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

Kumar Rishabh†

Click here for the latest version

Abstract

I study the informational value of interoperable payment data in lending, essential to global open banking initiatives. My focus is on the role of firm payment history in loan screening and monitoring, assessing how it interacts with credit bureau data, and the privacy-accuracy trade-offs involved. I utilize a unique dataset that links electronic payment history from a fintech company with both bank loans and fintech loans issued to the same set of Indian small businesses. I find that payment history parallels credit bureau data in predicting delinquency on bank loans and both sources of information complement each other. Integration of payment history in traditional lending models leads to a substantial increase in predictive accuracy. Payment history shows significant predictive strength across various borrower segments, especially aiding small businesses and those with a thin credit history. Using interpretable machine learning technique, I pinpoint critical variables like Aggregate Sales and Sales Growth in forecasting delinquency. However, I observe a potential trade-off between accuracy and privacy. I also show how payment data can be instrumental in early loan monitoring. In the context of sales-linked fintech lending, a nuanced tension emerges: the waning relevance of credit bureau data juxtaposed against increasing moral hazard concerns.

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 acknowledge the 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: Dec 2023.

†HEC Lausanne (University of Lausanne) & 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, 52% of countries, totalling 87 out of 168, had embarked on Open Banking initiatives (Babina, Buchak and Gornall, 2022). This model envisions a new era where financial products are built upon shared data from incumbent institutions. Foremost among its uses, as identified in industry surveys, is enhancing risk assessment in lending (Experian, 2022). Seen as a novel lending technology, Open Banking revolves around the interoperability of payment history.

The surge in interest in Open Banking as a lending technology is understandable, especially given the significant limitations of the currently dominant lending technology - bureau-based credit scoring. A glaring issue is the insufficient coverage of credit bureaus, leaving over half of the world's firms and individuals without representation and widening the financial access gap. Figure A1 in the appendix starkly illustrates this, especially in developing countries. Policymakers now view Open Banking as a promising tool to improve credit risk assessment and widen financial inclusion (BIS and World Bank, 2020; Plaitakis and Staschen, 2020). The rationale is that electronic payment transactions, ubiquitous in daily economic activities, generate valuable payment histories. These histories have the potential to bridge the information gap for those currently unrepresented in credit bureaus, offering a new dimension to financial inclusivity.

Is Open Banking simply advantageous for those beyond the reach of credit bureaus? Perhaps not. It may be beneficial for already covered population, because bureau scoring has another limitation. Traditional credit scores often fail to capture a borrower's current financial health, being inherently backward-looking. In contrast, payment histories are real-time and frequently updated, offering granular insights. This contrast prompts a reevaluation: might payment history provide a fuller picture, even when credit scores are available?

Yet, these favorable arguments for Open Banking rest on the premise that sharing payment information between financial institutions is intrinsically beneficial. It presumes that payment data from one source can enhance another institution's understanding of a borrower's financial health. While this seems logical, it requires further scrutiny. For instance, in specific market segments where relationship lending and soft information are crucial, payment data might have limited impact.

Thus, assessing the true value of interoperable payment data is crucial. This applies not only to those without a credit history but also to well-established credit users, including larger firms or those with a substantial credit history. My paper tackles several pivotal questions: Do payment histories complement or replace credit bureau information? Can they enrich the existing blend of a lender's soft and hard information? Are they effective for pre-loan screening and post-disbursal monitoring through early warning signals? How should the granularity of shareable data be balanced with privacy and technological costs? And finally, how do these dynamics shift when we move away from traditional bank loans to innovative contracts under fintech lending?

Addressing the questions I've raised presents a significant challenge, mainly due to the scarcity of scenarios where lending and payment data are derived from separate sources, directly reflecting data interoperability.[1] Further complicating this issue is what I term the BFP critique, after Berg, Fuster and Puri (2022). Their critique points out that most research designs assessing the effectiveness of alternative data are plagued by apples-to-oranges comparisons. This stems from the marked differences in borrower samples between bank loans and alternative data-dependent lenders, which hampers straightforward comparisons. Furthermore, even with harmonized borrower samples, the disparity in the nature of bank lending contracts versus those of alternative lenders poses another challenge, complicating comparisons and obscuring clear conclusions.

Addressing the interoperability challenge and circumventing BFP critique forms the backbone of my research design. I achieve this by linking the lending contracts of banks with the payment history data sourced from a payment fintech, focusing on small businesses in India. The fintech not only processes their electronic payments but also extends innovative loans with sales-linked repayments. However, it's the bank loans obtained by these fintech clients that form the foundation of my main analysis. This approach not only homogenizes the borrower samples, it checks the effectiveness of payment data at the level of the original bank loan contract. The role of fintech loans in my research is also significant, but I explore that in further detail later in the paper.

To tackle our questions on data interoperability, I focus on assessing the predictive effectiveness of various variable sets, termed as *models*, in forecasting loan delinquency. For the screening exercise, these models are constructed using data gathered before loan disbursal. For the monitoring exercise, they extend the most extensive screening model with the payment variables from the post-disbursal periods.

The first model is the Credit Bureau Model, which primarily uses the borrower's credit score and other bureau-related information, such as the number of inquiries before the loan and previous loan performance. The Traditional Model comes in two forms: one that combines the Credit Bureau data with borrower demographics like age, location, and industry, representing traditional hard information. The other variant of the Traditional Model further incorporates loan contract details such as amount, tenure, and interest rates, making it the most comprehensive traditional model.

In addition, there are two variants of the Payment History (PH) Models. The PHA (Payment History Aggregate) model consists of four pre-disbursal aggregate variables: total sales, sales growth, average daily transactions, and average transaction size. The PHG (Payment History Granular) model builds on the PHA by adding detailed, transaction-level data and comparisons with district-level averages. Lastly, I create combined models that amalgamate these traditional

[1] A notable exception to this is the study by Ghosh, Vallee and Zeng (2021). However, our studies differ in significant ways that I explain further below.

and payment history models. The combined model Traditional with loan terms + PHA (correspondingly PHG) is the most extensive screening model of the PHA (correspondingly, PHG) kind.

To predict loan delinquency, I employ the Random Forest machine learning algorithm, followed by making out-of-sample predictions. The predictive performance of various models is evaluated by plotting Receiver Operating Characteristics (ROC) curves and calculating the Area Under the ROC Curves (AUC). An AUC closer to one signifies better predictive performance, whereas an AUC of 0.5 implies a predictive accuracy no better than a random guess. Additionally, I calculate Average Precision (AP) as a complementary measure, indicating the accuracy of our delinquency predictions.

These performance comparisons offer insights into the key questions of my research. For instance, evaluating the PHA model against the Credit Bureau model sheds light on how aggregative payment data stack up against established lending technology. Crucially, by comparing the combined model of PHA and Credit Bureau against each standalone model, I assess whether these data sources share overlapping information about loan delinquency or capture distinct aspects, essentially determining if they are substitutes or complements.

Moreover, analyzing the extent to which PHA enhances the most extensive traditional model's performance helps gauge the value of interoperable payment data. I extend this analysis across various borrower sub-samples to explore potential heterogeneity, providing a comprehensive understanding of these models' effectiveness in different borrower contexts.

Our results relating to the baseline 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, showing a complementary relationship under bank lending.*
ii. *Integration of PHA with traditional models significantly improves predictive accuracy, demonstrating a 6% increase in AUC and 9% in AP over the traditional model with loan terms, highlighting the benefits of interoperable payment history data in loan screening.*
iii. *PHA shows notable predictive strength for small borrowers, outperforming its efficacy for larger borrowers, and improves traditional underwriting models for all borrower segments.*
iv. *Beneficial for both high-score and low-score borrowers, PHA is especially effective for thin-file borrowers, with a 4.6% rise in AUC, accentuating its impact in scenarios where lenders rely mainly on hard information.*

To pinpoint the key variables driving the predictive power of our models, I apply interpretable machine learning techniques. The first, out-of-bag (OOB) permutation measure, assesses a variable's importance based on the impact of its value permutation on prediction accuracy. Essentially, if random shuffling of a variable's values across observations worsens prediction error, it indicates the variable's significance. The second technique, Shapley Additive Explanation (SHAP), assigns importance to each variable for a given prediction, based on its

Shapley value within the model's variable coalition. This method not only highlights variable importance but also elucidates the direction of their relationship with the predicted outcome, enhancing our understanding of the variables' influence in the model. The OOB underscores the vital contribution of payment history variables, especially the PHA. This finding suggests that these aggregative variables already capture a major part of the loan screening capability in payment history. Additionally, SHAP analysis reveals that while higher values of certain PHA variables like Aggregate Sales and Sales Growth tend to decrease the probability of delinquency, Average Transaction Size increases it. Contrarily, traditional variables such as Credit Score decrease delinquency probability, unlike Bureau enquiries, which have a positive effect.

In open banking, banks face three intertwined challenges: safeguarding customer data privacy, handling diverse API standards, and managing technological intricacies of processing detailed data. Balancing privacy concerns with regulatory requirements limits the extent and type of data shared. Divergent API protocols further complicate data sharing, leading to inconsistencies. The complexity of handling granular data adds another layer of difficulty. A practical solution for banks might be to lean towards sharing aggregated data, simplifying the process and enhancing security and privacy compliance. In evaluating whether prioritizing aggregated (PHA) over granular (PHG) data compromises predictive performance, I compare both models and find that the PHG model exhibits improved predictive power. However, this introduces a nuanced trade-off: we must be mindful of the privacy and technological costs associated with granular data sharing. Further research is needed to fully understand these costs, but the effectiveness of PHA models alone, with their substantial predictive power, already presents a compelling argument for their application in loan screening.

