Can Open Banking Substitute Credit Bureaus?*

Kumar Rishabh†

Abstract

This study investigates whether open banking can serve as a substitute for credit bureaus. To address the identification problem that arises when lenders screen and monitor borrowers using financial information and credit scores together, I use a unique dataset to compare the predictive performance of payment footprints and credit scores. Using machine learning algorithms to predict defaults, I find that payment footprints have significantly higher predictive power than credit scores. In fact, my results suggest that payment footprints can substitute credit scores, as combining credit scores with payment footprints does not improve the predictive performance of the model beyond that of payment footprints alone. These findings have implications for the replacement of costly credit-reporting institutions with open banking. Additionally, by utilizing interpretable machine learning techniques, I identify the most important payment footprint variables that predict default. I also demonstrate the early warning capability of payment footprints by showing that they contain valuable information for predicting credit default as early as 28 days after loan disbursal. Preliminary analysis of traditional loans also suggests that payment footprints possess impressive predictive power for these types of loans.

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), Neharika Sobti (discussant), Boris Vallée (discussant), and the participants at the 2023 AFA Annual Meeting (New Orleans), 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, Chennai Mathematical Institute - 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. All errors are mine.

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

1 Introduction

Payments represent both income and expenses for transacting parties, making payment flows a valuable source of financial information for lenders to assess credit risk. Open banking, a system of bank-data interoperability that allows customers to share their payment history with other institutions, has been widely endorsed as a gateway to improving credit risk assessment and enhancing financial access (BIS and World Bank, 2020). In this paper, I compare the effectiveness of open banking and traditional credit bureaus in achieving these objectives. Credit bureaus and credit registries are considered the "holy grail" for enhancing credit market efficiency (Djankov, McLiesh and Shleifer, 2007), making them a natural benchmark for studying the significance of open banking. Specifically, I investigate whether open banking can fully substitute credit bureaus for credit risk assessment.

Establishing a credit reporting infrastructure, such as a public credit registry or a private credit bureau, comes with enormous costs. As a result, more than half of the world's firms and individuals remain uncovered by any kind of credit reporting system. This lack of coverage is especially pronounced in developing and underdeveloped countries, as shown in Figure 1.

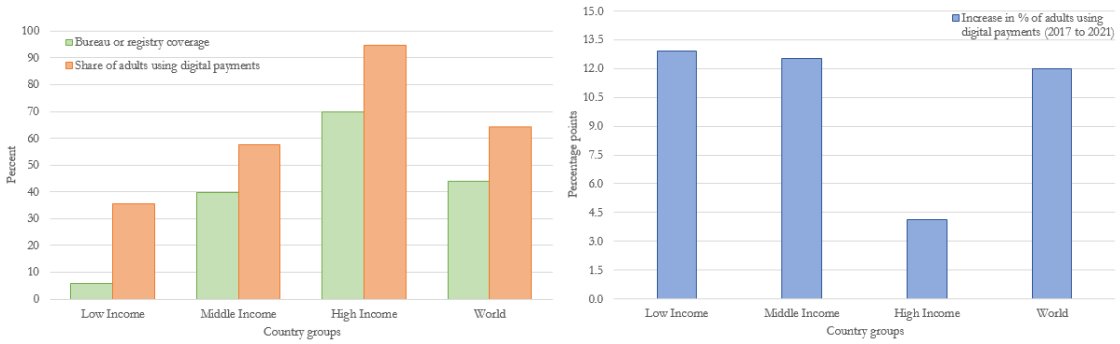
However, if the payment footprints of a borrower already contain the same information about their creditworthiness that their bureau credit score would convey, then payment footprints could serve as a substitute for costly credit bureaus. Electronic payments are increasingly prevalent and now account for a significant share of the economy, as shown in Figure 1. Additionally, payment footprints are a byproduct of a transaction rather than the primary product, making them relatively cheap to obtain, if not entirely costless.

Therefore, open banking presents a promising opportunity if we can demonstrate that payment footprints can indeed substitute for credit scores.

However, from the observable data it is quite challenging for a researcher to identify whether payment footprints substitute credit bureau. The reason is that lenders typically use both, credit scores and financial information such as payment history, to underwrite loans. In addition, lender may take into account a borrower's credit score while monitoring. For instance, the lender may monitor a borrower with low credit score more intensely than a borrower with high credit score. This difference in intensity of how the lender reacts to the different credit scores may reduce the performance difference between a low and a high-scored borrower. And that will eventually show up as a muted correlation between loan performance and credit scores, and thus, a low *ex-ante* discriminating power of credit scores.

Fortunately, I have a setting where these concerns do not apply. In the lending program that I am studying, the lender exclusively used payment footprints to underwrite loans. Specifically, it did not request borrower's credit score at any point in the life cycle of the loans under consideration. However, I have acquired the historical credit scores allowing me to study the relative predictive performance of payment footprints and credit scores, and to answer the question we set out to address. The focus of my study is the lending program of an Indian

Figure 1: Coverage Under Credit Bureau or Credit Registry and Use of Digital Payments



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

(b) Increase in the share of adults using 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.

payment fintech. Payment fintechs are technology based payment processing companies that are entering into lending markets. Payment fintechs screen borrowers based on the borrower's payment history. Banks, with access to borrower's transaction account data, should presumably also enjoy this screening ability. Payment fintechs are similar to the banks in the sense that, unlike other fintech lenders such as P2P platforms, payment fintechs offer both payment services and credit to the merchants using their payment services.

To answer our main question, I compare predictive performance of various *models* of credit default prediction. Where I define a model as a collection of variables. Credit score model contains two variables – credit score if a borrower is scored, and an indicator variable separating scored and unscored borrowers. The most basic of the payment footprint (PF) model is called *PF Basic* which includes high-frequency variables commonly available to a payment fintech. PF Basic model includes the following variables, among others: borrower's loan-sales ratio, average daily sales, average transaction size, and growth in average daily sales, and variability in sales. Where by *sales* I mean the value of electronic transactions. I also create a model which combines credit scores with other loan and business characteristics, that a typical lender using credit scores would, and call it a *Traditional extended* model.

I train models using the machine learning algorithm, Random forest, and make out-of-sample predictions. I compare predictive performances by plotting receiver operating charac-

teristics (ROC) curves for different models and calculating area under the ROC curves (AUC). AUC closer to one indicates better predictive performance, while AUC of 0.5 indicates that the model's predictive performance is only as good as a random guess. I find that in predicting default, credit score model has an AUC of 0.57 which is significantly higher than 0.5, however it is lower than the AUC of Basic PF model, which is 0.69. Even the Traditional extended model (AUC = 0.64) falls short in comparison to the Basic PF model. Basic PF model when combined with loan and business characteristics improves the AUC further to 0.72.

For the question on complement versus substitute, I append the PF Basic model with credit score model and check if it improves the predictive performance compared to the PF Basic model. I find that the combined model does not perform better than the PF Basic model. In fact, the point estimates of AUCs are the same across the PF Basic and the appended model, implying that credit score does not add any other information that is already captured in the basic payment footprints.

I also study heterogeneities across different types of borrowers. First looking across different firm sizes, I observe that PF based models continue to outperform traditional models for both large and relatively small firms. However, performance premium of the PF based models over credit scores is smaller for large borrowers compared to small borrowers. This indicates that primary beneficiaries of payment footprints are small firms. I run a similar test across low and high type of borrowers to test whether for borrowers with high credit score, credit score model has a higher predictability; arguably because because these borrowers may have longer credit history, for instance, and a better integration with the credit market. I find no difference across score buckets and compared to the baseline. Thus, payment footprints are more effective default predictors for both high scored and low scored borrowers. Our second baseline result about the substitutability of credit scores with payment footprints continues to hold across heterogeneous samples.

To identify the variables that are most important in yielding predictive power to our models, I use *interpretable machine learning* techniques. The first technique, called the out-of-bag permutation measure, works on the idea that if a feature (variable) is important then permuting its value (randomly shuffling the values of the feature across observations) will worsen the prediction error. Applying the method to the model that combines PF Basic variables with loan and business characteristics, I find that the top most important variables are predominantly payment footprint variables. Within PF variables, the most important ones are loan-sales ratio and growth of average daily sales. The variables on growth and level of average transaction size, and coefficient of variation of daily sales follow in importance. This shows that the informational advantage payment fintechs enjoy comes from their ability to use the short-term changes information on sales. Secondly, permuting the variables for the model combining PF Basic variables with credit scores we find that credit score variables show *negative importance* implying that credit score variables only add noise to the signal contained in the PF variables, confirming that they may be substituted by the PF variables.

The second technique, called Shapley additive explanation (SHAP) for a prediction instance assigns to each variable an importance according to its Shapley value in a coalition of variables (model) that generate the surplus defined as deviation of the prediction from the average prediction. SHAP confirms that the two most important determinants of defaults are growth of average sales and loan-sales ratio, followed by other variables mentioned above. SHAP also helps us gain insight about the direction of relationship between predicted variable and the features. SHAP shows that low growth and high loan-sales ratio are associated with positive contribution to the deviation, while CV of sales has a U-shape relationship.

Finally, another advantage that payment fintechs have is that they can observe changes in the borrower sales in real time even after loan disbursal. They can use these high frequency signals for an early warning exercise. To design one such exercise, I do the predictive analysis by augmenting all the PF based models with early-warning variables. Where early warning variables are the relative or absolute change in the PF variables calculated in the 28-day period after disbursal over their respective values calculated in the 28-day period before disbursal. I find that following payment footprints only 28 days after disbursal already improves the predictive performance of the models. Permutation importance measures also show that growth in average daily sales in the post-disbursal period acquires the status of the most important feature in the early warning exercise.

My results have important policy implication. First major result is that payment footprints have a potential to substitute credit scores as a screening device. Building credit information infrastructure such as credit bureau or credit registry is costly. Therefore, my results would indicate that promoting electronic payments and open banking could be more efficient than setting up or expanding the coverage of credit bureaus. Fintech innovation has made payment services accessible to micro, small and medium enterprises (MSMEs) creating a possibility to financially include the unbanked or under-banked parts of the economy (BIS and World Bank, 2020). Open banking derives the cost advantage over credit bureaus mainly from the fact that payment data are a by-product in the payment process. This combined with the fact that data is nonrival can create increasing returns to scale in data production (Jones and Tonetti, 2020; Bianchi et al., 2021). Further, since payment footprint is a hard data, it can be employed and interpreted independently from the entity that collected it (Liberti and Petersen, 2019).

One concern with open banking could be that borrowers may have to part with privacy if they need to share the details of their historical payments. However, my second set of results show that basic PF information are sufficient and borrowers do not need to reveal deep footprints that depend on transaction level information. PF basic variables are granular only in the sense that they are based on *short-term* data on PF. The basic PF variables are rather aggregative and therefore can easily conceal transaction level information.

Literature: My paper contributes to the following strands of literature: (i) Payment fintechs and SME lending, (ii) Fintech lending and application of big data, and (iii) relation between transactions and borrower screening and monitoring, and (iv) open banking. Payment fintechs

and bigtechs have disrupted the traditional payment industry (Bech and Hancock, 2020; Petralia et al., 2019; Philippon, 2016; Rysman and Schuh, 2017). Payment fintechs and bigtechs, all over the globe, have also entered into MSME lending. The most common examples of the US based payment fintechs that have ventured into lending are PayPal, Square, and Stripe. Square has even acquired a banking license to grow its merchant lending business. Tyro payments in Australia, also received a full banking license in 2016 with an authorization to operate as a deposit-taking institution. In Europe, iZettle, a Swedish payment company in the lending business, was acquired by PayPal in 2018. Examples of bigtechs into lending are Amazon, Mercado Libre, and Ant Financial. Payment fintechs and bigtechs may enjoy several advantages as lenders. My paper contributes in exploring one such advantage–informational edge from using borrower sales data.

A related paper that looks at the informational advantage for bigtechs is Frost et al. (2019). Frost et al. (2019) study the ability of a bigtech to use past transactions data to develop an internal credit score. They compare the predictive performance of the bigtech's internal credit score with the credit scores from credit bureau and find that bigtech's internal credit score performs better . My paper differs from Frost et al. (2019) in many respects. First, I work with the transaction level data that provides me much more granular information allowing me to identify the most important variables behind the superior performance of the transaction-based credit score. Second, my setting where lender's credit decision are orthogonal to credit bureau scores, allows to me answer the question to what extent credit bureaus can be substituted by transaction-based scoring system. Another paper on payment fintechs, Rishabh and Schäublin (2021), looks at the debt enforcement advantage of payment fintechs.

My paper also contributes to the literature on the use of big data and alternative data in loan underwriting. Berg et al. (2020) look at the predictive power of borrowers' *digital footprints*, such as the type of device, operating system, email host etc. in a consumer lending set-up in Germany. Similarly, Agarwal et al. (2021) use borrowers' *mobile footprints* such as number of calls, types of social media apps on the phone, and credit scores to predict default. My study differs from these studies in one important aspect that I use historical payments data as predictors and not alternative data such as digital footprints. Payments capture financial information of borrowers much more directly and that is why they are more likely to substitute credit scores. Second, my study focuses on SME lending in a developing country setting. Third, my setting allows me to separate the importance of credit scores from payment footprints owing to the fact that, the lender, in my setting ran the lending program oblivious to the credit scores. In the studies cited above, the lender always took into account the borrowers' credit scores. Fourth, in developing countries, several post-contracting frictions like poor enforcement are also big frictions. In these cases loan monitoring based on early-warning signals acquires an important role, which is also the focus of my study.

My paper also contributes to the long tradition of work connecting payments and lending. The idea, in the context of banks, is expressed as *checking account hypothesis* which states

that checking account transactions may contain useful information about the financial health of the borrowers (Black, 1975; Fama, 1985; Nakamura, 1993). Transactions—with payment fintechs or banks as intermediary—is potentially useful for both screening and monitoring of the borrowers. Screening advantage relates to the ability of the lender to use transaction accounts to derive creditworthiness of a (potential) borrower before the loan contract is signed. Monitoring advantage refers to the ability of the lender to use transaction accounts to obtain early warning signal on the loan performance after the loan disbursal.

Exploring the theme on monitoring Mester, Nakamura and Renault (2007) look at the response of a Canadian lender to monthly changes in transactions account and filings of the borrowing SMEs to the lender bank. They find that changes in the transaction accounts post-disbursal provide useful information about credit risk to which lenders respond by adjusting their monitoring intensity. The early warning exercise in my paper also confirms the usefulness of the transaction data similar to the analysis in Mester, Nakamura and Renault (2007). However, my analysis complements their findings in a few ways. First, I obtain my results in a developing country setting where the loans are uncollateralized, and monitoring relies completely on the transactions account activity. In contrast, in Mester, Nakamura and Renault (2007) the bank relies on both transaction account activity and monthly SME reports, and loans are collateralized with accounts receivable and inventories. Taken together, our results would imply that transactions are potentially useful in very different settings.

