NBER WORKING PAPER SERIES

FINTECH LENDING TO BORROWERS WITH NO CREDIT HISTORY

Laura Chioda
Paul Gertler
Sean Higgins
Paolina C. Medina

Working Paper 33208
http://www.nber.org/papers/w33208

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024, Revised April 2026

We gratefully acknowledge financial support from USAID and Digital Frontiers (Equitable AI Challenge), CEGA and the Bill & Melinda Gates Foundation (Digital Credit Observatory, DCO), and UC Berkeley's Lab for Inclusive Fintech (LIFT). We thank conference participants at CEGA, LIFT, NBER, Princeton/IFC, and We-Fi/EBRD/IDB for valuable feedback. We thank José Antonio Murillo, Luz Téllez, Roberto Arriaga, Pedro Armengol, and Jacobo Zetune from RappiCard Mexico (https://rappicard.mx/) for generously supporting this research project, providing access to the data, and answering questions. We thank Josh Blumenstock, Nitin Kohli, and Enrique Seira for helpful advice. We gratefully acknowledge the contributions of our research assistants. We especially thank Lucía Bertoletti, Ian Carbajal, Iñaki Fernández, Luis Roman, Haoyuan Song, and Tianyu Zhang for their substantial support on data and modeling pipeline development, dataset construction, and empirical analysis. We also thank Yusuf Abdul, Victoria Hollingshead, and Jora Li for research support during early stages of the project. Sophie van Genderen, Scott Coughlin, Sai Kumar Ramadugu, William Karl Thomson, and Matthew Gorby offered excellent computing and data service support. The authors declare that they have no financial or material interests in the findings of this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

FinTech Lending to Borrowers with No Credit History
Laura Chioda, Paul Gertler, Sean Higgins, and Paolina C. Medina
NBER Working Paper No. 33208
November 2024, Revised April 2026
JEL No. G23, G5, O16

ABSTRACT

Despite the promise of FinTech lending to expand credit access to populations without a formal credit history, FinTech lenders primarily lend to applicants with a formal credit history and rely on conventional credit bureau scores as an input to their algorithms. Using data from a large FinTech lender in Mexico, we show that alternative data from digital transactions through a delivery app are effective at predicting creditworthiness for borrowers with no credit history. Using account-by-month level data on revenues and costs, a machine learning model predicting profits generates similar profits as a model predicting default.

Laura Chioda
University of California, Berkeley
Haas School of Business
lchioda@berkeley.edu

Sean Higgins
Northwestern University
Kellogg School of Management
sean.higgins@kellogg.northwestern.edu

Paul Gertler
University of California, Berkeley
Haas School of Business
and NBER
gertler@haas.berkeley.edu

Paolina C. Medina
University of Houston
pcmedina@uh.edu

1 Introduction

Online FinTech lenders are an increasingly important source of credit for households and small businesses (Berg, Fuster, and Puri, 2022; Buchak, Matvos, Piskorski, and Seru, 2018; Gopal and Schnabl, 2022). The promise of FinTech lending is that by using alternative data sources to evaluate creditworthiness and reducing other frictions such as travel costs and loan processing time, FinTech lenders can expand access to credit to populations with limited or no credit history—i.e., the financially excluded. In practice, however, while FinTech lenders do improve their default prediction models leveraging alternative data sources, most of their lending algorithms still rely at least partly on conventional credit bureau scores (Johnson, Ben-David, Lee, and Yao, 2023) and do not substantially expand credit access to those traditionally excluded from the financial system (Fuster, Plosser, Schnabl, and Vickery, 2019).

Using data from a large FinTech lender in Mexico, we show that alternative data—digital transactions data—can be quite effective in predicting creditworthiness *even for borrowers with no credit history*. All applicants in our sample lack a traditional credit score from the credit bureau, because they have either no credit history or at best a limited credit history that the credit bureau deems as insufficient to use to generate a credit score.[1]

Our FinTech partner, RappiCard Mexico, started as a joint venture between Banorte, a large bank in Mexico, and Rappi, the leading on-demand delivery platform for food, goods, and services in Latin America.[2] RappiCard Mexico leverages digital footprints and transactions data to inform credit card lending decisions. The company lends to applicants both with and without credit history. When lending to individuals with a credit history and thus a credit score in the Mexican credit bureau, they combine credit bureau data with transaction-level data on delivery orders through the app and use a machine learning algorithm to assess risk.

At the onset of our collaboration, when lending to applicants with no credit history, RappiCard had not relied on a machine learning algorithm; instead, they used a set of parsimonious rules for various client segments to make their lending decisions. We use data on the subsequent repayment behavior of these borrowers to assess risk. Specifically, we combine the repayment information with transaction-level data on purchases made through the delivery app prior to credit application, data on these applicants' characteristics and "digital footprints" (Berg, Burg, Gombović, and Puri, 2020), and other data sources, to build machine learning models to predict creditworthiness.[3]

[1]For conciseness we refer to these borrowers with no credit bureau score as having "no credit history." None of the applicants in our sample had a credit card prior to applying for a credit card from our FinTech partner.

[2]In 2025, Banorte acquired the right to be the sole provider of financial services on the Rappi platform in Mexico.

[3]The other data sources include a "no-hit" score—developed by the credit bureau for those with no credit history or an insufficient credit history to report a traditional credit score—and socioeconomic characteristics at the census tract level. The "no-hit" score is reported by the credit bureau for all Mexican citizens with no credit history and thus no traditional credit score; it is independent of (i.e., not comparable to) the traditional credit scores reported for those

We find that machine learning models using alternative data predict the creditworthiness of borrowers without credit history with sufficiently high accuracy to generate substantial profits for the lender. In terms of predictive accuracy, our benchmark model achieves an area under the receiver operating characteristic curve (AUC) of 0.791. This exceeds the thresholds for desirable AUCs of 0.6 in data-scarce environments and 0.7 in data-rich environments (Iyer, Khwaja, Luttmer, and Shue, 2016), is at the upper end of AUCs estimated using alternative data (even when combined with credit bureau data) in middle-income countries, and surpasses the AUCs for populations with no credit history (see Table 1 for a comparison of samples and AUCs across studies). Furthermore, RappiCard has scaled this approach into production: after testing a model similar to our default model beginning in September 2024, they now apply it to roughly half of applicants with no credit history, where it has outperformed their previous rule-based lending decisions.[4]

To quantify the importance of each data source, we estimate the profits generated by models excluding the features from one data source and compare these to the profits generated by the "full" model using all features. We find that the digital transactions data from the delivery app have by far the highest marginal contribution to the model's profitability and predictive power: a model without the transactions data is only 31.4% as profitable and obtains an AUC that is 0.137 lower. Consistent with this pattern, the model excluding digital transactions data exhibits the largest reduction in approval rates (when using a profit-maximizing approval threshold): the full model approves 69.4% of applicants, while the model without access to transactions data approves only 49.5%. The performance of the model improves with the richness of the transactions history through the delivery app. Specifically, when we split the sample into quintiles based on the number of transactions they have completed through the app at the time of credit application, the model for the top quintile—those with 29 or more transactions through the delivery app—has an AUC of 0.828 while that for the lowest quintile—those with at most 2 transactions—has an AUC of 0.742 and generates only 45.4% as much profits as the model for the top quintile. Notably, proprietary digital transactions data are substantially more predictive than static digital footprint data—such as the operating system, device type, and email host of the applicant (Berg, Burg, Gombović, and Puri, 2020)—that can be obtained by any FinTech lender.[5]

Next, we ask how well a model predicting default performs relative to a model predicting

who do have a credit history, and is based on publicly available data sets aggregated at the local level merged with the location where the applicant lives; see CRIF (2018).

[4]In 2025, the approval rate of the two models was very similar but the machine learning model with alternative data resulted in half as much default as the parsimonious rules.

[5]Specifically, digital footprint data contribute 0.014 to the AUC and a model without these data is still 91.3% as profitable as a model with all data sources. As for other data sources, neither the "no-hit" scores generated by the credit bureau using publicly available geographic data for borrowers with no or limited credit history, combined with credit history data for those with a limited but insufficient credit history to generate a traditional credit bureau score, nor socioeconomic characteristics at the census tract level, increase the total profits generated by the model relative to a model without these data sources.

account-level profits, given the lender's ultimate objective of maximizing profits. Models that predict default, rather than profits, are used extensively by both traditional and FinTech lenders to make credit origination decisions (Rajan, Seru, and Vig, 2015; Johnson, Ben-David, Lee, and Yao, 2023). Given that firms—including financial firms—often forgo profitable opportunities (Gertler, Higgins, Malmendier, and Ojeda, 2025; Mishra, Prabhala, and Rajan, 2022), are lenders leaving substantial profits on the table by using default models to make lending decisions? To address this question, we use account-by-month level data on each component of the lender's realized revenues and costs for each originated credit card over time. Revenues to the lender include interest payments, buy now pay later (BNPL) fees, interchange revenue paid to the card issuer each time a purchase is made with the card, late payment fees, and other fees. Costs include charge-offs, cashback rewards, funding costs (calculated by RappiCard as the average daily balance over the month times its monthly cost of capital to borrow that money), costs of fraudulent transactions, and other costs such as the cost of replacing a lost card.

Using rich data on account-by-month level revenues and costs for the lender, we calculate account-level profits over the first 12 months since card origination and compare a model predicting profits to our benchmark model predicting default. In both cases, we derive the profit-maximizing approval threshold in the training sample and apply this threshold to determine credit allocation based on the predicted values of the target variable in the testing sample. The two models generate quite similar levels of profits for the lender: in particular, we cannot reject that the total profits generated by the two models are equal.

We next explore *why* the model predicting default generates as much profit as the model predicting profits, despite the latter being trained to predict what the lender ultimately cares about. Important patterns emerge when we plot average profits against predicted default probabilities and predicted profits. Realized profits follow an inverted-U pattern as the predicted probability of default rises: for the lowest-risk borrowers, average profits are low but on average positive, and average profits initially increase with risk of default, but then fall and become negative and continue decreasing as the predicted probability of default increases. The non-monotonic relationship between profits and the predicted probability of default is consistent with Agarwal, Chomsisengphet, Mahoney, and Stroebel (2015). However, a key difference is that in our setting, profits for the highest-risk borrowers cross zero and become negative, whereas the highest-risk borrowers approved for credit cards remain profitable in Agarwal, Chomsisengphet, Mahoney, and Stroebel (2015). In contrast to the non-monotonic relationship between realized profits and predicted default, average realized profits rise fairly steadily across bins of predicted profits from the profits model, suggesting that predicted profits track variation in average realized profits more systematically than predicted default does, which could favor the profits model in making profit-maximizing lending decisions. To explore why the default model nevertheless generates similar profits, we

turn to comparing borrowers approved by each model to understand why the two models generate similar profits.

While 58.2% of applicants are approved by both models and 27.7% are rejected by both models, 11.3% of applicants are approved only by the default model ("default-approved") and 2.9% are approved only by the profits model ("profits-approved"). The two models agree on the majority of applicants: 58.2% are jointly approved and 27.7% jointly rejected. However, they diverge for a meaningful share of the sample: 11.3% are approved only by the default model ("default-approved") and 2.9% only by the profits model ("profits-approved").

Intuitively, the profits model attempts to identify borrowers who generate substantial interest revenue (or other revenue such as through interchange fees net of rewards), but do not default. It succeeds at selecting users who generate substantial revenue: for example, 82% of profits-approved borrowers generate interest revenue compared to 62% of default-approved borrowers and average interest revenue of profits-approved borrowers is 1.6 times as large as that of default-approved borrowers. However, many of the borrowers who generate substantial interest revenue by carrying revolving debt (and making their interest payments) ultimately default, and the profits model fails to predict well which borrowers generating interest revenue do default: 28% of profits-approved borrowers are at least 60 days delinquent and 21% incur charge-offs over the first 12 months since origination, compared to only 17% delinquent and 12% with charge-offs among default-approved borrowers. Average charge-offs for profits-approved borrowers (including zeros for those without charge-offs) are 1.8 times as large as average charge-offs for default-approved borrowers.

Interchange fee revenue also plays a role. Those approved by both models are relatively high spenders—spending 3,291 Mexican pesos (MXN) per month—and their interchange fee revenue is approximately 6% higher than rewards expenses.[6] Borrowers approved exclusively by the default model spend slightly less than borrowers approved by both models—2,995 MXN per month—and their cost of rewards fully offsets interchange fees. In contrast, the applicants approved only by the profits model are riskier borrowers and spend less —spending 2,769 MXN per month—yet generate a more favorable interchange-to-rewards spread: their interchange fee revenue exceeds rewards by 17%. Thus, the profits model also selects borrowers who generate positive interchange revenue net of rewards, whereas this is not true for borrowers selected only by the default model. The lower-risk borrowers identified by the default model are likely more sophisticated and better at optimizing their shopping behavior to maximize rewards.

Finally, we assess which types of alternative data are best at predicting *each component* of borrower-level profits, and whether the importance of each type of data varies across sociodemographic groups. Digital transaction data from the delivery platform remain by far the most important data source when predicting each component of profits and across all subgroups we con-

[6]The purchasing power parity (PPP) adjusted exchange rate was 9.7 MXN per US dollar (USD) at the end of 2022.

sider. Interestingly, transactions data are relatively more important for approval decisions about the *less profitable* groups—which include men, older applicants, Android users, and those who use more cash—while the same does not always hold for digital footprint data.

We make three main contributions. First, we show that alternative data are effective in predicting creditworthiness for borrowers with *no credit history*. Most papers on the use of alternative data for FinTech lending estimate these models on a sample in which all or at least a majority of applicants *do* have a formal credit history and credit score (Table 1), perhaps because FinTech lenders primarily lend to applicants with formal credit histories and thus do not expand credit access to borrowers who have been completely excluded from formal credit markets (Berg, Fuster, and Puri, 2022; Fuster, Plosser, Schnabl, and Vickery, 2019). Some prior work has focused on subprime borrowers (Di Maggio and Yao, 2021) or borrowers with a thin credit file that is nevertheless sufficient for the credit bureau to generate a credit score (Blattner and Nelson, 2024), and has shown that FinTechs have the potential to expand credit access for these borrowers. However, better models for these groups do not address the exclusion of borrowers who have always been completely excluded from credit markets. Other studies evaluate machine learning models for borrowers with and without credit scores, but find substantially weaker model performance for those without credit history: for example, in Agarwal, Alok, Ghosh, and Gupta (2023), the AUC for those with no formal credit history in India is 0.674, compared to an AUC of 0.738 for those with formal credit histories (when the credit bureau data are included in the model).[7] In contrast, in our sample no applicants have sufficient credit histories for the credit bureau to generate a credit score for them, and we find an AUC of 0.791.

Second, we systematically compare machine learning models trained to predict *default* versus *profits* for applicants with no credit history. Given that many lenders predict default rather than profits to make credit allocation decisions (Rajan, Seru, and Vig, 2015; Johnson, Ben-David, Lee, and Yao, 2023), understanding whether this choice maximizes profits is an important empirical question. Several papers have diagnosed the profitability of different segments of credit card borrowers (e.g., Agarwal, Chomsisengphet, Mahoney, and Stroebel, 2015; Agarwal, Presbitero, Silva, and Wix, 2023; Drechsler et al., 2025; Krivorotov, 2023). Other contributions detailed in Table 1 assess the use of alternative data to predict *default*, but lack borrower-level profits data to assess whether lenders would do better by predicting profits. By combining alternative data for credit scoring with data not only on defaults but also on account-by-month level revenues and

[7]Two exceptions are Björkegren and Grissen (2020) and Lee, Yang, and Anderson (2026). In Björkegren and Grissen (2020), the best-performing model using mobile phone data for the 15% of their sample with no credit history has an AUC of 0.766, compared to an AUC of 0.770 for the best-performing model combining mobile phone data and credit bureau data for the 85% of their sample with credit histories. In Lee, Yang, and Anderson (2026), the model for the 18% of the modeling sample with no credit history has an AUC of 0.682, compared to an AUC of 0.677 for the 82% of the sample with credit histories.

costs for the lender, we compare the total profits generated by models predicting default and profits. While the two approaches yield similar aggregate profits, the profile of borrowers approved under each model differs meaningfully: the profits model approves riskier borrowers, suggesting that broad adoption of algorithms predicting profits could reduce financial stability if default shocks are correlated.[8] The choice of target variable by lenders may also affect financial inclusion: the profits model approves fewer borrowers, suggesting that a shift to such models would reduce credit access.

Third, we examine which data sources drive predictive performance in credit origination models, comparing their importance across borrower subgroups for predicting both default and profits, as well as each component of profits. While any FinTech lender could collect data on digital footprints from applicants' cell phones and credit applications, and while these data are predictive of creditworthiness (Berg, Burg, Gombović, and Puri, 2020), digital transactions would not be available to most FinTech lenders nor in most academic studies. Thus, our finding that proprietary digital transactions data greatly improve models predicting creditworthiness of applicants with no formal credit history has important policy implications. In particular, the finding suggests that recent open banking initiatives around the world could further expand access to credit for those traditionally excluded from the formal financial system, by enabling these applicants to share their transactions data (from bank accounts, delivery apps, e-commerce platforms, etc.) with lenders. Furthermore, while all countries with open banking policies require banks to share transactions data with FinTechs if the customer desires, only 18% require FinTechs to share their proprietary data with banks (Babina et al., 2025). Our results suggest a role for more expansive open banking policies that include proprietary data from online platforms. By expanding the set of lenders that will be able to assess creditworthiness using transactions data, these more expansive open banking policies could increase competition in consumer credit markets, further benefiting consumers (He, Huang, and Zhou, 2023; Nam, 2022).

2 Institutional Context

2.1 Financial Inclusion and Credit Cards

Only 49% of Mexicans have bank accounts, 44% have made or received digital payments, and 12% have credit cards, all significantly below the equivalent rates for countries with similar levels of development. Moreover, women are 14 percentage points (p.p.) less likely than men to have a bank account in Mexico, 11 p.p. less likely to have made or received digital payments, and 8 p.p. less likely to have a credit card. These gender gaps are significantly higher than those of other countries in Latin America and of other OECD countries (Demirgüç-Kunt, Klapper, Singer, and

[8] Krivorotov (2023) makes a similar point; in contrast to our paper, Krivorotov (2023) does not observe interchange fees, rewards, or funding costs, does not use alternative data, and focuses on borrowers with credit history.

Ansar, 2022).

Credit cards are one of the most common ways for new borrowers to access formal credit in Mexico, where credit cards are the first formal type of credit for 74% of all formal sector borrowers (Castellanos et al., 2025). Credit cards in Mexico share several features with credit cards in other countries. Much like in the US, they are both a form of payment and a source of financing. Cardholders use the card to purchase products at merchants that have a point-of-sale (POS) terminal. At the end of a 30-day billing cycle, cardholders receive a statement showing (among other things) the balance at the end of the cycle (statement balance), the required minimum payment, and a due date, which is typically 20 days after the end of the cycle. To stay current, cardholders need to pay at least the minimum payment by the due date. If they pay less than the full statement balance, they incur interest according to the interest rate assigned to the revolving line of credit. If they pay the statement balance in full, they do not incur any interest—effectively getting up to 50 days of free financing. As they make payments, their available credit line (i.e., the difference between their credit limit and outstanding balance) frees up.

Similar to the US, rewards tied to credit card transactions are a common instrument to promote card adoption and compete for consumers (Agarwal, Presbitero, Silva, and Wix, 2023; Wang, 2025). Our FinTech partner, for example, offers cashback of 1% of spending for most transactions and up to 5% of spending for transactions in certain establishments, with an average implicit cashback rate—calculated as cashback costs divided by spending—of 1.15% in the segment of borrowers without credit history.[9] In turn, card issuers receive an interchange fee charged to the merchant for every transaction, which vary by type of merchant and range from from 0% for non-profits and educational institutions to 1.76% for restaurants.[10] Our FinTech partner has an effective interchange rate—calculated as interchange fee revenue divided by spending—of 1.24% in the segment of borrowers without credit history.

In addition, unlike in the US, UK, and Nordic countries (Di Maggio, Williams, and Katz, 2022; Guttman-Kenney, Firth, and Gathergood, 2023; Laudenbach, Molin, Roszbach, and Sondershaus, 2025), but like in other countries including Turkey, South Korea, and Brazil (Aydin, 2022; Cho and Rust, 2017; Lima Junior, Silva, Altoé Junior, and Ruhe, 2021), credit cardholders in Mexico have access to "buy now, pay later" (BNPL) installment loans through their credit card. Cardholders can finance purchases of specific items from specific merchants with short-term loans that allow payment in installments over a set period, typically ranging from 3 to 12 months and, in some cases,

[9]BNPL transactions do not accrue cashback and as a result, implicit cashback rates are less than 1% for some consumers.

[10]Mexico's Central Bank publishes on its website the levels of interchange fees by merchant type that financial institutions have agreed upon, which have not changed since 2013; see `https://www.banxico.org.mx/servicios/cuotas-de-intercambio-por-el-uso-de-tarjetas-de-cr/cuotas-intercambio-tarjetas-c.html`. For merchants that use POS terminals from FinTech payments companies (as in Gertler, Higgins, Malmendier, and Ojeda, 2025), rather than POS terminals from banks, the interchange fee is 1.76% regardless of the type of establishment.

up to 18 months. These installment loans can be interest-free or carry interest rates lower than those applied to revolving balances. If the monthly installment is not fully covered by the payment against the corresponding bill, the unpaid amount is financed through the revolving line of credit, potentially leading to additional interest charges but not resulting in delinquency status. BNPL via credit cards accounts for 47% of credit card balances in Mexico (Banco de México, 2024) and 32% for our FinTech partner. The establishment selling the product bears the financing cost—either fully (for interest-free BNPL loans) or partially (for interest-bearing BNPL loans)—by paying an upfront fee to the credit card issuer, who in turn finances the transaction and bears the risk.

2.2 FinTech Lending

Fostering a dynamic FinTech environment has been part of regulators' strategy to promote financial inclusion in Mexico. In 2018, the Mexican Congress passed a FinTech law and, as of the end of 2023, Mexico is one of the largest FinTech markets in Latin America with 650 FinTech start-ups (CNBV, 2019; Department of Commerce, 2023). The most active segment of FinTech activity is lending, with 146 companies active in this space, followed by payments and remittances, personal financial management, and crowdfunding (Finnovista, 2023).

One of the main products through which FinTech companies lend is credit cards (CNBV, 2023). Traditionally, the credit card market in Mexico has been dominated by a few large banks: as of 2021, the top two largest banks controlled 57% of the cards issued by traditional financial institutions and the top five largest banks controlled 87% of them. However, in 2022 one of the main drivers of the growth in consumer credit was credit cards issued by FinTech lenders (CNBV, 2023), with the largest FinTech lender becoming the fifth-largest credit card issuer (by number of accounts) in the country.[11]

In contrast to the U.S. where regulatory agencies have been actively evaluating conditions in which alternative data can be used in credit origination decisions (Consumer Financial Protection Bureau, 2017, 2019), and where some lenders have explicitly requested no-action letters to ensure their models are compliant with appropriate regulations before using them (Di Maggio and Ratnadiwakara, 2025), in Mexico there is no regulation explicitly limiting the variables that FinTech lenders can use for credit origination. Both FinTech and traditional lenders increasingly use alternative data in their credit origination decisions (El Economista, 2024).

2.3 Delivery Platforms

RappiCard Mexico has access to transactions data from Rappi, the leading on-demand delivery platform of Latin America. An on-demand delivery platform connects customers with couriers

[11] See `https://www.bloomberglinea.com/2023/02/27/neobanco-nu-es-el-quinto-emisor-de-tarjetas-de-credito-en-mexico-moodys/`.

via mobile apps or websites for immediate or scheduled deliveries of goods or services to desired locations within set time frames. Rappi provides a variety of services through its mobile app, including the purchase of groceries, household items, restaurant food, alcoholic beverages, and pharmaceutical products, as well as booking of flights and hotels. It also allows users to request cash withdrawals and the execution of miscellaneous errands. Orders are completed by local couriers, typically within 30 minutes to one hour. Delivery apps are a growing business in Mexico. In the first quarter of 2023, 24% of mobile phone users had at least one delivery app installed on their phone, representing a 142% increase since 2019 (Trecone, 2023). The market is concentrated among three players that, as of the latest counts, operate in approximately 100, 80, and 57 cities in Mexico, respectively.[12]

3 Data

The data for our analysis were provided by RappiCard Mexico. To apply for a credit card, individuals must have an account with Rappi and complete the application through its mobile app or a web browser. There is no requirement for a minimum number of transactions nor a waiting period after account creation.[13] Applicants need only provide their full name, address, date of birth, and tax identification number, and consent to a credit check.

3.1 Sample

Our data set consists of information from 181,488 credit cards originated between January 2021 and May 2024 for borrowers with no credit history. This sample includes all borrowers who were flagged by the Mexican credit bureau as having null or insufficient credit history to have a traditional credit score, had no prior credit card, and applied for and received a RappiCard credit card over this time period.[14] For borrowers with no credit history during the sample period, RappiCard did not rely on a machine learning algorithm; instead, approval decisions were based on a set of parsimonious rules. These rules included a threshold on the no-hit score (derived solely from geographic location, with no individual-level data) and occasionally a threshold on a second variable. [15] The 181,488 credit cards originated between January 2021 and May 2024 are the applications that were approved using

[12]See https://www.forbes.com.mx/rappi-ya-rueda-en-100-ciudades-de-mexico-dolores-hidalgo-la-ultima-en-sumarse/, https://web.didiglobal.com/mx/conductor/ciudades, and https://www.uber.com/es-MX/newsroom/uber-eats-expansion-en-mexico/.

[13]Burlando, Kuhn, and Prina (2025) study the effects of a digital lender in Mexico imposing a waiting period.

[14]The filter of having no prior credit card removes those who had a credit card prior to applying for the RappiCard but who nevertheless had an insufficient credit history for the credit bureau to generate a credit score for them.

[15]For example, in the past RappiCard purchased scores based on cell phone records which are sold to lenders by independent local providers. RappiCard implemented a threshold using this variable in addition to the no-hit score threshold to determine which applicants received credit cards.

these parsimonious rules among 1,328,426 applications from borrowers with no credit history and no prior credit card.

For these borrowers, we also have data on balances and repayment from origination through May 2024. To not understate default rates associated with recent or completely inactive cardholders, we impose two additional restrictions on the analysis sample: cardholders (i) must have had the card for at least twelve months by the end of our repayment data in May 2024 and (ii) must have completed at least one transaction using their credit card within twelve months of origination. This leaves us with an analysis sample of 146,036 borrowers. Figure A.1 shows the number of applications per month and the number of originated contracts per month from the beginning of our sample period through May 2023 (given the restriction that the borrower must have the card for at least twelve months by the end of our repayment data in May 2024, which effectively filters out all applications after May 2023).

During the timing of our sample, from January 2021 to May 2024, Mexico had a relatively stable macroeconomic environment (Figure A.2). Unemployment, measured monthly, shows a slight downward trend over this time period, starting at 4.5% and reaching 2.7% in June 2024, with an average of 3.3% over the period; this can be compared with average unemployment of 4.1% from 2006 (the earliest we have unemployment data) to 2024. Annual inflation shows significant variation starting at 3.5% in January 2021, reaching a maximum of 8.7% in August 2022, and declining to 5.0% in June 2024; average inflation in Mexico was 8.1% from 1995–2024 and 4.5% from 2006–2024. Quarterly GDP growth averages 0.7%, ranging from -1.0% to 1.5% without a distinctive trend, which can be compared to average quarterly GDP growth of 0.6% from 1995–2024 and 0.4% from 2006–2024.

3.2 Data Sources for Features

For each approved applicant, we observe data from the following four information sources. For a more detailed list of variables and a description of how we transform these variables into features used in the machine learning algorithms, see Appendix B.

Transaction-level data from the delivery platform These data include the date and time of the order placed, a list of each item purchased, the quantity of each item purchased and its unit price, fees, discounts, tips, and total order cost. The data also include payment method (credit card, debit card, or cash), store name, and geographic identifiers for the store. This is more granular than traditional transaction-level data from credit or debit cards (as used in, e.g., Higgins, 2024), as it allows us to observe not only the shop where the order was placed, but the specific items purchased from that shop. This data source also includes other variables that are a function of the applicant's use of the Rappi delivery app, such as the user's loyalty level, whether they are a

Prime customer (a paid membership in exchange for various benefits), and additional categorization variables Rappi generates based on transactions and Rappi app engagement (e.g., lifetime activity, recent transactions, and risk of churn). These data are proprietary and only available to Rappi since they are generated when users place orders on their delivery app; this contrasts with digital footprints and applicant characteristics, which can be obtained by any FinTech lender when the applicant applies.

Digital footprints and applicant characteristics We explicitly requested the digital footprint variables from Berg, Burg, Gombović, and Puri (2020), and RappiCard provided the exact variables or close proxies for all of them, except for the timestamp of the credit card application. We have access to information about the operating system, desktop vs. mobile phone/tablet, acquisition channel, email provider, email privacy setting, and an array of email-string characteristics: name-in-email, number-in-email, lowercase conventions, and typo in the email address. A detailed comparison of our digital footprint measures and those of Berg, Burg, Gombović, and Puri (2020) is provided in Appendix B. This feature set also draws on standard applicant characteristics available on any loan application, such as gender, age, city, and state.

"No-hit" scores and credit history for those with limited credit history All of the applicants in our sample are referred to by the credit bureau as the "no-hit segment." This means that they have no formal credit history or too limited of a credit history for the credit bureau to use those data to provide a credit score. For them, the credit bureau issues a flag indicating that the traditional score (built from credit histories) is not applicable. Beginning in 2018, the credit bureau contracted a third party to develop a "no-hit" score for all Mexicans who do not have a traditional credit score. The no-hit score is based on geographic indicators merged with the location where the individual lives. The geographic indicators come from a variety of public records, including demographics, economic activity, public safety, social cohesion, and access to and use of credit at the local level (see CRIF, 2018). Traditional credit scores and no-hit scores are independent from each other, with traditional scores ranging 456 to 760, and no-hit scores ranging from 463 to 735. The no-hit segment is thus distinct from the subprime segment of the traditional market (studied in the US in Di Maggio and Ratnadiwakara, 2025)—identified by low values on the traditional credit score—and by those with thin credit files that are nevertheless sufficient for the credit bureau to generate a credit score (studied in the US in Blattner and Nelson, 2024).

For borrowers in our sample who do have a credit history—all of whom have an insufficient credit history for the credit bureau to assign a traditional credit score—we observe length of credit history and balances (if any). Our sample excludes anyone who had a credit card prior to applying for a card from our FinTech partner. While the credit bureau's rules on what constitutes a sufficient

credit history to generate a credit score are proprietary, these rules are unlikely to differ between Mexico and other countries such as the US since the credit bureau in Mexico is TransUnion. We include these credit variables for those with limited credit history in the same category as no-hit scores since both are provided by the credit bureau.

Socioeconomic characteristics at the census tract level We combine publicly available information from Mexico's National Institute of Statistics (INEGI) with location information collected by the delivery platform whenever a user logs in.

Furthermore, we observe transaction-level data and "no-hit scores" for applicants who were rejected based on RappiCard's parsimonious rules, but we did not receive the remaining data sources for rejected applicants, despite RappiCard observing them for all applicants.

3.3 Account-Level Data on Contract Terms, Use, and Profitability

We also observe the following variables related to credit card terms and use for accepted applicants:

Credit card terms assigned at loan origination, including the interest rate and credit limit.

Account-by-month level data on statement balances, minimum payments, and repayments. The minimum payment is the minimum amount the borrower must pay to avoid their balance being considered past due and, by regulation, must cover the interest charges of the period plus a fraction of the outstanding balance (Medina and Negrin, 2022).

Account-by-day level data on arrears. When a borrower does not pay the minimum balance on time, their statement balance is considered past due. We observe account-by-day records of arrears (tracking the number of days past due).

Transaction-level data on credit card spending, including the date, location, merchant category code of the merchant, interchange rate for the transaction, and the amount spent on the transaction.

For all cards issued during our observation period, we also observe account-by-month information on realized revenues and costs for the FinTech lender. These are the variables used in practice by our FinTech partner to evaluate the profitability of each account.

Interest revenue corresponds to the *realized* revenues from interest payments made by the borrower as part of their monthly payments.

BNPL revenue includes fees paid by merchants who offer installment payments (often interest-free) to their customers to increase demand (Berg, Burg, Keil, and Puri, 2025), as well as interest charges paid by the cardholder when the installment plan is not interest-free.

Interchange fee revenue from the interchange fees received by the card issuer for credit card transactions made by the borrower. As in the US, interchange fees vary by type of merchant (GAO, 2009); in our context, they vary from 0% to 1.76%.

Revenue from fees including revenue from late payment fees and revenue from other fees such as card replacement fees and cash advance fees for ATM withdrawals using the credit card. This lender does not charge an annual fee.

Charge-offs, which correspond to balances deemed not collectible by a lender and removed from its balance sheet. This measure captures two distinct regulatory concepts. The first, known as *quitas*, refers to partial reductions (or haircuts) on balances renegotiated with delinquent borrowers. These accounts may be reactivated depending on subsequent payment behavior, but the reduction amount is considered unrecoverable and written off. Renegotiations start at the discretion of the lender or borrower for accounts between 30 and 180 days delinquent. The second concept, *castigos*, refers to balances written-off on accounts that are permanently closed due to failure to repay. The time of write-off is at the discretion of the lender. The practice of our partner is to close accounts permanently when they reach 180 days in arrears.

Funding costs are defined, for each account-month, as average daily balances multiplied by the lender's cost of capital; our partner directly provided account-by-month level funding costs using their internal cost of capital. Funding costs are non-zero both for revolvers and for transactors—who pay off their balances in full each billing cycle and do not revolve debt—since revolvers carry balances month to month and transactors get up to 50 days of free credit due to the billing structure of credit cards (see Section 2.1).

3.4 Summary Statistics

Table 2 shows descriptive statistics for our modeling sample, i.e., the 146,036 applicants with no credit history and no prior credit card who were approved by RappiCard using their parsimonious approval rules for borrowers with no credit history, and who also had at least twelve months with the card by the end of our data period in May 2024 and made at least one transaction during their first twelve months with the card.[16]

[16]In addition, we lose less than 0.5% of the observations due to these borrowers not appearing in the profit components data set due to marginal inconsistencies in the samples included in different data tables.

Table 2, Panel A shows a subset of the features that are used in our machine learning models. The applicants in our sample are relatively young: the average user age in our sample is 25. Younger people are less likely to have a credit score (Cookson, Guttman-Kenney, and Mullins, 2025), more likely to use smartphones and delivery apps, and also more likely to consider a FinTech lender as a potential source of credit (Doerr, Frost, Gambacorta, and Qiu, 2022; Krivorotov, 2023), including being more willing to share transactions data with lenders (Armantier et al., 2024). Less than half of the sample (38%) uses an Apple product, which is an important predictor of creditworthiness (Berg, Burg, Gombović, and Puri, 2020). There is not a lot of variation in the no-hit score, which has a mean of 641, a standard deviation of 14, and an interquartile range of 635 to 649; this is not surprising given that the no-hit score is based only on publicly available geographic-level information merged with the location of the applicant.[17] There is also little variation in the census tract-level variables: for example, the marginality index has a mean of 0.96 and a standard deviation of 0.01.