Open Banking's potential in loan monitoring is significant yet underappreciated. Traditional credit scores often lag in reflecting a borrower's current status, usually updating for delays over 90 days and depending on other lenders' reporting efficiency. In contrast, payment history data, being independent and real-time, provides immediate and accurate insights into a borrower's financial health. My study examines the use of interoperable payment data in monitoring by adding post-disbursal PH variables to a comprehensive screening model. The findings reveal that payment data is exceptionally effective in generating early warning signals for loan monitoring. Specifically, the model incorporating PHA variables shows about an 11% increase in AUC within 90 days post-disbursal over the AUC of the pre-disbursal benchmark model.

Payment fintech loans, especially innovative sales-linked contracts where repayments are tied to merchant sales, provide unique insights into the information contents of the model in delinquency risk assessment. If fidn that the PHA model outperforms the Credit Bureau model, and combining both doesn't enhance predictive benefits. This suggests a potential redundancy of credit bureau data in such fintech contexts. Furthermore, fintech loans demonstrate a reduced reliance on lender soft information compared to traditional banking. Post-disbursal, PHA variables significantly boost predictive accuracy in fintech loans, but this may also highlight moral hazard issues unique to sales-linked contracts, as explored in studies by Rishabh and

Schäublin (2021) and Russel, Shi and Clarke (2023). This scenario presents a complex tension in fintech lending: while dependence on traditional credit bureau sources lessens, new challenges like moral hazard emerge.

Literature: My paper intersects with four major strands of literature: (1) Payment fintechs and SME lending, (2) Fintech lending and big data application, (3) Transaction-borrower screening and monitoring relationship, and (4) Open banking. Payment fintechs and bigtechs, such as PayPal, Square, Stripe, and Amazon, have transformed payment industries and ventured into MSME lending globally, offering a unique informational edge from borrower sales data (Bech and Hancock (2020); Petralia et al. (2019); Philippon (2016); Rysman and Schuh (2017)). My work uniquely contributes by exploring this informational advantage, particularly the use of transaction-level sales data for lending decisions.

While studies like Frost et al. (2019) have examined bigtechs' use of transaction data for internal credit scoring, my paper differs significantly. I delve into granular transaction data to identify key variables behind transaction-based scoring's effectiveness and assess the potential of transaction data in substituting credit bureau scores. Additionally, my focus on SME lending in a developing country context, alongside the unique setup where lender decisions are independent of credit scores, sets my study apart from works like Berg et al. (2020) and Agarwal et al. (2021), which explore digital and mobile footprints in consumer lending. My study stands out in its emphasis on using payment histories for loan monitoring, a critical aspect in environments with post-contractual challenges like poor enforcement. This focus aligns with the broader inquiry into the role and implications of open banking in enhancing lending practices.

My work contributes to the well-established connection between payment transactions and lending, a concept historically framed through the checking account hypothesis. This hypothesis suggests transaction accounts contain valuable insights for assessing borrower creditworthiness, both pre- and post-loan (Black (1975); Fama (1985); Nakamura (1993)). While studies like Mester, Nakamura and Renault (2007) and Norden and Weber (2010) have explored the monitoring role of transaction data in developed countries, my research extends these findings to a developing country context with uncollateralized loans, relying solely on transaction account activity. This approach underscores the versatility of transaction data in varied lending environments.

In the realm of loan screening, I build on the findings of Puri, Rocholl and Steffen (2017), which show bank customers with transaction accounts are more likely to obtain credit and less likely to default. Among more recent studies, Ouyang (2021) finds similar results in the context of household fintech loans in China. My study goes further by examining how lenders use transaction data for underwriting and determining its most informative aspects. A unique aspect of my research is its focus on transaction history from a source independent of the lender, crucial in the context of Open Banking. This novel approach allows me to directly investigate the value of interoperable payment data in lending, thus addressing a critical gap in the existing literature.

In a complementary study, Ghosh, Vallee and Zeng (2021) delve into how lenders not directly involved in the payments industry can leverage historical cashless transaction data for loan underwriting. This approach mirrors the concept of open banking, where lenders utilize past bank statements to gauge a borrower's creditworthiness. The study reveals that borrowers with bank statements reflecting a higher volume of cashless transactions have a better chance of obtaining loans, indirectly highlighting the value of interoperable payment data. My research builds on this by providing direct, quantifiable evidence of the utility of such data. Additionally, I expand the scope by examining loan monitoring using payment data, and separately exploring the unique aspects of sales-linked fintech loans.

My research contributes to the nascent but rapidly expanding literature on open banking. Theoretical works by He, Huang and Zhou (2023) and Parlour, Rajan and Zhu (2022) have advanced hypotheses on the potential impacts of open banking on payment service pricing and the structure of the credit market. Empirically exploring these theories, Nam (2023) investigates the German credit market, particularly focusing on personal loans. Unlike these studies, which presume the efficacy of payment data in default prediction, my paper empirically establishes this premise. Moreover, my analysis extends beyond personal loans and into the realm of firm lending within a developing country context, thus broadening the understanding of open banking's potential impact and applications.

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. In Section 4, I present the baseline results focusing on predictive analysis based on historical data in the context of loan screening. This section further delves into results pertaining to borrower heterogeneity, the privacy-accuracy trade-off, and the importance of various features. Additionally, it extends the analysis to include early warning exercises for loan monitoring. Section 5 revisits and adapts the analysis for sales-linked loans offered by fintech lenders. Finally, Section 6 provides a conclusion to the study.

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 company specializes in providing Point of Sale (POS) systems predominantly to Micro, Small, and Medium Enterprises (MSMEs). In this arrangement, the fintech firm processes payments received by client MSMEs through their POS devices. Additionally, the company operates a lending program in collaboration with various Non-Bank Financial Companies (NBFCs).[2] This study utilizes transaction data at the

[2]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

swipe level from all merchants utilizing the fintech's services from January 2015 to February 2019. I also accessed borrowing records from 2017 to 2019, encompassing loans obtained from the payment fintech as well as *bank loans* secured by the *same borrowers*. Below, I provide a detailed description of these two types of loans.

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. 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. This approach 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 this period. Given that the most 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 regulatory categories indicative of 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 payment history by merging their loan information with payment 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.

Within the limits of the available payment data, this study focuses on 11,972 bank loans issued from June 2015 to February 2019. To comprehensively understand the credit histories

regulated by the Reserve Bank of India.

of the borrowers in this study, I delve into the performance of 130,101 loans they obtained, stretching back to 1991. This in-depth historical analysis enables the calculation of key variables that reflect the borrowers' previous borrowing patterns at the point of taking out a new loan. This approach simulates the lender's review process of credit bureau records during loan application processing, acknowledging that only partial credit histories might have been available at that time. The variables calculated include the number of loan and credit card accounts the borrower had closed prior to receiving the loan under consideration. Additionally, the analysis categorizes these closed loans to ascertain how many were secured (such as mortgages, gold loans, or vehicle loans) and quantifies the number of active loans at the time the new loan was approved. For a detailed account of these variables, see 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. During the study period, one partnering NBFC contributed to over 80% of all loans. My analysis focuses on loans made by this predominant NBFC partner due to its more standardized sales-linked loan policies. 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 process initiated with the company evaluating merchants based on their historical transaction data, followed by sharing this information with the partnering NBFC. The NBFC then decided on the feasibility of extending a loan offer to a merchant, including the proposed loan amount. When a merchant was identified as a potential borrower, the payment fintech presented the loan offer on behalf of the NBFC. The merchant had the option to accept or decline this offer. Upon acceptance, the NBFC proceeded with the loan disbursement, usually within a few days, subject to any additional verifications deemed necessary.

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.[3] Additionally, merchants had the flexibility to repay the loan early, either in full or partially, through direct lump-sum payments to the company.[4]

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 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

[3] 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.

[4] 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).

scores in their lending processes.[5] Interestingly, Mishra, Prabhala and Rajan (2022) notes 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 cash inflows 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 comparing the predictive performances of different models, we delve into the heart of our research questions, which focus on the informational value inherent in various types of variables.

[5] 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).

For example, we determine the incremental insight payment history offers over traditional loan underwriting models by contrasting a traditional model's predictive performance with one that combine traditional model with payment history. An improvement in the latter's predictive performance would underscore the significance of payment history as a predictor of delinquency.

Building on this approach, we have developed several models to address our research questions, 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 w/o loan terms*, enriches the credit bureau data with demographic details of the borrowing firms, such as location, industry, and owner's age. The second, *Traditional w/ loan terms*, 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 second PH model, *Payment History Granular (PHG)* model, extends PHA, integrating transaction-level data and district-level sales benchmarks to provide a more detailed analysis. While PHG is inherently rich in information, its implementation may entail certain costs. For instance, the use of PHG could require more stringent regulatory oversight in data sharing and an increased effort from financial institutions for its effective collection and dissemination. Additionally, PHG's comprehensive nature might result in variability in data quality across different institutions, highlighting the importance of standardized APIs for consistent data sharing. A significant consideration is that the dissemination of granular transaction-level information could raise privacy concerns. These factors collectively pose a trade-off, balancing technological or privacy costs against the potential for enhanced predictive power.

