FinRegLab

Advancing the Credit Ecosystem: Machine Learning & Cash Flow Data in Consumer Underwriting

Empirical White Paper

JULY 2025

About FinRegLab

FinRegLab is a nonprofit, nonpartisan innovation center that tests new technologies and data to increase access to responsible financial services that help drive long-term economic security for people and small businesses. With our research insights, we facilitate discourse across the financial ecosystem to inform market practices and policy solutions.

Acknowledgments

This empirical white paper expands on FinRegLab's prior quantitative research on the use of cash flow data for credit underwriting in both consumer and small business markets and managing explainability and fairness concerns in connection with machine learning underwriting models. These prior reports, along with related data science, market, and policy analyses, are available at https://finreglab.org/category/projects/.

Support for this publication was provided by JP Morgan Chase & Co. and Capital One. Detailed information can be found on the inside back cover.

We would like to thank members of our project advisory board: Jay Budzik, Fifth Third Bank; Vickey Chang, Equifax; and Conrod Robinson, FinRegLab Advisor. Keith Ernst, Federal Deposit Insurance Corporation, also participated as a regulatory observer.

Thanks also to the following individuals who provided feedback in connection with this project and the report:

Erin Allard and Brian Duke, Prism Data
Charlie Costello, Alicia Dagosta, and Scott Weber, Upstart
Marsha Courchane and Steli Stoianovici, Charles River Associates
Hunter Hao, Lydia Huo, and Mike Petkun, Nova Credit
Esther Kahng and Sean Kamkar, Zest AI
Scott Nelson, University of Chicago Booth School of Business
Nicholas Schmidt, Solas AI
Sanjana Shellikeri, BLDS
David Silberman, FinRegLab Advisor
Michael Umlauf, TransUnion

Lastly, we would like to acknowledge Zishun Zhao for conducting the empirical analysis, Zishun Zhao and Kelly Thompson Cochran for writing this report, and other members of the FinRegLab team who worked on various elements of this research project, including Ali Bagherpour, Drew Bluethmann, and Sarah Davies.

When viewed with an Adobe Acrobat reader, elements listed in the Table of Contents or in blue text are links to the referenced section or feature. Functionality may be limited in non-Adobe readers. Adobe's reader can be downloaded for free at get.adobe.com/reader.

CONTENTS

1. EXECUTIVE SUMMARY

Consumer lending in the United States is undergoing its biggest evolution in decades, as lenders increasingly adopt machine learning (ML) techniques and new data sources to improve existing automated underwriting systems. These changes have the potential to deliver substantial increases in predictive accuracy and credit access among populations that are difficult to underwrite using traditional models and data sources, but they require appropriate risk management and investment in processes and systems changes. As a result, current adoption levels are uneven and stakeholder attitudes are mixed. Deciding whether and how to implement these changes can be daunting for small lenders, particularly in the face of conflicting research about their impacts. Attitudes among policymakers and consumer advocates also range from enthusiasm to substantial skepticism in trying to ensure that risks are mitigated and offset by concrete benefits.

To advance market practices and public policy discourse on these issues, this study assesses the impacts on model predictiveness and credit access of both adopting machine learning techniques and incorporating electronic bank account information (often called cash flow data) in consumer underwriting models.[1] Machine learning techniques can potentially detect more nuanced patterns in input data than traditional approaches, while information from checking, prepaid, and other accounts can provide insights into consumers' income, reserves, and expenses that do not appear in traditional credit reports.

We evaluated the two innovations by building a series of models based on an anonymized dataset that combines data from one of the main three nationwide credit bureaus with bank account information compiled by a data aggregator, and then comparing the models' predictions against actual credit performance on new accounts opened in 2018-2019. Our evaluation includes models that use only cash flow data, only credit bureau data, and "hybrid" models that combine bank account information with either a traditional credit score or credit bureau data. To allow a full set of comparisons, we built models for each data type using both logistical regression (LR) techniques, the most common traditional approach for consumer loan underwriting scorecards, and XGBoost, a widely used ML technique for building underwriting models.

1.1 Overview of results

Although our data and models are subject to certain limitations as discussed further below, we nonetheless find meaningful benefits for both the analytical and data innovations separately and even larger impacts when they are combined.

1.1.1 Machine learning vs. logistic regression models

- The machine learning models increased predictiveness relative to logistic regression models across all data sources, with particularly substantial improvements of about 2% in a commonly used metric called Receiver Operating Characteristic Area Under the Curve (ROC-AUC) for the models that used full data sets (cash flow data only, credit bureau data only, or the two sources combined). Prediction improvements were modest but still statistically significant for a machine learning model that combined cash flow data with a traditional credit score.
- In simulations at a conservative risk threshold that is likely to be used by mainstream lenders, the ML models built with credit bureau data (alone or when combined with cash flow data) both increased overall credit approvals by nearly 4% and reduced the percentage of defaulters that were approved by at least 9% compared to analogous LR models. To provide a sense of scale, in 2023 about 55 million new credit card accounts and 3.8 million new mortgages were originated using similar risk cutoffs. A 4% increase in borrowers who are deemed to be creditworthy at those thresholds would work out to roughly two million additional credit card accounts and 152,000 additional mortgages.[2]
- At higher risk thresholds likely to be used by near prime or subprime lenders, the ML models' impacts on overall approval rates became progressively smaller or negative relative to LR versions, but they made even larger percentage reductions in approvals of consumers who went on to default.

1.1.2 Hybrid models vs. models built with credit data only

- Adding cash flow data to traditional credit scores or credit bureau data produced modest increases in predictiveness compared to the credit only baseline models, including both logistic regression and machine learning models. For example, the hybrid models' ROC-AUC metrics were roughly .3% to .7% higher than the credit only baselines. The impacts were larger for models built with a traditional credit score than for models built with a range of credit bureau data.
- Simulations showed impacts on credit access that were similar directionally to the ML results but smaller in magnitude. At a conservative risk threshold that is likely to be used by mainstream lenders, the hybrid models increased credit approvals by a range of .6% to 1.6% when compared to credit only baseline models. At a high risk threshold likely to be used by higher cost lenders, impacts on approval rates were smaller or negative but the hybrid models reduced approvals of consumers who went on to default.

1.1.3 Impacts on subgroups and overall

- Across all models built for the study, the hybrid ML model that combined credit bureau data with cash flow data was the most predictive overall and across all subgroups. It also had the highest approval rates overall and for most subgroups at most risk thresholds, while also producing relatively low false positive rates (approvals of consumers who went on to default). The ML model built using only credit bureau data generally ranked second on both predictiveness and access.
- The ML and hybrid models' performance improvements benefitted all subgroups differentiated by past credit history, income level, and race/ethnicity. The models' impacts on credit access varied somewhat based on credit history and income levels. Increases in credit approval rates were particularly large and consistent across different thresholds for consumers with recent derogatory history, while reductions in approvals of consumers who went on to default were particularly large and consistent for low-to-moderate income populations.

FIGURE 1 INCREASES IN MODEL PREDICTIVENESS RELATIVE TO TRADITIONAL BASELINES

Figure 1 reports the ROC-AUC for a baseline model (either a logistic regression model built only with credit bureau data or a traditional credit score) in the top left cell of each grid. The other cells show the increase in ROC-AUC for models built with cash flow data (right column) and models built with machine learning techniques (bottom row). Models that combined both innovations (bottom right) are the most powerful. All of the increases are statistically significant.

Models built with credit bureau data

	CREDIT DATA ONLY MODELS	HYBRID (CREDIT + CASH DATA) MODELS
LOGISTIC REGRESSION MODELS	0.8653 (Baseline)	0.0029
MACHINE LEARNING MODELS	0.0178	0.0201

Models built with traditional credit score

	CREDIT SCORE ONLY	HYBRID (SCORE + CASH DATA) MODELS
LOGISTIC REGRESSION MODELS	0.8662 (Baseline)	0.0042
MACHINE LEARNING MODELS	NA	0.0061

FIGURE 2 INCREASES IN CREDIT APPROVAL RATES AT 3% RISK CUT OFF RELATIVE TO TRADITIONAL BASELINES

Figure 2 presents the results of a simulation using a conservative risk cut off that is most likely to be deployed by mainstream lenders. Credit approval rates for the traditional baseline models are presented in the top left cell of each grid. The other cells show the percentage point increase in approvals for models built with cash flow data (right column) and models built with machine learning techniques (bottom row). Among models built with credit bureau data, the ML hybrid model produced the largest gains. Among models built with a traditional credit score, the LR hybrid model produced the largest gains. All of the increases are statistically significant.

Models built with credit bureau data

	CREDIT DATA ONLY MODELS	HYBRID (CREDIT + CASH DATA) MODELS
LOGISTIC REGRESSION MODELS	65.18% (Baseline)	0.5 ppts
MACHINE LEARNING MODELS	2.5 ppts	3.0 ppts

Models built with traditional credit score

	CREDIT SCORE ONLY	HYBRID (SCORE + CASH DATA) MODELS
LOGISTIC REGRESSION MODELS	65.24% (Baseline)	1.0 ppts
MACHINE LEARNING MODELS	NA	0.4 ppts

1.2 Limitations

We note that these results are subject to certain limitations, including:

- As a result of the way that our dataset links available data sources, it has limited numbers of consumers with little or no traditional credit bureau data and is somewhat skewed toward consumers with prime credit scores and larger incomes. This limitation prevents precise estimates of the impacts on consumers who are likely to see the largest gains from using cash flow data for credit underwriting and means that the models are not as calibrated for their experiences as they would have been if we had had access to a more representative sample. We believe that the cash flow results would have been stronger with more observations of consumers with little or no credit history, given that they are often rejected or treated as extremely high risk by lenders using traditional models, but cannot say that definitively based on the available data.
- Because of sample limitations related to the COVID-19 pandemic, we built the models using a 12-month performance period and did not have immediate access to a comparable sample of data from a different, stable time period to use in validating the models. Instead, we divided the data to hold back observations in certain origination months to use solely for validating the models that had been built with the remaining data.
- Although we spent substantial time on developing input features based on cash flow data using a variety of approaches, our techniques and uses of the data may differ from other model developers and researchers. While credit bureau data is highly structured and familiar to model builders after decades of use, raw electronic cash flow data is less standardized. Accordingly, practices vary as we and other model builders continue to refine tools for processing the data and distilling predictive insights.
- The ML models described in this report have been built using large numbers of variables with an eye toward maximizing predictive performance. While some lenders use machine learning models with similar numbers of variables, others opt to use a much smaller number of inputs for a variety of reasons as discussed further below.