There is substantially more variation in measures from the transaction-level data. The average number of orders on the app is 24 with a standard deviation of 52 and interquartile range of 3 to 22, the average number of orders paid in cash is 8 with an interquartile range of 1 to 9, and the median amount per order is 350 MXN with a standard deviation of 345 MXN. The majority of purchases are orders from food establishments (19 orders per user), compared to 2 orders per user from supermarkets and 1 from pharmacies.

Table 2, Panel B, shows descriptive statistics on credit card terms and use of the cards. The average annual interest rate at origination is 78%. This is not substantially higher than the average interest rate across all credit cards in Mexico, which is 64% based on data from Mexico's Central Bank (and it is not surprising that RappiCard's interest rate for borrowers with no credit history would be higher than the average interest rate across all credit cards in Mexico). The lender has used different interest rates over time, but for the segment with no credit history has broadly followed a policy of using a single interest rate at origination at any point in time, except during periods of transition from one interest rate to another.[18] During the early part of our sample period in 2021, nearly all cards were originated with a 72% annual interest rate, while they were originated with an 87% annual interest rate in the first half of 2022 and an 80% annual interest rate in the second half of 2022 and throughout 2023 (Figure A.3). It is not uncommon for other FinTech lenders to similarly engage in uniform pricing.[19]

[17] We refer to this as little variation since no-hit scores range from 463 to 735.

[18] In contrast, the lender does use differentiated pricing for borrowers who do have a formal credit history; our sample does not include borrowers with a formal credit history, however.

[19] In practice, while many FinTech lenders do some form of differentiated pricing (as our partner does for loans to borrowers with a formal credit history), the variation in interest rates is often explained mostly by the applicant's formal credit score (Johnson, Ben-David, Lee, and Yao, 2023). It is also not uncommon for FinTech lenders to charge the same interest rate to all borrowers (as our partner does for loans to borrowers without a formal credit history): for

The average credit limit at origination is 5,171 MXN. In terms of card use, average spending is 3,090 MXN per month, the average statement balance 3,745 MXN, the average minimum payment 455 MXN, and the average repayment 2,835 MXN.

Table 2, Panel C, shows account-by-month level averages of the various revenues and costs incurred by the lender on that card. The average revenues from interest payments and BNPL are 64 pesos and 7 MXN per card per month, respectively, while the average amount lost to charge-offs is 58 MXN per card per month. Combining these, average interest and BNPL revenue net of charge-offs is 13 MXN per card per month. Average interchange fee revenue is 38 MXN per card per month, while average rewards on each card are 35 MXN per month, leaving 3 MXN per card per month in interchange fee net of rewards. Other revenues for the lender include late payment fees (13 MXN per card per month on average) and other fees (3 MXN), while other costs include costs from fraudulent transactions (1 MXN per card per month) and other costs (12 MXN). Finally, funding costs are 30 MXN per card per month on average.

3.5 Target Variables

Default model We define the target variable for the machine learning model predicting default as overdue for more than 60 days at any point during the first twelve months since origination, which we refer to throughout the paper as "default." We compare the definitions of default used by various papers in Table A.1, column 2, which shows that there is no standard definition across studies. This likely also reflects a large amount of heterogeneity across the definitions of default used by lenders in their credit scoring models.

Two choices must be made for the measure of default: how many days overdue must a borrower be, and over what time period should default be measured? We choose 60 days overdue because this is the threshold used by RappiCard in their credit scoring models, as they consider this to be an early warning of future charge-offs. Furthermore, 60 days was the regulatory definition of default for credit cards in Mexico through January 2022 (Banco de México, 2021). Table A.1 shows that 60 and 90 days delinquent are both commonly-used thresholds for credit cards and other consumer credit products in other studies.

Regarding the length of time over which delinquency is measured, we note that unlike for installment loans where there is a clear time period over which to measure default (i.e., the maturity of the loan), for credit cards there is no clear time period over which to measure default. Furthermore, there is a trade-off in the length of time we use: for shorter periods of time, we may have substantial measurement error as we label some borrowers who will default in the future as

example, the FinTech lenders in Berg, Burg, Gombović, and Puri (2020), Choi et al. (2025), and Yang (2025) charge the same rate to all borrowers. Furthermore, in the UK, even banks keep credit card interest rates nearly constant across consumers of varying default risk (Matcham, 2025).

non-defaulters. On the other hand, if we use a long time period, we limit the sample that can be included in the model. (For example, since our data end in May 2024, if we used 24 months as the relevant time window, we would only be able to include those who applied by May 2022 in the model.) This trade-off may be particularly acute for FinTech lenders, as they are relatively recent entrants to the market and may not have a sufficiently large sample to measure default over a longer time horizon. An additional drawback of using a longer time window is that it lengthens the cycle of model deployment.

Figure A.4 illustrates this trade-off. In panel (a), we plot the cumulative proportion of borrowers who are at least 60 days delinquent as of x months after card origination. Because this cumulative proportion continues increasing over time since origination, any threshold of x months since origination will have the drawback of mislabeling some borrowers who default after month x as non-defaulters, and the shorter the time period considered, the greater the extent of this mislabeling. On the other hand, panel (b) shows the size of the modeling sample based on different time periods. To determine the size of the modeling sample, we impose two restrictions based on the threshold of x months since origination: (i) the account must be observed for x months by the end of our data period in May 2024, and (ii) the card must have at least one transaction within x months since origination. The number of observations in the modeling sample initially increases because the second constraint dominates: some borrowers do not make a transaction until they have had the card for a few months. After four months since origination, the number of observations in the modeling sample begins decreasing because the first constraint dominates: for a given value of x, the sample can only include those who applied for a card at least x months before our data end in May 2024.[20]

We use a threshold of 12 months since origination, which leaves us with a modeling sample of 146,036 borrowers. With this definition, 20% of the modeling sample is delinquent; Table A.1, column 3, shows how this compares to delinquency rates in other studies.

Profits model To measure profits, we use the data on account-by-month level revenues and costs described in Section 3.2 to quantify the profits obtained by the lender on each card, restricted to the first 12 months since card origination to be comparable to the default model. Consistent with the practice of our FinTech partner, we define realized profits at the account-by-month level as the sum

[20] An alternative would be to consider defaults over the entire time period for which we observe data, which varies by card. For example, for a card originated in November 2022, we would use the 18 months of data on default, while for a card originated in November 2023, we would use the 6 months of available data on default. The drawback of this method is that we observe default over different time periods for different cohorts of applicants, and the type of people applying for cards in November 2022 might differ from the type of people applying for cards in November 2023. Nevertheless, we conduct a robustness test using this definition of default, and estimate a slightly lower AUC of 0.77 (Table A.2). (When using this alternative method of defining the time period over which we measure default, we impose a restriction that cardholders must have had their card for at least 4 months and made at least one transaction, but measure default over however many months there are between the card's month of origination and May 2024.)

of interest revenue, BNPL revenue, interchange fee revenue, revenue from late-payment fees, and revenue from other fees, minus charge-offs, cost of rewards, funding costs, cost of fraudulent transactions, and other costs. This definition is consistent with Agarwal, Chomsisengphet, Mahoney, and Stroebel (2015) and broadly consistent with Guttman-Kenney and Shahidinejad (2023).[21] We then aggregate information of the first 12 months since card origination to compute account-level profits as

$$\begin{aligned}\text{Profits}_i = \sum_{t=1}^{12} \Big[&\text{Interest revenue}_{it} + \text{BNPL revenue}_{it} + \text{Interchange fee revenue}_{it} \\ &+ \text{Late payment fee revenue}_{it} + \text{Other fee revenue}_{it} \\ &- \text{Charge-offs}_{it} - \text{Rewards}_{it} - \text{Funding costs}_{it} \\ &- \text{Costs from fraudulent transactions}_{it} - \text{Other costs}_{it}\Big],\end{aligned} \tag{1}$$

where t defines months relative to card origination. We use this measure of Profits_i as our target variable for the profits model.

We also test a model in Appendix Table A.3 predicting a binary outcome variable equal to 1 if $\text{Profits}_i > 0$.

4 Machine Learning Methods

4.1 Algorithm

We use data on default and profits to train machine learning models using extreme gradient boosting, or XGBoost (Chen and Guestrin, 2016). XGBoost is a tree-based ensemble method that builds models sequentially, with each iteration improving on prior prediction errors via gradient-based optimization. It is widely used in both industry and academic research due to its scalability, strong predictive performance, and ability to handle complex nonlinearities. See Table 1 for an overview of papers that employ machine learning to predict creditworthiness, including details on the data and machine learning methods used.

Our default models are trained by minimizing log-loss or, equivalently, cross-entropy loss, which penalizes deviations between predicted probabilities and realized outcomes (the more the predicted probability diverges from the actual value, the higher is the log-loss value). Because log-loss evaluates the quality of predicted probabilities rather than only predicted class labels, it is well suited to applications such as lending, where probability estimates—not just binary classi-

[21] In contrast to Guttman-Kenney and Shahidinejad (2023) we do not subtract customer acquisition costs, since those are sunk at the time of origination and, as a result, are not relevant in this setting.

fications (e.g., default vs. no default)—are of interest. Our profits and profit components models are trained by minimizing mean squared error. In both cases, we tune the model hyperparameters using Bayesian optimization (Bergstra, Yamins, and Cox, 2013), implemented with 3-fold cross-validation. Appendix C provides additional methodological information, including further discussion of the XGBoost framework, hyperparameter tuning, optimization metrics, and class imbalance considerations.

Our models learn on a training set and are evaluated on a testing set. The training set corresponds to a random 80% sample of the modeling data, stratified by gender and target variable (default).[22] Stratification guarantees that the incidence of each class (defaulted and did not default) is preserved in both sets. The testing set—i.e., the remaining 20% of the modeling data—permits us to assess model performance on data unseen by the algorithm, as well as to guard against overfitting.

4.2 Approval Threshold

We take an economic approach to determining the threshold to use for each model's approval decisions. In particular, we use detailed account-by-month level data on profits and combine predicted probabilities of default or predicted profits with the realized profits from each account to determine via grid search the profit-maximizing approval thresholds *in the training sample* (to avoid overfitting): these thresholds are a 25% predicted probability of default and predicted profits of -7 MXN (which, as expected, is very close to a threshold of zero profits). We then evaluate approval decisions *in the testing sample* using these profit-maximizing thresholds determined in the training sample to estimate the threshold-dependent performance metrics and various 2×2 matrices, to compare borrowers approved and rejected by each model, and to estimate the overall profits earned by the lender with each model.

4.3 Model Performance Measures

Taking advantage of our detailed data on account-by-month level revenues and costs for the lender, we use the total profits earned by the lender with a particular model as our main performance metric. In addition to being tightly linked to the lender's ultimate objective of maximizing profits, this performance metric has the advantage of being comparable across the default model—which uses a binary target variable—and the profits model—which uses a continuous target variable.

[22] To accurately compare performance and predictions for the default and profits models and ensure that observed differences are not driven by differences in training-testing splits, we always use the training-testing split for the default models, which is stratified by gender and default. We stratify by gender since we estimate gender-segmented models in Appendix D to test whether segmenting the model by gender can improve credit allocation fairness without a substantive effect on the model's predictive performance.

Default model For the binary default model, we report four additional measures of model performance: AUC-ROC, precision, recall, and F1 score. The AUC-ROC measures the area under the receiver operating characteristic (ROC) curve, which plots the true positive rate against the false positive rate for all thresholds. Thus, the AUC-ROC, often abbreviated simply as AUC, is a threshold-free measure. An AUC of 0.5 implies that the model performs no better than random guessing, while an AUC of 1 implies that the model makes perfect predictions.

An advantage of AUC for assessing the binary default model is that it is threshold-free and is reported in nearly all studies, which facilitates comparison. Even when considering performance across settings with different proportions of the positive class (which in our context means different default rates) and even in the presence of *class imbalance* (i.e., when the incidence of one class is significantly higher than the other, such as far fewer defaulters than non-defaulters), comparisons using AUC remain meaningful due to its mathematical and empirical properties. While AUC does not depend on the proportion of the positive class in theory based on its mathematical definition (Flach and Kull, 2015), some have argued that in practice AUC could mask performance issues when dealing with highly or extremely imbalanced data (e.g., Davis and Goadrich, 2006), such as in settings in which the prevalence of the positive class is below 10%. Richardson et al. (2024) carefully analyze this point and show empirically, using simulated and real-world data, that AUC is a robust and appropriate comparison metric in many scenarios with different class imbalance, and argue that AUC provides fair comparisons of models even across contexts with different default probabilities. Furthermore, our default rate of 20% does not place us in a regime typically considered highly or extremely imbalanced and, as such, we are less at risk of metric distortion (e.g., AUC inflation) or model instability.

We also assess the default model's performance using precision, recall, and F1 score. Because these measures depend on the approval threshold, however, comparisons across papers and across models are potentially misleading if the classification thresholds differ. Precision measures the proportion of predicted positive cases that are actually true positives (in our context, the proportion of predicted defaulters who actually default); this is calculated as the true positives divided by the sum of true positives and false positives. Recall measures the proportion of actual positive cases (actual defaulters) that were correctly identified by the model, calculated as true positives divided by the sum of true positives and false negatives. The F1 score is the harmonic mean of precision and recall.

Profits model As with the binary classification model, we summarize predictive performance using multiple metrics in addition to the profits generated by the model. For the continuous profits and profit components models, we report R^2 as a summary measure of the goodness of fit of the predicted continuous outcome compared to the observed outcome. For the profits model alone

we also report mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE). However, when discussing the results in Section 5.2, we also argue that measures of goodness of fit have limitations *when being used to assess the performance of a model that is being used to make lending decisions*; this highlights the usefulness of evaluating all of our models—for both binary and continuous outcomes—based on how much profits they generate for the lender.

5 Results

5.1 Default Model

The benchmark default model, which uses all available features, is profitable for the lender and achieves an out-of-sample AUC of 0.791 (Table 3, Panel A, column 2). Studies in highly data-rich environments, in which credit bureau data and scores are often included in the algorithm, obtain AUCs typically in the 0.66 to 0.88 range (e.g., Blattner and Nelson, 2024; Blattner, Nelson, and Spiess, 2024; Di Maggio and Ratnadiwakara, 2025; Duarte, Fonseca, Kohli, and Reif, 2025; Meursault, Moulton, Santucci, and Schor, 2025; Netzer, Lemaire, and Herzenstein, 2019). In contrast, AUCs estimated by studies in middle-income countries are lower, typically in the 0.61 to 0.82 range (e.g., Agarwal, Alok, Ghosh, and Gupta, 2023; De Cnudde et al., 2019; Frost et al., 2019; Gambacorta, Huang, Qiu, and Wang, 2024; Lee, Yang, and Anderson, 2024; Lee, Yang, and Anderson, 2026; Rishabh, forthcoming). The AUC of our model predicting default is at the upper end of those from middle-income settings—which include traditional credit bureau data and scores as inputs in their algorithm, unlike ours since our sample has no formal credit bureau score (Table 1).

Our model predicting default also performs well using threshold-dependent metrics (precision, recall, F1 score), and the area under the precision-recall curve (AUC-PR, reported in Table A.1). Columns 3–5 of Table 3, Panel A, report the threshold-dependent metrics for our model, where we use the profit-maximizing threshold (estimated in the training data) for the threshold-dependent performance metrics. The model has a precision of 0.425, recall of 0.644, and an F1 score of 0.512. Estimates of these metrics from other studies, albeit with different approval thresholds that complicate comparing performance across studies, are shown in Table A.1.

Table 4, Panel A, columns 1–2 show a confusion matrix for the default model, again using the profit-maximizing approval threshold. We report the percent of observations in the testing sample based on whether they would be approved or rejected by the default model and whether they ultimately default or not. The proportion approved by the default model who do not default (true negatives, since "default" is the positive class in our model) is 62.2%, while the proportion approved by the model who do default (false negatives) is 7.2%. The proportion rejected by the model who do not default (false positives) is 17.6%, while the proportion rejected by the model

who do default (true positives) is 13.0%.

We also assess whether those approved or rejected by the default model are profitable in Table 4, Panel B, columns 1–2. As a benchmark, overall, 51.0% of borrowers are profitable (Table 2); among the subset approved by the default model, 52.7% generate positive profits, while 48.1% of those rejected by the default model would have also generated positive profits. While this breakdown makes it appear that the default model is not performing much better than making approval decisions at random, the quality of the model's predictions becomes clearer when we examine *how profitable* those approved or rejected by the default model were (Figure A.5a). Importantly, 69.4% of borrowers who generate large negative profits of –200 MXN per month or more are rejected by the default model, while the unprofitable borrowers approved by the default model are typically only slightly unprofitable: 67.4% of the unprofitable default-approved borrowers generate profits of between –40 and 0 MXN per month. Figure A.6a shows why this is the case: the distribution of borrower-level profits is bimodal, and borrowers who defaulted mostly belong to the lower-profit mode of very unprofitable borrowers (although a small fraction belong to the more profitable mode despite defaulting). In contrast, borrowers who did not default belong to the higher-profit mode and are either profitable or—primarily due to funding costs exceeding interest revenues—slightly unprofitable.

To investigate this further, we examine the total profits generated by implementing different approval thresholds, as well as the average observed profits and components of profits for borrowers with different predicted probabilities of default. Figure 1a sums the profits obtained by the lender across all individuals who would be approved based on a given approval threshold (i.e., whose predicted probability of default is below the threshold), and repeats this exercise for each approval threshold. Figures 2a and 3 show binscatters of average realized profits and the main components of profits, respectively, against predicted probabilities of default.

As expected, total profits in Figure 1a—which are normalized such that the most profitable model across the default and profits models using the profit-maximizing threshold from the training sample generates profits in the testing sample equal to 1—are highest near the profit-maximizing approval threshold.[23] If the lender is too restrictive—setting a predicted probability of default threshold that is too low—it will not lend to any borrowers and will generate zero profits. As it becomes less selective, it begins lending to the borrowers who are the least likely to default, who are slightly profitable on average (Figure 2a). The reason the average profits of the lowest-risk (lowest predicted probability of default) borrowers are positive but close to zero can be seen in Figure 3: these are users who are very unlikely to default and generate charge-offs (panel c), but this is partly

[23]The peak of the total profits curve does not correspond exactly to the profit-maximizing threshold since the curve is based on lending decisions made in the testing sample, while the profit-maximizing thresholds are determined in the training sample.

because they carry less revolving debt on the card, and thus do not generate as much interest or late payment fee revenue for the lender (panels a and e). They are also sophisticated about optimizing their shopping to generate rewards (Figure A.7) and thus do not generate interchange fee revenue net of rewards (panel d).

As the lender continues becoming less selective by increasing the approval threshold, average per-borrower profits in Figure 2a increase due primarily to the newly-approved borrowers carrying more revolving debt and still making their interest payments while mostly not defaulting, thus generating interest revenue. However, after the probability of default increases above about 10%, average profits for borrowers with predicted probabilities of default above that threshold begin to fall, as interest revenue levels off and the negative effect of charge-offs begins to dominate (Figure 3). Average per-borrower profits fall below zero for predicted probabilities of default above around 21%, and thus for approval thresholds above this value, the lender's total profits begin to fall in Figure 1a. This reinforces the intuition of the lender's first-order condition: to maximize profits it wants to continue approving borrowers until marginal revenue equals marginal cost, or equivalently until the marginal borrower being approved generates zero expected profits.

How predictive is each data source? We next assess the importance of each data source by comparing the profits generated by our benchmark model using all of the data sources to those generated by separate models trained with all features but one data source (Table 3). We again normalize the total profits generated by each model, dividing by the total profits generated by the most profitable model using all data sources (which numerically is the profits model rather than the default model, even though statistically we cannot reject that the two models generate the same amount of profits). In addition to assessing the importance of each data source by comparing the total profits generated by models that do and do not have access to that data source, we compare these models' predictive performance using AUC and other performance metrics. We further assess how much profits and predictive accuracy could be achieved by models that have access to *only* that data source in Table A.4.

The transaction-level data from the delivery platform have by far the largest marginal contribution to the lender's total profits and the model's predictive performance: a model that uses all data sources *except* transactions only generates 31.4% as much profits as a model with access to all data sources including transactions (Table 3, column 1). The AUC of a model with all data sources except transactions is 0.654, a reduction in AUC of 0.137 compared to the benchmark model (column 2). Because of the deterioration in the profitability and predictive performance of the model when transactions data are removed, the approval rate using the profit-maximizing threshold also drops substantially from 69.4% for a model with all data to 49.5% for a model without transactions data (column 6).

Omitting any of the other data sources has a much smaller impact on the profits and predictive accuracy of the model. The digital footprint user characteristics, which include the set of features in Berg, Burg, Gombović, and Puri (2020), comprise the second most-important data set, as a model without that information generates 91.3% as much profits as a model with all data and reduces the AUC by 0.014.[24] Omitting the no-hit score and limited credit history leads to no statistically significant change in profits compared to a model with all data and to an AUC reduction of only 0.005. Omitting census tract-level socioeconomic characteristics does not reduce profits or AUC. The finding that the transactions data are by far the most important data source for the model's performance is robust to using precision, recall, or F1 to measure predictive accuracy (columns 3–5). Furthermore, a similar story emerges when assessing the profits and predictive power of models that use *only* one data source: a model using transactions data alone generates 91.0% of the profits of a model with all data and an AUC of 0.769, while a model using only digital footprints data generates just 25.7% as much profits and an AUC of 0.63 (Table A.4).

Given the importance of the transaction-level data in our FinTech lender's competitive advantage over other lenders, as well as in how profitable lending to borrowers with no credit history can be, we next assess how the predictive accuracy of the model varies by the "thickness" of a user's transaction history. The number of transactions made through the app may be analogous to a formal credit history in the sense that the model might perform more poorly for those with a "thin" transaction history (few transactions) compared to those with a "thick" transaction history (many transactions). To assess this, we segment the data into quintiles by number of transactions on the Rappi app prior to credit card application and estimate separate machine learning models for each quintile.[25] For this analysis, we now normalize profits relative to the total profits generated by the most-profitable group and model, which corresponds to a model predicting default for those in the top quintile of number of transactions.

The profitability of the models is broadly increasing in transaction history: for those in the first quintile with only 2 or fewer transactions, the model generates 45.4% as much profits as the model for those in the fifth quintile with 29 or more transactions, while for those in the middle quintile with 7–13 transactions the model generates 65.7% as much profits (Table 5, panel A, column 2). The AUCs and other performance metrics reveal a similar pattern (columns 3–6): the AUC is 0.742 for the first quintile, 0.785 for the middle quintile, and 0.828 for the fifth quintile. These differences across quintiles may be due to a combination of three forces: more transactions data improves the predictive power of the model, the type of people with more transactions have easier-to-predict default (even if the number of transactions fed to the model were equalized across groups), and the

[24] See Appendix B for a condensed list of the features constructed from this data source, as well as from the other data sources.

[25] Table A.5 shows summary statistics by quintile of number of transactions.

type of people with more transactions generate more profits regardless of approval decisions.

5.2 Profits Model

Given that the lender's ultimate objective is to maximize profits, does a model that uses a continuous measure of profits as its target variable outperform a model that predicts default? We test this by estimating a model that predicts borrower-level profits, as defined in equation (1). Table 3, column 1, shows that we cannot reject that the profits model and default model generate the same amount of profits for the lender: the 95% bootstrapped confidence interval for the ratio of total profits from the default model to that of the profits model is [0.900, 1.016].[26]

Overall, the profits model is more restrictive than the default model, approving 61.0% of applicants compared to 69.4% in the default model (Table 3, column 6). However, the proportions of borrowers approved by the profits model who default (Table 4, Panel A, columns 3–4) and who are unprofitable (Panel B, columns 3–4) are similar to those of the default model. The profits model makes more "mistakes" in rejecting borrowers who end up not defaulting (since it is not trained to predict default): 63.9% of applicants that the profits model recommends rejecting did not default, compared to 57.5% of the borrowers that the default model recommends rejecting. Nevertheless, this does not translate to more "mistakes" in terms of which borrowers end up being unprofitable: a very similar percent of those approved by the default (47.3%) and profits (46.5%) models are unprofitable.

The analysis in Table 4, Panel B, ignores *how profitable* or unprofitable those approved and rejected by each model are. We show this in Figure A.5. While the overall profitability distribution of those approved and rejected by the default and profits models look similar (panels a and b), this is due to the two models making the same approval decision for the majority of applicants (Table 4, Panel C). In Figure A.5c we show the borrower-level profitability of those approved by only one model. The profits model approves *relatively* more highly-profitable borrowers who generate at least 200 MXN per month in profits, but also approves relatively more highly-unprofitable borrowers who generate at least –200 MXN per month in profits. Furthermore, the profits model rejects far more applicants, including a large mass of somewhat profitable applicants (as well as some slightly unprofitable ones). On net, the average profit-approved applicant is 18.8% more profitable than the average default-approved applicant, but the default model approves a larger number of applicants, and these additional applicants approved by only the default model are profitable on average. Thus, using the default model increases access to credit relative to the profit model.

We next compare the predicted probabilities of default and predicted profits for each borrower,

[26]Figure A.8 shows a histogram of the 10,000 bootstrap estimates used to construct this confidence interval. For each bootstrap estimate, we draw a bootstrap sample from the testing sample and calculate the profits from each model for that bootstrap sample, then take the ratio.

and assess the types of borrowers approved by each model. Figure 4 shows the predicted outcomes from the two models for each borrower in the testing sample. The profit-maximizing thresholds of a 25% predicted probability of default and predicted profits of -7 MXN divide the graph into four quadrants; the fraction of observations in each quadrant is shown in Table 4, Panel C. The upper-left quadrant shows applicants who would be approved by both models (58.2% of the testing sample), the lower-right quadrant shows applicants who would be rejected by both models (27.7%), the lower-left quadrant shows applicants who would be approved by the default model but rejected by the profits model (11.3%), and the upper-right quadrant shows applicants who would be rejected by the default model but approved by the profits model (2.9%).

To explore how the borrowers in these four quadrants compare, Table A.6 shows summary statistics for each of these groups. Focusing on differences between borrowers approved by only the default model ("default-approved") and borrowers approved by only the profits model ("profits-approved"), default-approved borrowers are less likely to be women, are less likely to use an iPhone, have made fewer orders on the app, and are less likely to pay orders in cash.[27] Regarding their contract terms, default-approved borrowers are awarded higher credit limits and spend more, but have lower statement balances and make larger payments, reflecting that they roll over a smaller fraction of their balances from month to month.

Figure 5 and Table A.6, Panel C, compare the account-level revenues and costs for borrowers in these four quadrants. Applicants approved only by the default model generate less interest revenue than those approved only by the profits model (58 vs. 95 MXN per card per month). However, the default-approved applicants are much less likely to default (17% vs. 28%), and average charge-offs (including zeros for those who do not charge off) are 43 MXN per card per month for default-approved applicants and 77 MXN per card per month for profits-approved applicants. The net effect is that average interest and BNPL revenues net of charge-offs are similar across the two models, at 23 MXN per card per month for default-approved borrowers and 24 MXN for profits-approved borrowers. Thus, it appears that the profits model attempts to identify borrowers that generate substantial interest revenue (by carrying revolving debt and making their interest payments) and who do not generate charge-offs by defaulting. However, some of the high-interest-revenue applicants it lends to do end up defaulting (at higher rates than default-approved borrowers). Predicting who does not generate charge-offs by defaulting *among those who generate interest revenue* is a different prediction problem than the one undertaken by the default model, which is to predict who does not default among the full sample. The default model accomplishes this by approving borrowers who carry less debt and thus generate lower interest payments but also default

[27] While iOS users are less likely to default and are more profitable (Table A.7), most iOS users are accepted by both models (Table A.8); the iOS users accepted by only the profits model default at relatively high rates (28%) but nevertheless generate more interest revenue (94 MXN per account per month) than charge-offs (75 MXN per account per month).

less.

The applicants approved by each model also differ in the interchange revenue net of rewards that they generate for the lender. However, interchange revenue net of rewards is small relative to interest revenue net of charge-offs: for example, for those approved by both models, interchange net of rewards is 3 MXN per card per month, compared to 49 MXN per month for interest and BNPL revenue net of charge-offs. Applicants approved by both models spend 3,291 MXN per month on average and generate 46 MXN per month in interchange revenue, which is about 6% higher than their rewards expenses of 43 MXN per month. The default-approved borrowers spend less (2,995 MXN per month), and the cost of rewards for these borrowers (37 MXN per month) fully offsets interchange fees (also 37 MXN per month). The profitability-approved borrowers, on the other hand, spend only 2,769 MXN per month, but who also have a higher 17% spread between interchange fee revenue (30 MXN per month) and rewards (25 MXN per month).

The differences in the spread between interchange revenues and rewards between those approved by each model suggests that the lower-risk borrowers identified by the default model are likely more sophisticated and better at optimizing their shopping behavior to maximize rewards. We use the account-level data on interchange fee revenue and rewards expenses to test this in Figure A.7, and find that lower-risk borrowers indeed generate higher rewards *per dollar of spending*. While these borrowers also shop at stores that generate higher interchange revenue for the lender (given that interchange fees vary by merchant type), their sophistication about rewards dominates. Interchange revenue net of rewards is negative for the lowest-risk borrowers and initially increases as risk increases, then levels off.

Turning to goodness-of-fit measures, the profits model appears to perform poorly: it achieves an R^2 of 0.085, a RMSE of 173 MXN, and a MAE of 121 MXN (Table 3, Panel B, columns 2–5). The discrepancy between the profits model generating as much total profits as the default model but having a low R^2 highlights the limitations of using measures of goodness of fit to evaluate models being used to make lending decisions. In particular, the lender only cares about errors in predicted profits if they are large enough to cause it to make the wrong loan approval decision; since the profit-maximizing approval threshold is unsurprisingly very close to zero profits, the lender cares about whether it gets the sign wrong on a particular applicant (and it cares more about getting the sign wrong for those applicants when they would have been highly profitable or unprofitable than when they would have been slightly profitable or unprofitable). In other words, if a borrower's predicted profits are –20 MXN but realized profits are –200 MXN, the lender's profits are unaffected by this prediction error, whereas R^2 and other measures of goodness of fit *are* affected.

How predictive is each data source? As was the case for the default model, the transactions data comprise by far the most important set of features (Table 3, Panel B). Without the transactions

data, the profits model would only be 39.6% as profitable as a model predicting profits with all data sources. Interestingly, the transactions data are relatively less important for the profits model than for the default model, and if our lender did not have access to transactions data they would generate 31.9% more profits by implementing a profits model rather than a default model (but less than half of what they generate with either model *with* access to transactions data). Removing transactions data also reduces the approval rate of the profits model from 61.0% to 41.5%. Removing the digital footprint data—in contrast—only reduces profits by 7.8%, and removing the no-hit score and limited credit history data or the census tract-level socioeconomic characteristics does not lead to a statistically significant change in profits.[28] The patterns across profits models estimated separately for each quintile of number of transactions are also quite similar to those of default models: a profits model for borrowers in the top quintile of number of transactions is 72.8% more profitable than one for those in the bottom quintile (Table 5, Panel B).

5.3 Profit Component Models

The components of profits in equation (1) may be predictable by different aspects of the data. In this subsection, we estimate separate models for the six major components of profits from equation (1): interest revenue, BNPL revenue, charge-offs, interchange fee revenue net of rewards, late payment fee revenue, and funding costs.[29]

For the separate models to predict each component of profits, Table 6 shows the total profits generated for the lender if it were to base its credit origination decisions on a model predicting only that component of profits, as well as the R^2 of each model.[30] As before, we determine the profit-maximizing threshold for the corresponding profit component in the training sample and apply the threshold to predicted values in the testing sample. Users in the testing sample with predicted values above the threshold are approved.[31] The total profits metric is again normalized by the total profits generated by a model predicting profits with access to all data sources.

A model predicting the losses incurred to charge-offs generates just as much profits as a model

[28] The transactions data are similarly important when assessing the other performance metrics: the R^2 falls from 0.085 with all data to 0.022 without transactions data, and there is a corresponding increase in MSE and RMSE when removing transactions data. The only performance metric that does not change when removing transactions is MAE, which indicates that the transactions data are especially useful in helping to predict observations that would have a large prediction error if the model did not have access to those data.

[29] We combine interchange fee revenue and rewards to estimate a model for "interchange fee revenue net of rewards" given that these two measures are nearly perfectly negatively correlated (Figure A.7).

[30] Because R^2 is a function of the variance of the outcome being predicted, it should not be used to compare across the models that predict different components of profits (i.e., across columns); instead, it should only be used to diagnose how much predictive accuracy is lost when removing data sources for a particular outcome (i.e., within column across rows).

[31] The only exception is the model for late payment fee revenue. In that case, portfolio profits are maximized by approving users with predicted late payment fee revenue *below* the threshold.

predicting profits: the 95% bootstrapped confidence interval of the total profits generated by the model predicting charge-offs, normalized by the total profits generated by the profits model, is [0.95, 1.057]. Credit origination decisions based on predictions of other profit components generate far lower profits for the lender: for example, a model predicting interest revenue generates 11.7% as much profits as a model predicting profits.

To see why a model predicting interest revenue performs poorly, we turn to Figure A.6. Panel (b) shows that while all cardholders who generate large profits generate interest revenue, cardholders who generate interest revenue also account for a large share of the most negative profits, which, as shown in panel (a), are due to default. Furthermore, panel (c) shows that the most profitable borrowers generate substantial interest revenue on average (but do not default). Meanwhile, borrowers who are slightly unprofitable do not generate much interest revenue, while those who are more unprofitable are a mix of those who do generate some interest revenue by revolving debt and making some repayments but later defaulting and never-payers who borrow on the card but never make a payment. The interest-revenue model is not able to distinguish between the unprofitable and profitable borrowers who generate interest revenue, as they arguably share observable characteristics correlated with borrowing intensity: panel (d) shows that individuals with the highest predicted interest revenue also incur the largest charge-offs on average.

Figure A.6b also shows that those who do not generate any interest revenue fall into two distinct groups: transactors—also known as convenience users—who pay off their balance in full every month and thus generate no interest revenue, and never-payers, who accumulate debt on the card and are charged interest but default on that balance without ever making a repayment, thus never delivering that interest revenue to the lender. Unlike in the US context, transactors are typically slightly unprofitable for our lender because the lender incurs financing costs over the period of time between when a purchase is made and the due date of the repayment corresponding to that purchase, and the lender does not earn a large enough spread between interchange fee revenue and rewards costs to make up for these funding costs.