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, as illustrated in the example provided at the beginning of this section. When integrating a PH model with the Traditional with Loan Terms (Trad w/ loan terms) 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. To delve into the privacy-predictability trade-off, I contrast models incorporating the PHG with those using the PHA. For a detailed account of the variables and models employed in this analysis, please see Table A1 and Table A2.

To explore the utility of payment history in monitoring loans, I enhance the Trad w/ loan terms + 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 w/ loan terms + 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 w/ loan terms + 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) Required
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 w/ loan terms vs. Trad w/o loan terms
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?	(PH + Credit Bureau) vs. Credit Bureau; (PH + Credit Bureau) vs. PH
What additional insights do payment history variables bring to traditional models?	(PH + Trad) vs. Trad
What is the efficacy of payment history in early warning and loan monitoring?	(PH + Trad + PH DSD) vs. (PH + Trad)
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 w/ loan terms or Traditional w/o loan terms.

3.2 Predictive Methodology

To predict loan delinquency, I split the sample into a training set and a test set. I then train a machine learning algorithm on the training set, which includes the variables relevant to the selected model. After the training phase, I use the algorithm to predict delinquency, serving as the response variable, 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. I also calculate the 95% confidence interval for the AUC using bootstrapping methods with 1000 replicas of the test set.

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 in 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 recall at each threshold level. Recall, or the True Positive Rate (TPR), refers to the proportion of actual delinquent loans that the model correctly identifies. 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.

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.

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

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. For detailed variable description see Table A1. All nominal monetary variables are denominated in INR. While logged monetary variables may appear unitless, their underlying values are based on amounts in INR.

Variable	Mean	Median	Std Deviation	p10	p25	p75	p90
Payment Variables							
Sales growth	2.78	-0.04	15.01	-0.74	-0.41	0.50	1.91
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	0.10	0.06	2.06	-0.05	0.01	0.12	0.20
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	0.01	0.00	0.04	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	7.63	5.00	9.48	0.00	2.00	10.00	18.00
# 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	0.06	0.00	0.14	0.00	0.00	0.04	0.20
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)	16.85	14.50	8.15	9.00	10.00	24.00	25.00
Loan tenure (Months)	17.50	12.00	16.99	4.00	9.00	24.00	40.00
Outcome Variables							
Delinquent (1 = Yes)	0.09	0.00	0.29	0.00	0.00	0.00	0.00

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 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. While logged monetary variables may appear unitless, their underlying values are based on amounts in INR. 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

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 a subsequent section.

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 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 shows an AUC of 0.59, surpassing the random-guess benchmark of 0.5. The AP is 0.07[6].This indicates that credit bureau data has some predictive power regarding loan delinquency. However, what about traditional models? The Traditional model, which excludes loan terms (AUC = 0.62, AP = 0.14), demonstrates an improvement over the Credit Bureau-only model. This suggests that integrating demographic variables with credit bureau information can be beneficial for lenders.

In scenarios involving small business loans, lenders often lack detailed financial accounts of the borrowing firms, relying instead on hard information like credit bureau and demographic data. Yet, they also invest effort in gathering soft information about borrowers. This blend of soft and hard information plays a crucial role in shaping loan contractual terms. To gauge the impact of soft information, we compare the Traditional model with loan terms (encompassing both soft and hard information) to the Traditional model without loan terms (based solely on hard information). We find that the inclusion of lender soft information enhances predictability by 11% in terms of AUC and 57% in terms of AP. This underscores the significant contribution of lender soft information in the realm of bank loans.

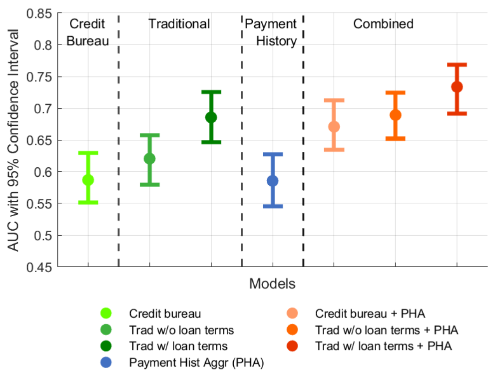
It is crucial to note a key limitation in our analysis of bank loans. We do not have access to the complete set of information utilized by the banks, which may include additional hard information. Consequently, the observed performance differences between the models might

[6]It should be noted that while the average precision may seem low, it is within expected ranges for this type of predictive modeling. For instance, in the study by Fuster et al. (2022), which also employs AP as a performance criterion, the most comprehensive model for predicting delinquency in the U.S. mortgage market reported an AP close to 0.06.

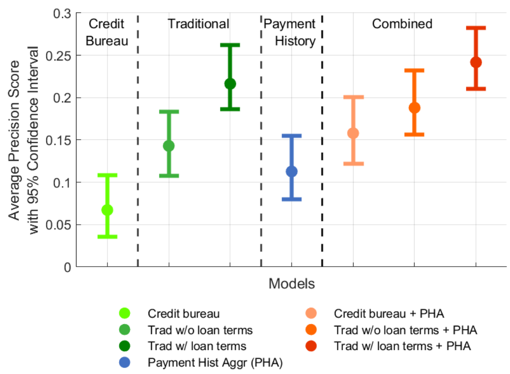
Figure 1: Bank Loans: Predictive Model Performance Comparison

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.

(a) Area Under the ROC Curve



(b) Average Precision



not solely reflect the impact of lender soft information; they could also be influenced by hard information that remains unobserved to us. However, given the context of small business lending, where there is typically limited additional hard information beyond what we have already accounted for, it is reasonable to infer that the predominant factor contributing to the performance disparity between the two traditional models is the lender's soft information.

Turning our attention to the value of aggregate payment data, we first examine the performance of the PHA model. Utilizing only the four PH variables, the PHA model achieves an AUC of 0.59, matching that of the Credit Bureau model, and an AP of 0.11, a 57% improvement over the Credit Bureau model. To determine if the Credit Bureau and PHA capture distinct types of information relevant for loan performance, we analyze a combined model of PHA and Credit Bureau and compare it to each model individually. The combined model enhances the AUC by approximately 13.5% compared to each model on its own. In terms of AP, the improvement is 45% over PHA alone and more than double compared to the Credit Bureau model alone. These figures suggest that payment history and credit bureau data are complementary in bank lending, as their combination enhances predictability beyond what each achieves individually.

Shifting focus to a critical comparison, we evaluate the integration of PHA with traditional models against their individual components. This analysis is pivotal to discern whether PHA adds unique information beyond what lenders traditionally acquire through hard, or both hard and soft, information. Initially, we juxtapose the Traditional model without loan terms, augmented with PHA, against the standalone Traditional model without loan terms. The results are noteworthy, revealing improvements of about 11% in AUC and 35% in AP — both considerable enhancements.

The subsequent comparison delves into the addition of PHA to the comprehensive Traditional model with loan terms. Here, we observe a boost in predictive performance: approximately 6% in AUC and 9% in AP. These findings suggest that lenders try to offset the lack of hard information with soft information, because the incremental gains from PHA are less pronounced in the model with loan terms compared to the model without loan terms. Crucially, however, our results also highlight that the current blend of lender's soft (and hard) information cannot fully substitute the value that payment history contributes. This is evidenced by the significant improvement when PHA is incorporated into the Traditional model with loan terms. The enhancement driven by PHA's inclusion implies that its absence in the initial models left room for considerable improvement in predictive accuracy.[7]

We summarize our main findings of this section as follows:

Takeaway 1 (a) *The Payment History Aggregate (PHA) model matches or outperforms the Credit Bureau model. When combined, they enhance predictability in bank loan delinquency, indicating their complementary nature.*

(b) *PHA adds significant value beyond traditional hard and soft information. Its integration into traditional models yields a 6% improvement in AUC and 9% improvement in AP compared to the traditional model with loan terms, underscoring the substantial potential benefits of interoperability of payment history data in loan screening.*

4.2 Predictive Performance Across Heterogeneous Borrowers

We next investigate if the outstanding performance of the Payment History (PH) models is attributable to specific types of firms. We particularly examine whether this performance consistency persists across different firm sizes and credit score categories. To explore variations among borrower sizes, I replicate the baseline analysis on two distinct subsets of borrowers—categorized as 'small' and 'large' based on their total transaction values within the 90-day period preceding loan disbursal. Small borrowers are those below the median transaction value, while large borrowers exceed it.

