WORKING PAPERS

N° 18

2024

"Beyond the Bureau: Interoperable Payment Data for Loan Screening and Monitoring"

Kumar Rishabh

FIT IN Initiative

Beyond the Bureau: Interoperable Payment Data for Loan Screening and Monitoring*

Kumar Rishabh†

Abstract

I investigate the informational value of interoperable payment data in lending, a key component of worldwide open banking initiatives. I utilize a unique dataset that links borrowers' electronic payment histories with *both* traditional bank loans and fintech loans issued to the same set of Indian small businesses. I find that interoperable payment history complements credit bureau data in predicting loan delinquency. Its value to lenders stems mainly from new signals of credit risk. Back-of-the-envelope calculations suggest payment history leads to a significant increase in the lender's return on investment. Post-disbursal, payment data refine delinquency forecasts by providing early warning signals for loan monitoring. Transitioning to payment-history-based underwriting has distributional consequences: while it benefits the majority, it puts those with low credit scores and limited payment histories at a disadvantage. In fintech lending with sales-linked loans, payment history emerges as a substitute for credit bureau data, albeit with pronounced moral hazard challenges.

JEL Classification: G20, G21, G23

Keywords: Payment, Machine learning, Prediction, Interoperability, Open banking

*I thank Gustavo Cortes (discussant), Hans Degryse, Quirin Fleckenstein, Thierry Foucault, Andreas Fuster, Zhiguo He (discussant), Jing Huang (discussant), Artashes Karapetyan (discussant), Yvan Lengwiler, David Martinez-Miera (discussant), Sophie Moinas (discussant), Evren Örs, Christophe Pérignon, Tarun Ramadorai, Larissa Schäfer (discussant), Amit Seru (discussant), Neharika Sobti (discussant), Boris Vallée (discussant), the seminar audience at HEC Paris, HEC Lausanne, and the participants at the 2023 AFA Annual Meeting (New Orleans), NBER Capital Markets, Technology, Financial Inclusion, and Economic Growth Conference (Hyderabad), Economics of Payments Conference XI (Ottawa), FIT IN Initiative Conference (Toulouse), University of Florida Conference on ML in Finance, China International Conference in Finance, 29th Finance Forum (Santiago de Compostela), EFI Research Network Workshop (Brussels), NSE-NYU Conference on Indian Financial Markets 2023, CMI-XKDR Forum Field Workshop on Firm Finance 2022, Workshop on Big and Smart data (Freiburg), Economics of Financial Technology Conference (Edinburgh), and Asian Economic Development Conference for their valuable comments and suggestions. I also thank Damien Krämer, Jorma Schäublin, and Alessandro Di Stefano for their help with the analysis. I acknowledge the generous funding support by the FIT IN Initiative of the Toulouse School of Economics, and NSE-NYU Stern Initiative on the Study of Indian Capital Markets. All errors are mine. This version: April 2024.

†HEC Lausanne (University of Lausanne) and University of Basel, Switzerland.
Email: `kumar.rishabh@unil.ch`

1 Introduction

The proliferation of digital payments has generated vast amounts of transactional data. However, this data often remains siloed within individual payment systems, limiting its use. As a result, there are growing calls for more open financial systems that allow payments data to be interoperable, i.e., sharable across institutions. Open Banking initiatives, which aim to achieve this goal, have quickly gained prominence across many economies, with 80 countries launching such efforts by October 2021 (Babina et al., 2024). One of the key objectives of these initiatives is to redefine risk assessment in lending by enabling financial institutions to utilize interoperable payment data for underwriting, thereby driving innovation in credit products (Experian, 2022). However, the effectiveness of this approach relies on the degree to which payment history from one source can meaningfully improve another institution's capacity to gauge a borrower's creditworthiness. This paper analyzes the value of such interoperable payment histories for lenders.

Specifically, in the context of small business lending, I ask the following questions. First, are interoperable payment histories effective in both pre-loan screening and post-disbursal monitoring? Second, how do they interact with credit bureau data—as complements or substitutes? Third, is there a trade-off between the granularity of shared data and privacy and technological costs? Fourth, are the gains to lenders from utilizing payment histories consistent across diverse borrower types, such as small versus large enterprises or firms with low versus high credit scores? Finally, which borrower types benefit and which lose when lenders employ payment history-based loan evaluation?

Addressing these questions is challenging due to the scarcity of situations where lending and payment data come from separate sources—a setup essential for understanding data interoperability.[1] The complexity is further compounded by what I call the BFP critique, based on concerns raised by Berg, Fuster and Puri (2022) regarding studies assessing the effectiveness of *alternative data* in lending. The challenge arises because alternative data is mainly used by fintechs and bigtechs, not traditional lenders. The BFP critique suggests that these lender types often serve distinct borrower bases, implying that conclusions about the utility of alternative data in one lending context may not apply broadly. Moreover, even when borrower samples are harmonized, differences in *lending contracts* between traditional lenders and alternative-data-reliant lenders can confound the estimated value of alternative data. To accurately assess the value of alternative data in traditional lending, it must be studied within the standard debt contracts issued by traditional banks.

My research design primarily addresses the interoperability challenge and tackles the BFP critique. This is accomplished by linking traditional bank lending contracts, specifically those for loans to small businesses in India, with payment history data from a collaborating fintech

[1] An exception is the study by Ghosh, Vallee and Zeng (2024), which, though relevant, differs in key aspects from my research; these differences will be detailed later.

company. In this setup, the loan and payment data originate from two separate entities, enabling an effective evaluation within the interoperability context. Importantly, this linkage between bank loans and payment flows allows me to assess the value of alternative data (payment history) in the screening and monitoring of *traditional* bank loans.[2]

To address our questions on the value of interoperable data, I assess the predictive power of different sets of variables, or *models*, in forecasting loan delinquency. These models are designed based on data available before loan disbursal for screening purposes, and for monitoring, they incorporate post-disbursal payment data into the most extensive screening model. The models differ in their information source and content. The three traditional screening benchmark models capture the credit bureau information, traditional hard information, and traditional hard and soft information, respectively, with each expanding on the previous one. The two payment-history (PH) models contain payment data at different granularity: PH aggregate captures broad information such as total transaction value (sales) and growth among others, while PH granular expands the aggregate model by incorporating detailed transaction-level data alongside comparisons with district-level payment averages.

By studying the predictive performance of the traditional benchmark models, PH models, and their combinations, we can answer our questions. For example, to determine whether aggregate payment data brings any new signals compared to those captured by the traditional lender, we compare the predictive performance of the composite model combining PH aggregate and traditional hard & soft information against the benchmark traditional hard & soft information. Similarly, to assess whether credit bureau and PH data are complements or substitutes, we compare the predictive performance of the composite model PH aggregate and credit bureau against each of its constituents. If the combined model significantly improves performance over its parts, it would imply that the two sources of information are complements. Section 3.1 describes in detail how each of our research questions can be answered using different models.

To predict loan delinquency, I employ the Random Forest machine learning algorithm. The predictive performance of various models is evaluated out-of-sample by plotting Receiver Operating Characteristics (ROC) curves and calculating the Area Under the ROC Curves (AUC).

[2]One might argue that payment histories cannot be considered *alternative data* in traditional lending, given that banks have historically possessed transaction data. However, despite collecting extensive payment data, banks do not appear to use it systematically for commercial purposes, such as loan underwriting. A survey of senior executives in 168 large banks across Europe, North America, and Asia by the financial advisory firm Celent found that only about 10% fully utilize payments data such purposes, often hindered by siloed data structures, regulatory constraints, and IT challenges. (Hines, 2021). In my Indian SME lending context, this is even less of a concern, as traditional loans were made by both banks and non-depository lenders called non-bank financial companies (NBFCs). While I do not observe the lender identity to separate out the two traditional lender types, NBFCs are active players in the small business loan segments and likely form an important part of my sample. Unlike banks, these NBFCs do not possess payment data. Moreover, Mishra, Prabhala and Rajan (2022) have shown that Indian banks are slow in adopting well-established lending technologies like credit scores despite easy access, due to organizational reasons. Given the challenges raised in the survey of bank executives, I expect this to be the case among Indian banks with payments data as well.

An AUC of one signifies perfect predictive performance, while an AUC of 0.5 implies a predictive accuracy no better than random guessing. As a complementary measure, I also calculate out-of-sample Average Precision (AP), which reflects the likelihood that a delinquent loan is correctly identified as such across various decision thresholds.

Our results relating to the screening value of interoperable data are as follows. Payment history and credit bureau data have equal predictive accuracy for loan delinquency, and when combined, they show significant synergy, suggesting that they capture distinct and complementary aspects of credit risk. Adding payment history to the most comprehensive traditional benchmark model improves prediction, highlighting its broad informativeness. The value of payment data is similar to that of the lender's soft information, with the payment history contributing majorly through unique, new signals, while a minor value also arises due to payment history making soft information more quantifiable. My back-of-the-envelope estimates indicate that using payment history for screening can lead to a significant increase in the lender's return on investment (ROI). Choosing the top 20% of loans identified by the PH model yields a 5% higher ROI compared to selecting the top 20% based only on traditional hard information.

Interoperable payment data holds significant, yet often overlooked, potential for loan monitoring. Unlike traditional credit scores that update with delays and depend on other lenders' reporting, payment history data provides real-time signals of a borrower's financial health, making it a more immediate and accurate early warning indicator. I explore how interoperable payment data could strengthen monitoring by adding post-disbursal payment variables to the extensive pre-disbursal screening model and updating predictions at thirty-day intervals post-loan issuance. Payment data significantly improves monitoring capabilities, with post-disbursal payment information contributing as much to AUC within 120 days (and to AP within 90 days) as pre-disbursal PH variables do in the screening phase. The dynamic real-time risk assessment capability of payment data is further demonstrated by the adjustment of delinquency probabilities—increasing for loans that finally default and decreasing for those that remain performing—as more payment data is integrated over time. Specifically, the probability of delinquency for loans destined to default increases by approximately 5 percentage points within 180 days after disbursal.

From a lender's perspective, payment history proves more effective for smaller firms and aids in evaluating both high-bureau-score and low-bureau-score borrowers. While the boost for thin-file borrowers (borrowers without a credit score or borrowing history at the time of the loan) is less pronounced, payment history still improves predictive accuracy for them, demonstrating its importance in assessing a wide range of borrower types.

Incorporating payment history-based screening is likely to create winners and losers among borrowers, identified by the change in probability of delinquency (PD) when moving from traditional models to PH-based models. Overall, 60% of borrowers benefit from the inclusion of payment history in lending decisions, favoring established firms and those with low bureau credit scores yet substantial payment activities. However, borrowers with both low credit scores and low payment flows are disadvantaged, with only 40% of such borrowers standing to benefit. Interestingly, moving from PH aggregate to PH granular helps not only lenders but also creates more winners within most borrower categories, implying that many borrowers will have to trade privacy for improved risk assessment by sharing their granular data.

As the discourse around payment data interoperability continues, the use of proprietary payment data in alternative lending by some fintechs and bigtechs has already taken root (BIS, 2019*a*; Liu, Lu and Xiong, 2022; Rishabh and Schäublin, 2021).[3] Many of these programs feature sales-linked repayment structures, where the lending entity receives a portion of the sales they process for the borrowing firm (Rishabh and Schäublin, 2021; Russel, Shi and Clarke, 2023; Liu, Lu and Xiong, 2022). While not directly addressing interoperable payment data, I find that in fintech loans with in-house data under sales-linked loans, payment history outperforms credit bureau data in predicting loan delinquency, and merging these sources doesn't improve predictive power, suggesting credit bureau data might be redundant in sales-linked lending contexts.

Literature and contribution: My research intersects with two strands of literature: the use of alternative data in lending and the interoperability of payment data. The initial wave of research on alternative data in *consumer lending* has seen diverse applications, from analyzing individuals' online behaviors (Berg et al., 2020), mobile phone usage (Agarwal et al., 2024), and grocery shopping patterns (Lee, Yang and Anderson, 2023), to using a broader set of conventional variables (Maggio and Ratnadiwakara, 2024; Jagtiani and Lemieux, 2019; Iyer et al., 2016). Recent inquiries extend into small business financing, particularly bigtechs' use of e-commerce transactions for credit assessments. Frost et al. (2019) compare alternative data-driven risk scores with traditional credit scores in the context of Mercado Libre's sales-linked loans in Argentina, while Huang et al. (2023) examines Alibaba's (MYbank) approach in China in a similar contractual environment.

Relative to these studies, my contributions are threefold. First, by linking transaction data to traditional bank loans, I avoid the BFP critique and assess how *traditional* lenders might benefit from incorporating payment history into risk assessment, extending the relevance of alternative data to the broader traditional lending landscape. Second, while these studies predominantly focus on the *screening* capabilities of alternative data, I explore both its *screening and monitoring* potential, offering a more comprehensive understanding of its value. Third, by analyzing the *same borrowers'* experiences with both bank and sales-linked fintech loans, I shed light on the relative importance of different information sources, including lender soft information and payment history, across diverse loan contractual contexts.[4]

[3] E-commerce giants like Amazon, Mercado Libre, and Ant Financial, and payment fintechs such as Paypal, Square, and Stripe, have notable payment-data-based lending programs.

[4] A wide array of research explores fintech and bigtech credit beyond the alternative data paradigm, focusing on various aspects such as intermediation costs, regulation, convenience, collateralized lending, financial inclusion, and moral hazard. For a thorough review, see Berg, Fuster and Puri (2022).

My work contributes to the evolving discourse on interoperable payment data within the Open Banking framework. Theoretical explorations by He, Huang and Zhou (2023) and Parlour, Rajan and Zhu (2022) have hypothesized about open banking's ramifications on payment service pricing and credit market architecture (see Bianchi et al. (2023) for a review). Concurrent empirical research, such as Ghosh, Vallee and Zeng (2024)'s analysis in India's small business lending sector, show that borrowers engaging in cashless transactions tend to secure loans under better terms, such as lower interest rates and larger loan amounts. Similar findings by Babina et al. (2024) in the UK and Nam (2023) in Germany's consumer lending market suggest that interoperable payment data indirectly benefits both lenders and borrowers.

Building on this foundation, my work offers a *direct* assessment of the value of interoperable payment data by contrasting it with the contributions of other informational sources, such as credit bureaus and lenders' hard and soft information. The granularity of the cashless transaction data I analyze allows me to identify specific payment history variables that significantly influence delinquency risk, shedding light on the particular attributes of payment data that lenders find valuable. This detailed view also informs discussions on Open Banking design policies, weighing privacy considerations against utility. Furthermore, my analysis extends to the post-disbursal phase, evaluating the efficacy of interoperable data in generating early warning signals for loan monitoring.

The checking account hypothesis, which underscores the value of *proprietary-in-house* payment data in loan underwriting (Black, 1975; Fama, 1985; Nakamura, 1993), provides an important historical context. Studies like Puri, Rocholl and Steffen (2017) demonstrate that customers with *in-house* transaction accounts are more likely to achieve favorable credit outcomes. Similarly, Mester, Nakamura and Renault (2007) and Norden and Weber (2010) have highlighted the value of *in-house* transaction data in loan monitoring. Diverging from these approaches, my study explores *interoperable digital payment* data, using its detailed granularity to identify the specific payment history attributes important for effective loan screening and monitoring. My approach, thus, significantly expands the scope of the checking account hypothesis in contexts where payment data originates outside of the lending institution.

Paper structure: The paper is organized as follows: Section 2 details the institutional background and data structure. Section 3 describes the data models and prediction algorithm used. Section 4 presents the results, starting with the assessment of payment data's predictive power in loan screening, exploring borrower heterogeneity, and the importance of different predictor variables. It also explores the utility of payment data in loan monitoring and identifies which borrowers win and which lose from payment data-based loan underwriting. Section 5 shifts focus to sales-linked loans provided by fintech companies. Finally, Section 6 concludes the paper, summarizing key findings and implications.

2 Institutional Set-up and Data

My study is based on a partnership with a leading Indian payment fintech, a significant player in the electronic payment industry, that offers Point of Sale (POS) systems mainly to Micro, Small, and Medium Enterprises (*merchants*). These merchants use the fintech's POS devices for electronic transactions. The analysis uses transaction data from *all* merchants with the fintech's services from January 2015 to February 2019.

Additionally, I have access to borrowing records of a subset of these transacting merchants. This subset represents those merchants that have also availed themselves of its sales-linked lending program. Importantly, the borrowing records include sales-linked loans obtained from the payment fintech as well as *traditional bank loans*, pertaining to the *same borrowers*. The subsequent sections detail these two types of loans. It's important to note that linking traditional bank loans with the transaction history of borrowing merchants from the payment fintech lays the foundation for analyzing *interoperable* payments data. This data is interoperable because it's produced and gathered by one entity (the payment fintech) but is potentially utilized as a screening tool by another (banks).

Bank Loans

The dataset on bank loans comes from borrowing merchants' credit records, obtained from TransUnion CIBIL, India's leading credit bureau. Compiled from financial institution reports, these records primarily document loans to small business owners. While the exact identities of the lending institutions are not disclosed, they include both commercial banks and Non Bank Financial Companies (NBFCs[5]). For clarity, I refer to these entities collectively as 'banks', recognizing their use of traditional standard debt contracts in contrast to the sales-linked loans provided by the payment fintech.

Small business lending often blurs the line between the personal liability of the owner and the business itself (Berger and Udell, 1998; Ang, Lin and Tyler, 1995; Briozzo and Vigier, 2014; Avery, Bostic and Samolyk, 1998). Thus, in my analysis, all loans to business owners, whether labeled as *personal* or *business* by lenders, are treated as business loans due to their interchangeable use. A small proportion of the loans are gold loans which I also treat as as business loans. Indian MSMEs are known to widely utilize gold loans for business financing (Asokan, 2020; Singh and Wasdani, 2016). I exclude specific loans like mortgages or vehicle loans.

Credit bureau records include detailed monthly repayment histories for each loan up to August 2020, covering up to 36 months or until loan closure within this period. With the latest loan in our dataset issued in February 2019, this ensures a minimum of 18 months of repayment

[5]NBFCs are financial institutions that, except for a few allowed to accept *non-demandable* deposits before 1997, do not have a deposit franchise. Since then, the Reserve Bank of India hasn't granted deposit franchises to new NBFCs. NBFCs are also not part of the payment and settlement system and are regulated by the Reserve Bank of India.

data for each loan, essential for pinpointing delinquency occurrences and their timing. I classify a loan as *delinquent* based on any repayment delay of 90 days or more, write-offs, or lender classifications indicating loss such as *Loss*, *Substandard*, *Doubtful*, or *Special Mention Account*.

