Beyond the Bureau: Loan Screening and Monitoring under Open Banking*

Kumar Rishabh†

Click here for the latest version

Abstract

I investigate the informational value of interoperable payment data in lending, integral to global 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. In analyzing traditional bank loans, I find that payment history complements credit bureau data in predicting loan delinquency. Quantitatively, the informational value of aggregate payment data equates to the value of lender's soft information. In a counterfactual scenario where traditional lenders incorporate payment history alongside their existing hard and soft information, substantial benefits are realized. However, while about 29% of the enhancement from adding payment history can be attributed to the hardening of soft information, the predominant value stems from its independent contribution. After loan disbursal, payment data markedly enhances delinquency predictions, affirming its role in generating timely early warning signals for monitoring loans. While there is a trade-off between accuracy and privacy in screening, this is less pronounced in monitoring. In the fintech lending with sales-linked loans, payment history emerges as a substitute for traditional credit bureau data, albeit with pronounced moral hazard challenges.

JEL Classification: G20, G21, G23

Keywords: Fintech, Payment, Machine learning, Big data, Prediction, Open banking

---

*I thank Gustavo Cortes (discussant), Hans Degryse, Andreas Fuster, Zhiguo He (discussant), Jing Huang (discussant), Artashes Karapetyan (discussant), Yvan Lengwiler, David Martinez-Miera (discussant), Larissa Schäfer (discussant), Amit Seru (discussant), Neharika Sobti (discussant), Boris Vallée (discussant), 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), 2022 University of Florida Conference on ML in Finance, 2022 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: Feb 2024.

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

1 Introduction

Open Banking has swiftly taken center stage in the global financial landscape. By October 2021, about 48% of countries, totalling 80 out of 168, had embarked on Open Banking initiatives (Babina et al., 2024). Open Banking envisions a new era where financial products are built upon interoperable payment data. Foremost among its promises, as identified in industry surveys, is enhancing risk assessment in lending (Experian, 2022).

The surge in interest in Open Banking as a lending technology is understandable, particularly given the significant drawbacks of the currently dominant technology—bureau-based credit scoring. The insufficient coverage of credit bureaus, leaving over half of the world's firms and individuals unrepresented, is a glaring issue. This is starkly illustrated in Figure A1 in the appendix, especially for developing countries. The limitations of credit bureaus aren't just about reach; their backward-looking approach is equally limiting. 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. These deficiencies have resulted in millions of small businesses and individuals, including those in advanced economies, being poorly served by traditional lenders(TransUnion, 2022).

Payment histories, emerging from widespread electronic payment transactions, sharply contrast with the backward-looking design of credit bureaus by offering a real-time and frequently updated perspective. These histories could potentially bridge the information gap, enhance credit pricing, and broaden financial access. Furthermore, the detailed and immediate nature of payment data makes them potential tools for providing early warning signals to lenders in post-lending loan monitoring. These presumed benefits of the payment histories underpin the strong global policy support for Open Banking initiatives (BIS and World Bank, 2020; Plaitakis and Staschen, 2020; Babina et al., 2024).

At the core, these initiatives are based on the idea that interoperability of payment data will enable new market entrants to innovate in credit products, utilizing borrower payment histories from incumbents for risk underwriting. This presupposes that data from one source can significantly enhance another institution's understanding of a borrower's financial health. However, research that conclusively supports or refutes these assumptions is notably lacking. With Open Banking's growing significance, a thorough evaluation of the value of interoperable payment data is imperative. The primary aim of this paper is to undertake this essential inquiry.

Specifically, my paper seeks to answer several key questions: Can payment histories enrich a lender's blend of soft and hard information? Or they only result in hardening of the soft information? How do they interact with credit bureau data—as complements or as substitutes? Are payment histories effective in both pre-loan screening and post-disbursal monitoring, particularly in providing early warning signals? What is the optimal balance between the granularity of shared data versus privacy and technological costs? My exploration encompasses studying role of payment histories for a diverse array of borrowers, including small borrowers, those lacking credit history, and larger firms.

Addressing these questions poses a significant challenge, primarily due to the rarity of scenarios where lending and payment data originate from separate sources—a set-up that is critical to understanding data interoperability.[1]

This complexity is compounded by what I refer to as the BFP critique, after the reservations raised by Berg, Fuster and Puri (2022) against the existing studies assessing the effectiveness of *alternative data* in lending. The challenge stems from the fact that alternative data is primarily utilized by fintechs and bigtechs, not traditional banks. BFP critique raises the issue that these different lender types often serve distinctly different borrower bases, suggesting that conclusions about the utility of alternative data in one lending context may not extend broadly. More importantly, even when borrower samples are harmonized, significant differences in the nature of *lending contracts* between traditional banks and alternative-data-reliant lenders can confound the estimated value of alternative data. This is because loan repayment behaviors are influenced not only by the borrower's risk profile prior to lending but also by the loan contract terms themselves. Therefore, different types of lending contracts create varied informational environments, complicating the generalization of the value of alternative data beyond alternative lending models. To accurately gauge the value of alternative data in traditional lending, it is essential to study its impact within the standard debt contracts issued by traditional banks.

The core of my research design addresses the interoperability challenge and navigates 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 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.

As the discourse around Open Banking 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). Notable examples of payment-data-based lending programs include those by E-commerce giants like Amazon, Mercado Libre, and Ant Financial, as well as by payment fintechs such as Paypal, Square, and Stripe. Another common characteristic of many of these lending programs is their sales-linked repayment structure, where the lending entity—be it a bigtech or fintech—receives a portion of the sales they process for the borrowing firm that utilizes their e-commerce or payment platform (Rishabh and Schäublin, 2021; Russel, Shi and Clarke, 2023; Liu, Lu and Xiong, 2022). This raises a critical question: what is the value of payment data in these sales-linked lending models?

Building on this context, my study extends to the valuation of data in such sales-linked lending. I am able to analyse sales-linked loans because my payment fintech collaborator not

[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.

only processes electronic payments for small businesses but also offers them sales-linked loans. It's important to clarify that the primary focus for addressing the interoperability question lies in examining *traditional bank loans* obtained by these fintech clients. The analysis of fintech loans is an exploration reserved for a subsequent section of the paper.

To address our questions on data interoperability, I assess the predictive power of different sets of variables, referred to as *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 first screening model I examine is the Credit Bureau Model, which relies on the borrower's credit score and related credit bureau data, including inquiry counts before the loan and past loan performance. The Traditional Model is developed in two variants: the first integrates credit bureau data with several borrower characteristics—including age, location, and industry—embodying what we typically identify as traditional *hard information*.

The second variant of the Traditional Model *expands* this approach by incorporating loan contract terms—amount, tenure, interest rates—into the hard information. Thus, the second traditional model combines *hard and soft information*, as loan contractual terms draw on a comprehensive range of sources, covering both observable and unobservable (soft) information. It is important to note that the traditional model with both hard and soft information, while analytically valuable, is not directly applicable for practical loan screening due to the inclusion of contractual variables. These variables, reflect the lender's choices based on a mix of hard and soft information. Nevertheless, this model is constructed to serve as a benchmark under various scenarios. In contrast, traditional model with hard information has practical applicability because it is built on the information available to the lender before they set the loan terms.

With the Credit Bureau and Traditional models set as our foundational benchmarks, we now turn to constructing models with payment history, that will eventually help us in pinning down the value of interoperable payment data. To this end, I introduce two distinct Payment History (PH) models: The Aggregate (PHA) model, which compiles broad payment indicators such as total sales, sales growth, and average transaction size from the 90 days preceding loan issuance, along with the average daily transactions metric calculated for the 30 days immediately before loan issuance, capturing recent financial activities. The Granular (PHG) model goes a step further by incorporating detailed transaction-level data alongside comparisons with district-level payment averages. To complete our suite of models, combined models, *Traditional Hard & Soft Information + PHA* (and, by extension, PHG), emerge as the most extensive screening models within their respective categories.

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). An AUC of one signifies perfect predictive performance, whereas an AUC of 0.5 implies a predictive accuracy no better than a random guess. Additionally, I calculate out-of-sample

Average Precision (AP) as a complementary measure. In simpler terms, AP reflects the likelihood that a delinquent loan is correctly identified as such across various decision thresholds.

The final step in valuing the interoperable data is to compare the predictive performance across various models. For instance, evaluating the PHA model against the Credit Bureau model sheds light on how aggregative payment data stack up against the established lending technology. This comparison, especially between the combined PHA and Credit Bureau model versus each individually, aims to discern whether these data sources provide overlapping or distinct insights, assessing their roles as substitutes or complements.

Further, the integration of PHA with the *Trad Hard Info* and *Trad Hard & Soft Info* models underscores the specific value of interoperable payment data to loan screening. Enhanced performance from adding PHA to the *Trad Hard Info* model, without similar gains when added to the *Trad Hard & Soft Info* model, would suggest PHA mainly hardens lender's soft information rather than introducing new insights. Conversely, significant performance gains from PHA's addition to *Trad Hard & Soft Info* would establish it as a distinct source of information. This would indicate that in a counterfactual scenario, banks employing PHA alongside traditional hard and soft information would see tangible benefits.

Our results relating to the screening exercise are summarized below:

i. The PHA model is on par with or superior to the Credit Bureau model in predicting bank loan delinquency, and their combination further enhances predictability, indicating their complementary relationship under bank lending.

ii. The improvement from integrating PHA into the *Trad Hard Info* model mirrors that achieved by incorporating loan terms, suggesting PHA's value is akin to lender soft information.

iii. Adding PHA to the *Trad Hard & Soft Info* model significantly boosts accuracy, with a 5 pp increase in AUC (6% relatively) and 4 pp in AP (20% relatively). However, this increase does not fully represent hardening of soft information. I estimate about 29% of the contribution of aggregate payment information is attribututed to hardening of soft information and remaining 71% to independent information.

iv. PHA particularly benefits small borrowers, surpassing its impact on larger counterparts, and enhances predictability across all borrower types. It proves especially potent for thin-file borrowers, indicating a 7% increase in AUC and 26% in AP, affirming its role as an effective screening tool for those with scant credit history.

Open Banking designers encounter three intertwined challenges: safeguarding customer data privacy, navigating diverse API[2] standards, and managing the technological complexities of processing granular data. The balance between privacy concerns and regulatory mandates restricts the scope and variety of shared data. Variability in API protocols complicates this data

[2]An API (Application Programming Interface) is a set of rules that allows different software programs to communicate with each other. In open banking, it enables banks and financial applications to share data securely and seamlessly, facilitating transactions and access to financial information across different platforms.

sharing, introducing inconsistencies (BIS, 2019*b*; Sia Partners, 2019). Moreover, intricacies of processing granular create challenges as many financial institutions report significant technological and financial burdens in establishing compliant data sharing frameworks, primarily because their data are stored on multiple, often unconnected, information technology systems, many of which are not currently interconnected with their core banking system (CFPB, 2023).

A pragmatic approach may involve favoring *aggregated data sharing*, which simplifies processes and bolsters security and privacy adherence. However, this choice can diminish the quality of risk signals from the interoperable payment data. To examine this trade-off, I compare the PHA and PHG models, discovering that integrating the PHG variables in the traditional model boosts predictive accuracy by approximately 4% in AUC and 12% in AP compared to the aggregate model. This suggests that prioritizing privacy and technological feasibility may compromise the optimal utilization of payment data. Nonetheless, it's important to recognize that the effectiveness of PHA models alone, with their substantial predictive power, already presents a compelling argument for their application in loan screening.

Open Banking holds significant, yet often overlooked, potential for loan monitoring. Traditional credit scores can be slow to reflect changes in a borrower's financial situation, typically updating only after delays exceeding 90 days and reliant on the reporting practices of other lenders. In contrast, payment history data offers real-time, independent insights into a borrower's financial health, making it a more immediate and accurate early warning indicator.

I explore how interoperable payment data can enhance monitoring by adding post-disbursal payment variables to the extensive pre-disbursal screening model and updating predictions at thirty-day intervals post-loan issuance. I discover that payment data significantly enhances monitoring capabilities, with post-disbursal payment information contributing as much to AUC within 120 days (and to AP within 90 days) post-disbursal as pre-disbursal PHA variables do in the screening phase. This 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 likelihood of delinquency for loans destined to default sees an approximate 5 percentage point rise within 180 days after disbursal.