1.2.1 Contributions to market practice and policy discourse

We believe these results make important contributions in light of the current state of flux in the market and in policy debates:

- **The results help to document the benefits of moving beyond status quo approaches to credit underwriting.** Continuing to use traditional models and data is in some ways the easiest option for lenders to select, since it does not require updating systems or processes. But existing systems do not predict default risk equally well for all groups of consumers, leaving lenders with blind spots where they are more vulnerable to risk or more conservative than they need to be in approving applicants who are in fact creditworthy. Although the sample is limited in certain respects, the analysis illustrates how increases in predictive accuracy can help both to expand credit access and reduce the number of consumers who obtain loans that they are not likely to succeed in repaying.

  Beyond these immediate impacts, having more accurate models that help lenders reduce losses and act with more confidence in underwriting consumers who they previously would have considered to be too high risk can have other positive downstream effects on credit access by prompting changes to lenders' marketing and pricing strategies. While we do not model those additional impacts due to data and resource constraints, our results emphasize the multiple benefits that can accrue to lenders and borrowers from more accurate credit models.

» **The results constitute the first public research that systematically compares varying data sources and analytical techniques, separately and together.** Prior research has concentrated primarily on one innovation or the other. While the majority of such studies have found increases in predictiveness, studies of ML models have reached differing conclusions concerning impacts on credit access. Although we cannot fully evaluate the impacts on consumers with little or no traditional credit history due to data limitations, constructing a careful set of comparisons across data and model types is still helpful to understand their respective impacts and contributions.

» **The results analyze the benefits of creating staged improvements for lenders who are reluctant to make both data and analytical changes at the same time.** Some lenders are adopting ML models but are focusing only on traditional data sources because they are worried about the compliance challenges involved in managing both adjustments at once. Smaller lenders may be nervous about their ability to manage ML models of any type and may prefer to incorporate traditional credit scores because they are validated on large representative populations, but may be open to using cash flow variables to get a more holistic picture of how applicants manage their finances. The results probe the potential benefits of various data and model combinations to help lenders consider which progression of changes may make the most sense for their situations.

» **The results underscore that combining both innovations produces the largest overall improvements in reducing credit defaults and increasing approvals of creditworthy consumers relative to traditional approaches.**. They also suggest that investing further resources to identify best practices for risk mitigation and facilitate responsible adoption of both ML analytics and cash flow data could have substantial benefits for consumers, lenders, and the broader national economy.

Toward that end, we believe that this report makes a second contribution that may be of use to practitioners, researchers, regulators, and other stakeholders by providing a detailed online appendix describing how the models were developed in light of general industry modeling practices, data management issues, compliance considerations, and other topics. While there is substantial variation in industry practices and we have not replicated every element of what practitioners would typically do in developing, validating, and preparing models for deployment, we believe this practically oriented discussion of our methodology may be helpful for audiences who want to understand more about both traditional and machine learning underwriting models and working with both credit bureau and cash flow data. The appendix cannot provide a complete roadmap for in developing models to meet companies' specific business strategies, customer bases, and other unique situations, but it may be useful to informing discussion about the potential adoption of these innovations and to academic researchers who want to build models that are more consistent with industry practice.

The paper is organized as follows: Section 2 summarizes market, research, and policy background that motivates the research. Section 3 describes how the models were built, while Section 4 compares the models based on predictive performance and credit access simulations. Section 5 and Section 6 provide supplemental discussions and the conclusion. Appendix A and Appendix B contain a longer discussion of research literature and additional results relating to models built solely with cash flow data. Appendix C (online) provides a more detailed discussion of model building issues and processes.

2. BACKGROUND

Adoption of machine learning techniques and cash flow data for consumer underwriting models has accelerated over the last several years primarily among large banks and fintech firms, which tend to have the most substantial resources for deploying new technologies. The changes are driven in significant part by recognition of the limitations of continuing to rely on prior generations of automated underwriting systems, which depend primarily on historical data collected by three nationwide credit bureaus and on logistic regression models that tend to be relatively inflexible in their assumptions about relationships between variables and subpopulations. Traditional automated systems had substantial advantages over more subjective judgmental underwriting systems when adoption first accelerated in the 1970s and 1980s, including the ability to provide faster, more consistent, and more nuanced predictions of default risk. While data sources and modeling practices have evolved to some extent over time, further improvements in predictiveness offer the potential to reduce losses for lenders and increase credit access among consumers who are hard to underwrite accurately using traditional approaches.

For example, traditional third-party scoring models often have a more difficult time predicting risk for consumers who do not have substantial recent experience with traditional credit products. Roughly a quarter of U.S. adults are estimated to have thin or no traditional credit files and another quarter to have "subprime" scores due to past delinquencies or periods of economic stability (See Box 2.1). These factors can affect both the price and availability of credit and are more likely to impact young adults, low-to-moderate income households, and Black and Hispanic consumers.[3]

These limitations make both machine learning techniques and cash flow data appealing because they have the potential to both identify additional creditworthy applicants who might otherwise be overlooked and help to detect default risks that are underestimated by traditional data and modeling approaches. When accessed electronically from banking platforms at consumers' direction, cash flow data can offer detailed and timely insights into after-tax income, reserves, and expenses that are not consistently and directly reflected in traditional credit reports since those focus primarily on prior use of mainstream credit products.[4] Machine learning models offer greater flexibility in detecting more nuanced interactions in input data, identifying patterns among smaller subgroups of consumers that may get drowned out in more simplified models, and managing larger numbers of input variables.[5] Essentially, a traditional logistic regression model that is separated into separate "scorecards" for a handful of key customer segments can be thought of as a two-layer model, whereas a machine learning model can add several additional layers to map additional patterns in the data.

BOX 2.1 ESTIMATING THE NUMBER OF CONSUMERS WHO STRUGGLE TO ACCESS CREDIT

Estimates of how many consumers are likely to struggle in accessing mainstream credit vary depending on economic conditions, the credit scoring model used, and other factors. Statistics published in 2015 by the Consumer Financial Protection Bureau have been widely cited for the last decade regarding the number of consumers who do not have credit history at the three major national credit bureaus (often called no files or credit invisibles) and the number of consumers who have credit files but insufficient history to be scored by particular third-party models (often called unscored).[6] The CFPB in 2025 announced a substantial adjustment to those estimates and updated numbers based on 2020 data suggesting that the number of consumers without credit scores has dropped by approximately a third. Nevertheless, concerns about access to credit remain substantial and ongoing.[7]

The CFPB's 2025 report factored in supplemental data that had been inadvertently omitted from its original credit bureau sample and made methodological adjustments. The changes did not substantially alter the original estimates of how many U.S. adults overall lacked credit scores as of 2010 (moving from 19.3% to 18.4%), but resulted in a substantial reduction in the percentage of consumers classified as credit invisible (11% to 5.8%) and a substantial increase in the percentage of consumers classified as unscored (8.3% to 12.7%). The CFPB also provided calculations based on December 2020 data estimating that about 2.7% of U.S. adults were credit invisible and 9.8% were unscored, for a total of 12.5%. Future reports are expected to analyze changes across different consumer groups and locations.

Comparing the two reports suggests that the overall number of consumers who could not be scored has dropped from about 45 million in 2010 to about 32 million in 2020. While this is a substantial reduction, it is not the only category of consumers who may struggle to access mainstream credit. For example, lenders may impose additional underwriting requirements or use different fraud, approval, or pricing models for consumers who are scoreable under third party models but are also considered "thin file" because their records indicate two or fewer credit accounts. Equifax estimates from 2022 indicate that 62 million consumers were thin file (regardless of whether they could be scored by particular models); if added to the CFPB's 2020 estimates for the number of consumers without credit files, this suggests about 69 million (26.6%) consumers are thin or no file.[8]

Consumers whose scores fall outside of the "prime" range—particularly those classified as subprime—due largely to past delinquencies, bankruptcies, and other derogatory history may also be rejected for credit or charged higher prices. This number fluctuates but was estimated to be about 25% as of 2022.[9]

Together or apart, the two innovations have increased hopes of achieving benefits that could benefit borrowers, firms, policymakers, and investors alike:

- Expanding access to more borrowers who are creditworthy and reducing the number of consumers who are offered credit on terms that they are unlikely to be able to repay,
- Reducing default rates and losses,
- Reducing mispricing based on inaccurate estimation of the likelihood of default and improving terms at which credit is offered to some applicants,
- Improving identification and mitigation of certain forms of discriminatory lending, and
- Facilitating less costly and faster model generation and updating.