The records also provide key details like loan disbursement and closure dates, loan types, and contractual terms including amounts, interest rates, and tenure. By integrating this loan information with electronic payment data from the payment fintech, I create a comprehensive payment history for each borrower. Further details on payment data and information from the credit bureau, including credit score are discussed in the subsequent sections.

This analysis covers 11,972 small business loans issued by banks from June 2015 to February 2019, aligned with the scope of the available payment data. To deeply explore the borrowers' credit histories, I evaluate the performance of 130,101 loans (including non-business types) disbursed to them since 1991. This comprehensive review facilitates the calculation of variables that reflect the borrowers' past borrowing activities at the point of acquiring a new loan within our study timeframe. This approach mimics the lenders' information set to replicate the information accessible to the lender at the time of loan sanctioning. Key variables calculated include the total number of loan and credit card accounts a borrower had closed before getting the new loan and the count of active loans at the time of the new loan's approval. For a detailed breakdown of these variables, see Table A1.

Fintech Loans

To analyze the sales-linked loans offered by the payment fintech, I reviewed its loan book up to February 2019, with an update in December 2019. All these loans were unsecured, carrying a standard interest rate of two percent per month. The repayment on these fintech loans was linked to the payments processed by the fintech company as the payment fintech deducted 10% of each sales towards loan repayment. In this context, I define *implied tenure* as the estimated duration (in days) for a borrower to repay the loan (both principal and interest), assuming their sales persist at the same average daily rate as the *pre-disbursal long-term average* and the 10% deduction rate. The long-term average sales are calculated as the daily average over a 90-day period, covering sales from 30 to 119 days *before* loan disbursal.[6] Additionally, merchants could make early repayments, fully or partially, via direct lump-sum payments.[7]

To determine delinquency for fintech loans, I use a snapshot as of December 31, 2019—ten months after issuing the final loan in our study. A loan is considered delinquent if it (i) exceeded its implied tenure and (ii) had a "large" repayment shortfall on the snapshot date. A shortfall is considered large if it surpasses five percent of the total repayment due by December 31,

[6]The days immediately preceding the disbursal date are excluded from the sales average calculation to avoid the inclusion of any short-term, abnormally high sales days that could inaccurately reflect the borrowers' financial stability.

[7]These loan policies share similarities with those of US-based payment fintechs like PayPal and Square. For more information, see Rishabh and Schäublin (2021).

2019. Some delinquent loans were written off, especially if the merchant had left the payment company's network.

The dataset includes 15,325 sales-linked loans issued between May 2017 and February 2019, featuring essential details like loan amounts, suggested repayment periods, and the dates of disbursal and closure, as well as any remaining balance as of December 2019. Using credit bureau records, I compute variables related to previous borrowing, adopting a similar method as with bank loans. Furthermore, I derive payment history variables from the fintech's transaction data.

Other Credit Bureau and Demographic Data

The credit bureau also provides information on credit inquiries and credit scores for the borrowing merchants, which I link with both bank and fintech loans. Credit inquiries represent each occasion a lending institution requested information about the merchant from the bureau, with a total of 346,079 instances. These inquiries signal the merchant's efforts or interest in securing financing, where a large volume of inquiries often suggests an urgent financial need. While the dates of these inquiries are recorded, the identities of the querying institutions are kept confidential.

The bureau assigns credit scores to borrowers on a scale of 300 to 900, with higher scores indicating better creditworthiness. Those without enough history to generate a score fall into the *unscored loans* category. In the lending market, scores above 700 are typically seen as favorable, and I use this threshold to distinguish between *high-score* and *low-score* borrowers.[8] I also gather demographic data on the borrowing merchants, obtained from either the credit bureau or directly from the payment fintech. This information aids in determining the business owner's age, their industry sector, and location.

Payment History Data

I define the term 'payment history' as the information obtained from merchant payment transactions. These histories were compiled from a dataset comprising 99.4 million transactions, each detailed at the card-swipe level, originating from electronic payments made through the fintech's POS devices. This dataset provides an in-depth view of the transactions between merchants and their customers but does not cover all merchant transactions, particularly lacking cash inflows and various outflows.

[8] Interestingly, the fintech lender in this study, like many payment fintechs including the prominent US-based PayPal and Square, did not use credit scores or other bureau data in their lending decisions. More details on PayPal and Square's credit scoring practices can be found at `https://www.paypal.com/workingcapital/faq` and `https://squareup.com/help/us/en/article/6531-your-credit-score-and-square-capital-faqs`, respectively. (Accessed: Dec 10, 2023). I assume banks accessed credit bureau records with the same thoroughness as we do. However, Mishra, Prabhala and Rajan (2022) observed that even traditional banks in India were initially slow to integrate credit scores into their lending criteria, missing out on useful information.

The anonymized transaction data spans from January 2015 to February 2019 and includes activity from approximately 270,000 merchants, encompassing both borrowers and non-borrowers of the fintech's POS systems. The dataset captures each transaction's details, including the amount, date, anonymized card number, and card type, featuring major providers like Amex, Visa, and Mastercard. The breadth of this dataset allows for the development of district-level benchmarks using data from merchants who haven't borrowed. Further details on this approach will be discussed subsequently.

3 Predictive Models and Methodology

3.1 Predictive Models

At the heart of our study are predictive *models* designed to predict loan delinquency. Each model is a combination of variables. By examining how different models perform in prediction, we dig into key questions about what types of information are most valuable for assessing loan risk. We've built several models with this goal in mind, each aimed at either the loan screening or monitoring process. Screening models use information available before the loan is given out, while monitoring models use information gathered after the loan is disbursed to spot early signs of trouble.

We begin by developing three benchmark screening models that capture various types of traditional information commonly utilized in small business lending. To assess the utility of interoperable payment data, we compare and integrate these benchmarks with subsequent payment history-based models. The initial benchmark, the *Credit Bureau* model, incorporates data such as credit scores, the number of inquiries, and past loan history. Expanding upon this, we create the *Traditional Hard Information* model, which augments the credit bureau data with additional details like the merchant's location, industry, and the business owner's age.

Our most comprehensive benchmark is the *Traditional Hard & Soft Information* model. This model includes all aspects of the hard information model plus the loan terms: loan amount, loan tenure, and interest rate. In small business lending, lenders assess borrowers using both concrete, measurable data ("hard information") and "soft information"—subtle cues and observations that aren't easily quantified but influence the loan terms decided by lenders after considering all available information (Agarwal and Hauswald, 2010; Berger and Udell, 2006; Petersen and Rajan, 2002). Thus, the most comprehensive benchmark model, which incorporates a broad spectrum of hard information along with loan terms, serves as the *Traditional Hard & Soft Information* benchmark.

I devise two payment history models. The *Payment History Aggregate (PHA)* model compiles a merchant's electronic sales from four key variables: total sales 90 days before disbursal, growth in average daily sales 30 days prior versus 30-60 days earlier, average transaction size over 90 days, and transaction count in the last 30 days before disbursal. The PHA model essentially provides a summarized perspective of a merchant's electronic sales. The *Payment History Granular (PHG)* model extends PHA by including detailed transaction data and district-level sales comparisons. Because of PHG's depth, it raises privacy and technical concerns. By comparing PHG and PHA models and their combinations, we assess the accuracy benefits of detailed versus aggregate payment data.

To evaluate the importance of payment data, I append the PH models with the traditional benchmarks—credit bureau and traditional models.[9] This assessment involves comparing the performance of these composite models against their individual counterparts. For example, evaluating the *Credit Bureau + PHA* model against the standalone *Credit Bureau* and *PHA* models helps determine whether aggregate payment history and credit bureau data provide complementary information to lenders. If the combined model outperforms the standalone versions, it suggests that integrating these data types offers significant advantages. Likewise, comparing the *Trad Hard Info + PHA* model with the *Trad Hard Info* model shows whether incorporating aggregate payment data into the lender's existing tools is advantageous. For an overview of the variables and models in this analysis, see Table A1 and Table A2.

To assess the utility of payment history in loan monitoring, I extend the *Trad Hard & Soft Info + PHA* (or PHG) pre-disbursal screening model by adding transaction variables across different days post-disbursal (days-since-disbursal or *dsd*). For instance, the PHA model at 30 dsd features metrics such as total sales, average transaction size, daily transaction count, and growth in average sales over the 30 days after loan issuance. A similar approach is applied to PHA 60 dsd, PHA 90 dsd, PHA 120, PHA 150 dsd, and PHA 180 dsd models, with each incorporating sales growth from all previous evaluation windows. For example, the PHA 90 dsd model integrates variables from (Trad Hard & Soft Info + PHA) with 90-day post-disbursal PHA data, including sales growth observed at 30 and 60 days after disbursal. The PHG variants follow a similar approach, adding granular payment variables at various post-disbursal intervals to the *Trad Hard & Soft Info + PHG* screening model. For detailed information on these monitoring models, refer to Table A3.

Table 1 outlines how various models and their comparisons answer specific research questions.

Table 1: Research Questions and Corresponding Models

Research Question	**Model(s) Utilized**
What is the predictive power of credit bureau data for delinquency?	Credit Bureau
What is the relative informativeness of payment history vs. credit bureau data?	PH vs. Credit Bureau
Do payment history and credit bureau data substitute or complement each other?	(Credit Bureau + PH) vs. Credit Bureau; (Credit Bureau + PH) vs. PH
What additional value does payment history bring to the lender in loan screening?	(Trad Hard Info + PH) vs. Trad Hard Info
Do payment histories harden the soft information or bring in new information?	(Trad Hard & Soft Info + PH) vs. Trad Hard & Soft Info
What is the value of granular payment history data?	Models with PHG vs. Models with PHA
What is the efficacy of payment history in early warning and loan monitoring?	(Trad Hard & Soft Info + PH + PH dsd) vs. (Trad Hard & Soft Info + PH)

PH models refer to payment history models, which can be Payment History Aggregate (PHA) or Payment History Granular (PHG). 'Trad' denotes Traditional models, which may include Traditional Hard Information or Traditional Hard & Soft Information.

[9] In merging a PH model with the *Trad Hard & Soft Info* model, I introduce a novel variable: the loan-to-sales ratio, on top of the set of variables that come from each model. This variable compares the loan amount to sales during the 90 days preceding disbursal.

3.2 Predictive Methodology and Evaluation

To predict loan delinquency, I partition the dataset into an 80% training set and a 20% test set. This setup facilitates training a supervised machine learning algorithm, specifically Random Forest. For each predictive model, The algorithm trains based on the variables part of the chosen model and labels indicating actual delinquency status. After training, the algorithm makes out-of-sample delinquency prediction in the test set utilizing the same model variables. I ensure results across models are comparable by maintaining the same random division between train and the test samples across models.

Random Forest, selected for its classification strength, operates as an ensemble of decision trees to increase prediction accuracy. It makes decisions at nodes, choosing the best split from a randomly selected subset of features to classify data effectively. This method and Bootstrap aggregation (bagging)—training each tree on different data samples—minimize correlations between trees, improving the overall prediction quality (Breiman, 2001).

For this study, the Random Forest algorithm is set to construct 400 trees. This quantity ensures a stable ensemble, as adding more trees offers minimal accuracy benefits for this dataset. Tree depth is refined through hyperparameter tuning via Bayesian optimization, parameters like the minimum leaf size, maximum number of splits, and the count of variables evaluated at each split (Hastie, Tibshirani and Friedman, 2008). The efficacy and straightforward parameter adjustment of Random Forests position them as a superior option for certain research tasks compared to other models, such as deep neural networks (Athey and Imbens, 2019; Hastie, Tibshirani and Friedman, 2008).

To evaluate the predictive models, I use the Receiver Operating Characteristic (ROC) Curve and the Area Under the ROC Curve (AUC). The ROC curve illustrates the trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR) across various decision thresholds for predicting delinquency. The TPR quantifies the correctly identified delinquencies, while the FPR reflects the proportion of non-delinquent loans incorrectly marked as delinquent. The AUC is a single-number summary of the model's overall performance. An AUC of 1 signifies perfect prediction; an AUC of 0.5 implies a prediction no better than a random classification.[10]

Despite AUC's utility, its informativeness may wane in datasets with class imbalance, such as when non-delinquent loans vastly outnumber delinquent ones. To address this, I follow Fuster et al. (2022) and calculate average precision (AP). This metric calculates precision (true positives over all positive predictions) adjusted by the increase in TPR (*Recall*) at each threshold, gauging the model's capability in spotting rare events. It provides a complementary measure to AUC, with higher values indicating more precise predictions of delinquency. For reliability, I compute 95% confidence intervals for AUC and AP using bootstrapping with 1000 test set replicas.

[10]AUC also has a probabilistic interpretation. It gives the probability that the model ranks a randomly selected delinquent loan higher than a non-delinquent one.

3.3 Summary Statistics

Table 2 offers summary statistics on bank loans, showing an average loan amount of INR 75,358 with a median of INR 99,708, and an average borrower age of 35 years, indicating a relatively young cohort. Over the 90 days prior to loan disbursal, the average aggregate sales amount to INR 73,865. Credit inquiries within 60 days before obtaining a loan average at 2.45, but a median of one suggests the average is influenced by a few borrowers with multiple inquiries; indeed, 25% of borrowers have three or more inquiries. Additionally, borrowers typically manage seven active loans or credit card accounts when securing a new loan, highlighting a tendency to tap into various credit sources for liquidity.

Table 2: Bank Loans: Summary Statistics on Borrower Payment, Demographic, and Loan Variables

Summary statistics based on 11972 loans made by banks to the merchants using payment services of the payment fintech. All nominal monetary variables are denominated in INR. CV refers to coefficient of variation. For detailed variable description see Table A1.

Variable	Mean	Median	Std Deviation	p10	p25	p75	p90
Payment Variables							
Sales growth	0.20	-0.07	1.15	-0.75	-0.42	0.40	1.28
Avg daily # transact (log)	0.79	0.65	0.64	0.07	0.32	1.12	1.67
Avg transact size (log)	7.48	7.36	1.27	6.04	6.67	8.19	9.17
CV daily sales	2.52	2.09	1.68	0.96	1.39	3.10	4.51
CV transact size	1.55	1.23	1.11	0.64	0.83	1.92	2.88
District aggregate sales (log)	20.05	20.58	1.68	17.54	19.12	21.38	21.67
Growth in district sales	0.05	0.06	0.10	-0.06	0.01	0.11	0.17
Median transact size	2823.80	800.00	7501.34	250.00	399.75	1500.00	5000.00
Aggregate sales (log)	11.21	11.97	3.15	9.45	11.17	12.66	13.32
Share of district sales	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Change in share of district sales	0.00	0.00	0.02	0.00	0.00	0.00	0.00
Share of transact through Visa or Master	0.87	0.89	0.12	0.73	0.82	0.96	1.00
Traditional Variables (Demographic, Bureau, and Loan terms)							
Owner age (Years)	35.41	34.12	7.73	26.85	29.83	39.37	45.96
Has credit score (1= Yes)	0.95	1.00	0.22	1.00	1.00	1.00	1.00
Length of credit history (Years)	6.19	5.07	4.77	0.89	2.25	9.83	13.01
# previously closed loans	6.28	5.00	5.98	0.00	2.00	9.00	15.00
# bureau enquiries	2.45	1.00	3.16	0.00	0.00	3.00	6.00
# active loans	7.34	6.00	5.25	2.00	4.00	10.00	14.00
Credit score	716.96	726.00	47.61	655.00	685.00	750.00	773.80
Share closed loans colltrl	0.46	0.44	0.37	0.00	0.10	0.80	1.00
Share closed loans non-perf	0.04	0.00	0.08	0.00	0.00	0.00	0.17
Share non-perf in active loans	0.02	0.00	0.09	0.00	0.00	0.00	0.00
Loan amount (log)	11.23	11.51	1.46	8.99	10.41	12.21	12.90
Rate of interest (Annual percent)	19.31	23.31	7.13	9.20	11.64	24.00	26.00
Loan tenure (Months)	17.03	12.00	14.69	4.00	12.00	24.00	36.00
Outcome Variables							
Delinquent (1 = Yes)	0.09	0.00	0.29	0.00	0.00	0.00	0.00

The average credit score for borrowers stands at 720, categorizing the average borrower within the ‘prime’ segment, typically defined by scores above 700. Yet, more than a quarter of borrowers do not meet this benchmark. Significantly, 10% of borrowers had no borrowing history prior to their current bank loan, and 5% lacked a credit score at borrowing, highlighting diverse credit histories among these merchants. On average, business loans from banks span a tenure of 18 months with an interest rate of 17%, aligning with standard conditions for small business lending.

Table 3 provides fintech loan statistics, revealing significant distinctions from bank loans. Fintech loans, on average smaller at INR 26,108, carry a higher interest rate of 24% annually. These loans are characterized by borrowers with shorter credit histories and fewer closed loans in comparison to those who take bank loans. Notably, every borrower in my sample with a bank loan has also engaged with fintech lending, indicating the difference in loan attributes is due to borrowing patterns rather than borrower diversity. This suggests borrowers with extensive banking relationships tend to take fewer fintech loans, highlighted by their shorter credit histories and fewer closed loans in the fintech context. Conversely, those newer to bank borrowing are more inclined towards fintech lending.

Table 3: Fintech Loans: Summary Statistics on Borrower Payment, Demographic, and Loan Variables

Summary statistics based on 15325 loans made by payment fintech to the merchants using its payment services. All nominal monetary variables are denominated in INR. CV refers to coefficient of variation. For detailed variable description see Table A1.

