Algorithmic Bias with Data Scarcity: Evidence from India

Abhiman Das Aarushi Kalra

Nirupama Kulkarni Advait Moharir*

June 17, 2026

ABSTRACT

FinTech lending has the potential to expand credit access by replacing relationship-based lending, where human discretion can lead to bias, with automated credit scoring based on large-scale data. Yet the effect of algorithms on bias is ambiguous in settings where many borrowers lack thick credit histories: when individual-level data are sparse, models may rely more heavily on aggregate signals that reflect past exclusion. This paper studies algorithmic bias in fintech consumer lending in India. Using a spatial difference-in-differences design and a market-level digitisation shock, we show that areas with historically thin credit bureau coverage receive significantly less fintech credit. This constraint is concentrated among first-time borrowers living in neighbourhoods with limited prior access to formal mortgage and consumer loans. These findings identify a distinct channel of algorithmic bias: unlike bias encoded in training data, this type of disparity arises when individual-level data are missing. Consistent with this mechanism, algorithmic bias is mitigated when thicker credit histories of minority groups in historically excluded

*Das: Indian Institute of Management Ahemdabad, abhiman@iima.ac.in; Kalra: University of Oxford, aarushi.kalra@economics.ox.ac.uk; Kulkarni: CAFRAL Reserve Bank of India, nirupama.kulkarni@cafral.org.in, Moharir: Michigan State University, moharir1@msu.edu

neighbourhoods become more readily available. Repayment behaviour provides suggestive evidence of misallocation: borrowers from excluded areas who receive credit repay at higher rates than comparable borrowers elsewhere. Our results are consisted with an imputed credit-scoring algorithm wherein lenders place greater weight on social identity markers when assessing borrowers from historically excluded geographies. Together, these results suggest that fintech algorithms misinterpret historical exclusion as current risk, reproducing disparities in credit access.

Keywords: digitisation, algorithms, inequality, redlining, banking

1 Introduction

Fintech lending has the potential to democratize credit access. By replacing relationship-based lending, where human discretion can lead to bias, with automated credit scoring based on large-scale data, fintech lenders may be able to reach borrowers historically underserved by traditional banks. Yet the effect of algorithms on bias is theoretically ambiguous (Kleinberg, Ludwig, Mullainathan, and Rambachan, 2018). While algorithms need not reproduce the same biases as human loan officers, they may generate new forms of exclusion when the data used to train and evaluate them reflect past inequalities (Kasy and Abebe, 2021; Fuster, Goldsmith-Pinkham, Ramadorai, and Walther, 2022; Bartlett, Morse, Stanton, and Wallace, 2022a).

Moreover, it remains unclear whether fintech lending diffuses historical bias or introduces new inequalities when borrower-level data that algorithms use to make lending decisions, is itself limited or unavailable. In such settings, limited individual-level data may push algorithms to rely more heavily on neighbourhood-level signals which may be correlated with borrowers' social identity markers and thus, a history of biased lending decisions (Fisman, Paravisini, and Vig, 2017; Asher, Jha, Novosad, Adukia, and Tan, 2026). This concern is especially important in less developed credit markets such as India, where many potential borrowers lack thick credit histories.

This paper asks whether lending algorithms reduce, reproduce or introduce bias in credit allocation, especially when borrower-level information is scarce. Studying fintech consumer lending in India, we show that areas with historically thin credit bureau coverage receive less fintech credit, even though borrowers from these areas who do receive loans repay at higher rates. The fact that new-to-credit (henceforth, NTC) borrowers are less likely to receive fintech credit than subprime borrowers within such neighbourhoods shows that AI introduces new disparities in credit access in data constrained environments. We are the first to provide causal evidence on augmented bias in lending algorithms when credit history is severely lacking.

In order to understand whether fintech lending reduces or reproduces historical exclusions, we first define a spatial measure of exclusion in lending markets as the excess lending in mortgages (or the "redlining" measure): the log difference in mortgages per capita at the pincode level and mortgages per capita at the district level. "Redlined" areas are those in the bottom quartile of the lending gap.[1] Our mortgage-level data comes from a large state-owned bank, which we use to define redlined neighbourhoods. The mortgage data is particularly suited for our purposes as it also records the location and identity of the borrowers at baseline.[2]

We leverage this variation in redlined pincodes in a spatial difference-in-differences design to causally estimate the effect of a market-level digitisation shock on AI-based fintech lending (Dube, Lester, and Reich, 2010; Raymond, 2023).[3] We construct neighbourhood pairs by identifying non-redlined pincodes lying within 5 km of the boundary of a redlined pincode. Our design then compares credit outcomes in redlined pincodes

[1]Historically, retail lending in India was predominantly mortgages and hence we focus on this loan segment to define redlined areas at baseline.

[2]This helps us characterize neighbourhoods with higher loan coverage as pincodes with lower share of minority population (SC, ST, OBC, or Muslim).

[3]The key identifying assumption in our design is that outcomes in redlined and non-redlined pincodes show similar trends, in the absence of the treatment, conditional on the above-mentioned fixed effects. An event study analysis shows that redlined and non-redlined neighbourhoods were on similar trends in line with the parallel trends assumption.

post-digitisation to *neighbouring* non-redlined pincodes. The data on outcomes comes from TransUnion which provides the universe of consumer and mortgage loans across pincodes by borrower types, including new-to-credit borrowers. For the main analysis we focus on fintech loans in the consumer lending segment, where most of the fintech lending in India is concentrated.[4]

After the digitisation shock, we show that redlined pincodes — that is, those with historically thin bureau coverage — receive significantly less fintech loans. New-to-credit (NTC) borrowers received 0.206 fewer loans per 1000 capita in redlined areas. Overall, our results indicate that fintech lenders reproduce historical disparities they were meant to disrupt in consumer loan credit markets. These disparities are also reproduced in prime and subprime segments and hence cannot be explained by differences in observable borrower creditworthiness.

In particular, the reduction in NTC lending is larger than reduction in subprime lending in redlined pincodes meaning that fintech algorithms amplify bias in lending when data on their lending histories is very limited. While algorithms perpetuate biases from training data where it is available (prime and sub-prime borrowers), this paper is the first to document algorithmic discrimination in a severely data-constrained environment. The algorithm is likely to use borrower data that is noisily correlated with credit history to make lending decisions as the credit history does not exist for NTC borrowers, by definition. In redlined pincodes, this amplifies bias as a data-constrained algorithm relies on credit histories of other borrowers in the same neighbourhoods.

Insofar as credit histories and social identities are spatially correlated, algorithms in data-scarce environments will further lead to discrimination. Limited availability of data on NTC borrowers can introduce bias as the algorithm then relies on other predictors of

[4]Consumer lending accounts for 94% of all fintech lending from the period January 2015-July 2022. We focus on the extensive margin, defined as the number of loans per 1000 capita, as the main dependent variable.

repayment, like users' social groups and neighbourhood, to arrive at the decisions. On the other hand, algorithmic bias in data constrained environments can be mitigated if data on the credit history of minorities in a neighbourhood is readily available.

Indeed, in redlined pincodes where minority borrowers have thicker credit histories, as evidenced by higher share of SC, ST, OBC or Muslim borrowers in that pincode at baseline, fintechs are more likely to lend to NTC borrowers in these neighbourhoods. This is consistent with algorithmic bias in data constrained environments, as we find evidence for externalities on NTC borrowers due to credit histories of minority communities that may be segregated into the same redlined neighbourhoods.

Finally, borrowers in excluded geographies who do receive credit default *less* than matched borrowers elsewhere indicating that fintech algorithms systematically misread historical exclusion as current risk. Thus, fintechs are overly conservative in lending in redlined areas which shows up in lower default rates across borrower types. Such biases are inconsistent with lender objectives. In subsequent work, we highlight these biases in an imputed algorithm that would have determined lending decisions by fintechs in two cases: when the algorithm maximizes long-run profits or when it optimizes short-run objectives like repayment rates (Dobbie, Liberman, Paravisini, and Pathania, 2021). A stronger reliance on neighbourhood characteristics or social identity markers would be consistent with the hypothesis that lending algorithms generate new types of bias in the absence of credit history data.

Overall, our results highlight the persistence of historical exclusion, by neighbourhood and social group, even as new technologies are introduced, and point towards the following implications for policy – expanding digital credit is not sufficient to ensure financial inclusion if the data used to allocate credit remain shaped by unequal past access. In markets with thin credit files, greater automation may require complementary investments in representative credit histories, alternative data governance, model auditing, and

evaluation based not only on default prediction but also on exclusion errors.

Related literature Our paper is draws on a large literature on discrimination in credit markets. Most of this literature has focused on the mortgage markets in the US. Bartlett, Morse, Stanton, and Wallace (2022b) and Ambrose, Conklin, and Lopez (2021) document racial disparities in mortgage and approval rates, while others (Bhutta and Hizmo, 2021; Bhutta, Hizmo, and Ringo, 2021) find no such disparity. In the Indian context, Fisman, Sarkar, Skrastins, and Vig (2020) and D'Acunto, Ghosh, Jain, and Rossi (2022), document evidence of in-group bias in lending in India. A part of this literature documents redlining — the historical practice of denying credit to areas with historically disadvantaged groups (Aaronson, Hartley, and Mazumder, 2021). Poor precision due to limited transactions in minority-concentrated neighbourhoods (Lang and Nakamura, 1993) can lead lenders to perceive minority borrowers as riskier and depress collateral values (Blackburn and Vermilyea, 2007; Ling and Wachter, 1998; Avery, Beeson, and Sniderman, 1999). While policy in India highlights such geographic disparity in access to credit, academic work is lacking (Sacchar, Hamid, Ooman, Basith, Basant, Majeed, and Shariff, 2006). We document how such disparities can persist even when credit becomes more readily available.

Second, our work also relates to the literature on segregation of minority households. These papers have documented how segregation can result in inequality in homeownership, public infrastructure (Asher, Jha, Novosad, Adukia, and Tan, 2026) and economic opportunities (Ananat, 2011; Cutler and Glaeser, 1997; Massey and Denton, 1993; Chetty, Hendren, and Katz, 2016; Shertzer, Twinam, and Walsh, 2018, 2021; Wilson, 2012; Kalra, 2021; Kulkarni, Bharadwaj, Das, and Kumar, 2024) perpetuating across generations (Derenoncourt, 2022). Our paper highlights how segregation can indirectly determine household access to credit, even if lenders do not actively discriminate borrowers.

Finally, this paper contributes to a burgeoning literature on algorithmic bias in eco-

nomic outcomes (Kleinberg, Ludwig, Mullainathan, and Rambachan, 2018; Kasy and Abebe, 2021; Raymond, 2023; Fuster, Goldsmith-Pinkham, Ramadorai, and Walther, 2022). This literature typically traces the source of the bias to the databases that AI models are trained on. Our paper shows an amplification of bias when no data is available for a class of borrowers, due to data-externalities imposed by similar borrowers in the neighbourhood. Future work will model this unevaluated source of bias explicitly.

