

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

Kumar Rishabh[†]

Abstract

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

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

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

Open Banking: A Global Shift

Open Banking marks a transformative change in financial services, bridging traditional banking with the data-driven era. By late 2021, it has gained global momentum, with 80 of 168 countries surveyed moving towards its adoption Babina et al. (2024). The industry anticipates its most significant promise to be the enhancement of risk assessments in lending. (Experian, 2022).

Credit Scoring Challenges and Payment Data Potential

Historically, credit scoring systems have relied heavily on credit bureau data, which captures an individual's past financial behavior to predict future creditworthiness. This approach, however, has significant limitations. Notably, it fails to represent over half of the world's population, especially in developing countries, leaving many individuals and small businesses invisible to lenders. Moreover, traditional credit scores often lag in reflecting current financial realities, disadvantaging even those within its coverage.

In contrast, payment history data presents a real-time snapshot of financial activity, offering a dynamic perspective that traditional models lack. This data, derived from the increasing volume of electronic payment transactions, holds the promise of closing the information gap inherent in bureau-based credit scoring. By providing up-to-date insights into a borrower's financial health, payment histories can potentially revolutionize risk assessment, credit pricing, and loan monitoring, making financial services more accessible to underserved populations.

Research Objectives and Challenges

The foundational premise of Open Banking is that by making payment data interoperable, new entrants to the financial market can develop innovative credit products. This assumes that information from one source can significantly improve another institution's understanding of a borrower's financial health. Despite the growing emphasis on Open Banking, there is a notable absence of conclusive research validating these assumptions. This gap underscores the need for comprehensive analysis to evaluate the true value of interoperable payment data in the lending process.

The overarching goal of this paper is to critically assess the value of interoperable payment data. This includes examining whether payment histories can enrich lenders' information mix, their interplay with credit bureau data, and their utility across the lending process—from pre-loan screening to post-disbursal monitoring. A unique aspect of this research lies in its analysis of data interoperability and the practical challenges

associated with integrating payment and lending data from disparate sources.

Addressing these inquiries is challenging due to the rarity of scenarios in which lending and payment data originate from separate sources—a setting that speaks to data interoperability. This complexity is further compounded by concerns over the use of alternative data in lending, which is predominantly utilized by fintechs and not traditional banks. The critique here is that the distinctive borrower bases and lending contracts of fintechs versus traditional banks may limit the generalizability of findings related to the utility of alternative data beyond the niche of alternative lending.

A Unique Setting Addressing the Challenges

My research, detailed in Rishabh (2024), navigates these complexities using a unique setting that helps us estimate the value of interoperable payment data in traditional as well as in fintech lending. My study leverages a novel dataset that combines *traditional bank lending* contracts with payment history data from a fintech company, focusing on small business loans in India. This setup provides a rare opportunity to examine the interoperability of data in a real-world context. The analysis extends beyond the traditional lending to the valuation of payment data in sales-linked lending programs, which have become increasingly common among fintech and bigtech companies.

Analytical Framework: Evaluating Model Performance Differences

The analytical framework is structured around several models designed to dissect the multifaceted relationship between payment histories and traditional lending metrics. These models serve as the foundation for evaluating the hypothesis that interoperable payment data can enrich lenders' information set, thereby enhancing the accuracy of delinquency risk assessment.

- **Credit Bureau Model:** This model relies on traditional credit scores and related data, serving as a benchmark for assessing the incremental value provided by payment histories.
- **Traditional Model:** Developed in two variants, the first incorporates hard information such as borrower demographics and financials, while the second integrates both hard and soft information, including loan terms, to capture a more holistic view of the borrower's profile.
- **Payment History Models:** Two distinct models are introduced to analyze payment data. The Aggregate (PHA) model focuses on broad payment indicators, while the Granular (PHG) model delves into detailed transaction-level data. These models are pivotal in

examining the predictive power of payment histories in isolation and in conjunction with traditional data sources.

In evaluating the predictive power of various data models, including those based on credit bureau information and payment histories, I employ machine learning algorithm Random Forest. The goal is to compare the effectiveness of different models in predicting loan delinquency. The primary metrics for evaluation include the Area Under the Receiver Operating Characteristics (ROC) Curve (AUC) and the Average Precision (AP) score. Analytically, we can answer our main questions as by comparing model performances as follows:

- **Credit bureau and payment history: Complement or Substitute?:** Investigating if integrating PH data with credit bureau information in a combined model would lead to better predictions than using each source independently. Notable improvement after combining data sources would suggest PH could complement traditional credit info, while equivalent or superior standalone PH model performance would suggest PH might substitute traditional credit bureau data in certain scenarios.
- **Does payment data harden the soft information?:** This analysis explores whether adding PH data into the Traditional Model, which includes hard and soft information, would result in observable predictive performance improvements. Increased accuracy, especially with soft information, would suggest that PH brings independent signals, small or no change in model performance would suggest hardening of soft information without any new signals being brought by the PH.
- **Does payment history help a lender in loan monitoring?:** Incorporating post-disbursal PH data into models to assess changes in predictive accuracy aims to determine if PH contributes to more effective real-time risk assessment. Improved performance with post-disbursal PH data would suggest its usefulness in ongoing loan monitoring.
- **Is there a granularity-accuracy trade-off?:** Comparing the predictive capabilities of models using detailed transaction-level PH data against aggregated PH data to see if increased detail improves predictions. A marked improvement with detailed PH data would suggest a trade-off between granularity and predictive accuracy, acknowledging that granular data incur privacy and technological costs.

Results

1. **PH Data Complements Credit Bureau Information in traditional lending:** The addition of PH data to Credit Bureau data improves prediction accuracy, indicating a complementary effect.

2. **Aggregative PH creates as much value as lender soft information:** Integration of PH data results in a 5 percentage point increase in AUC and a 4 percentage point increase in AP, which is quantitatively similar to our estimate of lender soft information. However, only 29% of this improvement by PH is due to hardening of soft information, and majority due to independent signals.
3. **Post-Disbursal PH Data Enhances Monitoring:** Incorporating PH data improves loan delinquency prediction significantly within 120 days, demonstrating its effectiveness in real-time monitoring.
4. **Granular PH Data Increases Accuracy highlighting a potential trade-off:** Comparing granular (PHG) versus aggregated (PHA) data shows a 4% AUC improvement and a 12% increase in AP for granular data, highlighting the trade-off between data granularity and privacy or technological costs.
5. **Fintech Loans and PH Data:** For sales-linked fintech loans, PH data is found to substitute credit bureau, suggesting that in fintech lending, credit data might be less relevant. However, sales-linked lending is found to be particularly prone to moral hazard issues.

Open Banking: Enhancing Traditional Credit Models and Fintech Innovations

In conclusion, my study highlights Open Banking's dual utility: it enhances credit evaluations in traditional settings and offers an alternative basis for credit decisions in fintech lending. It points towards the opportunity inherent in merging transactional and credit histories for a fuller financial picture in traditional setting, while also cautioning against the privacy and technological challenges inherent in handling granular data.

References

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

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

Rishabh, Kumar. 2024. "Beyond the Bureau: Loan Screening and Monitoring under Open Banking." *Working Paper*.