Variable	Mean	Median	Std Deviation	p10	p25	p75	p90
Payment Variables							
Sales growth	0.41	0.03	1.37	-0.54	-0.28	0.49	1.53
Avg daily # transact (log)	0.99	0.85	0.66	0.29	0.51	1.34	1.89
Avg transact size (log)	7.41	7.32	1.07	6.11	6.70	8.03	8.77
CV daily sales	2.00	1.71	1.19	0.86	1.18	2.48	3.47
CV transact size	1.55	1.22	1.05	0.69	0.86	1.90	2.85
District aggregate sales (log)	20.14	20.88	1.65	17.51	19.22	21.38	21.62
Growth in district sales	0.05	0.05	0.14	-0.07	-0.01	0.10	0.17
Median transact size	2141.23	845.00	5872.91	270.00	425.00	1500.00	3000.00
Aggregate sales (log)	12.26	12.23	0.94	11.25	11.69	12.80	13.37
Share of district sales	0.01	0.00	0.04	0.00	0.00	0.00	0.01
Change in share of district sales	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Share of transact through Visa or Master	0.86	0.88	0.10	0.73	0.81	0.94	0.98
Traditional Variables (Demographic, Bureau, and Loan terms)							
Owner age (Years)	36.32	34.81	8.87	26.59	29.75	41.06	48.22
Length of relationship w/ the lender (months)	15.15	13.77	8.64	5.22	8.31	20.07	27.17
Has credit score (1= Yes)	0.90	1.00	0.29	1.00	1.00	1.00	1.00
Length of credit history (Years)	3.96	2.03	4.72	0.00	0.05	5.99	11.69
# previously closed loans	3.80	1.00	6.12	0.00	0.00	5.00	11.00
# bureau enquiries	0.98	0.00	1.85	0.00	0.00	1.00	3.00
# active loans	2.72	2.00	2.98	0.00	0.00	4.00	7.00
Credit score	713.25	726.00	53.48	639.00	681.00	753.00	773.00
Share closed loans colltrl	0.41	0.33	0.37	0.00	0.00	0.70	1.00
Share closed loans non-perf	0.10	0.00	0.21	0.00	0.00	0.11	0.33
Share non-perf in active loans	0.10	0.00	0.25	0.00	0.00	0.00	0.50
Loan amount (log)	10.17	10.13	0.84	9.21	9.62	10.71	11.33
Loan tenure (Days)	112.82	90.00	43.96	90.00	90.00	180.00	180.00
Outcome Variables							
Delinquent (1 = Yes)	0.12	0.00	0.33	0.00	0.00	0.00	1.00

Tables A4 and A5 categorize bank and fintech loans by performance and include a two-sample t-test to highlight mean differences between performing and delinquent loans. This approach, while indicative of certain trends, only considers mean values and misses the broader data distribution and any non-linear patterns. For a detailed analysis that acknowledges these complexities, see Section 4.3.

4 Results on Interoperable Payment Data for Bank Loans

4.1 Interoperate Payment Data for Loan Screening

We start our analysis by establishing predictive performance benchmarks using traditional models, focusing on metrics such as the Area Under the Curve (AUC) and Average Precision (AP). Figure 1 displays the AUC and AP for various models, including their 95% confidence intervals, with detailed model performances provided in Appendix Table A6. Additionally, Figure A2 illustrates the Receiver Operating Characteristic (ROC) curves for calculating AUCs.

The Credit Bureau model, our initial benchmark, achieves an AUC of 0.59, surpassing the random-guess baseline of 0.5, and an AP of 0.07, which meets expectations for predictive modeling in this context. The next model broadens the dataset to include a more extensive set of hard information, with 14 variables related to borrowing history, credit scores, industry, etc. This *Traditional Hard Info* model shows improved performance (AUC = 0.62, AP = 0.14) compared to the Credit Bureau model alone, marking a notable increase in predictive capability. The most comprehensive traditional benchmark model, which integrates both hard and soft information, sees an additional 6 percentage points (pp) increase in AUC and a 5 pp increase in AP. These gains signify considerable relative improvements of 11% in AUC and 31% in AP compared to the model using hard information only.

Next, we evaluate payment history models relative to these traditional benchmarks, initially examining the Payment History Aggregate (PHA) models before moving to the granular models. The PHA model, incorporating four payment history variables, achieves an AUC of 0.59, which is equal to the Credit Bureau model, and an AP of 0.11, showing a 4 percentage point (pp) increase over the Credit Bureau model. The *Credit Bureau + PHA* model raises the AUC by 8 pp and the AP by 4 pp compared to the PHA model alone. This demonstrates that combining payment history with credit bureau information significantly improves prediction accuracy in bank loans, suggesting a complementarity between the two data sources.

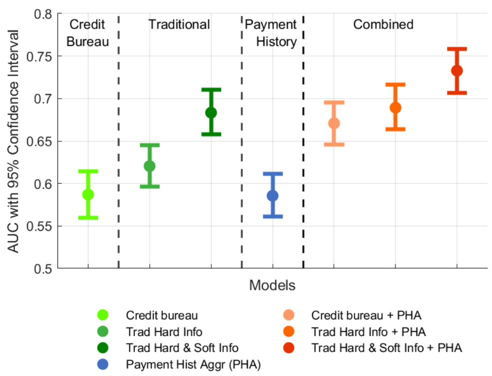
In examining a counterfactual scenario where a small-business lender includes PHA variables with traditional hard information, we aim to quantify the added value of this integration. Comparing the *Traditional Hard Info + PHA* model against the *Traditional Hard Info* benchmark, we find that adding PHA improves the AUC by 7 percentage points (to 0.69) and the AP by 5 percentage points (to 0.19) over the hard information benchmark.

The final test for the utility of payment history is its comparison with the most comprehensive traditional benchmark model, which encompasses all the hard and soft information a lender has access to. Should payment history improves this benchmark, it would indicate that it

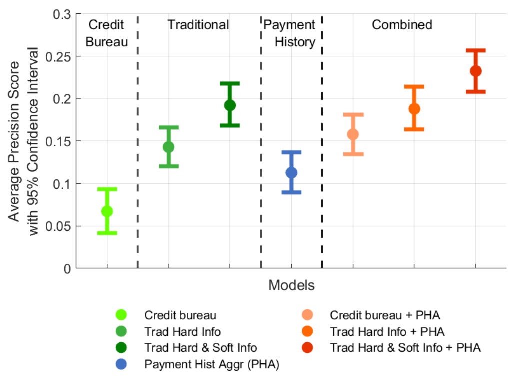
Figure 1: Bank Loans: Predictive Model Performance Comparison

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. The 95% confidence interval for the AUC and AP are calculated by bootstrapping with 1000 replicas of the test set. Among the payment-history based models, the aggregate (*PHA*) variant is employed. For detailed variable description see Table A1. For the composition of predictive models see Table A2.

(a) Area Under the ROC Curve



(b) Average Precision



acts as a distinct and complementary data source for lenders. Conversely, should there be no improvement, it would imply that payment history either does not provide additional information to the lender or that such information is already considered and captured in the contractual variables that are part of this traditional benchmark.

Our findings indicate that payment history serves as a novel information source for lenders. This is evidenced by the noticeable improvement in the predictive power of the *Traditional Hard & Soft Info + PHA* model compared to the *Traditional Hard & Soft Info* benchmark model. According to Figure 1, adding PHA to this benchmark model raises the AUC by 7% to 0.73 and the AP by 21% to 0.23. Figure A3 in the appendix presents the 95% confidence intervals for the differences in AUCs and APs among the models compared here. The figure confirms that the differences are statistically significant.

Transitioning to the Payment History Granular (PHG) model, which incorporates a total of 12 payment variables to add depth beyond the PHA model's four. Our aim is to evaluate the improvement in predictive accuracy that these granular variables offer over the PHA models, both in standalone and combined implementations. Table 4 reproduces the performance metrics of the Credit Bureau and Traditional models for as benchmarks, and introduces the results for PHG-based predictive models, along with the changes in performance metrics when moving from PHA to PHG versions. The standalone PHG model demonstrates a notable improvement over the PHA model, with a 4.7% increase in AUC and a 10.3% increase in AP. These improvements are also evident in the combined models, underscoring the significant value of granular payment data.

Table 4: Bank Loans: Out-of-Sample Predictive Performance with Granular Payment History

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models see Table A2. Granular payment variables, as opposed to aggregate payment variables, necessitate transaction-level information or are calibrated against district-level payment aggregates. The first four columns of this table replicate those in Table A6. The percentages indicate changes relative to the corresponding aggregate model detailed in Table A6.

		Traditional		Payment History		Models Combined with PHG		
Predicted var: Delinquency	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Granular (PHG) (5)	Mod (1) + Mod (5) (6)	Mod (2) + Mod (5) (7)	Mod (3) + Mod (5) (8)
Area Under the Curve (AUC)	0.59	0.62	0.68	0.59	0.61	0.69	0.70	0.76
% Δ compared to Agg model	-	-	-	-	4.73	2.19	2.09	3.65
Average Precision	0.07	0.14	0.19	0.11	0.12	0.20	0.20	0.26
% Δ compared to Agg model	-	-	-	-	10.28	27.85	7.86	12.23
N. Obs. Train	9578	9578	9578	9578	9578	9578	9578	9578
N. Obs. Test	2394	2394	2394	2394	2394	2394	2394	2394
N. Predictors	9	14	17	4	12	21	26	30

Before we examine the importance of our findings, let's set them against studies that use alternative data within different contexts. Papers such as Agarwal et al. (2024); Maggio and Ratnadiwakara (2024); Iyer et al. (2016); Berg et al. (2020), within consumer lending, find AUCs between 0.66 and 0.73 for alternative data models. In small-firm credit, Frost et al. (2019) and Huang et al. (2023) find AUCs of 0.76 and 0.87, respectively, in bigtech lending in Argentina and China, characterized by sales-linked repayments. Our model augmented with PHA shows an AUC of 0.73, and our model with PHG achieves 0.76, placing them within these observed ranges. However, drawing direct comparisons proves challenging because our analysis is centered on *interoperable payment data* and *bank lending*, distinct from the sales-linked contracts prevalent in bigtech scenarios.[11]

Value of the interoperable payment history: Our findings indicate that lenders can significantly benefit from improved predictive performance by incorporating payment history. What is the economic magnitude of these benefits? I investigate this from two perspectives: first, by examining the predictability provided by interoperable payment data compared to the lender's soft information; and second, by calculating the return on investment (ROI) for a portfolio of loans selected based on traditional metrics, and comparing this with the ROI from an equally sized portfolio assembled using PH-based metric.

In small business lending, lacking detailed financial records, lenders depend on other hard information but also collect *soft information* about borrowers. This combination influences contractual loan terms. Our comprehensive traditional benchmark model, by incorporating all hard information and contractual loan terms, thus, encapsulates both these types of information.

[11]Comparing other results, studies such as Agarwal et al. (2024) and Maggio and Ratnadiwakara (2024) find credit score predictability with AUCs around 0.51 to 0.53 in consumer lending, in India and the US. Iyer et al. (2016) finds a higher AUC of 0.62 in US peer-to-peer lending, while Berg et al. (2020) finds even higher AUCs in German delayed-payment schemes. Our credit bureau model, featuring extensive credit report analytics beyond credit scores, aligns with these higher findings.

Therefore, comparing the benchmark *Traditional Hard & Soft Info* model with the *Traditional Hard Info* model gives an estimate of the value of soft information. The results in Figure 1 show that soft information is worth about a 6 pp increase in AUC and a 5 pp rise in AP.[12] Interestingly, as shown above, adding PHA to the *Traditional Hard Info* model results in similar gains (7 pp in AUC and 5 pp in AP) as those seen with soft information. This finding implies that aggregate payment variables offer value on par with lenders' soft information.

This raises a question: Might PHA variables essentially reflect what is traditionally seen as soft information, effectively "hardening" it? If PHA primarily transforms soft information into hard data, integrating it with the comprehensive benchmark *Traditional Hard & Soft Info* model should not significantly improve performance. However, our results show otherwise, as the *Traditional Hard & Soft Info + PHA* model outperforms the comprehensive benchmark. Nonetheless, the performance gain with PHA (5 pp in AUC and 4 pp in AP) doesn't fully match its informational value, suggesting that PHA partially hardens soft information. For instance, the contrast between PHA's 7 pp lift over hard info alone and its 5 pp boost over a model combining hard and soft information indicates that approximately 29% (2/7) of PHA's value could be interpreted as hardening of soft information, with the remaining 71% (5/7) attributable to its unique, complementary contribution. A similar calculation for PHG models would suggest that granular payment history offers 20% to 30% greater value than the lender's soft information, varying with the performance metric applied.

In the second method to assess the economic significance of improvements in performance metrics, I conduct back-of-envelope calculations based on the methodology outlined in Netzer, Lemaire and Herzenstein (2019). I calculate the expected profits for each loan under the predictive model m, using the initial contractual variables set, as:

$$\mathbb{E}[\pi_i^m] = (1 - PD_i^m)(L_i r_i t_i) + PD_i^m(L_i r_i t_i \alpha - L_i), \tag{1}$$

where PD_i^m is the model m's assigned probability of delinquency for loan i. L_i, r_i, and t_i represent the loan amount, annual interest rate, and tenure in years, respectively, as recorded in the dataset. The parameter α reflects the proportion of tenure during which repayment occurs before delinquency. I calibrate α to 0.3, based on the average point to default observed in my dataset. I further assume that, in the event of delinquency, the lender loses the loan principal, barring the partial interest payment calculated above. Next, I rank loans based on their expected profits using two models: $m =$ *Trad Hard Info* and $m =$ *Trad Hard Info + PHG*. I create two portfolios from the top '*s*' percent of loans ranked by each of these models respectively, investing

[12]One might surmise that the differences in model performance discussed might not only reflect soft information but could also reflect hard information used by the lenders that eludes us as researchers. Yet, in the small business lending context we examine, the likelihood of us missing hard information available to the lender is minimal. In fact, the reality might be the reverse. Our models assume lenders fully utilize credit bureau data, included in our traditional benchmark. However, Mishra, Prabhala and Rajan (2022) indicates that lenders often overlook available credit bureau data.

one unit of capital in each loan within these equal-sized portfolios. I assume the unit capital is invested in each loan at its original interest rate and tenure.

As a final step, I calculate the *actual* return for each of the portfolio based on the real delinquency status of the constituent loans within each portfolio. This gives me the *actual* return on investment (ROI) for both the portfolios. By comparing the actual ROI for the portfolio formed using *Trad Hard Info + PHG* model against the ROI from the portfolio formed using the traditional *Trad Hard Info* model, I derive the incremental ROI of including payment history in screening. For instance, a selective portfolio comprising the top 20% of loans yields an incremental ROI of approximately 5% from including payment history. Meanwhile, a broader portfolio aiming to include nearly 90% of the originally granted loans shows a 1% incremental ROI from payment history inclusion. Figure A4 in the Appendix details these calculations across various selection scenarios, demonstrating that the ROI improvements attributable to payment history are economically significant.

We summarize our main findings of this section as follows:

Takeaway 1 (a) *Payment history matches the credit bureau in predictive accuracy for loan delinquency, significant synergy when combined, showing that both sources capture distinct, complementary risk facets.*

(b) *Incorporating payment history into the most comprehensive traditional benchmark model improves prediction, underlining its widespread applicability. The value from payment data is comparable to that of lender's soft information, with about 29% of PHA's contribution resulting from hardening soft information, and the remainder from unique, new signals.*

(c) *Rough estimates suggest that using payment history can significantly increase the lender's return on investment. Choosing the top 20% of loans identified by the PHA model results in a 5% greater ROI than selecting the top 20% using only traditional hard information.*

4.2 Comparative Informational Value of Interoperable Payment Data Across Heterogeneous Borrowers

Our analysis also investigates if the effectiveness of Payment History (PH) models holds uniformly across firms of different sizes and credit scores. We divide borrowers into 'small' and 'large' categories based on their transaction volumes 90 days before getting a loan—small firms have transactions below the median value, while large firms exceed it.

I perform a predictive analysis akin to the baseline, this time separately for each borrower size category. Table 5 details how models perform across firm sizes. Findings suggest payment history predicts small firm outcomes better than those of large firms (Column 4). This could be because small firms often use one payment system, unlike larger firms that might use several, possibly reducing the predictive power of data from any single source. Furthermore, for small firms, payment data effectively indicates operational cash flows, closely linked to their repayment ability. In contrast, larger firms have more financial resources and alternative financing options, helping them manage cash flow variations.

Table 5: Bank Loans: Out-of-Sample Predictive Performance with Aggregate Payment History

by Borrower Business Size

Small borrowers have sales in the 90-day pre-disbursal period that fall below the median, while large borrowers exceed it. AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models see Table A2.

		Traditional		Payment History	Model Combined with PHA		
Predicted var: Delinquency	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Mod (1) + Mod (4) (5)	Mod (2) + Mod (4) (6)	Mod (3) + Mod (4) (7)
Area Under the ROC Curve (AUC)							
Small	0.58	0.63	0.68	0.60	0.70	0.71	0.74
Large	0.61	0.60	0.68	0.57	0.64	0.67	0.72
Average Precision							
Small	0.09	0.18	0.26	0.13	0.24	0.24	0.27
Large	0.08	0.11	0.17	0.10	0.13	0.17	0.24
[Ntrain small, Ntrain large]	[4789, 4789]	[4789, 4789]	[4789, 4789]	[4789, 4789]	[4789, 4789]	[4789, 4789]	[4789, 4789]
[Ntest small, Ntest large]	[1197, 1197]	[1197, 1197]	[1197, 1197]	[1197, 1197]	[1197, 1197]	[1197, 1197]	[1197, 1197]
N. Predictors	9	14	17	4	13	18	22

Next, we assess how payment history compares with other data sources across firm sizes. Table 5 shows that while credit bureau data typically better predicts the loan outcomes of larger borrowers (column 1), expanding to the full range of hard information alters this trend across firm sizes (column 2), indicating additional hard data offsets credit bureau limitations for smaller firms. This context explains why combining PHA with traditional hard information (column 6 vs. column 2) benefits firms of any size, despite PHA's particular effectiveness for smaller firms. Furthermore, incorporating PHA into the *Trad Hard & Soft Info* model (column 7 vs. column 3) boosts performance for lenders serving different borrower categories. The distinct effects on small versus large firms, and whether these arise from refining soft information or introducing new, independent insights, remain somewhat unclear. However, it's evident that PHA integration provides clear advantages, either by delivering new signals or reducing the effort to collect soft information.

We further investigate how model performance differences vary according to borrowers' credit scores and histories. Borrowers are categorized into *low-scored*, with scores below 700, and *high-scored*, above this threshold, adhering to the creditworthiness standards prevalent in India's lending industry[13]. We aim to see if payment history yields novel signals for both high and low-score groups, particularly assessing its significance for *thin-file borrowers*—those with no credit score or borrowing history at the time of loan application.

[13]See `https://www.cibil.com/faq/understand-your-credit-score-and-report` (Accessed: December 10, 2023).