How predictive is each data source? We next test which kinds of new data help most to predict each component of profits (Table 6). For all components of profits, digital transactions data are the most important data source (with the caveat that for interchange revenue net of rewards, transactions data are most important based on R^2, whereas based on profits generated for the lender, we cannot reject that each data source is equally important). Removing proprietary transactions data eliminates the ability to generate any profits by making credit origination decisions based solely on predicted interest revenue, and it reduces the profits generated by making these decisions based on a model predicting charge-offs by 65.7%.

5.4 Heterogeneity

We next turn to assessing heterogeneity across sociodemographic groups. Do the default and profits models disagree in terms of which groups of applicants are approved? Are different types of data differentially predictive for different groups? And are some components of profits more accurately predicted for some groups, and other components of profits more accurately predicted for other groups? To investigate this, we focus on four sociodemographic characteristics (or, in some cases, proxies for sociodemographic characteristics): gender, age, Apple iOS or Android user, and the percent of transactions that are made in cash. We chose these characteristics because there is existing evidence in the literature documenting heterogeneity in FinTech adoption, use, and impacts across each of these dimensions. Table A.9 shows, for each of these groups, summary statistics about selected features used in the machine learning model, contract terms, and profit components.

Gender There are important gaps in FinTech use across gender (Chen et al., 2021), as well as differences in willingness to share transactions data (Armantier et al., 2024).

We first assess whether men or women are more likely to be approved by each model, as well as how men and women differ in their probability of default and profits. Women are 2.2 pp more likely than men to be approved by the default model, and 5.9 pp more likely than men to be approved by the profits model (Table A.8). This reflects that when assessing the full sample (not conditioning on the default or profits models' approval decisions), women are 0.6 pp less likely to default and generate 0.05 standard deviations (SD) more profits (Table A.7), driven both by both higher interest revenue and lower charge-offs (Figure A.9a).

Despite women being less likely to default and more profitable on average, women generate a smaller fraction of the lender's total profits, generating 45.7% of the profits from the default model and 44.1% of the profits from the profits model (Table A.10). This is largely because men make up a substantially larger share of applicants (61.1%) and thus of approved borrowers despite lower approval rates (60.4% of those approved by the default model and 58.5% by the profits model are men).

We next ask whether different data sources are differentially predictive for men and women. We find that while transactions data are extremely important for both men and women, they are *more important* for men: normalizing by the profits generated among men or women by a model using all data sources, Table A.11 shows that a model without transactions data generates only 22.9% as much profits for men compared to 41.4% as much profits for women. The higher importance of transactions data for men can also be seen by assessing the fraction of applicants approved by models that differ in the data to which they have access (Table A.12). While men have a 2.2 pp lower probability of approval than women (70.8% of whom are approved) using a model with access to all data, they have a 4.6 pp lower probability of approval than women (52.3% of whom are approved)

using a model without access to transactions data. For the other data sources, however, we cannot reject that the marginal contribution of each data source to profits is the same for men and women.

Are different components of profits differentially well-predicted for men and women? We find that this is indeed the case: although men generate 55.9% of the portfolio's overall profits using a profits model (Table A.10), we cannot reject that women and men generate the same amount of profits using an interest revenue model: based on the point estimates for a model predicting interest revenue, women generate 7.3% as much profits as the profits of the profits model across both men and women, compared to 4.3% for men. In contrast, men generate more profits than women in a model predicting charge-offs, consistent with the profits model (Table A.13). We note that these differences could be due either to the model predicting these components differentially well for men and women, or due to these components being more correlated with profits for one group than for the other.

Finally, while the models we have considered thus far are estimated pooling data from both men and women, we assess whether estimating separate models for men and women improves credit allocation fairness without a loss in the profits generated by the model or in predictive accuracy in Appendix D.

Age FinTech adoption and use is lower among older demographics (Crouzet, Ghosh, Gupta, and Mezzanotti, 2024; Jiang, Yu, and Zhang, forthcoming), and older people are also less willing to share transactions data (Armantier et al., 2024). In our sample, older users—measured as those above the median age—are 3.2 pp less likely than younger users to be approved by the default model and 7.6 pp less likely to be approved by the profits model, due to being 1.4 pp more likely to default and being 0.06 SD less profitable (Table A.7). Default is also better-predicted for older uses, with an AUC of 0.805 compared to an AUC of 0.773 for younger users (Table A.10). Despite making up 52.3% of applicants, older users generate only 46.6% of the total profits from using the default model and 48.7% of those of the profits model. This is partly due to older users being less profitable, as described above, and partly due to approval rates for older users being substantially lower (Table A.12).

As was the case with gender, transactions data are again more valuable for the less-profitable group that has lower approval rates (in this case, older applicants). A default model without transactions data only generates 16.6% as much profits from older individuals as a model with all data sources, compared to 44.3% of profits generated without transactions data for younger applicants. In contrast, the digital footprint data—although substantially less valuable than transactions data for both groups—is more valuable for older borrowers in the case of the default model but slightly more valuable for younger borrowers in the case of the profits model (Table A.11). This can also be seen when looking at approval rates, where for example a profits model without access to trans-

actions data would only approve 34.5% of older applicants, while a profits model without access to digital footprint data still approves 54.1% of these applicants (Table A.12).

Phone operating system The type of phone people use (Apple vs. Android) is highly correlated with income (Bertrand and Kamenica, 2023) and is an important predictor of default (Berg, Burg, Gombović, and Puri, 2020). We find that Apple iOS users are more likely to be approved by both the default and profits models (Table A.8), because they default less and are more profitable (Table A.7). Although iOS users represent only 38.0% of applicants and 41.9% and 45.1% of those approved by the default and profits models, respectively, they generate nearly half (49.8%) of the profits using either model (Table A.10). The digital transactions data from the Rappi platform are more important for Android users, as default and profits models without access to these data only generate 18.2% and 30.4% as much profits, respectively, as a model with all data sources, compared to 44.6% and 49.0% for iOS users (Table A.11). Furthermore, the approval rates for Android users drop by more when transactions data are removed, e.g., from 54.7% to 26.0% for the profits model (Table A.12), whereas the corresponding decline for iOS users is from 71.2% to 66.8%.

Turning to models predicting each component of profits (Table A.13), interesting patterns emerge. Models predicting interest revenue or interchange net of rewards fail miserably for Android users and do not generate any profits from them, while for iOS users they generate 15.2% and 11.3% as much total profits as a model predicting profits. In contrast, Android users generate *more* profits than iOS users in a model predicting late payment fee revenue and the same level of profits in a model predicting charge-offs.

Percent of transactions made in cash Cash is still widely used in Mexico (Gertler, Higgins, Malmendier, and Ojeda, 2025; Higgins, 2024), even for transactions on digital platforms such as Uber (Alvarez and Argente, 2022). Indeed, users can choose to make transactions in cash on the Rappi app—in which case they pay the courier upon delivery—and 32.9% of Rappi transactions made by users in our sample were paid in cash. In most contexts, cash payments are not observable, which prevents lenders from using them to predict creditworthiness; this motivates studying the impact of making payments verifiable through cashless technology (Ghosh, Vallee, and Zeng, forthcoming; Alok, Ghosh, Kulkarni, and Puri, 2025). In our context of transactions on a digital platform, however, even cash payments are observable and can be used to evaluate creditworthiness, as we do here. Looking at the impact of each model on people who rely heavily on cash allows us to see the impact on users whose transactions would otherwise be unobservable, even by a lender using digital payments for credit scoring.

We divide the sample into those whose percent of cash transactions is above or below the

median.[32] Those who make a higher fraction of their transactions in cash default more and are less profitable (Table A.7), and are thus substantially less likely to be approved by both models (Table A.8). Nevertheless, after this selection takes place in terms of approvals, around half of the total profits generated by each model are generated by those with above-median cash usage (Table A.10). The transactions data are *more important* for users with more cash usage: a model without access to those data only generates 29.9% as much profits as a model with access to those data, compared to 33.3% for those with below-median cash usage (Table A.11). Approval rates drop substantially for both groups if the model does not have access to transactions data: only 44.0% of heavier cash users are approved by a model without access to transactions data (down from 62.1% for a model with all data sources), while approval rates for those with below-median cash usage fall from 78.1% to 56.2% when the model loses access to transactions data (Table A.12).

6 Robustness

While the monitoring and maintenance of ML models have become integral parts of any deployment system to ensure their effectiveness and reliability, the choice of target variable and the relevant loan performance window (in our context, 12 months) have implications for the recency of the data entering the model, as well as for how quickly model performance can be assessed and models retrained. Concretely, our model trained on data up to the end of May 2023 using a 12-month default window could have been deployed, at the earliest, in June 2024. By then, the 2023 data might no longer be representative of the 2024 environment, because of changes in macroeconomic conditions, technologies, the pool of applicants, or the relationship between features and predictions.

Macroeconomic indicators in Mexico during the period January 2021 to June 2024 were relatively stable, with modest economic growth and mildly fluctuating levels of unemployment and inflation (Figure A.2). As discussed in Section 3.1, the levels of these variables from 2021–2024 were not unusual compared to the levels in previous decades going back to 1995. Our sample period does not include a macroeconomic downturn and, as a result, we are not able to evaluate how the performance of our models would change during a crisis. Prior research shows that data-driven lending models may perform poorly when economic conditions differ materially from those prevailing during the training period (Ben-David, Johnson, and Stulz, 2025). Whether and to what extent this holds for models trained with alternative data to assess borrowers without credit history remains an open question.

[32]The median percent of transactions made in cash, measured at the user level, is 50.0%. In addition, 5,294 users in our modeling sample had zero Rappi transactions; they are excluded from the "% cash above median" and "% cash below median" groups.

6.1 Out-of-Sample/Out-of-Time Test

To assess whether a model trained in the beginning of our sample period still provides reliable predictions at the end of the sample period, we implement an out-of-sample/out-of-time test, in the spirit of Berg, Burg, Gombović, and Puri (2020). Unlike our benchmark specification, which relies on a random training–testing split that ignores the time dimension, this exercise evaluates how a model trained on earlier data performs when applied to a later cohort of applicants. Specifically, we train a new model that splits the modeling sample of 146,036 borrowers into two periods based on date of origination: January 1st, 2021 to August 31st, 2022 and September 1st, 2022 to May 31st, 2023.[33] The first subperiod represents our training set; the second corresponds to the out-of-sample/out-of-time testing set. We chose the cut-offs to achieve close to an 80%/20% training-testing split, as in our benchmark model.

Table A.14 shows performance metrics for the out-of-sample/out-of-time test. The model trained on the earlier part of our sample period and tested on the later part of our sample period performs nearly as well as the benchmark model with a random training-testing split. In particular, the AUC is 0.779 (compared to 0.791 for the benchmark model), precision is 0.393 (compared to 0.425), recall is 0.676 (compared to 0.644), and the F1 score is 0.488 (compared to 0.512). Differences in performance metrics between these two models may reflect both differences in the models and differences in the testing samples, which come from different time periods. We therefore do not interpret this as a pure model performance comparison in the sense typically used in the machine learning literature, where the test sample is held fixed, but rather as an evaluation of temporal stability. Overall, this suggests that our model is reasonably robust to changes in applicant composition and macroeconomic conditions within the relatively stable range observed during our sample period.

6.2 Performance across Geographies with Different Economic Conditions

We also assess the robustness of the model performance across geographies with heterogeneous economic conditions. We use predictions from our benchmark model (see Section 4) on the original testing sample—not from models estimated separately for each state—and compute state-level AUCs for borrowers, who are assigned to a state based on the address provided at the time of application.

Figure 6 shows scatter plots of state-level AUCs (vertical axis) against state-level GDP per capita or its change (horizontal axis). Panel (a) uses GDP per capita for 2021 (the beginning of our analysis period), panel (b) uses GDP per capita for 2023 (the latest year for which state-level GDP

[33] As in our benchmark model, we can only include applicants through May 2023 because we measure the target variable over a 12-month window, and have performance data through May 2024.

is available), and panel (c) uses the change in GDP per capita between 2023 and 2021.[34]

We find no systematic relationship between economic activity and AUCs, suggesting that the performance of the model is stable across areas with different economic environments, within the relatively stable conditions observed during our sample period. We do not find evidence that the model has lower predictive power in states with negative or low GDP growth: the average AUC among the ten states with the lowest GDP growth (0.778) is not lower than the overall state-level average (0.777). Moreover, the AUC of the only state with negative GDP growth (0.797) is also higher than the overall average. Some states with high GDP per capita have narrower confidence intervals—due to a larger presence of our lender and hence a larger sample size. We note that pooling observations from states with low GDP per capita would lead to more comparable sample sizes and confidence intervals while preserving the null relation between economic activity and AUCs. This conclusion applies to the relatively stable macroeconomic conditions observed during our sample period, which does not include a national downturn.

Additional tests and robustness analyses, including formal data drift tests and results using penalized linear models rather than XGBoost, are reported in Appendix E.

7 Discussion

7.1 Why Do Lenders Use Default Models for Credit Origination Decisions?

Most FinTech lenders and many traditional lenders use default models for credit origination decisions in practice (Rajan, Seru, and Vig, 2015; Ben-David, Johnson, and Stulz, 2025). Given that the default and profits models generate very similar levels of total profits, why do so many lenders opt for default models?

One possible explanation is that, when the target variable is observed over short horizons (e.g. 12 months since origination, as in our analysis), default models may screen out borrowers who will be unprofitable over the lifetime of the relationship more effectively than profits models. This is because, empirically, most of those who default generate negative lifetime profits for the lender, while those who generate substantial interest revenue and profits over the first 12 months since origination (because they carry a substantial revolving balance but do not default) may experience a shock later on and default, making them unprofitable over the lifetime of the card. Because FinTech lenders are relatively new entrants, however, they face a binding trade-off between the sample size used in their models and the length of time over which they can measure profits to construct the

[34]In Figure A.10 we replicate the analysis using GDP per capita without revenue from oil. This is a common adjustment to state-level GDP in Mexico since revenue from oil is appropriated by the federal government and not by the states. Since default is measured up to May 2024 but the latest measurement available of state-level GDP is 2023, in Figure A.11 we replicate the analysis with a quarterly indicator of economic activity available for the first half of 2024. As before, there is no clear pattern between economic activity and model performance during that period.

profits model's target variable; in contrast, longstanding traditional lenders are able to implement models that seek to maximize a customer's lifetime value (Cowan, Mercuri, and Khraishi, 2023; Ekinci, Uray, and Ülengin, 2014).

A borrower who has a low probability of default but also low profits generated by the credit card—who will thus be approved by the default model but not the profits model—may also have a higher lifetime value to the lender due to cross-selling of other new or existing products in the future (Basten and Juelsrud, 2023; Li, Sun, and Montgomery, 2011). On the other hand, a borrower who carries high revolving debt and generates high interest revenue but has a higher probability of default—who will be approved by the profits model but not the default model—will likely be unprofitable even after accounting for cross-selling if they end up defaulting. This is an additional reason that, even if a profits model appears to generate similar profits from the credit card compared to a default model, a profit-maximizing lender might nevertheless implement a default model.

7.2 Endogenizing Interest Rates

A profit-maximizing lender can choose not only whom to approve or reject, but also what price to charge each borrower. Our FinTech partner generally charges the same interest rate at origination to all approved borrowers with no formal credit history at a given point in time (Figure A.3) and subsequently reprices the portfolio after origination.[35] Repricing over time allows the lender to incorporate information revealed by borrowers as they use and make repayments on the card (Petersen and Rajan, 1995; Nelson, 2025), and reduces the impact of the interest rate assigned at origination. It is also consistent with the industry practice of handling approval decisions and account management decisions (e.g., repricing and cross-selling) separately (Thomas, 2009).

Throughout this paper, we focus only on approval decisions using different predictive models and take prices as given. Our comparison of total profits under the default model and the profits model implicitly assumes that pricing policies (namely, the rate assigned at origination and subsequent adjustments) are identical under both models. Furthermore, the card-level profits we observe for each approved borrower reflect the pricing policy observed in our data and may differ under alternative pricing regimes.

A different pricing policy could influence the realized default and profits of each borrower, as well as the profit-maximizing approval thresholds, approval decisions, and ultimately—for all of these reasons—the portfolio profits of both the default and profits models. For example, for the default model, the set of borrowers could become more or less risky, both through a change in selection and through the causal effect of interest rates on credit card use and default (Karlan and Zinman, 2009). Further, the profit-maximizing threshold could change to approve riskier borrowers

[35]The data in Figure A.3 only show interest rates *at origination* of cards issued in a particular month; it does not show variation from the repricing of cards over time.

despite their higher default rates if the increase in portfolio interest revenue from lending to these additional borrowers at higher interest rates more than compensates the increase in portfolio charge-offs. For the profits model, individuals who were previously below the approval threshold may become profitable if priced optimally, and the level of profitability of those who were already above the approval threshold can also change.

To explore the relationship between interest rates, default, and profits in our data, we exploit the variation over time in the interest rates our lender assigned at origination (Figure A.3). Interest rates are not meaningfully correlated with default (Table A.15, column 1). This is consistent with evidence to date documenting that the causal effect of changes in interest rates on credit card default is relatively small (Castellanos et al., 2025; Nelson, 2025). Columns 2 and 3 of Table A.15 show that in our data, interest rates are correlated with interest revenue (and profits), although non-monotonically, and of course with the caveat that these correlations primarily exploit variation over time in the interest rates assigned by our lender at origination and are thus not well-identified causal estimates of the impact of changing the interest rate at origination.

To determine the optimal interest rate, a lender would need estimates of the elasticities with respect to interest rates of extensive-margin demand (take-up of the credit card), intensive-margin demand (spending on the card), repayment, and default, which would be difficult to obtain without running randomized experiments. An optimal pricing policy would require both setting an optimal price at origination and optimally repricing dynamically when new information about borrowers is revealed. Moreover, to make joint approval and optimal pricing decisions, these elasticities would then need to be incorporated into a model that predicts profits under different pricing policies. This would enable the lender to select the optimal pricing policy for each borrower based on the predicted profits associated with assigning each potential pricing policy to that borrower, and to then determine a profit-maximizing approval threshold.

The extent to which lenders engage in this complex optimization problem requiring many experiments with randomized interest rates both at and after origination to accurately set the optimal interest rate is not well-documented (Berg, Fuster, and Puri, 2022); because the variation in interest rates at origination in our data does not come from randomized experiments, we do not tackle jointly making approval decisions and setting optimal interest rates in our models. It remains an open question how this joint optimization would affect lenders' profits, and how it would affect the comparison between the total profits generated by a default model vs. a profits model.

7.3 Reject Inference

A limitation of our data is that repayment and default outcomes are observed only for applicants approved for credit by our FinTech partner, under its risk preferences and parsimonious approval rules designed to identify borrowers with low default risk (see Section 3). Consequently, repayment

behavior is unobserved for a substantial share of rejected applicants, who may differ systematically from those in our modeling sample (Blattner and Nelson, 2024; FinRegLab, 2023a).[36] This selection problem is known as *reject inference* in credit scoring (FinRegLab, 2023b; Caro, Gillis, and Nelson, forthcoming) and is closely related to the *selective labels* problem in machine learning (Lakkaraju et al., 2017).

One common approach to reject inference is to use *credit-bureau proxies*, i.e., to proxy repayment on the focal product for rejected applicants using repayment behavior on other credit products and/or in other periods observed in bureau data (FinRegLab, 2023b; Blattner and Nelson, 2024; Caro, Gillis, and Nelson, forthcoming). By definition, however, bureau-based proxies are not available for our target population, who lack meaningful credit history.[37]

For rejected applicants—and especially for those with limited or no credit history as in our setting—a feasible (though operationally more costly) solution would extend credit to a random subset of applicants who would otherwise be rejected by the model. This generates outcome labels in part of the rejected region, allowing future models to be trained on data that better represent the full applicant pool rather than only applicants accepted under the status quo decision rule. This approach is consistent with the practice of some lenders that periodically lend to a small subset of applicants below the approval threshold to learn about model error rates (Meursault, Moulton, Santucci, and Schor, 2025).

8 Conclusion

Traditional financial institutions such as banks typically do not lend to borrowers without formal financial history, and banks' past attempts to expand credit access to first-time formal borrowers with no credit history have often failed (Castellanos et al., 2025). Meanwhile, online FinTech lenders have rapidly proliferated around the world (Berg, Fuster, and Puri, 2022), and proponents argue that FinTech lending promises to expand access to credit and increase financial inclusion by using alternative data sources to evaluate creditworthiness. In other words, if alternative data sources such as call logs, social media interactions, and retail transactions can accurately predict credit *on their own* for people with no credit history, these potential borrowers would no longer necessarily be excluded from credit markets.

Many FinTech firms leverage alternative data to assess creditworthiness, and a growing academic literature studies the predictive performance of these data. However, most FinTech lending

[36] We did not receive the full set of data sources for the rejected pool (see Section 3 for details), nor did we receive any information for the riskiest rejected applications, per the lender's assessment. In the rejected-applicant data that we do observe, the median no-hit score is 625, compared to 638 among approved applicants.

[37] Other approaches impute missing outcomes for rejected applicants using observable characteristics and repayment behavior among approved applicants (FinRegLab, 2023b). These methods do not address selection on unobservables and, in some settings, have been found to be of limited use (Crook and Banasik, 2004).

algorithms still rely, at least in part, on conventional bureau scores (Johnson, Ben-David, Lee, and Yao, 2023), and existing evidence focuses largely on applicants with formal borrowing histories and bureau-reported scores. When lenders use bureau scores as model inputs and, in practice, approve only applicants with formal credit histories, they fall short of the FinTech promise to expand access for underserved and excluded populations.

We train machine learning models to assess credit risk for a population with no conventional credit score in the credit bureau, either because they have no credit history or an insufficient credit history for the credit bureau to generate a credit score. We show that a model trained on alternative data sources for this population is effective at predicting default. In particular, the predictive accuracy of our model is at the upper end of studies in middle-income countries (and is also higher than that of some studies in more data-rich environments such as the US), despite other studies focusing on populations that are already more financially included and that have conventional credit scores at the time of loan application, and despite these credit bureau scores being used as an input for credit assessment.

When we compare a model predicting default—as used in practice by many traditional and FinTech lenders—to a model predicting profits, we find that, with our modeling approach and under the pricing policy used by our lender, we cannot reject that the two models generate the same amount of profits. Digital transactions data—which are proprietary as they originate from applicants' order history on a delivery app—are far more important for models' predictive accuracy and profits generated than other data sources that could be theoretically acquired by any FinTech lender. This holds across across each subgroup we analyze and for each component of profits.

Our findings have three potential implications.

First, our results suggest that alternative data can be used to expand credit to underserved populations. In countries where alternative data are regularly used, this can have a direct impact on the strategy of FinTech companies considering serving this segment. In countries where regulatory agencies are evaluating the costs and benefits of using alternative data for credit origination (e.g., Consumer Financial Protection Bureau, 2017), our conclusions can inform the discussion by showing that predicting the default behavior and profitability of populations without a credit history is feasible. The distributional consequences of more granular predictions, privacy and fairness considerations, the performance of such models in more adverse macroeconomic environments, and the net effect of all of these forces on welfare require further research (Berg, Fuster, and Puri, 2022; Fuster, Goldsmith-Pinkham, Ramadorai, and Walther, 2022; Armantier et al., 2024).

Second, a shift by lenders from models predicting default to models predicting profits could have implications for aggregate credit access and financial stability. When approval thresholds are set to maximize lender profits in each model, the profits model lends to fewer borrowers, reducing aggregate credit access. Because it also selects riskier borrowers on average, who generate higher

interest revenue but default at higher rates, shifting to a profits model could reduce financial stability in the presence of correlated shocks across borrowers.

Third, our finding that delivery-app transactions data are by far the most informative source for predicting creditworthiness has implications for pricing, market power, and open banking regulations. In particular, the transactions data constitute a data source that is available to our lender through its affiliation with the Rappi delivery app, but not to other lenders. This contrasts with digital footprint data, which can be obtained by any FinTech lender from the applicant's phone and application. This can have implications for loan pricing, as lenders without proprietary transactions data may be unwilling to lend to borrowers without credit history, leading to market power for lenders with access to these data. Conversely, the importance of these proprietary data for credit assessment of borrowers without credit history suggests that open banking policies that enable consumers to send the data generated through their actions with one firm to various potential lenders—and that are expansive enough to include not only data from bank accounts but also data from online platforms such as delivery apps—could increase the number of lenders willing to lend to this segment, increase competition, and improve consumer welfare.

Table 1: Comparison of studies that predict creditworthiness

Citation	Country	Loan type	% with credit bureau data	Data	Methods	AUC
(1)	(2)	(3)	(4)	(5)	(6)	(7)
This paper	Mexico	FinTech credit card	0%	Delivery app transactions data, digital footprints, credit history for those with limited credit history (but no credit scores)	XGBoost	0.796
Agarwal, Alok, Ghosh, and Gupta (2023)	India	FinTech loan	81%	Digital data from mobile phones; call logs; demographics, address, bank statements, salary slips; traditional credit score (CIBIL)	Random forest, XGBoost, logit	0.738 for sample with credit history, 0.674 for sample without credit history
Albanessi and Vamossy (2024)	US	Credit card	100%	Credit bureau files	Hybrid deep neural network/gradient boosting	0.906
Berg, Burg, Gombović, and Puri (2020)	Germany	FinTech loan	94%	Digital footprints (device type, operating system, email service provider, writing style, etc.), credit scores	Logit	0.734
Björkegren and Grissen (2020)	A middle-income South American country	Mobile phone airtime credit	85%	Mobile phone call logs and text data, history of phone bill payment, credit bureau data and credit scores	Random forest, logit	0.772

Citation	Country	Loan type	% with credit bureau data	Data	Methods	AUC
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Blattner and Nelson (2024)	US	Mortgage	100%	TransUnion consumer credit report data and public and Infutor data on consumers' mortgage transactions, socio-economic characteristics, and lenders' information, Vantage credit scores	XGBoost, random forest, logit	0.840 for minority and 0.887 for non-minority sample
Blattner, Nelson, and Spiess (2024)	US	Credit card	100%	Credit bureau records	XGBoost, random forest, logit, elastic net, neural net	0.867
Butaru et al. (2016)	US	Credit card	100%	Account-level credit card data from 6 major commercial banks, macroeconomic variables, credit bureau data including credit score	Random forest, logit	Not reported
Caire and Vidal (2024)	India	FinTech loan to microenterprises	95%	Data from partner gig worker platforms on earnings, working hours, driver ratings, and other measures	Logit	0.710
De Cnudde et al. (2019)	Philippines	Microfinance loan	Not reported	Facebook data (sociodemographics, likes, comments, social network)	Linear support vector machine	0.825
Di Maggio and Ratnadiwakara (2025)	US	FinTech loan	100%	Age, annual income, debt-to-income ratio, FICO credit score	Random forest	0.659
Duarte, Fonseca, Kohli, and Reif (2025)	US	Mortgage, student loans, credit card	100%	Payment history data, debt and collections, credit utilization and credit limits, credit history length, recent credit activity	XGBoost	0.712
Frost et al. (2019)	Argentina	FinTech SME loan	100%	Sales data and internal rating from e-commerce platform, credit score	Logit, XGBoost	0.764

Citation	Country	Loan type	% with credit bureau data	Data	Methods	AUC
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2022)	US	Mortgage	100%	Income, loan-to-value (LTV) ratio, origination amount, FICO credit score, etc.	Random forest, XGBoost, logit	0.861
Gambacorta, Huang, Qiu, and Wang (2024)	China	FinTech loan	100%	Call data including frequencies, duration, etc., app use data, credit history, default history, frequency of credit card usage, credit scores produced by FinTech based on formal credit history	Logit	0.607
Hair, Howell, Johnson, and Matsumoto (2025)	US	Small business loans	100%	FICO score, business age, number of employees, requested loan amount, state, industry, region, loan type	RF, logit, OLS	0.663
Huang et al. (2023)	China	FinTech SME loan	100%	Asset data such as housing property, gender, age, and business type, data on provincial and municipal economy, MYbank credit histories	Random forest	0.841
Iyer, Khwaja, Luttmer, and Shue (2016)	US	FinTech P2P loan	100%	Borrower income, number of past delinquencies, maximum interest rate borrower is willing to pay, picture and text description in loan application, Experian credit score	OLS	0.714
Jagtiani and Lemieux (2019)	US	FinTech P2P loan	100%	Personal installment loan-level data from LendingClub's unsecured consumer platform, similar loan-level data from traditional lenders, FICO credit scores	Logit	0.689
Johnson, Ben-David, Lee, and Yao (2023)	US	FinTech loan	100%	Income, requested loan amount, loan purpose, credit bureau data, FICO credit score	Logit	0.665

Citation	Country	Loan type	% with credit bureau data	Data	Methods	AUC
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Khandani, Kim, and Lo (2010)	US	Credit card	100%	Customer transactions data and account balance data from a major commercial bank, credit bureau data, credit scores	Generalized classification and regression trees	0.952
Lee, Yang, and Anderson (2024)	Multiple countries in Asia	Credit card	50%	Supermarket's loyalty card data and credit card spending and payment history, sociodemographic data, credit scores	XGBoost	0.679 for sample with credit history, 0.647 for sample without credit history
Lee, Yang, and Anderson (2026)	Peru	Credit card	82%	Self reported socioeconomic characteristics, Equifax RP2 scores, credit history and retail transaction data	XGBoost	0.677 for sample with credit history, 0.682 for sample without credit history
Meursault, Moulton, Santucci, and Schor (2025)	US	Bank loan	100%	Credit bureau records	XGBoost, logit	0.883
Netzer, Lemaire, and Herzenstein (2019)	US	FinTech P2P loan	100%	Textual data from loan requests on Prosper, a FinTech P2P lending platform, plus financial and demographic information, Experian credit score	Random forest, logit	0.726
Rishabh (forthcoming)	India	Bank loans and FinTech loan	95% for banks, 90% for FinTech	Payment history data, demographic data, TransUnion credit scores	Random forest, logit	0.7 for banks, 0.68 for FinTech

Citation	Country	Loan type	% with credit bureau data	Data	Methods	AUC
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Sadhwani, Giesecke, and Sirignano (2021)	US	Mortgage	100%	Loan data and monthly performance records, local and national economic data from Zillow and the Federal Housing Administration (FHA), FICO credit scores	Deep learning neural network, logit	0.700
San Pedro, Proserpio, and Oliver (2015)	A Latin American country	Credit card	100%	Mobile phone usage logs from a telecommunications company, digital footprints, sociodemographics, credit bureau data	Regularized logit, support vector machines, gradient boosted trees	0.725

This table reports the country, loan type, percent with credit score, data sources, machine learning methods, and predictive performance (proxied by AUC) of other studies using machine learning models to predict creditworthiness. Appendix F includes additional details of how we extracted the percent with credit bureau data and AUC for various studies, as well as any assumptions we made to do so. AUC = area under the receiver operating characteristic curve; FICO = Fair Isaac Corporation; P2P = peer-to-peer.

Table 2: Summary statistics for modeling sample

	Mean (1)	Std dev (2)	25th perc. (3)	Median (4)	75th perc. (5)
Panel A: Subset of Features					
Woman - dummy	0.39	0.49	0.00	0.00	1.00
User age	25.11	8.01	20.00	23.00	27.00
User iOS (Apple) operating system - dummy	0.38	0.49	0.00	0.00	1.00
No-hit score	640.65	14.00	635.00	642.00	649.00
Number of orders on app	24.05	52.36	3.00	9.00	22.00
Number of orders paid in cash	7.91	14.35	1.00	3.00	9.00
Median amount per order (MXN)	349.84	344.87	173.00	250.00	403.00
Number of orders at supermarkets	1.88	9.25	0.00	0.00	1.00
Number of orders at pharmacies	0.83	3.20	0.00	0.00	0.00
Number of orders at food establishments	18.79	39.16	2.00	7.00	18.00
Marginality (SES) index of census tract	0.96	0.01	0.96	0.96	0.97
Years of schooling among age 15+ in census tract	12.42	1.56	11.71	12.40	13.30
Proportion households own a motor vehicle in census tract	0.64	0.16	0.57	0.64	0.72
Panel B: Credit Card Terms and Use					
Interest rate	0.78	0.06	0.72	0.80	0.80
Credit limit (MXN per month)	5,171.17	1,669.61	5,000.00	5,000.00	5,000.00
Spending (MXN per month)	3,090.40	2,090.25	1,677.12	2,563.88	3,937.17
Statement balance (MXN per month)	3,744.57	2,122.32	2,211.71	3,610.73	4,973.13
Minimum payment (MXN per month)	454.57	627.84	69.64	148.87	479.86
Repayment (MXN per month)	2,834.80	2,407.08	1,133.45	2,246.56	3,838.34
Default - dummy	0.20	0.40	0.00	0.00	0.00
Panel C: Profit Components (MXN per Month)					
Interest revenue	64.31	82.61	0.00	25.83	106.40
BNPL revenue	7.01	13.94	0.00	0.00	7.42
Charge-offs	58.42	142.37	0.00	0.00	0.00
Interest and BNPL revenue net of charge-offs	13.31	175.41	0.00	30.08	107.69
Interchange fee revenue	38.17	36.28	11.41	28.56	53.51
Cost of rewards	35.48	35.79	10.32	26.29	49.51
Interchange fee revenue net of rewards	2.84	15.92	-2.28	2.21	8.87
Late payment fee revenue	13.04	30.55	0.00	0.00	5.21
Funding costs	30.50	15.59	19.25	30.40	41.08
Costs from fraudulent transactions	0.53	9.72	0.00	0.00	0.00
Other fee revenue	2.69	8.00	-0.01	0.00	0.00
Other costs	12.32	13.03	3.46	8.58	16.38
Positive profits - dummy	0.51	0.50	0.00	1.00	1.00

This table shows summary statistics for the sample that we use in our machine learning modeling, which consists of applicants with no credit history and no prior credit card who were approved by RappiCard, had at least twelve months with the card by the end of our data period in May 2024, and made at least one transaction during their first 12 months with the card. Observations are at the user level. Panel A shows a subset of the features used by our machine learning models. For non-binary variables, upper-tail winsorization was performed at the 99.9th percentile consistent with the way features were winsorized for modeling. Census tract for each user is inferred based on login activity on the delivery app. The marginality (SES) index is a summary measure of economic vulnerability at the census track level, which takes values between 0 (high marginality) and 1 (low marginality). Panel B shows credit card terms and use. Interest rate and credit limit are measured at origination. Spending, statement balance, minimum payment, and repayment are averages over the first 12 monthly statements (measured in MXN per month). Default is measured as at least 60 days delinquent at any point over the first 12 months since origination. Panel C shows average monthly revenues obtained and costs incurred by the lender for each card. Borrower-level monthly averages are computed over the first 12 months since card origination. For panels B and C, non-binary variables are winsorized at the 1st and 99th percentiles, with the exception of "costs from fraudulent transactions," as positive values occur for this variable for less than 1% of users; borrower-level profits are computed using non-winsorized revenue and cost variables. For all variables we use the full modeling sample (N = 146,036), except for spending, statement balance, minimum payment, and repayment, which are only available for 143,203 users. Std. dev. = standard deviation; perc. = percentile; iOS = Apple device operating system; MXN = Mexican pesos; SES = socioeconomic status; BNPL = buy now pay later.