Figure 2 showcases the performance of select models segmented by firm size, while comprehensive results for all models are detailed in Table 4. Noteworthy observations emerge from the data. Initially, we find that payment history is a stronger predictor for small borrowers than for large ones. This disparity may stem from our data's reliance on electronic transactions from a single payment fintech, suggesting that larger firms likely engage in a broader array of electronic transactions beyond our dataset. Additionally, the Credit Bureau model exhibits enhanced performance for larger borrowers compared to smaller ones. Despite this, the PHA

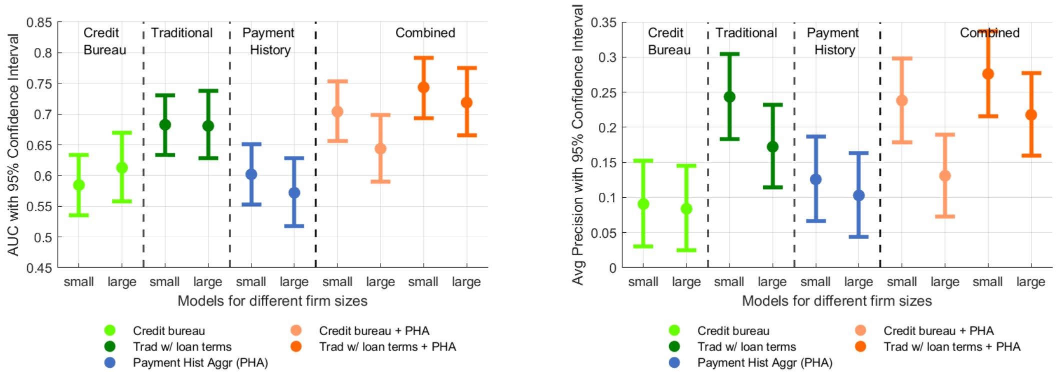
[7]The improvement driven by the inclusion of PHA also suggests that banks did not initially use payment history extensively, as its addition to the Traditional model with loan terms would have resulted in much smaller or negligible enhancement.

model, even with its relatively lower predictive strength for larger borrowers, still significantly boosts the effectiveness of traditional underwriting models. This is clearly demonstrated by the improvements in model performance when PHA data is integrated, as opposed to when it is excluded. Finally, the baseline results are consistent for small borrowers, with the influence of the PHA being notably more pronounced in this group.

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. 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



To study whether the performance differences are driven by differently scored borrowers, I divide the loan sample into those where the borrower had a *low* credit score and those with a *high* credit score. I classify the borrowers above 700—the threshold score[8] considered to be the demarcation between good and bad credit score—as high-scored borrowers and borrowers below 700 as low-scored borrowers. One may argue that for borrowers with high credit score, Credit Bureau model, and therefore the traditional models, should have a higher predictability because because these borrowers may have longer credit history for instance and a better integration with the credit market providing more information to the bureau and the lender. The question is does payment history brings in additional information for both high-score and low-score borrowers?

Due to the skewed distribution of credit score, with a notably smaller number of low-score borrowers, traditional training and test set splits could lead to biased model evaluations. To circumvent this, I utilize a five-fold cross-validation within the random forest algorithm. This method divides the dataset into five equal parts, where each part sequentially serves as the

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

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)	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)							
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	0.24	0.13	0.24	0.24	0.28
Large	0.08	0.11	0.17	0.10	0.13	0.17	0.22
[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

test set once and as part of the training set four times. This rotation ensures every data point is tested exactly once and appears in the training set four times, providing a balanced and comprehensive evaluation across all borrower categories. This approach not only prevents overfitting but also guarantees a fair representation of all groups in our predictions.

Figure 3: Bank Loans: Predictive Model Performance Comparison – by Credit Score

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.

(a) Area Under the ROC Curve

(b) Average Precision

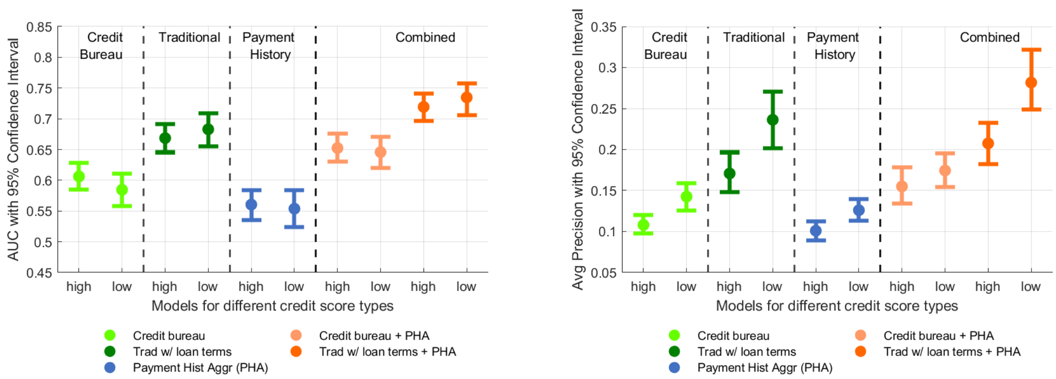


Figure 3 presents the predictive performance of selected models. Although no consistent patterns emerge in the relative effectiveness of credit bureau and traditional models between high-score and low-score borrowers, it's evident that the PHA model enhances predictability for both groups. This suggests that incorporating PHA is beneficial across the board, regardless of whether traditional lenders possess extensive or limited information about a borrower's credit history.

We are also keen to assess if the Payment History Aggregate (PHA) is beneficial for borrowers with minimal or no credit history. To explore this, our attention turns to loans granted to individuals who either had no credit score at the time of borrowing or lacked a history of past borrowing, a group we refer to as 'thin-file borrowers.' The findings of this analysis are detailed in Table 5. Our analysis shows that PHA integration into the traditional model, for thin-file borrowers does improve its performance in terms of the Area Under the Curve (AUC). However, the extent of this improvement (4.6%) is somewhat less than what we observed for both high-score and low-score borrowers (approximately 7.4%).

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)	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)			
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	0.73
Thin File		0.55	0.65	0.59		0.63	0.68
				Average Precision			
High Score	0.11	0.13	0.17	0.10	0.16	0.17	0.21
Low Score	0.14	0.16	0.24	0.13	0.17	0.23	0.28
Thin File		0.12	0.20	0.16		0.18	0.20
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

Interestingly, when we consider Average Precision (AP), the addition of PHA to the Traditional model without loan terms leads to improvement for thin-file borrowers, but this enhancement is not observed when PHA is added to the Traditional model with loan terms. This suggests that in terms of AP, while PHA generally benefits both low-score and high-score borrowers, its utility for thin-file borrowers is more pronounced when lenders depend predominantly on hard information.

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

Takeaway 2 (a) *The Payment History Aggregate (PHA) model demonstrates significant predictive power for small borrowers, surpassing its effectiveness for larger borrowers. Notably, PHA also enhances traditional underwriting models across both borrower segments, underscoring its wide-ranging utility.*

(b) *PHA is beneficial for both high-score and low-score borrowers, demonstrating its effectiveness across varying credit scores. For thin-file borrowers, who have minimal or no credit history, PHA's integration results in a 4.6% increase in AUC, slightly lower than the 7.4% improvement seen for scored borrowers. Notably, the PHA model's impact is particularly pronounced for thin-file borrowers in contexts where lenders predominantly utilize hard information.*

4.3 Loan Screening with Granular Payment Data

We have enhanced our payment history models by incorporating more granular payment variables, specifically those that require transaction-level data or are benchmarked against district-level payment aggregates. This enhanced model is referred to as the Payment History Granular (PHG) model, and contains a total of 12 variables, including the four PHA variables. Our goal is to assess the improvement in predictive performance when using PHG models compared to Payment History Aggregate (PHA) models.

Figure 4 displays the performance of Credit Bureau and Traditional models, and further compares them with selected PH models, including both PHA and PHG. Comprehensive model comparisons are provided in Table 6. Our findings reveal that the PHG model significantly outperforms the PHA model in terms of both Area Under the Curve (AUC) and Average Precision (AP). Specifically, the PHG model shows a 4.7% improvement in AUC and a 10.3% increase in AP compared to the PHA model. Additionally, when PHG is integrated into combined models, such as the Traditional model with loan terms, it demonstrates approximately a 3.7% enhancement over the corresponding PHA-inclusive combination.

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

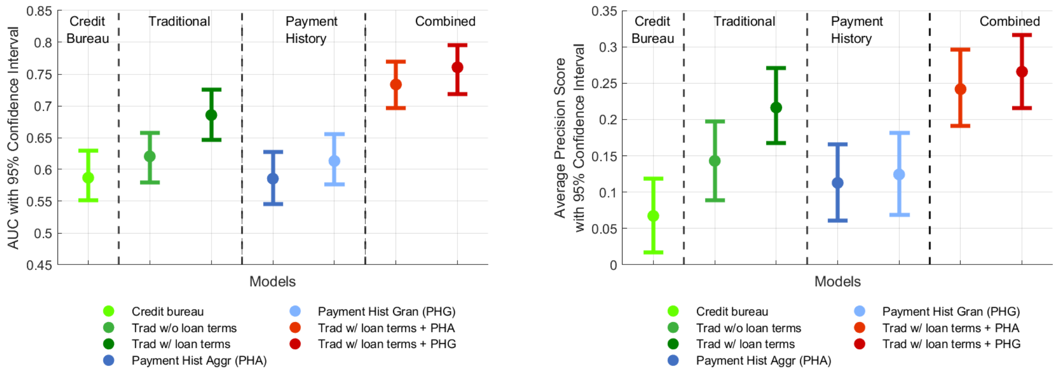
AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models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

		Traditional		Payment History		Models Combined with PHG		
Predicted var: Delinquency	Credit Bureau (1)	w/o Loan Terms (2)	w/ loan terms (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	0.69	0.59	0.61	0.69	0.70	0.76
% Δ compared to Agg model	-	-	-	-	4.73	2.19	2.09	3.68
Average Precision	0.07	0.14	0.22	0.11	0.12	0.20	0.20	0.27
% Δ compared to Agg model	-	-	-	-	10.28	27.85	7.86	9.94
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

Figure 4: 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. 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 extend PHA models with more detailed data, isn't surprising. However, their requirement for finer information raises privacy concerns and could incur significantly higher costs in terms of data compilation and sharing. Thus, it is crucial to weigh these potential costs against the benefits of transitioning from PHA to PHG models. Although a precise quantification of these costs might be the subject of separate research, it is clear that even if PHG models become less attractive due to an unfavorable cost-benefit balance, the argument for the interoperability of payment data remains strong. This is because PHA models already provide substantial predictive accuracy.