2 Data and Measurement

Our study uses two key regulatory and proprietary datasets. This section describes each of them, while explaining how the relevant variables used for analysis were constructed. Table B.1 provides a glossary of key variables and definitions used throughout the paper.

2.1 Outcomes

Credit bureau data: Our first key dataset is from TransUnion CIBIL, which is India's largest credit registry. This dataset covers the universe of loans across lenders in India, and provides detailed information on loan amount, number of loans as well as number of defaults. The data is available at the pincode-month-year level, and is disaggregated by lender, loan type, as well as credit score band.

For the purpose of our analysis, we focus on consumer loans by fintechs, covering the period 2015Q1-2019Q4. Our key variables of interest are (i) number of loans per 1000 capita, and (ii) the default rate, which is defined as the number of loans that are 90 days past due from one year of issuance, divided by total number of loans.

2.2 Redlining

Mortgage loan data: Our second key dataset consists of mortgage loan data from a large Indian public sector bank, covering the period 2010-2020. The data is at the loan-level, and for each loan, a rich set of information is made available including loan amount,

interest rate, collateral, property value, and property cost, along with the pincode, district and state where the mortgage was approved. Crucially, borrower characteristics like age, caste, gender and religion are also recorded.

The raw data consists of 67,13,287 unique mortgages. A number of steps were taken to clean the data for the purpose of our analysis. First, we only retain loans for which the unique account identifier, date of loan approval, and either borrower caste or religion is non-missing. Second, wherever the address is available, we use Google Maps API to geocode mortgages. The geographical distribution of the subsample of geocoded mortgages (N=108,248) is shown in Figure A.1. Finally, we drop all loans after 2019Q4 (the end of our period of analysis), and aggregate the data to the pincode and pincode $\times$ quarter level. Our final dataset consists of 4,145,061 loans.

Geospatial data: For normalizing the credit data, we obtain pincode-level population estimates for the year 2015 from Meta.[5] Finally, we also obtain pincode-level boundaries provided by the Department of Posts, Government of India.[6]

Measurement: We begin by defining excess lending, which measures the discrepancy in lending between a pincode and a district. Specifically, this is given by

$$\text{Excess Lending}_p = \text{Log}(\text{ loans per capita}_p) - \text{Log}(\text{ loans per capita}_d) \tag{1}$$

for pincode p in district d. This is constructed using the pre-period cross-section of mortgage loans from 2010Q2-2014Q4. The distribution of the lending gap across pincode is shown in Panel A of Figure 1. To identify the severity of this lending gap, we then discretize this measure by dividing it into quartiles, as demonstrated in Panel B of Figure 1. We then define a redlined pincode as follows

[5]The data is available at: https://ai.meta.com/ai-for-good/datasets/high-resolution-population-density-maps/

[6]The data can be freely downloaded at: https://www.data.gov.in/catalog/all-india-pincode-boundary-geo-json

$$\text{Redlined}_p = \mathbb{1}\{\text{Excess Lending}_p \leq Q_1\} \tag{2}$$

where Q_1 refers to the first quartile of the excess lending. A pincode or neighbourhood is classified as redlined if the excess lending between the pincode and its corresponding district is very low, that is, a pincode is severely underserved relative to the district that it belongs to. Our measure is identity-agnostic in the sense that it does not depend on whether the recipients of credit belong to a historically marginalized group. Our measure captures the spatial aspect of discrimination, where, for example, a Hindu/Muslim in a redlined pincode gets lesser loans relative to a Hindu/Muslim in a non-redlined pincode.

Since our measure is spatial, we want to compare outcomes for treated pincodes to their non-treated neighbours rather than all non-treated pincodes. To this end, we define pincode neighbourhoods as non-treated (i.e. non-redlined) pincodes that lie within 5 km of a treated (redlined) pincode. 4418 unique treated pincodes with 10465 such nrighbouts were identified. Each pincode-nrighbout pair is assigned a neighbourhood-id, yielding 26,832 such pairs. Finally, the dataset is reshaped to be unique at the pincode $\times$ neighbourhood-id level. This allows us to incorporate neighbourhood-id fixed effects in our analysis, which purge out characteristics common to pairs, and allow relevant comparison of a treated pincode to its control neighbour.

Finally, we construct balanced panels as follows. Using the pincode-neighbourhood-id dataset as our base, we merge the pincode level redlining measure, as well as the credit and default variables from TransUnion CIBIL. Our period of analysis is 2015Q1-2019Q4, and consists of 53,664 unique pincode neighbourhood-id combinations. The variables included are total number of loans per 1000 capita as well as default rate, in basis points, for all, subprime, new-to-credit and prime loans. Our data is unique at the pincode $\times$ neighbourhood-id $\times$ year-quarter level and consists of 10,73,280 observations.

2.3 Descriptive Statistics

Table 1 presents summary statistics for the consumer loan data. Default rates are similar across redlined and non-redlined pincodes, and across credit score bands, ranging from 2% for prime borrowers to 6% for subprime borrowers, on average. In non-redlined pincodes, there are 17.42 loans per 1000 capita on average (Panel C), almost 8x higher than redlined pincodes, which have 2.63 loans on average (Panel B). This gap persists across credit score bands – in non-redlined pincodes, median lending is about 3x higher for new-to-credit, subprime, and prime borrowers, relative to redlined pincodes. This indicates that redlined pincodes see reduced credit supply across the board.

Table B.2 shows the summary statistics for key variables for the mortgage data, over the period 2010Q2-2019Q1. In redlined areas, a borrower gets a loan of ₹1.13 million on average (Panel B), compared to ₹1.44 million for a non-redlined pincode (Panel C). Average property values are higher in non-redlined areas (₹4.66 million) relative to redlined ones (₹3.45), while project tenures and collaterals are lower. The share of SC, ST and OBC borrowers are higher in redlined pincodes, while Muslim borrowers are roughly the same (3% vs 4%).

Figure 2 shows the evolution of lending in redlined and non-redlined pincodes, for fintechs and Banks. For both lenders, the credit gap is by and large constant in the pre-digitisation era, and widens after the introduction of UPI. Figure A.2 presents the post-digitisation correlation between excess lending and loan shares for different groups. Loan shares for SC, ST, OBC and Muslims are negatively related to excess lending, while this correlation is strongly positive for General category borrowers. Overall, this indicates that our measure indeed captures disparity in access to mortgage credit across social groups.

3 Empirical Strategy

We begin by testing if our proposed measure of redlining captures disparity in lending, across social groups.[7] To do so, we use the following empirical specification

$$Y_{pn} = \alpha_p + \gamma_n + \beta \times \text{Redlined}_p + u_{pn} \tag{3}$$

for pincode p and neighbourhood-id n, where Y_{pn} refers to number of mortgage loans per 1000 capita for General, SC,ST, OBC, Muslim and Other[8] borrowers in pincode p with neighbourhood-id n. α_p and γ_n refer to pincode and neighbourhood fixed effects respectively. Redlined_p is the indicator for a redlined pincode as defined in Equation 2. The data is aggregated over the period 2010Q2-2015Q4. Table 2 shows that mortgage credit access is lower in redlined pincodes for all social groups, in the pre-digitisation period. The absolute reduction is largest for General borrowers, reflecting their higher baseline mortgage origination level. The decline relative to the non-redlined average is broadly similar across groups [9] ($\approx$ 100 % decline).

To account for spatial autocorrelation, and the fact that pincodes are repeated across neighbourhood-ids, we employ twoway-clustered standard errors at the pincode and neighbourhood-id level (De Chaisemartin and Ramirez-Cuellar, 2024; Cameron, Gelbach, and Miller, 2011). Since pincodes repeat multiple times in our data, it is possible that a few pincodes mechanically contribute more to the regression estimates on the account of appearing as neighbours to multiple pincodes. Hence, we follow Kroft, Kucko, Lehmann, and Schmieder (2020) in inverse-weighting each pincode-quarter observation by the number of neighbourhood-ids in which it appears in that quarter.

[7] We also find that our redlining measure is associated with lower mortgage loans across credit score bands in the pre-digitisation period. The results are presented in Table B.3.

[8] Other borrowers consist of Buddhist, Jain and residual undenominated borrowers.

[9] This is computed by dividing the coefficient by the control mean. For example, for SC borrowers this is $\frac{-0.0973}{0.099} = -0.98$.

We then estimate the effect of digitisation on credit in redlined pincodes using the following spatial difference-in-differences design, which is our preferred specification

$$Y_{pnt} = \alpha_p + \gamma_{nt} + \beta \times \text{Redlined}_p \times \text{Post}_t + \epsilon_{pnt} \tag{4}$$

where Y_{pnt} refers to the number of fintech consumer loans per 1000 capita or the default rate for All, Subprime, NTC, and Prime scorebands, for pincode p in neighbourhood n in year-quarter t. The time period covered is 2015Q1-2019Q4. The time variation now allows for neighbourhood-id $\times$ year-quarter fixed effects, which allows comparison of outcomes within a neighbourhood in the same quarter.

The coefficient of interest, β, identifies the average post-digitisation change in lending/default outcomes in redlined pincodes relative to neighbouring non-redlined pincodes within the same matched pincode pair, after accounting for time-invariant pincode characteristics and pair-specific quarter shocks. The key assumption underlying the spatial DiD setup, is that, conditional on pincode and neighbourhood-time varying unobservables, lending in treated pincodes would follow the same trend as those in non-treated pincodes. We test for pre-trends with the following event-study specification

$$Y_{pnt} = \alpha_p + \gamma_{nt} + \sum_{\tau \neq -1} \beta_\tau \times \text{Redlined}_p \times \mathbb{1}_\tau + \epsilon_{pnt} \tag{5}$$

for pincode p in neighbourhood n in year-quarter t. The dependent variables are loans per 1000 capita and default rates for NTC and non-NTC borrowers respectively. β_τ captures the treatment effect on redlined pincode relative to non-redlined neighbour for time period τ which lies in 2015Q1-2019Q4. Figures 3 and A.3 show that this assumption likely holds in our setting, as no clear linear trends emerge in the case of banks or fintechs with respect to our outcomes of interest (Rambachan and Roth, 2023).

Finally, we examine the heterogeneous impact of digitisation, by social groups, using the following specification

$$Y_{pnt} = \alpha_p + \gamma_{nt} + \beta_0 \left(\text{Redlined}_p \times \text{Post}_t\right) + \sum_{g \in \mathcal{G}} \beta_g \left(\text{Redlined}_p \times \text{Share}_{gp} \times \text{Post}_t\right) + \sum_{g \in \mathcal{G}} \delta_g \left(\text{Share}_{gp} \times \text{Post}_t\right) + \varepsilon_{pnt}. \tag{6}$$

for pincode p in neighbourhood n in year-quarter t, where $\mathcal{G}$ denotes the set of included social groups, so that $g \in \{\text{SC, ST, OBC, Muslim, and Others}\}$. The time period covered is 2015Q1–2019Q4. The dependent variables are loans per 1000 capita and default rates. Each group share, Share_{gp}, is defined as the number of mortgage loans approved for group g in a pincode, divided by total mortgage loans approved in that pincode, in the cross-section spanning 2010Q2–2015Q4. The coefficients β_g capture how the post-UPI effect of redlining varies with the pincode-level loan share of group g, relative to the omitted General category.[10] The coefficients δ_g absorb differential post-period changes associated with group shares in non-redlined pincodes. The specification includes pincode and neighbourhood $\times$ quarter fixed effects.