Table 3: Performance metrics and marginal contribution of each data source

Data sources used	Total profits (normalized) (1)	AUC (2)	Precision (3)	Recall (4)	F1 (5)	Approval rate (6)
Panel A: Default Model						
All	0.958	0.791	0.425	0.644	0.512	0.694
	[0.900, 1.016]	[0.784, 0.797]	[0.415, 0.435]	[0.632, 0.656]	[0.502, 0.522]	[0.689, 0.700]
All, but transactions	0.300	0.654	0.271	0.678	0.387	0.495
	[0.214, 0.381]	[0.647, 0.662]	[0.264, 0.278]	[0.666, 0.690]	[0.379, 0.396]	[0.489, 0.500]
All, but digital footprint	0.875	0.777	0.382	0.692	0.492	0.634
	[0.817, 0.935]	[0.770, 0.783]	[0.373, 0.392]	[0.680, 0.704]	[0.483, 0.502]	[0.629, 0.640]
All, but no-hit score	0.929	0.786	0.402	0.677	0.504	0.660
	[0.871, 0.989]	[0.779, 0.792]	[0.392, 0.411]	[0.665, 0.689]	[0.495, 0.513]	[0.654, 0.665]
All, but socioeconomic	0.992	0.791	0.421	0.668	0.516	0.679
	[0.934, 1.051]	[0.784, 0.797]	[0.411, 0.431]	[0.656, 0.680]	[0.506, 0.526]	[0.674, 0.685]
Data sources used	Total profits (normalized) (1)	R^2 (2)	MSE (3)	RMSE (4)	MAE (5)	Approval rate (6)
Panel B: Profits Model						
All	1.000	0.085	30,072	173	121	0.610
	[1.000, 1.000]	[0.078, 0.093]	[29,476, 30,685]	[172, 175]	[120, 123]	[0.605, 0.616]
All, but transactions	0.396	0.022	32,168	179	121	0.415
	[0.315, 0.473]	[0.018, 0.026]	[31,545, 32,814]	[178, 181]	[120, 123]	[0.410, 0.421]
All, but digital footprint	0.922	0.072	30,518	175	122	0.555
	[0.877, 0.968]	[0.065, 0.079]	[29,917, 31,127]	[173, 176]	[120, 123]	[0.549, 0.560]
All, but no-hit score	0.969	0.081	30,221	174	121	0.536
	[0.927, 1.014]	[0.074, 0.089]	[29,613, 30,817]	[172, 176]	[120, 123]	[0.530, 0.541]
All, but socioeconomic	1.013	0.085	30,078	173	121	0.589
	[0.978, 1.050]	[0.078, 0.093]	[29,473, 30,677]	[172, 175]	[120, 122]	[0.584, 0.595]

This table reports out-of-sample performance metrics for models trained to predict default or profits using all features or using features from all but one data source. Default is measured as at least 60 days delinquent at any point over the first 12 months since origination. Profits correspond to the average of monthly profits over the first 12 months since origination. Total profits (normalized) are the total profits from using each model, normalized by total profits from the profits model that uses all features. For the default model we also report AUC, precision, recall, and F1 score. For the profits model we also report R^2, MSE, RMSE, and MAE. Threshold-dependent measures use profit-maximizing thresholds determined separately for each model. Results use N = 146,036 users, split into training data to train the machine learning models and testing data to calculate performance metrics. Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples; normalizations of total profits are computed within bootstrap sample, which explains why the confidence interval for the profits model using all data sources is [1.000, 1.000]. AUC = area under the receiver operating characteristic curve. F1 score = the harmonic mean of precision and recall. MSE = mean squared error. RMSE = root mean squared error. MAE = mean absolute error.

Table 4: Percent of observations by default status, profitability, and model recommendation

	(1)	(2)	(3)	(4)
Panel A: Observed Default by Model Approval Recommendation				
	Default model		Profits model	
	Did not default	Defaulted	Did not default	Defaulted
Approved	62.2%	7.2%	54.9%	6.1%
Rejected	17.6%	13.0%	24.9%	14.1%
Panel B: Observed Profitability (Binary) by Model Approval Recommendation				
	Default model		Profits model	
	Profits > 0	Profits ≤ 0	Profits > 0	Profits ≤ 0
Approved	36.6%	32.9%	32.6%	28.4%
Rejected	14.7%	15.9%	18.6%	20.4%
Panel C: Matrix of Model Disagreement				
	Default model			
Profits model	Approved	Rejected		
Approved	58.2%	2.9%		
Rejected	11.3%	27.7%		

This table reports observed default and profitability by model approval recommendation, as well as agreement and disagreement between the default and profits models in the testing sample. Panel A shows the percentage of observations approved or rejected by each model that defaulted. Default is defined as at least 60 days delinquent at any point during the first 12 months since origination. Panel B shows the percentage approved or rejected by each model that generated positive or non-positive average monthly profits over the first 12 months since origination. Panel C reports the fraction of observations approved or rejected by each model. The results use N = 146,036 users, split into training data to train the machine learning models and testing data to evaluate default and profitability by approval recommendation. Approval recommendations of each model are based on profit-maximizing thresholds of a 25% predicted probability of default and -7 MXN predicted profits. The target variable of the default model is binary. The target variable of the profits model is continuous.

Table 5: Model performance by quintile of number of transactions through delivery platform

Quintile	Number of transactions (1)	Total profits (normalized) (2)	AUC (3)	Precision (4)	Recall (5)	F1 (6)
Panel A: Default Model						
1	2 or fewer	0.454	0.742	0.385	0.663	0.487
		[0.270, 0.636]	[0.728, 0.757]	[0.366, 0.405]	[0.637, 0.688]	[0.467, 0.506]
2	3-6	0.674	0.771	0.371	0.739	0.494
		[0.509, 0.840]	[0.757, 0.785]	[0.353, 0.390]	[0.715, 0.762]	[0.475, 0.513]
3	7-13	0.657	0.785	0.377	0.713	0.493
		[0.499, 0.816]	[0.770, 0.799]	[0.356, 0.397]	[0.686, 0.738]	[0.471, 0.513]
4	14-28	0.823	0.791	0.425	0.639	0.510
		[0.657, 0.987]	[0.776, 0.806]	[0.400, 0.450]	[0.609, 0.668]	[0.486, 0.533]
5	29 or more	1.000	0.828	0.526	0.539	0.533
		[0.787, 1.212]	[0.815, 0.842]	[0.496, 0.556]	[0.509, 0.570]	[0.506, 0.559]
Quintile	Number of transactions (1)	Total profits (normalized) (2)	R^2 (3)	MSE (4)	RMSE (5)	MAE (6)
Panel B: Profits Model						
1	2 or fewer	0.557	0.065	32,915	181	126
		[0.390, 0.725]	[0.052, 0.080]	[31,497, 34,355]	[178, 185]	[123, 129]
2	3-6	0.684	0.070	30,651	175	122
		[0.519, 0.845]	[0.056, 0.085]	[29,412, 31,946]	[171, 179]	[119, 125]
3	7-13	0.695	0.081	29,179	171	119
		[0.528, 0.855]	[0.065, 0.100]	[27,873, 30,473]	[167, 175]	[116, 122]
4	14-28	0.769	0.080	27,527	166	114
		[0.610, 0.926]	[0.062, 0.099]	[26,172, 28,911]	[162, 170]	[111, 117]
5	29 or more	0.962	0.078	31,360	177	125
		[0.774, 1.151]	[0.060, 0.096]	[29,940, 32,826]	[173, 181]	[122, 128]

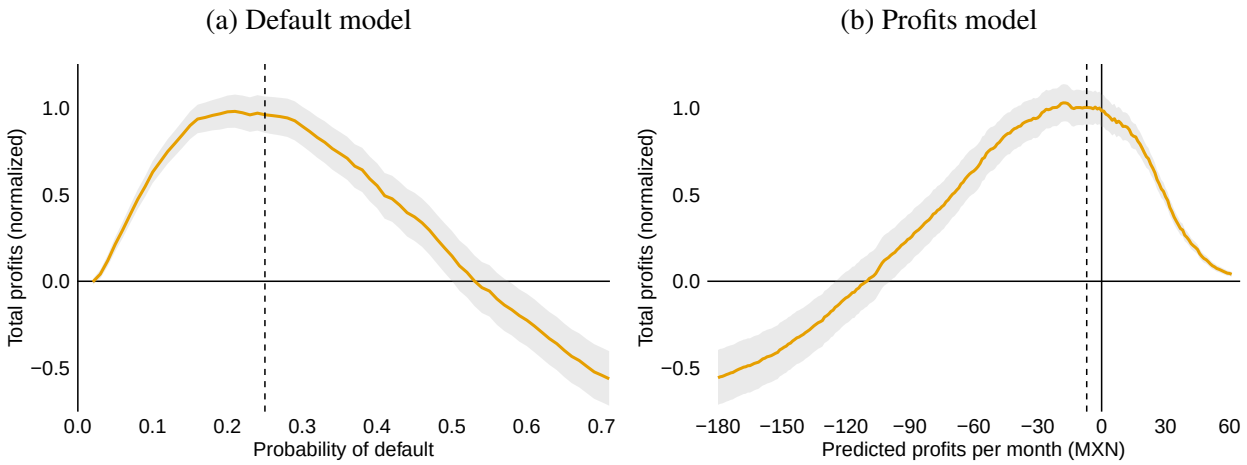
This table reports out-of-sample performance metrics for separate models estimated for each quintile of the distribution of number of transactions made through the delivery platform. Data are split into quintiles of the full modeling sample; machine learning models are then trained on the training data for each quintile and performance metrics are calculated on the testing data for each quintile. The results use N = 146,036 users, split into five quintiles based on number of transactions through the delivery platform, then into training data to train the machine learning model for that quintile and testing data to calculate out-of-sample performance metrics for that quintile. Default is measured as at least 60 days delinquent at any point over the first 12 months since origination. Profits correspond to the average of monthly profits over the first 12 months since origination. Total profits (normalized) are the total profits from using each model, normalized by total profits from the most profitable group and model, which is the profits model for quintile 5. Threshold-dependent measures use profit-maximizing thresholds determined separately for each model. Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples; normalizations of total profits are computed within bootstrap sample. AUC = area under the receiver operating characteristic curve. F1 score = the harmonic mean of precision and recall. MSE = mean squared error. RMSE = root mean squared error. MAE = mean absolute error.

Table 6: Performance and contribution of each data source for models predicting profit components

	(a) Interest revenue		(b) BNPL revenue		(c) Charge-offs	
	Total profits	R^2	Total profits	R^2	Total profits	R^2
	(1)	(2)	(3)	(4)	(5)	(6)
All	0.117	0.054	0.105	0.069	1.003	0.162
	[0.033, 0.196]	[0.048, 0.059]	[0.027, 0.176]	[0.062, 0.077]	[0.950, 1.057]	[0.152, 0.171]
All, but transactions	0.018	0.025	0.000	0.045	0.344	0.046
	[-0.025, 0.059]	[0.022, 0.029]	[0.000, 0.000]	[0.039, 0.052]	[0.257, 0.425]	[0.041, 0.051]
All, but digital footprint	0.101	0.040	0.171	0.042	0.923	0.141
	[0.014, 0.183]	[0.036, 0.045]	[0.086, 0.249]	[0.036, 0.047]	[0.868, 0.982]	[0.131, 0.150]
All, but no-hit score	0.165	0.047	0.068	0.065	0.974	0.153
	[0.081, 0.243]	[0.042, 0.052]	[-0.011, 0.141]	[0.059, 0.072]	[0.921, 1.031]	[0.143, 0.162]
All, but socioeconomic	0.107	0.053	0.087	0.068	1.001	0.163
	[0.017, 0.190]	[0.048, 0.058]	[0.009, 0.160]	[0.060, 0.076]	[0.946, 1.058]	[0.153, 0.173]
	(d) Interchange net of rewards		(e) Late payment fee		(f) Funding costs	
	Total profits	R^2	Total profits	R^2	Total profits	R^2
	(1)	(2)	(3)	(4)	(5)	(6)
All	0.134	0.198	0.142	0.036	0.000	0.150
	[0.078, 0.187]	[0.186, 0.211]	[0.104, 0.180]	[0.032, 0.041]	[0.000, 0.000]	[0.142, 0.158]
All, but transactions	0.152	0.121	0.000	0.009	-0.001	0.106
	[0.090, 0.213]	[0.113, 0.128]	[0.000, 0.000]	[0.007, 0.011]	[-0.004, 0.001]	[0.099, 0.114]
All, but digital footprint	0.109	0.166	0.231	0.032	0.000	0.087
	[0.055, 0.161]	[0.153, 0.179]	[0.177, 0.285]	[0.028, 0.037]	[0.000, 0.000]	[0.081, 0.093]
All, but no-hit score	0.116	0.196	0.137	0.035	0.000	0.128
	[0.067, 0.164]	[0.183, 0.208]	[0.099, 0.175]	[0.030, 0.039]	[0.000, 0.000]	[0.120, 0.135]
All, but socioeconomic	0.142	0.198	0.261	0.036	0.000	0.150
	[0.087, 0.196]	[0.185, 0.211]	[0.208, 0.314]	[0.032, 0.040]	[0.000, 0.000]	[0.142, 0.158]

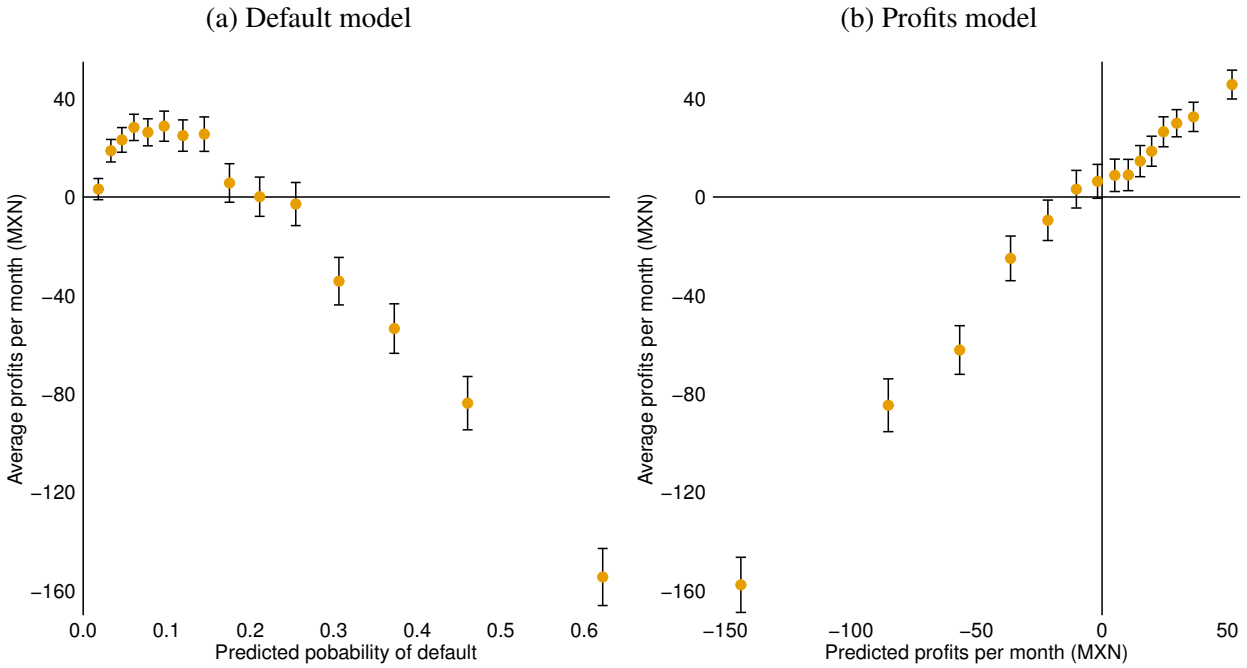
This table reports out-of-sample performance metrics for models trained to predict different profit components using all features or using features from all but one data source. Total profits are the total profits from using each model, normalized by total profits from the main profits model. Results use N = 146,036 users, split into training data to train the machine learning models and testing data to calculate out-of-sample total profits and R^2. For each model, we determine the profit-maximizing threshold for the corresponding profit component in the training sample and apply it to predicted values on the testing sample. Users in the testing sample with predicted values above the threshold for the corresponding profit component are approved (except for late payment fee revenue, for which portfolio profits are maximized when users with predicted value below the threshold are approved). Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples; normalizations of total profits are computed within bootstrap sample.

Figure 1: Profits generated by default and profits models across approval thresholds



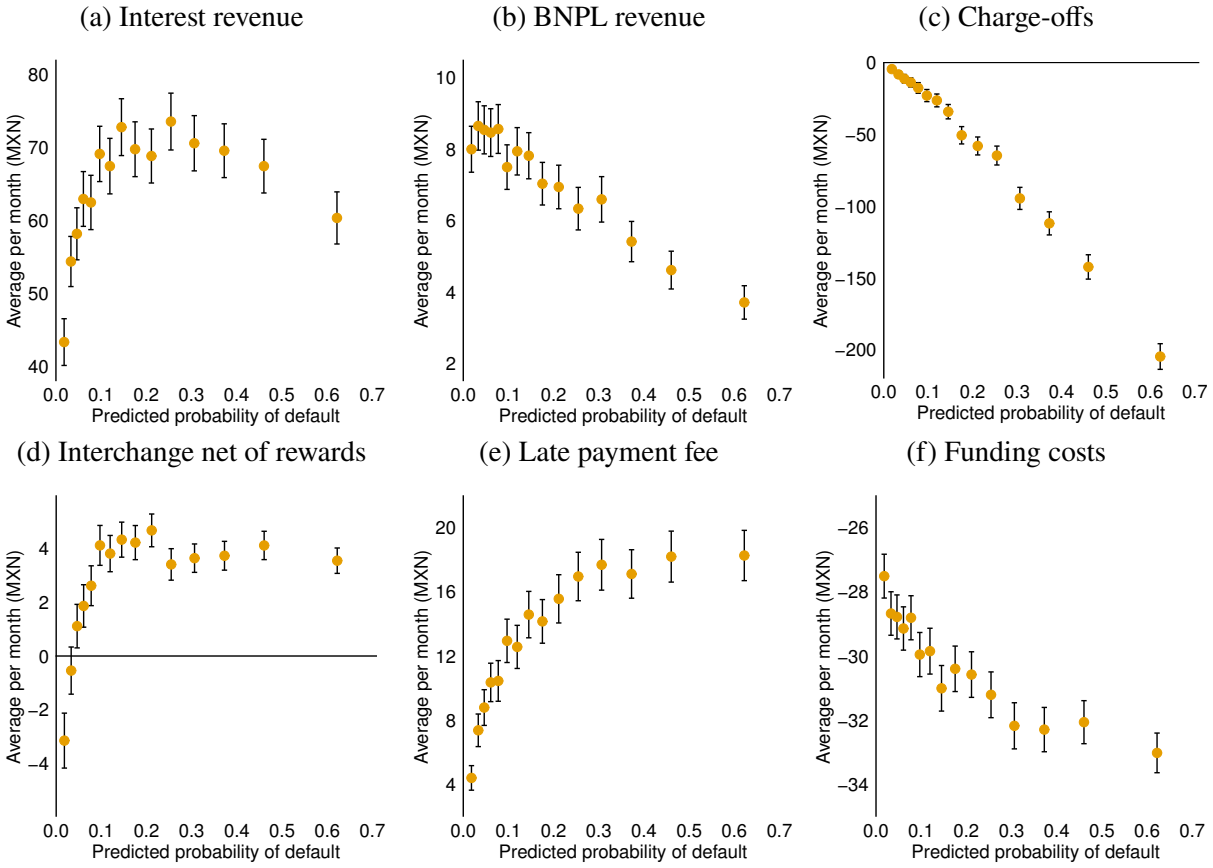
This figure shows the total profits obtained by the default and profits models when making approval decisions in the testing data for different approval thresholds. It is calculated by summing the profits obtained by the lender across all individuals in the testing data who would be approved based on a given approval threshold, i.e., whose predicted probability of default is below the threshold (default model) or whose predicted profit is above the threshold (profits model). The results use N = 146,036 users, randomly split into training and testing data. Profit-maximizing thresholds, estimated in the training data, are shown as dashed vertical lines (these do not necessarily correspond to the peak of the curves shown in the figures since the curves are based on lending decisions made in the testing data, while the profit-maximizing thresholds are determined in the training data to avoid overfitting). Total profits (normalized) are the total profits from a model using a given threshold, normalized by the total profits of the most profitable model across the default and profits models at the profit-maximizing threshold identified in the training data. Bootstrapped 95% confidence intervals are shaded in gray, using 10,000 bootstrap samples. Approval thresholds below the 1st percentile and above the 99th percentile of predicted probabilities or predicted profits are omitted from the graph for legibility. MXN = Mexican pesos.

Figure 2: Average realized profits by predicted probabilities



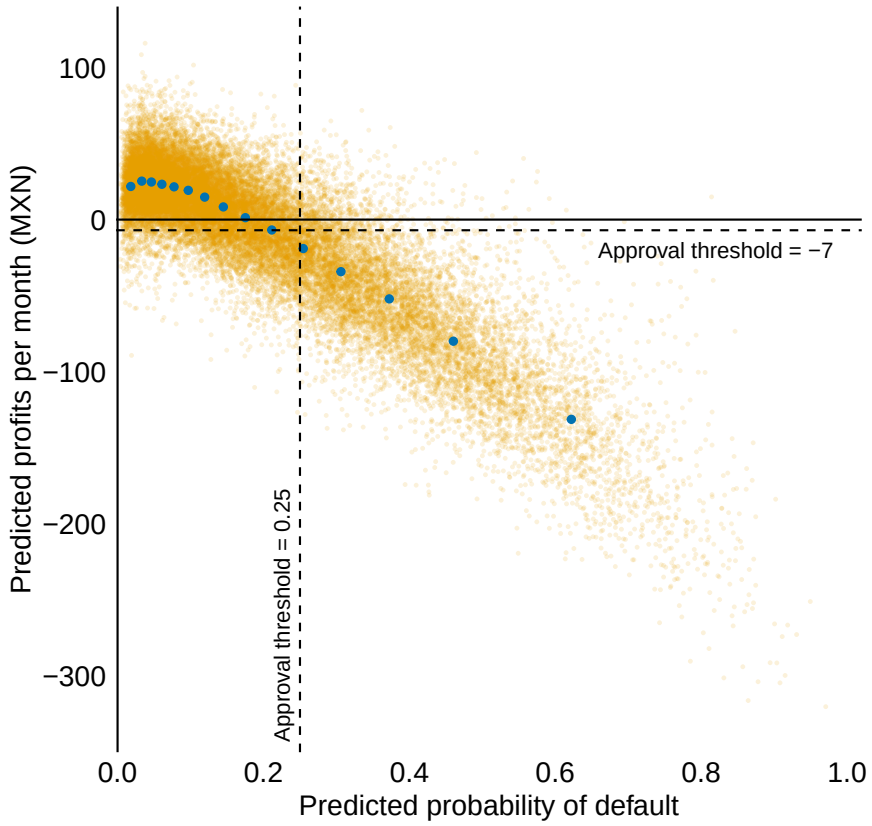
This figure plots a binscatter of average realized profits in the testing sample across 15 bins in the predicted probability of default or predicted profits per month, with approximately equal number of observations in each bin. The results use N = 146,036 users, randomly split into training and testing data. Average profits per month are constructed by summing non-winsorized average monthly revenues and costs variables at the account level, then winsorizing the account-level profits variable at the 1st and 99th percentiles. Whiskers represent 95% confidence intervals. MXN = Mexican pesos.

Figure 3: Profit components by predicted probability of default



This figure shows the average realized value per month of selected profit components in the testing sample across 15 bins in the predicted probability of default, with approximately equal number of observations in each bin. Monthly values are averaged over the first 12 months since account origination and then winsorized at 1st and 99th percentiles. Whiskers represent 95% confidence intervals. MXN = Mexican pesos.

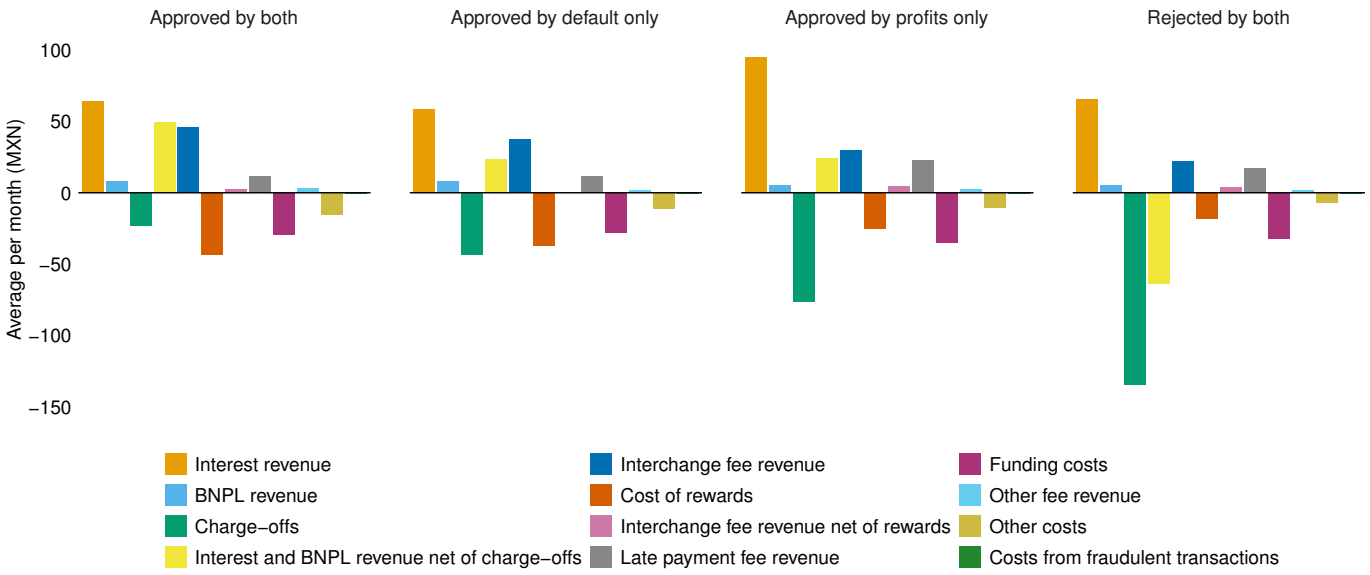
Figure 4: Predicted target variable in default and profits models



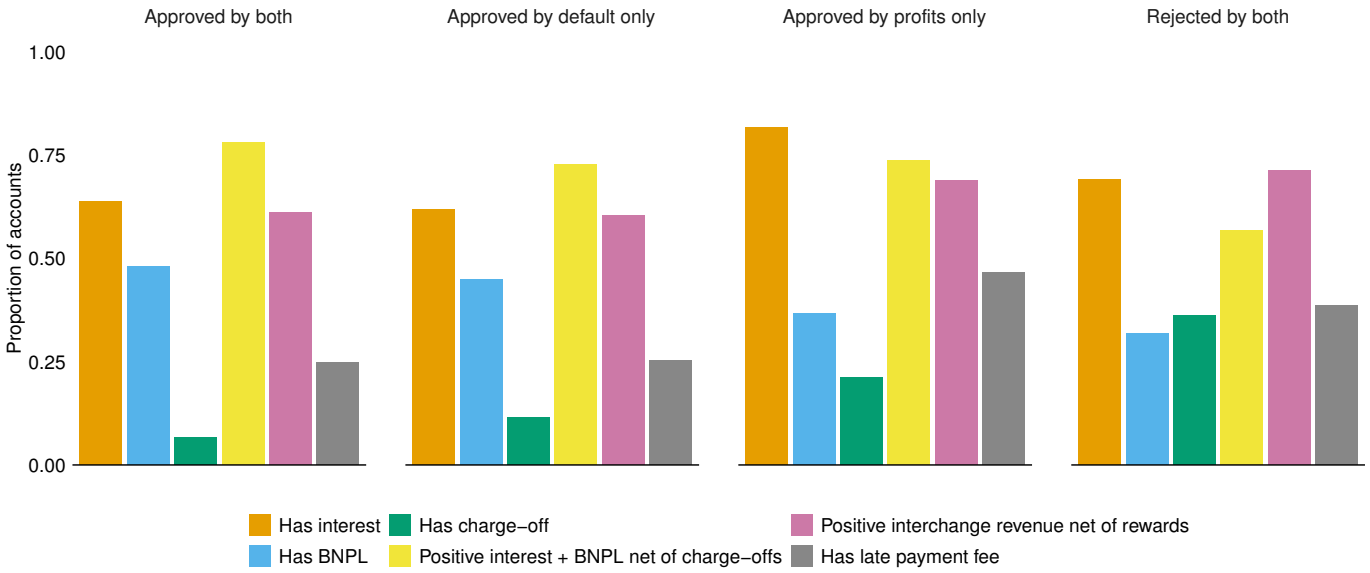
This figure shows the predicted probabilities of default and predicted profits per month for each observation in our out-of-sample testing data (opaque orange dots). It also shows the averages of these two measures within 15 bins of predicted probability of default, each containing an approximately equal number of observations (blue dots). The results use N = 146,036 users, split into training data to train the machine learning models and testing data to calculate out-of-sample predictions which are shown in the figure. Fixing the approval threshold to the profit-maximizing thresholds (25% predicted probability of default and -7 MXN predicted profits), the upper-left quadrant shows applicants who would be approved by both models, the lower-right quadrant shows applicants who would be rejected by both models, the lower-left quadrant shows applicants who would be rejected by the profits model but approved by the default model, and the upper-right quadrant shows applicants who would be rejected by the default model but approved by the profits model.

Figure 5: Revenues, costs, and profits by model disagreement

(a) Continuous measures of revenues, costs, and profits



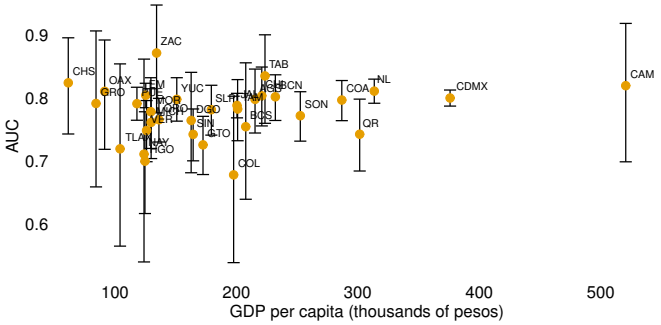
(b) Binary measures



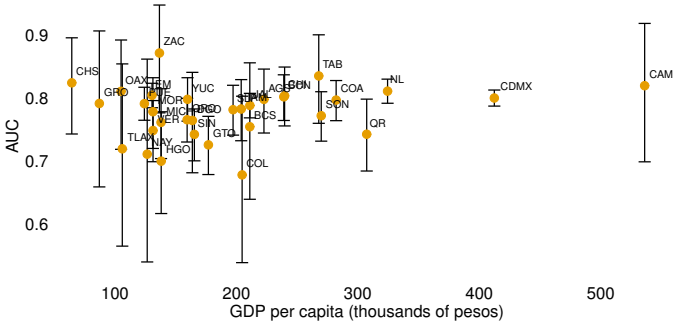
This figure shows the components of revenues, costs, and profits by model disagreement between the default and profits models for the testing sample: accepted by both models, accepted by default model only, accepted by profits model only, and rejected by both models. Approval decisions are based on profit-maximizing thresholds of 25% predicted probability of default and -7 MXN predicted profits. The results use N = 146,036 users, randomly split into training and testing data. In panel (a), variables are winsorized at the 1st and 99th percentiles, with the exception of "costs from fraudulent transactions," as positive values occur for this variable for less than 1% of users.

Figure 6: State-level AUCs and economic activity

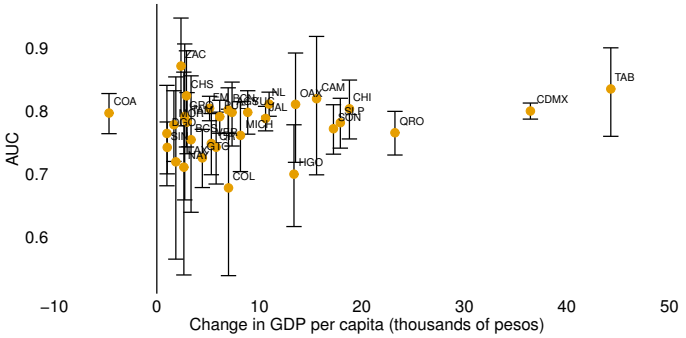
(a) 2021



(b) 2023



(c) 2023 - 2021



This figure shows scatterplots of state-level AUCs from our benchmark default model (vertical axis) against levels (or changes) in GDP per capita for each state (horizontal axis, in thousands of pesos from 2018). Borrowers are assigned to states based on the address provided at the time of credit card application. The results use N = 146,036 users, randomly split into training data and testing data. AUCs are computed on the testing data considering only observations from borrowers in the corresponding state. Predictions come from the default model of Section 4 (not from separate models by state). GDP per capita is calculated by dividing the state-level GDP, published by Mexico's National Institute of Statistics (INEGI), by the population in each state and year, as reported by Mexico's National Population Agency (CONAPO). Panel A uses GDP per capita for 2021. Panel B uses GDP per capita for 2023. Panel C uses the change in GDP per capita between 2023 and 2021. Whiskers represent 95% bootstrapped confidence intervals for AUCs, obtained with 10,000 bootstrap samples. Labels correspond to state names.

References

Agarwal, Sumit, Shashwat Alok, Pulak Ghosh, and Sudip Gupta (2023). "Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech."

Agarwal, Sumit, Souphala Chomsisengphet, Neale Mahoney, and Johannes Stroebel (2015). "Regulating Consumer Financial Products: Evidence from Credit Cards." *The Quarterly Journal of Economics* 130(1), 111–164.

Agarwal, Sumit, Andrea Presbitero, André F. Silva, and Carlo Wix (2023). "Who Pays for Your Rewards? Redistribution in the Credit Card Market." URL: `https://papers.ssrn.com/abstract=4347647` (visited on 05/18/2025). Pre-published.

Albanessi, Stefania and Domonkos Vamossy (2024). "Credit Scores: Performance and Equity."

Alok, Shashwat, Pulak Ghosh, Nirupama Kulkarni, and Manju Puri (2025). "Breaking Barriers to Financial Access: Cross-Platform Digital Payments and Credit Markets." URL: `https://www.nber.org/papers/w33259` (visited on 01/29/2026). Pre-published.

Alvarez, Fernando and David Argente (2022). "On the Effects of the Availability of Means of Payments: The Case of Uber." *The Quarterly Journal of Economics* 137(3), 1737–1789.