At the same time, adopting these innovations requires model builders and lenders to consider a number of potential business and compliance risks and to invest in changes to systems and processes to manage them.[10] Lenders, vendors, and regulators have become increasingly comfortable managing traditional models and data sources over several decades, so individual lenders frequently weigh potential transition costs against the potential benefits from changes in data or analytics to decide whether to proceed with adoption. These decisions are affected not just by business strategy considerations but by the potential impacts on customer relationships and compliance functions, since consumer lending is regulated more heavily than many other aspects of financial services due to the potential scale of its impacts on consumers, lenders, and the broader economy. Specific legal requirements include:

- **Adverse action disclosures:** The Equal Credit Opportunity Act requires lenders to provide individualized disclosures to credit applicants of the "principal reasons" for rejecting an application, terminating a credit line, or taking certain other adverse actions.[11] Where lenders rely on information from credit reports, the Fair Credit Reporting Act similarly requires them to provide consumers with a list of "key factors" that are negatively affecting their credit scores if the score was a factor in an adverse action or prompted the lender to charge higher prices.[12] These requirements (which we refer to collectively as adverse action requirements) force lenders to analyze which input variables play the biggest role in generating predictions for individual applicants.
- **Fair lending compliance:** The Equal Credit Opportunity Act (and Fair Housing Act in the mortgage context) prohibits discrimination on the basis of race, ethnicity, gender, and other protected characteristics in credit decisions.[13] These laws have been interpreted to prohibit both inconsistent treatment across protected classes ("disparate treatment") and the use of facially neutral criteria that have a disproportionately adverse impact on protected groups, unless the criteria further a legitimate business need that cannot reasonably be achieved through less impactful means ("disparate impact"). Lenders may scrutinize both the impact of their models as a whole and of individual input variables as part of fair lending compliance programs.[14] The Trump Administration has announced that it intends to eliminate disparate impact liability.[15]
- **General risk management and model governance:** To protect the safety and soundness of banks and the broader financial system, banks are expected to implement risk-based governance mechanisms for the development, deployment, and monitoring of underwriting and other predictive models. These processes include analyzing whether the models are relying on relationships in the data that are "conceptually sound" and assessing models' stability in changing data conditions.[16] Because approaches that have been developed in the context of logistic regression models do not always translate smoothly to machine learning models, lenders are having to evolve their practices to meet these regulatory expectations.
- **Third party risk management:** Under the Bank Service Company Act, where banks rely on a third party to design, develop, or operate tools or processes that are part of their lending operations, they are generally responsible for overseeing the vendor's compliance with applicable regulations. The Consumer Financial Protection Bureau has historically applied similar expectations to financial services providers that it supervises. These requirements include both due diligence in initial selection as well as ongoing monitoring of performance.[17] However, they can be particularly challenging for smaller institutions, who are more likely to turn to vendors for assistance in developing and deploying underwriting models because they face technology and other resource limitations in the first instance.

Particularly in light of these compliance requirements, lenders must generally update their processes and systems for risk and compliance management when adopting models that involve new data sources or more complex analytics. These changes are designed to test whether new underwriting models will perform as expected with regard to general predictiveness, fairness, and credit access, as well as to meet disclosure and vendor management requirements. The degree of changes required depends on a number of factors, but is often particularly large in the context of moving to machine learning underwriting models because such models provide a greater range of choices as to how they are constructed and often increase complexity relative to traditional techniques. These differences can help to improve their predictiveness but if the models are not developed and deployed thoughtfully, they can increase risks that performance will deteriorate rapidly in changing conditions, that the models will replicate or even exacerbate biases in training data, and that consumers, lenders, and regulators will have a more difficult time understanding the underlying models. Similarly, lenders frequently engage in compliance and validation testing and make adjustments to their processes for generating adverse action notices before adding new data elements to their underwriting models.

As summarized further in Appendix A, available public research about these two lending innovations is limited. FinRegLab's prior research and a few other published analyses suggest that incorporating cash flow data into credit underwriting can have substantial benefits for model predictiveness and credit access.[18] A number of academic studies have documented predictiveness benefits from applying machine learning techniques to traditional credit bureau data,[19] and a study conducted by FinRegLab and researchers at Stanford University evaluated advances in data science techniques that can be used to manage concerns about the explainability and fairness of ML models.[20] However, studies analyzing the potential impacts on credit access from ML models built solely with traditional data have produced a wide range of conclusions, including findings that they exacerbated gaps between some subgroups (Fuster et al, 2022), that they had relatively little impact because they could not overcome gaps in traditional data (Blattner and Nelson, 2024), and that they substantially increased credit access (Albanesi and Vamossy, 2024).[21]

Given continuing uncertainty among smaller lenders and various other stakeholders about the potential benefits and risks of these innovations, FinRegLab embarked on this research to build a series of credit underwriting models that allow a systematic comparison of the impacts of deploying models with or without cash flow data and with or without machine learning techniques relative to models built solely with traditional data sources and analytical approaches. To ensure the study's relevance to the current state of consumer lending, these models have been developed by a team of data scientists with many years of experience in developing and validating models in the retail banking industry. The team has drawn on the expertise of an advisory board consisting of industry leaders with expertise in machining learning, risk management, financial regulation, commercial credit score development, and consumer credit reporting.

The main objectives of this study are as follows:

- Examine how incorporating cash flow features affects the predictiveness of both traditional models and machine learning models.
- Compare the predictiveness of traditional and machine learning models applied to the same data sources, including cash flow data, consumer bureau data, and the two sources combined.
- Analyze the potential impacts of the models on predictiveness and credit access for different subgroups of consumers, including consumers with different credit histories, incomes, and races/ethnicities.

3. OVERVIEW OF MODEL DEVELOPMENT

A more detailed description of our model development process is provided in Appendix C (online) for readers who prefer additional technical detail and a more extended discussion of general modeling practices and specific considerations in working with cash flow data and machine learning models.

3.1 Data

3.1.1 Data sources

The data used in this study is derived from a comprehensive anonymized dataset compiled by one of the three nationwide credit bureaus for use by third-party score developers. It combines historical credit bureau data with bank account information from a large data aggregator for substantial numbers of consumers who opened a new credit account between April 2018 and March 2019. To build the models, we used a traditional credit score or detailed historical records provided by the credit bureau (including traditional credit metrics such as payment behavior, credit utilization, and account types), bank account transaction and balance data from the aggregator (which offers insights into income, expenses, and balances that are not directly reflected in credit bureau records), or both sources, up to the month before the new account opening. For the dependent variable and validation of the models, we used performance data on the newly opened account from the first 12 months after origination. These new tradelines provide a clear and objective measure of credit performance, such as the occurrence of serious delinquencies (90 days or more), charge-offs, or bankruptcies within the specified time frame.

While this dataset provided a robust foundation for our analysis, due to the way it was constructed it is not a nationally representative sample of consumers as discussed further below. The dataset was constructed starting on the data aggregator's side by focusing on consumers whose information included at least one bank account and one loan account as of April 2019. The loan account was used to link the aggregator data to the consumer's credit bureau profile using a hashing protocol to protect privacy. For simplicity, ambiguous matches—such as cases involving joint ownership of the loan used for linking—were excluded, leaving only unique one-to-one matches between the cash flow data and the credit bureau data. The final dataset included 750,266 consumers with new credit originations during the sample period, although we excluded a number of observations from the sample as described in Section 3.1.2.[22]

3.1.1.1 Credit bureau inputs

The credit report data provided by the credit bureau can be divided broadly into two categories: credit attributes/features that summarize a consumer's credit profile as of a snapshot date (Snapshot Attributes) and attributes/features that summarize variations across time for the same consumer (Trended Attributes) over a 24-month period up until the snapshot date. The combination of snapshot and trended attributes provides a comprehensive view of a consumer's credit history and behavior, and one or both types of attributes are widely used in traditional credit scoring and underwriting models.

In addition to the credit attributes/features, we obtained a conventional credit score (referred to as the credit score or CS) for the consumers in the sample calculated using credit bureau inputs as of the month-end before the new origination. We used these scores as inputs for some hybrid models and as benchmarks for assessing the performance of the models we built for the study.

3.1.1.2 Bank account history

The data aggregator files include transaction and balance history for various types of bank accounts, including checking, savings, certificates of deposit, money market accounts, flexible spending accounts, health spending accounts, and prepaid cards, covering the period from April 2016 to February 2019. Not all consumers may have linked all of the accounts they owned. The account history files for the individual accounts provide the account type and periodically refreshed balances, along with the refresh dates. The transaction history files include the transaction type, date, amount, currency, and a description that is provided by the underlying financial institution (such as the transaction description that appears on a checking account statement).

The aggregator also provided its own categorization of the transactions, based on models that it has built based on the other data elements from millions of transactions. The aggregator's classification scheme includes 64 categories designed to provide insight into the source or use of the funds involved. For example, a debit transaction could be classified as a purchase, loan payment, rent, etc., depending on the description and parties involved. Similarly, a credit transaction could be classified as a deposit, salary, refund, etc. Such characterization of the sources and uses of funds sheds light on the nature and stability of a consumer's income and liabilities, which can be valuable for predicting credit risk.

3.1.1.3 Implementation of the dependent variable

Since our models are designed to predict the probability of default, it is critical to define and implement a default definition that captures all material credit losses. For this study, we defined a default event as one or more of the following occurrences on the new account (or one of the new accounts) within 12 months of origination:

- **90+ Days Delinquency:** The account has been delinquent for 90 days or more, indicating a significant delay in payment.
- **Charge-Off:** The lender has written off the account as a loss, typically after 180 days of delinquency.
- **Repossession:** The lender has repossessed collateral (e.g., a vehicle) due to non-payment of the account.
- **Foreclosure:** The lender has initiated foreclosure proceedings on a property due to a default on the mortgage account.

- **Bankruptcy:** The account has been included in the borrower's filing for bankruptcy, which may include Chapter 7, Chapter 11, or Chapter 13 bankruptcy.

These events are widely recognized as indicators of severe credit distress and are commonly used by lenders and credit bureaus to assess credit risk.

3.1.1.4 Demographic data

For the purpose of analyzing model performance and its impact on underwriting outcomes across different demographic groups, we also obtained a separate dataset from the credit bureau that includes inferred information about race/ethnicity.[23] This enabled us to evaluate the impact of the model on different subgroups, conditional on the accuracy of the inferred demographic information.

3.1.2 Sampling methodology

As discussed further in Appendix C.2.2, the total sample of consumers with clean one-to-one matches between the credit bureau and data aggregator records included 750,266 loan originations. To further refine the population into a sample that is suitable for addressing our research objectives and representative of the information available for an underwriting model at the time of an application, we excluded some consumer records on three grounds:

- **No Credit Score:** To ensure that a third party credit score was available for benchmarking, we excluded a small number of consumers without a credit score.
- **Low-Quality Bank Account Data:** To ensure the reliability of cash flow data, we excluded consumers whose data aggregator files did not allow reliable analyses of transaction and balance information as described in Appendix C.2.2 and Appendix C.3.1.
- **Insufficient Data from a Primary Checking Account:** To ensure that the cash flow information included a transactional account that was being used relatively heavily on an ongoing basis, we defined criteria to identify primary checking accounts and excluded consumers from the sample if the data aggregator records did not include at least six months of history from one such account as described in Appendix C.2.2.[24] This ensures that any conclusions drawn regarding the effectiveness of cash flow data are representative of a real-world underwriting scenario where bank account data is provided.