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 PHA models alone, with their substantial predictive power, continues to make a strong case for their use in loan screening.*

4.4 Predictor Importance for Screening

Our primary focus is to identify which variables are most crucial in predicting loan delinquency. In a complex, black-box algorithm like random forest, establishing clear relationships isn't straightforward, as these algorithms leverage non-linear data 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) Out-of-bag variable importance through permutation, and (ii) Shapley additive explanations. In our analysis, we concentrate on the most encompassing screening model—PHG combined with Traditional with loan terms. This model selection enables a thorough ranking of predictors and facilitates an assessment of the relative predictive value of granular versus aggregate payment variables and traditional variables.

4.4.1 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 5 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 'Loan amount' standing out as significant predictors. The 'Loan to sales ratio', a combined variable, emerges as the foremost predictor across all categories. Additionally, district-level variables stand out among the granular payment history variables, emphasizing their relevance in the model.

Figure 6 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.

A caveat with the OOB permutation measure is its tendency to overstate the importance of correlated features. This occurs because permuting a variable forces the model, already

Figure 5: Bank 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 with loan terms + 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.

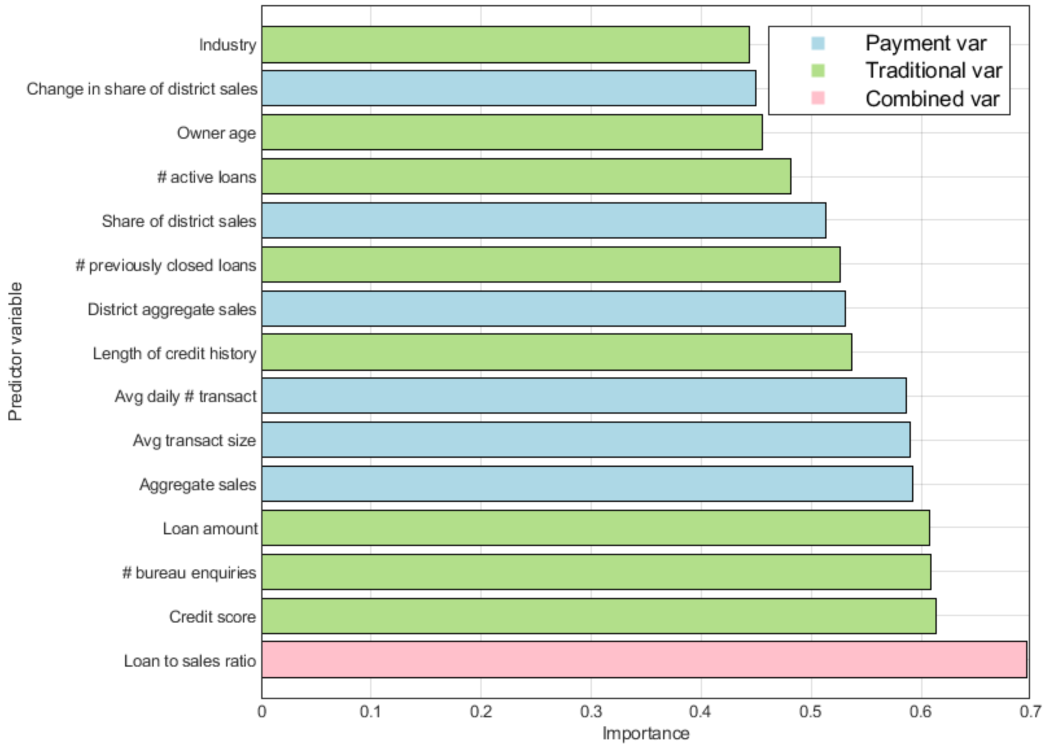
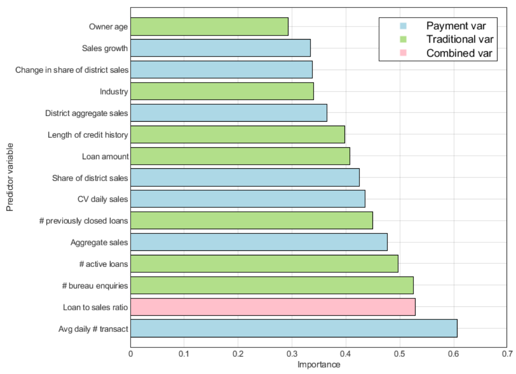


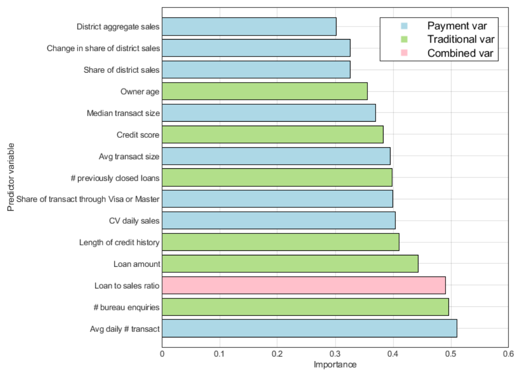
Figure 6: Bank Loans: Top 15 Predictors of Delinquency – by Size

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 with loan terms + 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



trained without such permutation, to make predictions in regions it has not been trained due to correlated features. This situation can lead to exaggerated prediction errors and, as a result, inflated perceived importance of the variable (Fisher, Rudin and Dominici, 2019; Gregorutti, Michel and Saint-Pierre, 2017). These 'unexplored' regions are less prevalent when features are not highly correlated. Additionally, to complement this approach, I utilize Shapley Additive Explanations (SHAP) as a measure. SHAP not only aids in gauging the importance of predictors but also elucidates the direction of their relationship with delinquency probability. We will delve into this aspect in the following discussion.

4.4.2 Shapley Additive Explanations (SHAP)

Shapley Additive Explanations (SHAP), developed by Lundberg and Lee (2017), is a local measure that, unlike global measures such as OOB permutation, explains individual predictions. SHAP values show how much each predictor pushes a prediction away from the average prediction. Features with larger SHAP values contribute more to the specific prediction. The concept of fair contribution computation is based on treating prediction as a cooperative game, where the players (predictors) work together to create a surplus (the deviation of the prediction from the average). The Shapley value quantifies each player's (predictor's) contribution to the prediction. A positive SHAP value for a predictor at a given instance suggests that the predictor increases the probability of delinquency for that specific instance. Conversely, a negative SHAP value indicates that the predictor decreases the probability of delinquency for that instance.

To manage computational demands, I compute SHAP for a random 50% sample of the test set, which encompasses roughly 1200 predictions. This approach provides a practical balance between computational efficiency and interpretive detail.

Figure 7: Bank Loans: Shapley Additive Explanations Summary Plot

SHAP values show how much each predictor pushes a prediction away from the average prediction. Predictors with larger SHAP values contribute more to the specific prediction. Computation is based on the most extensive screening model PHG + Traditional with loan terms. See Table A1 for variable details and Table A2 for model composition.