Armantier, Olivier, Sebastian Doerr, Jon Frost, Andreas Fuster, and Kelly Shue (2024). "Nothing to Hide? Gender and Age Differences in Willingness to Share Data." *SSRN Electronic Journal*.

Aydin, Deniz (2022). "Consumption Response to Credit Expansions: Evidence from Experimental Assignment of 45,307 Credit Lines." *American Economic Review* 112(1), 1–40.

Babina, Tania, Saleem Bahaj, Greg Buchak, Filippo De Marco, Angus Foulis, Will Gornall, Francesco Mazzola, and Tong Yu (2025). "Customer Data Access and Fintech Entry: Early Evidence from Open Banking." *Journal of Financial Economics* 169, 103950.

Banco de México (2021). "Adopción de Criterios Contables IFRS9."

Banco de México (2024). "Indicadores Basicos de Tarjetas de Credito."

Basten, Christoph and Ragnar Juelsrud (2023). "Cross-Selling in Bank-Household Relationships: Mechanisms and Implications for Pricing." *Review of Financial Studies*, 1–34.

Ben-David, Itzhak, Mark J. Johnson, and René M. Stulz (2025). "Models Behaving Badly: The Limits of Data-Driven Lending." *Review of Finance* 29(3), 711–745.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri (2020). "On the Rise of FinTechs: Credit Scoring Using Digital Footprints." *The Review of Financial Studies* 33(7), 2845–2897.

Berg, Tobias, Valentin Burg, Jan Keil, and Manju Puri (2025). "The Economics of "Buy Now, Pay Later": A Merchant's Perspective." *Journal of Financial Economics* 171, 104093.

Berg, Tobias, Andreas Fuster, and Manju Puri (2022). "FinTech Lending." *Annual Review of Financial Economics* 14(1), 187–207.

Bergstra, James, Dan Yamins, and David Cox (2013). "Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms." Python in Science Conference. Austin, Texas, 13–19.

Bertrand, Marianne and Emir Kamenica (2023). "Coming Apart? Cultural Distances in the United States over Time." *American Economic Journal: Applied Economics* 15(4), 100–141.

Björkegren, Daniel and Darrell Grissen (2020). "Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment." *The World Bank Economic Review* 34(3), 618–634.

Blattner, Laura and Scott Nelson (2024). "How Costly Is Noise? Data and Disparities in Consumer Credit."

Blattner, Laura, Scott Nelson, and Jann Spiess (2024). "Unpacking the Black Box: Regulating Algorithmic Decisions." *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC '22: The 23rd ACM Conference on Economics and Computation. Boulder CO USA: ACM, 559–559.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru (2018). "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks." *Journal of Financial Economics* 130(3), 453–483.

Burlando, Alfredo, Michael A. Kuhn, and Silvia Prina (2025). "Too Fast, Too Furious? Digital Credit Delivery Speed and Repayment Rates." *Journal of Development Economics* 174, 103427.

Butaru, Florentin, Qingqing Chen, Brian Clark, Sanmay Das, Andrew W. Lo, and Akhtar Siddique (2016). "Risk and Risk Management in the Credit Card Industry." *Journal of Banking & Finance* 72, 218–239.

Caire, Dean and Maria Fernandez Vidal (2024). "Leveraging Transactional Data for Micro and Small Enterprise (MSE) Lending."

Caro, Spencer, Talia Gillis, and Scott Nelson (forthcoming). "Differential Validity in Fair Lending." *Journal of Legal Analysis*.

Castellanos, Sara, Diego Jiménez Hernández, Aprajit Mahajan, Eduardo Alcaraz Prous, and Enrique Seira (2025). "Contract Terms, Employment Shocks, and Default in Credit Cards." *Review of Economic Studies*. 0th ser., 1–39.

Chen, Sharon, Sebastian Doerr, Jon Frost, Leonardo Gambacorta, and Hyun Song Shin (2021). "The Fintech Gender Gap."

Chen, Tianqi and Carlos Guestrin (2016). "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 785–794.

Cho, SungJin and John Rust (2017). "Precommitments for Financial Self-Control? Micro Evidence from the 2003 Korean Credit Crisis." *Journal of Political Economy* 125(5), 1413–1464.

Choi, Darwin, Zhenyu Gao, Sing-Sen Lam, Tian Li, and Wenlan Qian (2025). "Scared Away: Credit Demand Response to Expected Motherhood Penalty in the Labor Market." URL: https://papers.ssrn.com/abstract=4539887 (visited on 05/19/2025). Pre-published.

CNBV (2019). "Boletin Trimestral de Inclusion Financiera."

CNBV (2023). "Panorama Anual de Inclusion Financiera."

Consumer Financial Protection Bureau (2017). "Request for Information Regarding Use of Alternative Data and Modeling Techniques in the Credit Process." URL: https://www.consumerfinance.gov/rules-policy/notice-opportunities-comment/archive-closed/request-information-regarding-use-alternative-data-and-modeling-techniques-credit-process/.

Consumer Financial Protection Bureau (2019). "Interagency Statement on the Use of Alternative Data in Credit Underwriting."

Cookson, J. Anthony, Benedict Guttman-Kenney, and William Mullins (2025). "Immigration and Credit in America." URL: https://www.ssrn.com/abstract=5187062 (visited on 05/22/2025). Pre-published.

Cowan, Greig, Salvatore Mercuri, and Raad Khraishi (2023). "Modelling Customer Lifetime-Value in the Retail Banking Industry." URL: http://arxiv.org/abs/2304.03038 (visited on 05/25/2025). Pre-published.

CRIF (2018). "CRIF Desarrolla Nuevo Score No Hit Para Buró de Crédito En México."

Crook, Jonathan and John Banasik (2004). "Does Reject Inference Really Improve the Performance of Application Scoring Models?" *Journal of Banking and Finance*.

Crouzet, Nicolas, Pulak Ghosh, Apoorv Gupta, and Filippo Mezzanotti (2024). "Demographics and Technology Diffusion: Evidence from Mobile Payments." *SSRN Electronic Journal*.

Davis, Jesse and Mark Goadrich (2006). "The Relationship between Precision-Recall and ROC Curves." *Proceedings of the 23rd International Conference on Machine Learning - ICML '06*. The 23rd International Conference. Pittsburgh, Pennsylvania: ACM Press, 233–240.

De Cnudde, Sofie, Julie Moeyersoms, Marija Stankova, Ellen Tobback, Vinayak Javaly, and David Martens (2019). "What Does Your Facebook Profile Reveal about Your Creditworthiness? Using Alternative Data for Microfinance." *Journal of the Operational Research Society* 70(3), 353–363.

Demirgüç-Kunt, Asli, Leora Klapper, Dorothe Singer, and Saniya Ansar (2022). *Financial Inclusion, Digital Payments, and Resilience in the Age of COVID-19*. Washington, D.C.: World Bank Group.

Department of Commerce (2023). "Mexico - Financial Technologies (Fintech) Industry."

Di Maggio, Marco and Dimuthu Ratnadiwakara (2025). "Invisible Primes: Fintech Lending with Alternative Data." SSRN Scholarly Paper 3937438.

Di Maggio, Marco, Emily Williams, and Justin Katz (2022). "Buy Now, Pay Later Credit: User Characteristics and Effects on Spending Patterns." URL: https://www.nber.org/papers/w30508 (visited on 05/18/2025). Pre-published.

Di Maggio, Marco and Vincent Yao (2021). "Fintech Borrowers: Lax Screening or Cream-Skimming?" *Review of Financial Studies* 34(10), 4565–4618.

Doerr, Sebastian, Jon Frost, Leonardo Gambacorta, and Han Qiu (2022). "Population Ageing and the Digital Divide." *SUERF Policy Brief*.

Drechsler, Itamar, Hyeyoon Jung, Weiyu Peng, Dominik Supera, and Guanyu Zhou (2025). "Credit Card Banking." Staff Reports (Federal Reserve Bank of New York). Federal Reserve Bank of New York.

Duarte, Victor, Julia Fonseca, Divij Kohli, and Julian Reif (2025). "The Effects of Deleting Medical Debt from Consumer Credit Reports."

Ekinci, Yeliz, Nimet Uray, and Füsun Ülengin (2014). "A Customer Lifetime Value Model for the Banking Industry: A Guide to Marketing Actions." *European Journal of Marketing* 48(3/4), 761–784.

El Economista (2024). "Datos Alternativos Cada Vez Mas Presentes Para Otorgar Un Credito." URL: https://www.eleconomista.com.mx/sectorfinanciero/Datos-alternativos-cada-vez-mas-presentes-para-otorgar-un-credito-20240104-0070.html%7D.

Finnovista (2023). "Fintech Radar Mexico 2023."

FinRegLab (2023a). "Explainability and Fairness in Machine Learning for Credit Underwriting." URL: https://finreglab.org/research/explainability-fairness-in-machine-learning-for-credit-underwriting-policy-analysis/.

FinRegLab (2023b). "Machine Learning Explainability & Fairness: Insights from Consumer Lending."

Flach, Peter and Meelis Kull (2015). "Precision-Recall-Gain Curves: PR Analysis Done Right." *Advances in Neural Information Processing Systems*. Vol. 28. Curran Associates, Inc.

Frost, Jon, Leonardo Gambacorta, Yi Huang, Hyun Song Shin, and Pablo Zbinden (2019). "BigTech and the Changing Structure of Financial Intermediation." *Economic Policy* 34(100), 761–799.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2022). "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *The Journal of Finance* 77(1), 5–47.

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

Gambacorta, Leonardo, Yiping Huang, Han Qiu, and Jingyi Wang (2024). "How Do Machine Learning and Non-Traditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm." *Journal of Financial Stability* 73, 101284.

GAO (2009). "Credit Cards." URL: https://www.gao.gov/assets/gao-10-45.pdf.

Gertler, Paul, Sean Higgins, Ulrike Malmendier, and Waldo Ojeda (2025). "Do Behavioral Frictions Prevent Firms from Adopting Profitable Opportunities?" URL: https://www.nber.org/papers/w33387 (visited on 04/21/2025). Pre-published.

Ghosh, Pulak, Boris Vallee, and Yao Zeng (forthcoming). "FinTech Lending and Cashless Payments." *Journal of Finance* n/a(n/a).

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

Guttman-Kenney, Benedict, Chris Firth, and John Gathergood (2023). "Buy Now, Pay Later (BNPL) ...on Your Credit Card." *Journal of Behavioral and Experimental Finance* 37, 100788.

Guttman-Kenney, Benedict and Andres Shahidinejad (2023). "Unraveling Information Sharing in Consumer Credit Markets." *SSRN Electronic Journal*.

Hair, Christopher M, Sabrina T Howell, Mark J Johnson, and Siena Matsumoto (2025). "Modernizing Access to Credit for Younger Entrepreneurs: From FICO to Cash Flow."

He, Zhiguo, Jing Huang, and Jidong Zhou (2023). "Open Banking: Credit Market Competition When Borrowers Own the Data." *Journal of Financial Economics* 147(2), 449–474.

Higgins, Sean (2024). "Financial Technology Adoption: Network Externalities of Cashless Payments in Mexico." *American Economic Review* 114(11), 3469–3512.

Huang, Yiping, Zhenhua Li, Han Qiu, Sun Tao, Xue Wang, and Longmei Zhang (2023). "BigTech Credit Risk Assessment for SMEs." *China Economic Review* 81, 102016.

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

Jagtiani, Julapa and Catharine Lemieux (2019). "The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform." *Financial Management* 48(4), 1009–1029.

Jiang, Erica Xuewei, Gloria Yang Yu, and Jinyuan Zhang (forthcoming). "Bank Competition Amid Digital Disruption: Implications for Financial Inclusion." *Journal of Finance*.

Johnson, Mark J., Itzhak Ben-David, Jason Lee, and Vincent Yao (2023). "FinTech Lending with LowTech Pricing." URL: https://papers.ssrn.com/abstract=4396502 (visited on 03/12/2024). Pre-published.

Karlan, Dean and Jonathan Zinman (2009). "Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment." *Econometrica* 77(6), 1993–2008.

Khandani, Amir E., Adlar J. Kim, and Andrew W. Lo (2010). "Consumer Credit-Risk Models via Machine-Learning Algorithms." *Journal of Banking & Finance* 34(11), 2767–2787.

Krivorotov, George (2023). "Machine Learning-Based Profit Modeling for Credit Card Underwriting - Implications for Credit Risk." *Journal of Banking & Finance* 149, 106785.

Lakkaraju, Himabindu, Jon Kleinberg, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan (2017). "The Selective Labels Problem: Evaluating Algorithmic Predictions in the Presence of Unobservables." *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '17: The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax NS Canada: ACM, 275–284.

Laudenbach, Christine, Elin Molin, Kasper Roszbach, and Talina Sondershaus (2025). "Buy Now Pay (Less) Later: Leveraging Private BNPL Data in Consumer Banking." URL: `https://papers.ssrn.com/abstract=5117651` (visited on 05/20/2025). Pre-published.

Lee, Jung Youn, Joonhyuk Yang, and Eric Anderson (2024). "Using Grocery Data to Predict Credit Card Payments." *Management Science*, 1–25.

Lee, Jung Youn, Joonhyuk Yang, and Eric Anderson (2026). "Who Benefits from Alternative Data for Credit Scoring? Evidence from Peru." *Journal of Marketing Research* 63(1), 105–126.

Li, Shibo, Baohong Sun, and Alan L Montgomery (2011). "Cross-Selling the Right Product to the Right Customer at the Right Time." *Journal of Marketing Research* 48(4), 683–700.

Lima Junior, João Manoel de, Gabriela Borges Silva, José Egidio Altoé Junior, and Ana Paula Ruhe (2021). *Repercussões jurídicas e econômicas do mercado de cartões de crédito*.

Matcham, William (2025). "Risk-Based Borrowing Limits in Credit Card Markets."

Medina, Paolina C. and Jose L. Negrin (2022). "The Hidden Role of Contract Terms: The Case of Credit Card Minimum Payments in Mexico." *Management Science* 68(5), 3856–3877.

Meursault, Vitaly, Daniel Moulton, Larry Santucci, and Nathan Schor (2025). "One Threshold Doesn't Fit All: Tailoring Machine Learning Predictions of Consumer Default for Lower-income Areas." *Journal of Policy Analysis and Management* 44(3), 792–815.

Mishra, Prachi, Nagpurnanand Prabhala, and Raghuram G Rajan (2022). "The Relationship Dilemma: Why Do Banks Differ in the Pace at Which They Adopt New Technology?" *Review of Financial Studies* 35(7), 3418–3466.

Nam, Rachel J. (2022). "Open Banking and Customer Data Sharing: Implications for Fintech Borrowers." *SSRN Electronic Journal*.

Nelson, Scott T. (2025). "Private Information and Price Regulation in the US Credit Card Market." *Econometrica* 93(4), 1371–1410.

Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019). "When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications." *Journal of Marketing Research* 56(6), 960–980.

Petersen, M. A. and R. G. Rajan (1995). "The Effect of Credit Market Competition on Lending Relationships." *The Quarterly Journal of Economics* 110(2), 407–443.

Rajan, Uday, Amit Seru, and Vikrant Vig (2015). "The Failure of Models That Predict Failure: Distance, Incentives, and Defaults." *Journal of Financial Economics* 115(2), 237–260.

Richardson, Eve, Raphael Trevizani, Jason A. Greenbaum, Hannah Carter, Morten Nielsen, and Bjoern Peters (2024). "The Receiver Operating Characteristic Curve Accurately Assesses Imbalanced Datasets." *Patterns* 5(6), 100994.

Rishabh, Kumar (forthcoming). "Beyond the Bureau: Interoperable Payment Data for Loan Screening and Monitoring." *Journal of Financial Intermediation*.

Sadhwani, Apaar, Kay Giesecke, and Justin Sirignano (2021). "Deep Learning for Mortgage Risk*." *Journal of Financial Econometrics* 19(2), 313–368.

San Pedro, Jose, Davide Proserpio, and Nuria Oliver (2015). "MobiScore: Towards Universal Credit Scoring from Mobile Phone Data." *User Modeling, Adaptation and Personalization*. Ed. by Francesco Ricci, Kalina Bontcheva, Owen Conlan, and Séamus Lawless. Cham: Springer International Publishing, 195–207.

Thomas, Lyn C. (2009). *Consumer Credit Models: Pricing, Profit and Portfolios*. 1st edition. Oxford ; New York: Oxford University Press. 385 pp.

Trecone (2023). "La evolución del sector de delivery en México: un vistazo al pasado, presente y futuro." Trecone. URL: https://trecone.com/la-evolucion-del-sector-de-delivery-en-mexico-un-vistazo-al-pasado-presente-y-futuro/.

Wang, Lulu (2025). "Regulating Competing Payment Networks."

Yang, Jinpu (2025). "The Unintended Consequence of Interest Rate Caps on Small Dollar Loans."

Internet Appendix

A Appendix Tables and Figures

Table A.1: Default definition, default rates, and performance metrics for studies that predict creditworthiness

Citation (1)	Default definition (2)	Default rate (3)	AUC (4)	Precision (5)	Recall (6)	F1 (7)	AUC-PR (8)
This paper	60 days or more delinquent at any point during the first 12 months after origination	20%	0.796	0.421	0.666	0.516	0.513
Agarwal, Alok, Ghosh, and Gupta (2023)	Not specified	4%	0.738 for sample with credit history, 0.674 for sample without credit history	0.113 for sample with credit history, 0.115 for sample without credit history	0.348 for sample with credit history, 0.356 for sample without credit history	0.171 for sample with credit history, 0.174 for sample without credit history	Not reported
Albanessi and Vamossy (2024)	90 days or more past due on any debt within 8 quarters	18%	0.906	Not reported	Not reported	Not reported	Not reported
Berg, Burg, Gombović, and Puri (2020)	Unpaid purchase after 3 reminders after 14 days	3%	0.734	Not reported	Not reported	Not reported	Not reported
Björkegren and Grissen (2020)	More than 15 days overdue on the phone bill	11%	0.772	Not reported	Not reported	Not reported	Not reported
Blattner and Nelson (2024)	At least 90 days delinquent on the loan 24 months after the application	8%	0.840 for minority and 0.887 for non-minority sample	0.250	0.540	0.342	Not reported

Citation (1)	Default definition (2)	Default rate (3)	AUC (4)	Precision (5)	Recall (6)	F1 (7)	AUC-PR (8)
Blattner, Nelson, and Spiess (2024)	Default of any severity up to 24 months after origination	14%	0.867	Not reported	Not reported	Not reported	Not reported
Butaru et al. (2016)	90 days delinquent over the next 2, 3 or 4 quarters	From 1.36% to 4.36% (varies by bank)	Not reported	Many reported	Many reported	Many reported	Not reported
Di Maggio and Ratnadiwakara (2025)	90 days delinquent on a loan 12 months after origination	8%	0.659	Not reported	Not reported	Not reported	Not reported
Duarte, Fonseca, Kohli, and Reif (2025)	90 days or more past due within the next 18-24 months	13%	0.712	0.736	0.448	0.557	Not reported
Frost et al. (2019)	Loss rate: volume of outstanding credit that is 30 days or more past due over origination amount	1%	0.764	Not reported	Not reported	Not reported	Not reported
Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2022)	90 days or more delinquent at some point over the first 3 years after origination	1%	0.861	Not reported	Not reported	Not reported	0.062
Gambacorta, Huang, Qiu, and Wang (2024)	Not specified	16%	0.607	Not reported	Not reported	Not reported	Not reported

Citation (1)	Default definition (2)	Default rate (3)	AUC (4)	Precision (5)	Recall (6)	F1 (7)	AUC-PR (8)
Hair, Howell, Johnson, and Matsumoto (2025)	60 days past due, charged-off loan, or the borrower has received forbearance and a modified the loan	17%	0.663	Not reported	Not reported	Not reported	0.266
Huang et al. (2023)	Nonperforming loan ratio	2%	0.841	Not reported	Not reported	Not reported	Not reported
Iyer, Khwaja, Luttmer, and Shue (2016)	3 or more months late as of 3 years after the loan is initiated	31%	0.714	Not reported	Not reported	Not reported	Not reported
Jagtiani and Lemieux (2019)	At least 60 days past due within 24 months after origination	From 5% to 35%	0.689	Not reported	Not reported	Not reported	Not reported
Johnson, Ben-David, Lee, and Yao (2023)	Having a delinquent payment within 1 year since origination	8%	0.665	Not reported	Not reported	Not reported	Not reported
Khandani, Kim, and Lo (2010)	90 days or more delinquent on any credit card account	2%	0.952	0.839	0.688	0.756	Not reported
Lee, Yang, and Anderson (2024)	2 months delinquent	7%	0.679 for sample with credit history, 0.647 for sample without credit history	Not reported	Not reported	Not reported	Not reported

Citation (1)	Default definition (2)	Default rate (3)	AUC (4)	Precision (5)	Recall (6)	F1 (7)	AUC-PR (8)
Lee, Yang, and Anderson (2026)	At least 60 days delinquent on any consumer loan with any lender	7.6% for sample with credit history, 9.8% for sample without credit history	0.677 for sample with credit history, 0.682 for sample without credit history	Not reported	Not reported	Not reported	Not reported
Meursault, Moulton, Santucci, and Schor (2025)	90 or more days past due on at least one of the accounts within two years	22%	0.883	Not reported	0.860	Not reported	Not reported
Netzer, Lemaire, and Herzenstein (2019)	Loan status is "charge-off," "bankruptcy," or "delinquency"	35%	0.726	Not reported	Not reported	Not reported	Not reported
Rishabh (forthcoming)	90 days or more delinquent, write-offs, or lender classifications indicating a loss	9% for banks, 12% for FinTech	0.7 for banks, 0.68 for FinTech	Not reported	Not reported	Not reported	0.200
Sadhwani, Giesecke, and Sirignano (2021)	Transition states	34%	0.700	Not reported	Not reported	Not reported	Not reported
San Pedro, Proserpio, and Oliver (2015)	90 days or more delinquent within 9 months since card activation	13%	0.725	Not reported	Not reported	Not reported	0.292

This table reports the default definition, default rate, and predictive performance (proxied by AUC, precision, recall, F1 and AUC-PR) of other studies using machine learning models to predict creditworthiness. Appendix F includes additional details of how we extracted the default rate and performance metrics for various studies, as well as any assumptions we made to do so. AUC = area under the receiver operating characteristic curve; AUC-PR = area under the precision-recall curve; F1 score = harmonic mean of precision and recall.

Table A.2: Performance of default model with alternative definition of target variable

Model	Total profits (normalized) (1)	AUC (2)	Precision (3)	Recall (4)	F1 (5)
Alternative definition of target variable	0.886 [0.819, 0.952]	0.770 [0.764, 0.776]	0.620 [0.610, 0.629]	0.577 [0.568, 0.587]	0.598 [0.590, 0.606]

This table reports out-of-sample performance metrics for a model trained to predict when an account is at least 60 days delinquent at any point between origination and the end of our data period. The time window over which we measure default varies by user, rather than being fixed across users at the first 12 months since origination. Total profits (normalized) are the total profits from using each model, normalized by total profits from the profits model that uses all features. Threshold-dependent measures use the profit-maximizing threshold of a 44% predicted probability of default. Borrowers with predicted probabilities below this threshold are approved. The results use N = 146,036 users, split into training data to train the machine learning models and testing data to calculate out-of-sample measures of model performance. Total profits from this model are normalized by the total profits generated by the main profits model. Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples; normalizations of total profits are computed within bootstrap sample.

Table A.3: Performance of profitability model using binary target variable

Model	Total profits (normalized) (1)	AUC (2)	Precision (3)	Recall (4)	F1 (5)
Binary profitability model	0.533 [0.451, 0.611]	0.600 [0.594, 0.607]	0.553 [0.545, 0.560]	0.625 [0.617, 0.633]	0.586 [0.580, 0.593]

This table reports out-of-sample performance metrics for a model trained to predict the probability that a cardholder is profitable. The target variable is binary and equals one if average monthly profits over the first 12 months since origination are positive. Total profits (normalized) are the total profits from using the model, normalized by total profits from the profits model that uses all features. Threshold-dependent measures use the profit-maximizing threshold of a 52% predicted probability of positive profits. Borrowers with predicted probabilities above this threshold are approved. The results use N = 146,036 users, split into training data to train the machine learning models and testing data to calculate out-of-sample measures of model performance. Total profits from this model are normalized by the total profits generated by the main profits model (with continuous target variable). Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples; normalizations of total profits are computed within bootstrap sample.

Table A.4: Performance of models predicting default and profits with different data sources

Data sources used	Total profits (normalized) (1)	AUC (2)	Precision (3)	Recall (4)	F1 (5)
Panel A: Default Model					
Only transactions	0.872	0.769	0.370	0.713	0.487
	[0.812, 0.933]	[0.762, 0.775]	[0.361, 0.379]	[0.702, 0.725]	[0.478, 0.496]
Only digital footprint	0.246	0.630	0.239	0.808	0.369
	[0.176, 0.314]	[0.623, 0.638]	[0.233, 0.245]	[0.798, 0.818]	[0.361, 0.376]
Only no-hit score	0.068	0.588	0.220	0.879	0.351
	[0.007, 0.125]	[0.580, 0.596]	[0.214, 0.225]	[0.871, 0.888]	[0.344, 0.358]
Only socioeconomic	0.042	0.548	0.211	0.869	0.339
	[-0.021, 0.102]	[0.540, 0.557]	[0.205, 0.216]	[0.861, 0.878]	[0.332, 0.346]
Data sources used	Total profits (normalized) (1)	R^2 (2)	MSE (3)	RMSE (4)	MAE (5)
Panel B: Profits Model					
Only transactions	0.937	0.065	30,744	175	122
	[0.887, 0.988]	[0.058, 0.072]	[30,141, 31,372]	[174, 177]	[121, 123]
Only digital footprint	0.357	0.016	32,352	180	121
	[0.284, 0.430]	[0.013, 0.019]	[31,706, 33,003]	[178, 182]	[119, 123]
Only no-hit score	0.083	0.007	32,639	181	120
	[0.018, 0.142]	[0.005, 0.010]	[31,974, 33,295]	[179, 182]	[118, 122]
Only socioeconomic	0.052	0.003	32,791	181	120
	[-0.004, 0.107]	[0.002, 0.004]	[32,115, 33,455]	[179, 183]	[119, 122]

This table reports the out-of-sample performance of models trained to predict default or profits using data from only one data source. For each model, we determine the profit-maximizing threshold on the training sample and apply it to predicted values on the testing sample. Total profits (normalized) are the total profits from using the model, normalized by total profits from the profits model that uses all features. For the default model we also report AUC, precision, recall, and F1 score. For the profits model we also report R^2, MSE, RMSE, and MAE. The results use N = 146,036 users, split into training data to train the machine learning models and testing data to calculate out-of-sample performance metrics. Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples; normalizations of total profits are computed within bootstrap sample. AUC = area under the receiver operating characteristic curve. F1 score = the harmonic mean of precision and recall. MSE = mean squared error. RMSE = root mean squared error. MAE = mean absolute error.

Table A.5: Summary statistics by quintile of number of transactions

	Quintile 1 (1)	Quintile 2 (2)	Quintile 3 (3)	Quintile 4 (4)	Quintile 5 (5)
Panel A: Subset of Features					
Woman - dummy	0.37	0.38	0.39	0.40	0.40
User age	26.04	25.06	24.60	24.44	25.35
User iOS (Apple) operating system - dummy	0.28	0.32	0.37	0.42	0.53
No-hit score	640.50	640.18	639.78	640.18	642.62
Number of orders on app	1.19	4.33	9.59	19.67	87.50
Number of orders paid in cash	0.79	2.53	4.88	9.02	23.12
Median amount per order (MXN)	387.26	347.49	332.19	328.66	350.39
Number of orders at supermarkets	0.06	0.19	0.52	1.16	7.63
Number of orders at pharmacies	0.05	0.16	0.39	0.77	2.89
Number of orders at food establishments	0.99	3.65	7.90	15.99	67.03
Marginality (SES) index of census tract	0.96	0.96	0.96	0.96	0.97
Years of schooling among age 15+ in census tract	12.13	12.20	12.35	12.53	12.92
Proportion households own a motor vehicle in census tract	0.62	0.63	0.64	0.65	0.68
Panel B: Credit Card Terms and Use					
Interest rate	0.77	0.78	0.78	0.78	0.78
Credit limit (MXN per month)	5,301.63	5,026.48	5,094.83	5,153.15	5,285.54
Spending (MXN per month)	2,891.58	2,907.60	2,987.08	3,144.04	3,548.69
Statement balance (MXN per month)	3,796.36	3,671.41	3,674.99	3,723.46	3,858.50
Minimum payment (MXN per month)	482.53	462.22	452.79	441.86	430.76
Repayment (MXN per month)	2,456.94	2,570.22	2,690.26	2,913.82	3,582.27
Default - dummy	0.22	0.21	0.21	0.19	0.17
Panel C: Profit Components (MXN per Month)					
Interest revenue	60.27	60.72	62.42	65.31	73.39
BNPL revenue	7.67	7.14	7.15	6.98	6.07
Charge-offs	65.52	61.94	59.48	54.83	49.46
Interest and BNPL revenue net of charge-offs	3.11	6.27	10.47	17.75	30.31
Interchange fee revenue	33.04	34.57	36.44	39.62	47.77
Cost of rewards	29.64	30.70	32.94	36.48	48.34
Interchange fee revenue net of rewards	3.57	3.96	3.56	3.25	-0.24
Late payment fee revenue	12.45	12.13	12.51	12.96	15.26
Funding costs	29.44	29.92	30.18	30.89	32.17
Costs from fraudulent transactions	0.46	0.48	0.56	0.62	0.54
Other fee revenue	1.96	2.29	2.47	2.79	4.00
Other costs	9.97	11.08	11.74	12.92	16.14
Positive profits - dummy	0.51	0.51	0.51	0.51	0.51

This table shows means for the sample that we use in our machine learning modeling, which consists of applicants with no credit history and no prior credit card who were approved by RappiCard, had at least twelve months with the card by the end of our data period in May 2024, and made at least one transaction during their first 12 months with the card. The mean is calculated separately by quintile of number of transactions in the delivery app. Observations are at the user level. Panel A shows summary statistics for a small subset of the features used by our machine learning model. For non-binary variables, upper-tail winsorization was performed at the 99.9th percentile consistent with the way features were winsorized for modeling. Census tract for each user is inferred based on login activity on the delivery app. The marginality (SES) index is a summary measure of economic vulnerability at the census track level, which takes values between 0 (high marginality) and 1 (low marginality). Panel B shows summary statistics on the terms and use of the credit cards. Interest rate and credit limit are measured at origination. Statement balance, minimum payment, and repayment are averages over the first 12 monthly statements (measured in MXN per month). Default is measured as at least 60 days delinquent at any point over the first 12 months since origination. Panel C shows summary statistics for the average monthly revenues obtained and costs incurred by the lender for each card, where borrower-level monthly averages are computed over the first 12 months since card origination. For panels B and C, non-binary variables are winsorized at the 1st and 99th percentiles, with the exception of "costs from fraudulent transactions," as positive values occur for this variable for less than 1% of users; borrower-level profits are computed using non-winsorized revenue and cost variables. For all variables we use the full modeling sample (N = 146,036), except for spending, statement balance, minimum payment, and repayment, which are only available for 143,203 users. Std. dev. = standard deviation; perc. = percentile; iOS = Apple device operating system; MXN = Mexican pesos; SES = socioeconomic status; BNPL = buy now pay later.

Table A.6: Summary statistics, by model disagreement

	Approved by both (1)	Approved by default model only (2)	Approved by profits model only (3)	Rejected by both (4)
Panel A: Subset of Features				
Woman - dummy	0.41	0.33	0.48	0.36
User age	24.54	25.23	25.89	26.10
User iOS (Apple) operating system - dummy	0.44	0.28	0.46	0.28
No-hit score	642.54	642.02	636.68	636.74
Number of orders on app	26.63	27.65	38.73	14.57
Number of orders paid in cash	8.48	5.91	15.40	6.64
Median amount per order (MXN)	353.69	380.88	314.14	336.69
Number of orders at supermarkets	1.87	3.62	2.58	1.13
Number of orders at pharmacies	0.81	1.08	1.95	0.61
Number of orders at food establishments	21.25	19.73	29.48	10.94
Marginality (SES) index of census tract	0.97	0.96	0.96	0.96
Years of schooling among age 15+ in census tract	12.59	12.31	12.44	12.09
Proportion households own a motor vehicle in census tract	0.66	0.62	0.66	0.61
Panel B: Credit Card Terms and Use				
Interest rate	0.78	0.76	0.79	0.77
Credit limit (MXN per month)	5,164.83	5,376.03	4,965.23	5,129.11
Spending (MXN per month)	3,290.76	2,994.88	2,769.31	2,690.89
Statement balance (MXN per month)	3,368.45	3,573.51	4,127.20	4,526.20
Minimum payment (MXN per month)	285.50	374.96	602.11	822.36
Repayment (MXN per month)	3,290.68	2,765.51	2,513.57	1,848.32
Default - dummy	0.09	0.17	0.28	0.44
Panel C: Profit Components (MXN per Month)				
Interest revenue	64.22	58.23	94.87	65.25
BNPL revenue	7.83	8.15	5.38	5.21
Charge-offs	23.01	43.22	76.71	134.84
Interest and BNPL revenue net of charge-offs	49.35	23.49	23.87	-63.68
Interchange fee revenue	45.84	37.43	29.83	21.97
Cost of rewards	43.25	37.07	25.48	18.52
Interchange fee revenue net of rewards	2.66	0.57	4.68	3.71
Late payment fee revenue	11.29	11.78	22.87	17.12
Funding costs	29.74	28.36	35.38	31.97
Costs from fraudulent transactions	0.52	0.49	0.40	0.44
Other fee revenue	3.33	2.01	2.74	1.68
Other costs	15.12	10.87	10.37	6.76
Positive profits - dummy	0.53	0.51	0.63	0.47

This table shows means for the sample that we use in our machine learning modeling, which consists of applicants with no credit history and no prior credit card who were approved by RappiCard, had at least twelve months with the card by the end of our data period in May 2024, and made at least one transaction during their first 12 months with the card. The mean is calculated separately by group defined by model disagreement between the default and profits models: accepted by both models, accepted by default model only, accepted by profits model only, and rejected by both models. The results use N = 146,036 users, randomly split into training and testing data; summary statistics are shown for the testing sample of N = 29,208. Panel A shows summary statistics for a small subset of the features used by our machine learning model. For non-binary variables, upper-tail winsorization was performed at the 99.9th percentile consistent with the way features were winsorized for modeling. Census tract for each user is inferred based on login activity on the delivery app. The marginality (SES) index is a summary measure of economic vulnerability at the census track level, which takes values between 0 (high marginality) and 1 (low marginality). Panel B shows summary statistics on the terms and use of the credit cards. Interest rate and credit limit are measured at origination. Statement balance, minimum payment, and repayment are averages over the first 12 monthly statements (measured in MXN per month). Default is measured as at least 60 days delinquent at any point over the first 12 months since origination. Panel C shows summary statistics for the average monthly revenues obtained and costs incurred by the lender for each card, where borrower-level monthly averages are computed over the first 12 months since card origination. For panels B and C, non-binary variables are winsorized at the 1st and 99th percentiles, with the exception of "costs from fraudulent transactions," as positive values occur for this variable for less than 1% of users; borrower-level profits are computed using non-winsorized revenue and cost variables. Std. dev. = standard deviation; perc. = percentile; iOS = Apple device operating system; MXN = Mexican pesos; SES = socioeconomic status; BNPL = buy now pay later.