I also find that while initially granular (PHG) data exhibit a distinct advantage in monitoring, this benefit is temporary. As loan progresses, the effectiveness of aggregate (PHA) models converges with that of granular models, illustrating the diminishing privacy-accuracy trade-off in monitoring. This evolution also underscores that granularity-accuracy trade-off is relatively more pronounced in screening than in monitoring.

By applying interpretable machine learning techniques, I discover that payment history variables, notably aggregative ones, are significant in predicting loan delinquency. These variables not only enhance screening accuracy but also play a crucial role in monitoring, with sales growth emerging as a pivotal early warning indicator. I find a deterioration in sales growth post-disbursal to be directly linked to higher delinquency probabilities.

Shifting focus to sales-linked fintech loans, the landscape changes. The PHA model not only outperforms the Credit Bureau model in predicting loan delinquency but also shows that merging these data sources doesn't improve predictive power. This implies credit bureau data might be redundant in sales-linked lending contexts. Fintech loans, interestingly, depend less on the lender's soft information than traditional loans do. Post-disbursal, PHA variables significantly improve predictive performance, evidenced by a notable rise in AUC shortly after loan issuance in fintech lending. This quick uptick, however, may mirror the moral hazard challenges unique to sales-linked loans, as highlighted in my previous work (Rishabh and Schäublin, 2021) and more recently by Russel, Shi and Clarke (2023). Thus, my findings underscore a critical tension in fintech lending: while the dependence on traditional data sources like credit bureaus diminishes, the rise of moral hazard poses new risks.

Literature and contribution: My research intersects with two pivotal strands of literature: the use of alternative data lending, and the interoperability of payment data.

The initial wave of research concentrating on use of alternative data in *consumer lending* has seen diverse applications: from analyzing individuals' online behaviors (Berg et al., 2020) to mobile phone usage (Agarwal et al., 2024), grocery shopping patterns (Lee, Yang and Anderson, 2023), or simply 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' utilization of e-commerce transactions for credit assessments. Notably, 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 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, enabling a critical assessment of how *traditional* lenders might benefit from incorporating alternative data (payment history) into their risk assessment. Hence, my analysis extends the relevance of alternative data from niche markets to the broader traditional lending landscape that relies on standard debt contracts. Secondly, while these studies predominantly focus on the *screening* capabilities of alternative data, my work explores both its *screening and monitoring* potential, offering a more comprehensive understanding of its value. Third, by analyzing experiences of the *same borrowers* with both bank and sales-linked fintech loans, my study sheds light on the relative importance of different information sources, including lender soft information and payment history, across diverse loan contractual contexts.[3]

[3]A wide array of research explores fintech and bigtech credit beyond the alternative data paradigm, focusing on the influences of intermediation costs, regulation, and convenience in the rise of fintech and bigtech credit (Philippon, 2016; Buchak et al., 2018; Fuster et al., 2019; Liu, Lu and Xiong, 2022); their interplay with collateralized lending (Gambacorta et al., 2022; Beaumont, Tang and Vansteenberghe, 2023); their role as an alternative to traditional bank lending (Gopal and Schnabl, 2022; Tang, 2019; Eça et al., 2022); their distributional impact (Fuster et al., 2022), their impact on financial inclusion (Ouyang, 2021); and the issues of moral hazard they present (Rishabh and Schäublin, 2021; Russel, Shi and Clarke, 2023). For a thorough review, see Berg, Fuster and Puri (2022).

My work contributes to the evolving discourse on interoperable 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 the pricing of payment services and the broader credit market's architecture. Concurrent empirical research, such as Ghosh, Vallee and Zeng (2024)'s study in the Indian small business lending context, finds that borrowers who engage in cashless transactions are more likely to receive a loan and at more favorable conditions, evidenced by lower interest rates and higher loan amounts. This pattern is corroborated by Babina et al. (2024) in the UK's small business lending arena and by Nam (2023) in the consumer lending market of Germany, *indirectly* signaling the value of interoperable payment data for lenders and borrowers.

Building on this foundation, my work offers a *direct* assessment of the value of the 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 cashless transaction data at my disposal 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 *in-house* payment data in loan underwriting (Black, 1975; Fama, 1985; Nakamura, 1993), sets a vital 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, studies by 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 ventures into the realm of *interoperable digital payment* data, and using its detailed granularity to identify the specific payment history attributes crucial for effective loan screening and monitoring. This approach significantly expands the scope of the checking account hypothesis in contexts where payment data originates outside of the lending institution.

Paper structure: The organization of this paper is as follows: Section 2 details the institutional background and outlines the data structure. Section 3 describes the data models and the prediction algorithm employed. Results are presented in Section 4, starting with the assessment of payment data's predictive power in loan screening, exploring borrower heterogeneity, the trade-off between granularity and accuracy, and the significance of different data features. This section also explores the utility of payment data in loan monitoring, extending the analysis to highlight its effectiveness. Section 5 shifts focus to the application of these insights within the context of sales-linked loans provided by fintech companies. The paper concludes in Section 6, summarizing key findings and implications.

2 Institutional Set-up and Data

My collaboration with a leading Indian payment fintech, a key player in the electronic payment sector, forms the basis of this study. This fintech provides Point of Sale (POS) systems primarily to Micro, Small, and Medium Enterprises, hereafter referred to as *merchants*. These merchants utilize the fintech's POS devices to process various electronic payments. The analysis leverages transaction data at the swipe level from *all* merchants using the fintech's services, covering the period from January 2015 to February 2019.

Additionally, I have access to borrowing records of a subset of these merchants. This subset represents clients of the payment fintech 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 is pertinent to mention that linking traditional bank loans with the borrowing merchant's transaction history from the payment fintech forms the basis for studying the *interoperable* payments data.

2.1 Bank Loans

The dataset on bank loans is derived from the credit records of borrowing merchants, obtained from TransUnion CIBIL, a leading credit bureau in India. These records, compiled from financial institution reports, primarily focus on loans granted to small business owners. The exact identities of the lending institutions remain undisclosed; however, they encompass both commercial banks and NBFCs[4]. For the purpose of simplicity in this study, I collectively refer to these entities as 'banks', acknowledging their shared use of traditional standard debt contracts. These contracts contrasts with the sales-linked loans offered by the payment fintech.

A notable characteristic of small business lending is the often blurred 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). Therefore, in this study, all loans to business owners, irrespective of being labeled as 'personal' or 'business' by the lenders, are treated as business loans due to their interchangeable nature. Excluded from this categorization are distinctly non-fungible loans such as mortgages or vehicle loans. Additionally, gold loans, commonly used among Indian MSMEs as a financing method and secured against gold assets, are also classified as business loans (Asokan, 2020; Singh and Wasdani, 2016).

Bank loan records from the credit bureau also include a comprehensive monthly repayment history for each loan, compiled as of August 2020. These records cover up to 36 months, or conclude with the loan's closure if it occurs within the 36 months. Given that the most

[4]NBFCs are financial institutions without a deposit franchise, except for a few permitted to accept *non-demandable* deposits prior to 1997. Since then, the Reserve Bank of India has not granted deposit franchises to new NBFCs. NBFCs also remain outside the payment and settlement system, and are regulated by the Reserve Bank of India.

recent loan in our study was issued in February 2019, we have access to at least 18 months of repayment data for every loan. This extensive history is crucial for identifying instances of delinquency and their timing. I define a loan as *delinquent* if it exhibits any of the following: a repayment delay of 90 days or more, a write-off, or a classification by the lender under categories reflecting loss, such as *Loss*, *Substandard*, *Doubtful*, or *Special Mention Account*.

These records also include essential information such as the disbursement and closure dates of the loans, their types (as previously discussed), and key contractual terms like loan amounts, interest rates, and loan tenure. For each borrower, I compile a detailed electronic payment history by merging their loan information with transaction data from the payment fintech. More information on the payment data is provided in Section 2.4. Additionally, I describe the credit score and credit enquiries data obtained from the credit bureau in Section 2.3.

This study analyzes 11,972 bank loans issued from June 2015 to February 2019, within the constraints of the available payment data. To thoroughly examine the borrowers' credit histories, I analyze the performance of 130,101 loans (encompassing all types, including credit cards) they received, dating back to 1991. This historical review allows for the calculation of crucial variables mirroring the borrowers' past borrowing behaviors at the time they obtained a new loan within our study period. This method approximates the lenders' perspective, using credit bureau data to simulate the information available during the loan approval process. Key variables derived include the count of loan and credit card accounts a borrower had closed prior to the new loan, the number of active loans at loan approval, among others. For an exhaustive description of these variables, refer to Table A1.

2.2 Fintech Loans

To examine the sales-linked loans provided by the payment fintech, I accessed its loan book as of the end of February 2019, with a subsequent update in December 2019. Notably, all these loans were unsecured and had a uniform interest rate of two percent per month. This rate aligns with the typical charges imposed by NBFCs on high-risk borrowers in India and falls within the interest rate spectrum observed in the consumer credit markets of the US and the UK (Cornelli et al., 2020).

The loan repayment terms with the payment fintech were directly tied to sales, where 'sales' means the digital transactions processed by the fintech for the merchant. For loan amortization, the fintech deducted 10% from each transaction processed for the borrowing merchant, transferring the remaining balance (after any applicable charges) to the merchant. This unique repayment method meant the loans lacked a pre-defined tenure. However, the fintech typically suggested a repayment period of either three or six months. Surpassing the suggested tenure of the loan did not result in late penalties; however, borrowers were required to pay interest for the actual duration the loan was held.

Given this context, I introduce the concept of *implied tenure*–the number of days it would take for the borrower to repay the loan (principal + interest), assuming their sales continue at the same average daily level as the *pre-disbursal long-term average* with a 10% deduction rate. I define long-term average sales as the per-day average calculated over the 90-day window consisting of sales in 30 days to 119 days *before* disbursal.[5] Additionally, merchants had the flexibility to repay the loan early, either in full or partially, through direct lump-sum payments to the company.[6]

To define delinquency for fintech loans, I adopt a snapshot view of loan performance as of 31 December 2019—ten months following the disbursal of the last loan included in our analysis. A loan is categorized as delinquent if, (i) it ran beyond its implied tenure and, (ii) as of the snapshot date, it had a "large" shortfall in repayment. I deem a shortfall as large when it exceeds five percent of the total due repayment amount as of 31 December 2019. A minor segment of these delinquent loans was written off by the lender, particularly in cases where the merchant had exited the payment company's network.

The fintech-loan dataset consists of 15,325 sales-linked loans disbursed from May 2017 to February 2019. This dataset encompasses key information like the amount of each loan, its suggested repayment period, and the dates of disbursal and closure. It also includes the remaining balance, if any, as of December 2019. By leveraging credit bureau records, I calculate variables related to past borrowing, similar to the approach for bank loans. Additionally, payment history variables are derived using the payment transaction data.

2.3 Other Credit Bureau and Demographic Data

For both bank and fintech loans, the credit bureau provides additional data beyond the previously mentioned credit records of merchants. This supplementary information encompasses credit enquiries and credit scores. The credit enquiries represent each instance when a financial institution approached the bureau for information about a merchant. Numbering a total of 346,079, these enquiries indicate a merchant's pursuit of or interest in securing a loan. A high volume of enquiries often signals an urgent financing need from the merchant's side. While the dates of these enquiries are recorded, the identities of the inquiring financial institutions are kept confidential.

The bureau allocates credit scores to borrowers on a scale ranging from 300 to 900, where higher scores signify greater creditworthiness. Borrowers lacking adequate history for a score are classified under *unscored loans*. In the lending market, a credit score above 700 is generally regarded as good, and I use this benchmark to differentiate *high-score* borrowers from *low-score* ones. Notably, the fintech lender in this study did not utilize credit scores, or any other bureau

[5] I do not include the days close to the disbursal date in average sales calculations because some short-term, unusually high sales days that increase the probability of getting a loan might overstate the actual health of the borrowers.

[6] Many of these loan policies are similar to those adopted by US-based payment fintechs such as PayPal and Square. For more details, see Rishabh and Schäublin (2021).

data, for their lending decisions. This practice is consistent with the approach of many payment fintechs, such as the well-known US-based PayPal and Square, which also do not factor in credit scores in their lending processes.[7] Interestingly, Mishra, Prabhala and Rajan (2022) note that even traditional banks in India were initially slow to adopt credit scores in their lending decisions, thereby potentially overlooking valuable information.