I conduct separate analyses for loan sub-samples divided by credit-score categories. The uneven distribution of credit scores, particularly the smaller numbers of low-score and thin-file borrowers, requires a distinct evaluation approach, since smaller sample sizes for these sub-samples make it harder to train models. To address this, I employ five-fold cross-validation within the random forest framework for all sub-samples. This technique divides the sub-sample into five equal parts, each part alternately used as the test set once and as part of the training set four times. Despite being computationally-intensive, this method ensures every loan in the sub-sample is used for testing, facilitating out-of-sample predictions. It also results in tighter confidence intervals, crucial for sub-samples with fewer loans.

Table 6 presents our findings. Just as with the baseline sample, PHA's improve the prediction over traditional models across borrower categories. However, we also observe a positive yet somewhat smaller improvement in prediction accuracy for thin-file borrowers. In terms of the magnitude, the PHA's contribution mirrors that of soft information for both high-scored and low-scored borrowers, evident from comparing the AUC and AP increases in the *Trad Hard + PHA* model with the *Trad Hard Info* model, and those in the *Trad Hard & Soft Info* model. While the lenders also benefit from adapting PHA in predicting delinquency of the thin-file borrowers, the uplift is slightly less pronounced compared to soft information, as shown by comparing results between columns 7 and 3.

Table 6: Bank Loans: Out-of-Sample Predictive Performance with Aggregate Payment History

by Borrower Credit Score Status

High-score borrowers are those with credit scores above 700 on a scale of 300 to 900. Low-score borrowers have scores below 700. Thin-file borrowers either lacked a credit score at the time of borrowing or had no previous borrowing records. AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models see Table A2. Results are from out-of-sample predictions using five-fold cross-validation in the random forest algorithm, where each data subset is alternately used as a testing set and part of the training set, ensuring each observation is predicted out-of-sample once.

		Traditional		Payment History	Model Combined with PHA		
Predicted var: Delinqency	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Mod (1) + Mod (4) (5)	Mod (2) + Mod (4) (6)	Mod (3) + Mod (4) (7)
				Area Under the ROC Curve (AUC)			
High Score	0.61	0.63	0.67	0.56	0.65	0.67	0.72
Low Score	0.58	0.61	0.68	0.55	0.65	0.68	0.74
Thin File	-	0.55	0.65	0.59	-	0.63	0.67
				Average Precision			
High Score	0.11	0.13	0.17	0.10	0.16	0.17	0.22
Low Score	0.14	0.16	0.23	0.13	0.17	0.23	0.28
Thin File	-	0.12	0.18	0.16	-	0.18	0.21
Ntrain [high, low, thin]	[7643, 3714, -]	[7643, 3714, 1875]	[7643, 3714, 1875]	[7643, 3714, 1875]	[7643, 3714, -]	[7643, 3714, 1875]	[7643, 3714, 1875]
Ntest [high, low, thin]	[7643, 3714, -]	[7643, 3714, 1875]	[7643, 3714, 1875]	[7643, 3714, 1875]	[7643, 3714, -]	[7643, 3714, 1875]	[7643, 3714, 1875]
N. Predictors	9	14	17	4	13	18	22

To summarize, the key takeaways from our analysis in this section are as follows:

Takeaway 2 (a) *Payment history proves more effective for smaller firms, mainly due to their use of a single payment system and the direct link between their cash flows and* *repayment abilities. However, when combined with traditional data, payment history benefits firms of all sizes, demonstrating its broad applicability.*

(b) *Payment history aids the lenders in evaluating both high-score and low-score borrowers. While the boost for thin-file borrowers is less marked, PHA still improves prediction for them, showing its importance in assessing a wide range of borrower types.*

4.3 Which Predictors are the Most Important in Screening?

Our primary goal is to determine which *pre-disbursal* variables are key in predicting loan delinquency, a complex task due to the *black-box* nature of algorithms like random forests that rely on exploiting non-linear relationships. Thanks to advances in interpretable machine learning, highlighted by Molnar (2023), we can now unravel these complexities using two methods: Shapley additive explanations (SHAP) and Out-of-bag (OOB) variable importance through permutation. We focus on identifying critical variables using the most comprehensive screening model—*Traditional Hard & Soft Info + PHG*. This selection enables an in-depth comparison of the predictive power of both granular and aggregated payment data against traditional variables.[14] Our discussion will mainly focus on SHAP, with a detailed exploration of OOB variable importance provided in the appendix.

Shapley Additive Explanations (SHAP), developed by Lundberg and Lee (2017), analyze the impact of individual features on specific predictions, unlike broader methods like OOB permutation importance that evaluate feature importance across the entire dataset. SHAP values show how each feature influences a prediction's deviation from the model's average prediction. Features with higher absolute SHAP values have a greater impact on the prediction. This method, based on cooperative game theory, views prediction as a collective effort of features contributing to a "surplus"—the difference from the average prediction. The Shapley value calculates each feature's *fair share* of this surplus. A positive SHAP value means a feature increases the likelihood of delinquency, while a negative one suggests a decrease.

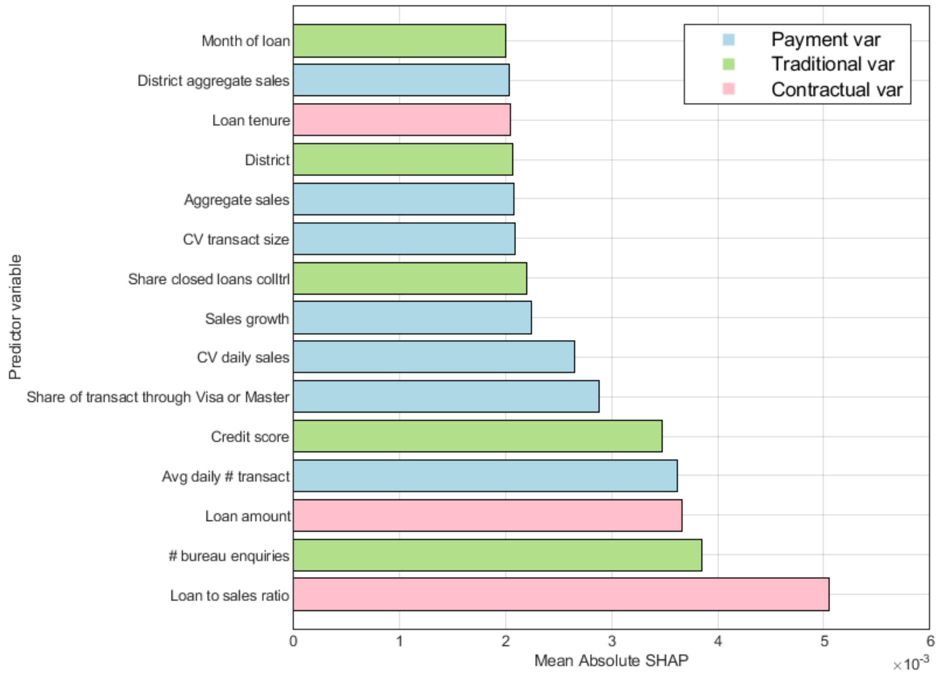
Given the substantial computational requirements for calculating SHAP values, I restrict the calculation to a random 50% sample of the test set, encompassing about 1200 out-of-sample predictions. This method strikes a balance between computational manageability and the depth of analysis. We use SHAP estimates to first identify variables by their overall (*global*) significance. Then, we explore directionality of the relationship between predictor values and the probability of delinquency.

To gauge each predictor's impact in our models, we calculate the mean absolute SHAP value for each variable across all predictions. This average provides a measure of the variable's

[14]This model, including contractual variables, serves as a benchmark for analysis rather than a practical screening exercise itself. By examining payment history and traditional predictors within this model, we uncover their predictive strength and understand their significance amidst a wide array of predictors, including contractual variables.

Figure 2: Bank Loans: Variable Importance Based on Mean Absolute SHAP

The figure displays the mean of absolute SHAP values for each predictor, where the mean is calculated over all predictions within a 50% random sample of the test set. Absolute SHAP values quantify the degree to which each predictor influences a prediction's deviation from the mean outcome, indicating the predictor's impact. A higher mean absolute SHAP value signifies greater *overall* importance of the variable. The SHAP analysis is conducted using the *Traditional Hard & Soft Info + PHG* model, which encompasses 30 predictors, including 12 payment history-related variables. Of these, 4 are aggregative PH variables (aggregate sales, average per-day transaction count, average transaction size, and sales growth), and the remaining 8 pertain to granular PH. The analysis also considers 14 traditional variables, 3 contractual terms (loan amount, tenure, interest rate), and the loan-to-sales ratio, derived from the loan amount. The figure displays the top 15 variables out of the 30 analyzed. For detailed information on the variables and model composition, refer to Tables A1 and A2.



overall importance, with higher values indicating greater influence. Figure 2 shows the top 15 of the 30 variables analyzed, grouping them into payment, traditional, and contractual categories. While we focus mainly on payment and traditional variables, the figure also includes contractual variables such as interest rates, loan tenure, loan amount, and the loan-to-sales ratio, which are set by lenders or calculated from the loan amount. Although included in the analysis for completeness, these contractual variables fall outside our core investigative scope due to their endogenous nature.

Figure 2 shows the main predictors of loan delinquency. Importantly, three aggregate payment variables—average daily transaction count, total sales, and sales growth—are among the top 15, marking their predictive importance. Additionally, from the granular payment history variables, the variability measures such as the coefficient of variation of daily sales emerge as crucial. Interestingly, the number of credit bureau enquiries stands out for its predictive power alongside other notable traditional variables like credit score, borrower location, and the month the loan was issued.

The Out-of-Bag (OOB) permutation method, explained in Appendix B, provides additional peek into variable importance in Random Forest models. It corroborates our previous results and expands the analysis to loan sub-samples based on firm size, showing variable importance variations. For example, *Credit score* is key for large borrowers but not as much for small ones. In essence, while payment variables, especially granular ones, are significant for large borrowers, the importance of aggregate payment variables is higher for small borrowers.

To analyze the direction of influence of various predictors on loan delinquency, we refer to SHAP feature dependence plots in Figure 3. These plots show the connection between each predictor's value and its effect on the probability of delinquency, represented by SHAP values. For each predictor, we also overlay a polynomial curve in the dependence plots. The selection of the curve's degree is determined by the best data fit, as indicated by the adjusted-R^2 value.

The findings are revealing: higher values of the payment history variables, such as total sales and sales growth, typically are associated with lower delinquency probability, indicating a negative correlation. In contrast, average transaction sizes has a positive correlation with the delinquency probability. The variability in sales and transaction sizes exhibits a U-shaped relationship with delinquency risk, implying that low and high variability is associated with increased risk. For traditional variables, better credit scores and longer credit histories correspond with lower delinquency risks, whereas a rise in bureau inquiries suggests an elevated risk.

As we wrap up this section, it's important to note a significant limitation of our approach. In predicting loan delinquency, the vital role of non-linearities and interactions among variables profoundly influences outcomes. Random Forest algorithms excel at identifying these complex interactions. While we model variable self-interactions using polynomial functions, fully grasping the impact of interactions between variables is still a challenge, despite the progress in interpretable machine learning methods.

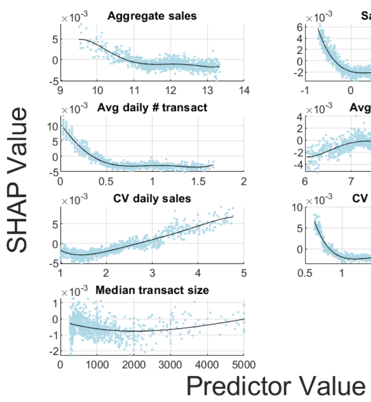
Takeaway 3 (a) *Three aggregate payment variables—average daily transaction count, total sales, and sales growth—stand out as significant predictors of loan delinquency. Additionally, granular variables, such as the variability in daily sales, are notably influential, particularly for larger borrowers. Traditional variables, including credit score and bureau inquiries, also display their anticipated predictive strength.*

(b) *Variables from aggregate payment history, such as total sales, average daily transactions, and sales growth, show a negative relationship with the likelihood of delinquency. In contrast, larger transaction sizes are associated with higher risk. Among traditional predictors, a higher credit score is associate with lower delinquency probability, while more bureau inquiries indicate a higher risk.*

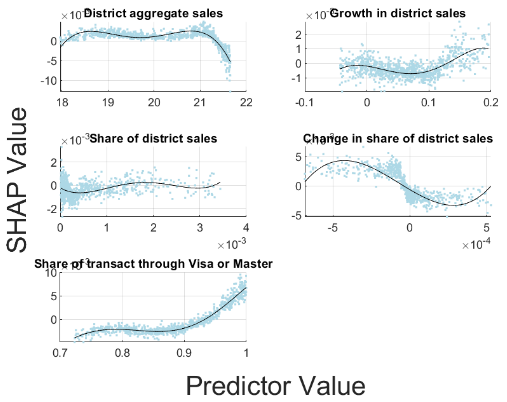
Figure 3: Bank Loans: SHAP Dependence Plots

The figure illustrates the relationship between each predictor's value and its corresponding SHAP value across a 50% random sample of the test set predictions. SHAP values indicate each predictor's influence in shifting a prediction from the average. Higher absolute SHAP values signify a greater contribution to a particular prediction. Positive SHAP values increase the probability of delinquency, while negative values decrease it. The SHAP analysis is conducted using the *Traditional Hard & Soft Info + PHG* model, which encompasses 30 predictors, including 12 payment history-related variables. Of these, 4 are aggregative PH variables (aggregate sales, average per-day transaction count, average transaction size, and sales growth), and the remaining 8 pertain to granular PH. The analysis also considers 14 traditional variables, 3 contractual terms (loan amount, tenure, interest rate), and the loan-to-sales ratio, derived from the loan amount. Among the traditional variables, the figure plots the dependence of 8 numerical variables, excluding the 6 categorical ones. Each plot includes a polynomial fit of degree N, ranging from 1 to 5, with the optimal degree selected based on the highest adjusted-R^2 value. For a detailed breakdown of variables, see Table A1; for model specifics, refer to Table A2.

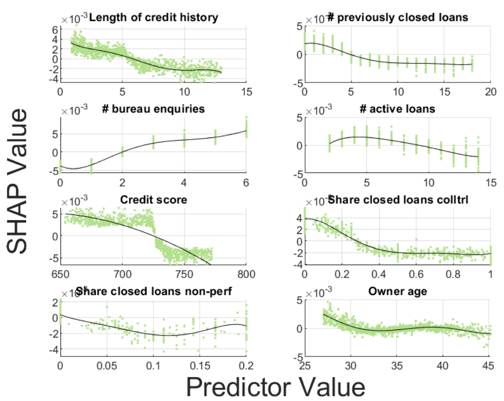
(a) Payment Variables Part-I



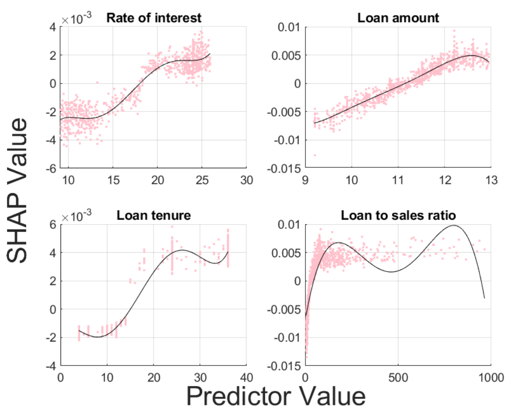
(b) Payment Variables Part-II



(c) Traditional Variables



(d) Loan Contractual Terms



4.4 Interoperable Payment Data for Loan Monitoring

We now turn to loan monitoring, specifically assessing delinquency risks post-disbursal. Real-time payment data may provide a frequent snapshot of a borrowing business's financial status. A key lender concern is identifying when loan repayment prospects start to worsen to initiate timely interventions. To explore the potential of payment history variables as early warning indicators, I perform predictive analysis at six consecutive 30-day intervals following loan disbursal. This method adds chosen post-disbursal payment history variables to the pre-disbursal *Trad Hard & Soft Info + PH model* for each specified period, as illustrated by an example monitoring model in Table A3. It examines both the aggregate (PHA) and detailed (PHG) versions of post-disbursal models, evaluating their predictive capability against the full pre-disbursal model (*Trad Hard & Soft Info + PH model*).

Figure 4: Bank Loans: Predictive Performance Comparison in Early Warning Models

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. Post-disbursal prediction involves augmenting the pre-disbursal *Traditional Hard & Soft Info model + PH* with additional post-disbursal payment history variables, calculated within each respective time window since disbursal (days-since-disbursal(dsd)). Granular payment (*PHG*) variables, as opposed to aggregate payment (*PHA*) variables, necessitate transaction-level information or are calibrated against district-level payment aggregates. The 95% confidence interval for the AUC and AP are calculated by bootstrapping with 1000 replicas of the test set. For detailed variable description see Table A1. For the composition of predictive models see Table A2.

(a) Area Under the ROC Curve (b) Average Precision

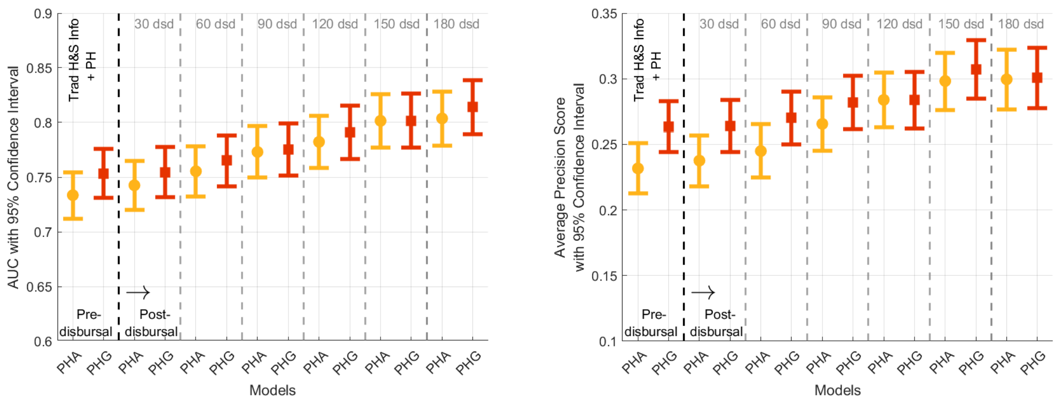


Figure 4 details the performance of our early warning system using PHA and PHG variables, showing consistent improvements in AUC and AP throughout the 180-day period following disbursal. Impressively, by 180 days post-disbursal, the PHA model achieves a 7 percentage point increase in both AUC and AP from pre-disbursal levels. To put these results into context, the contribution of payment history to AUC within the initial 120 days post-disbursal (roughly 5 pp) is comparable to the contribution of payment history over the lender's hard and soft information in the *pre-disbursal* period. Similarly, for AP, a comparable increase (about 4 pp) is observed within the first 90 days post-disbursal,

A practical method to gauge the monitoring utility of payment data is by tracking the updates in the estimated probability of delinquency over time within the test sample. Ideally, with the accrual of more payment data, the delinquency probability should increase for loans heading towards delinquency and decrease for those that remain performing. Figure 5 illustrates the change in the average out-of-sample estimated probability of delinquency for both delinquent and performing loans, providing several important observations.