Table A.7: Default rates and profits (z-scores) for different subpopulations

	Default				Profits (z-score)			
Panel A: Modeling Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.204***	0.195***	0.221***	0.163***	-0.020***	0.030***	-0.045***	0.048***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)	(0.004)
Women	-0.006***				0.051***			
	(0.002)				(0.005)			
Age above median		0.014***				-0.058***		
		(0.002)				(0.005)		
iOS			-0.051***				0.119***	
			(0.002)				(0.005)	
% cash above median				0.073***				-0.088***
				(0.002)				(0.005)
Observations	146,036	146,036	146,036	140,742	146,036	146,036	146,036	140,742
	Default				Profits (z-score)			
Panel B: Testing Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.204***	0.195***	0.221***	0.164***	-0.016**	0.032***	-0.043***	0.049***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.008)	(0.008)	(0.008)	(0.008)
Women	-0.006				0.041***			
	(0.005)				(0.012)			
Age above median		0.014***				-0.061***		
		(0.005)				(0.012)		
iOS			-0.050***				0.114***	
			(0.005)				(0.012)	
% cash above median				0.070***				-0.088***
				(0.005)				(0.012)
Observations	29,208	29,208	29,208	28,169	29,208	29,208	29,208	28,169

This table reports OLS regressions of default and profits (z-score) on indicator variables for selected subpopulations. In columns (1)–(4), the dependent variable is default, an indicator equal to one if an individual is late by 60 days or more on their credit card payments at at any time during the first 12 months since card origination. In columns (5)–(8), the dependent variable is the z-score of average monthly profits. Monthly profits are averaged over the first 12 months since card origination. The z-score is constructed by subtracting the cross-sectional mean and dividing by the cross-sectional standard deviation. Each column reports a separate regression on indicators for gender, being above the median age, using an iOS operating system, and having a share of cash transactions on the Rappi app above the median, respectively. Panel A uses all observations in the modeling sample (training and testing). Panel B uses only observations in the testing sample. Individuals with no transactions on the Rappi app are excluded from columns (4) and (8). Heteroskedasticity-robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A.8: Probability of approval by default and profits models for different subpopulations

	Approved by default model				Approved by profits model			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	0.686***	0.711***	0.654***	0.781***	0.587***	0.650***	0.547***	0.685***
	(0.003)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Women	0.022***				0.059***			
	(0.006)				(0.006)			
Age above median		-0.032***				-0.076***		
		(0.005)				(0.006)		
iOS			0.107***				0.165***	
			(0.005)				(0.006)	
% cash above median				-0.160***				-0.131***
				(0.005)				(0.006)
Observations	29,208	29,208	29,208	28,169	29,208	29,208	29,208	28,169

This table reports OLS regressions for approval by default or profit models on indicator variables for selected subpopulations. Columns (1)–(4) correspond to the default model. Columns (5)–(8) correspond to the profits model. For both models, approval is based on profit-maximizing thresholds determined in the training sample: a predicted probability of default of 25% for the default model and predicted profits of -7 MXN for the profits model. Each column reports a separate regression on indicators for gender, being above the median age, using an iOS operating system, and having a share of cash transactions on the Rappi app above the median, respectively. All regressions are estimated on the testing sample. Individuals with no transactions on the Rappi app are excluded from columns (4) and (8). Heteroskedasticity-robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Table A.9: Summary statistics for selected subpopulations

	Men	Women	Age above median	Age below median	Android	iOS	% cash above median	% cash below median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Subset of Features								
Woman - dummy	0.00	1.00	0.43	0.36	0.36	0.44	0.39	0.39
User age	24.54	26.02	31.09	20.60	25.82	23.97	24.67	25.39
User iOS (Apple) operating system - dummy	0.35	0.43	0.30	0.44	0.00	1.00	0.35	0.42
No-hit score	640.09	641.54	643.89	638.21	640.04	641.65	639.59	641.66
Number of orders on app	23.51	24.91	27.73	21.27	18.74	32.72	14.27	34.59
Number of orders paid in cash	7.71	8.22	7.50	8.22	6.68	9.92	11.05	5.65
Median amount per order (MXN)	355.70	340.60	376.52	329.69	354.98	341.45	347.75	376.73
Number of orders at supermarkets	1.63	2.28	2.64	1.32	1.46	2.57	0.63	3.15
Number of orders at pharmacies	0.69	1.06	1.03	0.69	0.65	1.13	0.44	1.25
Number of orders at food establishments	18.47	19.29	21.14	17.01	14.83	25.25	12.11	26.15
Marginality (SES) index of census tract	0.96	0.96	0.96	0.96	0.96	0.97	0.96	0.96
Years of schooling among age 15+ in census tract	12.39	12.46	12.41	12.42	12.24	12.72	12.25	12.58
Proportion households own a motor vehicle in census tract	0.64	0.65	0.64	0.65	0.62	0.67	0.63	0.65
Panel B: Credit Card Terms and Use								
Interest rate	0.77	0.78	0.77	0.78	0.77	0.78	0.78	0.78
Credit limit (MXN per month)	5,203.08	5,120.97	5,216.85	5,136.71	5,189.16	5,141.79	5,056.28	5,186.00
Spending (MXN per month)	3,192.79	2,929.25	3,125.62	3,063.81	2,942.53	3,331.64	2,837.26	3,316.05
Statement balance (MXN per month)	3,815.06	3,634.30	3,933.43	3,601.05	3,829.51	3,606.22	3,793.52	3,663.24
Minimum payment (MXN per month)	457.85	449.43	489.82	427.78	484.27	406.19	509.53	401.20
Repayment (MXN per month)	2,925.86	2,692.38	2,852.33	2,821.49	2,624.48	3,177.38	2,476.43	3,179.45
Default - dummy	0.20	0.20	0.22	0.19	0.22	0.17	0.24	0.17
Panel C: Profit Components (MXN per Month)								
Interest revenue	61.74	68.37	67.45	61.94	62.69	66.97	65.32	63.85
BNPL revenue	7.99	5.47	6.69	7.26	7.84	5.66	6.68	7.12
Charge-offs	60.01	55.93	64.85	53.57	65.49	46.90	69.65	47.53
Interest and BNPL revenue net of charge-offs	10.27	18.10	9.70	16.04	5.56	25.97	2.66	23.82
Interchange fee revenue	39.55	35.99	37.57	38.62	35.28	42.88	32.88	43.16
Cost of rewards	36.56	33.77	36.50	34.70	33.13	39.31	29.13	41.19
Interchange fee revenue net of rewards	3.08	2.45	1.34	3.97	2.32	3.68	3.84	2.16
Late payment fee revenue	12.76	13.50	12.37	13.55	12.54	13.87	13.91	12.19
Funding costs	30.75	30.10	31.28	29.90	30.68	30.20	30.93	30.32
Costs from fraudulent transactions	0.58	0.45	0.45	0.59	0.52	0.55	0.51	0.55
Other fee revenue	2.73	2.62	2.57	2.77	2.53	2.94	2.33	3.12
Other costs	12.58	11.92	12.05	12.53	11.38	13.86	10.58	14.20
Positive profits - dummy	0.50	0.52	0.50	0.52	0.51	0.52	0.52	0.50

This table shows means for the sample that we use in our machine learning modeling. The mean is calculated separately for selected subpopulations. When splitting by percent of transactions made with cash on the Rappi app, individuals with no transactions on the Rappi app are excluded. Panel A shows summary statistics for a small subset of the features used by our machine learning model. The marginality (SES) index is a summary measure of economic vulnerability at the census track level, which takes values between 0 (high marginality) and 1 (low marginality). Panel B shows summary statistics on the terms and use of the credit cards. Interest rate and credit limit are measured at origination. Statement balance, minimum payment, and repayment are averages over the first 12 monthly statements (measured in MXN per month). Default is measured as at least 60 days delinquent at any point over the first 12 months since origination. Panel C shows summary statistics for the average monthly revenues obtained and costs incurred by the lender for each card, where borrower-level monthly averages are computed over the first 12 months since card origination. iOS = Apple device operating system; MXN = Mexican pesos; SES = socioeconomic status; BNPL = buy now pay later. Winsorization is as described in Table 2.

Table A.10: Model performance, number of observations, and number approved, by subpopulation

		Default model			Profits model		
	N	Total profits	AUC	N approved	Total profits	R^2	N approved
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Women	11348	0.457	0.788	8031	0.441	0.080	7331
Men	17860	0.543	0.793	12250	0.559	0.088	10489
Age above median	15271	0.466	0.805	10369	0.487	0.098	8760
Age below median	13937	0.534	0.773	9912	0.513	0.068	9060
iOS	11109	0.498	0.785	8453	0.494	0.064	7914
Android	18099	0.502	0.790	11828	0.506	0.092	9906
% cash above median	14736	0.497	0.775	9147	0.488	0.096	8154
% cash below median	13433	0.504	0.806	10492	0.510	0.073	9196

This table reports out-of-sample performance metrics, number of observations, and number of observations approved, by selected subpopulations, for the main default and profits models. Performance metrics are calculated using only observations in the testing sample belonging to each subpopulation. Column (1) shows the number of observations in the testing sample belonging to each subpopulation. Columns (2)–(4) correspond to the default model. Columns (5)–(7) correspond to the profits model. In columns (2) and (5), total profits are normalized by the total profits generated across all subpopulations by the default and profits models, respectively. In columns (4) and (7), N approved denotes the number of observations in the testing sample belonging to each subpopulation that are approved by the corresponding model. Total profits and N approved are calculated using profit-maximizing thresholds of a 25% predicted probability of default and -7 MXN predicted profits. When splitting by percent of transactions made with cash on the Rappi app, individuals with no transactions on the Rappi app are excluded from the analysis. AUC = area under the receiver operating characteristic curve.

Table A.11: Total profits (normalized) for models trained with different data sources, by subpopulation

	All, but transactions (1)	All, but digital footprint (2)	All, but no-hit score (3)	All, but socioeconomic (4)
Panel A: Default Model				
Women	0.414	0.947	0.963	1.047
Men	0.229	0.884	0.977	1.025
Age above median	0.166	0.885	0.974	1.040
Age below median	0.443	0.937	0.967	1.031
iOS	0.446	0.957	0.978	1.020
Android	0.182	0.869	0.963	1.050
% cash above median	0.299	0.900	0.951	1.052
% cash below median	0.333	0.937	0.995	1.019
Panel B: Profits Model				
Women	0.480	0.964	1.003	1.039
Men	0.330	0.889	0.943	0.993
Age above median	0.307	0.935	1.003	1.009
Age below median	0.481	0.910	0.938	1.017
iOS	0.490	0.936	0.956	1.013
Android	0.304	0.908	0.982	1.013
% cash above median	0.382	0.905	0.949	0.985
% cash below median	0.418	0.933	0.995	1.033

This table shows the total profits obtained from different subpopulations using models trained to predict default or profits with different data sources, normalized by the total profits obtained from the same subpopulation predicting the same variable but using all data sources. Panel A considers models predicting default. Panel B considers models predicting profits. All models are trained using the full training sample (we do not train separate models for each subpopulation). Total profits are calculated in the testing sample using only observations that belong to each subpopulation. For each model, the profit-maximizing threshold is identified in the training sample and applied to predicted values in the testing sample (we do not use separate thresholds for each subpopulation). When splitting by percent of transactions made with cash on the Rappi app, individuals with no transactions on the Rappi app are excluded from the analysis.

Table A.12: Approval rates for models trained with different data sources, by subpopulation

	All	All, but transactions	All, but digital footprint	All, but no-hit score	All, but socioeconomic
	(1)	(2)	(3)	(4)	(5)
Panel A: Default Model					
Women	0.708	0.523	0.652	0.673	0.693
Men	0.686	0.477	0.623	0.651	0.671
Age above median	0.679	0.478	0.626	0.641	0.663
Age below median	0.711	0.513	0.644	0.681	0.697
iOS	0.761	0.651	0.695	0.734	0.746
Android	0.654	0.399	0.597	0.614	0.638
% cash above median	0.621	0.440	0.549	0.581	0.604
% cash below median	0.781	0.562	0.737	0.753	0.768
Panel B: Profits Model					
Women	0.646	0.519	0.589	0.580	0.628
Men	0.587	0.350	0.532	0.508	0.565
Age above median	0.574	0.345	0.541	0.496	0.552
Age below median	0.650	0.493	0.570	0.579	0.631
iOS	0.712	0.668	0.628	0.648	0.693
Android	0.547	0.260	0.510	0.467	0.526
% cash above median	0.553	0.384	0.495	0.481	0.534
% cash below median	0.685	0.466	0.635	0.609	0.661

This table shows approval rates for different subpopulations using models trained to predict default or profits with different data sources. Panel A considers models predicting default. Panel B considers models predicting profits. All models are trained using the full training sample (we do no train separate models for each subpopulation). For each model, the profit-maximizing threshold is identified in the training sample and applied to predicted values in the testing sample (we do not use separate thresholds for each subpopulation). When splitting by percent of transactions made with cash on the Rappi app, individuals with no transactions on the Rappi app are excluded from the analysis.

Table A.13: Total profits (normalized) for models trained to predict different profits components, by subpopulation

	Interest revenue	BNPL revenue	Charge-offs	Interchange net of rewards	Late payment fee	Funding costs
	(1)	(2)	(3)	(4)	(5)	(6)
Women	0.073	0.042	0.439	0.040	0.072	0.000
Men	0.043	0.063	0.564	0.093	0.070	0.000
Age above median	0.033	0.026	0.472	0.004	0.077	0.000
Age below median	0.084	0.080	0.531	0.130	0.065	0.000
iOS	0.152	0.046	0.490	0.113	0.031	0.000
Android	-0.035	0.059	0.514	0.021	0.111	0.000
% cash above median	0.038	0.098	0.499	0.063	0.058	0.000
% cash below median	0.103	0.050	0.506	0.071	0.084	0.000

This table shows the total profits obtained from different subpopulations using models trained to predict different profit components, normalized by total profits generated across all subpopulations using the main profits model trained to predict profits. All models are trained using the full training sample (we do not train separate models for each subpopulation). Portfolio profits are calculated in the testing sample using only observations that belong to each subpopulation. For each model, the profit-maximizing threshold is identified in the training sample and applied to predicted values in the testing sample (we do not use separate thresholds for each subpopulation). When splitting by percent of transactions made with cash on the Rappi app, individuals with no transactions on the Rappi app are excluded from the analysis.

Table A.14: Out-of-sample/out-of-time performance of default model

Model	AUC (1)	Precision (2)	Recall (3)	F1 (4)
Out-of-time	0.779	0.382	0.676	0.488
	[0.773, 0.786]	[0.373, 0.391]	[0.664, 0.688]	[0.479, 0.497]

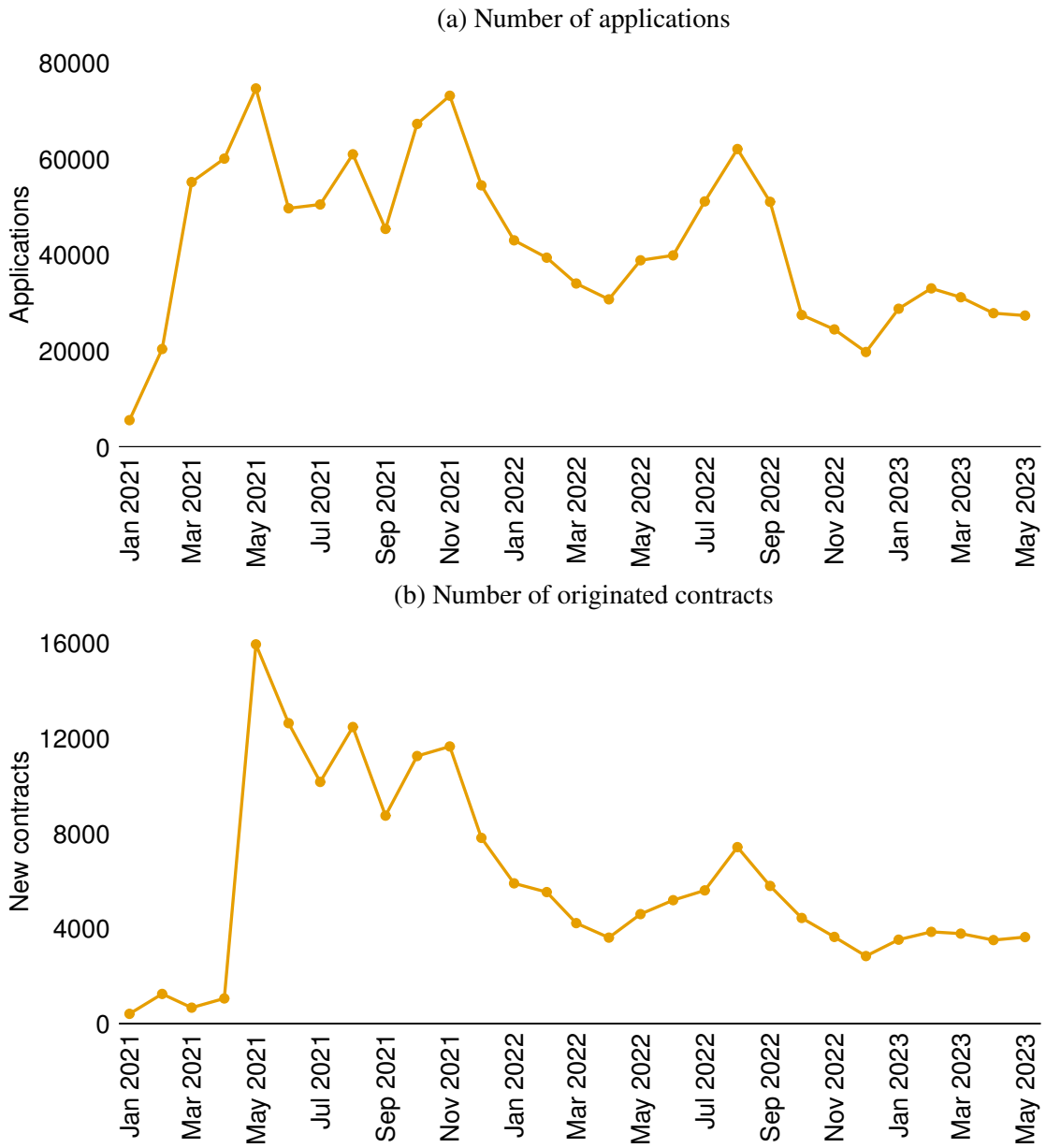
This table reports out-of-sample/out-of-time AUC, precision, recall, and F1 scores for a default model with alternative testing-training split. The training data correspond to individuals who applied from January 1st, 2021 to August 31st, 2022 (79% of modeling sample), and the testing data correspond to those who applied from September 1st, 2022 to May 31st, 2023 (21% of modeling sample). Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples. Threshold-dependent measures use the profit-maximizing threshold of a 25% predicted probability of default. Borrowers with predicted probabilities below this threshold are approved. AUC = area under the receiver operating characteristic curve. F1 score = the harmonic mean of precision and recall. Performance metrics are evaluated only on the testing set. Because the testing sample for this model differs from that of the main profits model, we do not compute total profits normalized by the profits of the main model. Differences in sample size would otherwise induce variation in total profits that is unrelated to model performance.

Table A.15: Default rates, profits, and interest revenue, by interest rate assigned at origination

Annual interest rate	Default (1)	Monthly profits (2)	Monthly interest revenue (3)
80%	-0.003	18.57***	11.45***
	(0.002)	(1.09)	(0.49)
87%	0.000	14.29***	9.56***
	(0.003)	(1.23)	(0.57)
Observations	146,030	146,030	146,030

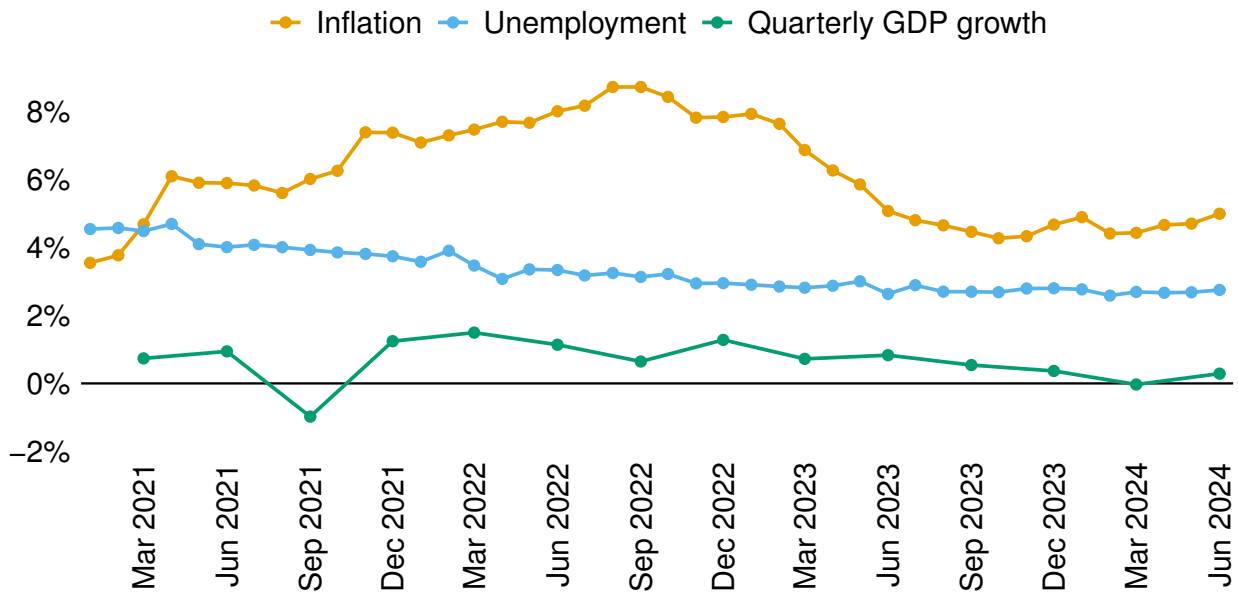
This table shows the results of regressing a binary measure of default and continuous measures of profits and interest revenue on categorical variables capturing interest rate at origination. Monthly profits and monthly interest revenue are measured in Mexican pesos (MXN). We group interest rates assigned at origination into three categories: 72% and below (omitted), 80%, and 87%, based on Figure A.3. Heteroskedasticity-robust standard errors are reported in parenthesis. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

Figure A.1: Number of applications and originated contracts over time



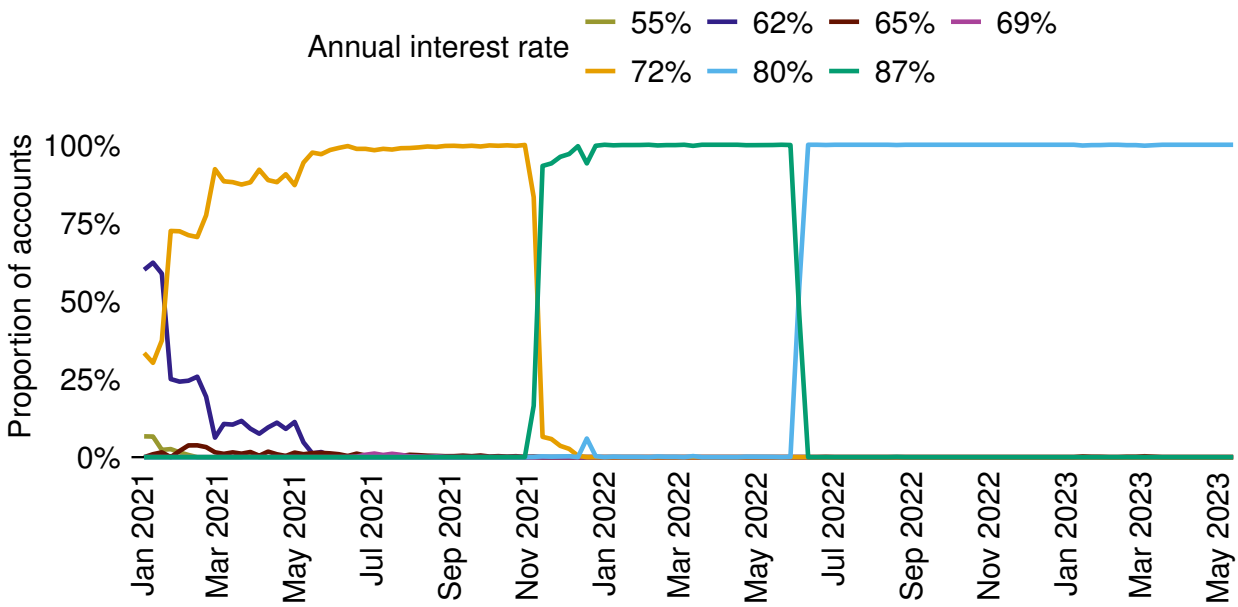
This figure shows the number of applications by month and the number of originated contracts by month from January 2021 through May 2023.

Figure A.2: Macroeconomic indicators for Mexico over period of analysis



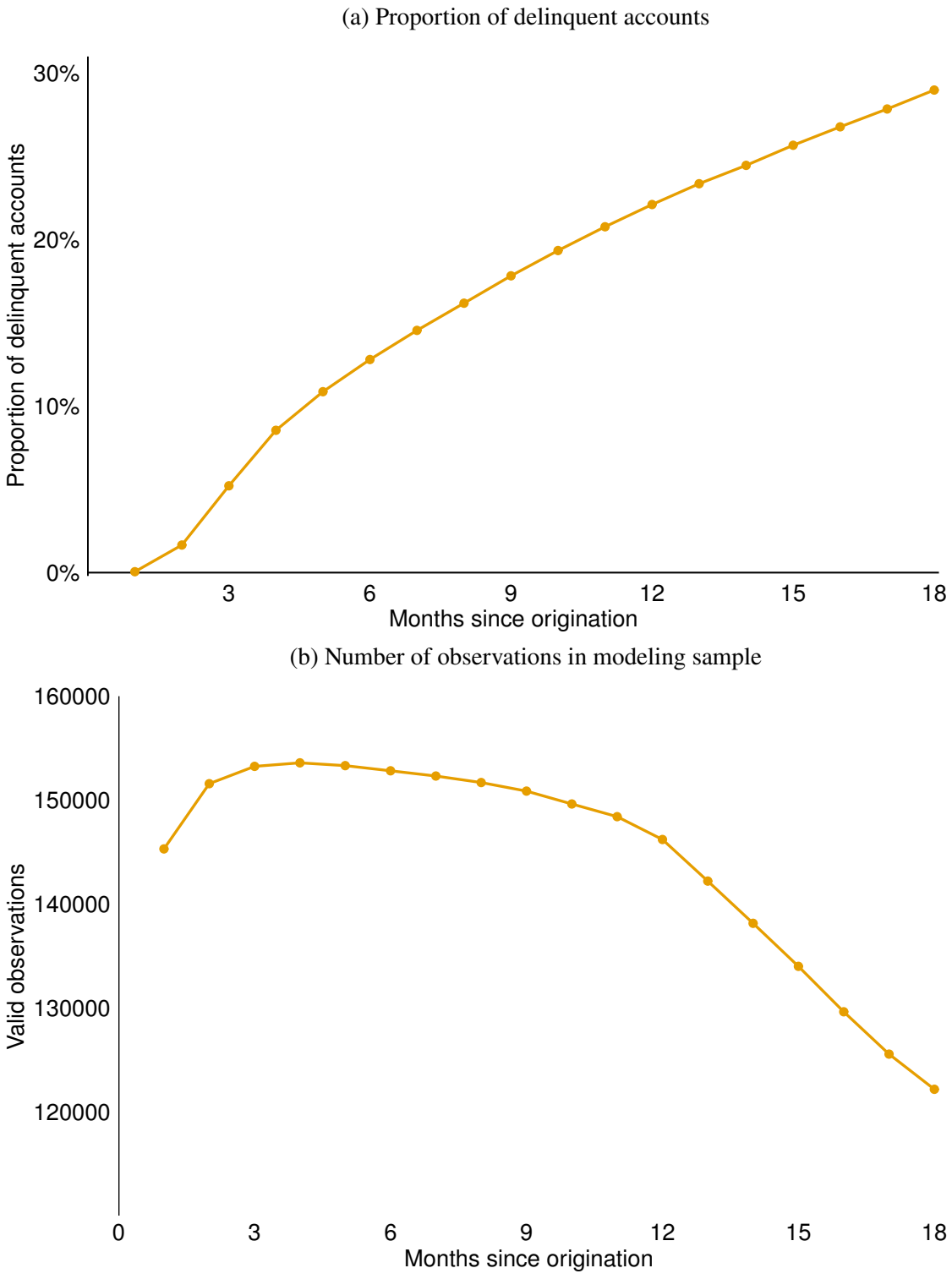
This figure shows annual inflation, unemployment, and quarterly GDP growth rates between January 2021 and June 2024. Annual inflation is published by Banco de México and is the difference in the consumer price index (CPI) on a given month, and the CPI of the same month one year earlier. Unemployment rate is expressed as a fraction of the labor force; it is seasonally adjusted and published by Mexico's National Institute of Statistics (INEGI). Quarterly GDP growth is published by INEGI, is expressed in real terms (GDP is measured in 2018 prices), and is seasonally adjusted.

Figure A.3: Distribution of interest rates assigned at origination, over time



This figure shows the proportion of accounts assigned each annual interest rate by week of origination. N = 146,030. For 6 accounts, we do not observe the interest rate assigned at origination, and thus these accounts are excluded from the figure.

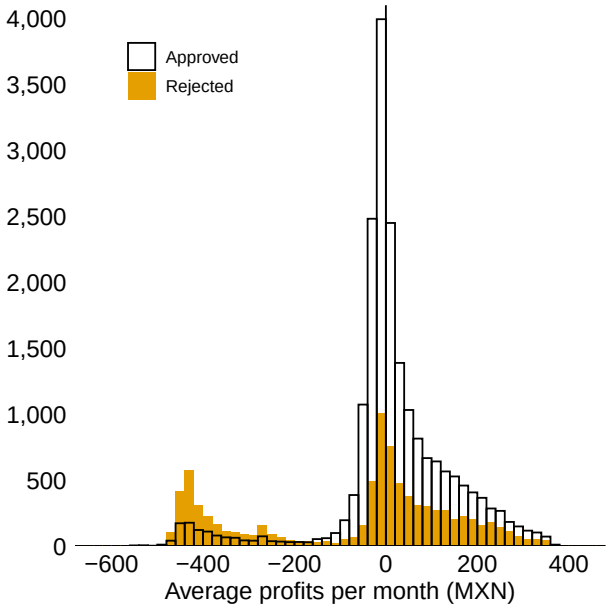
Figure A.4: Trade-offs in threshold of months since origination over which default is measured



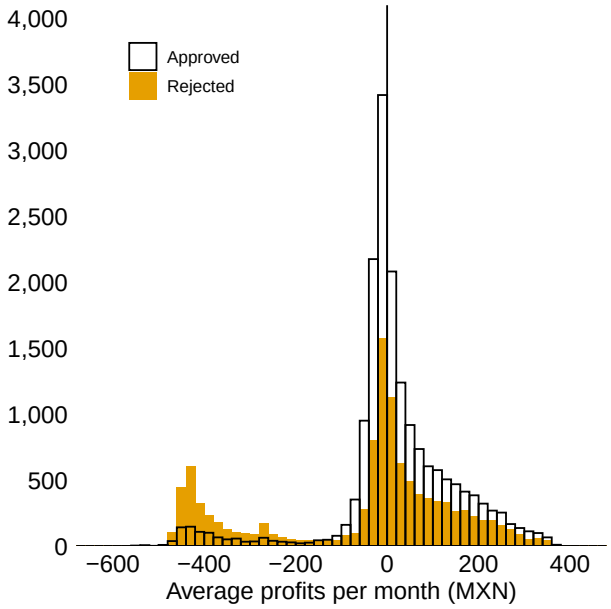
This figure illustrates the trade-offs underlying setting the threshold of number of months since origination over which default is measured. Panel (a) shows the proportion of accounts that are delinquent within x months since origination. Panel (b) shows the number of valid observations available to be included in the modeling sample based on different thresholds over which default is measured.