TABLE 3.1 SAMPLE WATERFALL

CATEGORY	# OF OBSERVATION	% POPULATION
TOTAL POPULATION	750,266	100.00%
NO CREDIT SCORE	2,588	0.34%
LOW QUALITY BANK ACCOUNT DATA	90,342	12.04%
INSUFFICIENT PRIMARY CHECKING ACCOUNT DATA	232,790	31.03%
SAMPLE FOR MODELS	424,546	56.59%

TABLE 3.2 DATA PARTITIONING

SAMPLE	# OBSERVATIONS	# DEFAULTERS	PROB. OF DEFAULT
OUT-OF-TIME VALIDATION	107,789	2,468	2.29%
VALIDATION	105,520	2,398	2.27%
DEVELOPMENT	211,237	4,706	2.23%

After applying these exclusions, we arrived at a final sample of 424,546 observations for model development and validation. We partitioned the data as follows:

- **Out-of-Time Validation Sample:** Approximately 25% of the data was set aside as a "stratified" out-of-time validation sample. This includes consumers with originations in July 2018, November 2018, and March 2019. This sample was not used in any part of the model development or validation process, and instead was used to calculate the performance and credit access metrics discussed in Section 4 as a measure of the extent to which the models are generalizable to unseen data.
- **Model Development and Validation Samples:** The remaining 75% of the data was randomly split into a model development sample and a validation sample in a 2:1 ratio. The model development sample was used for feature selection, hyperparameter tuning, and estimating the models, while the validation sample was used to validate and compare models before finalization.

While it is a common practice to validate model performance on samples that have not been involved in the model development or validation process, using an out-of-time sample for this purpose is typical among banks but may not be as universally used by other lenders or data scientists in other settings. The approach is designed as an extra check to ensure that results are robust and generalizable, although as discussed in the Section 3.1.3 we faced certain limitations due to the onset of COVID-19 pandemic.

3.1.3 Data limitations

While the dataset used in this study provides a robust foundation for evaluating the impact of cash flow data and machine learning techniques on credit risk prediction, there are several limitations that should be acknowledged. These limitations stem from the constraints of the data sources, the sample selection process, and the timing of the data collection.

3.1.3.1 Limited representation of thin and no-file consumers

Because the sample selection process focused on consumers who were using the data aggregator's services and had both an existing bank account and credit account, the sample includes relatively few consumers whose credit bureau data is so limited that they may have a difficult time being scored by traditional models. These include three primary categories of consumers:

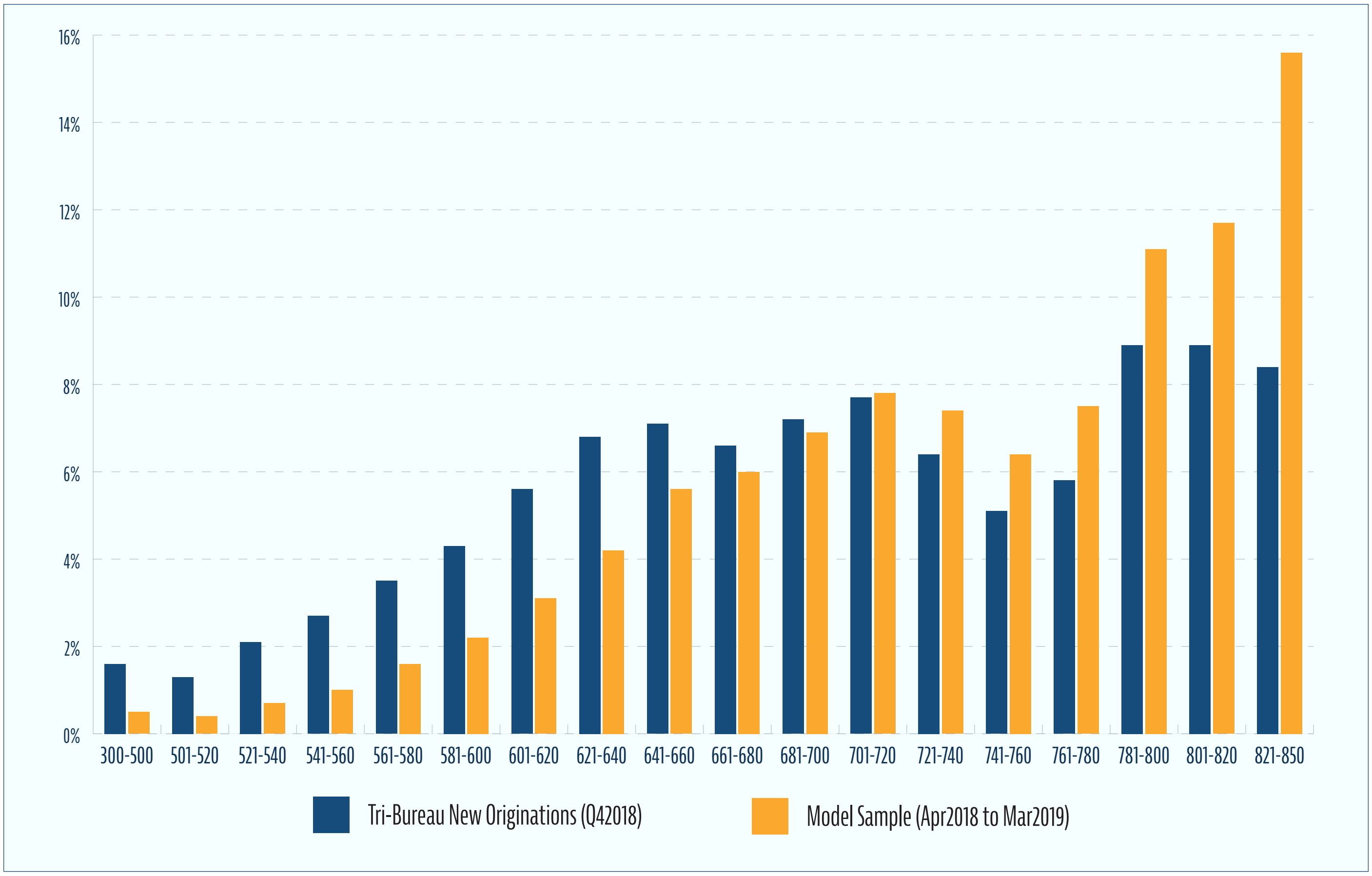
- **No file/credit invisible consumers:** Consumers who do not have a credit profile at all.
- **Consumers who are unscoreable under various third-party models:** The percentage of such consumers differs by model type. For example, FICO's general models will not generate scores for consumers with no trade activity older than 6 months or no trade activity in the past 6 months, while VantageScore will generate scores unless there are no trades, public records, or unpaid collections reported. General scores offered by individual credit bureaus and specialty scores may have other parameters.
- **Thin file:** Consumers with two or fewer credit accounts (often called tradelines) on their credit reports may be scoreable under one or more third-party models, but their likelihood of default is often more difficult to predict than consumers who have more extensive mainstream credit experience. Lenders may subject thin file consumers to special requirements or higher pricing due to risk and fraud concerns.

This limitation restricts our ability to assess the predictiveness and credit access impacts of our models on these specific subpopulations. In particular, it prevents precise estimates of the impacts on consumers who are likely to see the largest positive impacts from using cash flow data for credit underwriting and means that the models are not as calibrated for their experiences as they would have been if we had had access to a more representative sample. We believe that the cash flow results would have been stronger with more observations of consumers with little or no credit history, given that they are often rejected or treated as extremely high risk by lenders using traditional models, but cannot say that definitively based on the available data.

3.1.3.2 Skew toward prime borrowers

The sample selection process also skews the overall sample toward borrowers with prime credit scores and higher incomes compared to the overall tri-bureau new originations population distribution from the same time period. In addition, the exclusion of consumers without a primary checking account slightly added to the skew. Although there is adequate representation across the entire credit spectrum as reflected in Figure 3 to support model development and analysis, the measurement of impacts on populations with higher scores and incomes may tend to be the most precise.

FIGURE 3 SAMPLE DISTRIBUTION BY CREDIT SCORE



3.1.3.3 Validation sample constraints

To ensure that the sample and performance data were unaffected by the economic disruptions caused by the COVID-19 pandemic, we restricted the sample period to originations before March 2019 so that we could observe 12 months of performance. When the sample was compiled, originations from after the end of COVID-19 restrictions had not accumulated 12 months of performance. We therefore did not obtain a separate stable "out of time" sample to use for validation, and instead held back data on consumers whose originations fell in every fourth month of the performance period as described in Section 3.1.2. As a result, the validation sample is nearly contemporaneous with the model development sample, rather than being drawn from a later period as would be typical in a real-world setting.

This stratified out-of-time approach is commonly used by industry modelers faced with similar data constraints for an initial assessment, but nonetheless limits our ability to fully assess the model's performance under changing economic conditions. As noted in Section 1, the use of a stratified out-of-time sample tends to generate higher performance metrics. It may also impact comparisons to traditional credit scores that were developed in a different time period and applied to the study sample.

3.1.3.4 Absence of rejected applicants

Our sample consists only of consumers who opened a new credit account between April 2018 to March 2019 and does not include consumers who experienced rejections on all applications that they submitted during that same period. This introduces selection bias, since we lack evidence as to how those consumers would have performed on new accounts. Lack of data on rejected applicants is a well-documented limitation in observational credit risk modeling, absent experimental or reject-inference adjustments. As discussed further in Appendix C (online), different industry actors take a variety of approaches to reject inference concerns, and it is an area of interest for academic research as discussed in Appendix A. In this case, we believe the practical impact of such bias on our study is mitigated by two factors:

1. **Consumer-level risk assessment:** By analyzing all new originations per consumer, we are able to incorporate performance history for consumers who were rejected by some lenders but approved by others.
2. **Broad credit spectrum coverage:** Our sample includes sufficient representation across all credit score bands, unlike the situations faced by some individual lenders that use a score cutoff that requires them to consider extrapolating into a risk tier for which they have no actual observations.

As discussed further below we also weighted the sample to adjust the ratio of cases that did and did not involve defaults and in performing simulations of the impact on credit access to account for the lack of a representative distribution.