Additionally, I acquire demographic information about the borrowing merchants, sourced either from the credit bureau or directly from the payment fintech. This data facilitates the calculation of the owner's age, and the duration of their relationship with the fintech lender (a metric utilized solely in the analysis of fintech loans). It also includes the industry sector, as well as the district and state for each merchant.

2.4 Payment History Data

I refer to the information derived from merchant payment transactions as 'payment history'. I have constructed these histories using a comprehensive dataset of 99.4 million transactions, each recorded at the card-swipe level. This data comes from electronic payments processed through the payment fintech's POS devices, offering a detailed view of the transactions conducted between merchants and their customers. However, it's important to note that this dataset does not encompass the entirety of merchant transactions. It specifically lacks data on inflows in the form of cash and other types of outflows.

The anonymized transaction data, covering the period from January 2015 to February 2019, includes activities from about 270,000 merchants. This group comprises both those who have taken loans (borrowers) and those who haven't (non-borrowers), representing all users of the fintech's POS systems. Each transaction in this dataset is detailed, containing information such as the amount, date, anonymized card number, and the card type, which includes major providers like Amex, Visa, and Mastercard. The extensive nature of this dataset facilitates the creation of district-level benchmarks using data from non-borrowing merchants. A more detailed discussion on this methodology will be provided below.

3 Predictive Models and Methodology

3.1 Predictive Models

In our approach, predictive *models* are central to our analysis. These models, each a unique combination of variables, are specifically designed to forecast delinquency. Their out-of-sample predictive performance is crucial, as it assesses their ability to predict future delinquency. By

[7]For more on the credit scoring policies of PayPal and Square, see `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).

comparing the predictive performances of different models, we delve into the heart of our research questions, which focus on the informational value of various types of variables.

With our research questions in mind, we have developed several models focusing on both screening and monitoring aspects. For screening-related questions, models utilize pre-disbursal variables, while post-disbursal variables form the basis of models aimed at monitoring and early warning. The starting point for the screening model is the *Credit bureau* model, that incorporates variables based on the credit bureau data like credit score, number of enquiries, and loan history.

Expanding from this, we delve into the *Traditional* models: the first, *Traditional Hard Information*, enriches the credit bureau data with demographic details of the borrowing firms, such as location, industry, and owner's age. The second, *Traditional Hard & Soft Information*, is the most comprehensive within this category. It extends the first traditional model by including contractual loan terms, such as the amount, tenure, and interest rate of the loan. This particular model is pivotal as it encapsulates not only the hard information but also the 'soft' information that lenders gather about borrowers. In small business lending, where financial records are often not fully accessible, lenders rely on the business owner's credit reports and soft insights from loan officers. The additional information in the traditional model with loan terms is therefore likely reflective of this nuanced, soft information gathered by the lender.

I develop two payment history (PH) models. The first, the *Payment History Aggregate (PHA)* model, captures an aggregate view of a merchant's electronic sales. This model includes four variables: total sales in the 90 days before disbursal, the growth in average per-day sales in the 30 days preceding disbursal compared to the 30-60 days prior, average transaction size in the 90-day window, and the number of transactions in the final 30 days before disbursal.

The *Payment History Granular (PHG)* model builds upon the PHA by incorporating transaction-level details and district-level sales benchmarks. Despite PHG's rich informational content, its use comes with considerations, including privacy concerns and technological challenges. These aspects introduce a trade-off, necessitating a balance between the potential for greater predictive accuracy and the costs associated with privacy and technology. By comparing the predictive performance of PHG and its augmented models with those of the PHA and its augmented counterparts, we can evaluate the magnitude of accuracy improvements when transitioning from aggregate to granular payment data.

To further our understanding, I integrate the PH models with the Credit Bureau and Traditional models. This integration aims to ascertain the extent of information gained by combining these variables. The evaluation is based on comparing the performance of the joint models against the stand-alone models. When integrating a PH model with the *Trad Hard & Soft Info* model, I also introduce a new variable: the loan-to-sales ratio. This ratio measures the loan's size compared to the sales in the 90-day period before disbursal. For a detailed account of the variables and models employed in this analysis, see Table A1 and Table A2.

To explore the utility of payment history in monitoring loans, I enhance the *Trad Hard*

& Soft Info + PHA (or PHG) pre-disbursal screening model by adding various transaction-based variables calculated at different days since disbursal (dsd). These post-disbursal models mirror the structure of the pre-disbursal PH variables. For example, the post-disbursal PHA 30 dsd model includes the PHA variables from the 30-day period following loan issuance. This encompasses total sales, average transaction size, daily transaction count, and the relative growth in average per-day sales and average transaction size compared to the 30 days before the loan was disbursed. Similarly, the PHG 30 dsd model expands the pre-disbursal *Trad Trad Hard & Soft Info + PHG* model with PHG variables from the 30-day post-disbursal period.

This approach extends to PH 60 DSD, PH 90 DSD, and up to PH 180 DSD models. Each model incorporates sales growth from all previous assessment points. For instance, the PHA 90 DSD model combines variables from (Trad Hard & Soft Info + PHA) and PHA variables from the 90-day post-disbursal period, along with sales growth calculated at both 30 and 60 days after disbursal. To provide an overview of our discussion, Table 1 presents a concise summary, mapping specific research questions to the corresponding models used for answering the 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
How significant is lender's private (soft) information in lending decisions?	Trad Hard & Soft Info vs. Trad Hard Info
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 is the value of payment history data 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 efficacy of payment history in early warning and loan monitoring?	(Trad + PH + PH dsd) vs. (Trad + PH)
What is the value of granular payment history data?	Models with PHG vs. Models with PHA

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.

3.2 Predictive Methodology

To predict loan delinquency, I split the sample into a training set and a test set. I then train a supervised machine learning algorithm on the training set, which includes the variables relevant to the selected model and the label identifying whether the loan was delinquent or not. After the training phase, I use the algorithm to predict delinquency on the test set. This approach upholds the *firewall principle* (Mullainathan and Spiess, 2017), which dictates that the training data should not influence the evaluation of the model's performance. I allocate 80% of the data to the training set and reserve the remaining 20% for out-of-sample predictions. Once the random partition takes place, I consistently apply the same training and test sets across all models, ensuring comparability of results.

In our main analysis, the supervised machine learning algorithm *Random Forest* is employed for a classification task (Breiman, 2001). Operating as an ensemble of multiple decision trees, Random Forest boosts model accuracy and robustness through a majority voting system for predictions. Each tree in the ensemble makes decisions by splitting at points called nodes.

At these nodes, the tree divides the data based on values from a randomly selected subset of features, chosen to optimally classify the data. This method of feature selection, combined with Bootstrap aggregation (bagging)—where each tree is trained on a bootstrapped sample from the original data—reduces correlations between individual trees, thereby enhancing performance of their ensemble (the forest).

For this analysis, the Random Forest is configured to grow 400 trees. This number was selected to ensure a robust and stable ensemble, as increasing the count beyond 400 results in negligible improvements in accuracy for this dataset.

The depth of each tree is optimized using hyperparameter tuning with Bayesian optimization. This process identifies optimal values for parameters like minimum leaf size, maximum number of splits, and the number of variables considered at each node for splitting (Hastie, Tibshirani and Friedman, 2008). The ease of tuning and robust performance of Random Forests, establish them as a preferred choice over other methods, such as deep neural networks, in certain scenarios (Athey and Imbens, 2019; Hastie, Tibshirani and Friedman, 2008).

To gauge the performance of predictive models, I plot the Receiver Operating Characetristic (ROC) Curve, and calculate the Area Under the ROC curve (AUC). The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at different decision score (probability of delinquency) thresholds. TPR measures the proportion of actual delinquencies correctly identified, while FPR calculates the proportion of performing loans mistakenly classified as delinquent.

The AUC provides a comprehensive measure of a model's performance, effectively capturing the essence of the ROC curve in a single number. Crucially, the AUC also has a probabilistic interpretation: it represents the likelihood that a randomly chosen delinquent loan will be assigned a higher probability of delinquency than a randomly chosen performing loan by the model. An AUC of 1 indicates perfect prediction, while an AUC of 0.5 suggests no discriminative power, equivalent to random guessing.

AUC is a generally robust metric, yet its informativeness may diminish in scenarios of class imbalance, such as when delinquent loans are far outnumbered by performing ones. Hence, I include an additional performance metric, average precision as recommended by Fuster et al. (2022). Average precision provides an assessment of the model's ability to accurately identify actual delinquent loans among those predicted as delinquent, across various threshold levels. This metric is derived by weighting the precision (the ratio of true positives to all positive predictions) by the increase in TPR (also called *Recall*) at each threshold level. Average precision is particularly useful in evaluating the model's performance in detecting the minority class, offering a complementary perspective to the AUC. A higher average precision indicates a more accurate model in predicting delinquency. I also calculate the 95% confidence interval for the AUC and AP using bootstrapping methods with 1000 replicas of the test set.

3.3 Summary Statistics

Table 2 provides summary statistics relating to bank loans. The average loan amount is INR 75,358 (= exp(11.23)), with a median of INR 99,708. The mean borrower age is 35 years, indicating a comparatively young cohort of loan recipients. Transactional behavior is varied, with an average of 2.78 transactions per day but a high standard deviation, signifying a broad range of business activities among borrowers.

Aggregate sales, calculated for the 90 days preceding loan disbursal, average at INR 73,865 (=exp(11.12)). Borrowers exhibit an average of 2.45 credit inquiries within a 60-day window before acquiring a loan, denoting a proclivity for credit-seeking; yet, a median of one inquiry suggests that a few borrowers with a high number of inquiries skew this average, with 25% of borrowers registering three or more inquiries. An examination of credit accounts reveals that, on average, borrowers have seven loans or credit card accounts active at the time of a new loan, which may underscore a reliance on multiple credit sources for liquidity needs.

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)	0.76	0.61	0.66	0.00	0.26	1.11	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 among borrowers is 720, positioning the average borrower in the ’prime’ category, which is traditionally demarcated by a score above 700. Nonetheless, a substantial proportion—over a quarter—fall below this prime threshold. Notably, around 10% of the borrowers had no prior borrowing history before their current bank loan, and approximately 5% did not have a credit score at the time of borrowing. These figures indicate a nuanced landscape of creditworthiness and borrowing history among the merchant borrowers.

An average business loan given by the bank had a tenure of about 18 months and carried an interest rate of about 17%, which is typical of business loans in the small business lending.

Table 3 presents summary statistics for fintech loans. By comparing these with corresponding statistics from bank loans, we uncover key differences and trends. On average, fintech loans are smaller, typically amounting to around INR 26,108, and tend to be more expensive, as indicated by an annual interest rate of 24% (this uniform rate across borrowers is not included in the Table). A notable distinction of fintech loans, compared to bank loans, lies in their association with shorter credit histories and fewer previously closed loans. To understand this, it’s important to note that all borrowers in my sample who obtained bank loans had also taken at least one fintech loan. Thus, the observed differences in summary statistics between the two loan types are not attributable to borrower composition but rather to their repeat borrowing behaviors. This indicates that borrowers with longer banking histories tend to take fewer fintech loans, as shown by the shorter credit history and lower number of previously closed bank loans associated with fintech borrowing. It appears that borrowers with limited experience in bank borrowing demonstrate a stronger preference for repeat fintech loans.

Table A3 and Table A4 in the appendix provide summary statistics for bank and fintech loans, classified by loan performance status. Included in these tables are results from a two-sample t-test, aimed at identifying mean differences between performing and delinquent loans. However, this analysis, focusing solely on mean values, overlooks the full data distribution and potential non-linear relationships. While these mean differences can suggest possible relationships, caution is advised in their interpretation due to their inability to capture the complexity of the data. A more thorough examination of the relationships, considering these nuances, is conducted in Section 4.4.