Another related paper in this theme using similar monthly snapshots on credit line usage, limit violations, and cash inflows is Norden and Weber (2010). The authors do not distinguish between screening and monitoring roles of bank transactions. They find in the context of a German bank, that these variables show abnormal patterns approximately 12 months before default. One major difference between my paper, and Norden and Weber (2010) and Mester, Nakamura and Renault (2007) is that these papers mainly utilize the bank's monthly assessment of borrower's credit line usage, account receivables and inventories and observe aggregative changes in monthly balances. In contrast, I observe transaction level data and I can drill deeper to identify which variables provide the best predictive power to the early warning exercise. Indeed my results show that changes in transaction patterns as early as in 28 days provide useful early warning signal to the lenders.

A key contribution of my paper relates to the screening advantage of transactions. In this context, one complementary paper is Puri, Rocholl and Steffen (2017) which shows that bank customers who already hold a transaction account at a bank are more likely to obtain credit from their bank and are less likely to default. The finding relating to loan approval is consistent with the fact the lenders use the transactions history to underwrite loans. I build on that result by showing how exactly that information is used by the lender, which elements of transaction data are more useful and so on. In addition, I determine the relative strength of transaction data and credit scores in loan screening.

Closer to my setting is another complementary paper by Ghosh, Vallee and Zeng (2021).

They study how even lenders that are not into payments business can use historical cashless payments data for loan underwriting by requiring borrowers to submit their past bank account statements. This is in spirit close to open banking. The authors find that borrowers whose bank statements recorded more cashless transactions, were more likely to be granted loans. In addition, such borrowers were also less likely to default. Thus, Ghosh, Vallee and Zeng (2021) extend the point by Puri, Rocholl and Steffen (2017) to electronic payments. I complement this study by studying granular information on cashless payments and showing how exactly the information contained in cashless payments can be used by the lender and which aspects of cashless payments are useful in predicting delinquency. I also strengthen the result in Ghosh, Vallee and Zeng (2021) by showing open banking may have far reaching consequence for credit markets as it not only facilitates credit underwriting, it may also make costly credit bureaus redundant.

Finally, my paper contributes to the nascent yet growing literature on open banking. The two notable papers in this field are He, Huang and Zhou (2020) and Parlour, Rajan and Zhu (2020). These papers make theoretical predictions about how open banking might impact the pricing of the payment services and the industrial organization of the credit market. While these papers take it is as given that payment data improves the default prediction, I establish it empirically.

The rest of the paper is organized as follows. Section 2 presents the institutional set-up, structure of the data, and data models. Section 3 discusses the main results relating to prediction based on historical sales data and its comparison with credit scores. This section also discusses results relating to heterogeneous borrowers, and feature importance. Section 4 extends our analysis to early warning exercise. Section **??** repeats the analysis for the loans by traditional lenders. Section 6 concludes. At the end I also provide a short note on outlining the revisions I am currently working on.

2 Data and Methodology

2.1 Institutional set-up

I collaborate with an Indian payment fintech, which is a major player in the Indian electronic payment market. The payment company provides point-of-sale (POS) machines mainly to MSMEs. For a client MSME the company processes the payments the MSME receives the using the company's POS machine. The company runs a lending program in partnership with some non-bank financial companies (NBFCs).[1] For the period under study, one partnering NBFC

[1] NBFCs are financing companies that do not have a deposit franchise, barring a few that were allowed to collect *non-demandable* deposits before 1997. The Reserve Bank of India has not given a deposit franchise to any non-bank financial company since 1997. NBFCs are also not part of the payment and settlement system. NBFCs are regulated and supervised by the Reserve Bank of India.

accounted for more than 80% of all the loans. I work with loans made by this largest NBFC partner of the company because its loan policies were more standardized. All the loans were unsecured and carried two percent per month interest rate. The interest rate charged is about the standard rate NBFCs levy on risky borrowers in India, and is in the range of interest rates charged by fintech lenders in the consumer credit market in the US and the UK (Cornelli et al., 2020).

The company screens the merchants based on their historical transactions data and provides the information about the screened merchants to the partnering NBFCs. The NBFC then decides whether to make a loan offer to the merchant and the loan amount associated with the offer. Once the NBFC picks a merchant as a potential borrower, the payment fintech makes the loan offer on behalf of the NBFC. The merchant may then take or leave the offer. If the merchant accepts the offer, the NBFC makes the loan disbursal, typically in a few days after making any additional checks, if required.

The lender also *suggested* a tenure to the borrower. I use the term *suggested tenure* to highlight that going beyond the tenure did not carry a penalty because the loan repayment was sales linked. That is repayment depended on the amount processed by the payment company for the borrowing merchant. The payment company deducted 10% of each transaction it processed for the borrowing merchant towards amortizing the loan, and remitted the remaining (minus charges, if any) amount to the merchant. Therefore, I will define another notion of tenure and call it *implied tenure*–the number of days the borrower should take to repay the loan if the borrower's sales continue to be at the same as the long-term average sales pre-disbursal (see more discussion below). Most of the loans had a suggested tenure of 90 days while some carried the suggested tenure of 180 days. Merchants also had the option to pay the loan early as well as to repay fully or partially by direct payment in lump sum to the company.[2]

2.2 Data and Models

I have access to the de-identified loan data, credit bureau data, and POS transaction data from the payment fintech. The loan data comprises of the principal (*loan amount*), date of loan disbursal, interest rate, suggested tenure, date when the loan was fully repaid (*loan closure date*), and shortfall in the repaid amount compared to the amount owed, if any. I use the data on shortfall to identify loans that went into default. Our loan data covers loans made between May 2017 and Feb 12, 2019.

The anonymized historical credit bureau data is from India's largest credit bureau, TransUnion CIBIL. The bureau data corresponds to the owners of the business that borrow from the payment fintech. The bureau data includes credit score at the time of the loan and additional snapshot information about other credit accounts reported to the Bureau by financial

[2]Many of these policies are common with the loan policies of the US based payment fintechs such as PayPal and Square. See Rishabh and Schäublin (2021) for more details.

institutions. It was not in the lender's policy to use any bureau data (including credit scores) for the loans. The lender however acquired them from the Bureau at a latter date. Not using credit scores is not an exception but a norm among payment fintechs. For instance two most prominent US-based payment fintechs, Paypal and Square, also do not use credit scores to make lending decisions.[3]

The credit bureau assigns scores between 300 and 900 to the borrowers, with a higher score considered better. It also separates those borrowers who did not have a sufficient history to be assigned a score. I call these loans *unscored loans*. I could successfully map about 82% of the loans in our dataset with the bureau data. Out of the 82% of the mapped loans, about 10% loans were unscored.

The anonymized transaction data is at the card-swipe level. The data covers approximately 270,000 borrowing and non-borrowing merchants over the period from January 2015 to February 2019. These merchants constituted the universe of merchants using the payment fintech's POS machine to accept digital payments. Each observation in the transaction data records the amount and date of transaction. In addition, it also provides anonymized card number, type of card (Amex, Visa, Mastercard etc.), whether it was a fallback transaction, and whether the transaction was refunded. Information on some of the items in the latter set is not complete. However, basic information is available for all transactions. In total, I observe details for more than 99.4 million transactions for borrowing and non-borrowing merchants. I use transaction data for non-borrowing merchants to create district level benchmarks for the borrowing merchants (more discussion below).

I create predictive *models* by combining various variables. In other words, I refer to a collection of variables as a model. The simplest model based on credit bureau data only uses credit scores to predict loan performance. I construct an extended credit bureau model that uses not only credit score but also variables based on other information available in the credit bureau report such as number of enquiries by financial institutions and so on (see Table A3 and Table A4 for more details). The idea behind creating the credit bureau models is to check whether a lender, in the absence of other financial information, can rely only on the information from credit bureau to make an assessment about a prospective business borrower. It is a fairly common problem that MSMEs do not have credible and verifiable information about the business, but the owner might have a credit history and that might act as a useful screening device for the lender.

I also create a model that combines all the variables that a traditional lender would have access to when making a loan to an MSME lacking financial records. The lender would know the loan characteristics (e.g. loan amount), borrower characteristics (e.g. industry, location, performance in previous loans from the same lender etc.). The lender, in addition could look at

[3]For PayPal's statement about credit scores, see `https://www.paypal.com/workingcapital/faq` and for Square's, see `https://squareup.com/help/us/en/article/6531-your-credit-score-and-square-capital-faqs`. (Accessed: Oct 24, 2021).

the credit bureau record of the owner. I encapsulate all this information in a model that I call *Traditional extended* model. This model primarily covers all the non-payment related relevant variables for credit underwriting.

I also construct several payment footprint (PF) variables by aggregating basic and granular transaction data assessing varied aspects of the borrowing merchant's business. I categorize each of the resulting variable into different *models* depending on the *complexity of information* it captures. The first is *PF Basic* model that includes relatively simpler variables derived from transaction data, such as average per-day sales, loan amount to sales ratio, average transaction size, coefficient of variation (CV) of per-day sales and transaction amounts, growth in the average sales, and so on.

Another model of transaction related variables is *PF Deep*. This model joins all the variables from the PF Basic model with somewhat more 'complex' variables. These complex variables include number of unique customers, Herfindahl–Hirschman index of customer concentration in terms of transaction values, distribution of sales over days of the week, sales distribution over card types, share of fall back transactions etc.

I also obtain demographic information like industry and zip code (called postal index number, *PIN* in India) for each merchant. Using the PIN codes, I aggregate sales at the district levels by all (borrowing and non-borrowing) merchants. Then I benchmark a borrowing merchant's sales before disbursal to the district sales by looking at the share of the merchant in the district's total sales and the change in this share prior to disbursal.

All the PF variables are calculated over the 90-day period before disbursal and carry the subscript "lh". All the relative and absolute changes in the PF variables are changes in values calculated in the 28-day period before disbursal over values calculated in the same number of days between 63^{rd} day and 91^{st} day before disbursal. Change variables have the subscript "dr" for relative changes and "da" for the absolute changes. Finally, I mostly work with the log of variables that are neither ratios nor relative distance. Absolute change in a log variable approximates relative change in the variable. For a complete list of variables under each model, see Table A4. For a description of variables see Table A3.

Predicted variables: I predict three related adverse credit events—default, late and non-performance. To classify loans into these categories I use the snapshot view of loan performance as on 31 December 2019—ten months after the disbursal of the last loan included in our analysis. I define default loans as those loans that had a "large" shortfall (pending amount) as on 31 December 2019. I call a shortfall large if it is more than five percent of the due repayment amount as on 31 December 2019.[4]

I label a loan as *late* if the borrower took more than 30 days longer than the *implied tenure* to close (fully repay) the loan. Due to being sales-linked, all the loans had a *suggested* tenure and

[4]A small proportion of the default loans were written-off by the lender and no longer followed up due to merchant having left the payment company network. Majority of default loans were still being pursued and were obviously much late beyond their (implied) tenure as on 31 December 2019.

going beyond suggested tenure did not entail any penalty or late fees. The payment company offered loans with a suggested tenure of mainly 90 days and 180 days. To capture the idea of maturity better, I define a measure of tenure that I call *implied tenure*. Implied tenure is the number of days that would be required to repay the loan (loan amount plus interest amount), if the merchant continued to have same sales as their *pre-disbursal long-term average sales* and given the lender deducted 10% of each sale towards repayment. 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. 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. clearly, *default* and *late* are mutually exclusive categories, as *late* loans were fully repaid while the *default* loans were still pending or written-off.

Finally, a loan is non-performing if it was either *late* or in *default*.

2.3 Summary Statistics

Tables 1 provides summary statistics on select loan-related and credit-bureau related variables. The average loan made by the payment company is about INR 38,000, roughly about USD 570 in the nominal exchange rate or roughly about USD 1,900 in purchasing power parity exchange rate. As against the average suggestive tenure of 113 days, average implied tenure was 167 days. About 89% borrowing business owners were male. Of the 11526 loans that could be mapped with credit bureau data, 90% of the loans had borrower with a credit score while the remaining 10% were unscored without a recent credit history. Conditional on being scored, borrowers had an average score of 713, while the median score was 726. In Rishabh and Schäublin (2021) we show that compared to the benchmark credit market, the payment fintech serves mainly borrowers in the lower end of the score spectrum. Compare to the median score of 726 for the payment company loans, median score in the benchmark credit market is above 800.

Borrowers on average had a 6 year credit history, and about 15 months of relationship with the lender, at the time of the payment-company loan. These borrowers were mostly experienced borrowers having borrowed from the credit market previously.

Bottom half of Table 1 gives the summary statistics about the loan performance. 16% of the payment company loans were late and 12% went into default taking the share of non-performing loans to 28%. Days past due on an average loan is negative, indicating an average loan was repaid earlier than its implied due date.

Table 2 gives summary statistics according to loan status, dividing the sample into *performing* and *non-performing* loans. The table also presents a two-sample t-test for mean difference for each variable. The test already gives us a hint about the variables that may possess power to separate non-performing loans from performing loans. An average non-performing loan is

Table 1: Summary Statistics on Select Loan Related Variables

Variable	Short description	No. Loans	Mean	Median	SD	p10	p90
loanamount	Loan amount (INR)	15375	37643.22	25000.00	37688.73	10000.00	83000.00
tenure	Suggestive tenure (days)	15375	112.79	90.00	43.92	90.00	180.00
itenurelong	Implied tenure (days)	15375	166.62	119.05	235.07	56.49	269.24
relationage	Relationship length (months)	15375	15.15	13.77	8.64	5.22	27.20
age	Borrower age (years)	13565	36.31	34.80	8.88	26.58	48.21
gender	1 = Male, 0 = Female	13565	0.89	1.00	0.31	0.00	1.00
isscored	Credit history exists (1 = Yes)	12746	0.90	1.00	0.29	1.00	1.00
scorecond	Credit score	11526	713.30	726.00	53.46	639.00	773.00
lengthhist	Credit history length (years)	12390	6.28	5.06	5.23	0.44	13.64
nenq30	No. of bureau inquiries	13428	0.59	0.00	1.24	0.00	2.00
idpd	Implied days past due (days)	13465	-41.83	-12.68	201.66	-128.30	52.80
impislate30	Late (1 = Yes)	15375	0.16	0.00	0.37	0.00	1.00
isdefault	Default (1 = Yes)	15375	0.12	0.00	0.33	0.00	1.00
isnonperf	Non-performing (1 = Yes)	15375	0.28	0.00	0.45	0.00	1.00

For detailed description of the variables see Table A3. *Implied tenure* is the number of days it would take to repay the loan (loan amount + interest) given the 10% deduction rate and if the merchant, after disbursal continued to transact at their pre-disbursal long-term average. Pre-disbursal long-term average is average daily transaction value calculated over 91 days between day 120 and day 30 before loan disbursal. Relationship length is the number of months between the loan disbursal and the first ever transaction processed by the payment company for the borrowing merchant. Variables calculated using Bureau data (`age` through `shrunning`) refer to values as on the day of disbursal of the payment-company loan. CIBIL credit score ranges between 300 and 900, with score above 700 considered to be good by the financial institutions. *Implied days past due* is the number of days beyond *implied due date* that the borrower took to fully close the loan. Where *implied due date = date of disbursal + implied tenure*. *Default* is a binary variable that takes value 1, if the loan had a shortfall > 5% of repayment amount and it was either written off or still pending as of end 2019. *Late* is a binary variable that takes value 1 if a loan was non-defaulting and had an *implied days past due* of at least 31 days. *Non-performing* is a binary variable that takes value 1 when either the loan is categorized as Default or Late. *Number of credit accounts (nloans)* includes credit card and loan accounts reported to the credit bureau. Variables *shcollclosed* and *shrunning* represent shares of *nloans*. *shnpclosed* is share non-performing credit accounts as a share of closed credit accounts reported to the Bureau.

