = 0.5$, which we find to be qualitatively identical to those with $\mu_{ST} = 0.1$.

At each point in figure 1, the distribution of simulated credit histories H is given by μ_{non} along the y-axis and b along the x-axis. Thus, loan approval becomes more dependent on credit history as we move along the x-axis, and the proportion of non-STs with good credit histories rises as we move along the y-axis. The figure is constructed using Akima interpolation based on a grid of 558 points with 200 replications per point.[19]

---

[18]Of course everyone with a good credit history need not apply for a loan.

[19]The blank spaces arise from unfeasible combinations of of $\{\mu_{non}, \mu_{ST}, b\}$ following eq. 10.

How does our original result change once we account for unobservable credit histories? Not very much. The figure shows that ST loan approval rates remain at least 4-6% lower than those for non-STs, and this gap can potentially widen, for a range of distributional assumptions concerning H. The darker shades in subfigure (a) denote the area where our result remains valid, while subfigure (b) shows that the associated p-values are statistically significant. Regardless of the proportion of non-STs being multiple times that of STs, the approval gap still remains – indeed can increase – even if banks increasingly base loan approvals on credit histories; i.e. $0.5 \leq b \leq 0.9$.

The approval gap reduces and can approach zero if $b > 0.9$ provided $\mu_{non} \geq 0.2$. Thus, if loan approval is strongly conditional on credit histories ($b > 0.9$) *and if* the proportion of non-STs with good credit histories is at least *double* that of STs (0.2 vs 0.1), then the evidence towards taste-based discrimination weakens. Notice that the region where the result is weakened is relatively smaller, and it is here that loan approval gaps can be explained by statistical discrimination.

Do banks taste-discriminate against ST borrowers? The results from this simulation exercise suggest that likely answer is yes, since the residual effects of caste-group belonging are robust to the inclusion of hitherto unobserved credit histories for most, but not all specifications. Provided bank lending is not strongly predicated on good credit histories, the gap between STs and non-STs in most cases remains at least 4-6% even after accounting for credit histories.

To recap, our analysis so far suggests the presence of inter-caste differences in loan application and approval rates. However, these aggregate results can potentially mask diverse patterns across states and individuals. Below, we reexamine these results by dividing our sample along two specific dimensions. The first of these is land ownership, motivated by the RBI's policy directive to banks to enable access to small farmers, defined as those who own 5 acres or less. The second is the relative demographic presence of STs at the state level, motivated by possibility of potential gains that their greater numeric strength might bring about.

4.2   Small farmers

The Reserve Bank of India has long recognised the challenges faced by small and marginal farmers in accessing agricultural loans. Its policy guidance has encouraged banks to set aside a proportion of overall lending for the 'priority sector', and within this certain 'weaker sections', which include small farmers (Reserve Bank of India, 2007). Small
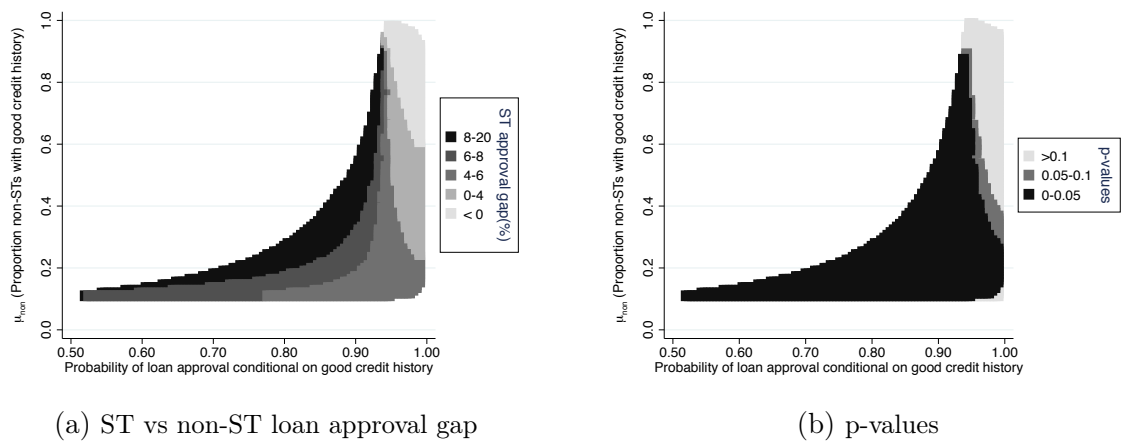
(a) ST vs non-ST loan approval gap   (b) p-values

Figure 1: Simulation results

Notes: This figures summarise the results from the simulation exercise where we hold μ_{ST}, viz. the proportion of STs with good credit histories, at 0.1. Values for μ_{non}, viz. the proportion of non-STs with good credit histories, are along the y-axis, while values for b, viz. the probability that banks will approve a loan conditional on good credit history, are along the x-axis. Each point thus corresponds to a given distribution of credit histories H according to combinations of $\{b, \mu_{ST} = 0.1, \mu_{non}\}$. For each combination, the simulation is performed 200 times to re-estimate model 4 in table 2 with caste simplified to ST and non-ST, and with H included as an additional regressor. For each regression, we calculate the difference in sample average predicted probabilities of loan approval for STs and non-STs. Subfigure (a) shows the average of these differences across simulations, i.e. the likelihood by which STs are less likely to have a loan approved compared to non-STs once credit histories are controlled for. Subfigure (b) shows the corresponding p-values calculated using equation 3. The simulations are based on a grid of 558 values with 200 simulations per point. These values are then interpolated using Akima interpolation. The blank areas are due to several combinations of $\{b, \mu_{ST} = 0.1, \mu_{non}\}$ being unfeasible (see eq. 10).