4 Results

In this section, we present the results from our spatial difference-in-differences specification to study bias in consumer lending as credit markets become more digitised, especially when algorithms do not have enough data on newer borrowers.

4.1 Consumer Loans

Table 3 shows that borrowers in redlined neighbourhoods receive fewer consumer loans on average after the digitisation shock, relative to neighbouring pincodes that are not

[10] All loan shares sum to one, making the excluded category, General, the base group.

redlined at baseline. Not only do banks and fintechs make fewer loans to subprime and NTC (new-to-credit) borrowers in redlined pincodes, prime lending is also lower in these neighbourhoods compared to neighbouring pincodes during the post digitisation period. This points towards the persistence of potential bias in redlined pincodes.

Notably, the reduction in NTC lending is larger than the reduction in subprime lending in redlined pincodes, as a percentage of lending in the same sub-group in control pincodes. This means that fintech algorithms elevate bias in lending when data on their lending histories is very limited or non-existent. While algorithms perpetuate biases from training data where it is available (prime and sub-prime borrowers), this paper is the first to document discrimination when individual level data, that algorithms require to make lending decisions, is unavailable.

Because NTC borrowers, by definition, lack individual credit histories, the algorithm is likely to base lending decisions on noisier proxy signals, including borrower characteristics correlated with the credit histories of other lenders' customers in the same neighbourhood. In historically excluded pincodes, this reliance on local credit histories can amplify bias: the algorithm may interpret the limited formal credit experience of the neighbourhood as evidence of individual borrower risk. Table B.4 shows that this is driven by fintechs and not banks, where the former type of institutions are more likely to use AI in making lending decisions for consumer loans.

Insofar as credit histories and social identities are spatially correlated, algorithms in data-scarce environments will further lead to discrimination (Fisman, Paravisini, and Vig, 2017; Fisman, Sarkar, Skrastins, and Vig, 2020). Limited availability of data on NTC borrowers can introduce bias as the algorithm then relies on other predictors of repayment, like users' social groups and neighbourhood, to arrive at the decisions. This would imply that bias is amplified for NTC borrowers in neighbourhoods that are segregated by caste or religion, and that are also redlined (Kulkarni, Bharadwaj, Das, and Kumar, 2024).

On the other hand, algorithmic bias in data constrained environments can be mitigated if data on credit history of minorities in a neighbourhood is readily available. Indeed, Table 4 shows that in redlined pincodes where minority borrowers have thicker credit histories, as evidenced by higher share of SC, ST, OBC or Muslim borrowers in that pincode at baseline, fintechs are more likely to lend to NTC borrowers.

This is consistent with algorithmic bias in data constrained environments, as we find evidence for externalities on NTC borrowers from credit histories of minority communities that may be segregated into redlined neighbourhoods. Table B.5 provides further evidence of algorithmic bias in the absence of data as no such patterns are found in the case of banks using human appraisers to make lending decisions.

4.2 Defaults

Table 5 shows that fintech NTC borrowers in excluded geographies repay consumer loans at a higher rate than NTC borrowers in neighbouring non-redlined pincodes. This indicates that fintech algorithms systematically misread group level traits as current risk.

This provides direct evidence of bias as not only does the algorithm rely on group level traits, it is also making incorrect decisions while relying on such data. In fact, Table B.6 shows that default rates are comparable across NTC borrowers in different neighbourhoods for public and private sector banks.

Table 6 shows that default rates on fintech consumer to NTC borrowers are comparable across redlined and non-redlined pincodes, irrespective of the amount of credit history data available for minorities in the respective neighbourhoods. This means that algorithms are wrongly inferring repayment risks from thick credit histories of other borrowers in historically excluded neighbourhoods.

4.3 Imputed Algorithm

We highlight these biases in an imputed algorithm that would have determined lending decisions by fintechs in two cases: when the algorithm maximizes long-run profits, and when it optimizes short-run objectives like repayment rates (Dobbie, Liberman, Paravisini, and Pathania, 2021). This exercise has been left for future work at the time of this submission.

4.4 Robustness Checks

Alternative measures: We check if our results are robust to alternative measures of redlining by defining three other measures: First, using mortgage loan data from TransUnion CIBIL, we define a redlined indicator using the exact procedure as the original measure. This measure uses mortgage loan data from multiple lenders, and has wider coverage than our baseline measure. Second, we use the continuous excess lending measure, as defined in Equation 1. Finally, we use the excess lending measure constructed using the CIBIL mortgage data.[11] The results for the impact of digitisation are presented in Figure A.4 for lending (Panel A) and default rate (Panel B). The results are consistent with our baseline, with similar magnitudes and the same sign.

Alternative specification: We re-estimate our full set of results with two changes: One, we cluster our standard errors only at the pincode level instead of the two-way clustering done for the baseline, and two, we estimate the specifications without weights. The results are presented in Tables B.8-B.11. Again, we find that the results are consistent with our preferred specification above.

[11] Since larger values of the continuous measures indicate lesser disparity, we multiply these measures with -1 so that these and the indicator measures have consistent interpretation.

5 Conclusion

This paper shows that, in data-scarce environments, digitisation can reproduce historical patterns of exclusion. Studying consumer lending in India, we find that areas with historically thin formal credit coverage receive significantly less fintech credit after a market-level digitisation shock. These effects are especially pronounced for new-to-credit borrowers, for whom individual credit histories are limited or unavailable.

The findings speak to a central ambiguity in the economics of algorithmic decision-making. Algorithms can reduce discretion and standardize decisions, but they also depend on the data environments in which they are trained and deployed. When borrower-level histories are sparse, predictive models may place greater weight on aggregate or neighbourhood-level signals. If these signals encode the legacy of unequal access to formal finance, automated underwriting can convert historical exclusion into contemporary credit rationing.

Our results are consistent with this mechanism: fintech credit falls most sharply for new-to-credit borrowers in historically underserved areas, while access improves in redlined neighbourhoods where minority groups have thicker baseline credit histories. In our setting, the relevant margin is not simply whether algorithms use protected characteristics directly. Rather, the concern is that seemingly neutral variables, such as credit history, geography, and nrighbouthood-level repayment patterns, may proxy for social identity and past exclusion when individual-level data are scarce.

Several limitations should guide the interpretation of the results. First, we do not observe the actual underwriting algorithms used by fintech lenders. The evidence is therefore indirect: we infer algorithmic behavior from lending and repayment patterns, as well as from an imputed scoring exercise, rather than from the production models themselves. Second, we cannot observe whether banks in the sample use artificial intelligence

or machine-learning tools in their own lending decisions. This limits our ability to separate algorithmic bias from other institutional differences across lenders, such as branch networks, risk appetite, product design, or customer acquisition strategies. Future work will address these limitations, and model this novel source of algorithmic bias explicitly.

The results point to a broader policy implication. Expanding digital credit is not sufficient to ensure financial inclusion if the data used to allocate credit remain shaped by unequal past access. In markets with thin credit files, greater automation may require complementary investments in representative credit histories, alternative data governance, model auditing, and evaluation based not only on default prediction but also on exclusion errors. More generally, fairness in digital lending depends not only on the formal removal of human discretion, but also on whether algorithmic systems can distinguish true borrower risk from the historical absence of opportunity.

References

AARONSON, D., D. HARTLEY, AND B. MAZUMDER (2021): "The Effects of the 1930s HOLC "Redlining" Maps," *American Economic Journal: Economic Policy*, 13(4), 355–92.

AMBROSE, B. W., J. N. CONKLIN, AND L. A. LOPEZ (2021): "Does Borrower and Broker Race Affect the Cost of Mortgage Credit?," *The Review of Financial Studies*, 34(2), 790–826.

ANANAT, E. O. (2011): "The Wrong Side(s) of the Tracks: The Causal Effects of Racial Segregation on Urban Poverty and Inequality," *American Economic Journal: Applied Economics*, 3(2), 34–66.

ASHER, S., K. JHA, P. NOVOSAD, A. ADUKIA, AND B. TAN (2026): "Residential segregation and unequal access to local public services in India: Evidence from 1.5 m neighborhoods," Discussion paper, National Bureau of Economic Research.

AVERY, R. B., P. E. BEESON, AND M. S. SNIDERMAN (1999): "Neighborhood Information

and Home Mortgage Lending," *Journal of Urban Economics*, 45(2), 287—-310.

BARTLETT, R., A. MORSE, R. STANTON, AND N. WALLACE (2022a): "Consumer-lending discrimination in the FinTech era," *Journal of Financial Economics*, 143(1), 30–56.

——— (2022b): "Consumer-lending discrimination in the FinTech Era," *Journal of Financial Economics*, 143(1), 30–56.

BHUTTA, N., AND A. HIZMO (2021): "Do Minorities Pay More for Mortgages?," *The Review of Financial Studies*, 34(2), 763–789.

BHUTTA, N., A. HIZMO, AND D. RINGO (2021): "How Much Does Racial Bias Affect Mortgage Lending? Evidence from Human and Algorithmic Credit Decisions," *Philadelphia Fed working paper*, (2021–067), 1–44.

BLACKBURN, M., AND T. VERMILYEA (2007): "The role of information externalities and scale economies in home mortgage lending decisions," *Journal of Urban Economics*, 61(1), 71—-85.

CAMERON, A. C., J. B. GELBACH, AND D. L. MILLER (2011): "Robust Inference With Multiway Clustering," *Journal of Business & Economic Statistics*, 29(2), 238–249.

CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): "The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment," *American Economic Review*, 106(4), 855–902.

CUTLER, D. M., AND E. L. GLAESER (1997): "Are Ghettos Good or Bad?," *The Quarterly Journal of Economics*, 112, 827–872.

D'ACUNTO, F., P. GHOSH, R. JAIN, AND A. G. ROSSI (2022): "How costly are cultural biases?," LawFin Working Paper Series 34, Goethe University, Center for Advanced Studies on the Foundations of Law and Finance (LawFin).

DE CHAISEMARTIN, C., AND J. RAMIREZ-CUELLAR (2024): "At what level should one cluster standard errors in paired and small-strata experiments?," *American Economic Journal: Applied Economics*, 16(1), 193–212.

DERENONCOURT, E. (2022): "Can You Move to Opportunity? Evidence from the Great Migration," *American Economic Review*, 112(2), 369–408.

DOBBIE, W., A. LIBERMAN, D. PARAVISINI, AND V. PATHANIA (2021): "Measuring Bias in Consumer Lending," *The Review of Economic Studies*, 88(6), 2799–2832.

DUBE, A., T. W. LESTER, AND M. REICH (2010): "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties," *The Review of Economics and Statistics*, 92(4), 945–964.