Table 2: Summary Statistics on Select Loan Related Variables – by Loan Repayment Status

		Mean Values		
Variable	Short description	Performing	Non-performing	Mean difference
loanamount	Loan amount (INR)	35810.08	42290.75	-6480.67***
tenure	Suggestive tenure (days)	112.81	112.72	0.10
itenurelong	Implied tenure (days)	169.85	158.43	11.42***
relationage	Relationship length (months)	15.19	15.05	0.13
age	Borrower age (years)	36.49	35.87	0.62***
gender	1 = Male, 0 = Female	0.89	0.89	0.00
isscored	Credit history exists (1 = Yes)	0.90	0.92	-0.02***
scorecond	Credit score	715.73	707.25	8.48***
lengthhist	Credit history length (years)	6.31	6.19	0.12
nenq30	No. of bureau inquiries	0.55	0.72	-0.17***
idpd	Implied days past due (days)	-68.29	77.49	-145.79***
impislate30	Late (1 = Yes)	0.00	0.56	
isdefault	Default (1 = Yes)	0.00	0.44	
isnonperf	Non-performing (1 = Yes)	0.00	1.00	
samplesize	Number of loans	11026	4349	

Non-performing loans are those that were either categorised as *late* or *default*. Performing loans are those that are not categorised as *non-performing*. For more details see notes for Table 1. For detailed description of the variables see Table A3. Mean difference test: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

about 18% larger in size than an average performing loan—a significant difference of about INR 6500. An average performing loan had a slightly higher credit score compared to an average non-performing loan. While both types of loans had roughly same number of credit accounts reported to the bureau, the non-performing loans had one percentage point higher rate of non-performance among their closed credit accounts.

Table 3 summarizes select variables included in the payment footprint models. Loan-sales ratio for an average loan stood at about 16. This means the merchant with the average loan-sales ratio, would repay her loan in about 16 days if she continued her pre-disbursal average sales in the post-disbursal period, and pledged 100% (instead of 10%) of her sales towards repayment. An average borrowing merchant had about 103 (=exp(4.63)) unique customers in the 90 days prior to loan disbursal. Before loan disbursal, a borrowing merchant made an average sales of about INR 2344 (=exp(7.76)) per day, and had an average transaction size of INR 1670 (=exp(7.42)). Sales are highly variable with coefficient of variation of two on average. Sales are also not concentrated as evident from the low HH index over customers' transaction values. A vast majority of transactions take place through Mastercard and Visa cards. On average, about one-third of the transactions in terms of values take place on weekends, and the remaining two-thirds equally distributed over the days of the remaining week (not show in the Table). Merchants experience growth in number of customer counts, in average transaction size, and in average daily sales before they are granted loans.

Table 3: Summary Statistics on Select Payment Footprint Variables

Variable	Short description	No. Loans	Mean	Median	SD	p10	p90
lsratio	Loan-sales ratio. Sales is average per day.	15375	15.66	11.27	22.26	5.36	25.09
rtsshort	Repayment to short-term sales ratio	15375	1.45	0.11	112.55	0.05	0.28
dayspast_freqcust_lh	Days since the visit of the most frequent cust	11645	27.98	19.00	25.26	3.00	69.00
dayspast_largcust_lh	Days since last visit of the largest cust	11645	31.47	25.00	25.10	4.00	71.00
log_custcount_lh	Log number of customer	11645	4.63	4.63	1.14	3.37	6.02
log_avgtrans_lh	Log average transaction size	15375	7.42	7.33	1.08	6.11	8.77
log_avg_transday_lh	Log average per day sales	15375	7.76	7.72	0.93	6.75	8.87
cv_trans_lh	Coeff. of variation of trans	15375	1.56	1.23	1.05	0.70	2.86
cv_transday_lh	Coeff. of variation of daily sales	15375	2.00	1.72	1.17	0.86	3.48
HH_cust_trans_lh	HHI of customers' transactions	11645	0.08	0.03	0.12	0.01	0.19
days_since_lasttrans_lh	Days since last transaction	15375	2.78	1.00	5.52	1.00	5.00
shr_Sun_trans_lh	Percent of sales happening on Sundays	15375	17.53	17.04	11.29	2.94	30.92
shr_Sat_trans_lh	…….. on Saturdays	15375	15.65	15.16	8.99	4.71	25.87
shr_AMEX_trans_lh	Percent of sales through Amex cards	15375	0.09	0.00	1.09	0.00	0.00
shr_MAESTRO_trans_lh	…… through Maestro cards	15375	2.21	1.09	3.63	0.00	5.63
shr_MASTERCARDVISA_trans_lh	…… through Mastercard or Visa	15375	86.34	88.35	10.68	73.01	97.75
log_custcount_da	Percent change in custcount	10257	12.58	6.16	75.14	-58.49	90.67
log_avgtrans_da	….. in avgtrans	15146	8.86	3.56	68.31	-50.17	75.53
log_avg_transday_da	….. in avg_transday	15146	9.31	7.04	97.58	-84.41	105.31
cv_trans_dr	….. in cv_trans	14947	4.40	2.33	58.47	-56.23	70.99
cv_transday_dr	….. in cv_transday	15146	2.75	1.90	44.24	-49.00	55.82
HH_cust_trans_da	Absolute change in HH_cust_trans	10257	0.00	0.00	0.17	-0.14	0.16
shr_Sun_trans_da	….. in shr_Sun_trans	15375	-0.13	0.00	18.57	-19.38	19.24
shr_Sat_trans_da	…… in shr_Sat_trans	15375	-0.25	0.00	17.81	-18.43	17.89
shr_AMEX_trans_da	…… in shr_AMEX_trans	15375	0.04	0.00	1.77	0.00	0.00
shr_MAESTRO_trans_da	…… in shr_MAESTRO_trans	15375	-0.31	0.00	6.25	-4.52	3.37
shr_MASTERCARDVISA_trans_da	….. in shr_MASTERCARDVISA_trans	15375	-0.10	0.00	18.22	-15.40	15.51
shr_dis_sumtrans_lh	Percent of district sales	15374	0.58	0.03	3.88	0.01	0.65
shr_dis_sumtrans_da	Absolute change in share of district sales	15374	0.02	0.00	1.68	-0.07	0.09

See Table A3 for variable descriptions. Variables ending with subscript "_lh" are calculated in the 91-day window before disbursal. Variable ending with "_dr" and "_da" represent relative and absolute changes in values of the payment variables calculated in the 28-day period before disbursal and values calculated in the 63-day to 91-day period before disbursal. Absolute changes in log variables approximate the percent change in the variable.

Table 4 compares the same variables across performing and non-performing loans. Interestingly, we see that almost all the variables in the Table show a significant mean difference between the two types loans, indicating that historical payment data can potentially help to predict adverse credit events. For instance, performing loans on average have 10% higher customers in the 90-day period before disbursal, than non-performing firms. Performing borrowers have lower variation in sales, lower average transaction size, higher growth in sales before credit uptake.

Table 4: Summary Statistics on Select Payment Footprint Variables – by Loan Repayment Status

		Mean Values		
Variable	Short description	Performing	Non-performing	Mean difference
lsratio	Loan-sales ratio. Sales is average per day.	15.96	14.91	1.04***
rtsshort	Repayment to short-term sales ratio	0.18	4.67	-4.50**
dayspast_freqcust_lh	Days since the visit of the most frequent cust	27.53	29.10	-1.56***
dayspast_largcust_lh	Days since last visit of the largest cust	30.73	33.33	-2.60***
log_custcount_lh	Log number of customer	4.66	4.55	0.10***
log_avgtrans_lh	Log average transaction size	7.35	7.58	-0.22***
log_avg_transday_lh	Log average per day sales	7.71	7.87	-0.16***
cv_trans_lh	Coeff. of variation of trans	1.53	1.65	-0.12***
cv_transday_lh	Coeff. of variation of daily sales	1.94	2.16	-0.23***
HH_cust_trans_lh	HHI of customers' transactions	0.07	0.09	-0.02***
days_since_lasttrans_lh	Days since last transaction	2.45	3.62	-1.18***
shr_Sun_trans_lh	Percent of sales happening on Sundays	17.66	17.22	0.44**
shr_Sat_trans_lh	…….. on Saturdays	15.97	14.85	1.12***
shr_AMEX_trans_lh	Percent of sales through Amex cards	0.08	0.13	-0.05***
shr_MAESTRO_trans_lh	…… through Maestro cards	2.30	1.97	0.33***
shr_MASTERCARDVISA_trans_lh	…… through Mastercard or Visa	85.90	87.47	-1.57***
log_custcount_da	Percent change in custcount	15.86	3.52	12.33***
log_avgtrans_da	….. in avgtrans	11.55	1.95	9.59***
log_avg_transday_da	….. in avg_transday	18.16	-13.47	31.63***
cv_trans_dr	….. in cv_trans	6.88	-2.04	8.92***
cv_transday_dr	….. in cv_transday	1.50	5.96	-4.46***
HH_cust_trans_da	Absolute change in HH_cust_trans	0.00	0.01	0.00
shr_Sun_trans_da	….. in shr_Sun_trans	-0.06	-0.29	0.22
shr_Sat_trans_da	…… in shr_Sat_trans	-0.40	0.13	-0.53*
shr_AMEX_trans_da	…… in shr_AMEX_trans	0.05	0.01	0.04
shr_MAESTRO_trans_da	…… in shr_MAESTRO_trans	-0.33	-0.24	-0.09
shr_MASTERCARDVISA_trans_da	….. in shr_MASTERCARDVISA_trans	0.29	-1.09	1.38***
shr_dis_sumtrans_lh	Percent of district sales	0.54	0.67	-0.12*
shr_dis_sumtrans_da	Absolute change share in district sales	0.07	-0.11	0.18***
samplesize	Number of loans	11026	4349	

See Table A3 for variable descriptions. Variables ending with subscript "_lh" are calculated in the 91-day window before disbursal. Variable ending with "_dr" and "_da" represent relative and absolute changes in values of the payment variables calculated in the 28-day period before disbursal and values calculated in the 63-day to 91-day period before disbursal. Absolute changes in log variables approximate the percent change in the variable. *Non-performing* loans are those that were either categorised as *late* or *default*. Performing loans are those that are not categorised as *non-performing*. Mean difference test: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

2.4 Methodology

Our objective is to predict the three adverse credit events using different sets of variables (*models*) and then to compare the predictive power across models. We want to check if payment footprint based models perform better than the traditional models in predicting these adverse events. Further, we will ask if combining the information in the traditional models and payment

footprint models gives us better predictive power. Moreover, we check if similar predictive power can be achieved by using rather simple variables which are easily observable and computable.

For each adverse-event prediction exercise, I divide the sample of observations into two *sets*—training and test sets. I use a machine learning algorithm that trains on the data in the training set corresponding to the variables contained in the model under consideration. I then use the trained algorithm to predict the response variable (indicator variable for the adverse event) for the observations in the *test set*. This ensures that I follow the *firewall principle* (Mullainathan and Spiess, 2017). That is, I do not use any of the data involved in training the algorithm to evaluate the predictive performance of the algorithm, neither in the main analysis nor in any robustness checks. I randomly allocate 80% of the observations for training (and cross-validation, if applicable) and the remaining 20% of the observations for the out-of-sample predictions.

For our main analysis, I employ supervised machine learning algorithm, *Random forest*. Random forest is an ensemble of several decision (regression) trees. To make predictions it takes an average over predictions from a large number of trees (Breiman, 2001). Each regression tree makes predictions by splitting into two at each of the *node*. Where the node represents some condition relating to a variable (feature) included in the model. However, not all trees are subject to the same data from the training set or include the same variables. In a process called Bootstrap aggregation (*bagging*), separate training data is generated for each tree from the original training data. To do that, for each of the tree, a bootstrapped sampling selects N observations from the original training sample of size N, with replacement. Finally, each tree employs only a random subset of features from the model. As suggested by practitioners, I randomly select one-third of the features for a tree and the nodes split until they are left with five observations (Breiman, 2001; Hastie, Tibshirani and Friedman, 2008). For each exercise, the random forest grows over 300 trees.

An advantage of using Random forests is that the algorithm requires relatively little tuning and generally performs better out-of-sample compared to methods such as deep neural networks (Athey and Imbens, 2019; Hastie, Tibshirani and Friedman, 2008). Nevertheless, I will also do a robustness check for our baseline results by estimating predictive power using neural network and logit regression.

I measure the relative predictive performance of different models by plotting Receiver Operating Characteristics (ROC) curves for the out-of-sample predictions. ROC curve is based on two metrics on predictive performance—false positive rate (FPR) and true positive rate (TPR). For illustration, let the outcome variable be the default indicator and therefore, the predicted value of this variable is a *score* between zero and one (essentially, the predicted probability of default). If we fix a predicted score as the threshold and classify all the loans that have a predicted score above the threshold to be default loans, then TPR determines how many of the loans we *correctly* classified as default out of the actual number of default loans. Similarly, FPR, at the given threshold score, counts out of all the non-default loans how many

we *incorrectly* classified as default loans. An ROC curve plots the FPR-TPR pairs obtained by varying the predicted scores as decision thresholds. A model that achieves a TPR of one and an FPR of zero for some threshold score, has a capability to perfectly predict default. On the other hand, a model that randomly predicts loans as default or non-default, will have the same TPR and FPR regardless of the threshold score.

A handy way to summarize the predictive power of a model is to calculate the area under the ROC curve (AUC). A model with perfect predictive power traces an area of one under the ROC curve, while the random-guess model has an AUC of 0.5. Therefore, the closer the AUC to one, better is the predictive power of the model. I also compute the 95% confidence interval of the AUC by bootstrapping over 1000 bootstrap replicas of the test set.