17

farmers are defined by land holdings of 5 acres or less. We now estimate our results separately for small farmers and for farmers who own more than 5 acres of land, to examine whether there is supportive evidence towards this policy's effectiveness.[20]

The top half of table 4 presents these results. As in the preceding section, the model specification includes all available covariates, and appendix A presents the full regression. These results show that inter-caste differences in application rates are stark only for large farmers. For small farmers there exist minor differences suggestive of the pattern we witness in the full sample, but their magnitude is far smaller. Similarly, looking at loan approvals, potential discrimination against STs is apparent only for large farmers; for small farmers the loan approval rates are essentially indistinguishable across caste groups.

This suggests that RBI guidance for lending to weaker sections may well be proving effective, even though our data do not allow us to examine this question directly. Not only are loan approval rates essentially equal across caste groups for small farmers, they are also slightly higher than the corresponding rates for large farmers. The same goes for loan application rates, in that small farmers are slightly *more* likely to apply for loans, and significant inter-caste differences exist only for large farmers. It would appear that banks use less discretion in lending to small farmers, and having experienced this willingness to lend over time, small farmers are more likely to apply for loans.

4.3   States with significant ST populations

STs are widely regarded as one of the most disadvantaged groups in Indian society, and according to the 2011 census they constitute nearly 9% of the total population. Their distribution is quite uneven across the states, and as table 1 affirms, STs are a substantive proportion in a few states but a very small minority in most others.[21] We might expect that in states where STs are a relatively larger minority, they would garner greater political and economic agency, and thus be less susceptible to discrimination. Alternatively, in these states, they might also become a more visible target for discrimination if they are competing with other caste groups for limited resources. Which of these phenomena dominates can be tested by dividing the sample into two groups based on the relative populations of STs. Five states in our sample have an ST population proportion of at least

---

[20]To be clear, we are not claiming to empirically test whether this policy guidance has worked. Since our data are cross-sectional, we cannot examine changes over time, and thus relate changes in policy to changes in access. Instead, our focus is on the presence of caste-wise differences at a given point in time, within a wider policy context that has long emphasised access to credit for disadvantaged groups.

[21]See Kijima (2006) for a detailed background and discussion of the socio-economic status of STs.

Table 4: Predicted probabilities for subpopulation estimations

	Loan applications				Loan approvals			
	Brahmin	**OBC**	**SC**	**ST**	**Brahmin**	**OBC**	**SC**	**ST**
Small farmers (N=2788)	0.479 (0.039)	0.479 (0.033)	0.436 (0.040)	0.436 (0.050)	0.965 (0.013)	0.947 (0.015)	0.943 (0.021)	0.931 (0.030)
	All castes equal: p=0.464				All castes equal: p=0.544			
					ST and non-ST equal: p=0.486			
Large farmers (N=5755)	0.413 (0.021)	0.438 (0.017)	0.374 (0.028)	0.318 (0.028)	0.930 (0.014)	0.956 (0.009)	0.953 (0.013)	0.896 (0.028)
	All castes equal: p=0.000				All castes equal: p=0.099			
					ST and non-ST equal: p=0.072			
States with significant ST population (N=2619)	0.378 (0.036)	0.360 (0.023)	0.270 (0.033)	0.286 (0.027)	0.966 (0.017)	0.924 (0.016)	0.914 (0.030)	0.838 (0.042)
	All castes equal: p=0.013				All castes equal: p=0.011			
					ST and non-ST equal: p=0.013			
Remaining states (N=5924)	0.417 (0.018)	0.441 (0.017)	0.384 (0.026)	0.349 (0.040)	0.937 (0.012)	0.953 (0.008)	0.949 (0.013)	0.924 (0.032)
	All castes equal: p=0.038				All castes equal: p=0.626			
					ST and non-ST equal: p=0.501			

Notes:

This table shows sample-average predicted probabilities by caste group for respective subpopulations from Logit models using all available covariates. Thus these are the same as model 4 in table 2. Standard errors are given in parentheses, and p-values are for Wald tests for respective null hypotheses.

20%: Jharkhand (48%), Chhattisgarh (41%), Madhya Pradesh (20%), Orissa (21%) and Gujarat (18%).[22] We now examine loan application and approval rates by caste in these states and, separately, in the remaining 13 states where STs are very small minorities.

The bottom half of table 4 summarises the results for these two groups of states. The specification used is the same as model 4 in table 2, which uses all available covariates. Full regression results are presented in appendix A. These probabilities suggest that the lower loan approval rates for STs in the full sample arise from states where STs *are* a substantial proportion of the population. States where STs are a very small part of the population do have lower approval rates for ST, but these are only marginally lower than those for non-STs, a trend confirmed by the corresponding p-values as well. In other words, of the two phenomena outlined above, the possibility of discrimination due to being a visible minority seems to dominate any political and economic agency that STs can muster. This result is similar to Das et al's (2012) finding that poverty reduction rates are slower amongst STs who reside in states with significant ST populations.