FISMAN, R., D. PARAVISINI, AND V. VIG (2017): "Cultural proximity and loan outcomes," *American Economic Review*, 107(2), 457–492.

FISMAN, R., A. SARKAR, J. SKRASTINS, AND V. VIG (2020): "Experience of Communal Conflicts and Intergroup Lending," *Journal of Political Economy*, 128(9), 3346–3375.

FUSTER, A., P. GOLDSMITH-PINKHAM, T. RAMADORAI, AND A. WALTHER (2022): "Predictably Unequal? The Effects of Machine Learning on Credit Markets," *Journal of Finance*, 77(1), 5–47.

FUSTER, A., P. GOLDSMITH-PINKHAM, T. RAMADORAI, AND A. WALTHER (2022): "Predictably Unequal? The Effects of Machine Learning on Credit Markets," *The Journal of Finance*, 77(1), 5–47.

KALRA, A. (2021): "A'Ghetto'of One's Own: Communal Violence, Residential Segregation and Group Education Outcomes in India," Discussion paper, Center for Open Science.

KASY, M., AND R. ABEBE (2021): "Fairness, Equality, and Power in Algorithmic Decision-Making," in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pp. 576–586, New York, NY, USA. Association for Computing Machinery.

KLEINBERG, J., J. LUDWIG, S. MULLAINATHAN, AND A. RAMBACHAN (2018): "Algorithmic fairness," in *Aea papers and proceedings*, vol. 108, pp. 22–27. American Economic

Association 2014 Broadway, Suite 305, Nashville, TN 37203.

KROFT, K., K. KUCKO, E. LEHMANN, AND J. SCHMIEDER (2020): "Optimal income taxation with unemployment and wage responses: A sufficient statistics approach," *American Economic Journal: Economic Policy*, 12(1), 254–292.

KULKARNI, N., P. BHARADWAJ, A. DAS, AND N. KUMAR (2024): "Disparities in Home Loans: Evidence from India," CAFRAL Working Paper.

LANG, W. W., AND L. I. NAKAMURA (1993): "A Model of Redlining," *Journal of Urban Economics*, 33(2), 223—-234.

LING, D. C., AND S. M. WACHTER (1998): "Information Externalities and Home Mortgage Underwriting," *Journal of Urban Economics*, 44(3), 317—-332.

MASSEY, D., AND N. A. DENTON (1993): *American Apartheid: Segregation and the Making of the Underclass*. Harvard University Press.

RAMBACHAN, A., AND J. ROTH (2023): "A more credible approach to parallel trends," *Review of Economic Studies*, 90(5), 2555–2591.

RAYMOND, L. (2023): "The market effects of algorithms," Discussion paper, Working Paper.

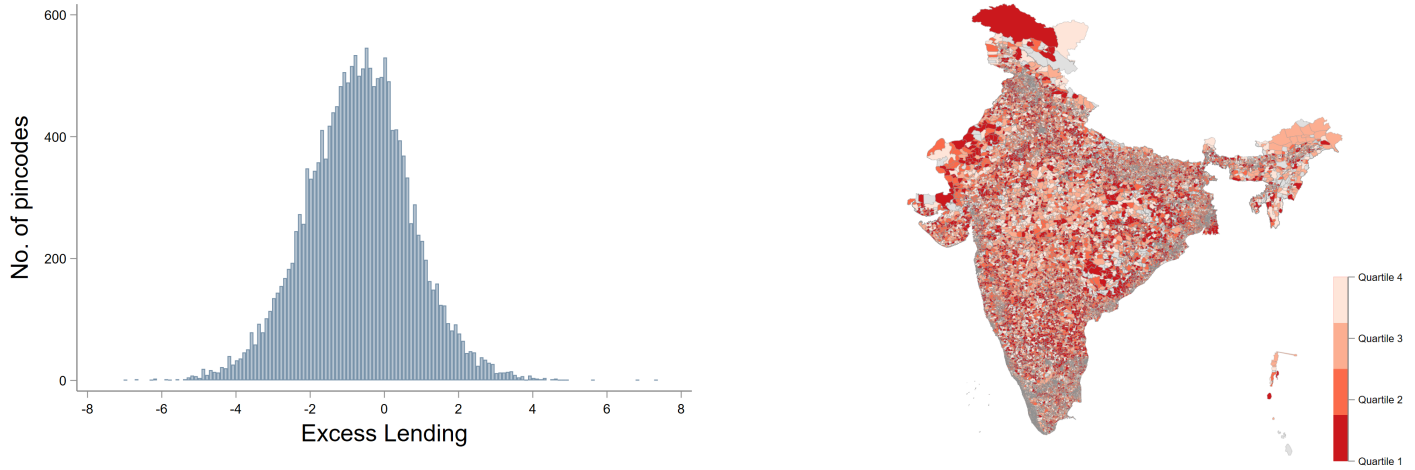
SACCHAR, R., S. HAMID, T. K. OOMAN, M. A. BASITH, R. BASANT, A. MAJEED, AND A. S. SHARIFF (2006): "Social, Economic and Educational Status of the Muslim Community of India. Technical report, Government of India," Discussion paper, Ministry of Education.

SHERTZER, A., T. TWINAM, AND R. P. WALSH (2018): "Zoning and the Economic Geography of Cities," *Journal of Urban Economics*, 105, 20–39.

——— (2021): "Zoning and Segregation in Urban Economic History," *Regional Science and Urban Economics*, p. 103652.

WILSON, W. J. (2012): *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*. University of Chicago Press.

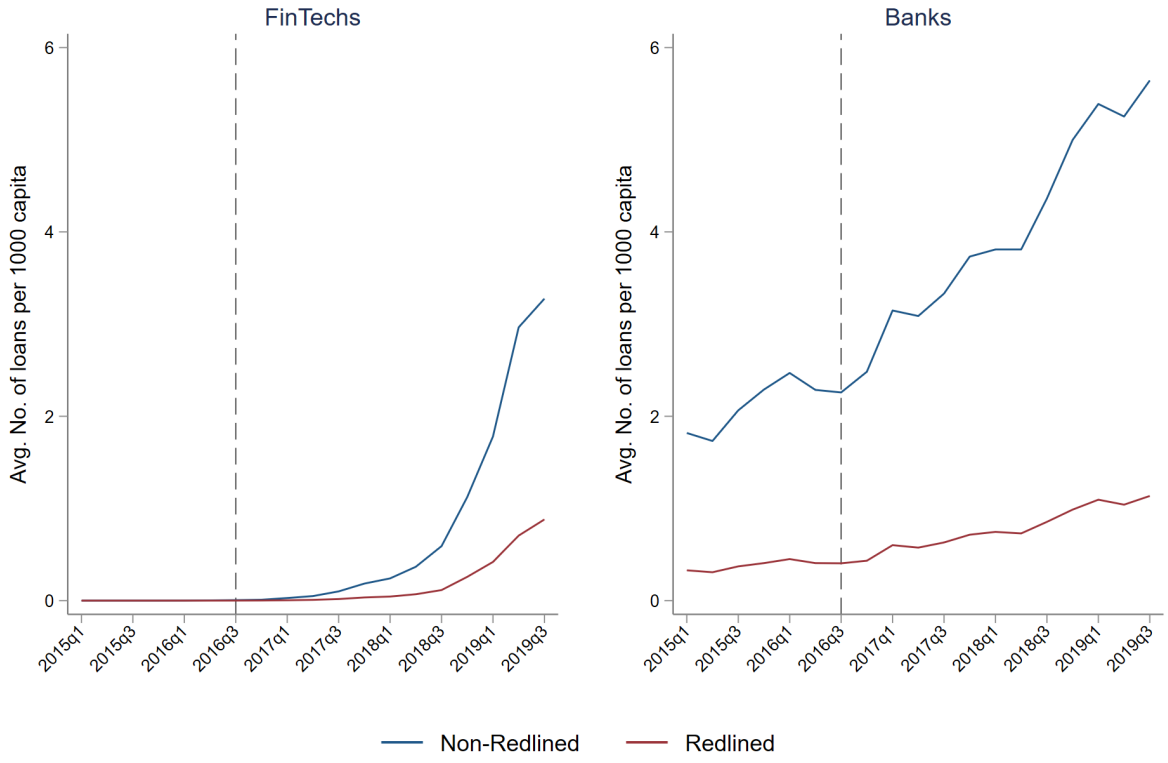
Figure 1
Redlining Measure



(a) Distribution of excess lending **(b) Severity of excess lending**

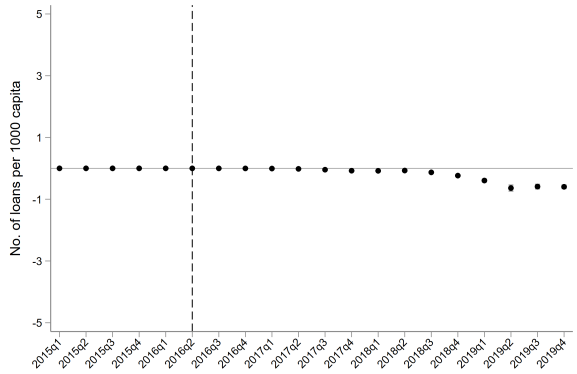
Notes: This figure shows the distribution of the redlining measure, as defined in Equation 1. Panel A shows the distribution of excess lending across pincodes. Panel B shows the severity of excess lending, by dividing it into quartiles. The darkest shade of red represents quartile one, corresponding to the most redlined pincodes, while the lightest shade represents quartile four, corresponding to the least redlined pincodes. Pincodes for which the redlining measure is not computed due to lack of data are shaded gray.

Figure 2
Loans over time, by redlining status

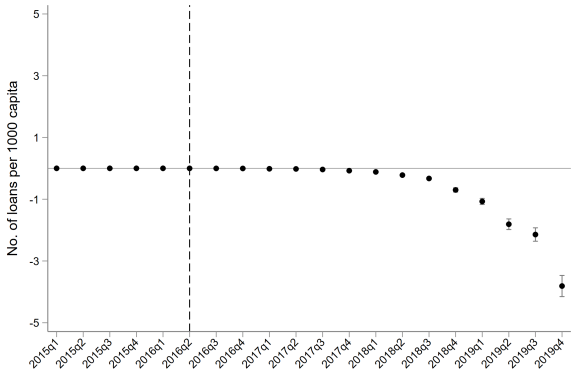


Notes: This figure shows the evolution of FinTech and Bank credit for redlined (red line) and non-redlined (blue line) areas. The vertical axis shows the average number of consumer loans approved per 1000 capita. The data covers the period 2015Q1-2019Q4, and is sourced from TransUnion CIBIL. The vertical dotted line at 2016Q3 marks the introduction of the Unified Payment Interface (UPI).