4 Results on Bank Loans

4.1 Payment Data for Loan Screening

Our analysis begins with a comparison of the Area Under the Curve (AUC) and Average Precision (AP) metrics across various screening models, all of which utilize pre-disbursal variables. For our foundational comparisons, we focus on the Payment History Aggregate (PHA) variables. The key aim is to gauge the utility of aggregated payment history data, especially when obtained from a financial institution different from the lending entity. Not only are these aggregated

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

payment histories easily accessible, but they also offer greater ease of standardization for sharing. Moreover, they generally present fewer privacy concerns. Considering these benefits, such aggregate metrics assume a vital role in the framework of open banking, marking an essential first step.

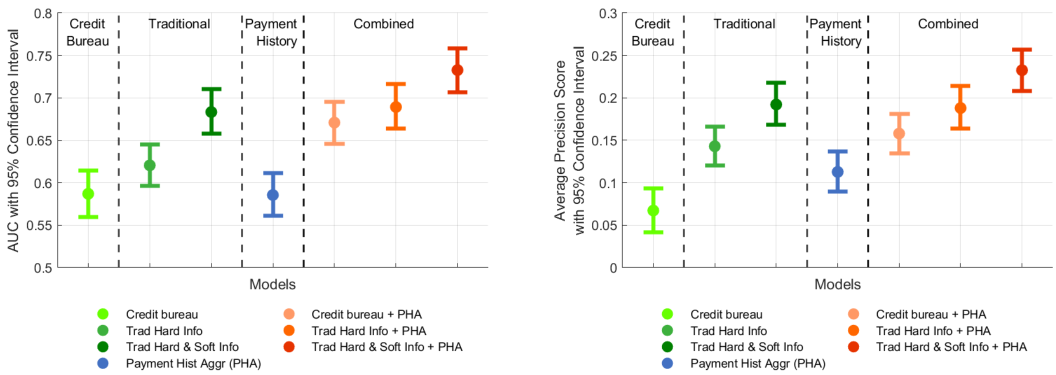
Figure 1 displays the Area Under the Curve (AUC) and Average Precision (AP) for various predictive models, accompanied by their 95% confidence intervals. Detailed performances of these models are tabulated in the appendix, specifically in Table A5. Additionally, Figure A2 illustrates the Receiver Operating Characteristic (ROC) curves, which form the basis for the calculated AUCs. Let's first examine the effectiveness of credit bureau data for traditional bank loans. The Credit Bureau model, with an AUC of 0.59, notably exceeds the random-guess baseline of 0.5, indicating a certain level of predictive accuracy for loan delinquency. Its AP stands at 0.07, which, though appearing modest, aligns with expectations for this type of predictive modeling.[8] This underscores the predictive capability of credit bureau data. Yet, how does it compare to a Traditional model enriched with a broader set of hard information, including credit bureau data? The *Traditional Hard Info* model, incorporating 14 diverse variables related to past borrowings, credit score, industry, and more, shows enhanced performance (AUC = 0.62, AP = 0.14) compared to the Credit Bureau model alone, indicating a significant predictive improvement.

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. For detailed variable description see Table A1. For the composition of predictive models see Table A2. The 95% confidence interval for the AUC and AP are calculated by bootstrapping with 1000 replicas of the test set.

(a) Area Under the ROC Curve

(b) Average Precision



In scenarios involving small business loans, lenders often lack detailed financial accounts

[8]For comparison, the study by Fuster et al. (2022) reports an AP of approximately 0.06 in predicting delinquency in the U.S. mortgage market with its most comprehensive model.

of the borrowing firms, relying instead on other hard information. Nonetheless, they also make concerted efforts to collect soft information about the borrowers, which, when combined with hard data, critically influences the formulation of loan contractual terms. To assess the impact of this soft information, we juxtapose the *Traditional Hard & Soft Info* model against the *Traditional Hard Info* model. The results reveal that integrating lender soft information into the predictive model enhances its accuracy, yielding a 6 percentage point (pp) increase in AUC and a 5 pp rise in AP. These enhancements represent significant relative improvements of 11% in AUC and 31% in AP, highlighting the valuable contribution of lender soft information to the predictive power in bank lending scenarios.

One might surmise that the disparities in model performance we've highlighted above aren't confined to the realm of soft information alone but may also encompass a wider array of hard information, which remains unobservable to us as researchers. This perspective introduces a legitimate caveat to our study. Nevertheless, in the domain of small business lending under scrutiny, the probability of overlooking significant hard information outside our dataset is slim. Small businesses often intermingle personal and business financial activities, making traditional financial statements like income statements and balance sheets rare. This backdrop strengthens our conviction that the differences in model outcomes primarily arise from the application of lender's soft information. Moreover, it catalyzes a critical question: In scenarios where detailed financial data on firms is sparse, does payment history reflect the same insights as lender soft information, fulfill a similar role as credit bureaus, or introduce an independent source of financial insight? Our next line of inquiry aims to dissect these aspects.

To answer these questions, we start by evaluating the standalone effectiveness of aggregate payment history. The PHA model, utilizing merely the four PH variables, records an AUC of 0.59, on par with the Credit Bureau model, and an AP of 0.11, marking a 2 pp increase over the Credit Bureau model. This sets the stage to explore whether Credit Bureau and PHA data encapsulate unique information for assessing loan performance. By examining the *Credit Bureau + PHA* combined model against each independently, we find the amalgamation boosts the AUC by 8 pp and the AP by 4 pp compared to PHA alone. These metrics underline the complementary nature of payment history and credit bureau information in bank loans, with their combination significantly elevating predictability beyond the capabilities of either source alone.

Next, we delve into a counterfactual scenario where a small-business lender considers incorporating PHA variables alongside traditional hard information. This exploration aims to quantify the additional value derived from such an integration. By contrasting the *Traditional Hard Info + PHA* model with the sole *Traditional Hard Info* model, we observe that the inclusion of PHA enriches the AUC by 7 pp and the AP by 5 pp. Interestingly, these enhancements due to PHA are quantitatively similar to the performance improvement we attributed to lender soft information. This result suggests that aggregate payment variables possess a value comparable to that of lenders' soft information, raising a pivotal question: Could PHA variables effectively encapsulate the essence of what is traditionally considered soft information by banks? In other words, do PHA harden the lender's soft information?

To examine that we study the predictive power of *Traditional Hard & Soft Info + PHA* model in comparison to its counterpart without PHA. Should PHA only be hardening soft information, adding it to the *Traditional Hard & Soft Info* model wouldn't markedly enhance performance. Conversely, a notable performance uplift equivalent to PHA's informational value discovered in the previous case would suggest that payment history constitutes an entirely distinct and complementary data source.

According to Figure 1, integrating PHA into the model with both hard and soft information elevates AUC to 0.73 and AP to 0.23, surpassing the traditional model's performance metrics.[9] Yet, the gain falls short of PHA's full informational value calculated above, hinting at PHA partially hardening soft information. For example, 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.

Comparing our results with other studies on alternative data reveals varied predictability across contexts. Agarwal et al. (2024) report credit score predictability with AUCs between 0.51 and 0.53 in Indian consumer lending, similar to Maggio and Ratnadiwakara (2024) in the US. In contrast, Iyer et al. (2016) finds a higher AUC of 0.62 of the credit score in peer-to-peer lending in the US, while Berg et al. (2020) observe even higher AUCs in German delayed-payment schemes. The performance of my credit bureau model, which extends beyond mere credit scores to include detailed credit report analytics, tends towards the upper end of these findings. Alternative data models in these studies on consumer lending show AUCs from 0.66 to 0.73. Delving into small-firm credit, Frost et al. (2019) and Huang et al. (2023) uncover AUCs of 0.76 and 0.87, respectively, within bigtech lending frameworks in Argentina and China, characterized by sales-linked repayment schemes. Our PHA-augmented model's performance (AUC = 0.73)[10] falls within these observed ranges. However, making direct comparisons are challenging, as our analysis centers on traditional bank lending, which significantly differs from the varied contractual structures examined in these studies.

We summarize our main findings of this section as follows:

Takeaway 1 (a) *Aggregate payment history matches the credit bureau in predictive accuracy for loan delinquency, with a notable synergy when combined—indicating that each captures unique, complementary aspects of borrower risk.*

(b) *We estimate aggregate payment data to have quantitatively similar value as the lender soft information in bank lending.*

[9]This improvement from PHA inclusion also hints at an underutilization of payment history in bank lending, as evidenced by the significant performance jump upon its addition to traditional models with hard and soft information.

[10]With PHG augmentation, AUC rises to 0.76 in bank lending, as detailed in Section 4.3.

(c) *Incorporating aggregate payment history into models with hard and soft data elevates predictive accuracy, showcasing its broader utility. This boost suggests a partial hardening of soft information—approximately 29% of PHA's value is attributable to hardening of soft information, while the rest is derived from independent, additional signals it generates.*

4.2 Comparative Informational Value of Payment Data Across Heterogeneous Borrowers

Our analysis extends to ascertain whether the notable performance of Payment History (PH) models is consistent across different types of firms, particularly across varying firm sizes and credit score categories. To explore variations among borrower sizes, I replicate the baseline analysis on two distinct subsets of borrowers—categorized into 'small' and 'large' based on their total transaction values within the 90 days prior to loan disbursal. Here, small firms are identified as those with transaction values below the median, whereas large firms are those above this threshold.

The performance of selected models, differentiated by firm size, is illustrated in Figure 2, with a detailed comparison available in Table 4. The result unveils that payment history serves as a more potent predictor for smaller firms compared to larger ones (column 4). This differential impact may stem from small firms' tendency to utilize a singular payment platform, in contrast to larger firms that might engage with multiple channels, potentially diminishing the predictive strength of data from a single source. Additionally, for small firms, operational cash flows which payment data are good at capturing, are more closely related with their debt repayment capabilities. In contrast, larger firms typically have a broader financial cushion and access to a variety of financing options to navigate cash flow fluctuations.

Next, we examine payment history's efficacy relative to other sources, across the firm size dimension. Table 4 shows while credit bureau data tends to favor larger borrowers (column 1), the full spectrum of hard information reverses the predictive performance across firm sizes (column 2). This balance suggests other hard data compensates for the less informative credit bureaus for smaller firms. It further explains why the integration of PHA with traditional hard information yields similar benefits across borrower sizes (as seen in column 6 vs. column 2), despite PHA's standalone effectiveness for small firms. Moreover, adding PHA to the *Trad Hard & Soft Info* model (column 7 vs. column 3) enhances performance for lenders across borrower categories. The differential impact on small versus large borrowers under various metrics remains ambiguous, particularly whether improvements stem from hardening soft information or offering new independent insights. Yet, unmistakably, PHA integration presents clear advantages—whether as novel signals or as cost savings in gathering soft information.

Exploring further, we delve into whether differences in model performance are influenced by borrowers' varying creditworthiness levels, as indicated by credit scores and histories. We

Figure 2: Bank Loans: Predictive Model Performance Comparison – by Size

Small borrowers have sales in the 90-day pre-disbursal period that fall below the median, while large borrowers exceed it. Large borrowers have above median sales. 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

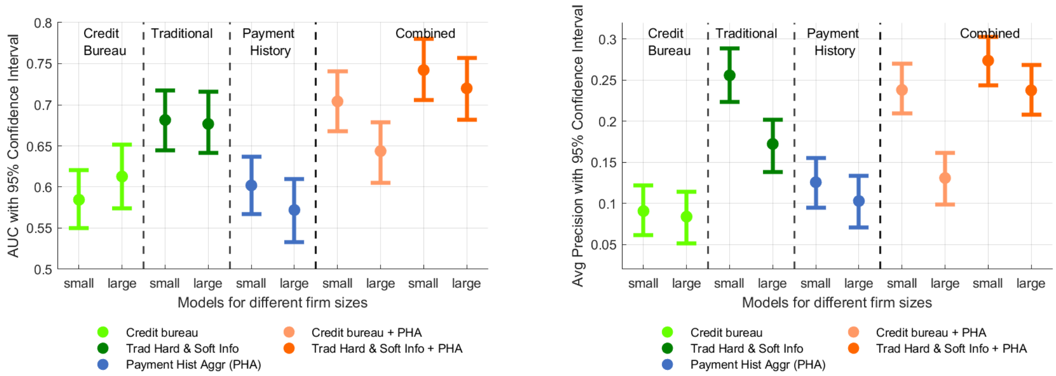


Table 4: Bank Loans: Out-of-Sample Predictive Performance with Aggregate Payment History

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. Large borrowers have above median sales. 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

categorize borrowers into *low-scored*, with scores below 700, and *high-scored*, surpassing this benchmark, aligning with India's lending industry standards for creditworthiness assessment[11]. One may argue that high-scored borrowers, boasting comprehensive credit histories and closer ties to the credit market, might benefit more from credit bureau and traditional models than from PHA models. Consequently, we explore if payment history introduces additional insights for both high-score and low-score borrowers, and particularly, its value for a third category of loans—made to the *thin-file borrowers*–those with no credit score or borrowing history at the time of loan application.