Unfortunately we are unable to consider North-Eastern states in our analysis, where STs are a significant proportion of the population and the majority in some. This limitation is due to the sample size available. There are only 33 farmer households in our data who applied for credit in these states, of which 17 received a loan, thus making it impractical to use regression analysis to model any caste-wise differences. That said, these states are also culturally and economically isolated from other parts of India, have sparser bank networks, and are thus likely to be distinct from the 18 states we are focusing on.

5 Discussion

Patterns of differential access to bank loans by caste-group bear structural similarities to the insights from studies on labour market discrimination by gender and race. Is there evidence of discrimination against certain caste-groups? If so, what is the nature of this discrimination? We investigate these questions in the Indian context, distinguishing between statistical discrimination and taste-based discrimination. Our paper finds that Scheduled Tribe borrowers are, after controlling for relevant characteristics, less likely to have a loan application approved relative to other three caste groups (89% versus 94-95%). Through counterfactual simulations, we argue that statistical discrimination stemming

---

[22]In our data, population proportions as a whole and not just for farmers, are: Jharkhand (36%), Chhattisgarh (29%), Madhya Pradesh (16%), Orissa (17%) and Gujarat (13%). These are still the five states with the highest proportion of STs.

from banks using caste belonging as a proxy for unobserved characteristics (e.g. credit histories) is unlikely. That is, ST borrowers likely face taste-based discrimination.

Burgess and Pande (2005) show that overall access to bank loans for SCs and STs improved during 1980s due to banking policy even as it slightly regressed during the 1990s. Despite the absence of legislation against discrimination in access to credit unlike the Equal Credit Opportunities Act in the USA for instance, the RBI has long propagated lending directives to banks emphasising improved access for SCs and STs. This paper finds that the magnitude of difference in loan *approvals* is indeed very little across caste groups. Overall, this suggests that RBI guidance emphasising inclusion has proved largely successful, even as we find that STs face persistent, though marginally ($\sim$5%) lower rates of loan approval. Our finding that caste-wise differences are significantly muted for small farmers – those who own less than 5 acres of land – is again consistent with the RBI's guidance encouraging lending to weaker sections of society. Not only do we find no evidence of caste discrimination in lending to small farmers, but crucially, small farmer SCs and STs are more likely to apply for loans relative to the full sample.

Another finding concerns the prevalence of discrimination across states. We find that lending in states where STs constitute very small minorities in terms of population is less likely to exhibit taste-based discrimination. On the other hand, we find that taste-based discrimination in loan approvals exists in states where STs have a significant population, though still a minority. This suggests that having a significant demographic presence does not necessarily guarantee that discriminatory practices will weaken.

The process of getting a loan involves applying for one, and having this application approved. Typically, studies have focused on overall access and hence fail to distinguish between these two steps (e.g. Burgess et al., 2005; Kumar, 2013). Our analysis shows that loan application rates account for a major component of inter-caste differences in loan access. SCs and STs are less likely to apply for bank loans compared to OBCs and higher castes, and this pattern remains even after controlling for various relevant characteristics. Thus, dynamic patterns of overall access need to decomposed into application and approval rates to better comprehend the mechanism behind any changes due to policy. Our study, though cross-sectional, demonstrates the importance of this distinction by showing that application rates differ substantially across caste-groups.

Why do loan application rates differ by caste? Studies on labour market discrimination potentially offer insight into the role of expectations and demand. In these studies , part of the difference in pay (say) by gender or race can stem from a failure to demand higher pay. In other words, driven by expectations potentially based on experience, individuals

21

belonging to certain groups can resign themselves and forego better outcomes. This occurs regardless of any form of discrimination. In the same vein, it is possible that STs and SCs might hesitate to apply for bank loans because they have learned, as a community if not as individuals, that they are less likely to receive them.[23] A corroborating observation is that education and income levels, arguably linked to expectations as well, are on average lower for SCs and STs than for higher castes and OBCs. In this light, the efficacy of all policies aimed at creating a level playing field including better financial access, might be consequently constrained in the presence of pessimistic expectations.[24] Unfortunately this explanation cannot be tested with the data at hand, because expectations themselves are unobserved.

References

Altonji, J. G. and R. M. Blank (1999). Race and gender in the labor market. Volume 3, Part C, pp. 3143–3259. Elsevier.

Akerlof, G.(1976).The Economics of Caste and of the Rat Race and Other Woeful Tales. *The Quarterly Journal of Economics 90*(4), 599–617.

Arrow, K. J. (1973). The theory of discrimination. In Orley Ashenfelter and Albert Rees (Eds.), *Discrimination in labor markets.*, pp. 3–33. Princeton, N.J.: Princeton University Press.

Arrow, K. J. (1998).What has economics to say about racial discrimination? *The Journal of Economic Perspectives 12*(2), 91–100.