3 Results

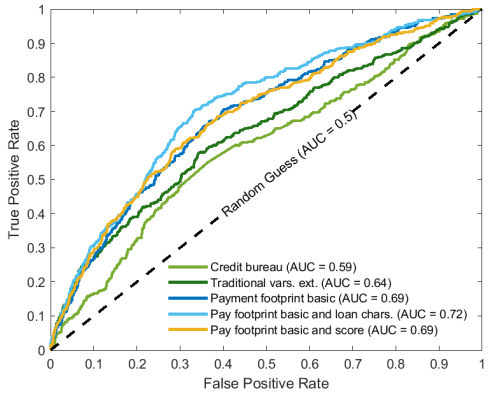
3.1 Baseline Comparison of Predictive Performance

The first major result our study is that payment footprint based models perform better than traditional variables in predicting any of adverse credit event. Figure 2 plots ROC curves for five different models: Credit bureau, Traditional, PF Basic, PF Basic + loan & business characteristics, and PF Basic + credit score. Table 5 reports the corresponding AUCs for these models as well as some more models. Table 5 also reports the 95% confidence interval for the AUCs. In predicting default, with an AUC of 0.59, Credit bureau model has predictive power significantly better than random guess but it is the lowest compared to other models. Traditional model, which builds on the Credit bureau model, is able to improve the predictive performance significantly up and the AUC point estimate reaches to 0.64. However, these models perform significantly worse than even the PF Basic model, which comprises of 9 transaction related variables. The point estimate of the AUC for the PF Basic model is 0.69—five percentage point higher than the Traditional model.

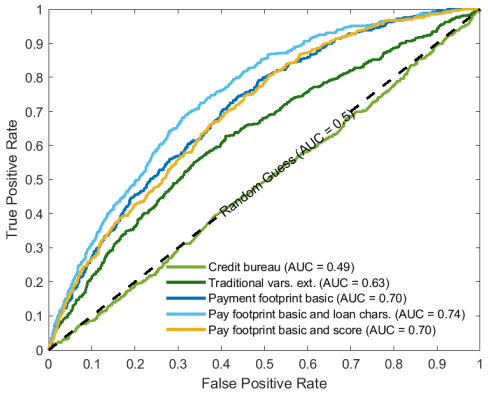
The PF based models achieve even higher performance when we combine them with business and loan characteristics. The point estimate of AUC for such a combined model stands at 0.72, highest compared to any other model. Specifically, it is higher than the AUC of the model that combines PF Basic with credit scores (0.69). This has two notable implications: First, PF Basic model already achieves high predictive performance compared to Credit bureau model. Second, the appended PF Basic + credit score model not improving performance implies that PF Basic variables capture most of the information contained in credit score. These two observations in turn imply that while credit bureau may have some signal about the credit quality it is much more noisy compared to the signal provided by payment footprints.

Finally, we find that these relative performances across models carry over to other adverse events – Late and Non-performing. The predictive performance in general is lower in predicting non-performing loans across models, nevertheless, PF based models outperform Traditional

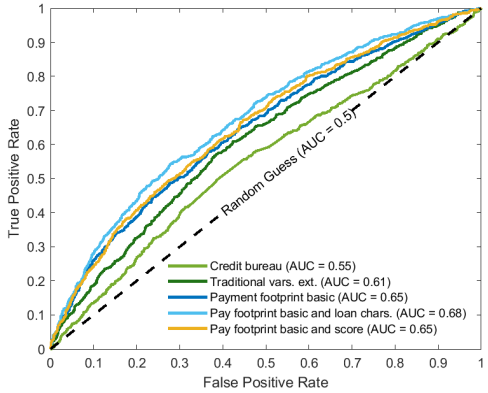
Figure 2: ROC Curves for Out-of-Sample Predictions Across Models



(a) Default



(b) Late



(c) Non-performing

For the figure on default predictions, false positive rate is the proportion of non-default loans incorrectly classified as default by the model predictions at the chosen threshold. True positive rate is the proportion of default loans correctly classified as default by the model predictions at the chosen threshold. Similar interpretation applies for Late and Non-performing predictions. *Default* is a binary variable that takes value 1, if the loan had a shortfall $>$ 5% of repayment amount and it was either written off or still pending as of end 2019. *Late* is a binary variable that takes value 1 if a loan was non-defaulting and was took at least 30 days more than the implied tenure to fully pay the loan off. Implied tenure is the number of days it would take to repay the loan (loan amount + interest) given the 10% deduction rate and if the merchant after disbursal continued to transact at their pre-disbursal long-term average. Pre-disbursal long-term average is average daily transaction value calculated over 91 days between day 120 and day 30 before loan disbursal. *Non-performing* is a binary variable that takes value 1 when either the loan is categorized as Default or Late. Credit bureau model includes credit score variables and other variables derived from credit report such as number of credit accounts reported in the bureau report, number of collateralized credit accounts, and number of non-performing accounts in the bureau report etc. Traditional model consists of variables traditionally used in loan underwriting and covers bureau variables, and variables relating to loan characteristics (e.g. loan amount), business characteristics (e.g. industry) and relationship characteristics (e.g. loan number). Basic payment footprint (PF) model captures basic variables on historical sales such as average daily sales, coefficient of variation of sales, loan-sales ratio etc. Combined models combine the elements of traditional variables with the PF variables. Basic PF + business & loan characteristics combines Basic PF variables with non-credit bureau variables of Traditional model. Basic PF + credit score model checks the predictability of easily constructible basic footprint variables in combination with credit score. For results on more models see Table 5. For details on models and variables see Table A3 and Table A4. Out-of-sample predictions were made after training a Random forest model. Random forest estimation is done with 300 trees.

Table 5: Area Under the ROC Curve for Out-of-Sample Predictions

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable		score + othr.	+ loan chr.		PF basic + other	+ loan chr.	+ score	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Default	0.57	0.59	0.64	0.69	0.68	0.72	0.69	0.67
	[0.54, 0.60]	[0.56, 0.63]	[0.61, 0.67]	[0.66, 0.72]	[0.64, 0.70]	[0.69, 0.74]	[0.66, 0.72]	[0.64, 0.70]
Late	0.50	0.49	0.63	0.70	0.71	0.74	0.70	0.73
	[0.48, 0.53]	[0.46, 0.52]	[0.60, 0.65]	[0.68, 0.73]	[0.69, 0.74]	[0.72, 0.76]	[0.67, 0.72]	[0.71, 0.76]
Non-perf	0.53	0.55	0.61	0.65	0.64	0.68	0.65	0.66
	[0.51, 0.56]	[0.53, 0.58]	[0.59, 0.63]	[0.62, 0.67]	[0.62, 0.66]	[0.65, 0.70]	[0.63, 0.68]	[0.64, 0.68]
No. of obs.	10183	11584	12300	12300	12300	12300	12300	12300
No. of vars.	2	9	22	10	51	22	12	76

Values in square brackets represent the 95% confidence interval of the AUC. We compute confidence interval by bootstrapping using 1000 bootstrap samples. Number of observations refers to number of observations in the training set. Number of variables does not include the number of dummy variables but only number of dummy classes (two). *Default* is a binary variable that takes value 1, if the loan had a shortfall > 5% of repayment amount and it was either written off or still pending as of end 2019. *Late* is a binary variable that takes value 1 if a loan was non-defaulting and was took at least 30 days more than the implied tenure to fully pay the loan off. Implied tenure is the number of days it would take to repay the loan (loan amount + interest) given the 10% deduction rate and if the merchant after disbursal continued to transact at their pre-disbursal long-term average. Pre-disbursal long-term average is average daily transaction value calculated over 91 days between day 120 and day 30 before loan disbursal. *Non-performing* is a binary variable that takes value 1 when either the loan is categorized as Default or Late. Number of observations refer to the number of observations used in training the model. Number of variables refers to the number of features (explanatory variables) included in the prediction model. Credit score model (column 1) consists of two features – (a) whether credit score exists at the time of borrowing, and (b) credit score. Extended credit bureau model (column 2) includes credit score variables and other variables derived from credit report such as number of credit accounts reported in the bureau report, number of collateralized credit accounts, and number of non-performing accounts in the bureau report etc. Traditional model (column 3) consists of variables traditionally used in loan underwriting and covers bureau variables, and variables relating to loan characteristics (e.g. loan amount), business characteristics (e.g. industry) and relationship characteristics (e.g. loan number). Basic payment footprint (PF) model (column 4) captures basic variables on historical sales such as average daily sales, coefficient of variation of sales, loan-sales ratio etc. Payment footprint deep model (column 5) covers deeper information about the business that can be gathered from the transaction data such as customer base, concentration of customers, card types swiped etc. Combined models combine the traditional variables with the PF variables. Basic PF + business & loan characteristics (column 6) combines variables for model in column (4) with non-credit bureau variables of model in column (3). Basic footprint + credit score model (column 7) checks the predictability of easily constructible basic footprint variables in combination with credit score. Model All (column 8) is the most comprehensive and includes all variables. For details on models and variables see Table A3 and Table A4. Out-of-sample predictions were made after training a Random forest model. Random forest estimation is done with 300 trees.

model and Credit Bureau model.

Robustness tests: I test whether the baseline results are sensitive to the algorithm I use. Therefore, I run the same sets of predictions by training alternative models on the same training set as in the baseline model. Table B2 in the appendix reports out-of-sample prediction performance for models trained with Neural network algorithm and Logit regression. We can see from the table that all our results on relative model performance still hold for these algorithms as well. We also find that, in terms of AUCs, Random forest model performs, in general, better than Neural net and Logit, in consonance with the previous literature (Athey and Imbens, 2019; Hastie, Tibshirani and Friedman, 2008).

The second test addresses the concern that Credit score model may have lower performance because there are more observations for non-credit score models, such as PF Basic and PF Basic augmented with loan and business characteristics. In this test, I restrict the sample only to those observations that have no data missing on *any* of the variable considered under any model. The results of this test, in Table B1 in the appendix, confirm our baseline findings. Although, performance does take a hit across all the models due to far fewer number of observations in the training set, yet the results about the superior performance of the PF based models still holds.

3.2 Relative Performance Across Heterogeneous Borrowers

Next, we want to test whether the superior performance of the PF models is driven by certain kind of firms or loans. Specifically, we check how the ordering of performance holds across firm sizes and across credit score types. To study heterogeneity across borrower sizes, I run the same analysis as our baseline but over two samples of borrowers—small and large. Small borrowers are defined as those with below median total transaction value in the 90-day window before disbursal and large borrowers as those above median values. Figure 3 presents the ROC curve according to firm size. There are several interesting findings here. First, for large merchants, even though the PF based models perform better than the Credit bureau model, the performance difference is smaller, compared to the baseline. For predicting default, Credit bureau model achieves an AUC of 0.60, only four percentage point below the PF Basic model. This difference in the baseline was 10 percentage points.

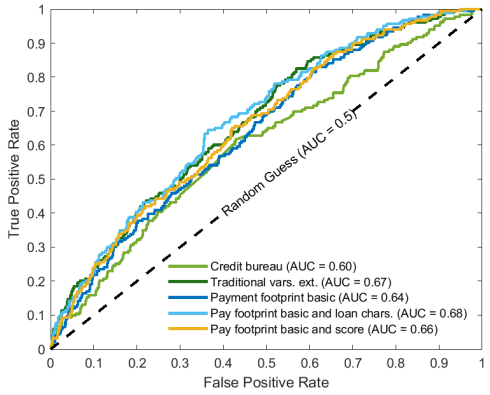
Second, the smaller difference for the large-merchant sample is mainly due to worse performance of the PF based models compared to the overall sample (baseline). One reason for the drop in the performance of the PF models could be that larger firms have multiple sources of revenue that payment footprints from one company cannot capture comprehensively. Third, for small merchants, the performance improvement of PF Basic model over Credit bureau model is even higher than in the baseline (by about 10 percentage points). This is driven by both, loss of performance for the Credit bureau model and a slight gain of performance for the PF Basic model. To summarize, our results suggest that payment footprints gauge credit quality better than credit bureau variables for firms of *all* sizes, but their comparative advantage is higher in assessing small firms. Table A1 in the appendix reports the AUCs for various models across different sizes, confirming our results above.

To study whether the superior performance of the PF models is 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[5] 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 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. However, in this case, we do not find any difference across score buckets and compared to the baseline. Thus, payment footprints seem to be more effective default predictors for both the types of borrowers. Table A2 in the appendix reports the AUCs for various models across different score buckets. The table also presents predictive performance of the PF based models for the unscored borrowers. These borrowers did not have a credit score because of a lack of a recent credit history. We can see

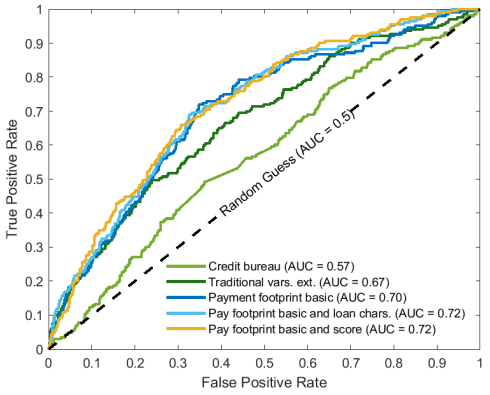
[5]See `https://www.cibil.com/faq/understand-your-credit-score-and-report`(Accessed: November 8, 2021).

Figure 3: ROC Curves for Out-of-Sample Predictions Across Models – by Size

Default

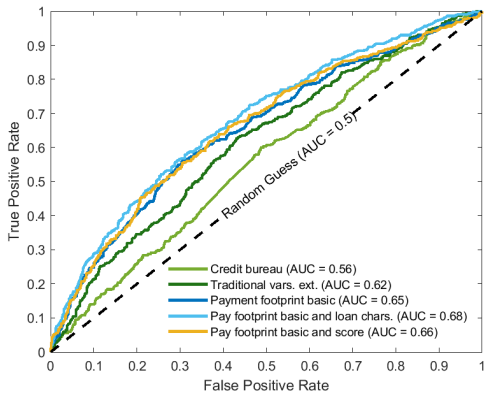


(a) Large Merchants

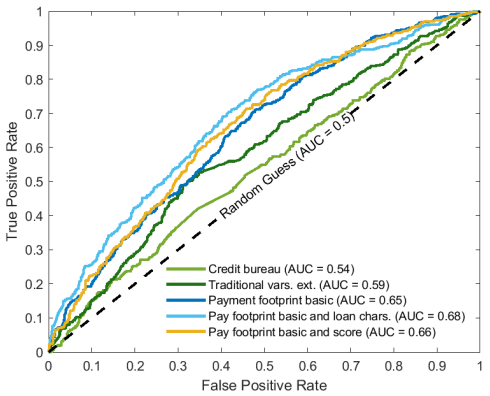


(b) Small Merchants

Non-performing



(c) Large Merchants



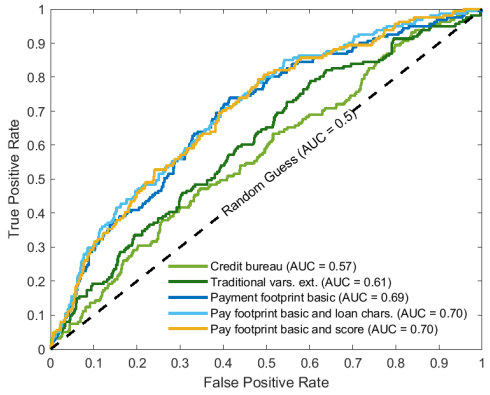
(d) Small Merchants

Large borrowers are those that had above median total value of transactions in the 90-day period before disbursal. Small borrowers had total value of transactions below median. For the figure on default predictions, false positive rate is the proportion of non-default loans incorrectly classified as default by the model predictions at the chosen threshold. True positive rate is the proportion of default loans correctly classified as default by the model predictions at the chosen threshold. Similar interpretation applies for Late and Non-performing predictions. For a short description about models, see notes for Figure 2. For the complete list of model variables see Table A4. For variable description see Table A3. Predictions are based on a model estimated with random forest algorithm. Random forest estimation is done with 300 trees.