Firstly, the effectiveness of pre-disbursal models is evident, as the average delinquency probability for loans identified as likely to default exceeds the overall sample delinquency rate (*unconditional delinquency probability*). This outcome underscores the strong predictive performance of the screening models we've discussed. Secondly, as expected, the delinquency probability for loans on a path to default gradually increases, though not a uniform rate. Specifically, for PHA-based models, this probability rises from about 11% pre-disbursal to 16% by 180 days post-disbursal, illustrating the dynamic updation of risk assessment.

Lastly, early warning models employing detailed (*PHG*) data initially outperform the aggregate (*PHA*) versions right after disbursal, a trend clearly depicted in Figure 5. After disbursal, PHG models begin with an advantage over PHA models, indicated by their differences prior to disbursal, and they update the delinquency probability at a slightly quicker pace early on. This results in the PHG model raising the delinquency probability to 14% within two months of disbursal, a mark the PHA model achieves a month later. Over time, the advantage of detailed data diminishes, highlighting the increasing significance of *aggregate* payment variables as more data accumulates. Interestingly, the PHG model's performance slightly declines at 180 days post-disbursal compared to 150 days, hinting that adding too many payment variables may complicate the model's learning process, possibly due to overfitting. This pattern suggests a diminishing difference between detail and accuracy in loan monitoring as time progresses.

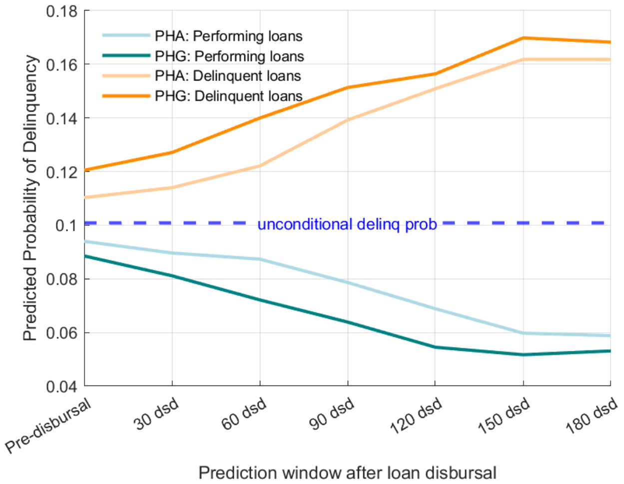
To identify key variables in early warning models for predicting delinquency—and the direction of their impact—we use a method similar to our screening analysis, but now focused on the post-disbursal phase. We start by determining the mean absolute SHAP value for the model 90 days post-disbursal (*90 dsd*). It's important to understand that the primary differences between post-disbursal models stem from the payment variables they include. Models with longer observation windows not only include the same PH variables as those with shorter windows but also additional variables up to their respective window. This logic leads us to choose the 90-dsd model as representative for our analysis of variable importance, positioned as it is midway through our observation windows.

Figure 6 shows the 15 most important variables, offering two key observations: first, post-disbursal PH variables significantly exceed their pre-disbursal equivalents in terms of importance. Second, among post-disbursal variables, those associated with sales growth stand out, highlighting their value in indicating potential financial distress. Traditional variables, such as bureau inquiries and credit score, also continue to be important predictors after disbursal, confirming their sustained significance throughout the loan lifecycle.

To discern the direction of relationships, we present SHAP dependence plots in Figure 7, focusing exclusively on post-disbursal payment variables due to spatial constraints. The direction and observations concerning pre-disbursal traditional and payment history variables

Figure 5: Bank Loans: Post-Disbursal Updating in Predicted Probability of Delinquency

Post-disbursal prediction involves augmenting the pre-disbursal *Traditional Hard & Soft Info model + PH* with additional post-disbursal payment history variables, calculated within each respective time window since disbursal (days-since-disbursal(dsd)). Granular payment (*PHG*) variables, as opposed to aggregate payment (*PHA*) variables, necessitate transaction-level information or are calibrated against district-level payment aggregates. For detailed variable description see Table A1. For the composition of predictive models see Table A2. The probabilities represent the out-of-sample average predicted probability of delinquency, calculated for distinct groups of borrowers categorized based on their eventual delinquency status.



are consistent with those identified in the screening analysis (Figure 3) and are thus not duplicated here. Significantly, Figure 7 shows that a decline in sales growth post-disbursal is associated with an increased delinquency risk. Conversely, higher total sales, a greater average number of daily transactions, and a larger share of district sales are associated with reduced delinquency risk after disbursal.

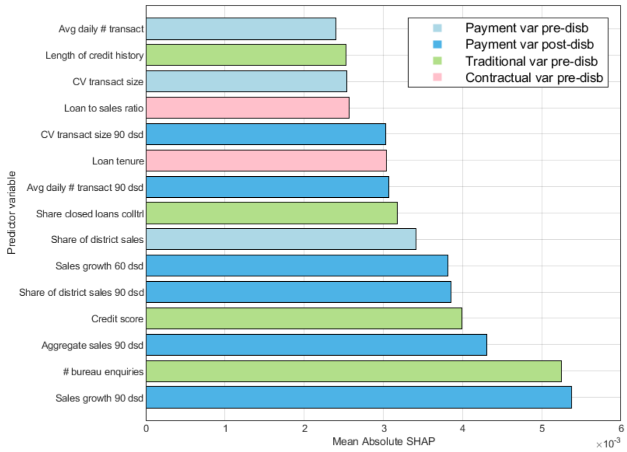
Takeaway 4 (a) *Payment data holds significant potential for generating early warning signals, aiding lenders in monitoring loans. Within 120 days after loan disbursal, post-disbursal payment history variables improve AUC to the same degree that pre-disbursal payment data does compared to traditional lending information under screening.*

(b) *The initial advantage of granular (PHG) data over aggregate (PHA) in early warning models is pronounced but transient. The performance gap narrows quickly, with aggregate model's effectiveness catching up over the life of the loan, highlighting a short-lived trade-off between privacy and accuracy in monitoring.*

(c) *Deterioration in sales growth post-disbursal is directly linked to higher delinquency probabilities, establishing sales growth metrics as essential early warning signals.*

Figure 6: Bank Loans: Variable Importance Based on Mean Absolute SHAP in Early Warning Model

The figure displays the mean of absolute SHAP values for each predictor, where the mean is calculated over all predictions within a 50% random sample of the test set. Absolute SHAP values quantify the degree to which each predictor influences a prediction's deviation from the mean outcome, indicating the predictor's impact. A higher mean absolute SHAP value signifies greater *overall* importance of the variable. The SHAP analysis is conducted using the *Traditional Hard & Soft Info + PHG* model augmented with payment variables at 90 dsd, which encompasses 44 predictors, including 26 payment history-related variables. Of these, 14 are post-disbursal PH variables, and the remaining 12 are pre-disbursal. The analysis also considers 14 traditional variables, 3 contractual terms (loan amount, tenure, interest rate), and the loan-to-sales ratio, derived from the loan amount. The figure displays the top 15 variables out of the 44 analyzed. For detailed information on the variables and model composition, refer to Tables A1 and A2.



4.5 Who Benefits and Who Loses from the Use of Payment History?

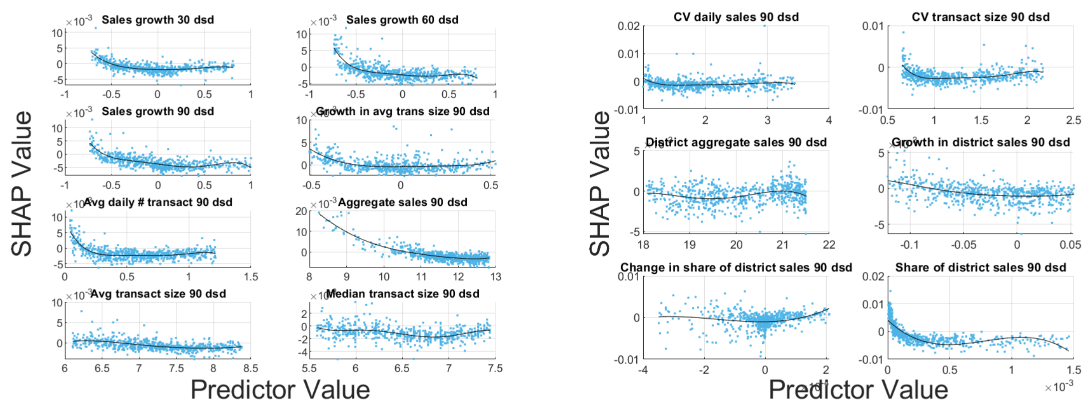
Shifting our analysis to the borrower's perspective, we explore the impact of incorporating payment histories into lenders' loan pricing strategies, specifically identifying which types of borrowers stand to benefit or lose from such a practice. Following the methodology outlined by Fuster et al. (2022), I compare the predicted probability of delinquency (PD) for borrowers based on lenders using a combination of traditional hard information and payment history against PDs calculated using only hard information. Borrowers experiencing a decrease in their PDs benefit from the inclusion of payment history, while those with an increase in PDs lose out in comparison to the traditional screening.

Specifically, I assess the relative changes in PDs by calculating the logarithmic differences between the PDs from the *Trad Hard Info + PH Granular* model and those from the *Trad Hard Info* model. To understand the distributional impacts of employing payment history in loan screening, I analyze these changes across four borrower categories. These categories are created by intersecting two binary classifications: borrower size and credit score types. In Figure 8, I depict the distributions of relative PD differences for both the individual borrower categories and the collective borrower pool. For interpretation, borrowers positioned left of the zero

Figure 7: Bank Loans: SHAP Dependence Plots for Post-disbursal Variables in Early Warning Model

SHAP values indicate each predictor's influence in shifting a prediction from the average. Higher SHAP values signify a greater contribution to a particular prediction. Positive SHAP values increase the probability of delinquency, while negative values decrease it. This analysis employs the 90-dsd post-disbursal model, which integrates various post-disbursal PHG variables into the pre-disbursal Traditional Hard & Soft Info + PHG model. Due to space constraints, the figure highlights SHAP dependence plots exclusively for post-disbursal variables. The findings concerning pre-disbursal PH and traditional variables are consistent with those presented in the pre-disbursal SHAP analysis depicted in Figure 3. For comprehensive details on variables, refer to Table A1; for an overview of model composition, see Table A2.

(a) Post-disbursal Payment Variables Part-I (b) Post-disbursal Payment Variables Part-II



line in this figure are considered "winners", benefiting from payment history-based screening, whereas those on the right are "losers", adversely affected by this approach.

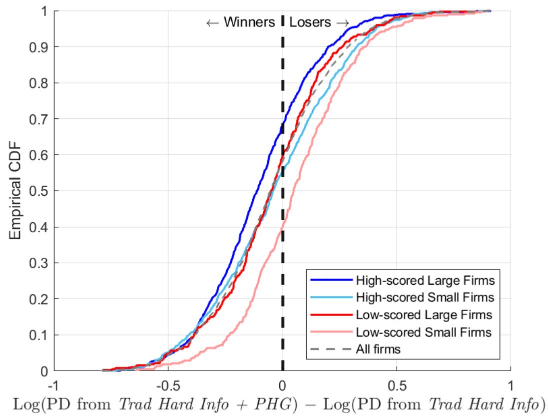
Several notable findings emerge from Figure 8 regarding the impact of including Payment History (PH) in loan screening. Firstly, PH inclusion positively affects approximately 60% of borrowers, as indicated by the point where the cumulative distribution function (CDF) for all firms intersects the zero line. Secondly, there's notable variation in this overall positive effect. High-scored-large firms see the greatest advantage, with 70% emerging as winners from PH-based screening. In contrast, only 40% of low-scored-small firms experience benefits, with most encountering increased delinquency probabilities. The other categories—high-scored-small and low-scored-large firms—also predominantly show improved creditworthiness with PH screening, reflecting diverse benefits across groups.[15]

These findings indicate that Open Banking is likely to favor established borrowers, yet they aren't the exclusive benefactors. Firms with low credit scores but high payment flows—termed as low-scored-"large" firms—also stand to gain, revealing themselves as *invisible primes*. An interesting point to note is the trade-off between the granularity of payment history and the

[15]The CDF for the three positively impacted borrower groups is consistently higher than that for low-scored-small firms, suggesting that even among borrowers who do not benefit, those from the three categories face less detriment than those from the low-scored-small firms.

Figure 8: Bank Loans: Relative Changes in Probability of Delinquency

This figure presents the Empirical Cumulative Distribution Function (CDF) of the differences in the logarithm of probabilities of delinquency (PD), comparing the models *Trad Hard Info + PH Granular* and *Trad Hard Info* for loans in our baseline test sample. A negative difference indicates a borrower benefits from incorporating PH variables into the traditional hard information model, as it leads to a lower delinquency probability. Conversely, a positive difference suggests a borrower faces a higher delinquency risk with the addition of PH variables. Borrowing firms are classified based on their sales volume in the 90-day period prior to loan disbursement and their credit scores. *small borrowers* have sales below the median, whereas *large borrowers* have sales above it. High-scored firms are those whose owner have credit scores above 700 on a scale of 300 to 900. Low-scores borrowers have scores below 700. For details on the predictive models' composition, refer to Table A2. For information on the variables, see Table A1.



degree of benefits in reducing the PD. As illustrated in Appendix Figure A5, transitioning from aggregated to granular payment history generally aids most borrower categories, indicating that more detailed data yields broader benefits.

Takeaway 5 Incorporating payment history into lending decisions benefits 60% of borrowers, notably favoring established firms and those with low credit scores yet substantial payment activities. Conversely, borrowers characterized by both low credit scores and low payment flows are at a disadvantage, with only 40% standing to benefit.

5 Payment Data and Fintech Loans

Payment fintech loans present a compelling case study in how altering contractual features impacts the information content of payment and traditional variables in assessing delinquency risk. Around the world, payment fintechs and bigtech entities are developing sales-linked loans, with loan repayments directly connected to the merchant's sales processed via the lending fintech or bigtech. This section explores the effectiveness of traditional and payment history variables in assessing the credit risk in these innovative loans. This exercise is particularly informative because the borrower samples for both bank and fintech loans are the same—each individual with a bank loan in our analysis has also obtained at least one fintech loan.

In examining fintech loans, we begin by assessing the screening process using pre-disbursement variables, integrating benchmark traditional models with Payment History Aggregate (PHA) variables. Our baseline findings, presented in Figure 9 and further detailed in Table A7 in the appendix, indicate notable differences from the bank loan screening results previously outlined in Section 4.1. Notably, the predictive effectiveness of the Credit Bureau, as gauged by the Area Under the Curve (AUC), diminishes for fintech loans. Yet, this pattern reverses when considering Average Precision (AP), leading to inconsistent evidence about the Credit Bureau's relative importance in different lending environments. Therefore, we will further investigate variable importance measures to clarify this aspect.

Additionally, traditional models show less predictive strength for fintech loans than for bank loans, suggesting that private, soft information plays a lesser role. This conclusion comes from noting that the traditional model incorporating both hard and soft information barely outperforms the model using only hard information in fintech contexts, unlike the significant boost observed in bank loan scenarios. Given that loan terms often reflect a lender's soft information, this pattern implies that fintech lenders collect fewer soft signals than banks.

The examination of the Payment History Aggregate (PHA) model in fintech loans yields important observations. With an AUC of 0.62 and an AP of 0.2, it outperforms the Credit Bureau. The central question is whether PHA complements or substitutes the Credit Bureau. Combining both models shows no increment in predictive accuracy beyond what the PHA model achieves on its own, suggesting that the PHA model captures all relevant information from the Credit Bureau concerning fintech loans.

This result has profound implications. The effectiveness of the PHA model in fintech loans shows it can adequately support credit decisions, potentially serving as an alternative where traditional credit reporting is challenging. Thus, in environments where setting up credit bureaus is costly, introducing sales-linked loans for businesses could be a strategic option to decrease dependence on these bureaus.

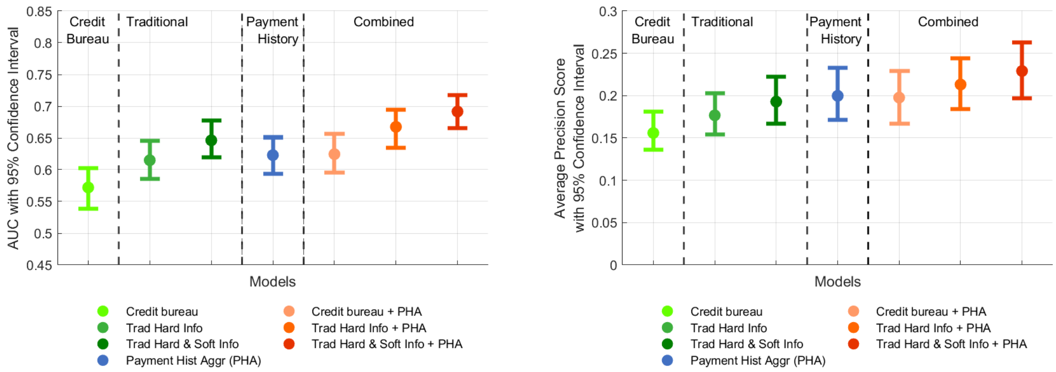
Incorporating Payment History Granular (PHG) variables into our models leads to significant improvements. The standalone PHG model, with an AUC of 0.65 as shown in Appendix Table A8, surpasses the standalone PHA model and matches the traditional comprehensive benchmark model, *Trad Hard & Soft Info*. While the success of PHG suggests potentially less reliance on other data types, integrating PHG with traditional loan terms (*Trad Hard & Soft Info + PHG*) increases prediction accuracy, indicating that loan terms provide unique signals beyond payment history even in fintech lending. Nonetheless, the additional value of loan terms, hence lender soft information, appears to be less in fintech than in bank lending. Despite fintech lending's improved performance with PH variables, a comprehensive model including PH, hard, and soft information doesn't reach the predictive capability of its bank loan equivalent, underscoring the vital contribution of soft information in bank lending contexts.