Asiedu, E., Freeman, J. A. and Nti-Addae, A. (2012). Access to credit by small businesses: How relevant are race, ethnicity, and gender? *American Economic Review 102*(1), 532–537.

Banerjee, A., M. Bertrand, S. Datta, and S. Mullainathan (2009). Labor market discrimination in Delhi: Evidence from a field experiment. *Journal of Comparative Economics 37*(1), 14-27.

---

[23]They might instead resort to informal sector borrowing, but these loans typically have higher interest rates and are on adverse terms compared to bank loans. In this paper our focus is on understanding access to formal bank loans.

[24]Dréze et al. (1997) have argued that the distinction between lender discrimination and differences in application rates is largely irrelevant, because any hesitation to apply for a loan in fact reflects differences in expectations that are based on experiences that might span several years or more.

Becker, G. (1971). *The economics of discrimination*. London: University of Chicago Press.

Berger,A. and G. Udell (2002).Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal 112* (477), F32-F53.

Bertrand, M. and S. Mullainathan (2004). Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review 94* (4), 991–1013.

Bertrand, M., D. Karlan, S. Mullainathan, E. Shafir, and J. Zinman (2010). What's advertising content worth? Evidence from a consumer credit marketing field experiment. *The Quarterly Journal of Economics 125* (1), 263–306.

Blanchard, L., B. Zhao, and J. Yinger (2008, March). Do lenders discriminate against minority and woman entrepreneurs? *Journal of Urban Economics 63* (2), 467–497.

Blanchflower, D. G., P. B. Levine, and D. J. Zimmerman (2003, November). Discrimination in the Small-Business Credit Market. *The Review of Economics and Statistics 85* (4), 930–943.

Burgess, R., R. Pande, and G. Wong (2005). Banking for the poor: Evidence from India. *Journal of the European Economic Association 3* (2-3), 268–278.

Chevalier, A. (2007). Education, occupation and career expectations: determinants of the gender pay gap for UK graduates*. *Oxford Bulletin of Economics and Statistics 69* (6), 819–842.

Conning, J. and C. Udry (2007). Rural financial markets in developing countries. In Robert Evenson, Prabhu Pingali, and T. Paul Schultz (Eds.), *Handbook of agricultural economics* (3rd ed.)., pp. 2857–2908. Amsterdam: Elsevier.

Desai, S. and R. Vanneman (2015). India Human Development Survey-II (IHDS-II), 2011-12. ICPSR36151-v2.

Deshpande, A. (2000, May). Does Caste Still Define Disparity? A Look at Inequality in Kerala, India. *The American Economic Review 90* (2), 322–325.

Dréze, J., P. Lanjouw, and N. Sharma (1997). Credit in rural India: a case study.

Duflo, E. and Rachel R. Glennerster and M. Kremer (2007).Using randomization in development economics research: A toolkit, chapter 61. In T. Paul Schultz and John A.

Strauss (Eds.), *Handbook of Development Economics, Vol. 4.*, pp. 3895–3962. Amsterdam: Elsevier

Government of India. (2007). *Report of the expert group on agricultural indebtedness.* New Delhi: Ministry of Finance.

Ichino, A., F. Mealli, and T. Nannicini (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics 23*(3), 305–327.

Ito, T. (2009, March). Caste discrimination and transaction costs in the labor market: Evidence from rural North India. *Journal of Development Economics 88*(2), 292–300.

Kijima, Y. (2006, January). Caste and Tribe Inequality: Evidence from India, 1983-1999. *Economic Development and Cultural Change 54*(2), 369–404.

Kochar, A. (1997). An empirical investigation of rationing constraints in rural credit markets in India. *Journal of Development Economics 53*(2), 339–372.

Kumar, S. M. (2013). Does Access to Formal Agricultural Credit Depend on Caste? *World Development 43*, 315–328.

LaLonde, R. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review 76*(4), 604-620.

Maitreyi Bordia Das, Gillette H. Hall, Soumya Kapoor, and Denis Nikitin (2012, April). India. In G. H. Hall and H. A. Patrinos (Eds.), *Indigenous Peoples, Poverty, and Development*, pp. 205–248. Cambridge University Press.

Mishra, S. (2008). Risks, farmers' suicides and agrarian crisis in India? *Indian Journal of Agricultural Economics 63*, 38–54.

National Sample Survey Organisation (2005). Household indebtedness in India as on 30.06.2002. Technical report, National Sample Survey Organisation, Ministry of Statistics and Programme Implementation, Government of India, New Delhi.

Pal, S. (2002). Household Sectoral Choice and Effective Demand for Rural Credit in India. *Applied Economics 34*(14), 1743–1755.

Prakash, A. (2015). *Dalit Capital: State, markets and civil society in urban India.* New Delhi: Routledge.

Ramaiah, A. (1992, June). Identifying Other Backward Classes. *Economic and Political Weekly 27*(23), 1203–1207.

24

Reserve Bank of India (2004). *Master circular: Priority sector lending facilities to Scheduled Castes (SCs) & Scheduled Tribes (STs)*. Mumbai: Reserve Bank of India.

Reserve Bank of India (2007). Report of the Working Group to Examine the Procedures and Processes of Agricultural Loans. Technical report, Reserve Bank of India, Mumbai.