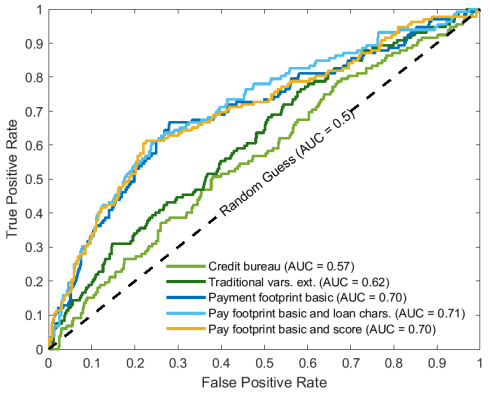
from the table that PF Basic model combined with loan and business characteristics has high predictive power. This shows that payment footprints can be used to underwrite loans to even those borrowers who may not have access to formal credit otherwise.

Figure 4: ROC curves for Out-of-Sample Predictions Across Models – by Score

Default

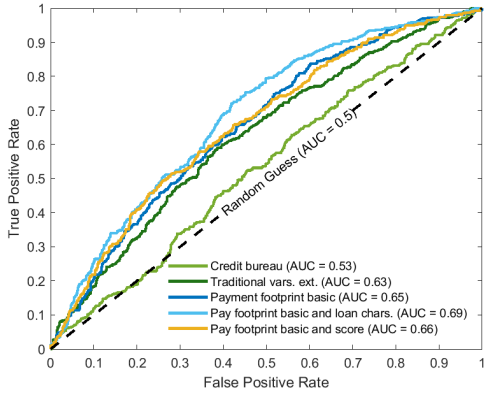


(a) High Score (Credit score $\geq$ 700)

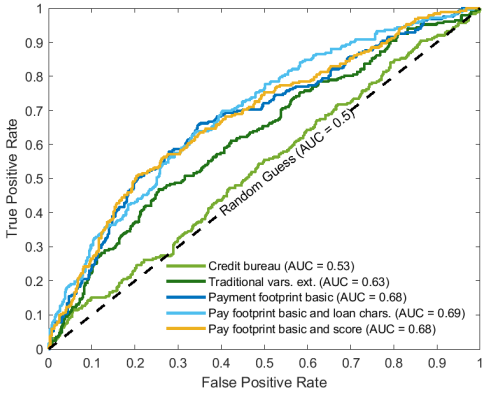


(b) Low Score (Credit score $<$ 700)

Non-performing



(c) High Score (Credit score $\geq$ 700)



(d) Low Score (Credit score $<$ 700)

Credit scores correspond to that of the merchant owning the business. Credit scores range between 300 and 900. Scores above 700 are assessed as good by the credit market. For the unscored loans, the borrowers did not have a long enough credit history at the time of the borrowing to have been assigned any score by the credit bureau. For the figure on default predictions, false positive rate is the proportion of non-default loans incorrectly classified as default by the model predictions at the chosen threshold. True positive rate is the proportion of default loans correctly classified as default by the model predictions at the chosen threshold. Similar interpretation applies for Late and Non-performing predictions. For short description of the model, see Figure 2. For results on more models see Table A2. For details on models and variables see Table A3 and Table A4. Out-of-sample predictions were made after training a Random forest model. Random forest estimation is done with 300 trees.

3.3 Feature Importance

Next we check which variables are the most important in the prediction. Since the model PF Basic + business & loan characteristics has the highest predictability, we will look for the most important variables from this model. Following, Molnar (2020), we will determine the most

important variables using two complementary measures: (i) Out-of-bag variable importance by permutation, and (ii) Shapley additive explanations.

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

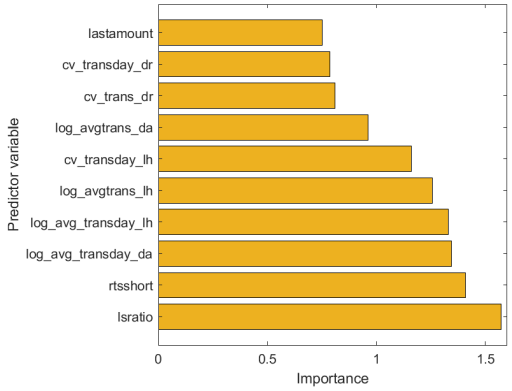
This method rests on the fact that in the process of bagging, due to sampling with replacement, a tree leaves out some (typically about 37%) of the observations from the original training set. Thus, for the tree, these left-out observations are *out-of-bag* (Breiman, 2001). To calculate the importance of a variable, this method first estimates the prediction error for the tree using these OOB observations. As a next step, it recalculates prediction error on the same OOB observations but after permuting (randomly reshuffling) the values of the variable across observations, but using the same trained model. After recording the change in prediction error, as a final step, the method averages the changes in prediction errors over all the trees where the variable appeared in some split. The variable importance is the average change in prediction errors divided by the standard error of the changes. If a variable is important in prediction, then permuting its values should increase the prediction error and lead to a higher value of the measure.

One drawback of the OOB permutation measure is that it tends to assign higher importance to features that are correlated. It is because post-permutation, it forces the model—that is trained *before* permuting the value of the correlated variable—to make predictions for regions where it did not train, leading to higher prediction errors and thus a higher variable importance(Hooker, Mentch and Zhou, 2021). Such an *unexplored* region is less likely to exist if the features are uncorrelated. To avoid this problem, I had already carefully picked variables and all the variables in our models are uncorrelated with each other. Therefore, we can put high degree of confidence in our OOB permuted measure of feature importance.

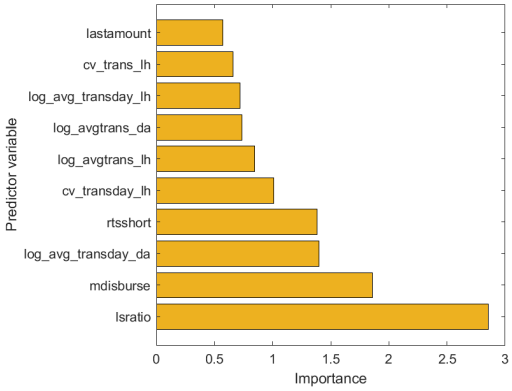
Figure 5 presents top 10 variables in predicting the three adverse credit events. Note that out of the total 20 variables and 2 sets of dummy variables in the model, there are nine PF Basic variables and remaining are business and loan characteristics. However, top 10 spots in the most important variables is completely dominated by the PF Basic variables. This is not surprising because we saw in previous sections most the performance improvement is driven by the PF Basic model. Among those in top 10, loan-sales ratio (`lsratio`) is the most important variable predicting default, followed by growth in average daily sales (`log_avg_transday_da`) and the level of average daily sales in the 90-day window before disbursal (`log_avg_transday_lh`). Average transaction size (`log_avgtrans_lh`) and variation in daily sales (`cv_transday_lh`) are also important. In predicting non-performing loans, 8 out of top 10 variables are payment footprint variables and the previously discussed variables remain important. Among the non-PF variables, month of disbursal (`mdisburse`) and days-past-due on the previous loan (`lastidpd`) are the important ones. I run the test of importance also on the model PF Basic + credit scores and find that neither credit score (`scorecond`) nor the indicator variable on whether the loan is scored (`isscored`) feature in top 10 variables (result not shown due to space reasons).

Figure 5: Top 10 Most Important Variables for Out-of-Sample Predictions

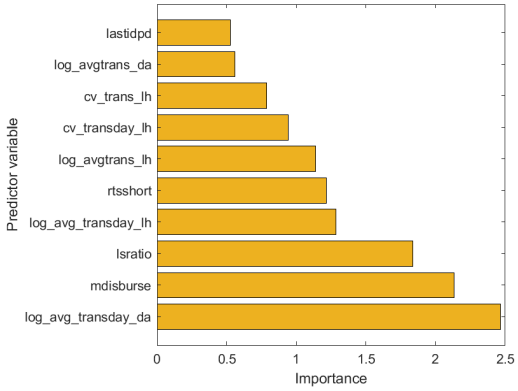
Model: *Payment Footprint Basic + Loan & Business characteristics*



(a) Default



(b) Late



(c) Non-performing

Variable importance is calculated as the increase in the out-of-bag prediction error of the random forest model following a random permutation of the predictor values. A higher *out-of-bag error* following random permutation indicates higher importance of the variable. Figure presents top 10 most important variables by this criteria. *Default* is a binary variable that takes value 1, if the loan had a shortfall $>$ 5% of repayment amount and it was either written off or still pending as of end 2019. *Late* is a binary variable that takes value 1 if a loan was non-defaulting and was took at least 30 days more than the implied tenure to fully pay the loan off. Implied tenure is the number of days it would take to repay the loan (loan amount + interest) given the 10% deduction rate and if the merchant after disbursal continued to transact at their pre-disbursal long-term average. Pre-disbursal long-term average is average daily transaction value calculated over 91 days between day 120 and day 30 before loan disbursal. *Non-performing* is a binary variable that takes value 1 when either the loan is categorized as Default or Late. Basic PF + business & loan characteristics combines Basic PF variables with non-credit bureau variables of Traditional model. For details on models and variables see Table A3 and Table A4.

3.3.2 Shapley Additive Explanations (SHAP)

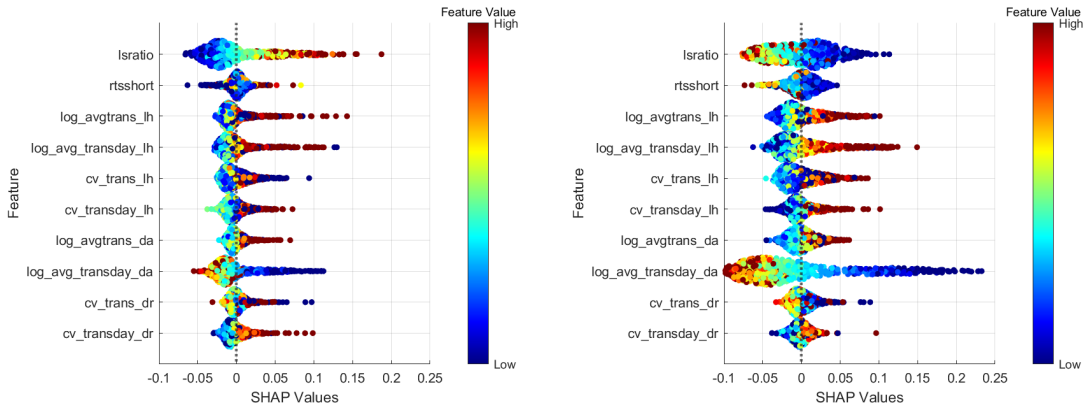
SHAP gives, for each individual prediction, the *contribution* of each feature in the deviation of the prediction from the average prediction (Lundberg and Lee, 2017). The higher the contribution of a feature, the higher the importance of the feature for that particular prediction.

The idea of computing fair contribution is based on treating prediction as a cooperative game where players (features) combine to produce a surplus (deviation of the prediction from average prediction). Shapley value gives the contribution of each player (feature) in the prediction.

Since calculating Shapley value is computationally very demanding, I take the following two steps to keep the computations manageable. First, from our main results we note that most of the performance gain is already captured in the *PF Basic* model, and other PF based models improve on it only marginally. Therefore, we will calculate SHAP values only for the predictions under PF basic model as it contains fewer variables than other PF based models. Second, instead of calculating SHAP values for all the predictions in the test set, we will do so for a 20% random sample of the test set, still covering more than 600 predictions.

Figure 6: Variable Importance – SHAP Values

Model: *Payment Footprint Basic*



(a) Default (b) Non-performing

Shapley Additive Explanations (SHAP) for a variable and for a given observation captures the contribution of the variable in the deviation of the prediction for the observation from the average prediction. A higher SHAP value for a variable implies higher importance of the variable in the prediction of that instance. The figure plots, for each variable, SHAP values calculated for 615 predictions. These 615 predictions are 20% random sample of the testing set of the random forest estimation applied on the Basic PF model. Each point is colored according the value of the variable (feature) at that observation. For details on models and variables see Table A3 and Table A4.

Figure 6 gives the SHAP values for more than 600 out-of-sample predictions. Values are jittered to show the concentration of values at a point. In order to assess which feature is more important, we should look for those variables that are most spread out as higher absolute values indicate larger (contribution to the) deviation from the average. Much in line with the permutation measures, we find that loan-sales ratio, growth in average daily sales, and level of average daily sales are crucial variables for predicting default and non-performing loans.

SHAP is a local measure of importance in the sense that SHAP values are contributions calculated for individual predictions. To calculate global importance for a variable, I take the average of the absolute Shapley values of the variable across all predictions. Figure B1 in the

appendix again confirms our previous results that loan-sales ratio and growth in average daily sales are the most important variables.

Another use of SHAP values is that we can relate the SHAP values for a feature with feature values. Such an exercise informs us of the direction of relationship between the feature and its impact on the predicted probability of the adverse event. Figure 6 color codes the SHAP values with the underlying feature value. From the figure we can see that higher value of loan-sales ratio is associated with higher probability of default (higher SHAP values). Similarly, lower values of average daily sales growth is associated with higher positive deviation of default probability from average predicted default probability.

4 Early Warning Exercise

Payment fintechs observe financial health of a borrowing business at a high frequency through the transactions they process for the borrowing merchant. Thus, payment fintechs as lenders are positioned to take advantage of the high frequency signal on borrowers even after loan disbursal. Specifically, they can detect any early signs of negative shocks to borrower's business. In addition, moral hazard—borrower actions after the loan disbursal—is also a key determinant of loan performance (Karlan and Zinman, 2009). Rishabh and Schäublin (2021) show that borrowers may manipulate their sales immediately after loan disbursal in order to default and delay on their repayment. Thus, an early warning system that captures changes in sales may signal an imminent default might and may help lenders to take corrective actions.

Whether and how can post-disbursal transaction data be useful in informing the lender about a default (or a delay)? To study that, I predict the adverse credit events as before, but now by augmenting the historical payment footprints with post-disbursal changes in sales related payment footprints. For example, I augment PF Basic model with post-disbursal changes in following variables: average daily sales, average transaction size, CV of daily sales, and CV of transaction sizes. I augment all other PF based models in a similar way by adding new variables on changes in PF variables. I calculate changes as relative or absolute difference in the variable calculated in 28-day window *after* disbursal over its value calculated in the 28-day window *before* disbursal. Thus, our early warning exercise is an exercise at the 28^{th} day after disbursal.