3.2 Cash flow feature engineering

The integration of cash flow data into credit risk prediction models necessitates a robust framework for feature engineering that captures critical dimensions of consumer financial behavior while adhering to regulatory and practical constraints. Unlike credit bureau data, where predictive features (e.g., payment history, credit utilization) have been distilled into standardized metrics over time, bank statement data are provided as raw transactional records. To leverage this data in a predictive model, we first engaged in pre-processing to match transaction and balance information so that we could understand the account history, as discussed further in Appendix C.3.1. We then developed features representing meaningful aggregates of transaction and balance history, focusing on solvency, liquidity, and behavioral patterns.

We took some account of compliance considerations in this process as described below, although we have not conducted all of the analyses that lenders would typically perform for compliance purposes in preparing a new model for deployment. In addition, while we spent substantial time on feature engineering activities, our techniques and uses of the data may differ from other model developers and researchers. Given the relatively short amount of time that developers have been working with electronic cash flow information, practices vary as we and other model builders continue to refine tools and strategies for processing the data and distilling predictive insights.

3.2.1 Basic cash flow features

As described in Section 3.2.1, the dataset included balance history, account types (e.g., checking, savings, certificates of deposit), and transactional records, such as amounts, directions (debit/credit), currencies, descriptions provided by the bank or financial institution where the account is held, and the aggregator's standardized categorizations.[25] We initially developed a basic set of cash flow features using only structured data fields, without using the institutions' transaction descriptions or the aggregator's categorizations. The transaction descriptions are more complicated to process and interpret because their contents are less standardized, and some lenders may prefer not to use those fields for simplicity when generating adverse action disclosures for individual consumers. Three primary categories of features were created to assess both long-term stability measures and shorter-term liquidity.

3.2.1.1 Long-term stability features

Solvency metrics were designed to evaluate long-term financial stability. These included net worth (total balance divided by credit bureau-reported debt), total assets (sum of account balances), and debt-to-asset ratios. Income and expense proxies were derived from median monthly inflows and outflows, respectively, while savings rates were calculated as the ratio of median monthly net outflow to median inflow. Savings and investment metrics, such as balances in certificates of deposit and money market funds where reflected in the data, provided additional insights into financial reserves.

3.2.1.2 Short-term liquidity features

Short-term financial stability was assessed through liquidity indicators. The current ratio (balance divided by median monthly outflow) and expense-to-income ratio (median outflow divided by median inflow) quantified cash flow adequacy. Cash and liquid asset totals were aggregated across accounts to measure immediate resource availability.

3.2.1.3 Other financial features

Behavioral features captured dynamic financial habits and stress signals. Metrics such as time since the first account activity were computed by account type. Transactional activity was quantified through recency, frequency, and ticket size, while affluence was inferred from the number of large purchases. Financial stress indicators included counts of days with negative balances and deviations from historical balance trends. Trend analyses, such as current-to-average balance ratios and percentile rankings of recent balances, provided temporal insights into financial stability.

3.2.2 Advanced feature engineering

While the basic features were exclusively based on structured data fields, we spent additional time and effort to develop an advanced feature set using information derived from either the transaction descriptions or the categorizations provided by the aggregator to identify more nuanced patterns in regular and irregular income and expenses as well as liquidity events. This process involved extensive data analysis to inform reasoned judgments about how to define key concepts and classifications and extract information from less structured data fields with an eye toward ensuring that the features would be verifiable and justifiable, particularly in light of adverse action requirements.

To prevent internal transfers from inflating debt and income calculations, internal transfers were identified and removed by matching debit-credit pairs of identical amounts occurring within ±3 days across a consumer's accounts. This threshold balanced completeness (capturing weekend processing delays) and precision, minimizing over-exclusion.

3.2.2.1 Features related to recurring transactions

Recurring transactions in a consumer's bank statements are of particular interest because recurring credit transactions often represent regular income, while recurring debit transactions typically reflect debt payments or non-discretionary liabilities. This may provide a more nuanced picture of the consumer's capacity to take on additional obligations than consideration of only annual income and debts appearing on traditional credit reports.

Through extensive case reviews of cash flow data, we defined criteria for identifying transactions that were sufficiently similar that we classified them as recurring transactions. To start, we focused on transactions within the same account that occurred at regular intervals (weekly, biweekly, or monthly—common pay and billing cycles) with a minimum frequency (at least one three-month period with at least three similar transactions) and an amount exceeding $50. Three similarity definitions were applied:

- **By Transaction Amount:** Transactions with identical amounts. In addition, we created a separate version by excluding internal bank transfers and fees.
- **By Transaction Description:** Transactions descriptions that indicated they involved the same source, excluding bank transfers and fees.
- **Hybrid Criteria:** Transactions meeting either the amount- or description-based similarity criterion, with bank transfers and fees filtered out.

3.2.2.2 Classification of income and expenses based on aggregator transaction categorization

We also used the aggregator transaction categorizations as a second approach to distinguishing between the regularity of income sources and the degree of discretion over types of expenses. For example, we constructed debit transaction definitions based on the aggregator categories that were structured to progress from core obligations (e.g., loans, rent) to expanded categories (e.g., utilities, insurance), and indirect obligations (e.g., checks, credit card payments) to all outflows. This approach moved from narrowly defined, highly predictive categories with limited transaction coverage to broader, more inclusive definitions that encompassed most transactions, albeit with some ambiguity. To improve signal-to-noise ratios, we applied a $50 threshold to exclude insignificant transfers and checks. Similarly, income proxies ranged from conventional sources (e.g., salaries, retirement benefits) to expanded, less regular categories (e.g., tax refunds), to total inflows.

3.2.2.3 Liquidity events

We also used the financial institutions' description fields to identify indications of overdrafts and insufficient funds (NSF) through keyword patterns (e.g., "NSF FEE"). Recognizing the widely different practices in the frequency of a check being re-presented and the amount of fees charged by different financial institutions, we only count the number of days an NSF event occurred and the time since the last NSF event to minimize idiosyncratic information content in these features. This information complemented the liquidity signals we extracted from the first stage of feature engineering by analyzing raw balances and other structured fields. The first stage information alone may not be sufficient to detect returned checks or overdrafts that are covered from linked accounts.

3.2.3 Final feature set

For this study, we ultimately decided to use both the simple and advanced features, computing them across account types (checking, savings, other) and various time windows (30–365 days). Key variables included net flow (credits minus debits), credit/debit sums, and ratios of recurring debt payments to income. Liquidity stress was further quantified through recency metrics, such as days since the last overdraft or sub-threshold balance. The feature engineering process yielded a comprehensive suite of 1,976 features.

Both sets of features were also made available for the cash only and hybrid logistic regression models to select from. Following the process described in Section 3.3.1, the final models typically use 10 to 20 features covering areas such as low or negative balance events, stability of cash inflows, balance trend over time, number and amount of recent cash outflows, and debt-to-income proxies.

3.3 Model development

To address the research objectives, we developed eight distinct models, each employing different methodologies and input configurations. These configurations included: (1) cash flow data only, (2) cash flow data combined with a traditional credit score, (3) cash flow data combined with credit bureau data, and (4) credit bureau data only. For each configuration, we constructed both a logistic regression model—a traditional approach widely used in consumer loan underwriting—and an XGBoost model—a machine learning algorithm known for its ability to capture complex, non-linear relationships. We used the traditional credit score as a general benchmark for comparative analysis across all of the models.

Following a common practice to balance the classes in the sample, we applied sample weights to adjust the ratio of "good" to "bad" cases to 3:1 during the model development process for both LR and ML models. This approach is standard in credit risk modeling, as it enhances the model's ability to distinguish between default and non-default cases while maintaining stability during training. The 3:1 ratio strikes a balance between emphasizing the minority class (defaults) and leveraging the majority class (non-defaults), aligning with the business objective of minimizing default-related losses.

3.3.1 Logistic regression models

The development of logistic regression models followed generally established practices in retail banking, emphasizing statistical validity and parsimony, which refers to the principle of creating a model that is as simple as possible while still effectively capturing the relationships between the predictors and the target variable.[26] A key aspect of this process was segmentation, which builds separate mini-models or scorecards for key subpopulations based on risk profiles, such as for consumers who have filed for bankruptcy, consumers with thin files, or consumers who have never been delinquent. Segmentation is frequently used in building traditional underwriting models for a mix of operational and statistical considerations. Operationally, it allows for tailored model specifications that are simpler and more interpretable to major subgroups. Statistically, it accounts for the possibility that the same feature may have different effects across segments, reducing the need for complex interaction terms.

For models using credit bureau data, we adopted a segmentation scheme similar to those used by many third-party credit scores. This scheme categorizes consumers based on credit history characteristics, such as the presence of derogatory events, credit utilization, and trade age. For example, consumers with a history of severe delinquencies or bankruptcies were modeled separately from those with clean credit histories, as the drivers of risk differ significantly between these groups. For both the credit only and credit + cash hybrid LR models, the population with sufficient credit bureau information were divided into 5 segments. The hybrid LR model has an extra cash only segment to cover the population with very limited credit bureau information.[27] This approach ensures that the models are both interpretable and aligned with industry standards.

In contrast, for the other two logistic regression models (cash flow data only and cash flow data + credit score), we did not segment but rather built models covering the entire sample for each configuration. This decision was driven in part by the fact that all cash flow features tend to be densely populated for all consumers, so there was no apparent gain in parsimony from creating separate segments only for subgroups of consumers that have certain data elements that the majority lack.[28] Neither was there compelling evidence in support of segmentation for performance gain, as will be seen in the comparison results between the logistic regression and XGBoost models for different configurations.

The logistic regression development process included several key steps: exploratory data analysis (EDA), feature selection, and model finalization. EDA involved identifying and addressing missing values, outliers, and feature distributions, ensuring the robustness of the input data. Feature selection employed a two-step process, combining LASSO regularization for initial screening with backward elimination based on p-values as described in greater detail in Appendix C (online). Model finalization involved manual refinement of features, including transformations to address non-linearity and adjustments to missing value treatments, ensuring that the final models were both interpretable and robust.