To address this, I analyze each model separately for loan sub-samples based on credit-score categories. The skewed distribution of credit scores, especially the smaller number of low-score an thin-file borrowers, necessitates a different evaluation method than the traditional test-train split, which might yield less reliable results for these groups. To overcome this, I utilize five-fold cross-validation within the random forest algorithm for all sub-samples. This method partitions the sub-sample into five equal parts, with each serving in turn as the test set once and as part of the training set four times. This approach, while being computationally intensive, ensures comprehensive utilization of all loans in the sub-sample for testing, and still enabling out-of-sample predictions. It provides narrower confidence intervals, essential even in sub-samples with fewer loans.

Table 5 delineates our findings, relating to these questions. I find, analogous to trends observed in the overall sample, the informational enhancement PHA provides equates to the boost from soft information for both high-scored and low-scored borrowers. This equivalence is observed by comparing the increments in AUC and AP between *Trad Hard + PHA* and *Trad Hard Info* models, to those between *Trad Hard & Soft Info* and *Trad Hard Info* models. While thin-file borrowers also benefit from PHA, the uplift is slightly less pronounced compared to soft information. Pondering a counterfactual where traditional lenders augmented their evaluation criteria with aggregate payment history in addition to the usual hard and soft information, our analysis posits a positive impact on predictability for thin-file borrowers, albeit to a lesser extent as evidenced by the comparison between column 7 and column 3. In essence, payment history enriches the predictive landscape for all borrower categories, yet its influence is particularly more marked for those with high and low credit scores than for thin-file borrowers.

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

Takeaway 2 (a) *Payment history emerges as a notably stronger predictor for smaller firms than larger ones, attributed to smaller firms' reliance on a single payment platform and their operational cash flows being more directly tied to their debt repayment capabilities. Despite PHA's distinct advantage for small firms, its integration with traditional information sources offers consistent benefits across all firm sizes, highlighting its universal value.*

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

Table 5: 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

(b) *Payment history significantly benefits high-score and low-score borrowers by providing insights that complement traditional hard and soft information. For thin-file borrowers, although PHA's uplift is relatively smaller than the other two categories, it still offers valuable predictive increments, underscoring payment history's role in broadening the scope of financial assessment for diverse borrower profiles.*

4.3 Is there a Granularity–Accuracy Trade-off?

A critical issue is the impact on predictive accuracy when transitioning from granular to aggregated data sharing, motivated by the significant privacy and technological challenges of handling granular data. To address this, our study enhances the payment history model by incorporating granular payment variables, which require transaction-level data or benchmarking against district-level aggregates. This enhancement forms the Payment History Granular (PHG) model, which includes a total of 12 variables, adding depth beyond the four found in the Payment History Aggregate (PHA) models. The objective of our analysis is to assess how these granular variables improve predictive accuracy compared to the PHA models, in both standalone and combined implementations.

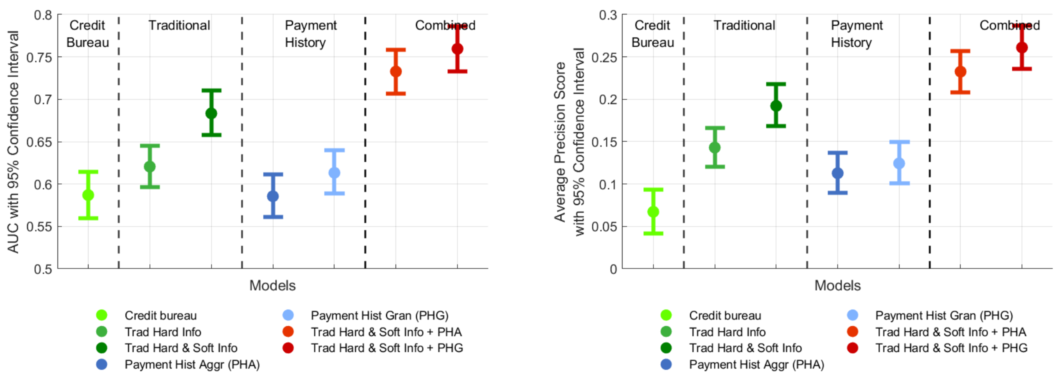
Figure 3 showcases the performance of the Credit Bureau and Traditional models in comparison with our PH models—both the PHA and PHG variants. For a comprehensive model comparison, refer to Table 6. We have reproduced the performance metrics of the Credit Bureau and Traditional models here to show them as benchmarks. It's pertinent to remember our baseline findings: aggregated payment history (PHA) was approximately as valuable as the lender's soft information, as detailed in Section 4.1. This foundational comparison informs our current analysis. Adhering to the methodology from our baseline scenario—which concentrated on aggregated data—our recent calculations reveal that granular payment history offers 20% to 30% greater value than the lender's soft information, varying with the performance metric applied. This enhancement is underscored by the PHG model's significant superiority over the PHA model, marked by a 4.7% improvement in AUC and a 10.3% leap in AP. These findings robustly confirm the substantial added value that granular payment data provides, markedly refining the accuracy of predictive models.

Figure 3: Bank Loans: Predictive Model Performance Comparison – Aggregate v/s 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. 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. Granular payment variables, as opposed to aggregate payment variables, necessitate transaction-level information or are calibrated against district-level payment aggregates.

(a) Area Under the ROC Curve

(b) Average Precision



The improved performance of PHG models, which build upon PHA models by incorporating more granular data, is a logical outcome. However, the key lies in understanding the extent of the trade-offs involved due to the significant value of the granular payment history. Our analysis reveals that harnessing the full potential of payment data might elevate privacy concerns and escalate data management costs. Thus, weighing these costs against the benefits of transitioning from PHA to PHG models is paramount. Although detailed cost quantification warrants further research, the inherent value of data interoperability is undeniable. This holds true even if PHG models become less favored due to cost implications since PHA models already deliver considerable predictive accuracy.

Table 6: 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 A5. The percentages indicate changes relative to the corresponding aggregate model detailed in Table A5.

		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

Takeaway 3 *The Payment History Granular (PHG) model offers improved predictive performance compared to the PHA model. However, the additional costs and privacy challenges associated with PHG may outweigh its benefits. The effectiveness of aggregate payment history models alone, with their substantial predictive power, continues to make a strong case for their use in loan screening.*

4.4 Which Predictors are the Most Important in Screening?

Our primary focus is to identify which *pre-disbursal* variables are most crucial in predicting loan delinquency. The task is challenging within the confines of complex, "black-box" algorithms like random forests, as these algorithms leverage non-linear relationships for predictions. Advances in interpretable machine learning, however, offer new methods for elucidating these complexities. Guided by Molnar (2023), we employ two complementary measures to identify critical variables: (i) Shapley additive explanations (SHAP), and (ii) Out-of-bag (OOB) variable importance through permutation. We identify the most important variables from most encompassing screening model—*Traditional Hard & Soft Info + PHG* . Selecting this model allows us to rank predictors comprehensively, comparing the predictive strength of granular versus aggregated payment data alongside traditional variables.[12] We will primarily concentrate on our first measure, SHAP, and relegate extensive discussions on our complementary measure to the appendix.

Shapley Additive Explanations (SHAP), developed by Lundberg and Lee (2017), offer a

[12]It's worth repeating that since this model incorporates contractual variables, it is not proposed for screening, rather it stands as a comprehensive benchmark. By evaluating the importance of payment history and traditional predictors within this extensive model, we gain insights into their predictive power, even when juxtaposed with the potentially more informative loan contractual terms. This analysis not only clarifies which variables hold sway but also enhances our understanding of their role against a broader predictor set.

method for analyzing how individual features influence specific predictions. This approach differs from global interpretability methods like OOB permutation importance, which assess feature significance across the entire dataset. SHAP values specifically illustrate how each feature *contributes* to a prediction's deviation from the model's mean prediction. Features wielding higher absolute SHAP values are identified as having a more pronounced effect on the prediction in question. This approach is grounded in cooperative game theory, conceptualizing the prediction task as a collective endeavor among features to produce a "surplus"—the deviation of a particular prediction from the mean prediction. Within this framework, the Shapley value determines each feature's *fair contribution* to the surplus generated. A positive SHAP value signifies that a feature elevates the probability of delinquency for a given instance, whereas a negative SHAP value indicates that it reduces it.

Given the computational intensity of calculating SHAP values, I compute them for a random 50% sample of the test set, covering approximately 1200 out-of-sample predictions. This method strikes an optimal balance between computational efficiency and the richness of interpretive detail. We utilize SHAP estimates firstly to rank variables by their overall importance, extending the analysis beyond individual predictions. Secondly, we assess the directionality of the relationship between predictor values and the probability of delinquency.

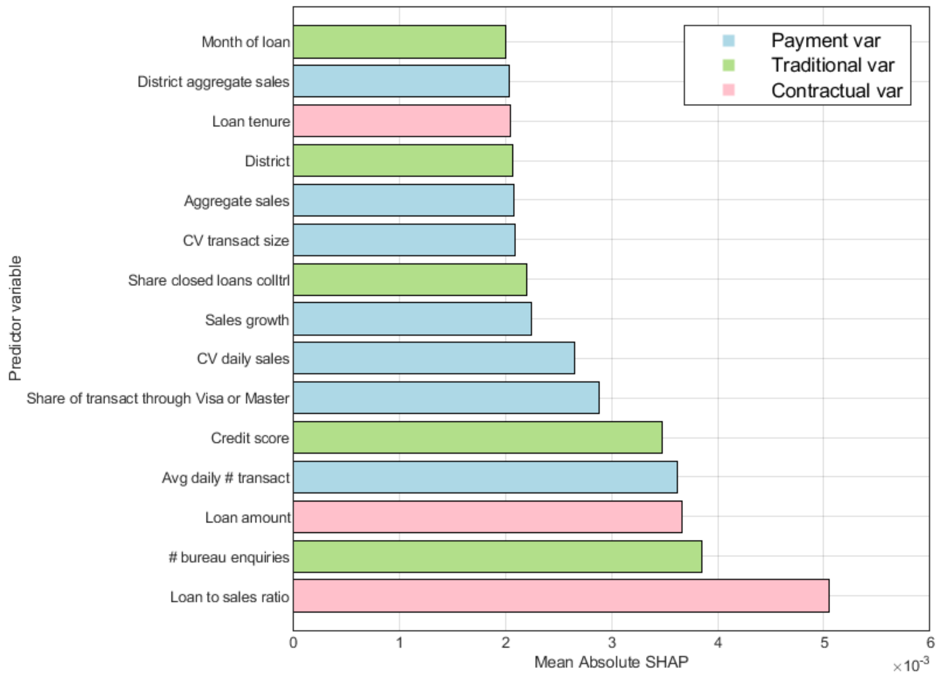
To understand the overall influence of each predictor within our models, we average the absolute SHAP values for each variable across all predictions. This mean absolute SHAP value acts as a barometer of a variable's global impact, with higher figures pointing to greater significance. Figure 4 delineates the top 15 of the 30 variables studied, categorizing them into payment, traditional, and contractual types. While our primary interest lies in the payment and traditional variables, the figure also enumerates contractual variables. The contractual variables include interest rates, loan tenure, and amount—variables determined by lender discretion—as well as the loan-to-sales ratio which is derived from loan amount. Although included in the analysis for completeness, these contractual variables fall outside our core investigative scope due to their endogenous nature.