Reserve Bank of India (2008). *Master circular - Lending to priority sector*. Mumbai: Reserve Bank of India.

Reserve Bank of India (2011). *Master Circular - Priority Sector Lending - Credit Facilities to SCs and STs*. Mumbai: Reserve Bank of India.

Reserve Bank of India (2014). *Priority Sector Lending - targets and classification*. Mumbai: Reserve Bank of India.

Rosenbaum, P. R. and D. B. Rubin (1983). Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *Journal of the Royal Statistical Society. Series B (Methodological)*, 212–218.

Sriram, M. S. (2007, October). Productivity of Rural Credit. *International Journal of Rural Management 3*(2), 245–268.

Swain, R. B. (2007, December). The demand and supply of credit for households. *Applied Economics 39*(21), 2681–2692.

Thorat, S. and P. Attewell (2007). The Legacy of Social Exclusion: A correspondence study of job discrimination in India. *Economic and Political Weekly*, 4141–4145.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. Mass: MIT Press.

Appendix A   Logit regression results

Tables 5 and 6 show the logit regression results for the full sample loan applications and approvals, respectively. All models contain state fixed effects and robust standard errors that account for village-level clustering.

Tables 7 and 8 shows subpopulation-wise Logit regression results for loan applications and approvals, respectively. These subpopulations are defined according to land ownership and ST population proportions. These are the full regression results for sections 4.3 and 4.2. All models contain state fixed effects and robust standard errors that account for village-level clustering.

Table 5: Logit regression results: loan applications

	Model 1	Model 2	Model 3	Model 4
Brahmin			(base)	
OBC	-0.115	0.0403	0.0732	0.0719
	(0.0905)	(0.0911)	(0.0916)	(0.0917)
SC	-0.633***	-0.295**	-0.244*	-0.245*
	(0.117)	(0.124)	(0.126)	(0.127)
ST	-0.890***	-0.528***	-0.444***	-0.419***
	(0.135)	(0.144)	(0.147)	(0.150)
(log) land owned		0.476***	0.463***	0.463***
		(0.0521)	(0.0514)	(0.0507)
(log) land rented in		0.109**	0.109**	0.105**
		(0.0478)	(0.0473)	(0.0480)
(log) land rented out		-0.103	-0.102	-0.104
		(0.0628)	(0.0639)	(0.0638)
(log) HH cons. exp.		0.537***	0.444***	0.450***
		(0.0628)	(0.0747)	(0.0750)
HH income		-0.000000315	-0.000000363*	-0.000000366*
		(0.000000193)	(0.000000189)	(0.000000188)
age HH head			0.0000993	0.000200
			(0.00275)	(0.00275)
male HH head			0.564***	0.567***
			(0.194)	(0.195)
Years of edu highest male			0.0399***	0.0396***
			(0.00900)	(0.00895)
Years of edu highest female			0.00714	0.00630
			(0.00851)	(0.00849)
HH size			-0.0141	-0.0139
			(0.0179)	(0.0179)
HH propn adult males			0.0798	0.0671
			(0.243)	(0.243)
HH propn adult females			-0.241	-0.238
			(0.286)	(0.286)
(log) distance nearest town				0.135**
				(0.0688)
% HHs with electricity				0.00360**
				(0.00172)
(log) distance closest bank branch				0.0302
				(0.0483)
Constant	-0.768***	-7.581***	-7.305***	-8.183***
	(0.159)	(0.717)	(0.844)	(0.906)
N	8543	8543	8543	8543

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All models contain state dummies

Table 6: Logit regression results: loan approvals

	Model 1	Model 2	Model 3	Model 4
Brahmin		(base)		
OBC	0.0605	0.110	0.199	0.187
	(0.256)	(0.249)	(0.244)	(0.248)
SC	-0.195	-0.0529	0.141	0.144
	(0.351)	(0.347)	(0.331)	(0.333)
ST	-1.000***	-0.832**	-0.727**	-0.648*
	(0.355)	(0.343)	(0.345)	(0.339)
(log) land owned		0.352**	0.365**	0.388**
		(0.160)	(0.160)	(0.166)
(log) land rented in		0.297**	0.307**	0.313**
		(0.145)	(0.142)	(0.143)
(log) land rented out		-0.130	-0.155	-0.174
		(0.166)	(0.179)	(0.174)
(log) HH cons. exp.		0.0882	0.173	0.147
		(0.168)	(0.168)	(0.166)
HH income		-4.73e-09	1.68e-08	-1.83e-08
		(0.000000449)	(0.000000448)	(0.000000403)
age HH head			-0.0182**	-0.0191**
			(0.00781)	(0.00788)
male HH head			0.422	0.439
			(0.350)	(0.344)
Years of edu highest male			-0.00757	-0.00964
			(0.0268)	(0.0273)
Years of edu highest female			0.0441	0.0415
			(0.0307)	(0.0311)
HH size			-0.0523	-0.0464
			(0.0399)	(0.0407)
HH propn adult males			2.088**	2.135**
			(0.868)	(0.886)
HH propn adult females			-0.0474	-0.0348
			(0.927)	(0.936)
(log) distance nearest town				0.0370
				(0.171)
% HHs with electricity				0.00457
				(0.00436)
(log) distance closest bank branch				-0.159
				(0.124)
Constant	3.981***	2.415	1.314	1.361
	(0.739)	(1.963)	(2.168)	(2.420)
N	3619	3619	3619	3619