3.3.2 XGBoost models

Unlike logistic regression, which relies heavily on manual feature engineering and model specification, XGBoost is designed to automatically capture complex, non-linear relationships and interactions between features. This makes it particularly well-suited for high-dimensional datasets. Since the learning algorithm is largely automated, the development process for XGBoost models shifts focus away from manual intervention during estimation and instead prioritizes three key areas: (1) ensuring the relevance of input features, (2) optimizing hyperparameters to suit the specific learning task, and (3) implementing early stopping to prevent overfitting and safeguard generalization performance.[29]

Because most of the input features from either the credit bureau side or the cash flow side are relevant for predicting credit risk by design, we allowed the XGBoost models to select as many variables as possible. The only features excluded from the candidate set were certain age-related features in the credit bureau data that can only be used under certain limited conditions under the Equal Credit Opportunity Act.[30] This approach, which includes all eligible features without prior filtering, is often used to explore the full predictive potential of machine learning models. We note that some practitioners conduct feature selection to narrow down the set of candidate features on which an XGBoost model to be trained. For example, weak features—those with low correlation to the dependent variable—are often removed to reduce the risk of overfitting to noise in the data. Similarly, redundant features, which are highly correlated with one another or with a group of features, are excluded to create simpler and more interpretable models. However, our primary objective was to study the predictive potential of machine learning models, particularly in the context of integrating cash flow data with traditional credit bureau attributes. During experimentation, we observed that removing the least predictive features generally resulted in worse performance on unseen data. As a result, we decided to retain all eligible features as candidate inputs, allowing the XGBoost algorithm to select the most relevant features during training. While this approach may reduce interpretability, it aligns with our goal of maximizing predictive performance and ensuring that the models generalize well to unseen data.

Hyperparameter optimization was conducted using the Tree-structured Parzen Estimator (TPE) method, a Bayesian optimization technique that iteratively refines hyperparameter selections. Key hyperparameters included L2-regularization (lambda), maximum tree depth, minimum child weight, and column and subsampling rates. The optimization process involved 300 trials, evaluated using 4-fold cross-validation on the model development sample, with the Receiver Operating Characteristic Area Under the Curve (ROC-AUC) as the primary performance metric.

The final models were trained with a reduced learning rate (0.01 or 0.05) and an early stopping rule to ensure robust convergence. This approach, while computationally intensive, minimizes the risk of overfitting and enhances generalization to unseen data. The resulting models contained thousands of decision trees, but real-time scoring remained feasible with modern computational infrastructure at negligible cost.

TABLE 3.3 KEY HYPERPARAMETERS FOR THE FINAL MODELS

HYPERPARAMETER	CASH ONLY	CS + CASH	CREDIT + CASH	CREDIT ONLY
NUMBER OF TREES	8,539	456	1,704	1,414
LEARNING RATE	0.01	0.01	0.05	0.05
LAMBA	2,958	3,224	3,319	4,041
MAX DEPTH	3	7	7	7
MINIMUM CHILD WEIGHT	18	3	1	2
COLUMN SAMPLING RATE	0.16	0.77	0.50	0.06
SUBSAMPLING RATE	0.25	0.92	0.89	0.92

***

Additional background is provided in Appendix C (online) on these processes and parameters. While we have worked to follow industry practice in various aspects of model development, it is important to note that those practices are continuing to evolve particularly in the context of developing machine learning models. There can be variations in practice between institutions—including general differences between banks and non-banks on certain topics such as the number of inputs included in machine learning models. For this research, we did not constrain the number of inputs to the degree that banks often do or conduct the full range of validation analyses and compliance related testing of the models that lenders frequently perform prior to deployment.

4. EMPIRICAL ANALYSIS

4.1 Model evaluation methodology and metrics

To address the research objectives, we calculated various metrics relating to predictiveness and credit access for each model as a whole and for various subgroups as described in this subsection, using the out-of-time validation sample. We then compared the various models' results to each other and to the traditional credit score for models built with that score as a component. The results are presented in Section 4.2, which separately discusses the impact of each innovation on performance and credit access:

- **Cash Flow Data:** We compared the metrics for the various models built with and without cash flow data, reporting results separately for logistic regression models and for machine learning models. The analyses of the various hybrid models built with both data sources are benchmarked to a comparable credit only model (including a traditional credit score where it was a component of the hybrid model).
- **Machine Learning Analytics:** For each data type, we compared the metrics for the machine learning model against the metrics for the logistic regression model built.

The performance statistics are based on the out-of-time sample, excluding observations without a prediction from the credit only LR model. Because we have a directional hypothesis as to the impact of both innovations, we used a one-sided 5% test for statistical significance when comparing one model to another or to traditional credit scores, based on 100 bootstrapped samples of the out-of-time validation sample.[31]

While the traditional credit score can in some sense provide a general benchmark for comparison to all of the models built for the study, such comparisons could be affected by the fact that we used a stratified validation sample that was nearly contemporaneous with the sample used to develop the models for the study. As discussed in Section 3.1.3, using this kind of sample tends to generate higher performance metrics for the models developed for the study as compared to applying a traditional credit score model that was developed using data from an earlier time period.

4.1.1 Predictiveness metrics

To evaluate overall model performance, we employed two widely used metrics in credit risk modeling: the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) and the Kolmogorov-Smirnov (KS) statistic.

- **Area Under the ROC Curve (ROC-AUC):** The ROC-AUC evaluates a model's ability to distinguish between defaulters and non-defaulters across all possible classification thresholds. Specifically, the receiver operating characteristic (ROC) curve plots the true positive rate (sensitivity) against the false positive rate (1-specificity) at various thresholds, and the area under the curve (AUC) metric provides a single summary statistic of the model's ability to distinguish or separate the two classes, with a value of 1.0 representing perfect separation and 0.5 indicating random performance. The ROC-AUC is particularly valuable in credit risk modeling because it provides a consistent metric for predictive power, regardless of what approval threshold is used by an individual lender.
- **Kolmogorov-Smirnov (KS) Statistic:** The KS statistic measures the maximum separation between the cumulative distribution functions of predicted probabilities for defaulters and non-defaulters across the entire score range generated by the specific dataset. It quantifies a model's ability to separate these two groups, with a value of 1.0 representing perfect separation and 0 indicating random performance. In credit risk applications, a higher KS statistic reflects a model's effectiveness in ranking borrowers by risk, which is critical for decision-making in underwriting and portfolio management.

Together, these metrics provide complementary insights: ROC-AUC aggregates a model's separation power across all potential decision boundaries, while the KS statistic measures the model's maximum point of separation between defaulters and non-defaulters across the distribution of the specific dataset.

4.1.2 Simulations of impacts on credit access

While ROC-AUC and KS provide insight into overall model performance, they do not directly quantify the impact of model adoption on underwriting decisions by individual lenders working with different pools of applicants. To address this, we conducted simulation studies to evaluate how incorporating cash flow data and machine learning techniques would affect three metrics: (1) overall approval rates; (2) changes in declines of consumers who did not go on to default; and (3) changes in approvals of consumers who did go on to default. We ran three sets of similiuations for lenders who seek to limit losses to different thresholds depending on their risk appetites. (Lenders with higher risk appetites will typically charge higher prices to cover their higher losses.) These simulations reflect the fact that impacts of a new underwriting model on credit access will vary depending on what risk cutoff a particular lender uses, its distribution of applicants, and other factors.

As a first step in the simulations, we applied all of the models, including the traditional credit score for use as a benchmark, to the test sample and used the defaults reflected there as an anchor to calibrate the models relative to each other to ensure consistency in interpreting score cutoffs across models.[32] We then simulated approval and decline decisions for each model using the out-of-time validation sample and applying three different probability of default (PD) thresholds, corresponding to loss rates of 3%, 5%, and 7%.

- **3% Loss Rate:** Conservative risk appetite, typical for prime lenders.
- **5% Loss Rate:** Moderate risk appetite, common for near-prime lenders.
- **7% Loss Rate:** Aggressive risk appetite, more likely to be used by subprime lenders.

Thus, simulation results provide measures of which consumers would be approved or rejected by each individual model if it was adopted by a lender seeking to keep its losses consistent to a set percentage of its portfolio. The simulations are risk neutral in the sense that we assume that the lender is generally seeking to maintain the same credit risk appetite by requiring all approvals to have a predicted risk less than or equal to the same probability of default threshold under each model being compared.

Since the out-of-time validation sample is not representative of the general U.S. population, as previously noted, we weighted the results to match the distribution by credit score bands reflected in a national dataset of new credit account originations across all three nationwide credit bureaus—a benchmark reflecting the risk profile of new credit applicants across the three major credit bureaus during the study period. This adjustment ensures that the simulation results estimate effects on the general population of U.S. consumers who had effective demand for credit during the study period, though an individual lender that used the models would not necessarily have a nationally representative applicant pool either and thus would experience a different distribution of "through-the-door" applicants.

The specific metrics reported below from the simulations are defined as:

- **Approval Rates:** The proportion of the dataset that were approved by the given model at the particular risk threshold.
- **Approved Defaulters:** The percent of consumers that were approved by the given model at the particular risk threshold that went on to default on the credit account that they opened during the study period, expressed as a proportion of all defaulters in the dataset. This is also known as the Type I error rate or false positive rate. False positives lead to credit losses from a lender's perspective, and can have negative consequences for consumers as well due to collections activity, foreclosures, credit score declines, and other negative financial consequences from failing to repay the loan on time.

For simplicity, we did not report changes in the percentage of declined non-defaulters, which is the percentage of consumers that were declined by the given model at the particular risk threshold but did not in fact default on the credit account that they opened during the study period, expressed as a proportion of all non-defaulters in the dataset. This is also known as the Type II error rate or false negative rate. False negatives lead to missed profit opportunities from a lender's perspective and missed utility (e.g., inability to access credit to bridge a cash flow gap or for other purposes) from a consumer's perspective. Our analysis showed that increases in overall approval rates are closely associated with reductions in false negatives.

It should be noted that the simulations are based on several implicit assumptions, adopted for simplicity and to focus on the main research objectives. Specifically, the analysis assumes that a more powerful model, as measured by global performance metrics such as ROC-AUC or KS, will always deliver higher profit than a less powerful one when the same probability of default threshold is applied for underwriting. Consequently, a profit-maximizing lender is assumed to always adopt the more powerful model. This conclusion, however, depends on a few underlying assumptions:

- **Constant Reward for Non-Defaulted Loans:** The profit from a non-defaulted loan is assumed to be constant, regardless of the borrower's risk level.
- **Constant Loss for Defaulted Loans:** The loss from a defaulted loan is assumed to be constant, irrespective of the borrower's risk profile.
- **Unbiased PD Predictions:** The calibrated probability of default (PD) predictions are assumed to be unbiased estimators of the true PD at the consumer level.
- **ROC Dominance:** The more powerful model has an ROC curve that strictly dominates (is everywhere above) the lower performing model over the entire portfolio.