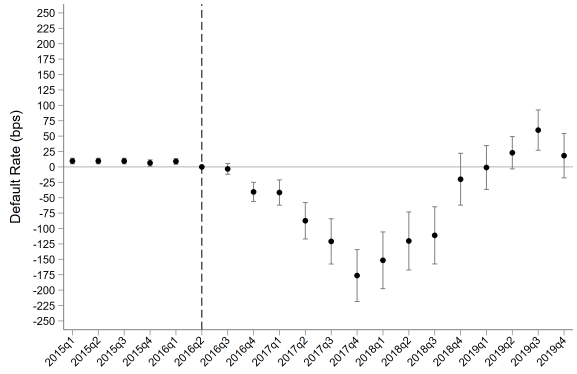
Figure 3
Event Study: Effect on FinTech Lending and Default



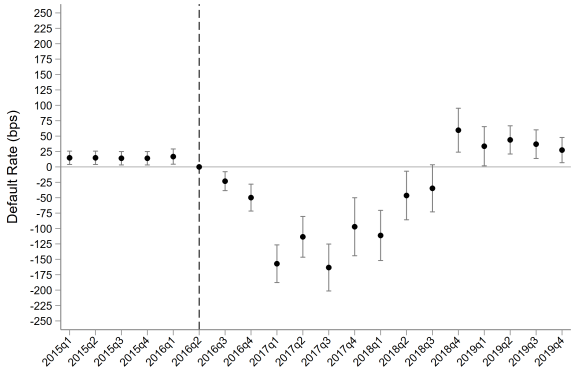
(a) FinTech: NTC Loans



(b) FinTech: Non-NTC Loans



(c) FinTech: NTC Default



(d) FinTech: Non-NTC Default

Notes: This figure presents the results of the dynamic impact of redlining on lending, estimated using the event study specification given in Equation 5 for Fintech new-to-credit (Panel A), non new-to-credit (Panel B) loans, and new-to-credit (Panel C) and non-new-to credit (Panel D) default. The dependent variable is number of consumer loans per 1000 capita and default rate. The treatment variable is the redlining indicator defined in Equation 2. Observations are at pincode $\times$ neighbourhood id-quarter level, and cover the period 2015Q1-2019Q4. Each black dot shows the point estimate for the relevant year-quarter, and the light grey capped lines show the 95 % confidence interval. Pincode and neighbourhood $\times$ quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood level. The dashed vertical line represents the pre-treatment quarter (2016Q2).

Table 1
Summary Statistics: Consumer Loans

Panel A: Full Sample				
	Mean	Median	St. Dev	N
New to credit	2.90	0.74	22.83	301240
Prime	7.15	1.06	55.95	301240
Subprime	0.72	0.12	5.54	301240
Total	13.06	2.35	98.85	301240
New to credit	0.03	0.02	0.05	301360
Prime	0.02	0.01	0.04	301360
Subprime	0.06	0.00	0.12	301360
Total	0.03	0.02	0.03	301360

Panel B: Redlined Sample				
	Mean	Median	St. Dev	N
New to credit	0.75	0.36	1.80	88780
Prime	1.23	0.49	2.89	88780
Subprime	0.16	0.06	0.35	88780
Total	2.63	1.12	5.52	88780
New to credit	0.03	0.00	0.06	88780
Prime	0.02	0.00	0.05	88780
Subprime	0.06	0.00	0.14	88780
Total	0.03	0.03	0.04	88780

Panel C: Non-Redlined Sample				
	Mean	Median	St. Dev	N
New to credit	3.80	1.02	27.10	212460
Prime	9.62	1.55	66.44	212460
Subprime	0.96	0.17	6.58	212460
Total	17.41	3.34	117.38	212460
New to credit	0.03	0.02	0.05	212580
Prime	0.02	0.01	0.03	212580
Subprime	0.06	0.00	0.11	212580
Total	0.03	0.02	0.03	212580

Notes: This table presents summary statistics for the full, redlined, and non-redlined samples of consumer loans, sourced from TransUnion CIBIL. The data are at the pincode-quarter level and cover the period 2015Q1–2019Q4.

Table 2
Cross-Sectional Differences in Mortgage Credit Access by Group

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	No. of loans per 1000 capita					
Group	General	SC	ST	OBC	Muslim	Others
Redlined	-2.6902***	-0.0973***	-0.0321***	-0.1533***	-0.1320***	-0.1333***
	(0.0828)	(0.0035)	(0.0012)	(0.0040)	(0.0297)	(0.0087)
R^2	0.522	0.526	0.610	0.533	0.501	0.528
Pincode FE	Y	Y	Y	Y	Y	Y
Neighbourhood FE	Y	Y	Y	Y	Y	Y
Control mean	2.645	0.099	0.036	0.157	0.123	0.132
t-test vs General	.	31.275	32.086	30.590	29.074	30.700
p-value vs General	.	0.000	0.000	0.000	0.000	0.000
N	51102	51102	51102	51102	51102	51102

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that mortgage credit access is lower in redlined neighbourhoods across social groups. The estimates are from a cross-sectional spatial specification comparing redlined pincodes to neighbouring non-redlined pincodes for General (Col 1), SC (Col 2), ST (Col 3), OBC (Col 4), Muslim (Col 5), and Other (Col 6) borrowers. The data are at the pincode $\times$ neighbourhood-id level and cover the period 2015Q1–2015Q4. The dependent variable is the number of mortgage loans per 1000 capita. The treatment variable is the redlining indicator defined in Equation 2. Pincode and neighbourhood fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level. The t-tests of equality of coefficients for each social group against the General category are reported, along with the corresponding p-values.

Table 3
Impact of digitalisation on FinTech consumer lending in redlined pincodes

	(1)	(2)	(3)	(4)
Dependent variable	No. of loans per 1000 capita			
Score Band	Total	Subprime	NTC	Prime
Redlined × Post	-0.946***	-0.092***	-0.206***	-0.322***
	(0.048)	(0.005)	(0.013)	(0.015)
R^2	0.928	0.926	0.923	0.932
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	1.146	0.115	0.249	0.389
N	1073040	1073040	1073040	1073040

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that FinTech lending to borrowers, across credit-score bands, is lower in redlined pincodes. The estimates are from a spatial difference-in-differences specification to test the impact of redlining on FinTech credit for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data is at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the number of consumer loans per 1000 capita. The treatment variable is the redlining indicator defined in Equation 2. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

Table 4
Heterogeneous Impact on Lending by Social Groups: Triple Differences

	(1)	(2)	(3)	(4)
Dependent variable	No. of loans per 1000 capita			
Score Band	Total	Subprime	NTC	Prime
Redlined × SC Share × Post	1.9772***	0.1847***	0.4340***	0.6733***
	(0.3635)	(0.0345)	(0.0992)	(0.1103)
Redlined × ST Share × Post	2.0461***	0.2060***	0.4172***	0.7217***
	(0.2504)	(0.0236)	(0.0673)	(0.0767)
Redlined × OBC Share × Post	1.6896***	0.1684***	0.3407***	0.6059***
	(0.2577)	(0.0216)	(0.0728)	(0.0834)
Redlined × Muslim Share × Post	1.2335***	0.1400***	0.2497**	0.3849**
	(0.3953)	(0.0275)	(0.1057)	(0.1580)
Redlined × Others × Post	0.4482	0.0471	0.1393*	0.1037
	(0.3411)	(0.0326)	(0.0815)	(0.1142)
SC Share × Post	-1.6360***	-0.1523***	-0.3450***	-0.5749***
	(0.2270)	(0.0211)	(0.0641)	(0.0678)
ST Share × Post	-1.5879***	-0.1593***	-0.3174***	-0.5686***
	(0.1650)	(0.0153)	(0.0441)	(0.0510)
OBC Share × Post	-1.1654***	-0.1161***	-0.2249***	-0.4326***
	(0.1868)	(0.0164)	(0.0516)	(0.0593)
Muslim Share × Post	-0.6905***	-0.0803***	-0.1264*	-0.2310**
	(0.2652)	(0.0186)	(0.0699)	(0.1054)
Others × Post	-0.4020*	-0.0518**	-0.1016*	-0.1087
	(0.2345)	(0.0216)	(0.0570)	(0.0792)
Redlined × Post	-1.4372***	-0.1417***	-0.3063***	-0.4924***
	(0.1121)	(0.0098)	(0.0324)	(0.0336)
R^2	0.9445	0.9515	0.9412	0.9490
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	1.2226	0.1229	0.2633	0.4169
N	866220	866220	866220	866220

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that the negative impact of redlining on FinTech lending is attenuated in pincodes with higher shares of minority-group borrowers. The estimates are from a spatial triple-differences specification that tests the impact of redlining on FinTech credit by social group for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the number of consumer loans per 1000 capita. SC, ST, OBC, Muslim, General, and Others category shares are calculated as the number of approved mortgage loans from 2015Q1–2015Q4 for the respective category in a pincode divided by total mortgage loans in that pincode over the same period. The triple-interaction coefficients are interpreted relative to the omitted General category. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

Table 5
Impact of digitalisation on FinTech consumer default in redlined pincodes

	(1)	(2)	(3)	(4)
Dependent variable	Default Rate (bps)			
Score Band	Total	Subprime	NTC	Prime
Redlined × Post	-54.570***	-93.064***	-62.558***	-44.705***
	(4.799)	(5.392)	(5.003)	(3.868)
R^2	0.703	0.720	0.705	0.692
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	341.779	325.611	325.625	194.644
N	1073280	1073280	1073280	1073280

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that FinTech default rates, across credit-score bands, are lower in redlined pincodes. The estimates are from a spatial difference-in-differences specification to test the impact of redlining on FinTech default rates for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the default rate, measured in basis points, calculated as the number of defaults divided by the number of loans in a pincode-quarter bucket. The treatment variable is the redlining indicator defined in Equation 2. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