To test whether payment footprints contain early warning (EW) signals, we compare the AUCs in the baseline models with the AUCs of the EW models. Figure 7 and Table 6 presents the ROC curves for the early warning exercise. We note that compared to the baseline, point estimates of AUCs are higher with the early warning PF models. Notably, the largest improvement compared to the baseline is in predicting late and non-performing loans. For example, consider the baseline model PF Basic + business & loan characteristics that achieves the AUC of 0.68. The same model with EW variables achieves the AUC of 0.75. Therefore, we conclude that PF variables are a useful early warning indicators too.

Another way to check the extent to which the EW variables add to the predictive power is to

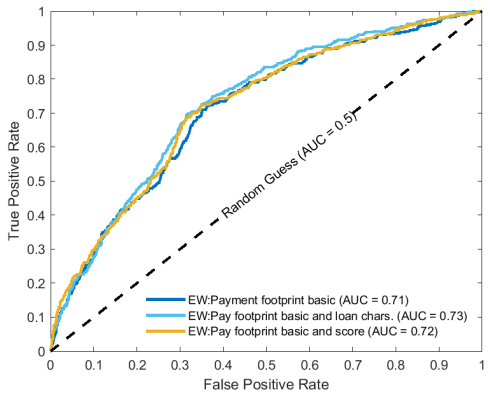
Table 6: Early Warning Exercise – AUC for Out-of-Sample Predictions

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	-	-	-	0.71 [0.68, 0.74]	0.70 [0.67, 0.73]	0.73 [0.70, 0.75]	0.72 [0.69, 0.75]	0.70 [0.67, 0.73]
Late	-	-	-	0.77 [0.75, 0.79]	0.77 [0.75, 0.80]	0.80 [0.78, 0.82]	0.77 [0.75, 0.79]	0.79 [0.77, 0.81]
Non-perf	-	-	-	0.72 [0.70, 0.74]	0.71 [0.69, 0.73]	0.75 [0.73, 0.77]	0.73 [0.70, 0.75]	0.73 [0.71, 0.74]
No. of obs.	-	-	-	11964	11964	11964	11964	11964
No. of vars.	-	-	-	14	72	26	16	99

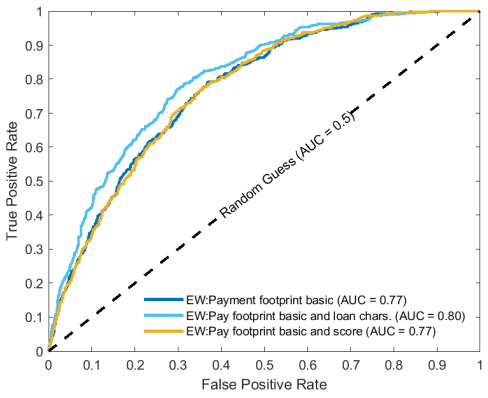
Early warning exercise appends the models in the baseline with the relative or absolute changes in the payment variables post-disbursal. Changes are measured between values calculated in the 28-day window *after* disbursal over the values measured in the 28-day window *before* disbursal. Values in square brackets represent the 95% confidence interval of the AUC. We compute confidence interval by bootstrapping using 1000 bootstrap samples. Number of observations refers to number of observations in the training set. Number of variables does not include the number of dummy variables but only number of dummy classes (two). For a description about models see notes for Table 5. For the complete list of model variables see Table A4. For variable description see Table A3. Predictions are based on a model estimated with random forest algorithm. Random forest estimation is done with 300 trees.

study the importance of EW variables in comparison to the historical variables. To do that I look at OOB permutation measure for the EW augmented model PF Basic + business & loan characteristics. Figure 8 presents top 10 variables from the model in predicting different credit events. Interestingly the post-disbursal growth in average daily sales (`log_avg_transday_pda`)occupies top spot in the list of important variables. For default predictions, four out of top 10 variables are now EW PF variables while remaining are the original (historical) PF variables. In predicting non-performing loans three out of top 10 places are acquired by EW PF variables, six places by historical PF variables, and one spot by business & loan characteristics variable.

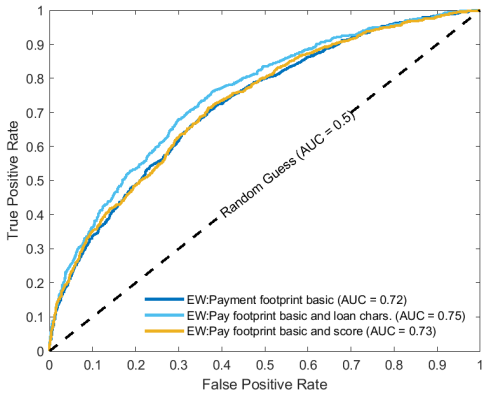
Figure 7: Early warning exercise – ROC Curves for Out-of-Sample Predictions Across Early Warning Models



(a) Default



(b) Late

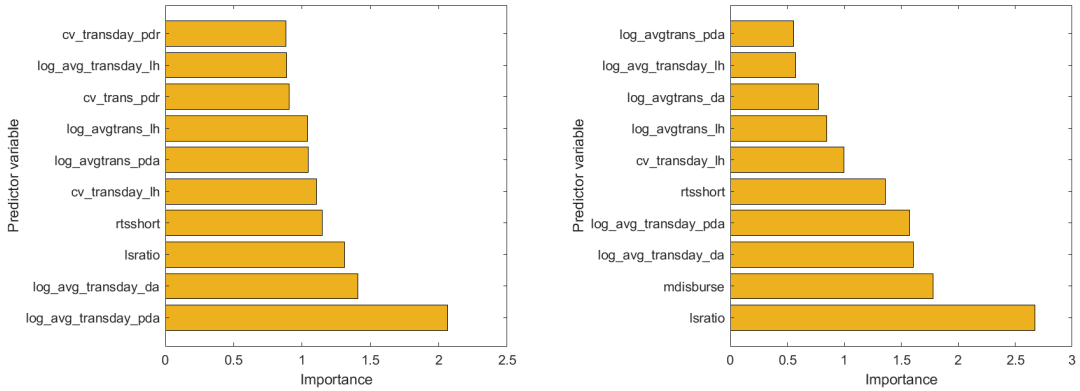


(c) Non-performing

Early warning exercise appends the models in the baseline (Figure 2) with the relative or absolute changes in the payment variables post-disbursal. Changes are measured between values calculated in the 28-day window *after* disbursal over the values measured in the 28-day window *before* disbursal. For a description about models see notes for Figure 2. For the complete list of model variables see Table A4. For variable description see Table A3. Predictions are based on a model estimated with random forest algorithm. Random forest estimation is done with 300 trees.

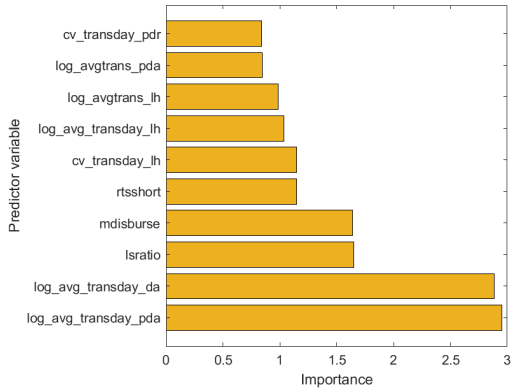
Figure 8: Early Warning Exercise – Top 10 Most Important Variables for Out-of-Sample Predictions

Early Warning Augmented Model: *Payment Footprint Basic + Loan & Business characteristics*



(a) Default

(b) Late



(c) Non-performing

Variable importance is calculated as the increase in the out-of-bag prediction error of the random forest model following a random permutation of the predictor values. Early warning exercise appends the models in the baseline (Figure 2) with the relative or absolute changes in the payment variables post-disbursal. Changes are measured between values calculated in the 28-day window *after* disbursal over the values measured in the 28-day window *before* disbursal. For a description about models see notes for Figure 2. For the complete list of model variables see Table A4. For variable description see Table A3. For this figure the baseline model Basic PF + business & loan characteristics is augmented with early warning variables. For details on models and variables see Table A3 and Table A4.

5 Payment Footprints and Traditional Loans

The loans we considered in our analysis so far are made by a payment fintech. For the payment-fintech loans, the repayment on the loans is linked to the sales that the fintech processes. Therefore, it is possible that payment footprints capture signals that are useful only for these specific type of loans. To study the usefulness of payment footprints in other contexts, I now conduct a predictive analysis for loans made by traditional lenders, such as banks and non-banks.

I obtain the data on non-traditional loans from the credit bureau. These loans were taken by the same borrowers that we considered in the previous analysis. The credit bureau data does not provide lender identity. However, given the period under consideration (2015-2019), we can be reasonably certain that the lenders were banks and non-bank credit institutions. More importantly, sales-linked loans by payment fintechs are a more recent phenomenon, and given that the payment fintechs do not report to credit bureau, we can be sure that the loans we consider from the credit bureau are not sales-linked.

For each traditional loan in my dataset, I observe information such as loan disbursal date, loan amount, loan type (e.g. personal loan, business loan, credit card, mortgage etc.), loan tenure, and loan repayment history as reported by the lenders to the bureau. The reported repayment history allows me to create measures of loan performance, and adverse credit events. I define two adverse credit events. I categorize a loan as *default* if the loan was either written off, or was classified as *Loss*, *Substandard*, *Doubtful*, or *Special Mention Account* by the lender. I call a loan *non-performing* when it is categorized as Default or had days past due of at least 90 days.

I construct the payment footprint variables just as before, but now in relation to the loan disbursal date of the traditional loans. I restrict my analysis to business loans. Where I define business loan as any loan where the reported loan type is personal loan by the owner, or business loan, or *gold loan*. Gold loans are a common form of financing among Indian MSMEs (Asokan, 2020; Singh and Wasdani, 2016). These are loans that are collateralized by gold (mostly gold jewellery). I perform the preliminary analysis with Business loans so as to keep the set of loans comparable to the previous analysis. I provide a snapshot information on the traditional loans in the summary statistics presented in Table 7. The table presents summary statistics on Business loans including gold loans, without gold loans, and all loans.

I perform the predictive analysis as before. This analysis is on going. Therefore, these results are very preliminary. Figure 9 reports the ROC curves and their corresponding AUCs for different predictive models. Some interesting observations are in order. First, credit scores perform better for traditional loans than fintech loans. The AUC of the credit score model in predicting default or non-performing loan is now around 0.6, which is higher than its AUCs when predicting credit events for payment-fintech loans (0.53 to 0.57).

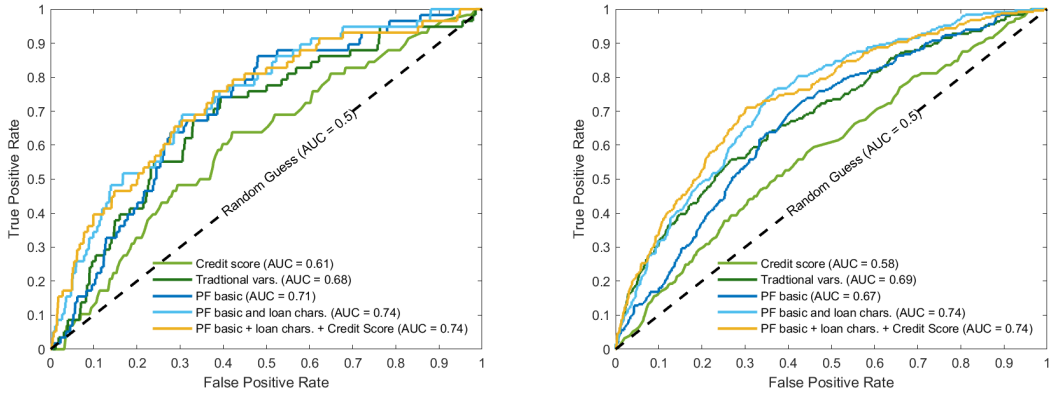
Second, payment footprints continue to perform very well, achieving an AUC of 0.71 for

Table 7: Summary Statistics on Traditional Loans

	All loans	Business loans	Business Loans w/o gold loans
No. of Loans	32011	12406	7984
Loan amount ('000s)	187.3	192.5	195.3
Tenure	27.9	18.1	23.3
Interest rate (% p.a.)	25.9	19.3	26.1
Loan is collateralized (1 = yes)	0.28	0.46	0.16
Loan is Non-performing (1 = Yes)	0.101	0.098	0.125
Loan is Default (1 = Yes)	0.032	0.022	0.030
Credit Score	720.5	716.9	714.8

Values reported are the mean, except for number of loans. Traditional loans refer to loans given by banks and non-bank financial institutions, and reported to the credit bureau. CIBIL credit score ranges between 300 and 900, with score above 700 considered to be good by the financial institutions. *Default* is a binary variable that takes value 1, if the loan was either written off or was classified as either *Loss*, *Substandard*, or *Doubtful* by the lender. *Non-performing* is a binary variable that takes value 1 when either the loan is categorized as Default or had days past due of at least 90 days.

Figure 9: ROC Curves for Out-of-Sample Predictions Across Models for Traditional Loans



(a) Default (b) Non-performing

For the figure on default predictions, false positive rate is the proportion of non-default loans incorrectly classified as default by the model predictions at the chosen threshold. True positive rate is the proportion of default loans correctly classified as default by the model predictions at the chosen threshold. Similar interpretation applies for Late and Non-performing predictions. *Default* is a binary variable that takes value 1, if the loan was either written off or was classified as either *Loss*, *Substandard*, or *Doubtful* by the lender. *Non-performing* is a binary variable that takes value 1 when either the loan is categorized as Default or had days past due of at least 90 days. Credit score model includes credit score and a variable identifying whether the borrower had a credit score. Traditional model consists of variables traditionally used in loan underwriting and covers credit scores, and variables relating to loan characteristics (e.g. loan amount), business characteristics (e.g. industry). Basic payment footprint (PF) model captures basic variables on historical sales such as average daily sales, coefficient of variation of sales, loan-sales ratio etc. Combined models combine the elements of traditional variables with the PF variables. The model *Basic PF + loan chars. + credit score* includes business & loan characteristics in combination with Basic PF variables and credit scores. Out-of-sample predictions were made after training a Random forest model. Random forest estimation is done with 500 trees. Loans considered for this exercise include business loans, personal loans, and loan collateralized by gold (*gold loans*).

default prediction, and 0.67 for predicting non-performance. Finally, adding credit scores to payment footprint model, does not improve predictability. Therefore, we conclude that payment footprints provide a useful screening tool even for traditional loans (albeit, with the admission that the results are preliminary).

***Note on traditional-loan analysis**:* Section on traditional loans will assume more prominence (it may even become the baseline) as I develop the analysis further. Since I acquired this data more recently, the analysis on this dataset is ongoing, and currently I have only some preliminary results. If payment footprints show high predictability for the traditional loans (preliminary results suggest that they do), it establishes their usefulness outside of the payment fintech loans. Another, useful feature of the traditional loan dataset from bureau is that there is heterogeneity also on the dimension of loan types. For instance, we observe business loans, consumer loans, collateralized loans, and credit cards. We can then test the relative usefulness of payment footprints in for different loan markets.