To identify the key factors influencing delinquency in fintech loans, I analyzed the most comprehensive screening predictive model,*Trad Hard & Soft Info + PHG*, using SHAP values,

Figure 9: Fintech Loans: Predictive Model Performance Comparison

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. The 95% confidence interval for the AUC and AP are calculated by bootstrapping with 1000 replicas of the test set. For detailed variable description see Table A1. For the composition of predictive models see Table A2.

(a) Area Under the ROC Curve (b) Average Precision



similar to the approach for bank loans. Results, shown in Figure A7 in the appendix, present the top 15 variables based on mean absolute SHAP value. Sales variability stands out as the most impactful payment variable, even more so than traditional ones. Interestingly, the credit score doesn't make it to the top 15, while the number of bureau inquiries still shows some relevance, though not as much as in bank loan contexts. Coupled with earlier observations on the Credit Bureau model's efficacy, this indicates a diminished role for credit bureau data in fintech lending. Further observations from SHAP dependence plots in Appendix Figure A7 indicate that higher sales variability increases delinquency risks, while transaction size and its variability exhibit a U-shaped relationship with delinquency.

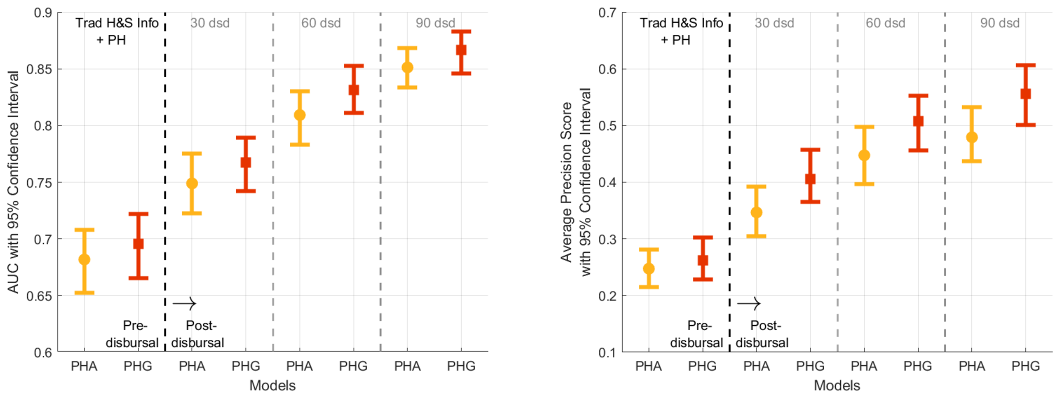
Shifting our focus to assessing the monitoring capabilities of payment variables in sales-linked fintech loans, I augment the pre-disbursal benchmark model—*Trad Hard & Soft Info + PH*—with post-disbursal payment variables. This study carries out predictive analyses across three specific 30-day periods post-loan issuance, a departure from the six intervals examined for bank loans. This adjustment accounts for the generally shorter durations of fintech loans compared to traditional bank loans. The analysis, shown in Figure 10, demonstrates the significant and swift impact of post-disbursal PH variables on prediction accuracy. Notably, there's a 10% rise in AUC at 30 days and a 20% uplift at 60 days post-disbursal, relative to the pre-disbursal benchmark.

To summarize, while combined models for *screening* in sales-linked fintech lending may not be as predictive as those in traditional bank loans, fintech-specific monitoring models quickly close this gap post-disbursal. The changing nature of post-disbursal sales data is particularly

Figure 10: Fintech Loans: Predictive Performance Comparison in Early Warning Models – Aggregate and Granular Payment History

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. Post-disbursal prediction involves augmenting the pre-disbursal *Traditional Hard & Soft Info model + PH* with additional post-disbursal payment history variables, calculated within each respective time window since disbursal (days-since-disbursal(dsd)). Granular payment (*PHG*) variables, as opposed to aggregate payment (*PHA*) variables, necessitate transaction-level information or are calibrated against district-level payment aggregates. The 95% confidence interval for the AUC and AP are calculated by bootstrapping with 1000 replicas of the test set. For detailed variable description see Table A1. For the composition of predictive models see Table A2.

(a) Area Under the ROC Curve (b) Average Precision



pivotal for these loans. This aligns with findings from studies like Rishabh and Schäublin (2021) and Russel, Shi and Clarke (2023), highlighting moral hazard in fintech lending, where merchants might manipulate sales to defer repayments. Thus, while sales-linked loans reduce dependence on traditional, backward-looking data sources, they could amplify moral hazard issues. This necessitates further exploration to grasp the broader consequences of these loan agreements.

We can summarize our findings regarding the fintech loans as below:

Takeaway 6 (a) *In fintech loan screening, the Payment History Aggregate (PHA) model surpasses the Credit Bureau model in effectiveness, with no added advantage from combining them. This indicates traditional backward-looking credit bureau data may be redundant in sales-linked fintech contexts.*

(b) *The significant improvement in prediction accuracy from post-disbursal PH variables in fintech lending could indicate moral hazard problems associated with sales-linked repayment strategies designed to postpone repayment.*

6 Conclusion

The significant drawbacks of bureau-based credit scoring, the currently dominant lending technology, have fueled the growing interest in interoperable payment data through Open Banking initiatives. The insufficient coverage of credit bureaus, leaving over half of the world's firms and individuals unrepresented, is a glaring issue.[16] Moreover, traditional credit scores, primarily focused on past borrowing actions, often fail to accurately reflect a borrower's current financial situation, even for those covered by bureaus. In contrast, payment histories, emerging from widespread electronic payment transactions, offer a real-time and frequent signals.

My findings show that payment histories offer valuable information to lenders that is distinct from what lenders typically possess. However, in traditional lending, payment histories serve to complement credit bureau data rather than replace it. My findings support the creation of credit bureaus that combine credit and transaction histories, with Open Banking initiatives potentially facilitating this integration. Moreover, in situations where creating credit bureaus is prohibitively costly, using payment history for underwriting is still a better choice than having no access to either bureau or payment data. Interestingly, sales-linked loans offer yet another alternative, because they render credit bureaus obsolete. Yet, these contracts suffer from moral hazard challenges. This necessitates further research to fully uncover the informational requirements and incentive issues in sales-linked lending.

My results also have implications for Open Banking design policies. First, while most borrowers are likely to benefit from underwriting based on payment history, certain groups, especially those with low credit scores and sparse payment histories, may be negatively affected, raising concerns about distributional impacts. Second, policy-makers should consider clear trade-offs. Granular payment data improve lender capabilities beyond aggregate payment data, but they might increase technological costs for entities sharing this data due to more sophisticated standardized data collection, storage, and sharing protocols, exacerbating the challenges the current system already faces (BIS, 2019*b*; Sia Partners, 2019). Moreover, while granular data could potentially benefit most borrowers in the form of better credit terms, they also raise privacy concerns, implying that some borrowers might need to choose between improved credit terms and maintaining data privacy.

[16] This is illustrated in Figure A1 in the appendix.

References

Agarwal, Sumit, and Robert Hauswald. 2010. "Distance and Private Information in Lending." *Review of Financial Studies*, 23(7): 2757–2788.

Agarwal, Sumit, Shashwat Alok, Pulak Ghosh, and Sudip Gupta. 2024. "Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech." *Journal of Money, Credit, and Banking*. Forthcoming.

Ang, James S, James Wuh Lin, and Floyd Tyler. 1995. "Evidence on the lack of separation between business and personal risks among small businesses." *Journal of Small Business Finance*, 4(2/3): 197–210.

Asokan, Mamtha. 2020. "Gold loans shine among small businesses." *Times of India (Jun 9, 2020)*.

Athey, Susan, and Guido W Imbens. 2019. "Machine learning methods that economists should know about." *Annual Review of Economics*, 11: 685–725.

Avery, Robert B, Raphael W Bostic, and Katherine A Samolyk. 1998. "The role of personal wealth in small business finance." *Journal of Banking & Finance*, 22(6-8): 1019–1061.

Babina, Tania, Saleem Bahaj, Greg Buchak, Filippo De Marco, Angus Foulis, Will Gornall, Francesco Mazzola, and Tong Yu. 2024. "Customer Data Access and Fintech Entry: Early Evidence from Open Banking." *Stanford University Graduate School of Business Research Paper*.

Berger, Allen N, and Gregory F Udell. 1998. "The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle." *Journal of Banking & Finance*, 22(6-8): 613–673.

Berger, Allen N, and Gregory F Udell. 2006. "A more complete conceptual framework for SME finance." *Journal of Banking & Finance*, 30(11): 2945–2966.

Berg, Tobias, Andreas Fuster, and Manju Puri. 2022. "Fintech lending." *Annual Review of Financial Economics*, 14: 187–207.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri. 2020. "On the rise of FinTechs: Credit scoring using digital footprints." *Review of Financial Studies*, 33(7): 2845–2897.

Bianchi, Milo, Matthieu Bouvard, Renato Gomes, Andrew Rhodes, and Vatsala Shreeti. 2023. "Mobile payments and interoperability: Insights from the academic literature." *Information Economics and Policy*, 65: 101068.

BIS. 2019*a*. "Big tech in finance: opportunities and risks." *BIS Annual Economic Report*.

BIS. 2019*b*. "Report on open banking and application programming interfaces (APIs)." Basel Committee on Banking Supervision. Accessed on 1 Feb 2024.

Black, Fischer. 1975. "Bank funds management in an efficient market." *Journal of Financial Economics*, 2(4): 323–339.

Breiman, Leo. 2001. "Random forests." *Machine Learning*, 45(1): 5–32.

Briozzo, Anahi, and Hernán Vigier. 2014. "The role of personal loans in the financing of SMEs." *Academia Revista Latinoamericana de Administración*, 27(2): 209–225.

Experian. 2022. "Business and Consumer Insight Report." Report.

Fama, Eugene. 1985. "What's different abount Banks?" *Journal of Monetary Economics*, 15: 29–39.

Frost, Jon, Leonardo Gambacorta, Yi Huang, Hyun Song Shin, and Pablo Zbinden. 2019. "BigTech and the changing structure of financial intermediation." *Economic Policy*.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. "Predictably unequal? The effects of machine learning on credit markets." *Journal of Finance*, 77(1): 5–47.

Ghosh, Pulak, Boris Vallee, and Yao Zeng. 2024. "FinTech Lending and Cashless Payments." *Journal of Finance*. Forthcoming.

Hastie, T., R. Tibshirani, and J. Friedman. 2008. *The Elements of Statistical Learning, Second Edition.* New York: Springer.

He, Zhiguo, Jing Huang, and Jidong Zhou. 2023. "Open banking: Credit market competition when borrowers own the data." *Journal of Financial Economics*, 147(2): 449–474.

Hines, Kieran. 2021. "Expectation versus reality for payments data monetisation." *Report by Celent, a division of Oliver Wyman, Inc.*

Huang, Yiping, Zhenhua Li, Han Qiu, Sun Tao, Xue Wang, and Longmei Zhang. 2023. "BigTech credit risk assessment for SMEs." *China Economic Review*, 81: 102016.

Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F. P. Luttmer, and Kelly Shue. 2016. "Screening Peers Softly: Inferring the Quality of Small Borrowers." *Management Science*, 62(6): 1554–1577.

Jagtiani, Julapa, and Catharine Lemieux. 2019. "The roles of alternative data and machine learning in fintech lending: evidence from the LendingClub consumer platform." *Financial Management*, 48(4): 1009–1029.

Lee, Jung Youn, Joonhyuk Yang, and Eric Anderson. 2023. "Using Grocery Data for Credit Decisions." *Wharton Customer Analytics Research Paper*.

Liu, Lei, Guangli Lu, and Wei Xiong. 2022. "The big tech lending model." National Bureau of Economic Research.

Lundberg, Scott M, and Su-In Lee. 2017. *Proceedings of Advances in Neural Information Processing Systems*, , ed. I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan and R. Garnett Vol. 30, Chapter A Unified Approach to Interpreting Model Predictions. Curran Associates, Inc.

Maggio, M. Di, and D. Ratnadiwakara. 2024. "Invisible primes: Fintech lending with alternative data." Working Paper No. w29840. National Bureau of Economic Research.

Mester, Loretta J, Leonard I Nakamura, and Micheline Renault. 2007. "Transactions accounts and loan monitoring." *Review of Financial Studies*, 20(3): 529–556.

Mishra, Prachi, Nagpurnanand Prabhala, and Raghuram G Rajan. 2022. "The relationship dilemma: Why do banks differ in the pace at which they adopt new technology?" *The Review of Financial Studies*, 35(7): 3418–3466.

Molnar, Christoph. 2023. *Interpretable machine learning.* . Second ed.

Nakamura, Leonard I. 1993. "Commercial bank information: Implications for the structure of banking." In *Structural Change in Banking*. , ed. M. Klausner and L. White.

Nam, Rachel J. 2023. "Open banking and customer data sharing: Implications for fintech borrowers."

Netzer, Oded, Alain Lemaire, and Michal Herzenstein. 2019. "When words sweat: Identifying signals for loan default in the text of loan applications." *Journal of Marketing Research*, 56(6): 960–980.

Norden, Lars, and Martin Weber. 2010. "Credit line usage, checking account activity, and default risk of bank borrowers." *Review of Financial Studies*, 23(10): 3665–3699.

Parlour, Christine A, Uday Rajan, and Haoxiang Zhu. 2022. "When fintech competes for payment flows." *The Review of Financial Studies*, 35(11): 4985–5024.

Petersen, Mitchell A, and Raghuram G Rajan. 2002. "Does distance still matter? The information revolution in small business lending." *Journal of Finance*, 57(6): 2533–2570.

Puri, Manju, Jörg Rocholl, and Sascha Steffen. 2017. "What do a million observations have to say about loan defaults? Opening the black box of relationships." *Journal of Financial Intermediation*, 31: 1–15.

Rishabh, Kumar, and Jorma Schäublin. 2021. "Payment Fintechs and Debt Enforcement." *WWZ Working Paper Series*.

Russel, Dominic, Claire Shi, and Rowan P Clarke. 2023. "Revenue-Based Financing." *Available at SSRN*.

Sia Partners. 2019. "About PSD2 and APIs: opportunities and challenges." `https://www.sia-partners.com/en/insights/publications/about-psd2-and-apis-opportunities-and-challenges`, Accessed on 1 Feb 2024.

Singh, Charan, and Poornima Wasdani. 2016. "Finance for micro, small, and medium-sized enterprises in India: Sources and challenges."

Appendices

A Additional Figures and Tables

A.1 Additional Tables

Table A1: Variable Description

Variable	Description
Payment Variables	
Sales growth	Relative change in the average per-day transaction value 30 days pre-disbursal compared to that in the window 30-60 days pre-disbursal
Avg daily # transact (log)	Total number of transaction in the 30-day window prior to disbursal / 30
Avg transact size (log)	Total value of transaction / Number of transaction; calcualted in the 90-day window before disbursal
CV daily sales	Coefficient of variation of daily value of transactions in the 90-day window before disbursal
CV transact size	Coefficient of variation of transaction values in the 90-day window before disbursal
District aggregate sales (log)	Total value of transactions by all the merchants in the district of the borrowing merchant in the 90-day window pre-disbursal
Growth in district sales	District level growth in value of transactions over the same period as sales growth
Median transact size	Median transaction amount in the 90 days leading up to disbursal
Aggregate sales (log)	Total value of transactions in the 90-day window before disbursal
Share of district sales	Aggregate sales / District sales
Change in share of district sales	Change in share of district sales between the windows same as in sales growth
Share of transact through Visa or Master	Share of transactions done through Visa or Mastercard
Traditional Variables (Demographic, Bureau, and Loan terms)	
Owner age (Years)	Age of the business owener
Length of relationship w/ the lender (months)*	Months since the first transaction recorded by the Payment Fintech
Has credit score (1= Yes)	Indicator variable = 1, if the merchant had a credit score available at the time of borrowing
Length of credit history (Years)	Years since the first loan in the bureau records
# previously closed loans	Number of loans (including credit card accounts) closed prior to the loan
# bureau enquiries	Number of enquiries made to the bureau in the 60 days prior to loan disbursal
# active loans	Number of loans by the borrower (including credit card accounts) that were running at the time of the loan
Credit score	TransUnion CIBIL score. Ranging between 300 and 900. 700+ considered high credit score
Share closed loans colltrl	Fraction of closed loans that were collateralized
Share closed loans non-perf	Fraction of previously closed loans that were delinquent
Share non-perf in active loans	Proportion of active loans classified as delinquent at disbursal
District	District of the borrower
State	State of the borrower
Industry	Borrower industry indentified by the SubGroup of Merchant Category Codes (MCC) Classification
Month of Loan Disbursal	Calendar month of the loan disbursal
Loan amount (log)	Loan amount
Rate of interest (Annual percent)**	Rate of interest
Loan tenure (Months)	Tenure of the loan
Combined Variables	
Loan-sales ratio	Loan amount / Average per-day transaction value in the 90-day window pre-disbursal
Outcome Variables	
Delinquent (1 = Yes) Bank Loans	Indicator for loans 90+ days overdue or classified under regulatory loss categories: Written off, Loss, Substandard, Doubtful, or Special Mention Account
Delinquent (1 = Yes) Fintech Loans	Indicator for loans that were delayed and had a "large" shortfall (pending amount ≥ 5% of due amount) as on the cut-off date of 31 December 2019.

* Variables used as a predictor only in fintech loan analysis.
** Variables used as a predictor only in bank loan analysis.
All monetary values are in Rupees. Transactions refer to the electronic transactions processed by the payment fintech for the merchants.