Table 6
Heterogeneous Impact on Default by Social Groups: Triple Differences

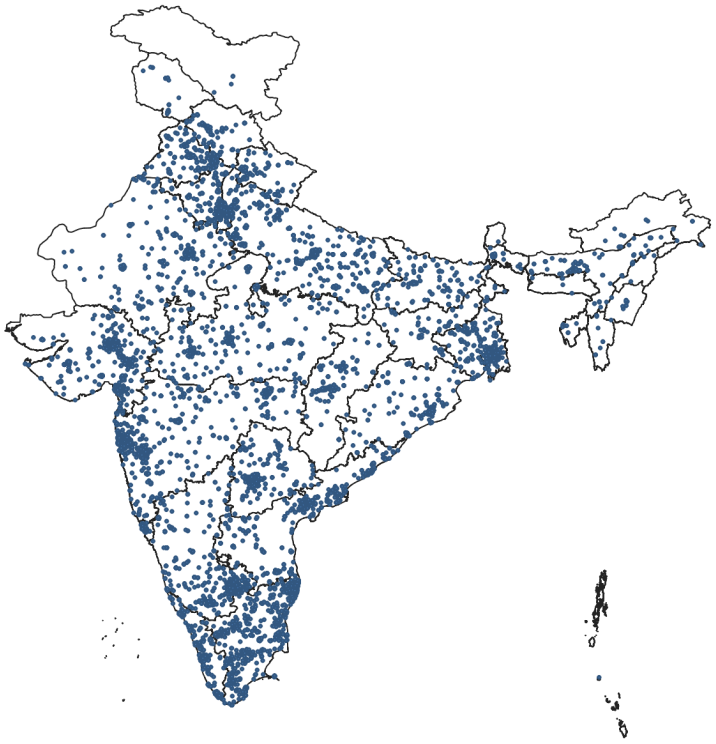
Dependent variable	(1)	(2)	(3)	(4)
	Default Rate (bps)			
Score Band	Total	Subprime	NTC	Prime
Redlined × SC Share × Post	93.0549*	199.2981***	13.0691	87.0547**
	(48.8574)	(55.5015)	(51.7531)	(39.7214)
Redlined × ST Share × Post	118.6981**	143.9410**	13.8502	149.6851**
	(58.1018)	(70.4476)	(57.3743)	(63.9336)
Redlined × OBC Share × Post	70.8233*	161.3778***	77.3978**	15.9721
	(37.2407)	(39.6042)	(38.2327)	(27.5983)
Redlined × Muslim Share × Post	87.4596**	62.5037	69.5832	47.0556
	(42.5228)	(54.7647)	(45.5671)	(39.6246)
Redlined × Others × Post	49.6895	76.3542	15.5038	56.7624
	(49.4379)	(50.7594)	(52.9968)	(39.1944)
SC Share × Post	-31.1366	-134.2243***	-39.8097	-53.9613*
	(38.0521)	(41.4899)	(42.1282)	(29.3861)
ST Share × Post	-93.1532**	-84.6342	-71.2529	-57.8975
	(44.7498)	(60.3739)	(50.4054)	(40.2048)
OBC Share × Post	9.5806	-76.7252**	7.5665	-15.5848
	(30.0902)	(32.8788)	(29.7407)	(21.6188)
Muslim Share × Post	-43.5168	7.6105	-42.9860	-23.7177
	(37.4445)	(41.0331)	(39.1140)	(29.9032)
Others × Post	-15.3534	-63.1033	-18.8512	-13.4279
	(44.7646)	(44.6581)	(46.9810)	(34.3063)
Redlined × Post	-76.5658***	-121.4974***	-74.2151***	-59.2838***
	(7.9189)	(8.8946)	(8.2095)	(6.4499)
R^2	0.7789	0.7879	0.7794	0.7685
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	350.8371	340.4697	334.3484	201.0901
N	866360	866360	866360	866360

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that redlined pincodes with higher minority-group shares experience relatively higher post-UPI default rates. The estimates are from a spatial triple-differences specification that tests the impact of redlining on FinTech default rates by social group for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the default rate, measured in basis points. SC, ST, OBC, Muslim, General, and Others category shares are calculated as the number of approved mortgage loans from 2015Q1–2015Q4 for the respective category in a pincode divided by total mortgage loans in that pincode over the same period. The triple-interaction coefficients are interpreted relative to the omitted General category. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

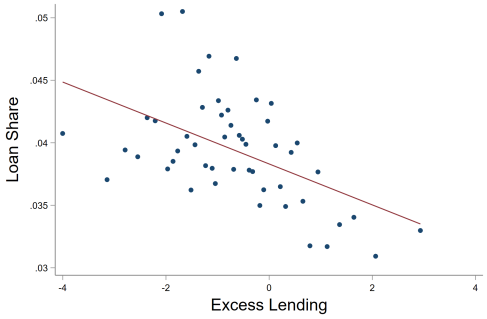
A Supplementary Figures

Figure A.1
Distribution of mortgages

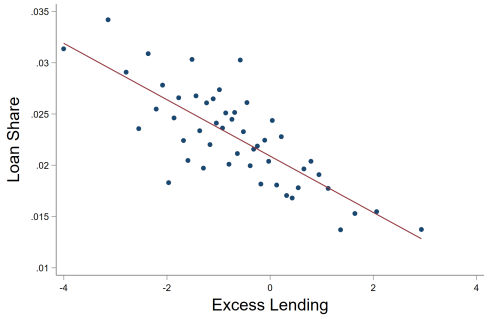


Notes: This figure shows the distribution of 108,428 geocoded mortgages across India over the period 2010-2020. Each blue dot in the mortgage data is a subsample of loans by a large public sector bank.

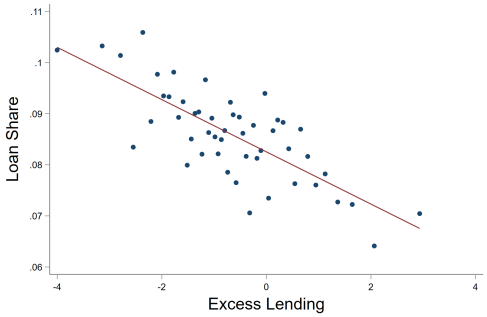
Figure A.2
Identity Correlates of Redlining



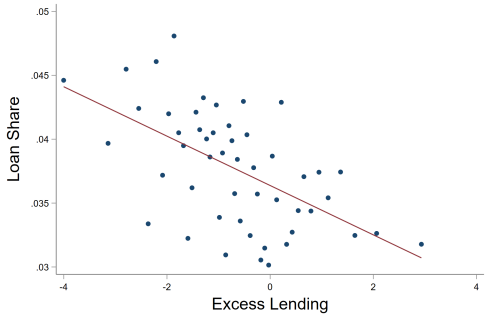
(a) SC



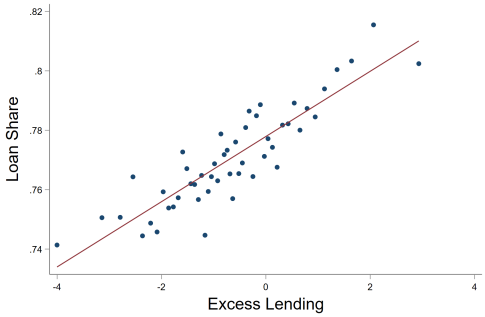
(b) ST



(c) OBC



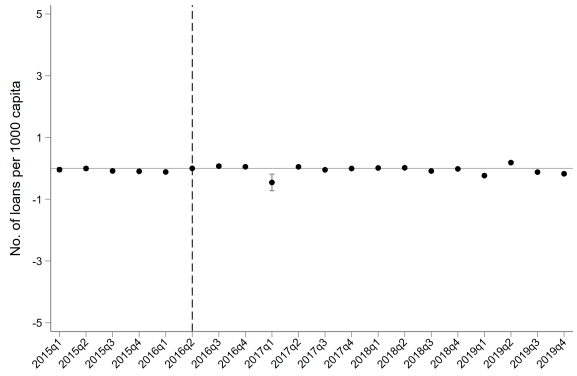
(d) Muslim



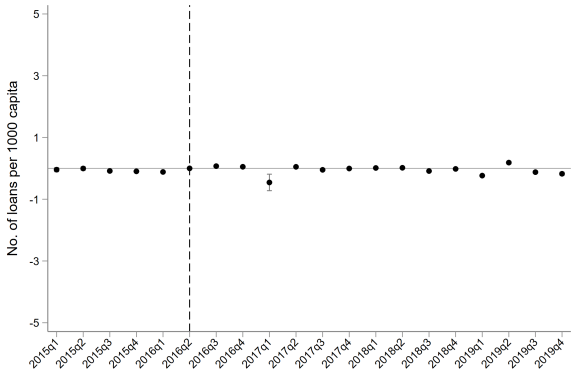
(e) General

Notes: This figure shows the relationship between lending gap and group-wise mortgage loan share, for SC (Panel A), ST (Panel B), OBC (Panel C), Muslim (Panel D), and General (Panel E) category borrowers respectively. The horizontal axis shows the log of the lending gap, as defined in Equation 2. The vertical axis shows each group's loan share, calculated as the number of loans approved for the group in a pincode divided by total loans in that pincode. Each dot represents roughly 300 pincodes. District fixed effects are included. The loan data is aggregated over the period 2016Q3–2019Q4.

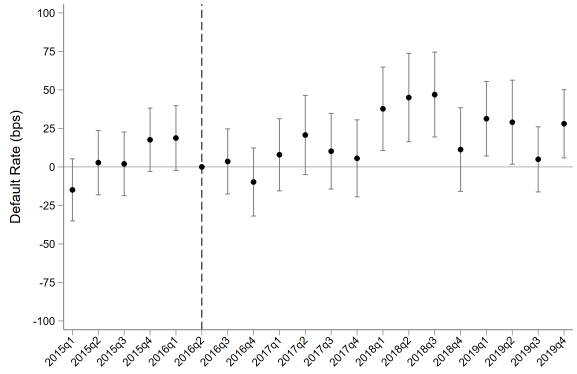
Figure A.3
Event Study: Effect on Bank Lending and Default