The insights from Figure 4 unveil key findings regarding the most important predictors of loan delinquency. It's noteworthy that three out of four aggregative payment variables—average per-day count of transactions, aggregate sales, and sales growth—rank among the top 15, underscoring their significance in predicting delinquency. Additionally, from the granular payment history variables, the variability measures such as the coefficient of variation of daily sales emerge as crucial. Intriguingly, 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 offers another perspective on variable importance within Random Forest models, detailed in Appendix B. This approach supports our earlier findings and extends the analysis to loan sub-samples by firm size, highlighting how variable importance shifts. For instance, *Credit score* emerges as crucial for large borrowers but less so

Figure 4: 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.



for small ones. In summary, while payment variables dominate the list of important features for large borrowers—especially granular ones—for small borrowers, the significance of aggregative payment variables is more pronounced.

To study how various predictors influence loan delinquency, we turn to SHAP feature dependence plots, as depicted in Figure . These plots illuminate the relationship between each predictor's value and its contribution to the probability of delinquency, reflected through SHAP values. For each predictor, we overlay a polynomial curve—ranging from linear (degree 1) to more complex (up to degree 5)—to pinpoint the exact nature of its influence, choosing the degree that best aligns with the data as indicated by the adjusted R-square value.

The findings are revealing: higher values of certain payment history variables, like total sales and sales growth, generally lower probability of delinquency, showing a clear negative link. Conversely, the impact of average transaction size on delinquency risk becomes positive as transaction values rise. Notably, the variability in sales and transaction sizes shows a U-shaped relationship with delinquency risk, suggesting that extremes in variability correlate with higher delinquency risk. Among traditional variables, higher credit scores and extended credit histories are linked with reduced delinquency risks. In contrast, an increase in bureau inquiries signals a higher risk of delinquency.

As we conclude this section, it's pertinent to acknowledge a key limitation in our approach. In loan delinquency prediction, the critical role of non-linearities must be acknowledged. Interactions within and between variables significantly affect outcomes, and Random Forest algorithms are adept at detecting such complexities. Although we represent variable self-interactions with polynomial functions, fully capturing interdependencies between variables remains challenging. Advancements in machine learning have not yet fully surmounted this hurdle, leaving Random Forests to manage non-linear decision boundaries effectively but at the cost of interpretive clarity.

To evaluate the significance of non-linearities and interactions in our data, we compare the predictive performance of the random forest algorithm with that of a linear logit model. The results, presented in Table A6 in the appendix, show the AUC scores from both algorithms. We find that as the complexity of models increases, the random forest algorithm more effectively captures interactions, resulting in a substantially higher AUC compared to the linear logit model, particularly in models with a larger number of variables.

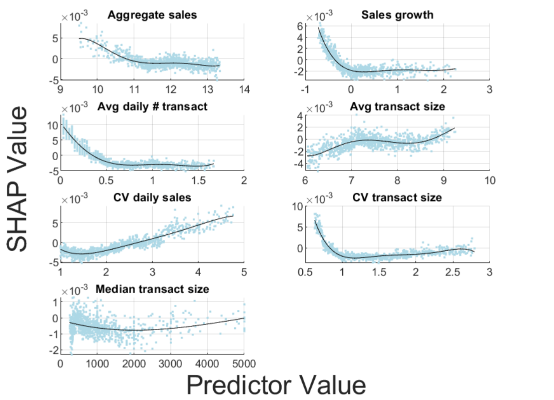
Takeaway 4 (a) *Three out of four aggregative payment variables—average per-day count of transactions, aggregate sales, and sales growth—emerge as significant predictors of loan delinquency. Granular variables, such as the coefficient of variation in daily sales, also play a critical role, especially for larger borrowers. Meanwhile, traditional variables like credit score and the number of bureau inquiries maintain their expected predictive power.*

(b) *Aggregate payment history variables like aggregate sales, average per-day number of transactions, and sales growth negatively correlate with delinquency*

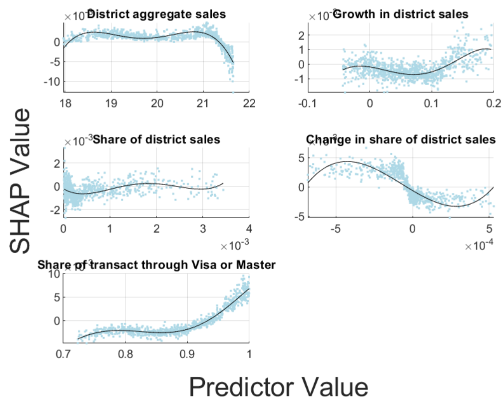
Figure 5: Bank Loans: SHAP Dependence Plots

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.

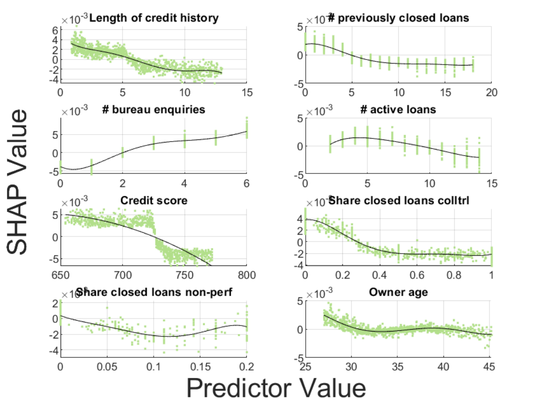
(a) Payment Variables Part-I



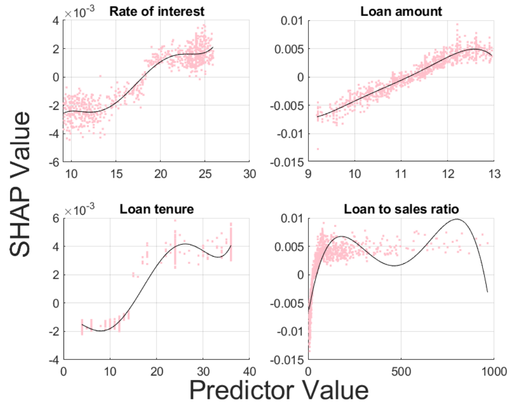
(b) Payment Variables Part-II



(c) Traditional Variables



(d) Loan Contractual Terms



probability, while larger transaction sizes elevate risk. Credit Score, a traditional variable, negatively influences delinquency probability, in contrast to the positive effect of Bureau enquiries.

4.5 Payment Data for Loan Monitoring

We now turn to loan monitoring, specifically assessing delinquency risks post-disbursal. Real-time payment data can offer insights into a borrowing business's financial health at frequent intervals. A critical question for lenders is determining how early they can detect a deterioration in loan repayment probabilities to take corrective action. To explore the potential of payment history variables as early warning signs, I perform predictive analysis at six consecutive 30-day intervals following loan disbursal. This analysis appends the pre-disbursal *Trad Hard & Soft Info + PH model* with post-disbursal payment history variables, calculated for each time window post-disbursal. I examine both the aggregate (PHA) and granular (PHG) versions of the post-disbursal models, comparing their predictive power against the corresponding comprehensive pre-disbursal model (*Trad Hard & Soft Info + PH model*).

Figure 6: 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

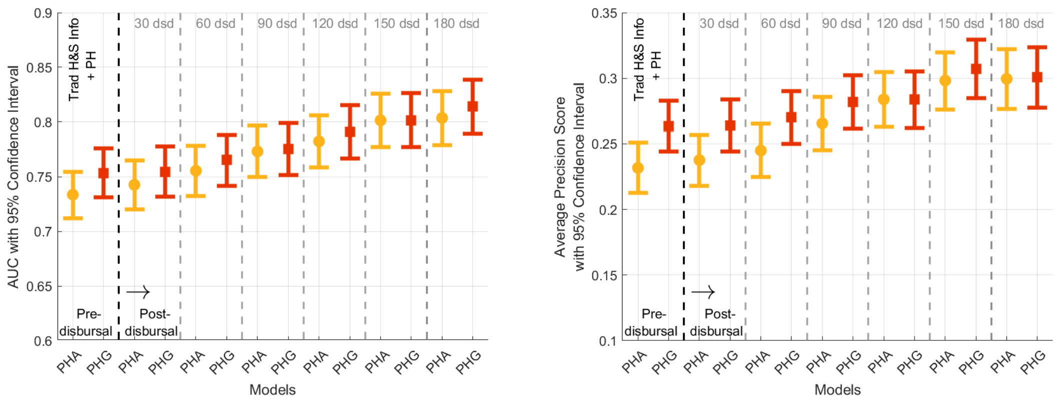


Figure 6 presents the outcomes of our early warning predictive analysis using PHA and PHG variables, revealing a consistent increase in both AUC and AP metrics throughout the 180-day post-disbursal period. Notably, by 180 days post-disbursal, the PHA model registers a 7 percentage point rise in both AUC and AP from their pre-disbursal benchmarks. To contextualize these improvements, the *post-disbursal* payment history's contribution to AUC within the initial 120 days post-disbursal (or roughly 5 percentage points) mirrors the contribution of payment history over the lender's hard and soft information in the *pre-disbursal* period. For AP, a similar comparative value (of roughly 4 pp) is observed within the first 90 days post-disbursal.

An more practical approach to understanding the monitoring value of payment data involves observing how the estimated probability of delinquency for loans updates over time within the testing sample. Ideally, as more payment information becomes available, the probability of delinquency should adjust upward for loans that eventually become delinquent and downward for those that do not. Figure 7 displays the evolution of mean out-of-sample estimated probability of delinquency for both delinquent and performing loans, offering several key insights.

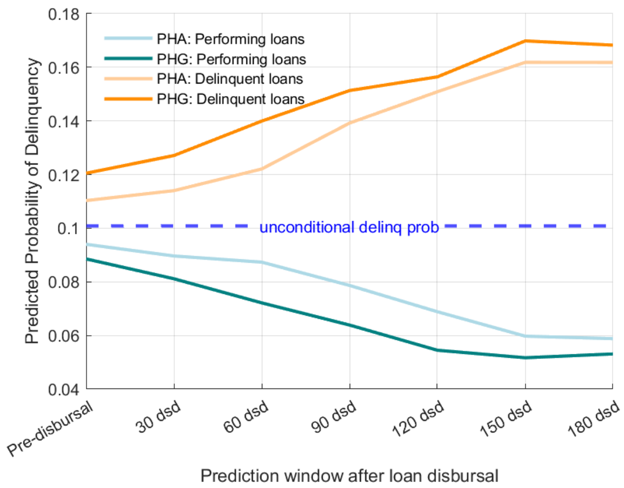
Firstly, the pre-disbursal models' effectiveness is clearly demonstrated, as the average probability of delinquency for loans flagged as delinquent surpasses the sample's overall delinquency rate (*unconditional delinquency probability*). This finding confirms the robust predictive capability of the screening models previously discussed. Secondly, consistent with expectations, the delinquency probability for loans destined to default progressively rises, albeit at an uneven 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 over time.

Finally, early warning models that utilize granular (*PHG*) data initially surpass their aggregate (*PHA*) counterparts soon after disbursal, a trend clearly illustrated in Figure 7. Post-disbursal, PHG models not only start with an advantage over PHA models, as highlighted by their differences before disbursal, but they also exhibit a marginally faster update rate in the initial period. This dual effect leads the PHG model to increase the delinquency probability to 14% within two months of disbursal—a threshold the PHA model reaches only a month later. However, the edge held by granular data fades over time, indicating a growing importance of *aggregate* payment variables as data volume builds. Notably, the performance of the PHG model at 180 days post-disbursal slightly regresses compared to its performance at 150 days, suggesting that the inclusion of an increasing number of payment variables might complicate model learning, potentially due to overfitting. Consequently, this evolution points to a quick narrowing of the granularity-accuracy trade-off in monitoring over a loan's life.

To unravel which variables play pivotal roles in the early warning models' ability to predict delinquency—and in which direction—we adopt an approach akin to the one used in the screening exercise, focusing this time on the post-disbursal scenario. We commence by calculating the mean absolute SHAP measure for the model 90 days after disbursal (90 dsd). It's important to note that, unlike the screening models, the post-disbursal models primarily differ in the payment variables included. A model covering a longer window incorporates the same PH variables as one with a shorter window but also adds variables up to the specified

Figure 7: 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.



window. This rationale guides the selection of the 90-dsd model as a representative model for our variable importance analysis, because it as at the midpoint among the windows we examine.