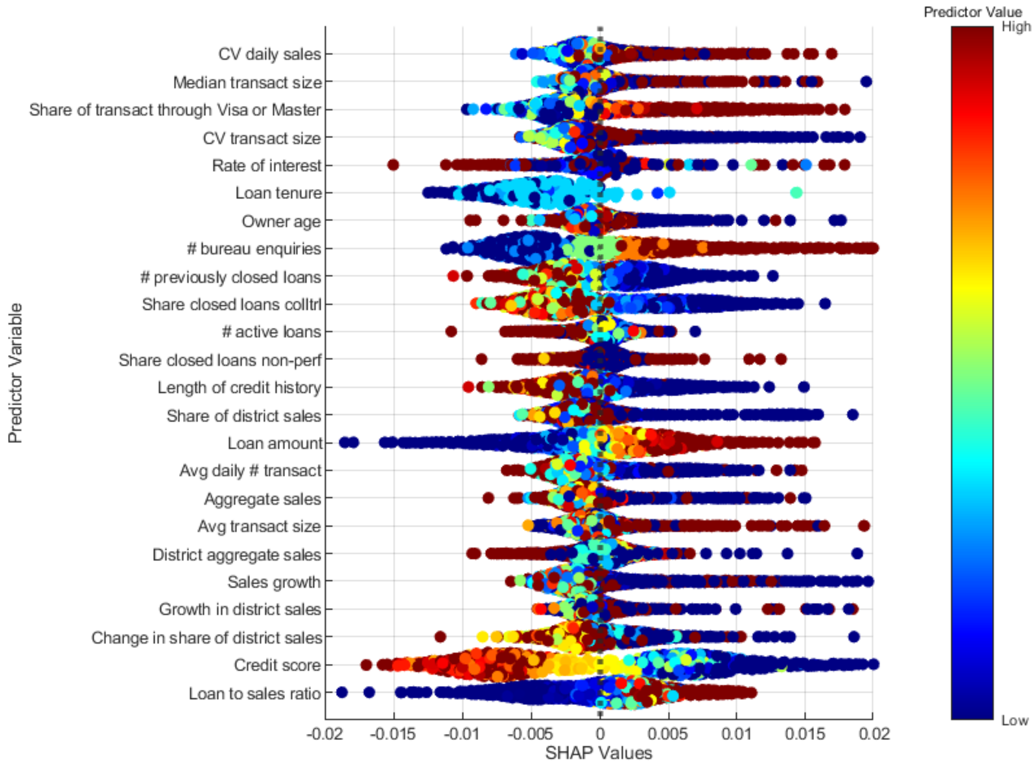


Figure 7 displays a SHAP summary plot for around 1200 out-of-sample predictions, merging predictor importance with their effects. Each point represents a Shapley value for a predictor at a specific instance. The color gradient illustrates the predictor's value from low to high. Points are jittered vertically to show the distribution of Shapley values for each feature. To identify key predictors, we look for those with wider spreads along the x-axis, as larger absolute Shapley values indicate a more significant deviation from the average prediction. Aligning with permutation measures, predictors like loan-sales ratio, sales growth, credit score, number of credit bureau enquiries, loan amount, and share of transactions through Visa or Mastercard are crucial for predicting delinquency.

To investigate the direction of relationship between predictor values and the probability of delinquency, we scrutinize the color coding on the predictor values in Figure 7, relating it to their SHAP values. For instance, the figure reveals that a higher loan-sales ratio, marked in red, is associated with elevated SHAP values, indicating a positive direction of relationship with an increased probability of delinquency. The summary plot provides initial insight into the directional relationship between the value of a feature and its effect on the prediction. For a more precise delineation of this relationship, however, SHAP dependence plots are more useful.

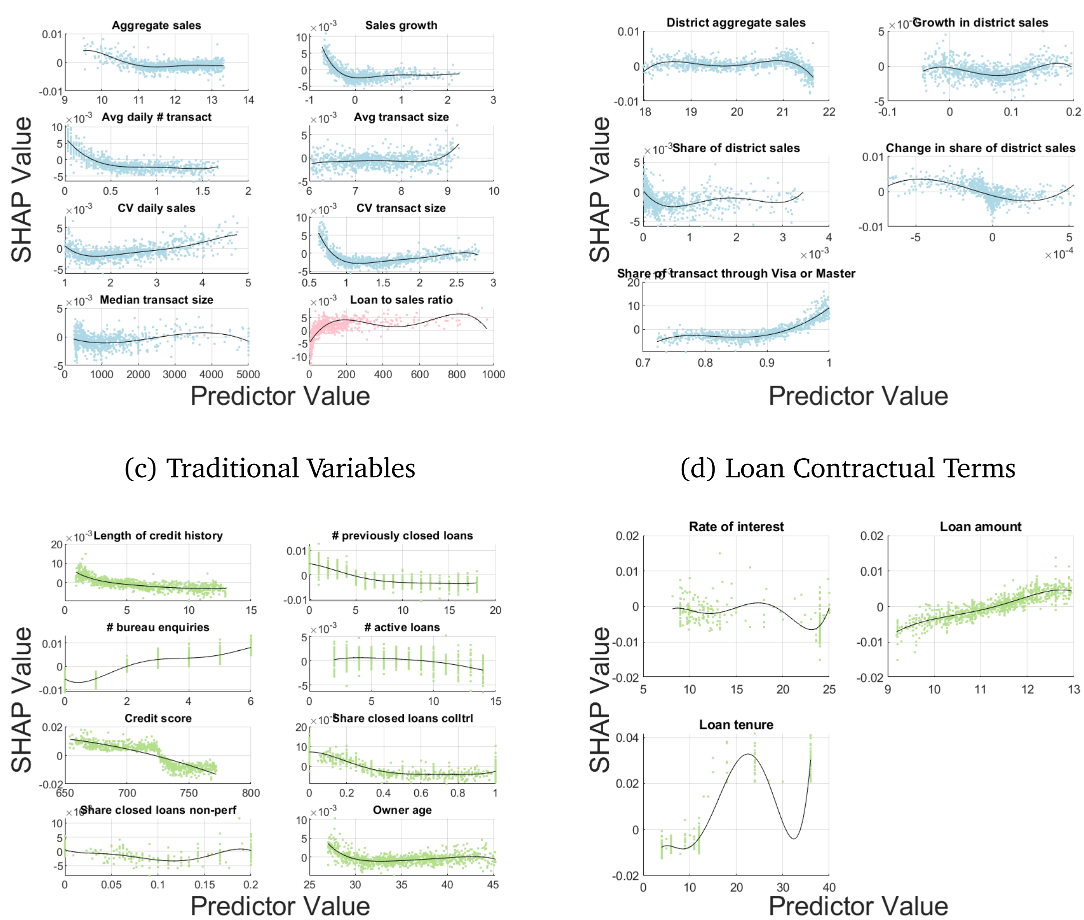
Figure presents feature dependence plots that more explicitly map SHAP values against

Figure 8: Bank Loans: SHAP Dependence Plots

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 is based on the comprehensive PHG + Traditional model with loan terms. For detailed variable information, see Table A1, and for model composition, refer to Table A2

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

(c) Traditional Variables (d) Loan Contractual Terms



predictor values. For each predictor, I have fitted a polynomial of degree N, where N could take a value from the range 1 to 5, selecting the optimal degree based on the adjusted R-square. A polynomial degree above 1 indicates a non-linear relationship. The plots reveal that payment history aggregate variables, such as aggregate sales, sales growth, and the average daily number of transactions, exhibit a negative relationship with the probability of delinquency. Conversely, while average transaction size has a lesser effect, its impact turns positive at higher values. Sales variability, represented by the coefficient of variation of daily sales and transaction size, shows a 'U-shaped' relationship with delinquency probability. Among traditional variables, credit score demonstrates a negative relationship with delinquency, whereas the number of bureau inquiries has a positive relationship.

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 variable interdependencies 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) *The Out-of-Bag (OOB) method highlights the significant role of payment history variables, particularly the Payment History Aggregate (PHA) variables like Aggregate Sales, Average Transaction Size, and Average Daily Number of Transactions, in predicting delinquency. This observation reinforces the notion that a significant portion of the loan screening capability within payment history variables is already encapsulated in these aggregative variables.*

(b) *SHAP analysis clarifies that PHA variables like Aggregate Sales, Average per-day number of transactions, and Sales Growth negatively correlate with delinquency probability, while Average Transaction Size positively impacts delinquency probability. Credit Score, a traditional variable, negatively influences delinquency probability, in contrast to the positive effect of Bureau enquiries. Higher value of the combined variable, Loan to Sales Ratio, significantly heightens delinquency probability.*

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. To evaluate how payment history variables might serve as early warnings for ongoing loans, I conduct a predictive analysis at six 30-day intervals following loan disbursal. This involves augmenting the pre-disbursal 'PH + Traditional w/ loan terms' model with additional post-disbursal payment history variables, calculated within each respective time window since disbursal. The analysis includes both aggregate (PHA) and granular (PHG) versions of these post-disbursal payment variables, which are then compared to the benchmark pre-disbursal model.

Figure 9: Bank 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 'PH + Traditional w/ loan terms' model with additional post-disbursal payment history variables, calculated within each respective time window since disbursal (days-since-disbursal(dsd)).