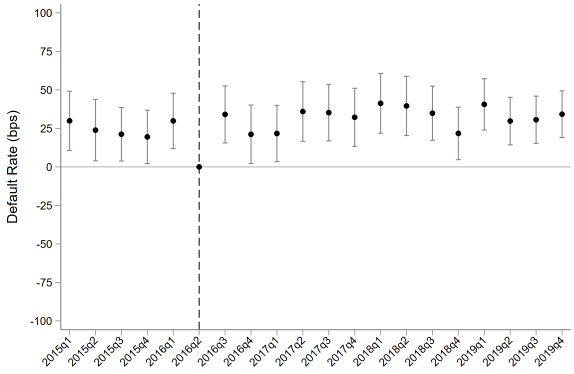
(a) Banks: NTC Lending



(b) Banks: Non-NTC Lending



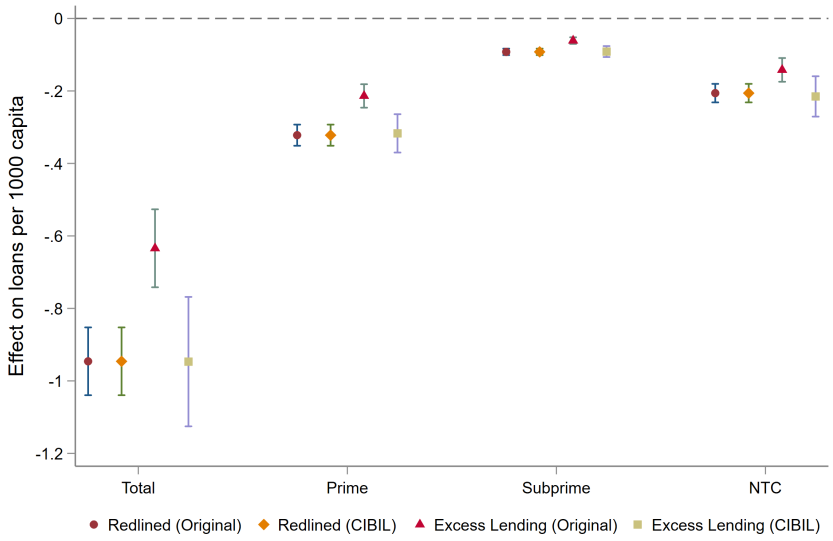
(c) Banks: NTC Default



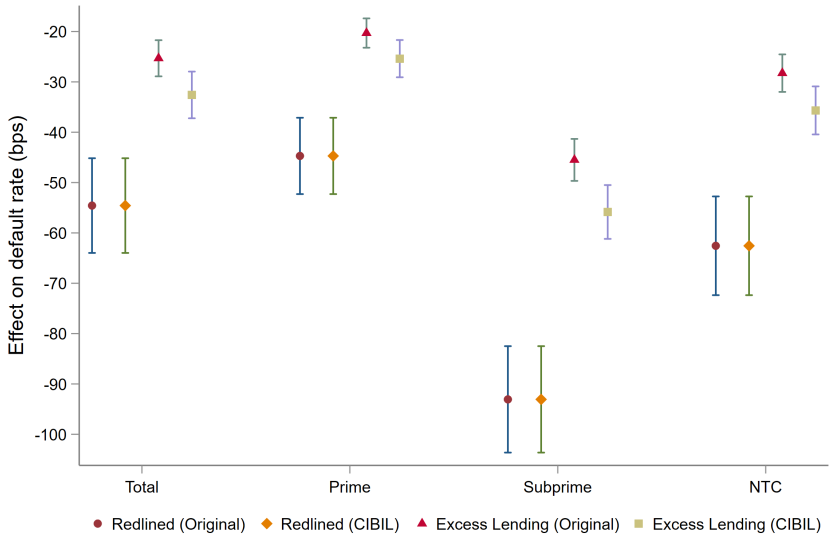
(d) Banks: Non-NTC Default

Notes: This figure presents the results of the dynamic impact of redlining on lending, estimated using the event study specification given in Equation 5 for Bank new-to-credit (Panel A), non new-to-credit (Panel B) loans, and new-to-credit (Panel C) and non-new-to credit (Panel D) default . The dependent variable is number of consumer loans per 1000 capita and default rate. The treatment variable is the redlining indicator defined in Equation 2. Observations are at pincode $\times$ neighbourhood id-quarter level, and cover the period 2015Q1-2019Q4. Each black dot shows the point estimate for the relevant year-quarter, and the light grey capped lines show the 95 % confidence interval. Pincode and neighbourhood $\times$ quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood level. The dashed vertical line represents the pre-treatment quarter (2016Q2).

Figure A.4
Robustness to Alternative Redlining Measures



(a) Consumer loans



(b) Default rates

Notes: This figure shows robustness of the spatial difference-in-differences estimates to alternative measures of redlining. Panel A reports effects on FinTech consumer loans per 1000 capita. Panel B reports effects on FinTech default rates, measured in basis points. The four redlining measures are: the baseline redlining indicator, the CIBIL redlining indicator, the baseline excess lending measure, and the CIBIL excess lending. The continuous measures are multiplied by -1, so that higher values correspond to greater redlining. Each coefficient is estimated from a separate regression using Equation 4. The data cover the period 2015Q1–2019Q4. All specifications include pincode and neighbourhood $\times$ quarter fixed effects. Standard errors are two-way clustered at the pincode and neighbourhood-id level. Bars show 95 percent confidence intervals.

B Supplementary Tables

Table B.1
Glossary

Variable	Definition
Excess Lending	Difference between log loans per capita in a pincode and log loans per capita in a district.
Banks	Scheduled Commercial Banks that includes private and public sector banks.
FinTechs	CIBIL classification based on their operational structure.
Prime	Credit score category combining Prime (731-770), Prime Plus (771-790) and Super Prime (790-900) categories, as per CIBIL
Subprime	Credit score category, for score range 300–680, as per CIBIL.
New-to-Credit	Credit score category for first-time borrowers with no credit score, as per CIBIL

Notes: This table defines variables and common terms used in the paper.

Table B.2
Summary Statistics: Mortgage Loans

Panel A: Full Sample				
	Mean	Median	St. Dev	N
Loan Amount (Million INR)	1.43	0.90	11.86	3852218
Loan-to-value Ratio	0.58	0.62	0.22	2673280
Interest Rate (per cent)	6.43	7.40	2.98	3911562
Property Value (Million INR)	4.61	3.47	23.65	989155
Project Tenure (Months)	208.44	214.00	84.77	3713382
Project Cost (Million INR)	11.21	0.00	11849.26	3907687
Collateral (Million INR)	0.08	0.00	23.35	2875558
Carpet Area (sqft.)	1239.40	722.00	63865.05	518350
General	0.88	1.00	0.32	3693272
OBC	0.07	0.00	0.25	3693272
SC	0.03	0.00	0.18	3693272
ST	0.02	0.00	0.12	3693272
Hindu	0.91	1.00	0.28	3951476
Muslim	0.03	0.00	0.17	3951476
Panel B: Redlined Sample				
	Mean	Median	St. Dev	N
Loan Amount (Million INR)	1.13	0.80	1.40	138604
Loan-to-value Ratio	0.60	0.63	0.22	99362
Interest Rate (per cent)	6.62	7.40	2.78	141277
Property Value (Million INR)	3.45	2.80	3.37	44646
Project Tenure (Months)	211.90	228.00	87.01	137972
Project Cost (Million INR)	1.38	0.00	3.99	140773
Collateral (Million INR)	0.06	0.00	0.55	109204
Carpet Area (sqft.)	906.83	760.00	1400.47	24358
General	0.84	1.00	0.37	132901
OBC	0.09	0.00	0.29	132901
SC	0.04	0.00	0.21	132901
ST	0.03	0.00	0.17	132901
Hindu	0.90	1.00	0.30	141775
Muslim	0.04	0.00	0.19	141775
Panel C: Non-Redlined Sample				
	Mean	Median	St. Dev	N
Loan Amount (Million INR)	1.44	0.90	12.08	3713614
Loan-to-value Ratio	0.58	0.62	0.22	2573918
Interest Rate (per cent)	6.42	7.40	2.99	3770285
Property Value (Million INR)	4.66	3.50	24.19	944509
Project Tenure (Months)	208.31	213.00	84.68	3575410
Project Cost (Million INR)	11.58	0.00	12068.63	3766914
Collateral (Million INR)	0.08	0.00	23.80	2766354
Carpet Area (sqft.)	1255.80	720.00	65419.87	493992
General	0.89	1.00	0.32	3560371
OBC	0.07	0.00	0.25	3560371
SC	0.03	0.00	0.18	3560371
ST	0.02	0.00	0.12	3560371
Hindu	0.91	1.00	0.28	3809701
Muslim	0.03	0.00	0.16	3809701

Notes: This table presents summary statistics for the full, redlined and non-redlined sample of mortgage loans, sourced from a large public sector bank. The data is at the loan-level covers the period 2010Q2-2019Q4.

Table B.3
Cross Sectional Estimates of Impact on Mortgage Loans

	(1)	(2)	(3)	(4)
Dependent variable	No. of loans per 1000 capita			
Score Band	Total	Subprime	NTC	Prime
Redlined	-2.290***	-0.050***	-1.732***	-0.397***
	(0.111)	(0.003)	(0.085)	(0.020)
R^2	0.907	0.898	0.906	0.912
Neighbour FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
Control mean	3.427	0.077	2.619	0.570
N	53414	53414	53414	53414

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that redlined pincodes get lower mortgage loans, across credit score bands. These estimates are from OLS regressions of all (Col 1), subprime (Col 2), new to credit (Col 3) and Prime (Col 4) mortgage loans. The data is at pincode $\times$ neighbourhood id level, and covers the year 2015. The dependent variable is number of consumer loans per 1000 capita. The treatment variable is the redlining indicator defined in Equation 2. District and neighbourhood fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood level.

Table B.4
Impact of digitalisation on bank consumer lending in redlined pincodes

	(1)	(2)	(3)	(4)
Dependent variable	No. of loans per 1000 capita			
Score Band	Total	Subprime	NTC	Prime
Redlined × Post	-1.630***	-0.095***	0.003	-1.245***
	(0.086)	(0.013)	(0.017)	(0.066)
R^2	0.979	0.953	0.971	0.975
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	5.525	0.327	1.273	3.093
N	1073040	1073040	1073040	1073040

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that bank lending to borrowers, across credit-score bands, is lower in redlined pincodes. The estimates are from a spatial difference-in-differences specification to test the impact of redlining on bank credit for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the number of consumer loans per 1000 capita. The treatment variable is the redlining indicator defined in Equation 2. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

Table B.5
Heterogeneous Impact on Bank Lending by Social Groups: Triple Differences

	(1)	(2)	(3)	(4)
Dependent variable	No. of loans per 1000 capita			
Score Band	Total	Subprime	NTC	Prime
Redlined × SC Share × Post	3.0246***	0.1491***	-0.1005	2.4149***
	(0.5479)	(0.0305)	(0.1068)	(0.4410)
Redlined × ST Share × Post	3.2069***	0.1382***	-0.0350	2.5482***
	(0.4144)	(0.0252)	(0.0810)	(0.3208)
Redlined × OBC Share × Post	3.6520***	0.1285***	0.0705	2.8528***
	(0.4109)	(0.0253)	(0.0650)	(0.3348)
Redlined × Muslim Share × Post	-2.7927	-1.6281	1.6108	-1.5344
	(3.7687)	(1.0258)	(1.0515)	(2.6569)
Redlined × Others × Post	0.6853	0.0526*	0.0161	0.5273
	(0.5125)	(0.0312)	(0.1093)	(0.4172)
SC Share × Post	-2.7225***	-0.1279***	0.0339	-2.1316***
	(0.3369)	(0.0215)	(0.0759)	(0.2732)
ST Share × Post	-2.8233***	-0.1095***	-0.0980*	-2.1619***
	(0.2666)	(0.0218)	(0.0543)	(0.2089)
OBC Share × Post	-2.6168***	-0.0768***	-0.0828	-2.0410***
	(0.3180)	(0.0222)	(0.0555)	(0.2470)
Muslim Share × Post	1.2821	1.0490	-1.0771	0.5574
	(2.4577)	(0.6674)	(0.6816)	(1.7307)
Others × Post	-0.8097**	-0.0447*	-0.1032	-0.5524*
	(0.3937)	(0.0237)	(0.0972)	(0.3211)
Redlined × Post	-2.2796***	-0.0587**	-0.0555	-1.7908***
	(0.1873)	(0.0274)	(0.0407)	(0.1515)
R^2	0.9857	0.9716	0.9823	0.9815
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	5.8418	0.3382	1.3319	3.2916
N	866220	866220	866220	866220

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that the negative impact of redlining on bank lending is attenuated in pincodes with higher shares of minority-group borrowers. The estimates are from a spatial triple-differences specification that tests the impact of redlining on bank credit by social group for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the number of consumer loans per 1000 capita. SC, ST, OBC, Muslim, General, and Others category shares are calculated as the number of approved mortgage loans from 2015Q1–2015Q4 for the respective category in a pincode divided by total mortgage loans in that pincode over the same period. The triple-interaction coefficients are interpreted relative to the omitted General category. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