Figure 8 reveals the top 15 variables by importance, leading to two significant insights: first, the post-disbursal PH variables consistently outrank their pre-disbursal counterparts in importance. Second, within the realm of post-disbursal variables, those related to sales growth emerge as paramount, signifying their effectiveness in flagging potential financial distress. Traditional variables, including the number of bureau inquiries and credit score, remain significant predictors post-disbursal, reaffirming their continuous relevance across different stages of the loan lifecycle.

To elucidate the direction of relationships among variables, SHAP values are calculated for each variable within the 90-dsd representative model. Due to spatial constraints, Figure 9 specifically showcases SHAP-dependence plots for post-disbursal payment variables only. The directionality and qualitative results concerning pre-disbursal traditional and payment history variables align with those observed in the screening analysis depicted in Figure 5, and are therefore not repeated. Notably, Figure 9 illustrates that loans with diminished sales growth post-disbursal see an increased probability of delinquency. Conversely, a higher total sales volume, an elevated average daily transaction count, and a greater share of district sales correlate with a decreased delinquency likelihood in the post-disbursal phase.

Figure 8: 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.

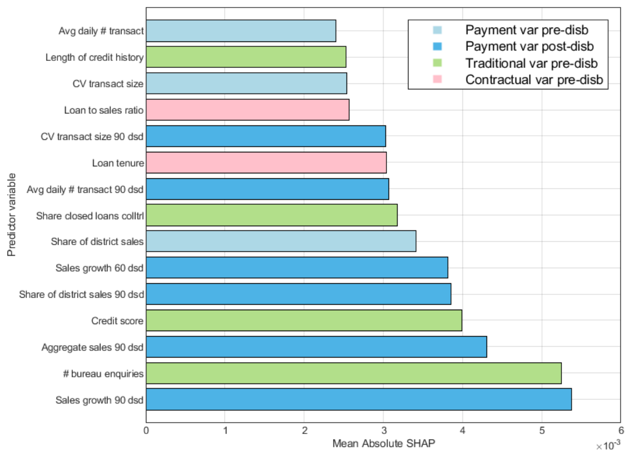
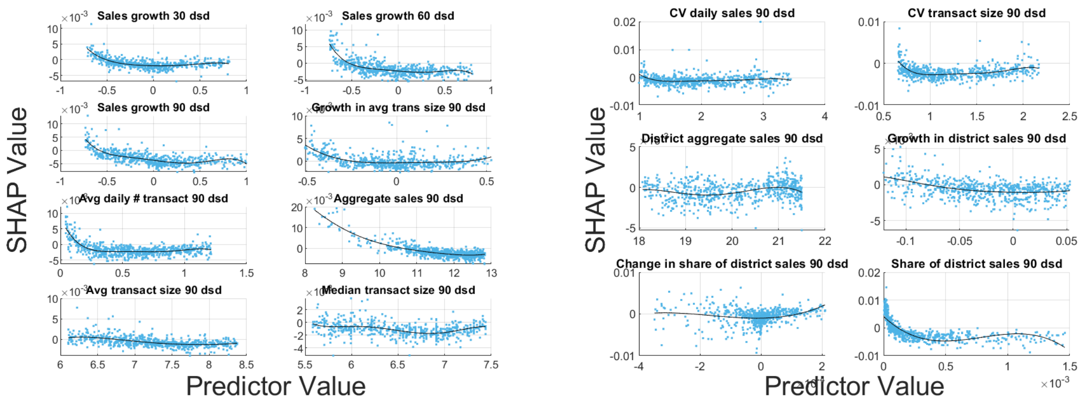


Figure 9: Bank Loans: SHAP Dependence Plots for Post-disbursal Variables in Early Warning Model

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 5. 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



Takeaway 5 (a) *Payment data holds significant potential for generating early warning signals, aiding lenders in monitoring loans. Within 120 days post-disbursal, post-disbursal payment history variables enhance AUC to the same degree that pre-disbursal payment data does compared to traditional lending information under screening. For AP, this equivalent enhancement occurs within 90 days.*

(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 from PHA as essential early warning signals.*

5 Payment Data and Fintech Loans

Payment fintech loans present a fascinating case study in how altering contractual features impacts the information content of payment and traditional variables in assessing delinquency risk. This exploration is particularly insightful because the borrower samples for both bank and fintech loans are identical—every bank borrower in our study has also taken at least one fintech loan. Globally, payment fintechs and bigtech platforms are innovating with sales-linked loans, where repayments are directly tied to the merchant's sales processed by the lender. This section delves into how traditional and payment history variables fare in screening and monitoring these novel loan types.

Embarking on a path parallel to our exploration of bank loans, we first confront the screening challenge for fintech loans, employing pre-disbursal variables. Our approach begins with traditional models, enriched by the integration of Payment History Aggregate (PHA) variables. Figure 10 reveals the baseline results for these fintech loans, while Table A7 in the appendix provides a more detailed analysis. The findings illuminate notable differences when juxtaposed with bank loans. A striking initial observation is that the Credit Bureau's predictive effectiveness in terms of the Area Under the Curve (AUC) is somewhat diminished for fintech loans compared to bank loans. Intriguingly, this pattern reverses when we pivot to consider Average Precision (AP), underscoring a nuanced dynamic in predictive performance between the two loan types.

Furthermore, traditional models exhibit diminished predictive power in the realm of fintech loans compared to their bank counterparts, hinting at a reduced influence of private, soft information. This inference is drawn from the observation that the traditional model with hard and soft information offers only marginal improvement over its counterpart with only hard information in the fintech scenario, as opposed to a more pronounced enhancement in the bank loan context. Since loan terms are generally indicative of a lender's soft information, this suggests that fintech lenders contribute less in terms of soft signals compared to banks.

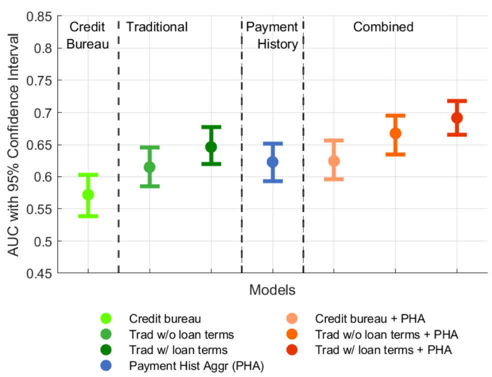
The performance of the Payment History Aggregate (PHA) model in the context of fintech loans is quite revealing. With an AUC of 0.62 and an AP of 0.2, the PHA model outperforms the Credit Bureau. However, the critical question is whether the PHA complements or substitutes the Credit Bureau. To explore this, we combine both models and discover that this amalgamation does not enhance predictive performance beyond the standalone PHA model. This suggests that the PHA model already encapsulates the information provided by the Credit Bureau regarding fintech loans.

This finding carries significant implications. In economies where establishing credit bureaus is an expensive endeavor, one potential solution to reduce reliance on these bureaus could be the introduction of sales-linked loans for businesses. The efficacy of the PHA model in fintech loans demonstrates its capacity to sufficiently inform credit decisions, possibly making it a viable alternative in contexts where traditional credit reporting mechanisms are less feasible.

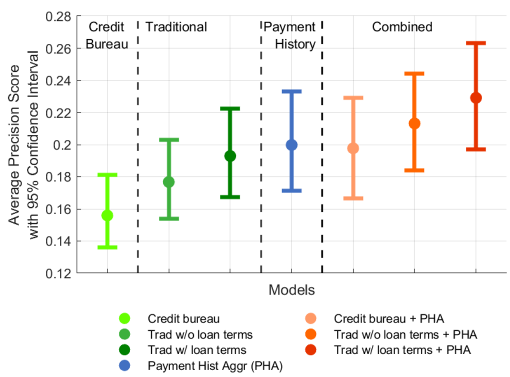
Figure 10: Fintech Loans: Predictive Model Performance Comparison

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



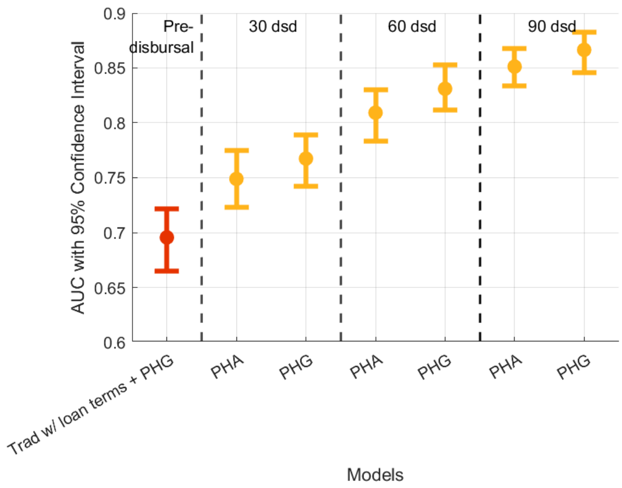
Incorporating Payment History Granular (PHG) variables into our fintech loan analysis reveals notable improvements. The PHG model, achieving an AUC of 0.65 as shown in Table **??**, outperforms the PHA model and even the traditional model with hard and soft information. This raises questions about the role of lender soft information. While PHG's superior performance suggests it might overshadow lender's soft information, combining PHG with traditional loan terms actually enhances predictive accuracy. This suggests that lender soft information remains valuable and synergizes well with PHG. However, despite these improvements, the comprehensive fintech loan model doesn't match the predictive power of its bank loan counterpart, indicating a more substantial contribution of soft information from bank lenders.

Exploring the effectiveness of payment history in monitoring sales-linked fintech loans is essential, especially considering moral hazard—a critical factor in loan performance post-disbursal (Karlan and Zinman, 2009). In this context, I extend the pre-disbursal benchmark model Traditional Hard & Soft Info + PH for fintech loans, performing predictive analyses across three 30-day intervals post-loan issuance. This duration aligns with the generally shorter tenures of fintech loans compared to traditional bank loans.

Figure 11: Fintech Loans: Predictive Performance Comparison in Early Warning Models – Aggregate v/s Granular Payment History

Area Under the ROC Curve

AUC varies between 0.5 (random guess) and 1 (perfect prediction) 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. Post-disbursal prediction involves augmenting the pre-disbursal Traditional Hard & Soft Info + PH model with additional post-disbursal payment history variables, calculated within each respective time window since disbursal (days-since-disbursal(dsd)). The 95% confidence interval for the AUC and AP are calculated by bootstrapping with 1000 replicas of the test set.



The findings, presented in Figure 11, reveal the swift impact of PH variables on predictive accuracy, with an impressive 10% increase in AUC at 30 days and a 20% surge at 60 days post-disbursal, relative to the pre-disbursal benchmark.

To summarize, these results indicate that in the realm of fintech lending, with sales-linked repayments, the performance of combined models in screening is less effective than in bank loans. However, the fintech monitoring models quickly offset this gap post-disbursal. Given the sales-linked nature of these loans, the evolution of post-disbursal sales data becomes increasingly critical. This aligns with findings by Rishabh and Schäublin (2021) and Russel, Shi and Clarke (2023), who have highlighted moral hazard in fintech lending by showing that merchants in sales-linked loans have tendencies to divert sales away from the lending platform, aiming to delay repayments. This scenario presents a trade-off: while sales-linked loans reduce dependence on traditional, backward-looking data sources like credit bureaus, they also potentially exacerbate moral hazard issues. Further research is necessary to fully comprehend the broader implications of such loan contracts.

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

Takeaway 6 (a) *The payment history aggregate (PHA) model in fintech loan screening outperforms the Credit Bureau model and shows that combining both does not yield additional predictive benefits. This dominance of PHA suggests potential redundancy of traditional credit bureau information in sales-linked fintech lending contexts. Furthermore, the reduced effectiveness of models with hard and soft information in fintech, as opposed to bank loans, highlights the greater relevance of lender soft information in traditional banking compared to fintech lending.*

(b) *Post-disbursal, PHA variables significantly improve predictive performance, evidenced by a notable rise in AUC shortly after loan issuance in fintech lending. This quick uptick, however, may mirror the moral hazard challenges unique to sales-linked loans, as identified in recent studies. Our findings underscore a critical tension in fintech lending: while the dependence on traditional data sources like credit bureaus diminishes, the rise of moral hazard poses new risks, necessitating further research into the implications of such lending contracts.*

6 Conclusion

I utilize a distinctive setting that enables a clear valuation of interoperable payment data of small businesses for lending. My investigation spans a wide range of queries related to open banking, demonstrating its vast potential in developing a new lending technology. However, my results also unveil several nuances, offering important policy implications.