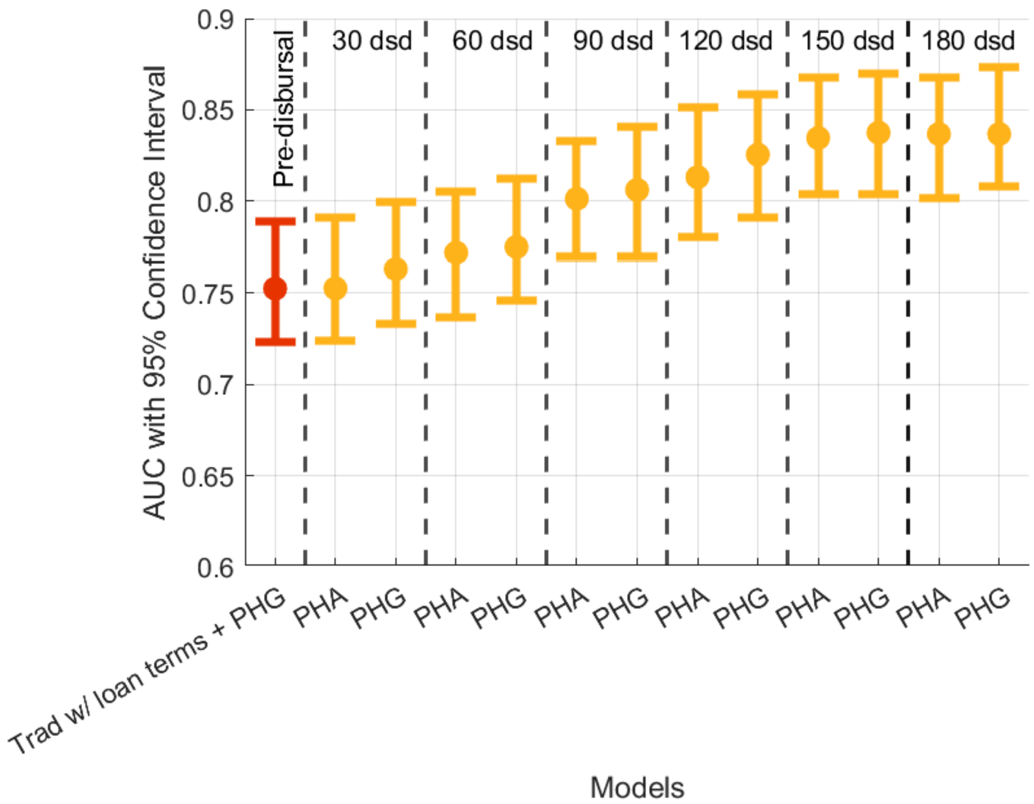
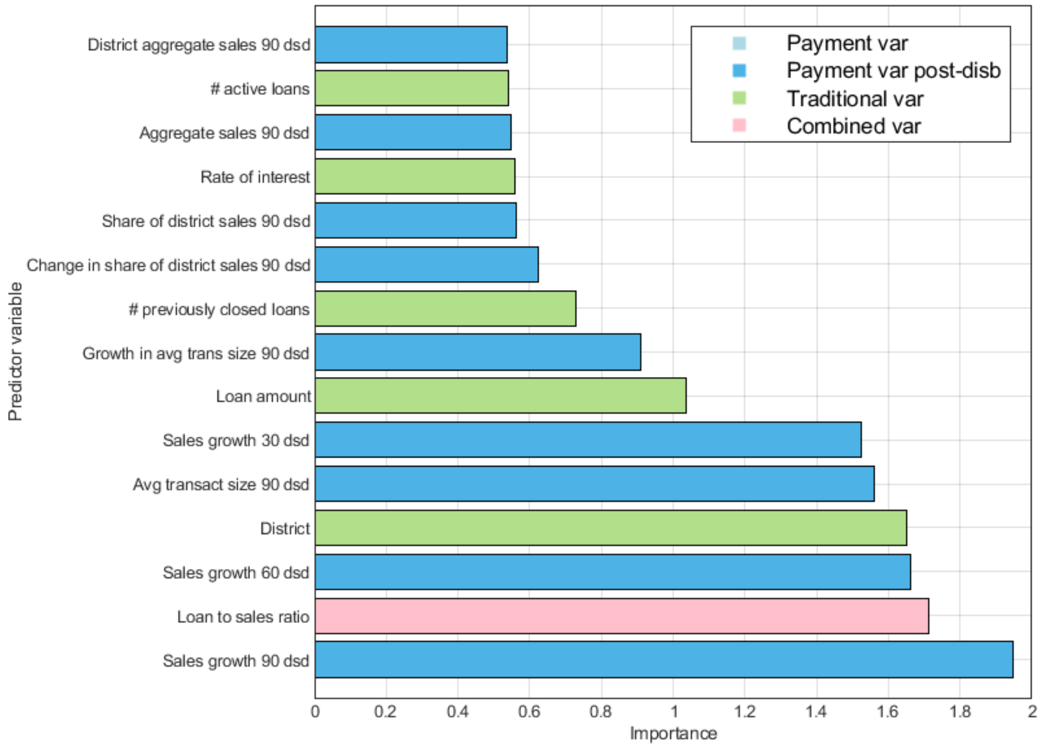


Figure 9 displays the outcomes of early warning predictive analysis using both PHA and PHG variables. The analysis shows a consistent increase in AUC over the post-disbursal period, with the rate of improvement diminishing around 150 days since disbursal (dsd). This plateau in performance enhancement can be attributed to the already high levels of accuracy achieved by this stage, where significant further improvements may require more extensive data sources. Notably, the PHA model exhibits an approximate 11% improvement in AUC by 90 dsd compared to pre-disbursal.

Figure 10: 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 'Traditional with loan terms + 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.



In assessing the contribution of post-disbursal variables, the OOB permutation measure for the PHG model at 90 dsd + Traditional with loan terms reveals their importance. This model includes 44 variables. Figure 10 highlights the top 15 variables by importance. Two key findings are evident: firstly, post-disbursal payment variables rank higher than pre-disbursal PH variables; secondly, post-disbursal PHA variables, particularly growth in sales, demonstrate the greatest importance, underscoring their value as early warning indicators.

Takeaway 5 (a) *Payment data proves to be highly effective for generating early warning signals, offering valuable assistance to lenders in monitoring loans. The post-disbursal* *early warning model based on payment history aggregate (PHA) variables shows a significant improvement in predictive accuracy, with an approximate 11% increase in AUC within the first 90 days following disbursal, compared to the pre-disbursal benchmark model.*

(b) *In post-disbursal analysis, PHA variables, notably post-disbursal sales growth, prove to be key predictors.*

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 11 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 loan terms offers only marginal improvement over its counterpart without loan terms 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.

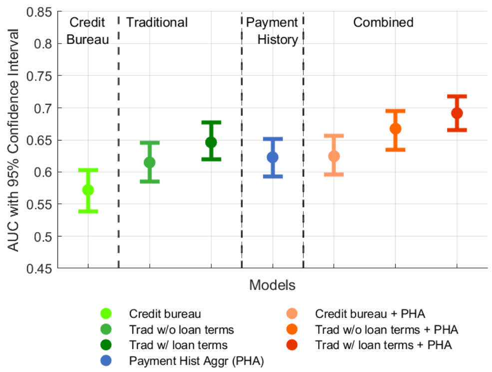
The performance of the Payment History Aggregate (PHA) model in the context of fintech loans is quite revealing. With an AUC of 0.62 and an AP of 0.2, the PHA model outperforms the Credit Bureau. However, the critical question is whether the PHA complements or substitutes the Credit Bureau. To explore 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.

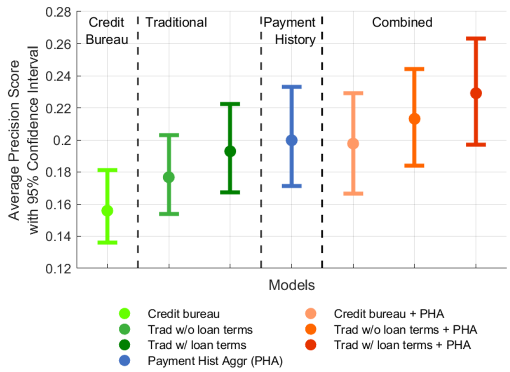
Figure 11: Fintech Loans: Predictive Model Performance Comparison

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. 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



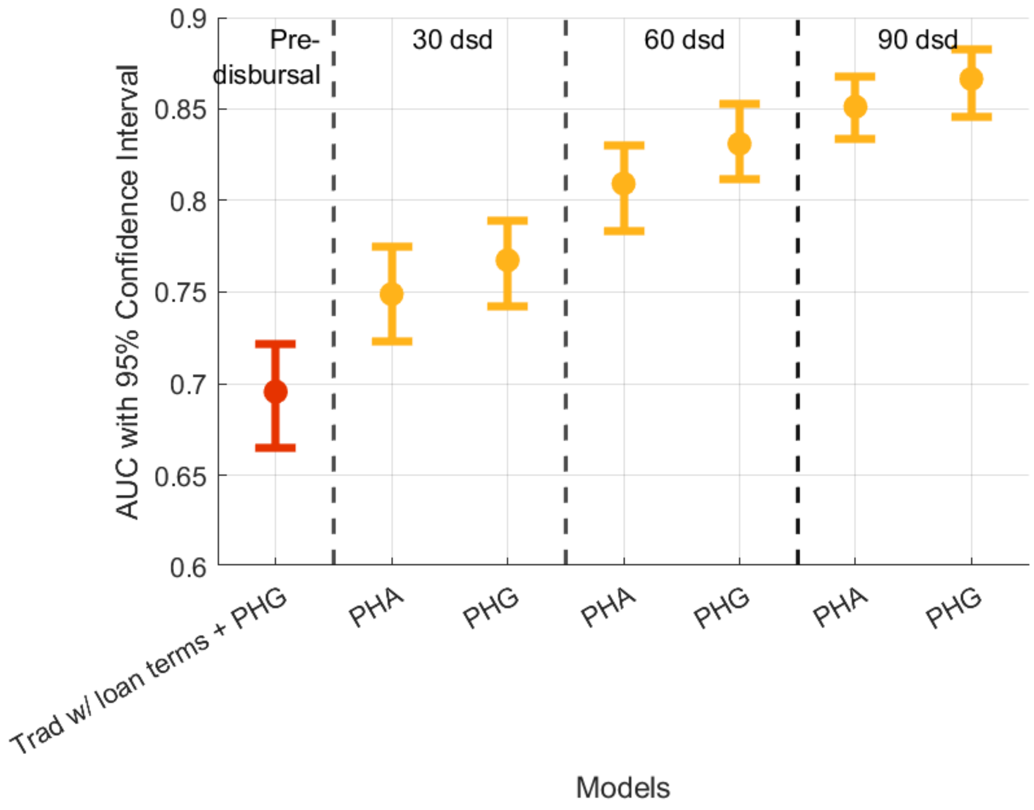
Incorporating Payment History Granular (PHG) variables into our fintech loan analysis reveals notable improvements. The PHG model, achieving an AUC of 0.65 as shown in Table **??**, outperforms the PHA model and even the traditional model with loan terms. This raises questions about the role of lender 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 with loan terms + 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 12: 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 'PH + Traditional w/ loan terms' model with additional post-disbursal payment history variables, calculated within each respective time window since disbursal (days-since-disbursal(dsd)).