Figure A.5: Distribution of borrower-level profits, by model approval recommendation

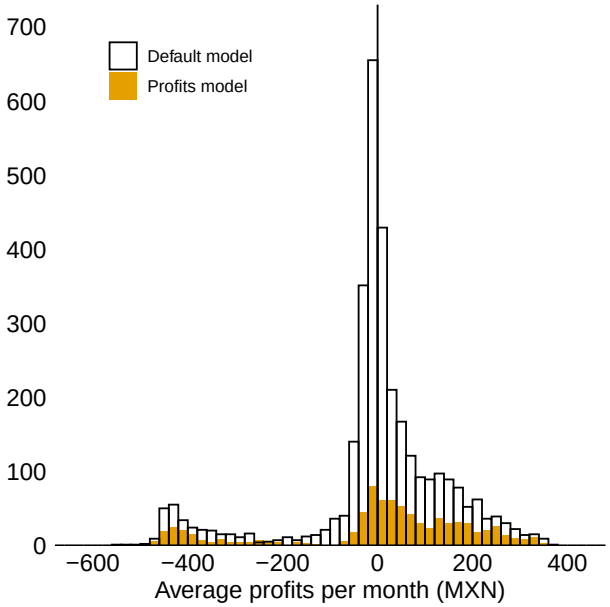
(a) Approved/rejected by default model



(b) Approved/rejected by profit model

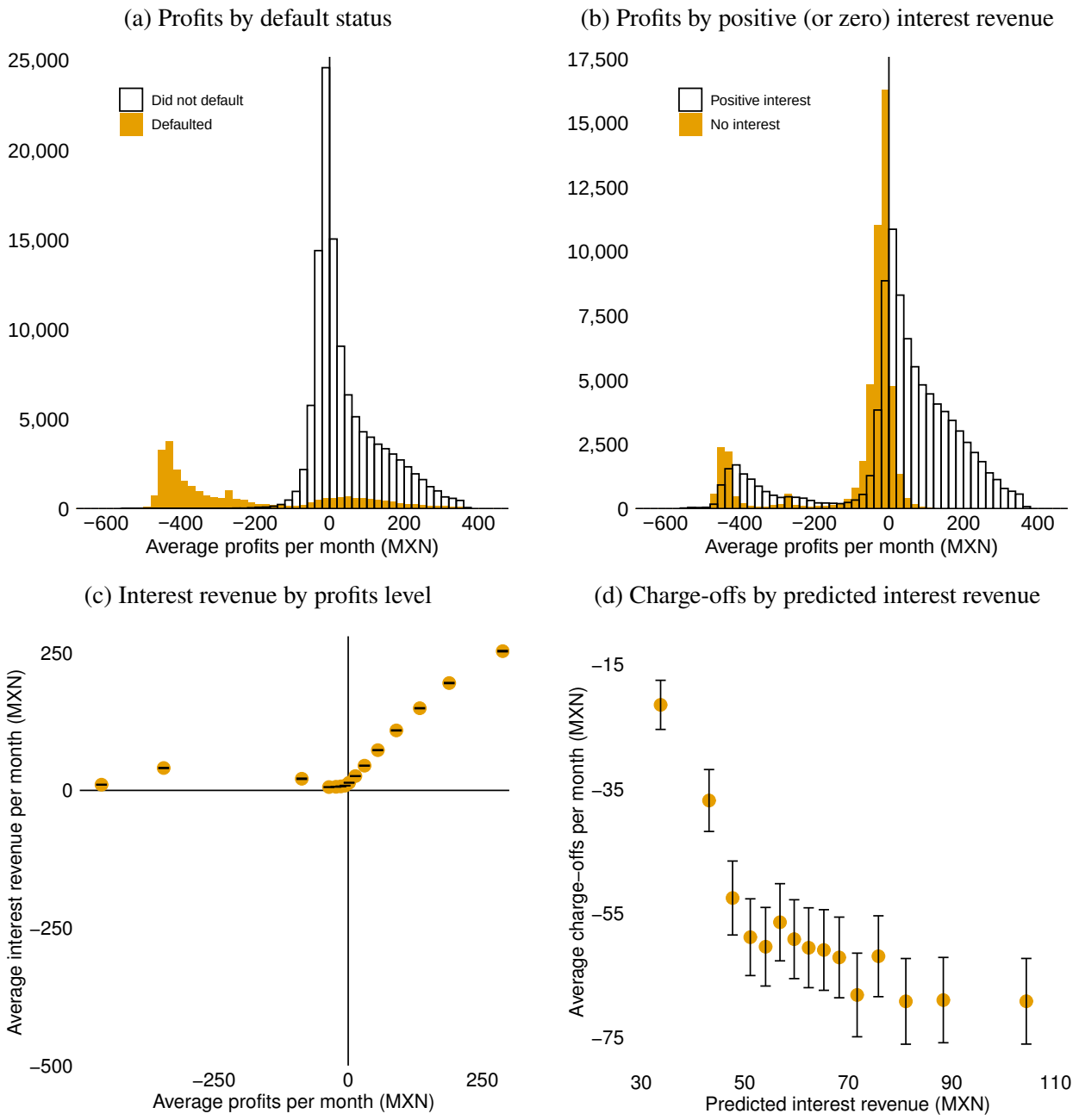


(c) Approved by one model only



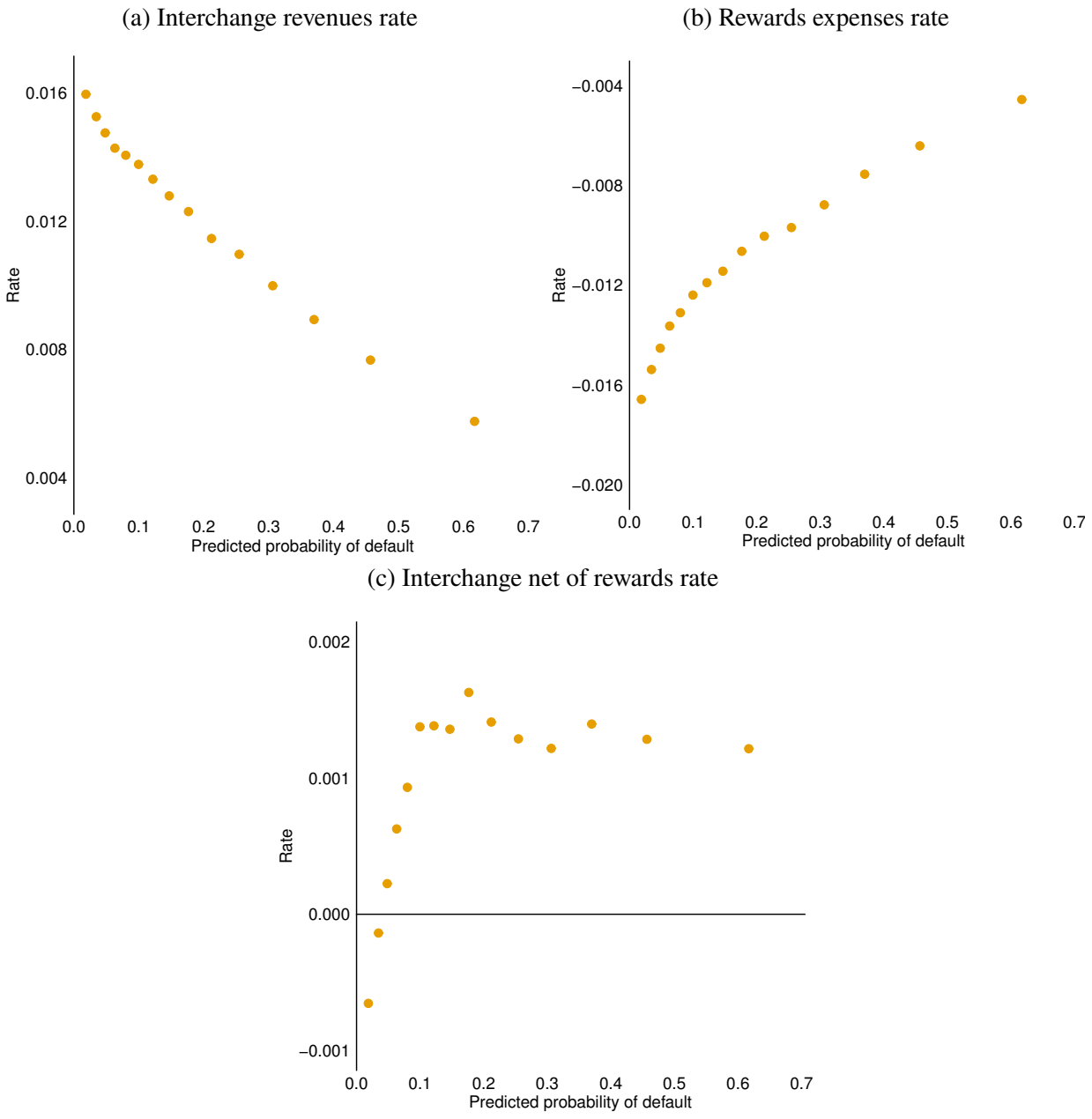
This figure shows the distribution of borrower-level average monthly profits in the testing sample, by approval recommendation of the main default and profits models. For each individual, monthly profits are averaged over the first 12 months since credit card origination. Approval recommendations are computed using the profit-maximizing thresholds of 25% predicted probability of default and -7 MXN predicted profits. Panel (a) considers observations approved (white) or rejected (orange) by the default model. Panel (b) considers observations approved (white) or rejected (orange) by the profit model. Panel (c) considers observations approved only by the default model (white) or approved only by the profits model (orange). The vertical axis plots the number of observations in the testing sample at each level of profits and for each model approval-recommendation.

Figure A.6: Profits, default, interest revenue and charge-offs



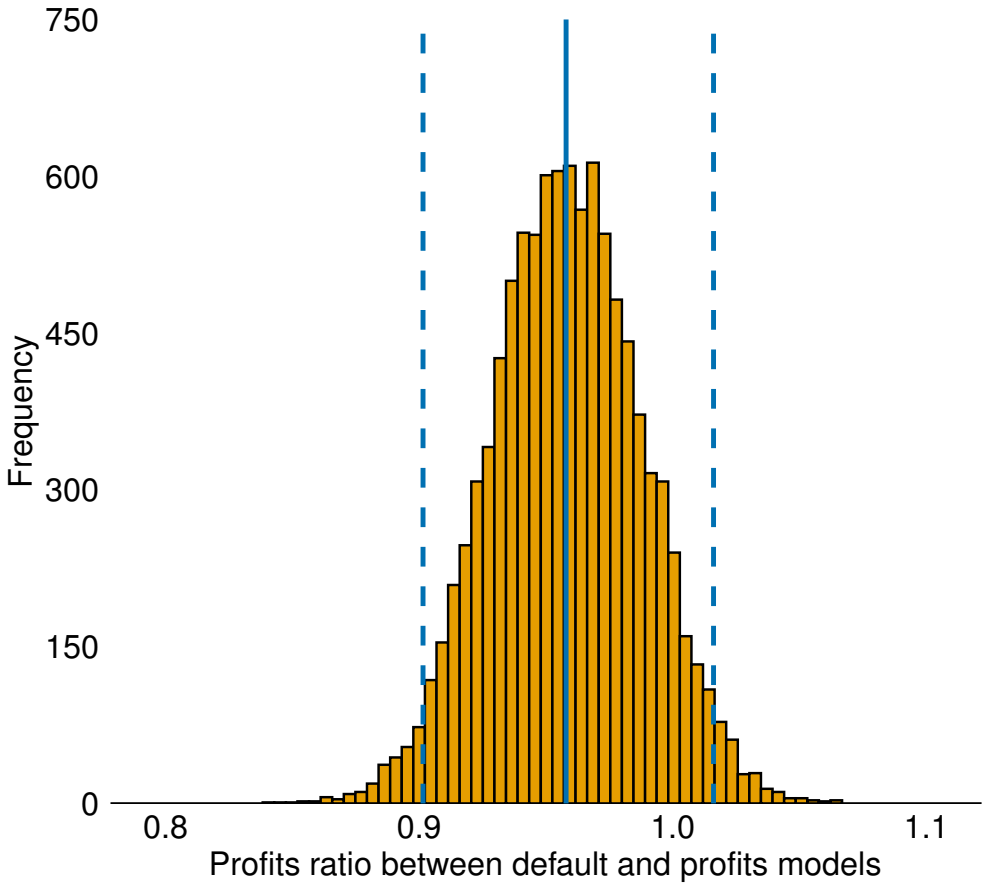
This figure shows various aspects of the relationship between profits, default, interest revenue, and charge-offs. Panels (a) and (b) show the frequency distribution of realized average profits per month, splitting the sample by default status and by positive (vs. zero) interest revenue, respectively. Default is a binary variable that takes the value one if an individual is at least 60 days delinquent at any point. Positive interest revenue is a binary variable that takes the value one if an individual incurs any interest revenue at any point. The vertical axis shows the number of individuals at each profit level. Panel (c) shows realized average interest revenue per month across 15 bins of realized average profits per month, with approximately equal number of observations in each bin. Panel (d) shows realized average charge-offs per month across 15 bins of predicted interest revenue, with approximately equal number of observations in each bin. All variables are measured over the first 12 months since credit origination. Panels (a), (b) and (c) uses the full modeling sample. Panel (d) uses the testing sample. In panels (c) and (d), whiskers represent 95% confidence intervals.

Figure A.7: Interchange fees and rewards as a proportion of spending



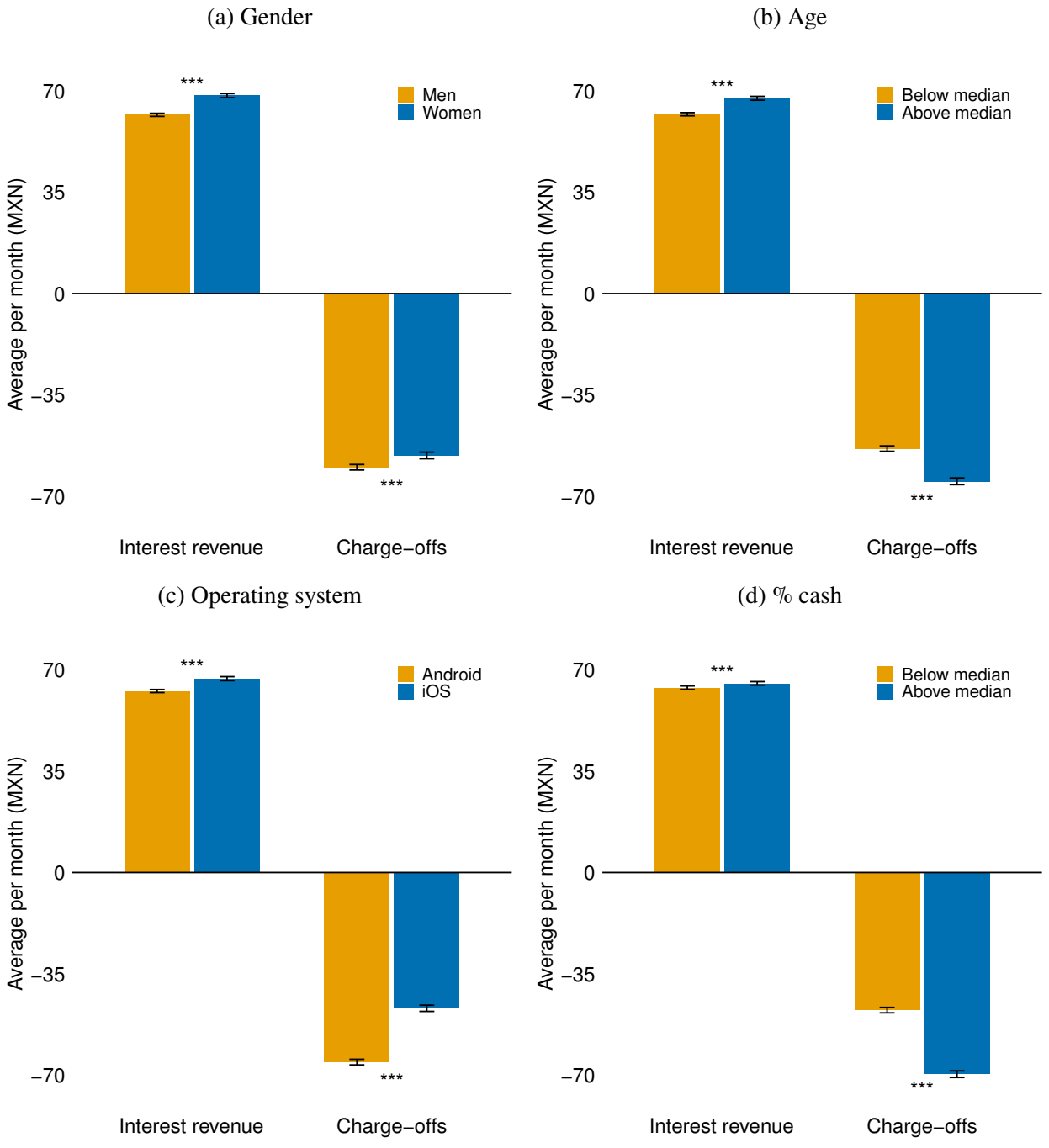
This figure plots a binscatter of interchange revenues, rewards expenses, and interchange revenues net of rewards expenses, all measured as a rate (i.e., as a proportion of spending). The results use N = 146,036 users, randomly split into training and testing data. All estimates are in the testing sample across 15 bins in the predicted probability of default, with approximately equal number of observations in each bin. We construct each measure at the account level by dividing interchange fees, rewards, or interchange fees net of rewards by spending, all over the first 12 months since card origination. We then calculate spending-weighted averages of the rates within bin. (This is equivalent to summing total interchange fees, rewards, or interchange fees net of rewards within bin and dividing by the sum of spending within the bin.)

Figure A.8: Bootstrap estimates of ratio of profits between default and profits models



This figure plots estimates from 10,000 bootstrap samples of the ratio of total profits from the default model to total profits from the profits model, all within the testing sample, using the profit-maximizing threshold for each model. The results use N = 146,036 users, randomly split into training and testing data. The solid vertical blue line shows the estimated ratio in the full testing sample, while the dashed vertical lines show the upper and lower bounds of the 95% bootstrapped confidence interval.

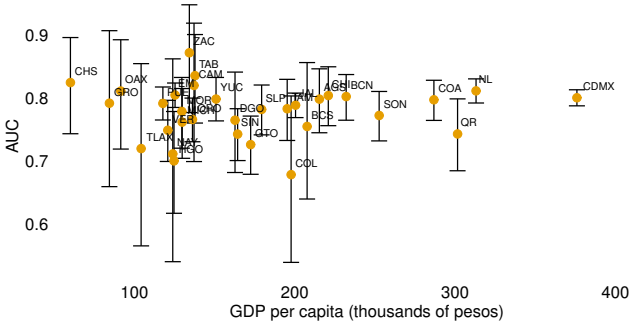
Figure A.9: Interest revenue and charge-offs, by subpopulation



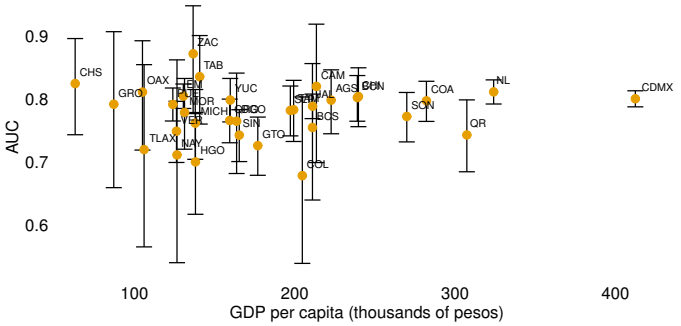
This figure shows average interest revenue and average charge-offs for selected subpopulations. Averages are calculated using the full modeling sample (training and testing). Whiskers represent 95% confidence intervals. Stars between bars indicate statistical significance of the difference between two groups, with heteroskedasticity-robust standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively. When splitting by percent of transactions made with cash on the Rappi app, individuals with no transactions on the Rappi app are excluded from the analysis.

Figure A.10: State-level AUCs and economic activity (without revenue from oil)

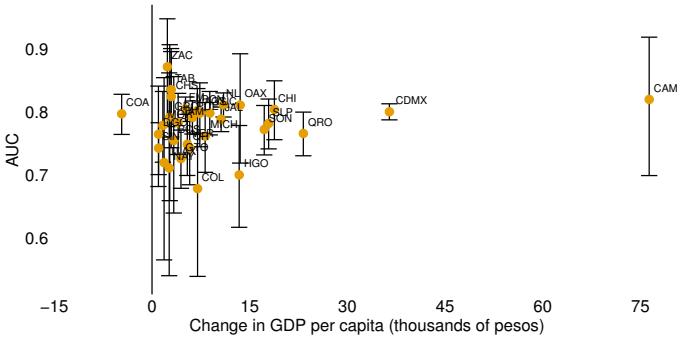
(a) 2021



(b) 2023

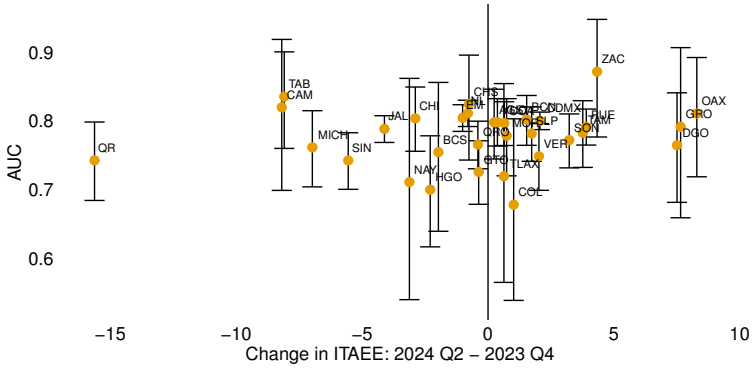


(c) 2023 - 2021



This figure shows scatterplots of state-level AUCs from our benchmark default model (vertical axis) against levels (or changes) in GDP per capita, excluding revenue from oil, for each state (horizontal axis, in thousands of 2018 pesos). Borrowers are assigned to states based on the address provided at the time of credit card application. The results use N = 146,036 users, randomly split into training data and testing data. AUCs are computed on the testing data considering only observations from borrowers in the corresponding state. Predictions come from the default model of Section 4 (not from separate models by state). GDP without revenue from oil is calculated by substracting the contribution of revenue from oil to state-level GDP, as published by Mexico's National Institute of Statistics (INEGI). This value is then divided by the population in each state and year, as reported by Mexico's National Population Agency (CONAPO). Panel A uses GDP per capita for 2021. Panel B uses GDP per capital for 2023. Panel C uses the change in GDP per capita between 2023 and 2021. Whiskers represent 95% bootstrapped confidence intervals for AUCs, obtained with 10,000 bootstrap samples. Labels correspond to state names.

Figure A.11: State-level AUCs and economic activity during the first semester of 2024



This figure shows scatterplots of state-level AUCs from our benchmark default model (vertical axis) against the change in the Quarterly Index of Economic Activity (ITAEE) between 2024Q2 and 2023Q4 (horizontal axis). ITAEE is published by Mexico's National Institute of Statistics (INEGI) and takes the value of 100 for every state in 2018. Borrowers are assigned to states based on the address provided at the time of credit card application. The results use N = 146,036 users, randomly split into training data and testing data. AUCs are computed on the testing data considering only observations from borrowers in the corresponding state. Predictions come from the default model of Section 4 (not from separate models by state). Whiskers represent 95% bootstrapped confidence intervals for AUCs, obtained with 10,000 bootstrap samples. Labels correspond to state names.

B Data Details

B.1 Raw Data

Table B.1 lists a subset of the variables included in the raw data we received from the four data sources described in Section 3.2.

B.2 Feature Construction

We now describe how we construct features from the raw data.

Transaction-level data from the delivery platform We constructed applicant-level features by aggregating *high-frequency transaction records* into a compact set of summary statistics defined over the pre-application transaction history on the Rappi delivery app. The features are designed to capture seven dimensions of behavior:

(i) *platform usage and activity intensity:* intensive and extensive margins of digital purchase behavior, including total order counts and counts in weekly bins prior to the credit card application date (within 1, 2, . . . , 6 weeks before application, as well as more than 6 weeks prior);

(ii) *recency/timing:* the minimum, mean, and maximum number of days between observed transactions and the credit card application date;

(iii) *monetary value and basket characteristics:* measured using a common set of distributional summaries (mean, median, and maximum) applied to per-order spend and basket composition—order value, items per order, and product prices (unit and basket-total)—and complemented by totals of product prices;

(iv) *transaction patterns across payment methods, spending categories, and timing,* represented by modal ("favorite") spending categories, the number of distinct categories observed, and the allocation of activity across time-of-day bins and weekday versus weekend, using both order counts and conditional summaries of order value within each bin/category;

(v) *transaction quality and potential risk signals:* proxied by counts and shares of completed, pending-review, and canceled orders, including cancellation reasons (canceled by user, fraud, payment error, charge-related, and other);

(vi) *discount and tip behavior:* counts and rates as well as level and percentage summaries (means/medians and maxima), while retaining explicit "missing/other" groupings where applicable so that missingness is represented as information rather than implicitly dropped;

(vii) *loyalty and engagement markers* available from the platform, including the customer's loyalty level, an indicator for Prime customer status (a paid membership costing approximately 109 MXN that entails delivery benefits, priority support, discounts, cashback and other rewards, etc.), and an additional categorization constructed by Rappi from transaction history and app engagement (lifetime activity, recent transactions, and inactivity/churn risk).

Overall, the feature construction separates extensive-margin participation, frequency/recency, spending level and composition, and transaction resolution states (completion, review, and cancellation outcomes, including reason codes), while maintaining a transparent mapping from raw transactions to model inputs.

In addition to transactions data, we use data from three additional data sources (our lender's internal application system, the credit bureau, and geographic context), which capture complementary information.

Digital footprints and applicant characteristics We explicitly requested that RappiCard share all variables in the Berg, Burg, Gombović, and Puri (2020) digital footprint, and we received analogous variables for all of these except the time of day that the application was submitted.

The measures we were provided map most directly to the acquisition channel, operating system, email-provider/host, and family of email-string indicators (e.g., name-in-email, number-in-email, lowercase conventions, and typo/error) in Berg, Burg, Gombović, and Puri (2020). For device, we infer whether the user applied through a desktop rather than mobile app based on whether the operating system we observe is "web" rather than "iOS" or "Android," as we confirmed in a separate data set that this is highly correlated with applying on a desktop; we are unable to separate phones from tablets and unable to identify the operating system for desktop users. In contrast, Berg, Burg, Gombović, and Puri (2020) have access to a more granular device variable, with an explicit *device type* classification (desktop, tablet, or mobile) in addition to operating system. Finally, Berg, Burg, Gombović, and Puri (2020) include a *"do-not-track" setting* which restricts the collection of certain tracking-based metadata (i.e., it simultaneously masks device, operating system, and acquisition channel information). We do not observe such a comprehensive privacy flag; instead, the privacy setting we observe and use is whether the user registers via Apple ID using Private Relay which masks the user's email address.

Berg, Burg, Gombović, and Puri (2020) additionally consider *time of credit application*. In their context, this feature is called "*checkout time*" because the credit product they consider is a short-term deferred-payment installment loan to finance a furniture purchase, and thus the purchase of the good being financed occurs at the same time as the credit application. In our setting, the closest indicator would be the application timestamp, which we do not have; we have the application date, but Berg, Burg, Gombović, and Puri (2020) create indicators for time-of-day of the application. We

emphasize the difference between their application-time marker—which simultaneously measures the checkout time for the purchase being financed and the time of the application to finance that purchase—and the repeated timestamps in applicants' pre-application digital purchase history on the Rappi app, which we do have and use to construct the transactions feature set.

In addition to the digital footprint features, we have applicant characteristics from the internal application system. These capture demographics and administrative information observed at application (city/state, gender, age, coarse bins of estimated income, and the applicant's prior application count), compliance/eligibility flags (e.g., potential foreign-identifier and nationality fields, validation that the applicant has not been blacklisted for fraud), and platform-specific policy rules.

"No-hit" scores and credit history for those with limited credit history None of the applicants in our sample have a sufficient credit history for the credit bureau to generate a credit score for them. For this "no-hit" segment, the credit bureau provides a no-hit score calculated by a third party using only the location where the individual lives and publicly available characteristics of that location. For those who have a limited credit history that is nevertheless insufficient for the credit bureau to calculate a traditional credit score, we observe past-due balances split by repayment type (installment vs. revolving), a default indicator, and the length of the recorded credit history. We emphasize that we find these data have essentially no predictive power, which underscores that for those with a limited credit history, the credit bureau was correct to deem these insufficient to generate a traditional credit bureau score.

Socioeconomic characteristics at the census tract level These include census tract-level measures of education, employment/occupation composition, and housing/assets/services (which include access/ownership markers such as computer, cellphone, internet connectivity, vehicles, utility services, and other dwelling characteristics). These data also include a vulnerability index measure measuring structural deficiencies along four dimensions: education, housing, income, and remoteness, which acts as a proxy for socioeconomic development.

B.3 Missing Data and Imputation

We imputed missing covariates using a k-nearest-neighbors (KNN) procedure designed to respect both the evaluation split and the feature-block structure used in our analyses of the importance of each set of features. Imputation was performed separately within the training set and within the testing set, and within gender strata. In addition, we implemented imputation separately within each pre-specified feature block—transactions, digital footprint and applicant characteristics, no-hit scores and limited credit history, and census tract-level socioeconomic characteristics—so that missing values in a given block were imputed using information from that same block only

(i.e., without borrowing information across feature sets). We use the default KNN imputation settings ($K = 5$ neighbors and missing-aware Euclidean distance). We treat data imputation as a preprocessing step rather than an object of optimization; accordingly, we did not tune imputation hyperparameters to maximize predictive performance. The current approach and stratification avoid information flow across the train–test boundary, preserves comparability across gender strata (that is, gender-specific comparisons are not mechanically affected by borrowing information across groups), and maintains alignment between the imputation step and the feature grouping used for assessing the marginal contribution of each set of features.

Table B.1: Summary of variables in raw data from each data source

Panel A: Transaction-level data from the delivery platform
Order status (whether order was completed, pending, canceled)
If canceled, reason for cancellation (e.g., canceled by user, fraud, payment error)
Spending category
Date and time of transaction
Number of items purchased
Total and unit prices for each item purchased
Order value
Discount
Tip amount
Payment method
Loyalty level (category)
Prime customer (paid membership for delivery benefits, discounts, priority support)
Additional categorization variables based on app use (e.g., lifetime activity, risk of churn)
Panel B: Digital footprints and applicant characteristics
Digital footprint features similar to those in Berg, Burg, Gombović, and Puri (2020)
Operating system (iOS vs. Android, only for phone and tablet users)
Desktop vs. mobile phone/tablet
Acquisition channel
Email domain
Email address has number; user's name; lowercase; typo in domain
Privacy setting
Applicant characteristics from application system
Woman - dummy
User age
City and state
Cumulative number of applications for a RappiCard
Nationality
Coarse bins of estimated income
Panel C: "No-hit" scores and credit history for those with limited credit history
No-hit score (generated for everyone in our sample, as no one has a traditional credit bureau score)
Past due balances on installment loans and revolving credit (for those with limited credit history)
Never 90 days late on any credit - dummy (for those with limited credit history)
Length of credit history (for those with limited credit history)
Panel D: Socioeconomic characteristics at the census tract level
Marginality (socioeconomic status) index
Years of schooling, illiteracy rate, and percent achieving various schooling levels among age 15+
Employed population aged 12+
Proportion of households with a motor vehicle and with computing devices (computers, laptops, or tablets)
Dwellings with only one room; electricity, piped water, and drainage; a mobile phone; internet
Average occupants per dwelling and per room

This table lists a subset of the variables included in the raw data we received from four data sources.

C Algorithm Details

C.1 XGBoost Details

We use data on default and profitability to train machine learning models using extreme gradient boosting, or XGBoost (Chen and Guestrin, 2016). Like random forests (Breiman, 2001), XGBoost is an ensemble learner. Ensemble learning is a process that combines several base predictors to produce improved accuracy or stability (Yin and Li, 2022). However, XGBoost and random forests differ in the way they merge predictions from multiple weak models to produce more accurate predictions. Random forests train multiple independent models in parallel and combine the results of multiple classifiers modeled on different subsamples of the data.

XGBoost, like other boosting methods, adds new models into the ensemble sequentially, where each subsequent model attempts to correct the errors of the previous one. In particular, with boosting methods, the training data for each subsequent classifier increasingly focuses on instances misclassified by previously generated classifiers.

XGBoost has become the standard in industry and academic settings due to its scalability and accuracy. It has been shown to outperform other machine learning algorithms in many predictive modeling tasks (Mienye and Sun, 2022). The combination of ensemble learning, gradient descent optimization, and regularization techniques are some of the elements that explain XGBoost's performance and popularity.

XGBoost is the algorithm of choice in other recent work that relies on machine learning to predict creditworthiness (Agarwal, Alok, Ghosh, and Gupta, 2023; Blattner and Nelson, 2024; Blattner, Nelson, and Spiess, 2024; Duarte, Fonseca, Kohli, and Reif, 2025; Lee, Yang, and Anderson, 2024; Lee, Yang, and Anderson, 2026). Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2022) use random forests as their main method, but also use XGBoost in robustness tests. Table 1 presents an overview of papers that employ machine learning to predict creditworthiness, including country, target populations (and in particular the fraction of the target population with a conventional credit score from the credit bureau), data, and methods.

C.2 Training Objectives and Evaluation Metrics

Classification models are trained by minimizing *log-loss* or, equivalently, *cross-entropy loss*. Intuitively, log-loss measures how close the predicted probability is to the corresponding actual/true value (0 or 1 in the case of binary classification). The more the predicted probability diverges from the actual value, the higher is the log-loss value. Log-loss penalizes highly confident incorrect predictions; it takes into account the quality of the predicted probabilities, not only the predicted class labels. As such, log-loss is well-suited to problems for which probability estimates are an object of

interest—as they are for a lender that will allocate credit to everyone with predicted probabilities of default below a threshold (or credit scores above a threshold). In addition, log-loss allows for a more nuanced evaluation of uncertainty. This contrasts with specifying that the model's objective function is to maximize metrics such as the area under the receiver operating characteristic curve (AUC-ROC) or the area under the precision-recall curve (AUC-PR), as these metrics solely focus on the model's ability to discriminate between defaulters and non-defaulters, without penalizing miscalculations in predicted probabilities.

Log-loss is a natural objective when the goal is probabilistic prediction; nevertheless, calibration is an empirical property and is not guaranteed in finite samples (Guo, Pleiss, Sun, and Weinberger, 2017). Accordingly, we examine calibration directly. In our setting, where the label is not subject to misclassification and in combination with 3-fold cross-validation, regularization, and careful data splitting to avoid overfitting, our calibration curves, which compare predicted and true frequencies of the positive label in Figure C.1a, suggest that the predicted probabilities are well-calibrated.[38]

Models that predict continuous outcomes (profits and profit components) are trained by minimizing a squared error metric, so that predictions are optimized to approximate the mean of profits conditional on features. The mean squared error (MSE) places greater weight on large errors in predictions than alternatives such as mean absolute error. This is a desirable when large errors are especially consequential for the decisions informed by the model.

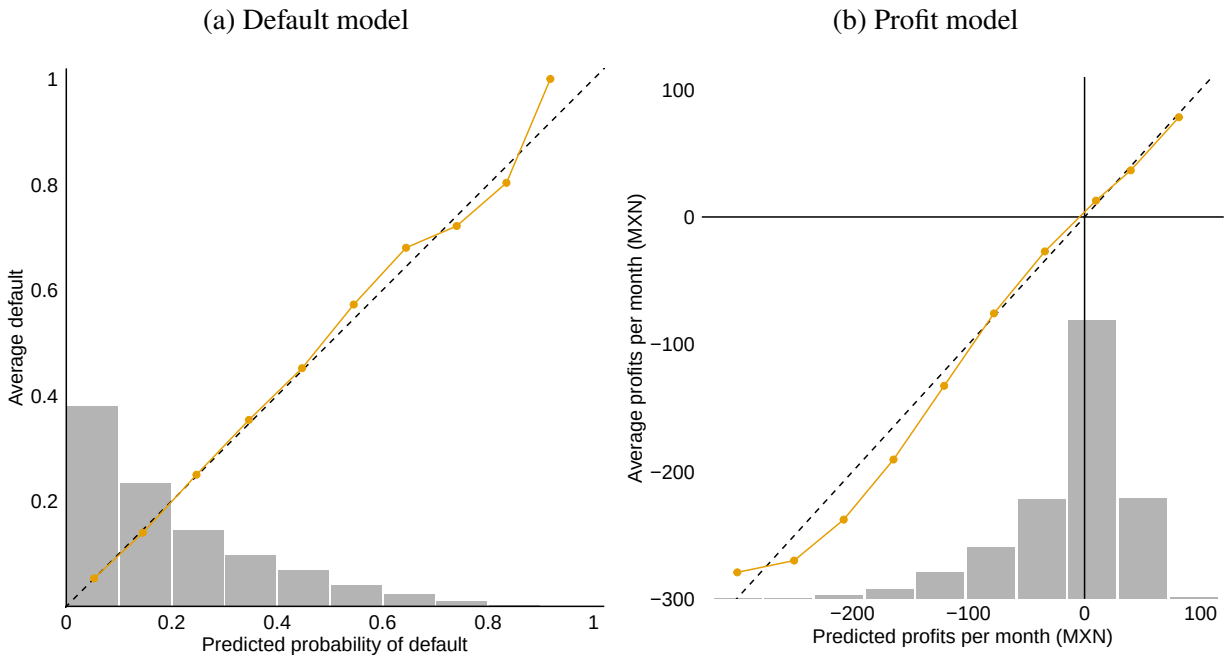
C.3 Hyperparameter Tuning

While manual hyperparameter tuning is essential and time-consuming in many machine learning algorithms, it is especially so in XGBoost. We use Bayesian optimization to tune hyperparameters (for both our default and profits models), relying on sequential model-based optimization as in Bergstra, Yamins, and Cox (2013). Bayesian optimization is more efficient than grid or random search because it attempts to balance exploration and exploitation of the search space. It is also well-suited for cases with a large number of hyperparameters and large search space. Details on the search space we adopt can be found in Table C.1. The Bayesian optimization algorithm was implemented with the aid of 3-fold cross-validation.