Firstly, establishing credit information sharing institutions like credit bureaus is an expensive endeavor. Currently, more than half of the global firms and individuals remain unlisted in any bureau or public registry. In traditional lending with standard debt contracts, I find that payment histories complement rather than substitute the information from bureaus. This suggests an optimal strategy could be to develop credit bureaus that integrate traditional credit history with transaction history. Such a synthesis could be facilitated by Open Banking policies.

Secondly, in contexts where establishing bureaus is prohibitively costly and credit information sharing is challenging, credit markets could still function effectively. They can rely on standard debt contracts underwritten based on payment history. While the outcomes may not be as robust as those with comprehensive bureau data, they are certainly more favorable than having neither bureau data nor Open Banking. However, an intriguing alternative in the absence of bureaus is the adoption of sales-linked loan contracts. These contracts could render bureaus redundant but introduce their own set of moral hazard challenges.

Finally, the design of Open Banking systems carries critical implications, particularly regarding the balance between accuracy and privacy and technological costs. My study indicates a clear trade-off in this respect. More granular data may enhance predictive accuracy but raises significant privacy concerns in loan screening. There is a need for further research to thoroughly investigate these trade-offs and inform the design of effective and responsible Open Banking policies.

References

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*.

Beaumont, Paul, Huan Tang, and Eric Vansteenberghe. 2023. "Collateral Effects: The Role of FinTech in Small Business Lending."

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.

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.

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.

BIS and World Bank. 2020. "Payment aspects of financial inclusion in the fintech era." Bank for International Settlements and World Bank.

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.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2018. "Fintech, regulatory arbitrage, and the rise of shadow banks." *Journal of Financial Economics*, 130(3): 453–483.

CFPB. 2023. "Rulemaking on Personal Financial Data Rights." *Consumer Financial Protection Bureau*.

Cornelli, Giulio, Jon Frost, Leonardo Gambacorta, P Raghavendra Rau, Robert Wardrop, and Tania Ziegler. 2020. "Fintech and big tech credit: a new database." *BIS Working Papers*.

Eça, Afonso, Miguel A Ferreira, Melissa Porras Prado, and Antonino Emanuele Rizzo. 2022. "The Real Effects of Fintech Lending on SMEs: Evidence from Loan Applications."

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, Matthew Plosser, Philipp Schnabl, and James Vickery. 2019. "The Role of Technology in Mortgage Lending." *Review of Financial Studies*, 32(5): 1854–1899.

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.

Gambacorta, Leonardo, Yiping Huang, Zhenhua Li, Han Qiu, and Shu Chen. 2022. "Data versus Collateral*." *Review of Finance*, 27(2): 369–398.

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

Gopal, Manasa, and Philipp Schnabl. 2022. "The Rise of Finance Companies and FinTech Lenders in Small Business Lending." *Review of Financial Studies*, 35(11): 4859–4901.

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.

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.

Karlan, Dean, and Jonathan Zinman. 2009. "Observing unobservables: Identifying information asymmetries with a consumer credit field experiment." *Econometrica*, 77(6): 1993–2008.

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.

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.

Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine learning: an applied econometric approach." *Journal of Economic Perspectives*, 31(2): 87–106.

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."

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.

Ouyang, Shumiao. 2021. "Cashless payment and financial inclusion." *Available at SSRN 3948925*.

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

Philippon, Thomas. 2016. "The fintech opportunity." *NBER Working Paper*, , (No. 22476).

Plaitakis, Ariadne Ingrid, and Stefan Staschen. 2020. "Open Banking: How to Design for Financial Inclusion." *World Bank Report*.

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."

Tang, Huan. 2019. "Peer-to-peer lenders versus banks: substitutes or complements?" *Review of Financial Studies*, 32(5): 1900–1938.

TransUnion. 2022. "More than 45 Million Americans are Either Credit Unserved or Underserved." *TransUnion Newsroom*.

Appendices

A Additional Figures and Tables

A.1 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: 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)	0.77	0.65	0.12***
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 A4: 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 A5: 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 A6: Bank Loans: Out-of-Sample Predictive Performance Non-linear vs. Linear Algorithms

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.

						Models Combined with:	
		Traditional		Payment History		PHA	PHG
Predicted var: Delinquency	Credit Bureau (1)	Hard Info (2)	Hard & Soft Info (3)	Aggregate (PHA) (4)	Granular (PHG) (5)	Mod (3) + Mod (4) (8)	Mod (3) + Mod (5) (9)
				Area Under the ROC Curve			
Non-Linear (Random Forest)	0.59	0.62	0.69	0.59	0.61	0.73	0.76
Linear (Logit)	0.56	0.56	0.65	0.55	0.57	0.66	0.65
% Δ Non-linear over Linear	4.9	11.1	5.7	6.9	7.6	11.6	17.3
				Average Precision			
Non-Linear (Random Forest)	0.07	0.14	0.22	0.11	0.12	0.24	0.27
Linear (Logit)	0.10	0.12	0.18	0.10	0.10	0.18	0.15
% Δ Non-linear over Linear	-34.6	18.2	18.9	17.9	21.3	37.3	74.8
N. Obs. Train	9578	9578	9578	9578	9578	9578	9578
N. Obs. Test	2394	2394	2394	2394	2394	2394	2394
N. Predictors	2	14	17	4	12	22	30

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

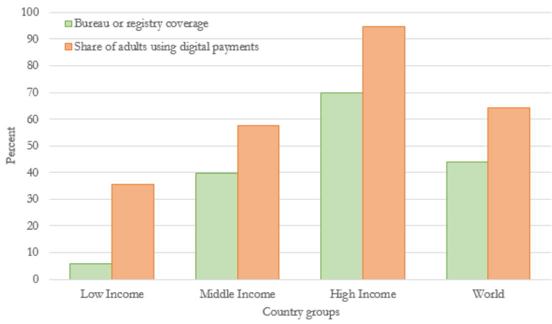
AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models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

		Traditional		Payment History	Models Combined with PHA		
Predicted var: Delinquency	Credit Bureau (1)	w/o Loan Terms (2)	w/ loan terms (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

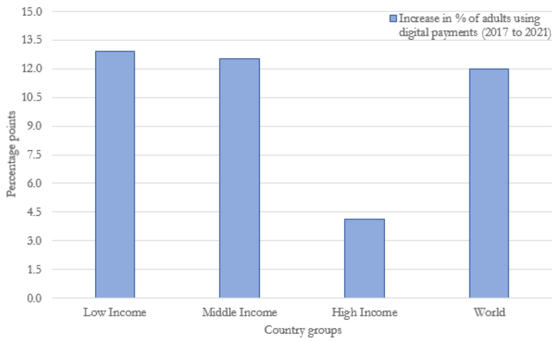
A.2 Figures

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)

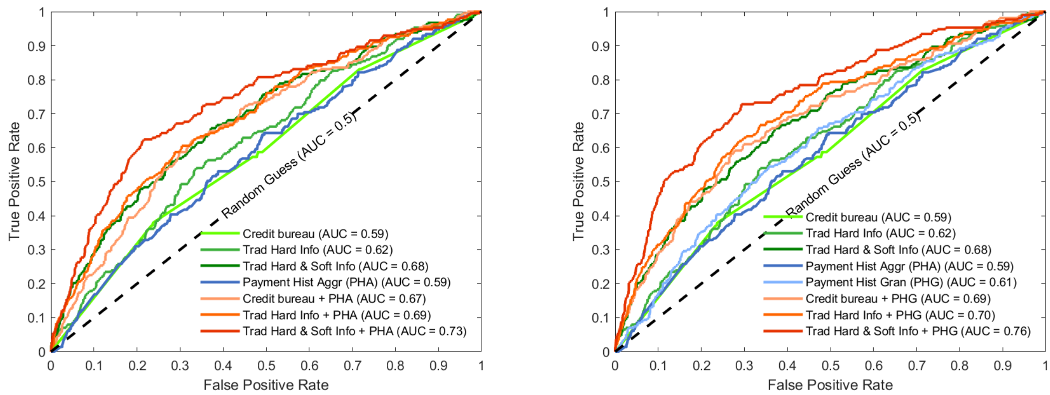
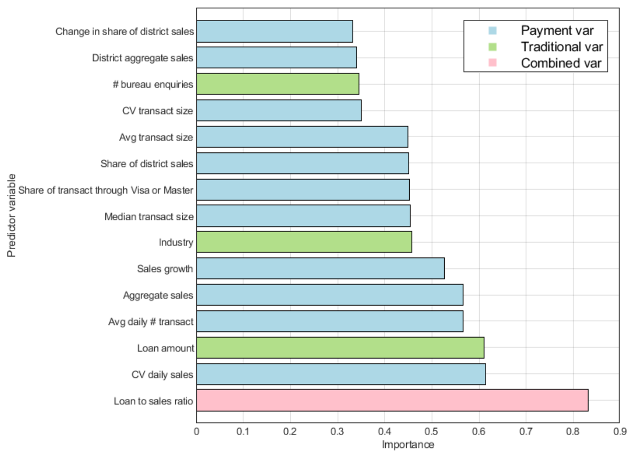


Figure A3: Fintech Loans: Top 15 Predictors of Delinquency

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, 15 traditional, 2 contractual loan terms (Loan amount, suggested tenure), 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.



B Out-of-bag (OOB) Variable Importance by Permutation

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 B4 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 B5 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 B6 showcases the 15 most impactful variables in this model. Remarkably, the lineup of top predictors mirrors closely those identified through the absolute SHAP values in Section 4.5, underscoring consistent findings across both measures.

Figure B4: 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.

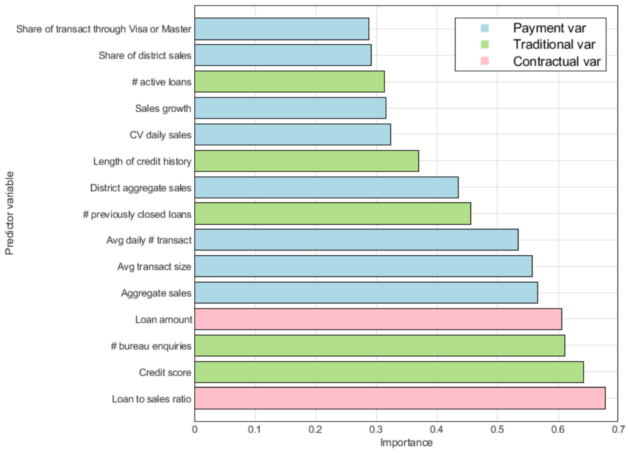
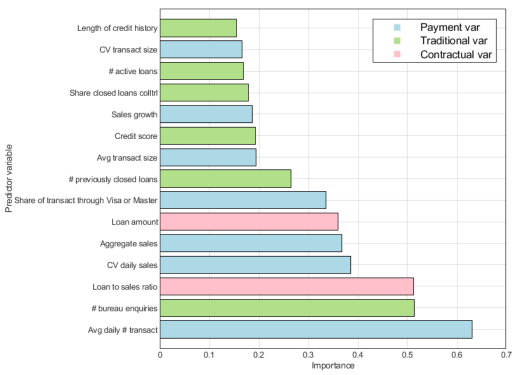


Figure B5: Bank Loans: Top Predictors of Delinquency Based on OOB Permutation – by Size

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

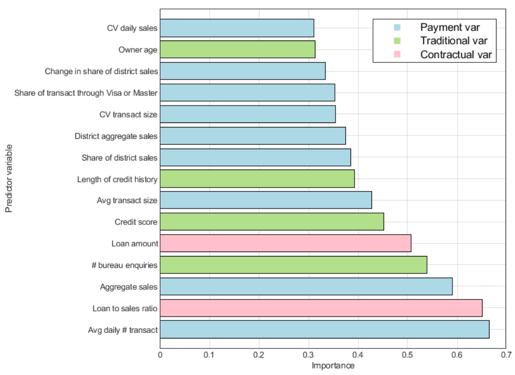


Figure B6: Bank Loans: Top 15 Predictors of Delinquency in Post-Disbursal Predictive Models

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.