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All models contain state dummies

Table 7: Subpopulation logit regression: loan applications

	Small farmers	Large farmers	Tribal states	Remaining states
Brahmin		(base)		
OBC	-0.000356	0.125	-0.0938	0.117
	(0.180)	(0.104)	(0.175)	(0.104)
SC	-0.238	-0.198	-0.602**	-0.165
	(0.206)	(0.162)	(0.245)	(0.143)
ST	-0.238	-0.496***	-0.503**	-0.349
	(0.267)	(0.178)	(0.233)	(0.223)
(log) land owned	0.963***	0.360***	0.572***	0.436***
	(0.177)	(0.0754)	(0.0839)	(0.0612)
(log) land rented in	0.163*	0.128**	-0.103	0.168***
	(0.0894)	(0.0576)	(0.0760)	(0.0590)
(log) land rented out	-0.208	-0.0990	-0.0853	-0.113
	(0.373)	(0.0658)	(0.102)	(0.0781)
(log) HH cons. exp.	0.334***	0.490***	0.663***	0.382***
	(0.128)	(0.0925)	(0.126)	(0.0886)
HH income	0.000000176	-0.000000418**	-0.00000104***	-0.000000165
	(0.000000333)	(0.000000196)	(0.000000390)	(0.000000180)
age HH head	0.00568	-0.00345	-0.00140	0.000110
	(0.00547)	(0.00309)	(0.00483)	(0.00330)
male HH head	0.275	0.665***	0.474*	0.594**
	(0.331)	(0.239)	(0.286)	(0.232)
Years of edu highest male	0.0327**	0.0422***	0.0569***	0.0365***
	(0.0158)	(0.0112)	(0.0137)	(0.0106)
Years of edu highest female	0.00987	0.00410	-0.0340**	0.0174*
	(0.0155)	(0.0108)	(0.0168)	(0.00961)
HH size	-0.00201	-0.0174	-0.0292	-0.0118
	(0.0360)	(0.0206)	(0.0295)	(0.0209)
HH propn adult males	0.409	-0.126	-0.473	0.236
	(0.474)	(0.285)	(0.419)	(0.294)
HH propn adult females	-0.533	-0.0808	0.594	-0.552
	(0.517)	(0.347)	(0.525)	(0.341)
(log) distance nearest town	0.0130	0.184**	-0.0631	0.215**
	(0.107)	(0.0793)	(0.104)	(0.0850)
% HHs with electricity	0.00186	0.00474**	0.00737**	0.00178
	(0.00264)	(0.00210)	(0.00292)	(0.00209)
(log) distance closest bank branch	-0.0317	0.0635	0.0186	0.0328
	(0.0760)	(0.0547)	(0.0871)	(0.0567)
Constant	-6.922***	-8.454***	-10.77***	-7.504***
	(1.528)	(1.154)	(1.399)	(1.077)
Observations	2788	5755	2619	5924

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All models contain state dummies

Table 8: Subpopulation logit regression: loan approvals

	Small farmers	Large farmers	Tribal states	Remaining states
Brahmin		(base)		
OBC	-0.535	0.526*	-1.004	0.321
	(0.465)	(0.296)	(0.655)	(0.273)
SC	-0.640	0.453	-1.161	0.220
	(0.577)	(0.394)	(0.759)	(0.356)
ST	-0.874	-0.471	-2.072***	-0.216
	(0.680)	(0.388)	(0.635)	(0.528)
(log) land owned	0.834**	0.235	0.872**	0.297*
	(0.354)	(0.242)	(0.434)	(0.179)
(log) land rented in	0.611***	0.321*	0.444	0.324*
	(0.236)	(0.189)	(0.351)	(0.166)
(log) land rented out	0.0957	-0.127	-0.169	-0.203
	(0.658)	(0.190)	(0.241)	(0.209)
(log) HH cons. exp.	0.0467	0.287	-0.332	0.150
	(0.355)	(0.194)	(0.331)	(0.194)
HH income	0.00000463*	-0.000000375	-0.00000186**	0.000000283
	(0.00000243)	(0.000000617)	(0.000000899)	(0.000000808)
age HH head	-0.0172	-0.0184**	-0.0383**	-0.0146
	(0.0138)	(0.00916)	(0.0167)	(0.00898)
male HH head	1.220**	0.0747	-0.0642	0.562
	(0.488)	(0.465)	(0.773)	(0.367)
Years of edu highest male	-0.0277	0.00948	-0.0464	-0.000203
	(0.0388)	(0.0359)	(0.0534)	(0.0310)
Years of edu highest female	0.138***	-0.00825	0.0641	0.0369
	(0.0483)	(0.0377)	(0.0543)	(0.0371)
HH size	-0.215***	0.0168	0.192**	-0.0908**
	(0.0796)	(0.0548)	(0.0753)	(0.0447)
HH propn adult males	2.996**	1.373	2.388*	2.162*
	(1.346)	(1.175)	(1.229)	(1.149)
HH propn adult females	0.202	0.268	1.867	-0.555
	(1.592)	(1.145)	(1.600)	(1.129)
(log) distance nearest town	-0.340	0.291	-0.197	0.154
	(0.271)	(0.216)	(0.324)	(0.199)
% HHs with electricity	-0.00954	0.0102**	0.00795	0.00257
	(0.00765)	(0.00501)	(0.00754)	(0.00499)
(log) distance closest bank branch	-0.138	-0.183	-0.237	-0.129
	(0.185)	(0.165)	(0.198)	(0.149)
Constant	3.474	13.63	3.664	1.158
	(4.676)	(.)	(4.342)	(2.766)
Observations	744	2845	1014	2605