Table A2: Predictive Model Description

	Traditional			Relating to Payment History: Aggregate				Relating to Payment History: Granular			
					Models Combined with PHA				Models Combined with PHG		
	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Mod (1) + Mod (4) (5)	Mod (2) + Mod (4) (6)	Mod (3) + Mod (4) (7)	Granular (PHG) (8)	Mod (1) + Mod (8) (9)	Mod (2) + Mod (8) (10)	Mod (3) + Mod (8) (11)
Sales growth				✓	✓	✓	✓	✓	✓	✓	✓
Avg daily # transact (log)				✓	✓	✓	✓	✓	✓	✓	✓
Avg transact size (log)				✓	✓	✓	✓	✓	✓	✓	✓
CV daily sales								✓	✓	✓	✓
CV transact size								✓	✓	✓	✓
District aggregate sales (log)								✓	✓	✓	✓
Growth in district sales								✓	✓	✓	✓
Median transact size								✓	✓	✓	✓
Aggregate sales (log)				✓	✓	✓	✓	✓	✓	✓	✓
Share of district sales								✓	✓	✓	✓
Change in share of district sales								✓	✓	✓	✓
Share of transact through Visa or Master								✓	✓	✓	✓
Owner age (Years)		✓	✓			✓	✓			✓	✓
Length of relationship w/ the lender (months)*		✓	✓			✓	✓			✓	✓
Has credit score (1= Yes)	✓	✓	✓		✓	✓	✓		✓	✓	✓
Length of credit history (Years)	✓	✓	✓		✓	✓	✓		✓	✓	✓
# previously closed loans	✓	✓	✓		✓	✓	✓		✓	✓	✓
# bureau enquiries	✓	✓	✓		✓	✓	✓		✓	✓	✓
# active loans	✓	✓	✓		✓	✓	✓		✓	✓	✓
Credit score	✓	✓	✓		✓	✓	✓		✓	✓	✓
Share closed loans colltrl	✓	✓	✓		✓	✓	✓		✓	✓	✓
Share closed loans non-perf	✓	✓	✓		✓	✓	✓		✓	✓	✓
Share non-perf in active loans	✓	✓	✓		✓	✓	✓		✓	✓	✓
District		✓	✓			✓	✓			✓	✓
State		✓	✓			✓	✓			✓	✓
Industry		✓	✓			✓	✓			✓	✓
Month of Loan Disbursal		✓	✓			✓	✓			✓	✓
Loan amount (log)			✓				✓				✓
Rate of interest (Annual percent)**			✓				✓				✓
Loan tenure (Months)			✓				✓				✓
Loan-sales ratio							✓				✓
Number of Variables	9	15@	17	4	13	19	22	12	21	27	30

* Variables used as a predictor only in fintech loan analysis.
** Variables used as a predictor only in bank loan analysis.
@ Number of variables in bank loans analysis is 14. It is 15 in fintech loan analysis.
All monetary values are in Rupees. Transactions refer to the electronic transactions processed by the payment fintech for the merchants. For description of variables see Table A1.

Table A3: Predictive Monitoring Model Description

This table outlines the components of the post-disbursal models at 90 days since disbursal (dsd), for both the payment history aggregate (PHA) and payment history granular (PHG) versions. Models for other monitoring periods are constructed in a similar manner. It also compares these with the comprehensive pre-disbursal models, noting that post-disbursal models augment these by incorporating payment variables from various post-disbursal windows. The pre-disbursal models are same as those detailed in columns (7) and (11) under the screening model description in Table A2.

	Pre-disbursal		Post-disbursal	
	Trad Hard & Soft Info + PHA	Trad Hard & Soft Info + PHG	90-dsd PHA	90-dsd PHG
	(1)	(2)	(3)	(4)
Sales growth	✓	✓	✓	✓
Avg daily # transact (log)	✓	✓	✓	✓
Avg transact size (log)	✓	✓	✓	✓
CV daily sales		✓		✓
CV transact size		✓		✓
District aggregate sales (log)		✓		✓
Growth in district sales		✓		✓
Median transact size		✓		✓
Aggregate sales (log)	✓	✓	✓	✓
Share of district sales		✓		✓
Change in share of district sales		✓		✓
Share of transact through Visa or Master		✓		✓
Owner age (Years)	✓	✓	✓	✓
Length of relationship w/ the lender (months)*	✓	✓	✓	✓
Has credit score (1= Yes)	✓	✓	✓	✓
Length of credit history (Years)	✓	✓	✓	✓
# previously closed loans	✓	✓	✓	✓
# bureau enquiries	✓	✓	✓	✓
# active loans	✓	✓	✓	✓
Credit score	✓	✓	✓	✓
Share closed loans colltrl	✓	✓	✓	✓
Share closed loans non-perf	✓	✓	✓	✓
Share non-perf in active loans	✓	✓	✓	✓
District	✓	✓	✓	✓
State	✓	✓	✓	✓
Industry	✓	✓	✓	✓
Month of Loan Disbursal	✓	✓	✓	✓
Loan amount (log)	✓	✓	✓	✓
Rate of interest (Annual percent)**	✓	✓	✓	✓
Loan tenure (Months)	✓	✓	✓	✓
Loan-sales ratio	✓	✓	✓	✓
Aggregate sales between 0 and 90 days post-disb (log)			✓	✓
Average transact size between 0 and 90 days post-disb (log)			✓	✓
Avg daily # transact between 0 and 90 days (log)			✓	✓
Sales growth 0 to 30 days post-disb over 30 days pre-disb			✓	✓
Sales growth 30 to 60 days post-disb over 0 to 30 days post-disb			✓	✓
Sales growth 60 to 90 days post-disb over 30 to 60 days post-disb			✓	✓
Relative change in avg transact size 0 to 90 days post-disb over pre-disb			✓	✓
CV daily sales between 0 and 90 days post-disb				✓
between 0 and 90 days post-disb				✓
CV transact size between 0 and 90 days post-disb				✓
Median transact size between 0 and 90 days post-disb				✓
District aggregate sales between 0 and 90 days post-disb (log)				✓
Growth in district sales 60 to 90 days post-disb over 30 to 60 days post-disb				✓
Change in share of district sales 60 to 90 days post-disb over 30 to 60 days post-disb				✓
Number of Variables	22	30	29	44

* Variables used as a predictor only in fintech loan analysis.
** Variables used as a predictor only in bank loan analysis.
All monetary values are in Rupees. Transactions refer to the electronic transactions processed by the payment fintech for the merchants. For description of variables see Table A1.

Table A4: Bank Loans: Summary Statistics on Borrower Payment, Demographic, and Loan Variables – by Loan Repayment Status

Summary statistics based on 11972 loans made by banks to the merchants using the payment services of the payment fintech. For detailed variable description see Table A1. All nominal monetary variables are denominated in INR. Mean difference test: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

	Mean		Mean difference
Variable	Performing (N = 10854)	Delinquent (N = 1118)	Perf − Delinquent
Payment Variables			
Sales growth	0.20	0.14	0.06
Avg daily # transact (log)	0.80	0.69	0.10***
Avg transact size (log)	7.47	7.57	-0.10**
CV daily sales	2.49	2.82	-0.33***
CV transact size	1.56	1.53	0.03
District aggregate sales (log)	20.04	20.07	-0.03
Growth in district sales	0.05	0.06	0.00
Median transact size	2764.85	3418.85	-653.99***
Aggregate sales (log)	11.26	10.71	0.55***
Share of district sales	0.00	0.00	0.00
Change in share of district sales	0.00	0.00	0.00
Share of transact through Visa or Master	0.87	0.89	-0.02***
Traditional Variables (Borrower information and Loan terms)			
Owner age (Years)	35.49	34.68	0.81***
Has credit score (1= Yes)	0.95	0.96	-0.01
Length of credit history (Years)	6.27	5.44	0.83***
# previously closed loans	6.39	5.23	1.16***
# bureau enquiries	2.31	3.86	-1.55***
# active loans	7.40	6.76	0.64***
Credit score	718.21	704.93	13.28***
Share closed loans colltrl	0.47	0.40	0.06***
Share closed loans non-perf	0.04	0.03	0.00
Share non-perf in active loans	0.02	0.02	0.01*
Loan amount (log)	11.19	11.65	-0.46***
Rate of interest (Annual percent)	19.19	20.45	-1.26***
Loan tenure (Months)	16.35	23.67	-7.32***

Table A5: Fintech Loans: Summary Statistics on Borrower Payment, Demographic, and Loan Variables – by Loan Repayment Status

Summary statistics based on 15325 loans made by payment fintech to the merchants using its payment services. For detailed variable description see Table A1. All nominal monetary variables are denominated in INR. Mean difference test: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

	Mean		Mean difference
Variable	Performing (N = 13444)	Delinquent (N = 1881)	Perf − Delinquent
Payment Variables			
Sales growth	0.38	0.59	-0.21***
Avg daily # transact (log)	1.01	0.88	0.14***
Avg transact size (log)	7.39	7.57	-0.18***
CV daily sales	1.94	2.38	-0.43***
CV transact size	1.54	1.61	-0.07***
District aggregate sales (log)	20.14	20.13	0.01
Growth in district sales	0.05	0.05	0.00
Median transact size	1970.27	3363.11	-1392.84***
Aggregate sales (log)	12.27	12.20	0.07***
Share of district sales	0.01	0.00	0.00
Change in share of district sales	0.00	0.00	0.00*
Share of transact through Visa or Master	0.86	0.88	-0.02***
Traditional Variables (Demographic, Bureau, and Loan terms)			
Owner age (Years)	36.50	35.02	1.48***
Length of relationship w/ the lender (months)	15.17	15.00	0.17
Has credit score (1= Yes)	0.90	0.92	-0.01*
Length of credit history (Years)	3.99	3.69	0.31***
# previously closed loans	3.80	3.78	0.03
# bureau enquiries	0.93	1.36	-0.43***
# active loans	2.68	3.01	-0.33***
Credit score	715.25	699.34	15.91***
Share closed loans colltrl	0.41	0.41	0.00
Share closed loans non-perf	0.10	0.13	-0.03***
Share non-perf in active loans	0.10	0.13	-0.03***
Loan amount (log)	10.14	10.41	-0.27***
Loan tenure (Days)	112.17	117.53	-5.36***

Table A6: Bank Loans: Out-of-Sample Predictive Performance with Aggregate Payment History

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models see Table A2.

		Traditional		Payment History	Models Combined with PHA		
Predicted var: Delinqency	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Mod (1) + Mod (4) (5)	Mod (2) + Mod (4) (6)	Mod (3) + Mod (4) (7)
Area Under the ROC Curve (AUC)	0.59	0.62	0.68	0.59	0.67	0.69	0.73
Average Precision	0.07	0.14	0.19	0.11	0.16	0.19	0.23
N. Obs. Train	9578	9578	9578	9578	9578	9578	9578
N. Obs. Test	2394	2394	2394	2394	2394	2394	2394
N. Predictors	9	14	17	4	13	18	22

Table A7: Fintech Loans: Out-of-Sample Predictive Performance with Aggregate Payment History

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models see Table A2.

		Traditional		Payment History	Models Combined with PHA		
Predicted var: Delinquency	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Mod (1) + Mod (4) (5)	Mod (2) + Mod (4) (6)	Mod (3) + Mod (4) (7)
Area Under the ROC Curve (AUC)	0.57	0.61	0.65	0.62	0.62	0.67	0.69
Average Precision	0.16	0.18	0.19	0.20	0.20	0.21	0.23
N. Obs. Train	12260	12260	12260	12260	12260	12260	12260
N. Obs. Test	3065	3065	3065	3065	3065	3065	3065
N. Predictors	9	15	17	4	13	19	22

Table A8: Fintech Loans: Out-of-Sample Predictive Performance with Granular Payment History

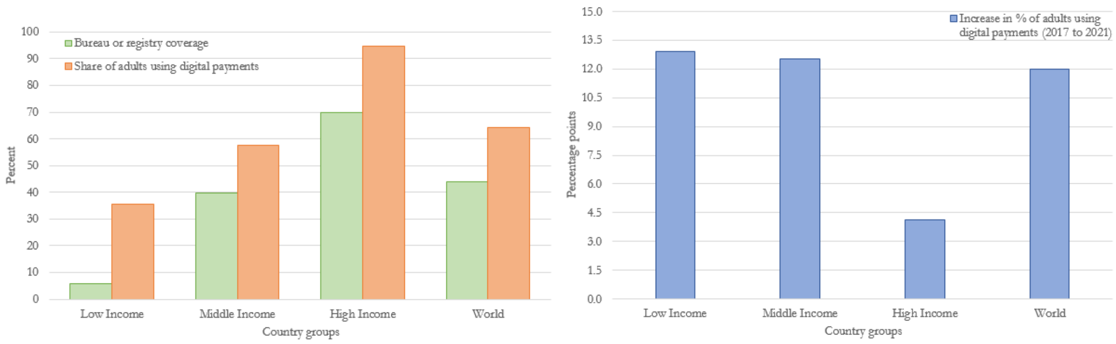
AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models see Table A2. Granular payment variables, as opposed to aggregate payment variables, necessitate transaction-level information or are calibrated against district-level payment aggregates.

		Traditional		Payment History		Models Combined with PHG		
Predicted var: Delinqency	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Granular (PHG) (5)	Mod (1) + Mod (5) (6)	Mod (2) + Mod (5) (7)	Mod (3) + Mod (5) (8)
Area Under the Curve (AUC)	0.57	0.61	0.65	0.62	0.65	0.65	0.68	0.70
% Δ compared to Agg model	-	-	-	-	3.80	3.94	2.05	1.64
Average Precision	0.16	0.18	0.19	0.20	0.21	0.21	0.23	0.24
% Δ compared to Agg model	-	-	-	-	3.61	4.81	6.80	3.79
N. Obs. Train	12260	12260	12260	12260	12260	12260	12260	12260
N. Obs. Test	3065	3065	3065	3065	3065	3065	3065	3065
N. Predictors	9	15	17	4	12	21	26	30

A.2 Additional Figures

Figure A1: Coverage Under Credit Bureau or Credit Registry and Use of Digital Payments

In Panel (a), coverage refers to number of firms and individuals covered either under a private credit bureau or a public credit registry, expressed as a percent of adult (15+) population. The number for a country group is derived in two steps. First, for each country, coverage is calculated as the maximum of the share of adults covered under a bureau, and the share of adults covered under a registry. Second, for a country group, coverage is the arithmetic mean of the coverages of the constituent countries obtained in the first step. The coverage statistics is for the year 2019 and is obtained from World Bank's World Development Indicators. Share of adults using digital payments refers to the percent of adults (15+) who used digital means of payments in the past 12 months. The data on digital payments is for the year 2021 and is obtained from the World Bank's Global Findex database. Panel (b) plots the *increase* in the share of adults using digital payments between the years 2017 and 2021, expressed in percentage points. Country groups are formed based on the income classification of the World Bank.



(a) Adults covered under credit bureau or registry, and adults using digital payments

(b) Increase in the share of adults using digital payments

Figure A2: Bank Loans: ROC Curves for Out-of-Sample Predictions Across Models

AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models see Table A2. Granular payment variables, as opposed to aggregate payment variables, necessitate transaction-level information or are calibrated against district-level payment aggregates.

(a) with Payment History: Aggregate (PHA) (b) with Payment History: Granular (PHG)

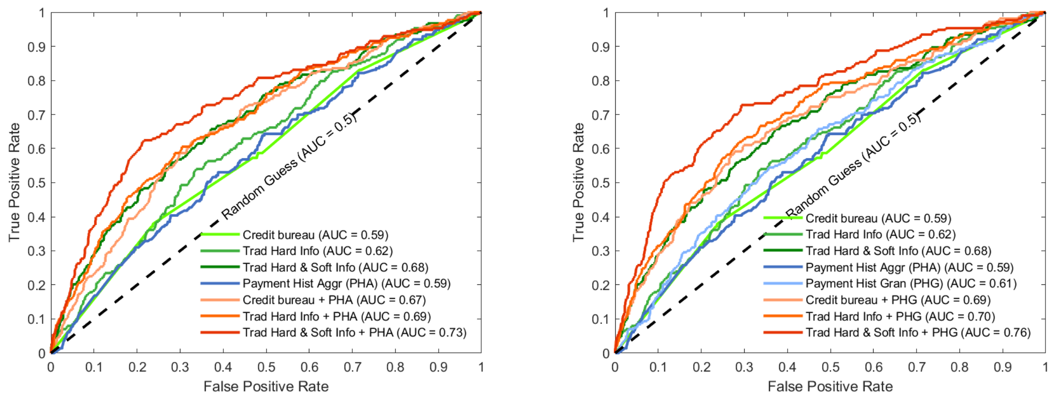


Figure A3: Bank Loans: Difference in Performance Metrics Across Models

The figure illustrates the differences in Areas under the ROC curve (AUCs) and Average average precisions (APs) between select pairs of models. For instance, the difference labeled *PHA vs. Credit Bureau* represents the AUC of the PHA model minus the AUC of the Credit Bureau model for the test sample, including the 95% confidence interval of this difference. These 95% confidence intervals are derived by bootstrapping AUC-differences and AP-differences across 10,000 replicas of the test set. For details on the composition of predictive models, refer to Table A2.

(a) Difference in AUC (b) Difference in AP

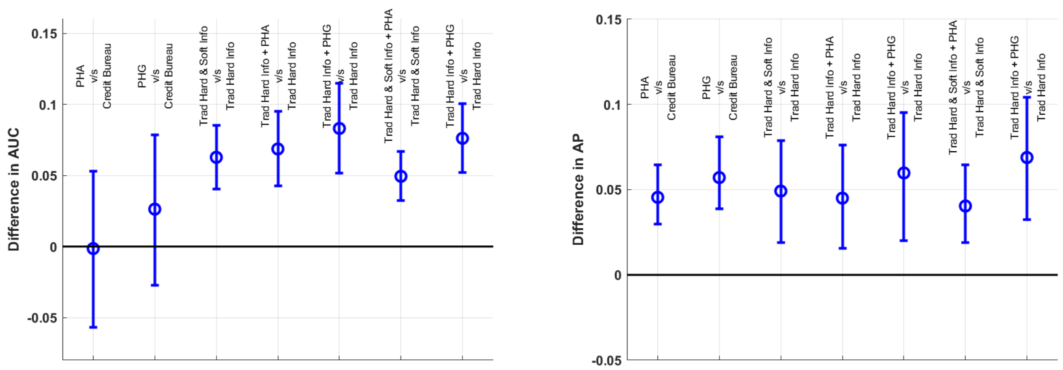


Figure A4: Bank Loans: Incremental Return on Investment from Payment-history Based Screening

The figure reports the incremental actual Return on Investment (ROI) for portfolios formed by selecting the top s% of loans based on expected returns, using the Probability of Delinquency (PD) from a model augmented with payment history, compared to the ROI from a portfolio of the top s% of loans calculated using PDs from the traditional hard information model. The figure reports these back-of-the-envelope calculations for various levels of s. The calculations assume that the lender charges the same interest as the original loan. Therefore, in these calculations the benefit to the lender accrues solely due to being able to select better quality loans by incorporating payment history.