6 Conclusion

Using a unique setting that allows me to cleanly identify the relative predictive performance of credit scores and payment footprints, I find that the latter substitute for the former. This result has policy implications. Setting up credit information sharing institutions such as credit bureau and credit registries is costly. More than 50% of the firms and individuals globally are still unlisted under any bureau or public registry. On the other hand electronic payments are on a rise in their own right everywhere. Payment footprints are cheaply obtainable because they are the by-product of the the electronic transactions. Given the substitutability of credit scores with payment data, and the relative cost efficiency of the latter, my results imply that promoting electronic payments and open banking should be taken up as policies to expand financial access.

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*.

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.

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

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

Bianchi, Milo, Matthieu Bouvard, Renato Gomes, Andrew Rhodes, and Vatsala Shreeti. 2021. "Mobile Payments and Interoperability: Insights from the Academic Literature."

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.

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*.

Djankov, Simeon, Caralee McLiesh, and Andrei Shleifer. 2007. "Private credit in 129 countries." *Journal of Financial Economics*, 84(2): 299–329.

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

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

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

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. 2020. "Open banking: credit market competition when borrowers own the data." National Bureau of Economic Research.

Hooker, Giles, Lucas Mentch, and Siyu Zhou. 2021. "Unrestricted Permutation forces Extrapolation: Variable Importance Requires at least One More Model, or There Is No Free Variable Importance."

Jones, Charles I, and Christopher Tonetti. 2020. "Nonrivalry and the Economics of Data." *American Economic Review*, 110(9): 2819–58.

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

Liberti, José María, and Mitchell A Petersen. 2019. "Information: Hard and soft." *Review of Corporate Finance Studies*, 8(1): 1–41.

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.

Molnar, Christoph. 2020. *Interpretable machine learning*. Lulu. com.

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.

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

Parlour, Christine A, Uday Rajan, and Haoxiang Zhu. 2020. "When fintech competes for payment flows." *Available at SSRN 3544981*.

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).

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*.

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

Table A1: Area Under the ROC Curve for Out-of-Sample Predictions – by Size

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	0.56	0.60	0.67	0.64	0.65	0.68	0.66	0.67
	[0.51, 0.60]	[0.56, 0.64]	[0.62, 0.71]	[0.60, 0.68]	[0.61, 0.70]	[0.64, 0.72]	[0.61, 0.69]	[0.63, 0.71]
Late	0.48	0.51	0.61	0.66	0.69	0.70	0.67	0.71
	[0.45, 0.52]	[0.47, 0.54]	[0.58, 0.65]	[0.63, 0.69]	[0.66, 0.72]	[0.67, 0.73]	[0.63, 0.70]	[0.68, 0.74]
Non-perf	0.53	0.56	0.62	0.65	0.66	0.68	0.66	0.67
	[0.50, 0.56]	[0.53, 0.59]	[0.59, 0.65]	[0.62, 0.68]	[0.63, 0.69]	[0.65, 0.71]	[0.63, 0.69]	[0.65, 0.70]
No. of obs.	5087	5756	6151	6151	6151	6151	6151	6151
No. of vars.	2	9	22	10	51	22	12	76

(a) Large Borrowers

	Credit Bureau		Traditional	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	0.57	0.57	0.67	0.70	0.69	0.72	0.72	0.70
	[0.52, 0.61]	[0.53, 0.61]	[0.63, 0.71]	[0.66, 0.74]	[0.65, 0.72]	[0.68, 0.75]	[0.69, 0.75]	[0.66, 0.74]
Late	0.50	0.53	0.62	0.72	0.73	0.75	0.71	0.74
	[0.46, 0.54]	[0.50, 0.58]	[0.57, 0.66]	[0.69, 0.76]	[0.69, 0.76]	[0.72, 0.78]	[0.68, 0.75]	[0.70, 0.77]
Non-perf	0.55	0.54	0.59	0.65	0.63	0.68	0.66	0.66
	[0.52, 0.59]	[0.51, 0.57]	[0.56, 0.62]	[0.62, 0.68]	[0.60, 0.66]	[0.65, 0.71]	[0.63, 0.69]	[0.63, 0.69]
No. of obs.	5095	5812	6150	6150	6150	6150	6150	6150
No. of vars.	2	9	22	10	51	22	12	76

(b) Small Borrowers

Large borrowers are those that had above median total value of transactions in the 90-day period before disbursal. Small borrowers had total value of transactions below median. Values in square brackets represent the 95% confidence interval of the AUC. We compute confidence interval by bootstrapping using 1000 bootstrap samples. Number of observations refers to number of observations in the training set. Number of variables does not include the number of dummy variables but only number of dummy classes (two). For a description about models see notes for Table 5. For the complete list of model variables see Table A4. For variable description see Table A3. Predictions are based on a model estimated with random forest algorithm. Random forest estimation is done with 300 trees.

Table A2: Area Under the ROC Curve for Out-of-Sample Predictions – by Credit Score

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	0.57	0.57	0.61	0.69	0.67	0.70	0.70	0.68
	[0.52, 0.61]	[0.53, 0.62]	[0.56, 0.65]	[0.65, 0.73]	[0.63, 0.71]	[0.66, 0.75]	[0.65, 0.74]	[0.64, 0.72]
Late	0.50	0.51	0.65	0.70	0.73	0.75	0.70	0.75
	[0.47, 0.54]	[0.47, 0.55]	[0.61, 0.68]	[0.67, 0.74]	[0.69, 0.75]	[0.72, 0.78]	[0.67, 0.73]	[0.73, 0.78]
Non-perf	0.55	0.53	0.63	0.65	0.67	0.69	0.66	0.69
	[0.51, 0.58]	[0.50, 0.57]	[0.59, 0.66]	[0.62, 0.68]	[0.64, 0.70]	[0.65, 0.72]	[0.63, 0.69]	[0.65, 0.71]
No. of obs.	6092	6092	6092	6092	6092	6092	6092	6092
No. of vars.	2	9	22	10	51	22	12	76

(a) High Score

	Credit Bureau		Traditional	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	0.49	0.57	0.62	0.70	0.69	0.71	0.70	0.71
	[0.45, 0.55]	[0.51, 0.62]	[0.56, 0.67]	[0.64, 0.75]	[0.63, 0.74]	[0.66, 0.76]	[0.65, 0.75]	[0.66, 0.76]
Late	0.51	0.50	0.61	0.74	0.77	0.80	0.74	0.80
	[0.44, 0.56]	[0.44, 0.55]	[0.56, 0.67]	[0.69, 0.78]	[0.73, 0.81]	[0.75, 0.83]	[0.69, 0.78]	[0.75, 0.83]
Non-perf	0.51	0.53	0.63	0.68	0.69	0.69	0.68	0.70
	[0.47, 0.56]	[0.49, 0.57]	[0.58, 0.67]	[0.63, 0.72]	[0.65, 0.73]	[0.65, 0.73]	[0.64, 0.72]	[0.67, 0.74]
No. of obs.	3129	3129	3129	3129	3129	3129	3129	3129
No. of vars.	2	9	22	10	51	22	12	76

(b) Low Score

	Credit Bureau		Traditional	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	-	-	0.65	0.59	0.53	0.61	-	0.58
	-	-	[0.54, 0.75]	[0.49, 0.69]	[0.41, 0.65]	[0.50, 0.70]	-	[0.47, 0.70]
Late	-	-	0.63	0.74	0.74	0.76	-	0.74
	-	-	[0.52, 0.73]	[0.64, 0.82]	[0.63, 0.81]	[0.65, 0.83]	-	[0.64, 0.81]
Non-perf	-	-	0.61	0.65	0.63	0.66	-	0.67
	-	-	[0.52, 0.69]	[0.57, 0.73]	[0.55, 0.71]	[0.58, 0.74]	-	[0.59, 0.74]
No. of obs.	-	-	976	976	976	976	-	976
No. of vars.	-	-	13	10	51	22	-	67

(c) Unscored

Credit scores correspond to that of the merchant owning the business. Credit scores range between 300 and 900. Scores above 700 are assessed as good by the credit market. For the unscored loans, the borrowers did not have a long enough credit history at the time of the borrowing to have been assigned any score by the credit bureau. For more details on models and variables see Table 5. Values in square brackets represent the 95% confidence interval of the AUC. We compute confidence interval by bootstrapping using 1000 bootstrap samples. Number of observations refers to number of observations in the training set. Number of variables does not include the number of dummy variables but only number of dummy classes (two). For a description about models see notes for Table 5. For the complete list of model variables see Table A4. For variable description see Table A3. Predictions are based on a model estimated with random forest algorithm. Random forest estimation is done with 300 trees.

Table A3: Variable Description

S. No.	Variable Name	Variable Description
1	clolast	(For repeat borrowers) Number of days between disbursal of the repeat loan and closure of the previous loan
2	lastamount	(For repeat borrowers) Principal amount of the previous loan
3	lastidpd	(For repeat borrowers) implied days past due for previous loan
4	lastimpislate30	(For repeat borrowers) 1 if the previous loan was late on implied criteria (lastidpd $>$ 30)
5	loannumber	(For repeat borrowers) Loan number (first loan, second loan, ...)
6	relationage	Age of relationship with the payment company at the time of disbursal (in months)
7	age	Age of the business owner at the time of the payment company loan
8	gender	Dummy variable assigning Male $=$ 1
9	mdisburse	Month of loan disbursal
10	loanamount	Principal amount
11	tenure	Suggested tenure (90 or 180 days)
12	hasimproved	(For repeat borrowers) Dummy if score improved over previous loan. Also considers qualitative changes like improvement from unscored and deterioration to unscored.
13	isscored	Dummy that takes value 1 If bureau record could be accessed and credit history was sufficient to be assigned a credit score
14	lengthhist	Credit history length. Days since the first credit record in the bureau report for the borrower at the time of the payment company loan.
15	lsratio	Loan-sales ratio. Loan amount divided by average daily sales calculated in the 91-day period before disbursal.
16	nenq30	Number of enquiries by financial institutions from the bureau 30 days prior to the payment company loan.
17	nloans	Number of credit accounts on the bureau report. Credit accounts include loans and credit cards accounts.
18	rtsshort	Loan-sales ratio calculated with average daily sales calculated in the 28-day period before disbursal.
19	scorecond	CIBIL credit score conditional on being scored. Credit score varies between 300 and 900.
20	shcollclosed	Share of collateralized loan accounts in closed loan accounts
21	shnpclosed	Share of non-perf loan accounts in all closed loan accounts
22	shrunning	Share of loan accounts reported in the bureau report that were running at the time of payment-company loan.
23	max_trans_dt_lh	(Relative to disbursal) Day with the largest transactions value.
24	max_transcount_dt_lh	(Relative to disbursal) Day with most number of transactions.
25	dayspast_freqcust_lh	Days since last transaction of most frequent customer at the time of disbursal.
26	dayspast_largcust_lh	Days since last transaction of largest customer at the time of disbursal.
27	log_custcount_lh	Number of distinct customers within period (in log).
28	log_avgtrans_lh	Average transaction sizes (in log).
29	log_avg_transday_lh	Average of total daily transaction value (in log).
30	cv_trans_lh	Coefficient of variation of transaction sizes.
31	cv_transday_lh	Coefficient of variation of total daily transaction values.
32	HH_cust_trans_lh	Herfindahl-Hirschmann index of customers total transaction value.
33	avg_inactivity_lh	Average of periods without any transactions (# of inactive days/total number of transaction days)
34	days_since_max_inactivity_lh	Days since the period of longest inactivity
35	days_since_lasttrans_lh	Days since last transaction
36	shr_Sun_trans_lh	Share of total transaction value conducted on Sundays
37	shr_Mon_trans_lh	Share of total transaction value conducted on Mondays
38	shr_Tue_trans_lh	Share of total transaction value conducted on Tuesdays
39	shr_Wed_trans_lh	Share of total transaction value conducted on Wednesdays
40	shr_Thu_trans_lh	Share of total transaction value conducted on Thursdays
41	shr_Fri_trans_lh	Share of total transaction value conducted on Fridays
42	shr_Sat_trans_lh	Share of total transaction value conducted on Saturdays

(Table continued...)

Table A3: Variable description (continued)

S. No.	Variable Name	Variable Description
43	shr_AMEX_trans_lh	Share of total transaction value transacted with AMEX cards
44	shr_MAESTRO_trans_lh	Share of total transaction value transacted with Maestro cards
45	shr_MASTERCARDVISA_trans_lh	Share of total transaction value transacted with Visa or Mastercard
46	shr_OTHER_trans_lh	Share of total transaction value transacted with Other cards (or Missing)
47	shr_fallbacktrans_lh	Share of fallback transaction value in total transaction value
48	lastfallback_days_lh	Days since last fallback transaction.
49	lastfallback_amt_relavg_lh	Amount of last fallback transaction, relative to average transaction size.
50	shr_refundtrans_lh	Share of transaction value of refund transactions in total transacted value.
51 to 71	"..._da", "..._dr", "log_...._da"	Relative changes (with subscript _dr) in payment variables measured in levels above. Absolute changes (with subscript _da) in payment variables measured as a share above. Changes are calculated between 28-day period prior to disbursal and 28-day period spanning 63-day to 91-day before disbursal. Absolute change in log variables (log_...._da) also approximate relative change.
72	log_dis_avg_transday_lh	Average per-day total transactions at the district level. Total district leve transactions are transactions value summed over all the merchants in the borrower's district.
73	shr_dis_sumtrans_lh	Share of the borrowing merchant in the district total transactions.
74 and 75	"log_...._da", "..._dr",	Absolute and relative changes respectively in the above two variables.
76 to 106	sg1-sg31	Dummy for mcc subgroups (industry)
107 to 141	s1-s35	Dummy for state and union territories of India

Variables ending with subscript "_lh" are calculated in the 91-day window before disbursal. Variable ending with "_dr" and "_da" represent relative and absolute changes in values of the payment variables calculated in the 28-day period before disbursal and values calculated in the 63-day to 91-day period before disbursal.
Absolute changes in log variables approximate the relative change in the variable.