However, in real-world scenarios, deviations from these assumptions—such as variable rewards and losses, biased PD estimates, or ROC crossovers—could lead to situations where the lower performing model delivers higher profit for the lender. Therefore, while the simulation results provide valuable insights regarding the adoption of cash flow data and machine learning techniques, lenders typically will consider their specific business contexts, risk appetites, and product-level risk-reward profiles when evaluating alternative models.

For example, a model that delivers better separation power within a segment that aligns with the lender's most profitable risk-reward profile might become the preferred choice, even if its global performance metrics (e.g., ROC-AUC or KS) are inferior to an alternative model. Consider a credit card portfolio designed to profit from revolvers (borrowers who carry balances) rather than transactors (borrowers who pay off balances monthly). Such a portfolio might benefit more from a model that excels at differentiating default risk among borrowers within the near-prime credit tier, where revolver behavior is more prevalent, even if another model performs better globally or in the super-prime tier. Conversely, a portfolio focused on high transaction volumes might prioritize a model that better differentiates borrowers in the super-prime tier, where transaction activity is typically higher.

These examples illustrate that the "best" model for a given lender depends not only on global performance metrics but also on the specific profitability dynamics of its portfolio and target segments. As such, lenders will evaluate models based on their alignment with business objectives and the unique risk-reward profiles of their products.

4.1.3 Subgroup analyses

In addition to evaluating performance and credit access impacts for the models as a whole, we also measured those impacts for specific subgroups:

1. **By credit history:** We used three subgroups that are frequently used by model developers and lenders to differentiate among consumers depending on whether and when they have experienced delinquencies on credit accounts or major derogatory events such as accounts being moved to internal collections status, charged off by lenders, foreclosures, or bankruptcies. Time periods are measured from the month before the consumer opened the new credit account reflected in the dataset.
   - **Recent Derogatory:** Consumer has experienced at least one delinquency of 60 days or more in the past 6 months or at least one major derogatory event in the past 24 months.
   - **Older Derogatory:** Consumer experienced at least one 60+ delinquency more than 6 months before or at least one major derogatory event more than 24 months before.
   - **Never Derogatory:** No 60+ days delinquencies or major derogatory events reflected in the consumer's credit file.
2. **By income:** For purposes of defining the subgroups, we measured total inflows within the previous 365 days across all accounts reflected in the data aggregator information, excluding internal transfers. We benchmarked to national median household income in 2019 assuming that 25 percent had been deducted for taxes. While income levels are often assessed based on metropolitan area averages or similar benchmarks, that was not possible for this study because we did not have any geographic information about the sample members.
   - **Low-to-Moderate Income (LMI):** Income less than or equal to $51,000, which is roughly 80% of national median household income.
   - **Middle- and Upper-Income (MUI):** Income greater than $51,000.

3. **By race/ethnicity:** The demographic information supplied in the credit bureau data focuses contains about 14 categories divided by race/ethnicity. Due to sample size issues, we combined various categories to concentrate our analysis on a majority reference group of Caucasians and a combined minority group of Black and Hispanic consumers. We did not analyze several categories of Asian consumers, who tend to fall between the two groups analyzed with regard to financial metrics.
    - **Caucasian:** Consumers identified as of Mediterranean, Scandinavian, Middle Eastern, Jewish, Western European, or Eastern European backgrounds.
    - **Black/Hispanic:** Consumers identified as Black or Hispanic.

A summary of the sample distributions and default rates for these subgroups groups and their cross tabulations are provided in Table 4.1.

TABLE 4.1 SAMPLE DISTRIBUTION BY SUBGROUPS

	COUNTS INCOME LEVEL			% OF ROW TOTAL INCOME LEVEL			% OF COLUMN TOTAL INCOME LEVEL			DEFAULT RATE INCOME LEVEL		
DEROGATORY STATUS	LMI	MUI	TOTAL	LMI	MUI	TOTAL	LMI	MUI	TOTAL	LMI	MUI	TOTAL
RECENT	8,805	26,951	35,756	25%	75%	100%	16%	7%	8%	12.90%	9.22%	10.12%
OLD	18,186	80,779	98,965	18%	82%	100%	33%	22%	23%	5.98%	3.20%	3.71%
NEVER	28,302	261,523	289,825	10%	90%	100%	51%	71%	68%	1.95%	0.66%	0.79%
TOTAL	55,293	369,253	424,546	13%	87%	100%	100%	100%	100%	5.02%	1.84%	2.25%

	COUNTS INCOME LEVEL			% OF ROW TOTAL INCOME LEVEL			% OF COLUMN TOTAL INCOME LEVEL			DEFAULT RATE INCOME LEVEL		
RACE/ETHNIC GROUP	LMI	MUI	TOTAL	LMI	MUI	TOTAL	LMI	MUI	TOTAL	LMI	MUI	TOTAL
BLACK/HISPANIC	10,497	52,011	62,508	17%	83%	100%	19%	14%	15%	5.10%	2.44%	2.89%
WHITE	29,862	238,864	268,726	11%	89%	100%	54%	65%	63%	4.53%	1.59%	1.92%
OTHER	14,934	78,378	93,312	16%	84%	100%	27%	21%	22%	5.94%	2.20%	2.80%
TOTAL	55,293	369,253	424,546	13%	87%	100%	100%	100%	100%	5.02%	1.84%	2.25%

	COUNTS DEROG STATUS				% OF ROW TOTAL DEROG STATUS				% OF COLUMN TOTAL DEROG STATUS				DEFAULT RATE DEROG STATUS			
RACE/ETHNIC GROUP	RECENT	OLD	NEVER	TOTAL	RECENT	OLD	NEVER	TOTAL	RECENT	OLD	NEVER	TOTAL	RECENT	OLD	NEVER	TOTAL
BLACK/HISPANIC	7,255	19,433	35,820	62,508	12%	31%	57%	100%	25%	24%	14%	17%	9.52%	3.74%	1.08%	2.89%
WHITE	20,107	58,429	190,190	268,726	7%	22%	71%	100%	71%	73%	75%	74%	9.56%	3.38%	0.66%	1.92%
OTHER	8,394	21,103	63,815	93,312	9%	23%	68%	100%	29%	27%	25%	26%	11.98%	4.60%	1.00%	2.80%
TOTAL	28,501	79,532	254,005	362,038	8%	22%	70%	100%	100%	100%	100%	100%	10.12%	3.71%	0.79%	2.25%

4.2 Results of model comparisons

In this study, we evaluate the performance of various credit underwriting models, focusing on traditional logistic regression and machine learning models that incorporate different data sources. We separately analyze the impact of cash flow data (across all model types) and the impact of model types (across all data sources). While there is some overlap, we believe that the results are often easier to understand when presented separately for each innovation. To help synthesize the combined impact of both innovations, we also discuss which models across all of the various permutations have the strongest predictiveness and access metrics.

Our findings reveal that both the data and analytical innovations have generally positive effects although there are some variations in impacts on predictiveness and credit access depending on the data type, model, risk threshold, and subgroup:

- **Our underwriting models built solely with cash flow data were able to rank order credit risk well, although they were not as powerful as a traditional credit score or our credit only and hybrid models.** The cash only machine learning version was more powerful than the logistic regression version but still did not match the predictiveness of the models that incorporated credit bureau data for the full sample population. For example, the ML cash only model's ROC-AUC was .7987, compared to ROC-AUCs ranging from .8662 to .8854 for credit scores and for models that incorporated credit bureau data. This result suggests that cash only models are likely to be most useful for scoring consumers with limited or no credit history, a segment often underserved by traditional credit scoring and underwriting methods, or other niche situations. However, our sample limitations prevented us from quantifying the exact benefits of these models for these subgroups.
- **Combining cash flow data with either credit scores or credit bureau data produced modest improvements in performance relative to analogous credit only models.** For example, the hybrid models' ROC-AUC metrics were roughly .3% to .7% higher than the most similar credit only benchmarks. The impacts of adding cash flow data were somewhat larger for models that incorporated a traditional credit score than the models built with credit score data.
  - **In simulations using a 3% risk cutoff that is likely to be used by mainstream lenders, the hybrid models increased approval rates by .6% to 1.6% over analogous scores or models that relied solely on credit bureau data.** Compared to a LR credit only baseline, the ML hybrid models also reduced approvals of defaulters, but the LR hybrids did not.
  - **At a 7% risk threshold likely to be used by higher cost subprime lenders, the hybrid models reduced approvals of consumers who went on to default relative to credit only analogues.** Most models also increased overall approval rates, though the size of the gains was smaller than at 3%.
- **Machine learning models outperformed logistic regression models across all data sources, with particularly substantial improvements for the models using all available features (cash flow, credit bureau, or the two sources combined).** ROC-AUC improvements were about 2% across all three data types. Prediction improvements were more modest but still statistically significant when comparing ML and LR models that combined cash flow information with credit scores.

- › **Similar directionally to the hybrid results, the ML models produced the biggest and most consistent increases in credit approvals in simulations at a conservative risk threshold and larger and more consistent reductions in approvals of consumers who went on to default at higher risk thresholds.**
- › **The ML models built with credit bureau data (alone or when combined with cash flow data) showed the largest and most consistent improvements in access at multiple risk thresholds.** For example, at the most conservative risk cut off the models both increased credit approvals by nearly 4% and reduced approvals of consumers who went on to default by at least 9% relative to LR analogs. Approval rates were higher for the LR hybrid built with a traditional credit score and cash flow data than for the ML version, though the ML model substantially reduced approvals of defaulters.

» **Improvements in predictiveness for the hybrid and ML models relative to credit only and LR models were generally widespread across subgroups based on past credit history, income level, and race/ethnicity.**

- › **Changes in credit access varied somewhat based on credit history and income levels.** Increases in credit approval rates were particularly large and consistent across different thresholds for consumers with recent derogatory history, while reductions in approvals of consumers who went on to default were particularly large and consistent for low-to-moderate income populations.