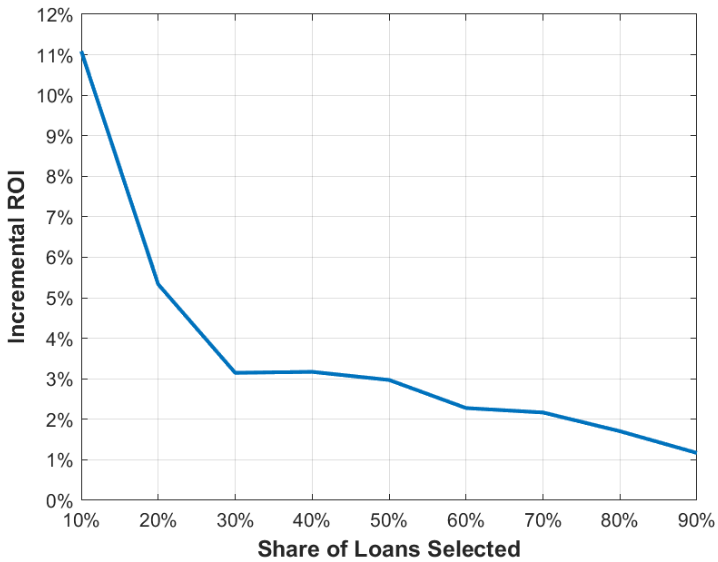


Figure A5: Bank Loans: Relative Changes in Probability of Delinquency: Aggregate vs. Granular Payment History

This figure presents the Empirical Cumulative Distribution Function (CDF) of the differences in the logarithm of probabilities of delinquency (PD), comparing the models *Trad Hard Info + PH* and *Trad Hard Info* for loans in our baseline test sample. Here, PH represents Payment History, which could be either Aggregate (PHA) or Granular (PHG). A negative difference indicates a borrower benefits from incorporating PH variables into the traditional hard information model, as it leads to a lower delinquency probability. Conversely, a positive difference suggests a borrower faces a higher delinquency risk with the addition of PH variables. Borrowing firms are classified based on their sales volume in the 90-day period prior to loan disbursement and their credit scores. *small borrowers* have sales below the median, whereas *large borrowers* have sales above it. High-scored firms are those whose owner have credit scores above 700 on a scale of 300 to 900. For details on the predictive models' composition, refer to Table A2. For information on the variables, see Table A1.

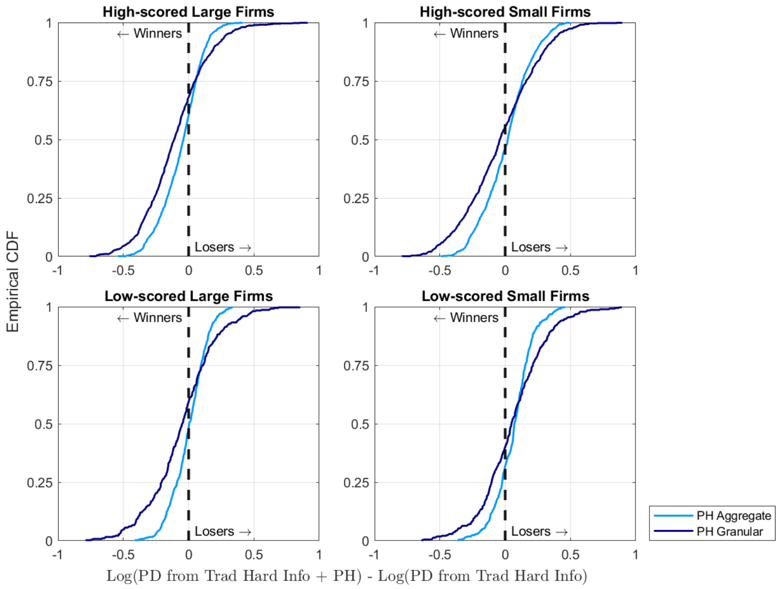


Figure A6: Fintech Loans: Variable Importance Based on Mean Absolute SHAP

The figure displays the mean of absolute SHAP values for each predictor, where the mean is calculated over all predictions within a 50% random sample of the test set. Absolute SHAP values quantify the degree to which each predictor influences a prediction's deviation from the mean outcome, indicating the predictor's impact. A higher mean absolute SHAP value signifies greater *overall* importance of the variable. The SHAP analysis is conducted using the *Traditional Hard & Soft Info + PHG* model, which encompasses 30 predictors, including 12 payment history-related variables. Of these, 4 are aggregative PH variables (aggregate sales, average per-day transaction count, average transaction size, and sales growth), and the remaining 8 pertain to granular PH. The analysis also considers 14 traditional variables, 3 contractual terms (loan amount, tenure, interest rate), and the loan-to-sales ratio, derived from the loan amount. The figure displays the top 15 variables out of the 30 analyzed. For detailed information on the variables and model composition, refer to Tables A1 and A2.

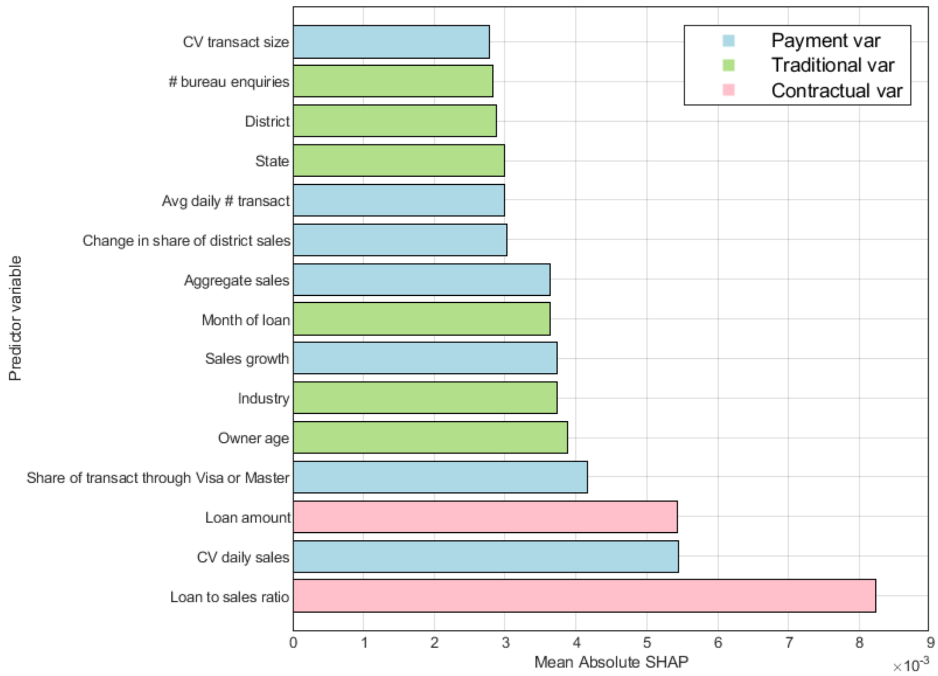
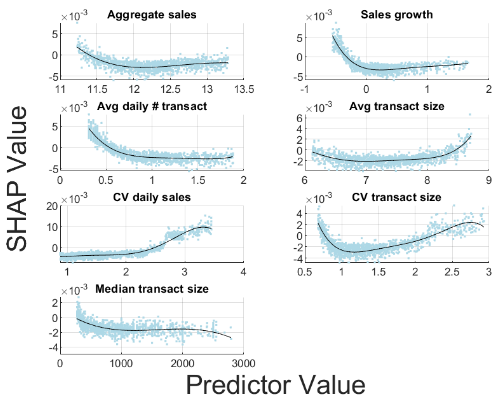


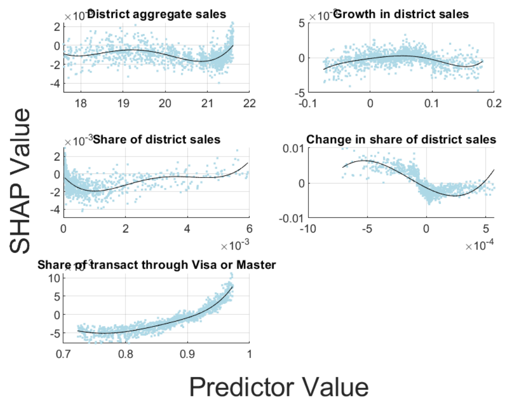
Figure A7: Fintech Loans: SHAP Dependence Plots

The figure illustrates the relationship between each predictor's value and its corresponding SHAP value across a 50% random sample of the test set predictions. SHAP values indicate each predictor's influence in shifting a prediction from the average. Higher absolute SHAP values signify a greater contribution to a particular prediction. Positive SHAP values increase the probability of delinquency, while negative values decrease it. The SHAP analysis is conducted using the *Traditional Hard & Soft Info + PHG* model, which encompasses 30 predictors, including 12 payment history-related variables. Of these, 4 are aggregative PH variables (aggregate sales, average per-day transaction count, average transaction size, and sales growth), and the remaining 8 pertain to granular PH. The analysis also considers 14 traditional variables, 3 contractual terms (loan amount, tenure, interest rate), and the loan-to-sales ratio, derived from the loan amount. Among the traditional variables, the figure plots the dependence of 8 numerical variables, excluding the 6 categorical ones. Each plot includes a polynomial fit of degree N, ranging from 1 to 5, with the optimal degree selected based on the highest adjusted-R^2 value. For a detailed breakdown of variables, see Table A1; for model specifics, refer to Table A2.

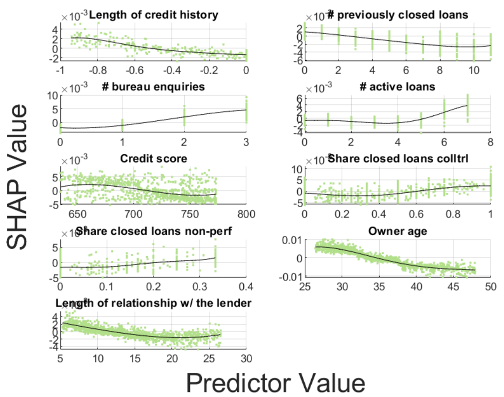
(a) Payment Variables Part-I



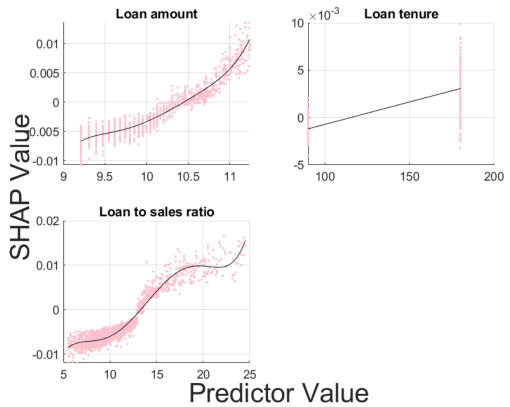
(b) Payment Variables Part-II



(c) Traditional Variables



(d) Loan Contractual Terms



B Out-of-bag (OOB) Variable Importance by Permutation for Bank and Fintech Loans

The out-of-bag (OOB) method leverages the fact that in the process of bagging, approximately 37% of observations are not used to train any given tree within an ensemble when sampling with replacement (Breiman, 2001). To estimate the importance of a variable, the method first calculates the prediction error on these OOB observations. It then shuffles the values of the variable across the OOB observations and measures how this permutation affects the error rate, using the same ensemble of trees. The increase in error rate, due to the permutation, indicates the importance of the variable. This process is repeated across all trees that include the variable. The significance of the variable is quantified by the average increase in prediction error, normalized against the standard error of these increases. A significant variable is one that, when shuffled, leads to a substantial increase in the prediction error, indicating its high importance in the model.

Figure B8 plots the OOB importance measures for the top 15 predictors, categorizing them into payment history variables, traditional variables, and combined variables. Notably, within the payment history category, the three most impactful variables—Aggregate sales, Average transaction size, and Average daily number of transactions—are aggregative, highlighting their strong contribution to prediction accuracy. Traditional variables also play a crucial role, with *Credit score* and standing out as a significant predictor. Additionally, district-level variables stand out among the granular payment history variables, emphasizing their relevance in the model.

Figure B9 provides an Out-of-Bag (OOB) importance measure for the top 15 predictors, this time segmented by the size of the borrowing businesses involved in the prediction exercise. It reveals that *Credit score* is a significant variable for large borrowers but not for small borrowers. In the case of large borrowers, payment variables claim eight of the top 15 positions, predominantly granular payment variables, with only two aggregative payment history variables appearing. Conversely, for small borrowers, three out of the four aggregative payment variables are among the top 15, with the most influential feature being an aggregative payment history variable.

Building on our screening analysis, we also evaluate the OOB permutation importance for the early warning model, assessing performance 90 days post-disbursal. Figure B10 showcases the 15 most impactful variables in this model. The lineup of top predictors mirrors closely those identified through the absolute SHAP values in Section 4.4, underscoring consistent findings across both measures.

In our analysis of fintech loan screening, OOB-based evaluations reveal a dominant performance by payment variables over traditional ones, as detailed in Figure B11. This dominance aligns with our findings from the SHAP-based importance assessment for fintech loans.

Figure B8: Bank Loans: Top Predictors of Delinquency Based on OOB Permutation

Variable importance determined using out-of-bag permutation, where higher values indicate greater importance due to increased prediction error after variable permutation. Importance assessed for Traditional Hard & Soft Info + PHG model, comprising 30 predictors: 12 payment history-related, 14 traditional, 3 contractual loan terms (Loan amount, tenure, interest rate), and 1 combining both (loan-sales ratio). See Table A1 for variable details and Table A2 for model composition.

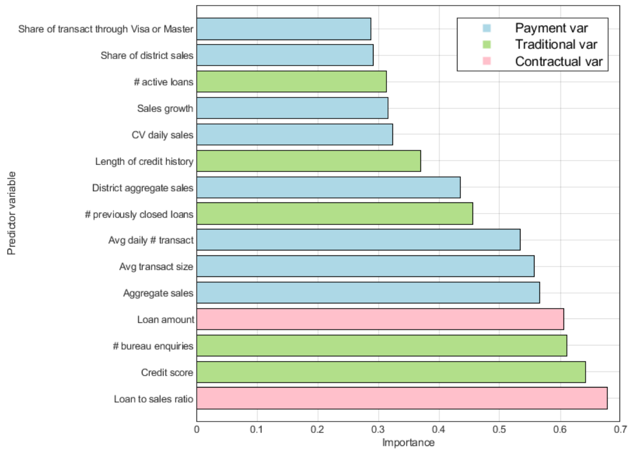
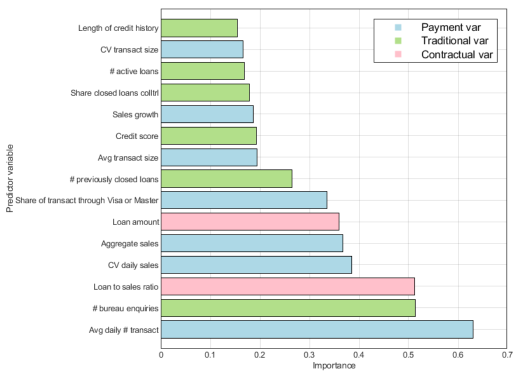


Figure B9: Bank Loans: Top Predictors of Delinquency Based on OOB Permutation – by Size

Variable importance determined using out-of-bag permutation, where higher values indicate greater importance due to increased prediction error after variable permutation. Importance assessed for Traditional Hard & Soft Info + PHG model, comprising 30 predictors: 12 payment history-related, 14 traditional, 3 contractual loan terms (Loan amount, tenure, interest rate), and 1 combining both (loan-sales ratio). See Table A1 for variable details and Table A2 for model composition. Small borrowers are defined by sales below the median in the 90-day pre-disbursal period; large borrowers exceed this median.

(a) Small Borrowing Merchants



(b) Large Borrowing Merchants

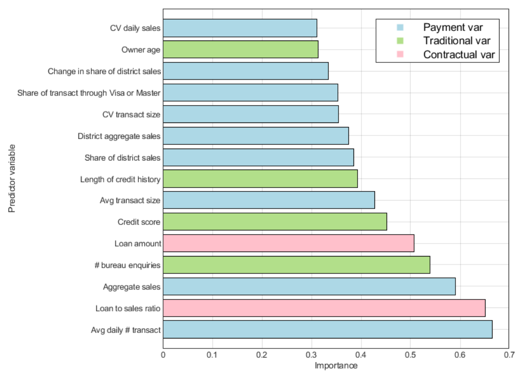


Figure B10: Bank Loans: Top 15 Predictors of Delinquency in Post-Disbursal Predictive Models

Variable importance determined using out-of-bag permutation, where higher values indicate greater importance due to increased prediction error after variable permutation. Importance assessed for Early warning PHG model at 90 days-since-disbursal. This model appends the variables in Traditional Hard & Soft Info + PHG model with the post-disbursal PHG variables calculated up to the 90 days since disbursal. See Table A1 for variable details and Table A2 for model composition.

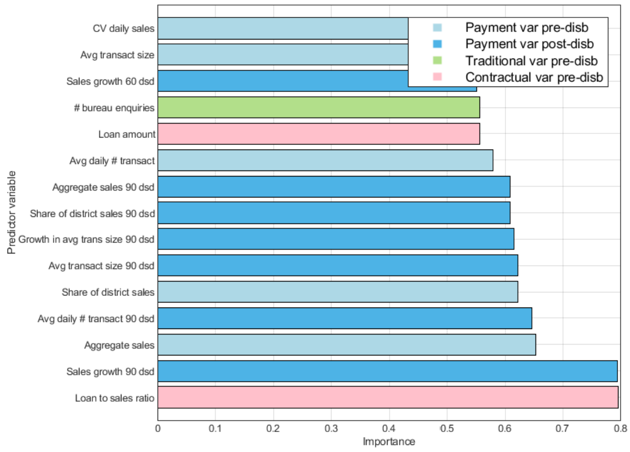


Figure B11: Fintech Loans: Top 15 Predictors of Delinquency

Variable importance determined using out-of-bag permutation, where higher values indicate greater importance due to increased prediction error after variable permutation. Importance assessed for Traditional Hard & Soft Info + PHG model, comprising 30 predictors: 12 payment history-related, 14 traditional, 3 contractual loan terms (Loan amount, tenure, interest rate), and 1 combining both (loan-sales ratio). See Table A1 for variable details and Table A2 for model composition.