The findings, presented in Figure 12, 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 loan terms 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 comparison of the predictive effectiveness of firm payment histories and traditional variables. This investigation spans a wide range of queries related to open banking, demonstrating its vast potential as a lending technology for both loan screening and monitoring. However, my results also unveil several nuances, offering important policy implications.

Firstly, establishing credit information sharing institutions like credit bureaus and registries 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 payment history data. 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. My study indicates a clear trade-off in this respect. More granular data may enhance predictive accuracy but raises significant privacy concerns. 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. 2021. “Fintech and credit scoring for the millennials: evidence using mobile and social footprints.” *Working Paper*.

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, Greg Buchak, and Will Gornall. 2022. “Customer data access and fintech entry: Early evidence from open banking.”

Bech, Morten L, and Jenny Hancock. 2020. “Innovations in payments.” *BIS Quarterly Review, March*.

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 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.

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*.

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

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

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. 2019. "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously." *J. Mach. Learn. Res.*, 20(177): 1–81.

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

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

Ghosh, Pulak, Boris Vallee, and Yao Zeng. 2021. "FinTech Lending and Cashless Payments." *Available at SSRN*.

Gregorutti, Baptiste, Bertrand Michel, and Philippe Saint-Pierre. 2017. "Correlation and variable importance in random forests." *Statistics and Computing*, 27: 659–678.

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.

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

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

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.

Petralia, Kathryn, Thomas Philippon, Tara Rice, and Nicolas Veron. 2019. "Banking disrupted? Financial intermediation in an era of transformational technology." *22nd Geneva Report on the World Economy*.

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*.

Rysman, Marc, and Scott Schuh. 2017. "New innovations in payments." *Innovation Policy and the Economy*, 17(1): 27–48.

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

Appendices

A Additional Figures and Tables

A.1 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 $\geq$ 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 the 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)	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)	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 the 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. While logged monetary variables may appear unitless, their underlying values are based on amounts 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	2.67	3.86	-1.19**
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	0.10	0.09	0.02
Median transact size	2764.85	3418.85	-653.99***
Aggregate sales (log)	11.26	10.71	0.55***
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	0.87	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	7.83	5.64	2.19***
# 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	0.06	0.05	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)	16.89	16.43	0.46
Loan tenure (Months)	16.66	23.76	-7.10***

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. While logged monetary variables may appear unitless, their underlying values are based on amounts 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)	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)	0.59	0.62	0.69	0.59	0.67	0.69	0.73
Average Precision	0.07	0.14	0.22	0.11	0.16	0.19	0.24
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

Area Under the 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.

						Models Combined with:	
		Traditional		Payment History		PHA	PHG
Predicted var: Delinquency	Credit Bureau (1)	w/o Loan Terms (2)	w/ loan terms (3)	Aggregate (PHA) (4)	Granular (PHG) (5)	Mod (3) + Mod (4) (8)	Mod (3) + Mod (5) (9)
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
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

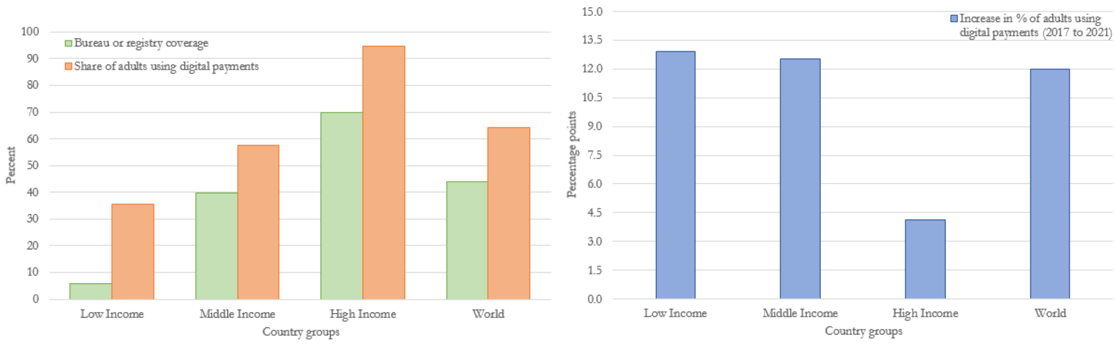
AUC varies between 0.5 (random guess) and 1 (perfect prediction) with higher values indicating better predictive power. Average Precision varies between 0 and 1, with higher values indicating better predictive power. For detailed variable description see Table A1. For the composition of predictive models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



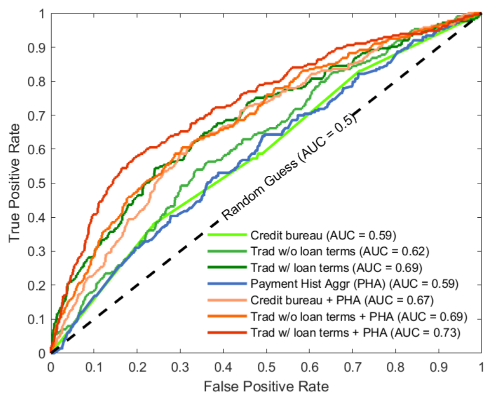
(a) Adults covered under credit bureau or registry, and adults using digital payments

(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)

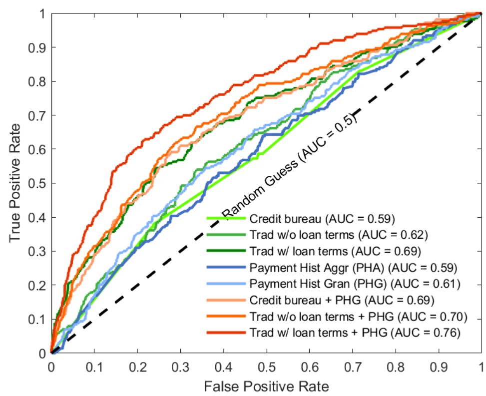


Figure A3: Fintech Loans: Top 15 Predictors of Delinquency

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 with loan terms + 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.

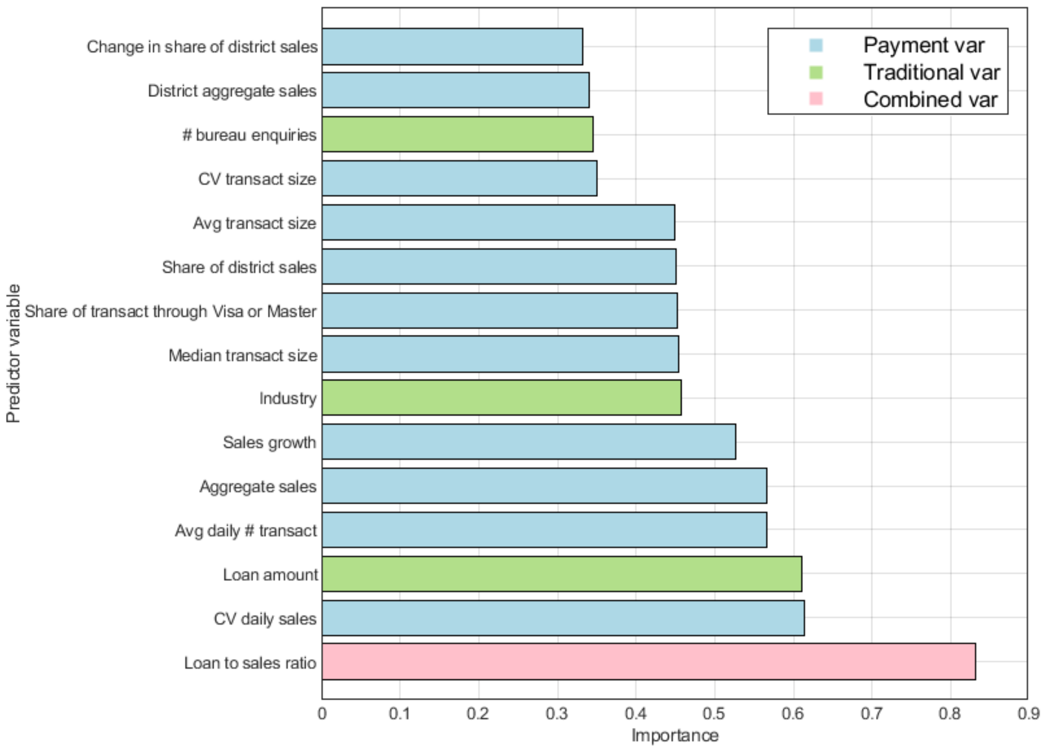


Figure A4: Fintech Loans: Shapley Additive Explanations Summary Plot

SHAP values show how much each predictor pushes a prediction away from the average prediction. Predictors with larger SHAP values contribute more to the specific prediction. Computation is based on the most extensive screening model PHG + Traditional with loan terms. See Table A1 for variable details and Table A2 for model composition.