For hyperparameter tuning, we use the tree-structured Parzen estimator (TPE) algorithm implemented in Hyperopt, executed in parallel via `SparkTrials`. Trials are distributed across available workers, where a "worker" corresponds to a Spark executor core that fits and evaluates

[38]Calibration techniques like isotonic regression (Zadrozny and Elkan, 2002) and platt-scaling (Böken, 2021) require a validation set, typically 10% of the modeling sample, which can be costly in terms of sample size and can result in less efficient use of the data with possible implications for model performance. As a robustness check, we ran a version of our models in which we split the data into 80% training, 10% validation, and 10% testing sets and used platt-scaling to calibrate probabilities. Our calibration curves in these robustness tests were very similar to the calibration curves shown in Figure C.1.

Figure C.1: Model calibration



This figure shows calibration curves for default and profit models in panels (a) and (b), respectively. Observations in the testing sample are split into 10 equal-width bins between 0 and 1 for the default model, and between -321 and 116 for the profit model. For each bin, we plot average predicted probability of default (or average predicted profits) against actual default rates (or average actual profits). Dotted lines denote 45-degree lines. The height of the gray bars represents the fraction of observations in each bin.

one hyperparameter configuration at a time. Because TPE is sequential, each trial informs subsequent exploration of the hyperparameter space—parallelization introduces a non-deterministic element in the optimization path (whichever worker finishes first determines which completed trial evaluation—hyperparameters and objective value—is incorporated next). Exact replication of hyperparameter values across runs is therefore not guaranteed, even with fixed random seeds. Single-core execution would ensure reproducibility, but would be computationally prohibitive.

C.4 Class Imbalance Considerations

In line with recent recommendations, we do not correct for class imbalance. The observed imbalance in our data arises naturally in the context of our problem and is not related to biased sampling and/or incorrect labels (i.e., misclassification).[39] Furthermore, our default rate/imbalance is 20%, which can be considered moderate compared to settings with very low default rates such as mortgages

[39] In other words, because we observe the ground truth of whether someone is at least 60 days delinquent over the first 12 months since card origination, there is no misclassification. In contrast, in other contexts (e.g., cancer presence based on imaging, mental health condition based on self-reported symptoms, content mediation, and sentiment analysis), labels can be subject to measurement error misclassification.

Table C.1: Search space used in machine learning algorithm

Panel A: XGBoost	
Classification	minimizing log-loss
Regression	minimizing squared error loss
Tuning	hyperopt, max eval 1,250
Panel B: Hyperparameter Space	
Tree-specific hyperparameters	
max_depth	hp.quniform('max_depth', 1, 100, 1)
min_child_weight	hp.loguniform('min_child_weight', -2, 3)
subsample	hp.uniform('subsample', 0.5, 1)
colsample_bytree	hp.uniform('colsample_bytree', 0.5, 1)
n_estimator	hp.quniform('n_estimators', 100, 1000, 1)
Learning and regularization hyperparameters	
η (learning rate)	hp.loguniform('learning_rate', -9, 0)
γ	hp.loguniform('gamma', -10, 10)
α (L1)	hp.loguniform('reg_alpha', -10, 10)
λ (L2)	hp.loguniform('reg_lambda', -10, 10)

This table shows the search space used for hyperparameters in our XGBoost machine learning algorithm. XGBoost = extreme gradient boosting.

(e.g., Fuster, Goldsmith-Pinkham, Ramadorai, and Walther, 2022). Finally, recent work has pointed to the potential unintended consequences of class imbalance corrections for risk prediction models: these corrections can yield poorly calibrated models, where the probability of belonging to the minority class is strongly overestimated, without delivering higher AUC-ROCs compared to models trained without class imbalance correction (van den Goorbergh, van Smeden, Timmerman, and Van Calster, 2022; Piccininni et al., 2024).

D Gender-Segmented Models

Gender gaps in access to credit have persisted despite the automation of creditworthiness evaluations and the entry of many FinTech lenders into the market (IFC, 2024). We hypothesize that concerns about the fairness and equity of algorithmic credit decisions—which have become increasingly important points of discussion and regulation for the use of machine learning models to predict creditworthiness (Bartlett, Morse, Stanton, and Wallace, 2022; Fuster, Goldsmith-Pinkham, Ramadorai, and Walther, 2022)—can be addressed by adopting gender-segmented models without meaningful losses in the profits generated by the model or its predictive accuracy (as proxied by AUCs), nor a deterioration in the portfolio default rates.

We compare credit allocation decisions, performance, and portfolio default rates of the pooled models that combine data on both men and women presented in the main text with gender-segmented models that first split the data by gender prior to training them. It is worth noting that the pooled models are not gender-blind: that is, they are allowed to access the gender variable. By training gender-segmented models, we allow all aspects of the XGBoost algorithm to vary. These include *initialization of the base learner* (i.e., a simple prediction for all observations: the log odds of default) as well as the *learning path and aggregation* of weak learners into the ensemble model, some aspects of which are governed by hyperparameters.[40] As such, even when we allow the pooled model to access the gender variable, the learning and aggregation process may look quite different between the gender-segmented and pooled models.

In a previous version of this paper (Chioda, Gertler, Higgins, and Medina, 2024), we found that 12.3% of women who would be rejected by a standard pooled machine learning model would instead be approved by the gender-segmented model. In contrast, only 4.0% of women who were approved by the pooled model would be rejected by the gender-segmented model. This was achieved with virtually no loss in predictive power: the gender-segmented model had an AUC of 0.750 compared to 0.752 for the pooled model. Portfolio default rates were also nearly identical at 10%, and the allocation of credit to men was largely unchanged, with 2.8% approved by the pooled model but rejected by the gender-segmented model and 2.6% approved by the gender-segmented model but rejected by the pooled model.

We made several changes to the model from Chioda, Gertler, Higgins, and Medina (2024), including using data covering a longer time period, a larger sample, higher-quality data, and a consistent definition of default across cohorts. Table D.1 shows that after these changes, the AUC of the pooled model (0.791) is similar to that of the men-only (0.79) and women-only (0.785) models, as before. Turning to credit allocation decisions, panel A of of Table D.2 shows that, unlike in the earlier version, the proportion of women approved by the gender-segmented model but rejected by the pooled model is very similar to proportion of women approved by the pooled model but rejected by the gender-segmented model (both around 3%). Panel B shows that, as in the previous version, allocation of credit to men is largely unchanged. Panel C shows that all changes occur without meaningful differences in portfolio default risk, with both models leading to a default rate of 10.1% and 10.4%. Figure D.1 shows that this is because—unlike in Chioda, Gertler, Higgins, and Medina (2024)—there is no longer a large mass of women receiving lower predicted probabilities of default in the gender-segmented model compared to the pooled model.

This is likely due to improvements in the model's accuracy which shrink the differences in

[40] Key elements of the iterative learning process include residual errors for each weak learner; the direction in which predictions should be modified to reduce the loss in subsequent learners (gradient descent step); and the learning rate and minimum loss reduction required to make a further partition on a leaf node of the tree.

Table D.1: AUC of pooled and gender-segmented models

Model	Full sample (1)	Women only (2)	Men only (3)
Pooled model	0.791	0.788	0.793
	[0.785, 0.797]	[0.777, 0.798]	[0.785, 0.801]
Gender-segmented model	0.788	0.785	0.790
	[0.782, 0.794]	[0.775, 0.796]	[0.781, 0.798]

This table shows the out-of sample AUC for both the pooled and the gender-segmented models, calculated for three samples: the full sample (all observations), men only, and women only. That is, for the gender-segmented model predictions from the men-only and women-only samples, we estimate separate models and use them to make predictions on the segmented testing data. For the gender-segmented predictions on the full sample, we estimate separate models on the segmented training data and use these models to make predictions on the full sample of testing data (including both men and women, according to their gender). For the pooled model predictions on the full sample, we estimate one model on the pooled (men and women) training data and use it to make predictions on the full sample of testing data. For the pooled model predictions on the men-only and women-only samples, we estimate one model on the pooled (men and women) training data and use it to make predictions on the segmented samples of testing data, according to the gender of the applicant. The results use N = 146,036 users (N = 89,298 men and N = 56,738 women), split into training data to train the machine learning models and testing data to calculate out-of-sample measures of model performance. AUC = area under the receiver operating characteristic curve.

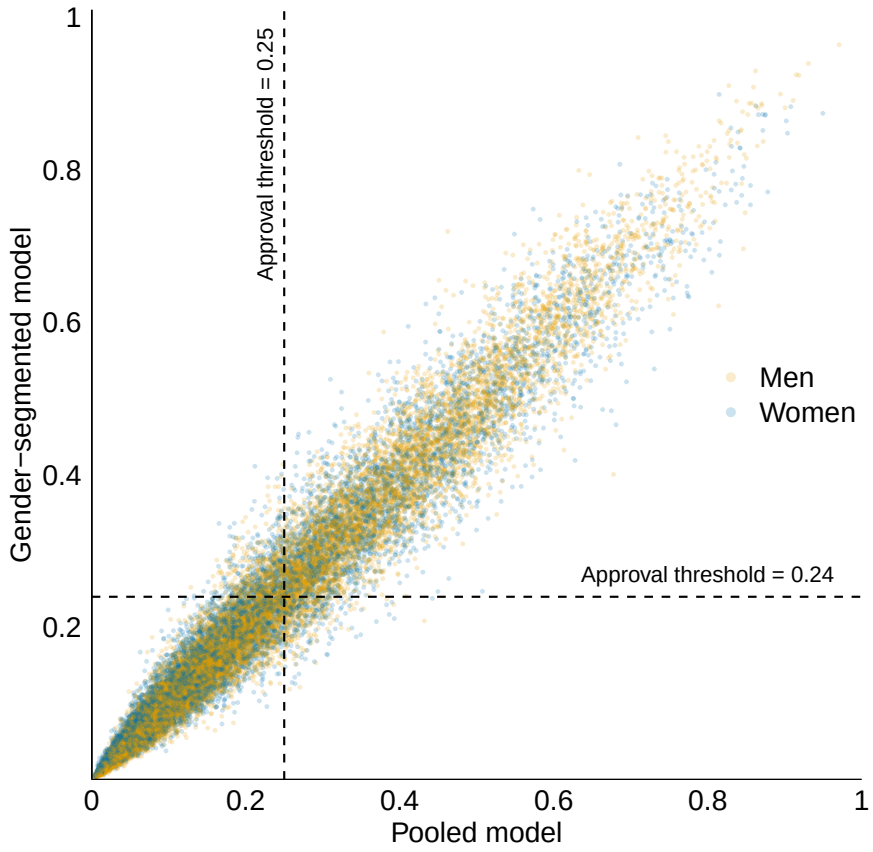
allocation decisions across the two models. Intuitively, if both models perfectly predicted default, neither would make mistakes in their credit allocation decisions and both would approve the same set of borrowers, so there would be no fairness gains from moving from a pooled to a gender-segmented model. The model's improved accuracy relative to the model in our previous version likely stems from using data covering a longer time period, a larger sample, and higher-quality data (because for the updated version of the paper we worked extensively with our FinTech partner to mitigate data quality issues and missing values across features that arose during the prior data transfer process).

Table D.2: Model Disagreement and default rates

	(1)	(2)
Panel A: Model Disagreement for Women		
	Gender-segmented model	
Pooled model	Approved	Rejected
Approved	67.2%	3.6%
Rejected	1.9%	27.3%
Panel B: Model Disagreement for Men		
	Gender-segmented model	
Pooled model	Approved	Rejected
Approved	65.4%	3.1%
Rejected	1.7%	29.7%
Panel C: Default Rates of Overall Portfolio		
Pooled model	10.4%	
Gender-segmented model	10.1%	

This table shows agreements and disagreements between the pooled and gender-segmented models, expressed as proportions of all applicants. It also shows the overall portfolio default rates of the two models (panel C), i.e., the default rate of all applications approved under each model. The results use N = 146,036 users (N = 56,738 women and N = 89,298 men), split into training data to train the machine learning models and testing data to calculate the measures reported in the table out-of-sample. Approval recommendations are computed using profit-maximizing thresholds calculated separately for the pooled and gender-segmented models; the profit-maximizing threshold is a 25% predicted probability of default for the pooled model and 24% for the gender-segmented model (we do not use different thresholds for men and women).

Figure D.1: Predicted probabilities of default in pooled and gender-segmented models



This figure shows the predicted probabilities of default for each observation in our out-of-sample testing data under both the pooled and gender-segmented models. Blue dots represent women and orange dots represent men. The results use N = 146,036 users (N = 56,738 women and N = 89,298 men), split into training data to train the machine learning models and testing data to calculate out-of-sample predicted default probabilities which are shown in the figure. Approval recommendations are computed using profit-maximizing thresholds calculated separately for the pooled and gender-segmented models; the profit-maximizing threshold is a 25% predicted probability of default for both the pooled model and 24% for the gender-segmented model (we do not use different thresholds for men and women). The lower-left quadrant shows applicants who would be approved by both the pooled and gender-segmented models, the upper-right quadrant shows applicants who would be rejected by both models, the upper-left quadrant shows applicants who would be rejected by the gender-segmented model but approved by the pooled model, and the lower-right quadrant shows applicants who would be rejected by the pooled model but approved by the gender-segmented model.

E Additional Tests and Robustness

E.1 Data Drift

Because changes in the composition of the applicant pool over time may affect the stability of model inputs, we assess whether the distribution of covariates shifts over the course of the sample using a data drift analysis (Lu et al., 2018). Using the raw covariates *prior to imputation*, for every quarter starting in January 2021, we test whether the data from a new, adjacent quarter's contracts present a statistically significant shift in the distribution of any of the covariates/features relative to all previous quarters. We conduct this test across all binary, categorical, and numerical covariates and all quarters, resulting in 1,863 comparisons; across all of these tests, only 42 comparisons (2.25%), spanning only 19 variables out of 207, are significant at the 5% level.[41] Details on data drift testing procedures for numeric (continuous and discrete), binary, and categorical variables are provided in Table E.1. Each group of variables is tested using three testing procedures. A comparison is deemed significant if at least two of the three tests yield a statistically significant drift for each group of variables.

Table E.1: Data drift test statistics

Data drift test	Variables	Definition	Reference
(1)	(2)	(3)	(4)
Population stability index (PSI)	Numerical, categorical, binary	Compares distributions by variables and computing a log-ratio of frequencies between expected (baseline) and observed (updated data) datasets.	(Siddiqi, 2017)
Jensen-Shannon (JS) divergence	Numerical, categorical, binary	Jensen-Shannon is a symmetric divergence metric that measures the relative entropy or difference in information represented by two probability distributions. It is always finite and interpretable as a measure of similarity. It is derived from the Kullback-Leibler divergence but smoother.	(Lin, 1991)
Chi-squared test	Categorical, binary	Tests for independence between distributions using a comparison of observed vs. expected category frequencies.	(Pearson, 1900)
Kolmogorov-Smirnov (KS) test	Numerical	Non-parametric test comparing empirical cumulative distributions from two samples. Sensitive to differences in shape, location, and spread.	(Massey, 1951)

This table describes the data drift test statistics that we use.

[41] The period with the largest number of significant comparisons (11) is Q3 2021. This quarter is toward the beginning of the sample and is a period of rapid growth and likely experimentation.

E.2 Penalized Linear Models

A potential concern with highly flexible algorithms is linked to their ability to uncover unseen and intricate data patterns which do not generalize well in unseen data. In the context of XGBoost, this behavior could arise due to non-robust data (see Section 6.1 and Appendix E.1), outlier sensitivity, overfitting, and other issues. The best-known and most extreme version of this phenomenon has been studied in the context of neural networks which extract their own features, some of which may be non-robust even when derived from patterns in the data distribution (Athalye, Engstrom, Ilyas, and Kwok, 2018; Biggio et al., 2013; Ilyas et al., 2019).

To further examine the stability and robustness of our results, we implemented two regularized logistic regression models, using L1 regularization (LASSO; Tibshirani, 1996) and elastic net regularization (GLMNET; Friedman, Hastie, and Tibshirani, 2010), with penalty parameters tuned via cross-validation. We compared LASSO and GLMNET with our original XGBoost implementation using a consistent evaluation pipeline, including performance metrics with bootstrapped confidence intervals.[42] Empirically, we do not see significant differences between the two regularization methods, with GLMNET regularization behaving almost like a pure LASSO, assigning 90% weight to the L1 regularization parameter and 10% to L2.[43]

Table E.2 shows performance metrics for the two penalized linear models. While XGBoost outperforms both penalized linear models in terms of both total profits generated by the model and predictive performance, the performance of all three models is broadly consistent. XGBoost has a higher AUC (0.791 vs. 0.763 for LASSO vs. 0.761 for GLMNET). XGBoost achieves a better F1 score (0.512 vs. 0.477 for LASSO vs. 0.478 for GLMNET) by better balancing precision and recall. Precision is higher for XGBoost (0.425 vs. 0.351 for LASSO vs. 0.366 for GLMNET), while recall is lower for XGBoost (0.644 vs. 0.743 for LASSO vs. 0.688 for GLMNET). These results offer additional confidence that the XGBoost model is uncovering relevant nonlinear data patterns that deliver better precision and sharper class separation, while having lower recall than the logistic regression models—which are notoriously more cautious and tend to favor recall by smoothing predictions towards the observed default rate and generating more inclusive classification boundaries.

Overall, the results across the three algorithms also offer additional evidence on the stability of our benchmark model. The model's robustness can be explained by choices made when training

[42] The performance of logistic regression classifiers coupled with L1/L2 norms is affected by the scale of features. In order to allow for an adequate comparison across features with different scales and improve the classifier performance, features were centered and scaled to have zero mean and unit variance (separately in the training/testing sets) to ensure coefficients could be directly compared in magnitude and to improve model convergence. Standardization is not needed for XGBoost models, since tree splits depend on the order or rank of values, not on the magnitude or scale of inputs.

[43] L1 regularization tends to be more aggressive in feature selection but can be unstable when features are correlated. Elastic net, which combines L1 (LASSO) and L2 (Ridge), tends to offer stability in terms of feature selection.

Table E.2: Penalized linear models

Model	Total profits (normalized) (1)	AUC (2)	Precision (3)	Recall (4)	F1 (5)
LASSO	0.809	0.763	0.351	0.743	0.477
	[0.744, 0.878]	[0.757, 0.770]	[0.342, 0.359]	[0.732, 0.754]	[0.468, 0.486]
GLMNET	0.784	0.761	0.366	0.688	0.478
	[0.715, 0.854]	[0.755, 0.768]	[0.358, 0.375]	[0.676, 0.700]	[0.469, 0.487]

This table reports out-of-sample total profits, AUC, precision, recall, and F1 scores for penalized linear models predicting default. LASSO and GLMNET are penalized logistic regression models using L1 regularization and elastic net regularization (GLMNET), respectively. The results use N = 146,036 users, split into training data to train the machine learning models and testing data to calculate out-of-sample measures of model performance. Total profits (normalized) are the total profits from using each model, normalized by total profits from the profits model that uses all features. Threshold-dependent measures use the profit-maximizing thresholds 21% for LASSO and 23% for GLMNET. Bootstrapped 95% confidence intervals are included in square brackets, using 10,000 bootstrap samples; normalization of total profits are computed within each bootstrap sample. AUC = area under the receiver operating characteristic curve. F1 score = the harmonic mean of precision and recall. Performance metrics are evaluated only on the testing set.

XGBoost. Not only does our implementation include L1 and L2 regularization (with the tuned relative importance giving more prominence to the L1 relative to the L2 norm, in line with the above GLMNET results), but the tuned hyperparameters result in additional layers of regularization along several dimensions. That is, stochastic training (50% features, 75% sample), conservative learning (1.9% learning rate), and structural constraints prevent individual trees from becoming too complex (allowing for many trees with shallow depth, and penalizing too many splits). By design, our modeling approach had the goals of avoiding overly precise but non-generalizable predictions and of guarding against possible brittleness.

F Comparison of Studies

Tables 1 and A.1 compare studies that predict creditworthiness. This appendix provides additional detail about decisions we made when extracting numbers from those papers for the tables.

F.1 Predictive Performance

Agarwal, Alok, Ghosh, and Gupta (2023) use both random forest (RF) and XGBoost; we report the AUC from their best-performing model, which is RF. Agarwal, Alok, Ghosh, and Gupta (2023) do not report an overall AUC for the full sample including those with and without credit scores. For Berg, Burg, Gombović, and Puri (2020), we report the out-of-sample AUC using credit bureau scores, digital footprints, and fixed effects. Björkegren and Grissen (2020) use both RF and logistic

regression; we report the AUC from their best-performing model which is logistic regression. For Blattner and Nelson (2024) we report the AUC of the XGBoost baseline model. For Blattner and Nelson (2024) we calculated F1 score from reported recall and precision measures. For Blattner, Nelson, and Spiess (2024), we report the AUC of the XGBoost model. For Caire and Vidal (2024), we use the AUC for KarmaLife borrowers; we do not use the AUC for Fundfina borrowers since prior loan repayment with the same lender was used as an input to the Fundfina model, and this input would not be available to lenders seeking to lend to new applicants rather than repeat borrowers. For De Cnudde et al. (2019), we report the AUC from the best-performing model, which is an ensemble using a network-only link-based classifier to process the Facebook network data. Di Maggio and Ratnadiwakara (2025) report the AUC of the FinTech platform's model for the full sample as well as those with subprime and prime credit scores; we use their AUC for the full sample.

For Frost et al. (2019), we report the AUC for the XGBoost model. For Fuster, Goldsmith-Pinkham, Ramadorai, and Walther (2022), we report the AUC for RF with race as a variable. The AUC we report for Gambacorta, Huang, Qiu, and Wang (2024) is the one for the baseline model using all information except the interest rate, as the interest rate would not be available at the time of loan application. For Hair, Howell, Johnson, and Matsumoto (2025), we report the AUC for their preferred default model using FICO and cash flow data. For Iyer, Khwaja, Luttmer, and Shue (2016), we report the AUC combining all data. For Jagtiani and Lemieux (2019), we report the AUC for the best-performing model, which uses rating grades and other control factors. Khandani, Kim, and Lo (2010) report a range of AUCs without additional detail (and do not report if they are estimated in-sample or out-of-sample); we report the upper end of the range they report. For Lee, Yang, and Anderson (2024), we report the AUC for the best-performing model, which uses all data sources predicting ever-delinquent.

For Lee, Yang, and Anderson (2026), we report the AUC for the model using baseline and retail data. For Meursault, Moulton, Santucci, and Schor (2025), we report the AUC for the overall XGBoost model, averaged over all years. For Netzer, Lemaire, and Herzenstein (2019), we report the AUC of the model with text, financial, and demographic data. For Rishabh (forthcoming), we report the AUC of the model using "traditional hard information" and granular payments data. Sadhwani, Giesecke, and Sirignano (2021) report AUCs for going from each potential state this month to each potential state next month, where the potential states are current, 30 days delinquent, 60 days delinquent, 90 days delinquent, and foreclosure; we use the AUC for predicting transitioning from 60 days delinquent to 90 days delinquent in their best-performing model. For San Pedro, Proserpio, and Oliver (2015), we report the AUC for default at 90 days using all data sources.

F.2 Percent of Applicants with Credit Bureau Data

For Gambacorta, Huang, Qiu, and Wang (2024), the "% with credit data" is based on the percent with a credit score produced by the FinTech based on formal borrowing histories, as the paper does not have access to credit bureau data (though the sample is likely to have a credit score in the credit bureau). For Huang et al. (2023), the "% with credit data" is based on the presumed percent with MYBank credit scores based on formal credit histories, as the paper does not have access to credit bureau data (though the sample is likely to have a credit score in the credit bureau). San Pedro, Proserpio, and Oliver (2015) do not report the "% with credit data", but the authors report an AUC using credit bureau data, so we assume it is 100%.

F.3 Default Rate

For Blattner and Nelson (2024) we report a weighted average default rate constructed using the default rate for accepted loans (2.2%) and the default rate for rejected loans (11.5%). For Blattner, Nelson, and Spiess (2024) we report a weighted average default rate constructed using the default rate for accepted loans (1.2%) and the default rate for rejected loans (24.2%). For Frost et al. (2019), we report the default rate as the weighted average of loss rate across internal ratings. For Sadhwani, Giesecke, and Sirignano (2021) we report as the default rate the percentage of loans transitioning from 60 days delinquent to 90 days delinquent.

Appendix References

Agarwal, Sumit, Shashwat Alok, Pulak Ghosh, and Sudip Gupta (2023). "Financial Inclusion and Alternate Credit Scoring: Role of Big Data and Machine Learning in Fintech."

Albanessi, Stefania and Domonkos Vamossy (2024). "Credit Scores: Performance and Equity."

Athalye, Anish, Logan Engstrom, Andrew Ilyas, and Kevin Kwok (2018). "Synthesizing Robust Adversarial Examples." *Proceedings of the 35th International Conference on Machine Learning*. International Conference on Machine Learning. PMLR, 284–293.

Bartlett, Robert, Adair Morse, Richard Stanton, and Nancy Wallace (2022). "Consumer-Lending Discrimination in the FinTech Era." *Journal of Financial Economics* 143(1), 30–56.

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri (2020). "On the Rise of FinTechs: Credit Scoring Using Digital Footprints." *The Review of Financial Studies* 33(7), 2845–2897.

Bergstra, James, Dan Yamins, and David Cox (2013). "Hyperopt: A Python Library for Optimizing the Hyperparameters of Machine Learning Algorithms." Python in Science Conference. Austin, Texas, 13–19.

Biggio, Battista, Igino Corona, Davide Maiorca, Blaine Nelson, Nedim Srndic, Pavel Laskov, Giorgio Giacinto, and Fabio Roli (2013). "Evasion Attacks against Machine Learning at Test Time." Vol. 7908, 387–402.

Björkegren, Daniel and Darrell Grissen (2020). "Behavior Revealed in Mobile Phone Usage Predicts Credit Repayment." *The World Bank Economic Review* 34(3), 618–634.

Blattner, Laura and Scott Nelson (2024). "How Costly Is Noise? Data and Disparities in Consumer Credit."

Blattner, Laura, Scott Nelson, and Jann Spiess (2024). "Unpacking the Black Box: Regulating Algorithmic Decisions." *Proceedings of the 23rd ACM Conference on Economics and Computation*. EC '22: The 23rd ACM Conference on Economics and Computation. Boulder CO USA: ACM, 559–559.

Böken, Björn (2021). "On the Appropriateness of Platt Scaling in Classifier Calibration." *Information Systems* 95, 101641.

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

Butaru, Florentin, Qingqing Chen, Brian Clark, Sanmay Das, Andrew W. Lo, and Akhtar Siddique (2016). "Risk and Risk Management in the Credit Card Industry." *Journal of Banking & Finance* 72, 218–239.

Caire, Dean and Maria Fernandez Vidal (2024). "Leveraging Transactional Data for Micro and Small Enterprise (MSE) Lending."

Chen, Tianqi and Carlos Guestrin (2016). "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA: ACM, 785–794.

Chioda, Laura, Paul Gertler, Sean Higgins, and Paolina Medina (2024). "FinTech Lending to Borrowers with No Credit History."

De Cnudde, Sofie, Julie Moeyersoms, Marija Stankova, Ellen Tobback, Vinayak Javaly, and David Martens (2019). "What Does Your Facebook Profile Reveal about Your Creditworthiness? Using Alternative Data for Microfinance." *Journal of the Operational Research Society* 70(3), 353–363.

Di Maggio, Marco and Dimuthu Ratnadiwakara (2025). "Invisible Primes: Fintech Lending with Alternative Data." SSRN Scholarly Paper 3937438.

Duarte, Victor, Julia Fonseca, Divij Kohli, and Julian Reif (2025). "The Effects of Deleting Medical Debt from Consumer Credit Reports."

Friedman, Jerome H., Trevor Hastie, and Rob Tibshirani (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33, 1–22.

Frost, Jon, Leonardo Gambacorta, Yi Huang, Hyun Song Shin, and Pablo Zbinden (2019). "BigTech and the Changing Structure of Financial Intermediation." *Economic Policy* 34(100), 761–799.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2022). "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *The Journal of Finance* 77(1), 5–47.

Gambacorta, Leonardo, Yiping Huang, Han Qiu, and Jingyi Wang (2024). "How Do Machine Learning and Non-Traditional Data Affect Credit Scoring? New Evidence from a Chinese Fintech Firm." *Journal of Financial Stability* 73, 101284.

Guo, Chuan, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger (2017). "On Calibration of Modern Neural Networks." Version 2. URL: https://arxiv.org/abs/1706.04599 (visited on 02/24/2026). Pre-published.

Hair, Christopher M, Sabrina T Howell, Mark J Johnson, and Siena Matsumoto (2025). "Modernizing Access to Credit for Younger Entrepreneurs: From FICO to Cash Flow."

Huang, Yiping, Zhenhua Li, Han Qiu, Sun Tao, Xue Wang, and Longmei Zhang (2023). "BigTech Credit Risk Assessment for SMEs." *China Economic Review* 81, 102016.

IFC (2024). "Her Fintech Edge: Market Insights for Inclusive Growth."

Ilyas, Andrew, Shibani Santurkar, Dimitris Tsipras, Logan Engstrom, Brandon Tran, and Aleksander Madry (2019). "Adversarial Examples Are Not Bugs, They Are Features." URL: http://arxiv.org/abs/1905.02175 (visited on 05/24/2025). Pre-published.

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

Jagtiani, Julapa and Catharine Lemieux (2019). "The Roles of Alternative Data and Machine Learning in Fintech Lending: Evidence from the LendingClub Consumer Platform." *Financial Management* 48(4), 1009–1029.

Johnson, Mark J., Itzhak Ben-David, Jason Lee, and Vincent Yao (2023). "FinTech Lending with LowTech Pricing." URL: https://papers.ssrn.com/abstract=4396502 (visited on 03/12/2024). Pre-published.

Khandani, Amir E., Adlar J. Kim, and Andrew W. Lo (2010). "Consumer Credit-Risk Models via Machine-Learning Algorithms." *Journal of Banking & Finance* 34(11), 2767–2787.

Lee, Jung Youn, Joonhyuk Yang, and Eric Anderson (2024). "Using Grocery Data to Predict Credit Card Payments." *Management Science*, 1–25.

Lee, Jung Youn, Joonhyuk Yang, and Eric Anderson (2026). "Who Benefits from Alternative Data for Credit Scoring? Evidence from Peru." *Journal of Marketing Research* 63(1), 105–126.

Lin, J (1991). "Divergence Measures Based on the Shannon Entropy." *IEEE Transactions on Information Theory* 37(1), 145–151.

Lu, Jie, Anjin Liu, Fan Dong, Feng Gu, Joao Gama, and Guangquan Zhang (2018). "Learning under Concept Drift: A Review." *IEEE Transactions on Knowledge and Data Engineering*, 1–18.

Massey Jr., Frank J. (1951). "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association* 46(253), 68–78.

Meursault, Vitaly, Daniel Moulton, Larry Santucci, and Nathan Schor (2025). "One Threshold Doesn't Fit All: Tailoring Machine Learning Predictions of Consumer Default for Lower-income Areas." *Journal of Policy Analysis and Management* 44(3), 792–815.

Mienye, Ibomoiye Domor and Yanxia Sun (2022). "A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects." *IEEE Access* 10, 99129–99149.

Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019). "When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications." *Journal of Marketing Research* 56(6), 960–980.

Pearson, Karl (1900). "On the Criterion That a given System of Deviations from the Probable in the Case of a Correlated System of Variables Is Such That It Can Be Reasonably Supposed to Have Arisen from Random Sampling." *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 50(302), 157–175.

Piccininni, Marco, Maximilian Wechsung, Ben Van Calster, Jessica L. Rohmann, Stefan Konigorski, and Maarten van Smeden (2024). "Understanding Random Resampling Techniques for Class Imbalance Correction and Their Consequences on Calibration and Discrimination of Clinical Risk Prediction Models." *Journal of Biomedical Informatics* 155, 104666.

Rishabh, Kumar (forthcoming). "Beyond the Bureau: Interoperable Payment Data for Loan Screening and Monitoring." *Journal of Financial Intermediation*.

Sadhwani, Apaar, Kay Giesecke, and Justin Sirignano (2021). "Deep Learning for Mortgage Risk*." *Journal of Financial Econometrics* 19(2), 313–368.

San Pedro, Jose, Davide Proserpio, and Nuria Oliver (2015). "MobiScore: Towards Universal Credit Scoring from Mobile Phone Data." *User Modeling, Adaptation and Personalization*. Ed. by Francesco Ricci, Kalina Bontcheva, Owen Conlan, and Séamus Lawless. Cham: Springer International Publishing, 195–207.

Siddiqi, Naaem (2017). *Intelligent Credit Scoring*. John Wiley & Sons, Ltd.

Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58(1), 267–288.

Van den Goorbergh, Ruben, Maarten van Smeden, Dirk Timmerman, and Ben Van Calster (2022). "The Harm of Class Imbalance Corrections for Risk Prediction Models: Illustration and Simulation Using Logistic Regression." *Journal of the American Medical Informatics Association* 29(9), 1525–1534.

Yin, Jiangning and Nan Li (2022). "Ensemble Learning Models with a Bayesian Optimization Algorithm for Mineral Prospectivity Mapping." *Ore Geology Reviews* 145, 104916.

Zadrozny, Bianca and Charles Elkan (2002). "Transforming Classifier Scores into Accurate Multiclass Probability Estimates." *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD02: The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton Alberta Canada: ACM, 694–699.