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
All models contain state dummies

Appendix B   Loan amounts and interest rates

Our main analysis focuses on the binary event of obtaining a loan. In this appendix, we examine whether loan amounts or interest rates differ systematically by caste. There is inherent sample selection in the data for loan amounts and interest rates because the survey questionnaire asks about these attributes only for the largest loan borrowed by the household in the five years preceding the survey. 70% (survey weighted) of the farmer households in our sample who received a bank loan declare this to be their largest loan and thus provide information on amounts and interest rates. 62% of these (i.e. 43% of the total) declare that this loan was for agriculture or agricultural equipment purchase.

The top panel of table 9 summarises average and regression-adjusted annualised interest rates according to caste-group. The regression-adjusted values are calculated as linear predictions from a model that includes all the covariates used in model 4 in table 2, that is, caste, land and income, education and household composition, village characteristics, and state dummies. Interest rates on agricultural bank loans do differ by caste group, with the lower caste-groups, viz. SCs and STs, reporting *lower* interest rates than Brahmins or OBCs. This pattern persists even after borrower characteristics are accounted for, though the inter-caste differences become smaller. This should not be surprising, because the RBI encourages agricultural lending to SCs and STs at lower rates of interest under the 'Differential Rate of Interest Scheme' (Reserve Bank of India, 2011). For bank loans not restricted to agricultural purposes, the same pattern is visible but inter-caste differences are no longer statistically significant.

Loan amounts are summarised in the bottom panel of table 9. In contrast to interest rates, these reflect the conventional caste-group hierarchy, with Brahmins receiving larger loans, followed by OBCs, SCs and STs in decreasing order. This order is visible for all bank loans. Regression-adjustment in this table is performed using Poisson regression models, where the dependent variable is loan amounts and the covariates used are as in model 4 in table 2.[25] The regression-adjusted results show that inter-caste differences are no longer statistically significant once socio-economic characteristics are controlled for.

Overall, these results suggest that conditional on household and village characteristics, loan amounts and interests rates do not differ significantly by caste-group with the possible exception of interest rates on agricultural loans.

---

[25]See section 19.2 of Wooldridge (2010) for a discussion on the rationale behind using Poisson regression when the dependent variable has a highly skewed distribution, as our data for loan amounts does.

Table 9: Interest rates and loan amounts by caste

	Unadjusted means				Regression adjusted			
	Brahmin	**OBC**	**SC**	**ST**	**Brahmin**	**OBC**	**SC**	**ST**
Interest rates (annual %)								
All loans	8.350	6.160	6.228	5.729	9.162	6.686	5.797	5.200
	(28.072)	(4.873)	(4.514)	(5.164)	(2.281)	(0.664)	(0.620)	(1.050)
	All castes equal: p=0.270				All castes equal: p=0.482			
Agricultural loans	7.082	5.999	6.209	5.434	6.921	6.003	6.100	5.702
	(4.459)	(4.692)	(4.282)	(4.983)	(0.297)	(0.266)	(0.501)	(0.594)
	All castes equal: p=0.028				All castes equal: p=0.052			
Loan amount (Rs. '000)								
All loans	140.906	129.985	73.342	61.891	102.173	113.576	97.204	83.678
	(265.917)	(347.325)	(120.262)	(109.888)	(8.831)	(16.249)	(15.868)	(18.228)
	All castes equal: p=0.000				All castes equal: p=0.538			
Agricultural loans	128.837	108.926	47.138	56.661	76.243	76.286	59.769	79.261
	(303.032)	(195.627)	(69.693)	(104.808)	(6.189)	(6.659)	(8.540)	(17.370)
	All castes equal: p=0.000				All castes equal: p=0.355			

Notes:
This table shows caste-wise average interest rates and loan amounts for households. Columns 1-4 provide unadjusted means with standard deviations in parentheses, while columns 5-8 provide regression adjusted predictions with standard errors in parentheses. The p-values are for Wald tests for respective hypotheses. The regression model for interest rates is OLS using all available covariates, as in model 4 in table 2. For loan amounts, we use a poisson regression model using the same covariates. N=2548 for bank loans for any purpose, and N=1516 for agricultural loans.