Table A4: Model Variables

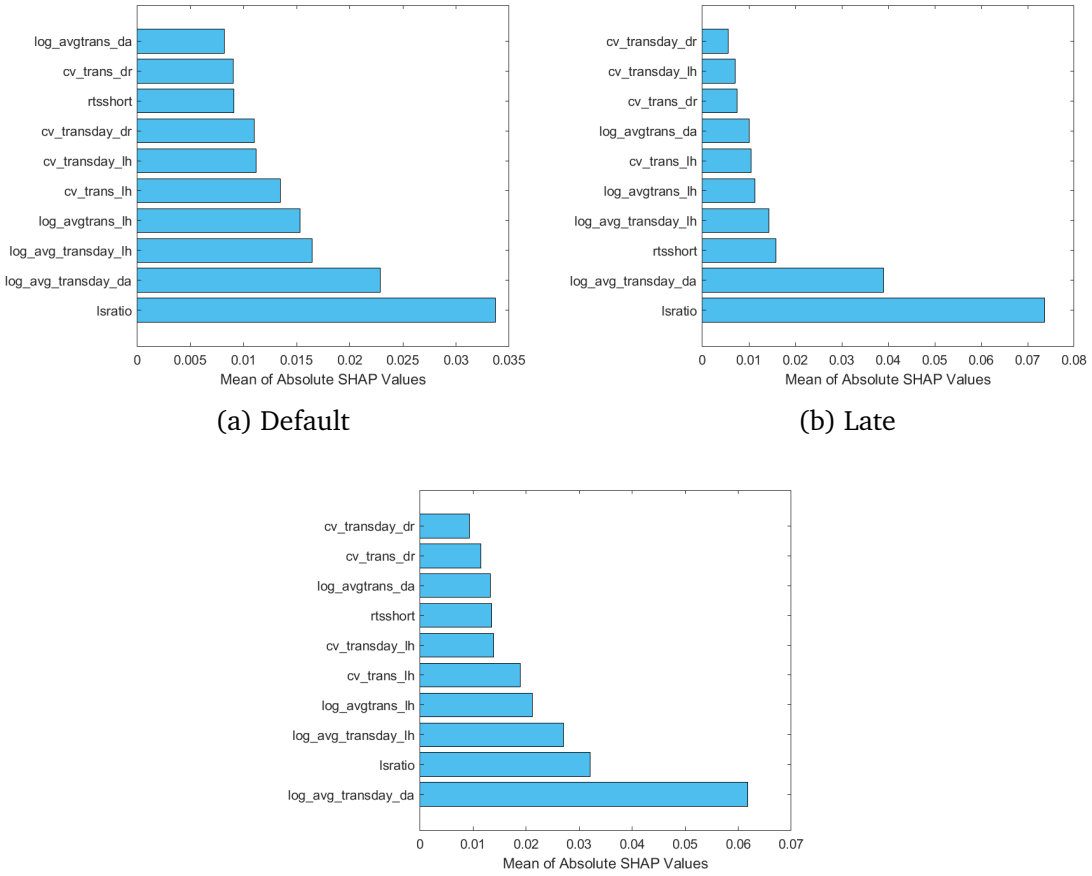
		Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
		Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
S. No.	**Variable Name**	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
1	clolast			✓			✓		✓
2	lastamount			✓			✓		✓
3	lastidpd			✓			✓		✓
4	lastimpislate30			✓			✓		✓
5	loannumber			✓			✓		✓
6	relationage			✓			✓		✓
7	age			✓			✓		✓
8	gender			✓			✓		✓
9	mdisburse			✓			✓		✓
10	loanamount			✓			✓		✓
11	tenure			✓			✓		✓
12	hasimproved		✓	✓					✓
13	isscored	✓	✓	✓				✓	✓
14	lengthhist		✓	✓					✓
15	lsratio				✓	✓	✓	✓	✓
16	nenq30		✓	✓					✓
17	nloans		✓	✓					✓
18	rtsshort				✓	✓	✓	✓	✓
19	scorecond	✓	✓	✓				✓	✓
20	shcollclosed		✓	✓					✓
21	shnpclosed		✓	✓					✓
22	shrunning		✓	✓					✓
23	max_trans_dt_lh					✓			✓
24	max_transcount_dt_lh					✓			✓
25	dayspast_freqcust_lh					✓			✓
26	dayspast_largcust_lh					✓			✓
27	log_custcount_lh					✓			✓
28	log_avgtrans_lh				✓	✓	✓	✓	✓
29	log_avg_transday_lh				✓	✓	✓	✓	✓
30	cv_trans_lh				✓	✓	✓	✓	✓
31	cv_transday_lh				✓	✓	✓	✓	✓
32	HH_cust_trans_lh					✓			✓
33	avg_inactivity_lh					✓			✓
34	days_since_max_inactivity_lh					✓			✓
35	days_since_lasttrans_lh					✓			✓
36	shr_Sun_trans_lh					✓			✓
37	shr_Mon_trans_lh					✓			✓
38	shr_Tue_trans_lh					✓			✓
39	shr_Wed_trans_lh					✓			✓
40	shr_Thu_trans_lh					✓			✓
41	shr_Fri_trans_lh					✓			✓
42	shr_Sat_trans_lh					✓			✓
43	shr_AMEX_trans_lh					✓			✓
44	shr_MAESTRO_trans_lh					✓			✓
45	shr_MASTERCARDVISA_trans_lh					✓			✓
46	shr_OTHER_trans_lh					✓			✓
47	shr_fallbacktrans_lh					✓			✓
48	lastfallback_days_lh					✓			✓
49	lastfallback_amt_relavg_lh					✓			✓
50	shr_refundtrans_lh					✓			✓
51 to 71	"…_da", "…_dr", "log_…._da"				✓	✓	✓	✓	✓
72	log_dis_avg_transday_lh								✓
73	shr_dis_sumtrans_lh								✓
74 and 75	"log_…._da", "…_dr",								✓
76 to 106	sg1-sg31			✓			✓		✓
107 to 141	s1-s35			✓			✓		✓

See Table A3 for variable descriptions. Variables ending with subscript "_lh" are calculated in the 91-day window before disbursal. Variable ending with "_dr" and "_da" represent relative and absolute changes in values of the payment variables calculated in the 28-day period before disbursal and values calculated in the 63-day to 91-day period before disbursal. Absolute changes in log variables approximate the relative change in the variable.

B Robustness

Figure B1: Variable Importance – Mean Absolute SHAP Values

Model: *Payment Footprint Basic*



(a) Default

(b) Late

(c) Non-performing

Shapley Additive Explanations (SHAP) for a variable and for a given observation captures the contribution of the variable in the deviation of the prediction for the observation from the average prediction. A higher SHAP value for a variable implies higher importance of the variable in the prediction of that instance. The figure plots, for each variable, mean of the absolute SHAP values calculated across 615 predictions. These 615 predictions are 20% random sample of the testing set of the random forest estimation applied on the Basic PF model. For details on models and variables see Table A3 and Table A4.

Table B1: Area Under the ROC Curve for Out-of-Sample Predictions with Same Sample Across Models

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted		score + othr.	+ loan chr.		PF basic + other	+ loan chr.	+ score	
variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Default	0.56	0.58	0.60	0.64	0.66	0.65	0.65	0.67
	[0.52, 0.61]	[0.53, 0.64]	[0.55, 0.65]	[0.60, 0.70]	[0.61, 0.71]	[0.60, 0.70]	[0.60, 0.70]	[0.63, 0.72]
Late	0.52	0.50	0.59	0.75	0.76	0.79	0.75	0.78
	[0.48, 0.57]	[0.46, 0.55]	[0.55, 0.64]	[0.72, 0.78]	[0.73, 0.79]	[0.75, 0.81]	[0.71, 0.78]	[0.75, 0.81]
Non-perf	0.52	0.52	0.57	0.66	0.68	0.69	0.66	0.68
	[0.49, 0.56]	[0.49, 0.56]	[0.54, 0.61]	[0.62, 0.69]	[0.65, 0.72]	[0.66, 0.72]	[0.63, 0.69]	[0.65, 0.72]
No. of obs.	4840	4840	4840	4840	4840	4840	4840	4840
No. of vars.	2	9	22	10	51	22	12	76

Only those observations that have complete data on *all* the variables are admitted in the exercise. Values in square brackets represent the 95% confidence interval of the AUC. We compute confidence interval by bootstrapping using 1000 bootstrap samples. Number of observations refers to number of observations in the training set. Number of variables does not include the number of dummy variables but only number of dummy classes (two). For a description about models see notes for Table 5. For the complete list of model variables see Table A4. For variable description see Table A3. The predictions are based on a model estimated with random forest algorithm. Random forest estimation is done with 300 trees.

Table B2: Area Under the ROC Curve for Out-of-Sample Predictions – Neural Network and Logit Regression

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	0.58	0.59	0.63	0.68	0.66	0.69	0.69	0.64
	[0.55, 0.62]	[0.55, 0.63]	[0.59, 0.67]	[0.65, 0.71]	[0.62, 0.69]	[0.65, 0.72]	[0.66, 0.72]	[0.60, 0.69]
Late	0.50	0.53	0.58	0.70	0.72	0.72	0.70	0.72
	[0.46, 0.53]	[0.49, 0.57]	[0.54, 0.61]	[0.68, 0.72]	[0.69, 0.75]	[0.69, 0.74]	[0.68, 0.73]	[0.68, 0.75]
Non-perf	0.55	0.52	0.60	0.63	0.62	0.64	0.63	0.62
	[0.52, 0.57]	[0.49, 0.55]	[0.57, 0.62]	[0.61, 0.65]	[0.59, 0.65]	[0.61, 0.66]	[0.61, 0.66]	[0.59, 0.66]
No. of obs.	9213	7129	7123	11946	8105	10274	8938	4822
No. of vars.	2	9	22	10	51	22	12	76

(a) Neural Network

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	0.58	0.60	0.64	0.66	0.63	0.67	0.66	0.64
	[0.55, 0.61]	[0.56, 0.64]	[0.60, 0.67]	[0.62, 0.69]	[0.59, 0.67]	[0.64, 0.70]	[0.62, 0.69]	[0.59, 0.68]
Late	0.50	0.50	0.57	0.70	0.73	0.72	0.70	0.72
	[0.47, 0.53]	[0.47, 0.54]	[0.54, 0.61]	[0.68, 0.72]	[0.70, 0.75]	[0.69, 0.74]	[0.68, 0.73]	[0.69, 0.76]
Non-perf	0.55	0.57	0.60	0.64	0.63	0.64	0.64	0.64
	[0.52, 0.57]	[0.54, 0.60]	[0.57, 0.63]	[0.61, 0.66]	[0.60, 0.65]	[0.62, 0.67]	[0.61, 0.66]	[0.60, 0.67]
No. of obs.	9213	7129	7123	11946	8105	10274	8938	4822
No. of vars.	2	9	22	10	51	22	12	76

(b) Logit Regression

Values in square brackets represent the 95% confidence interval of the AUC. We compute confidence interval by bootstrapping using 1000 bootstrap samples. Number of observations refers to number of observations in the training set. Number of variables does not include the number of dummy variables but only number of dummy classes (two). *Default* is a binary variable that takes value 1, if the loan had a shortfall > 5% of repayment amount and it was either written off or still pending as of end 2019. *Late* is a binary variable that takes value 1 if a loan was non-defaulting and was took at least 30 days more than the implied tenure to fully pay the loan off. Implied tenure is the number of days it would take to repay the loan (loan amount + interest) given the 10% deduction rate and if the merchant after disbursal continued to transact at their pre-disbursal long-term average. Pre-disbursal long-term average is average daily transaction value calculated over 91 days between day 120 and day 30 before loan disbursal. *Non-performing* is a binary variable that takes value 1 when either the loan is categorized as Default or Late. Number of observations refer to the number of observations used in training the model. Number of variables refers to the number of features (explanatory variables) included in the prediction model. Credit score model (column 1) consists of two features – (a) whether credit score exists at the time of borrowing, and (b) credit score. Extended credit bureau model (column 2) includes credit score variables and other variables derived from credit report such as number of credit accounts reported in the bureau report, number of collateralized credit accounts, and number of non-performing accounts in the bureau report etc. Traditional model (column 3) consists of variables traditionally used in loan underwriting and covers bureau variables, and variables relating to loan characteristics (e.g. loan amount), business characteristics (e.g. industry) and relationship characteristics (e.g. loan number). Basic payment footprint (PF) model (column 4) captures basic variables on historical sales such as average daily sales, coefficient of variation of sales, loan-sales ratio etc. Payment footprint deep model (column 5) covers deeper information about the business that can be gathered from the transaction data such as customer base, concentration of customers, card types swiped etc. Combined models combine the traditional variables with the PF variables. Basic PF + business & loan characteristics (column 6) combines variables for model in column (4) with non-credit bureau variables of model in column (3). Basic footprint + credit score model (column 7) checks the predictability of easily constructible basic footprint variables in combination with credit score. Model All (column 8) is the most comprehensive and includes all variables. For details on models and variables see Table A3 and Table A4. Neural Network estimation is done with one hidden layer.

Table B3: Early Warning Exercise – AUC for Out-of-Sample Predictions – Neural Network and Logit Regression

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	-	-	-	0.72	0.70	0.72	0.72	0.68
				[0.69, 0.75]	[0.66, 0.74]	[0.68, 0.75]	[0.68, 0.75]	[0.63, 0.73]
Late	-	-	-	0.77	0.79	0.79	0.76	0.77
				[0.75, 0.79]	[0.76, 0.81]	[0.77, 0.81]	[0.74, 0.79]	[0.73, 0.80]
Non-perf	-	-	-	0.72	0.72	0.73	0.72	0.69
				[0.70, 0.74]	[0.69, 0.74]	[0.71, 0.75]	[0.69, 0.74]	[0.66, 0.72]
No. of obs.	-	-	-	11412	7940	9866	8567	4727
No. of vars.	-	-	-	14	72	26	16	99

(a) Neural Network

	Credit Bureau		Traditional Ext.	Payment Footprint (PF)		Combination of Models		
	Cr. Score	Extended	Bureau. + buss. chr.	Basic	Deep	PF basic + buss. chr.	PF basic	All
Predicted variable	(1)	score + othr. (2)	+ loan chr. (3)	(4)	PF basic + other (5)	+ loan chr. (6)	+ score (7)	(8)
Default	-	-	-	0.71	0.69	0.72	0.72	0.68
				[0.67, 0.74]	[0.65, 0.73]	[0.68, 0.75]	[0.68, 0.75]	[0.63, 0.74]
Late	-	-	-	0.77	0.79	0.80	0.76	0.78
				[0.75, 0.79]	[0.77, 0.81]	[0.78, 0.82]	[0.74, 0.79]	[0.74, 0.81]
Non-perf	-	-	-	0.71	0.71	0.73	0.70	0.69
				[0.69, 0.73]	[0.68, 0.74]	[0.70, 0.75]	[0.68, 0.73]	[0.65, 0.73]
No. of obs.	-	-	-	11412	7940	9866	8567	4727
No. of vars.	-	-	-	14	72	26	16	99

(b) Logit Regression

Early warning exercise appends the models in the baseline with the relative or absolute changes in the payment variables post-disbursal. Changes are measured between values calculated in the 28-day window *after* disbursal over the values measured in the 28-day window *before* disbursal. Values in square brackets represent the 95% confidence interval of the AUC. We compute confidence interval by bootstrapping using 1000 bootstrap samples. Number of observations refers to number of observations in the training set. Number of variables does not include the number of dummy variables but only number of dummy classes (two). For a description about models see notes for Table **??**. For the complete list of model variables see Table A4. For variable description see Table A3. Neural Network estimation is done with one hidden layer.