Table B.6
Impact of digitalisation on bank consumer default in redlined pincodes

	(1)	(2)	(3)	(4)
Dependent variable	Default Rate (bps)			
Score Band	Total	Subprime	NTC	Prime
Redlined × Post	14.570***	11.153	15.088***	15.825***
	(2.702)	(8.365)	(4.250)	(3.371)
R^2	0.695	0.673	0.665	0.656
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	245.087	445.299	257.563	188.034
N	1073280	1073280	1073280	1073280

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that bank default rates, across credit-score bands, are higher in redlined pincodes. The estimates are from a spatial difference-in-differences specification to test the impact of redlining on bank default rates for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the default rate, measured in basis points, calculated as the number of defaults divided by the number of loans in a pincode-quarter bucket. The treatment variable is the redlining indicator defined in Equation 2. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

Table B.7
Heterogeneous Impact on Bank Default by Social Groups: Triple Differences

	(1)	(2)	(3)	(4)
Dependent variable	Default Rate (bps)			
Score Band	Total	Subprime	NTC	Prime
Redlined × SC Share × Post	11.2680	90.7607	37.7184	-33.9908
	(28.5498)	(93.2272)	(42.1018)	(39.5653)
Redlined × ST Share × Post	2.4074	126.6501	-52.7451	-44.8770
	(35.4523)	(118.9637)	(43.4372)	(59.3627)
Redlined × OBC Share × Post	-9.4399	-7.1348	-37.8361	27.6317
	(20.0528)	(66.3741)	(30.3278)	(24.9640)
Redlined × Muslim Share × Post	41.3893	97.5467	-45.2093	43.2371
	(27.2735)	(87.0820)	(42.4214)	(29.6065)
Redlined × Others × Post	-39.6286	-84.7450	-15.6665	-29.7400
	(25.6249)	(100.6220)	(48.0361)	(35.3112)
SC Share × Post	-38.9918*	-27.9336	-59.6380*	-8.2384
	(21.6562)	(73.9854)	(30.5042)	(26.3022)
ST Share × Post	16.0166	-44.5189	79.0174**	32.4941
	(28.0883)	(94.9128)	(35.1171)	(36.0643)
OBC Share × Post	15.8602	16.9772	-19.7199	8.9395
	(16.4067)	(52.2534)	(24.6980)	(21.2303)
Muslim Share × Post	6.4332	68.7200	20.9710	-4.7846
	(19.6433)	(73.2766)	(31.1622)	(24.3513)
Others × Post	39.0787	176.2894**	-7.5644	8.7200
	(24.4537)	(72.1001)	(46.7789)	(29.6890)
Redlined × Post	11.4994***	6.0956	23.3728***	9.3526*
	(4.0707)	(13.2310)	(6.7398)	(5.1157)
R^2	0.7666	0.7452	0.7372	0.7387
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	244.6542	457.4860	259.2531	187.0998
N	866360	866360	866360	866360

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that redlined pincodes with higher minority-group shares do not experience relatively higher post-UPI bank default rates relative to the General category. The estimates are from a spatial triple-differences specification that tests the impact of redlining on bank default rates by social group for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the default rate, measured in basis points. SC, ST, OBC, Muslim, General, and Others category shares are calculated as the number of approved mortgage loans from 2015Q1–2015Q4 for the respective category in a pincode divided by total mortgage loans in that pincode over the same period. The triple-interaction coefficients are interpreted relative to the omitted General category. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are two-way clustered at the pincode and neighbourhood-id level.

Table B.8
Impact of digitalisation on FinTech consumer lending in redlined pincodes (Alternative clustering)

	(1)	(2)	(3)	(4)
Dependent variable	No. of loans per 1000 capita			
Score Band	Total	Subprime	NTC	Prime
Redlined × Post	-1.213***	-0.121***	-0.253***	-0.425***
	(0.069)	(0.007)	(0.016)	(0.024)
R^2	0.653	0.631	0.656	0.652
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	1.146	0.115	0.249	0.389
N	1073040	1073040	1073040	1073040

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that FinTech lending to borrowers, across credit-score bands, is lower in redlined pincodes. The estimates are from a spatial difference-in-differences specification to test the impact of redlining on FinTech credit for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the number of consumer loans per 1000 capita. The treatment variable is the redlining indicator defined in Equation 2. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are clustered at the pincode level.

Table B.9
Heterogeneous Impact on FinTech Lending by Social Groups: Triple Differences (Alternative clustering)

	(1)	(2)	(3)	(4)
Dependent variable	No. of loans per 1000 capita			
Score Band	Total	Subprime	NTC	Prime
Redlined × SC Share × Post	2.4439***	0.2363***	0.5183***	0.8581***
	(0.5326)	(0.0557)	(0.1242)	(0.1779)
Redlined × ST Share × Post	2.8027***	0.2907***	0.5539***	1.0077***
	(0.3324)	(0.0335)	(0.0788)	(0.1115)
Redlined × OBC Share × Post	2.3619***	0.2454***	0.4520***	0.8626***
	(0.3617)	(0.0354)	(0.0877)	(0.1234)
Redlined × Muslim Share × Post	1.4781***	0.1776***	0.2775**	0.4714**
	(0.5498)	(0.0377)	(0.1368)	(0.2363)
Redlined × Others × Post	0.4708	0.0520	0.1340	0.1140
	(0.6920)	(0.0738)	(0.1449)	(0.2442)
SC Share × Post	-2.2520***	-0.2192***	-0.4608***	-0.8067***
	(0.3263)	(0.0339)	(0.0791)	(0.1080)
ST Share × Post	-2.0877***	-0.2156***	-0.4072***	-0.7598***
	(0.2350)	(0.0236)	(0.0558)	(0.0781)
OBC Share × Post	-1.6770***	-0.1736***	-0.3137***	-0.6240***
	(0.2600)	(0.0259)	(0.0626)	(0.0886)
Muslim Share × Post	-0.8250**	-0.1022***	-0.1407	-0.2823*
	(0.3829)	(0.0259)	(0.0937)	(0.1649)
Others × Post	-0.9683**	-0.1132***	-0.2059**	-0.3174**
	(0.3771)	(0.0373)	(0.0803)	(0.1358)
Redlined × Post	-1.8248***	-0.1855***	-0.3735***	-0.6396***
	(0.1470)	(0.0145)	(0.0362)	(0.0489)
R^2	0.7282	0.7410	0.7380	0.7332
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	1.2226	0.1229	0.2633	0.4169
N	866220	866220	866220	866220

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that the negative impact of redlining on FinTech lending is attenuated in pincodes with higher shares of minority-group borrowers. The estimates are from a spatial triple-differences specification that tests the impact of redlining on FinTech credit by social group for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the number of consumer loans per 1000 capita. SC, ST, OBC, Muslim, General, and Others category shares are calculated as the number of approved mortgage loans from 2015Q1–2015Q4 for the respective category in a pincode divided by total mortgage loans in that pincode over the same period. The triple-interaction coefficients are interpreted relative to the omitted General category. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are clustered at the pincode level.

Table B.10

Impact of digitalisation on FinTech consumer default in redlined pincodes (Alternative clustering)

	(1)	(2)	(3)	(4)
Dependent variable	Default Rate (bps)			
Score Band	Total	Subprime	NTC	Prime
Redlined × Post	-56.744***	-98.189***	-66.669***	-46.357***
	(5.684)	(6.559)	(5.904)	(4.575)
R^2	0.579	0.587	0.582	0.559
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	341.779	325.611	325.625	194.644
N	1073280	1073280	1073280	1073280

Standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that FinTech default rates, across credit-score bands, are lower in redlined pincodes. The estimates are from a spatial difference-in-differences specification to test the impact of redlining on FinTech default rates for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the default rate, measured in basis points, calculated as the number of defaults divided by the number of loans in a pincode-quarter bucket. The treatment variable is the redlining indicator defined in Equation 2. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are clustered at the pincode level.

Table B.11
Heterogeneous Impact on FinTech Default by Social Groups: Triple Differences (Alternative clustering)

Dependent variable	(1)	(2)	(3)	(4)
	Default Rate (bps)			
Score Band	Total	Subprime	NTC	Prime
Redlined × SC Share × Post	126.7030**	209.0648***	7.5376	94.5412**
	(60.6307)	(67.0801)	(62.8066)	(47.8926)
Redlined × ST Share × Post	121.0499*	137.3815	23.4805	156.7126**
	(70.2779)	(85.7530)	(72.0579)	(68.8589)
Redlined × OBC Share × Post	77.3567*	179.4454***	70.9355	26.3562
	(46.3205)	(49.3283)	(47.4006)	(34.0978)
Redlined × Muslim Share × Post	79.7363	101.7395	53.7606	57.9891
	(50.3010)	(71.4821)	(52.6519)	(47.1738)
Redlined × Others × Post	76.6182	170.2308**	48.8194	51.7853
	(62.2913)	(68.5549)	(73.4017)	(44.6125)
SC Share × Post	-57.4015	-149.9213***	-40.7807	-60.5950*
	(47.0805)	(49.3625)	(51.8619)	(34.9826)
ST Share × Post	-93.1386	-80.0316	-75.9937	-59.3434
	(57.5873)	(76.0279)	(61.9008)	(46.4195)
OBC Share × Post	4.3740	-84.9816**	12.9161	-27.0652
	(36.6120)	(40.5565)	(36.7257)	(26.8304)
Muslim Share × Post	-27.3606	3.4871	-23.6882	-17.8873
	(44.5582)	(50.6456)	(45.7992)	(36.0656)
Others × Post	-27.1367	-81.2872	-13.9432	-22.2964
	(56.7061)	(53.7245)	(59.6916)	(41.6168)
Redlined × Post	-78.2802***	-128.6068***	-74.2221***	-59.4120***
	(10.5279)	(12.2198)	(10.5781)	(8.7754)
R^2	0.6628	0.6652	0.6658	0.6449
Pincode FE	Y	Y	Y	Y
Neighbour-time FE	Y	Y	Y	Y
Control mean	350.8371	340.4697	334.3484	201.0901
N	866360	866360	866360	866360

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Notes: This table shows that redlined pincodes with higher minority-group shares experience relatively higher post-UPI FinTech default rates. The estimates are from a spatial triple-differences specification that tests the impact of redlining on FinTech default rates by social group for all borrowers (Col 1), subprime borrowers (Col 2), new-to-credit borrowers (Col 3), and prime borrowers (Col 4). The data are at the pincode × neighbourhood-id-quarter level and cover the period 2015Q1–2019Q4. The dependent variable is the default rate, measured in basis points. SC, ST, OBC, Muslim, General, and Others category shares are calculated as the number of approved mortgage loans from 2015Q1–2015Q4 for the respective category in a pincode divided by total mortgage loans in that pincode over the same period. The triple-interaction coefficients are interpreted relative to the omitted General category. Post is an indicator taking value 1 from 2016Q3 onwards. Pincode and neighbourhood × quarter fixed effects are included. Standard errors are clustered at the pincode level.