Appendix C   Additional simulation results

(a) ST vs non-ST loan approval gap ($\mu_{ST} = 0.25$)

(b) p-values ($\mu_{ST} = 0.25$)

(c) ST vs non-ST loan approval gap ($\mu_{ST} = 0.5$)
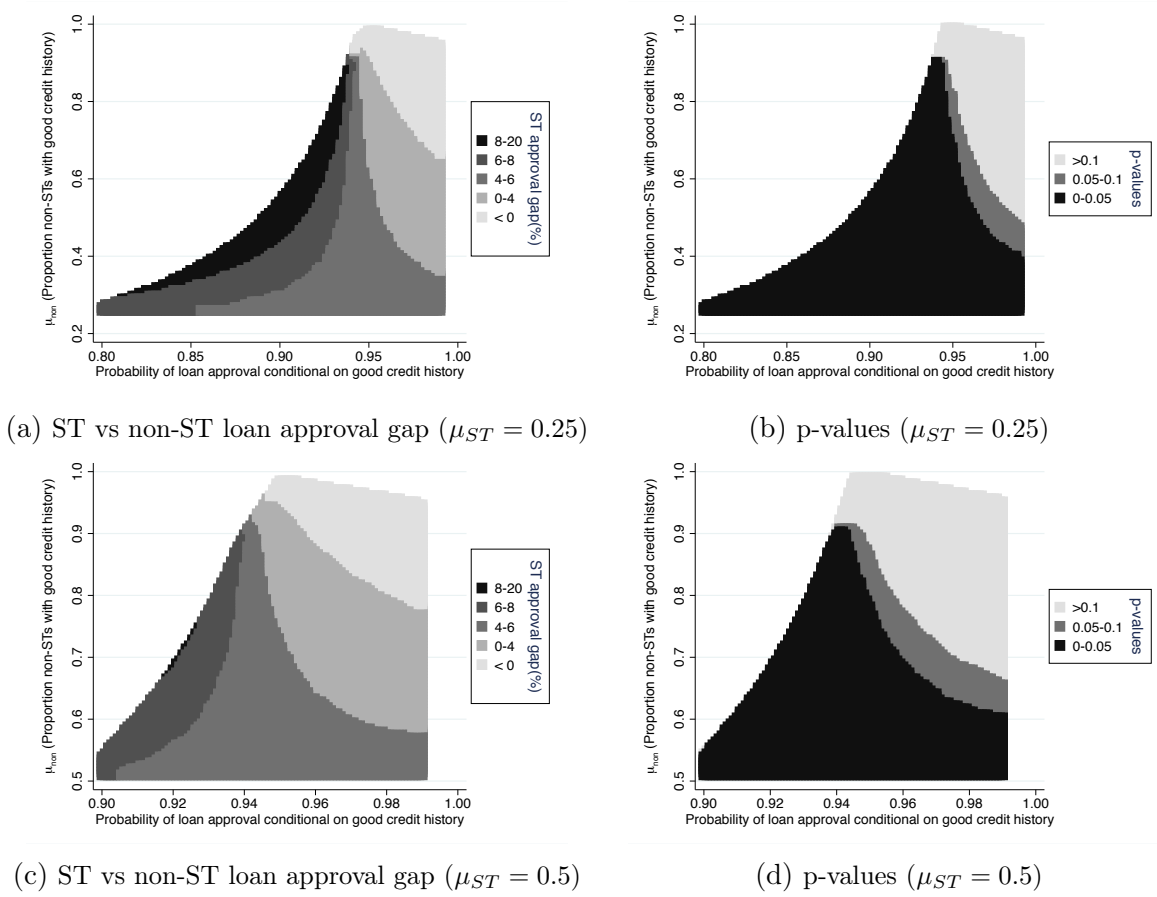
(d) p-values ($\mu_{ST} = 0.5$)

Figure 2: Simulation results for $\mu_{ST} = 0.25$ and $\mu_{ST} = 0.5$

Notes: This figures summarise the results from the simulation exercise. $\mu_{ST} = 0.25$ for subfigures (a) and (b), while $\mu_{ST} = 0.5$ for subfigures (c) and (d) where μ_{ST} denotes the proportion of STs with good credit histories. Values for μ_{non}, viz. the proportion of non-STs with good credit histories, are along the y-axis, while values for b, viz. the probability that banks will approve a loan conditional on having a good credit history, are along the x-axis. Each point in both figures thus corresponds to a given distribution of credit histories H according to combinations of $\{b, \mu_{ST} = 0.25, \mu_{non}\}$. For each combination, the simulation is performed 200 times to re-estimate model 4 in table 2 with caste simplified to ST and non-ST, and with H included as an additional regressor. For each regression, we calculate the difference in sample average predicted probabilities of loan approval for STs and non-STs. The left-hand side subfigures (a) and (c) show the average of these differences across simulations, i.e. the likelihood by which STs are less likely to have a loan approved than are non-STs once credit histories are controlled for. Subfigures (b) and (d) show the corresponding p-values calculated using equation 3. The simulations for $\mu_{ST} = 0.25$ are based on Akima interpolation using a grid of 694 values, while those for $\mu_{ST} = 0.5$ use a grid of 334 values, with 200 simulations for each point. The blank areas are due to several combinations of $\{b, \mu_{ST}, \mu_{non}\}$ being unfeasible since all four probabilities that characterise the distribution of H must lie between 0 and 1 using eq. 10.