» **Overall, the hybrid ML model that combined cash flow data with credit bureau data had the highest predictiveness metrics in general and across subgroups.** It also had the highest approval rates overall and for most subgroups at most risk thresholds, while also producing relatively low false positive rates (approvals of consumers who went on to default).

- › **The ML model built using only credit bureau data frequently ranked second on both predictiveness and access metrics.** Metrics for the credit only ML were higher than for the ML hybrid in a few scattered cases depending on the threshold and subgroup, but overall the ML hybrid model produced the most widespread improvements.
- › **The two hybrid models built with cash flow data and credit scores were relatively close together in overall predictiveness and both expanded access at a 3% cut off, but had different patterns on credit access at higher risk thresholds.** The LR credit score hybrid generally had higher approval rates and higher false positive rates, while the ML version had lower approval rates and lower false positive rates.

The following subsections describe the empirical results in greater depth. Section 5 discusses implications and supplemental analyses.

4.2.1 Predictiveness impacts

We start by analyzing model performance. Section 4.2.1.1 focuses on the models overall while Section 4.2.1.2 focuses on model performance by subgroup.

4.2.1.1 Overall model performance

Table 4.2 reflects the overall ROC-AUC and KS statistics for all of the models and scores used in the study, with the cash only models presented at the top of the chart, the models built with credit bureau data in the middle rows, and the models built with credit scores at the bottom. The yellow cells reflect the main performance metrics for the given model, while the white cells reflect both raw and percentage changes for comparisons of various models, using italics to note statistically significant differences. The comparisons between logistic regression models and machine learning models built with the same data are presented vertically while the comparisons between credit only and hybrid models that also incorporate cash flow information are presented horizontally. (For the ML hybrid model that is built by combining credit scores with cash flow data, the metrics are benchmarked against the credit score alone.) The same format is used in subsequent tables, except that results for the cash only models are presented separately in Appendix B for the reasons discussed below.

TABLE 4.2 OVERALL MODEL PERFORMANCE METRICS

	ROC-AUC						KS					
MODEL			**CASH ONLY**						**CASH ONLY**			
LR			.7822						.4326			
ML			.7987						.4661			
Δ			*.0165*	*2.11%*					*.0335*	*7.74%*		
MODEL	**CREDIT ONLY**		**CREDIT + CASH**		**Δ**		**CREDIT ONLY**		**CREDIT + CASH**		**Δ**	
LR	.8653		.8682		*.0029*	*0.34%*	.5777		.5801		.0024	0.42%
ML	.8831		.8854		*.0023*	*0.26%*	.6157		.6218		.0061	0.99%
Δ	*.0178*	*2.06%*	*.0172*	*1.98%*			*.0380*	*6.58%*	*.0417*	*7.19%*		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**		**CS ONLY**		**CS + CASH**		**Δ**	
LR	.8662		.8704		*.0042*	*0.48%*	.5798		.5906		*.0108*	*1.86%*
ML	NA		.8723		*.0061*	*0.70%*	NA		.5996		*.0198*	*3.41%*
Δ			*.0019*	*0.22%*					*.0090*	*1.52%*		

The chart reflects several key findings. With regard to **overall model performance by data type,** the first band of the chart demonstrates that both versions of the cash only models rank ordered risk well, for instance with ROC-AUCs of .7822 for the LR version and .7987 for the ML version. However, neither model performed as well as the models reflected in the second and third bands of the chart, including the credit only and hybrid models that we built for the study and the traditional credit score benchmark (ROC-AUCs ranging from .8662 to .8854).

Given that substantial gap in performance, lenders would be highly unlikely to use the kinds of cash only models we built to evaluate a broad pool of applicants for the kinds of traditional credit products where those applicants have substantial traditional credit histories.[33] However, the cash only models could still potentially be powerful for niche situations, such as more specialized small dollar products or for evaluating more specialized groups of applicants, such as consumers with little or no information in their credit files. Because our dataset only included about 2000 consumers who did not have traditional credit scores under the mostly widely used third party models and about 3000 consumers who had scores but only two or fewer credit accounts listed in their files, we do not present performance results for those populations because they would not be reliable. (Implications for this group are further discussed in Section 5.1.). We also do not present performance results for the cash only models in subsequent main text or tables, given their lower performance. (See Appendix B for additional results.)

Among the remaining models, the ML hybrid model built with cash flow data and credit bureau data (which we shorthand as the ML credit + cash model) performed the best, with a ROC-AUC of .8854 and KS of .6218. The ML credit only model came in second and the ML hybrid model built with credit scores and cash flow data came in third, although the LR credit score hybrid was not far behind and performed the best of all the LR models.

Focusing specifically on the performance of **hybrid models relative to analogous credit only models,** moving horizontally across the table shows modest improvements in performance from adding cash flow data. The improvements range from roughly .3% to roughly .7% for ROC-AUC and .4% to as much as 3.4% for KS. Improvements are generally larger for the hybrid models built with credit scores than for the hybrid models built with credit bureau data.

Shifting to performance of **machine learning models relative to analogous logistic regression models,** moving vertically within the table shows that the ML models that used full data sets (cash only in the top band and credit only and credit + cash in the second band) increased ROC-AUCs by roughly 2% and K-S statistics of roughly 7% over comparable LR models. As reflected in the third band, the difference between the two hybrid models built with credit scores and cash flow data was modest but still statistically significant at .2% ROC-AUC and 1.5% K-S.

4.2.1.2 Subgroup analyses

Performance results for different consumer segments are presented in Table 4.3 (credit history), income (Table 4.4), and race/ethnicity (Table 4.5).

Moving horizontally to compare the performance of **hybrid models relative to analogous credit only models** shows positive gains for all of the subgroups except for consumers with older derogatory history under the LR credit + cash model and the KS metric for the ML credit + cash model. Similar to the results for the sample as a whole, the gains for most subgroups from adopting cash flow data tend to be somewhat larger under the credit score hybrids than the hybrids built with credit bureau data.

Moving vertically to compare the performance of **machine learning models relative to analogous logistic regression models** shows relatively large gains across all subgroups from using ML models built with credit bureau features (with or without cash flow information) compared to LR analogues. As between the two credit score hybrids, the ML model performed slightly better across all of the subgroups on most metrics.

Similar to the patterns for overall performance, the ML credit + cash model generally performs the best across all subgroups, followed by the ML credit only model. The ML credit score hybrid generally came in third, though the LR credit score hybrid performed better on a few metrics for consumers with recent derogatory history and no derogatory history.

TABLE 4.3 MODEL PERFORMANCE BY PAST CREDIT HISTORY

RECENT DEROG

ROC-AUC

MODEL	CREDIT ONLY		CREDIT + CASH		Δ	
LR	.7293		.7308		.0015	0.21%
ML	.7577		.7617		.0040	0.53%
Δ	.0284	3.89%	.0309	4.23%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**	
LR	.7277		.7301		.0024	0.33%
ML	NA		.7285		.0008	0.11%
Δ			-.0016	-0.22%		

KS

MODEL	CREDIT ONLY		CREDIT + CASH		Δ	
LR	.3510		.3533		.0023	0.66%
ML	.3858		.4085		.0227	5.88%
Δ	.0348	9.91%	.0552	15.62%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**	
LR	.3436		.3600		.0164	4.77%
ML	NA		.3607		.0171	4.98%
Δ			.0007	0.19%		

OLDER DEROG

ROC-AUC

MODEL	CREDIT ONLY		CREDIT + CASH		Δ	
LR	.7655		.7640		-.0015	-0.20%
ML	.8057		.8067		.0010	0.12%
Δ	.0284	3.89%	.0309	4.23%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**	
LR	.7277		.7301		.0024	0.33%
ML	NA		.7285		.0008	0.11%
Δ			-.0016	-0.22%		

KS

MODEL	CREDIT ONLY		CREDIT + CASH		Δ	
LR	.4004		.3983		-.0021	-0.52%
ML	.4715		.4683		-.0032	-0.68%
Δ	.0348	9.91%	.0552	15.62%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**	
LR	.3436		.3600		.0164	4.77%
ML	NA		.3607		.0171	4.98%
Δ			.0007	0.19%		

NEVER DEROG

ROC-AUC

MODEL	CREDIT ONLY		CREDIT + CASH		Δ	
LR	.8331		.8460		.0129	1.55%
ML	.8565		.8610		.0045	0.53%
Δ	.0234	2.81%	.0150	1.77%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**	
LR	.8372		.8480		.0108	1.29%
ML	NA		.8483		.0111	1.33%
Δ			.0003	0.04%		

KS

MODEL	CREDIT ONLY		CREDIT + CASH		Δ	
LR	.5330		.5547		.0217	4.07%
ML	.5732		.5951		.0219	3.82%
Δ	.0402	7.54%	.0404	7.28%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**	
LR	.5376		.5668		.0292	5.43%
ML	NA		.5608		.0232	4.32%
Δ	NA		-.0060	-1.06%		

TABLE 4.4 MODEL PERFORMANCE BY INCOME LEVEL

LMI	ROC-AUC						KS					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	.8180		.8215		.0035	0.43%	.4968		.5032		.0064	1.29%
ML	.8425		.8489		.0064	0.76%	.5284		.5369		.0085	1.61%
Δ	.0245	3.00%	.0274	3.34%			.0316	6.36%	.0337	6.70%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	.8208		.8257		.0049	0.60%	.5112		.5195		.0083	1.62%
ML	NA		.8286		.0078	0.95%	NA		.5243		.0131	2.56%
Δ			.0029	0.35%					.0048	0.92%		

MUI	ROC-AUC						KS					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	.8665		.8700		.0035	0.40%	.5813		.5840		.0027	0.46%
ML	.8838		.8856		.0018	0.20%	.6123		.6216		.0093	1.52%
Δ	.0173	2.00%	.0156	1.79%			.0310	5.33%	.0376	6.44%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	.8669		.8718		.0049	0.57%	.5835		.5964		.0129	2.21%
ML	NA		.8732		.0063	0.73%	NA		.6053		.0218	3.74%
Δ			.0014	0.16%			NA		.0089	1.49%		

TABLE 4.5 MODEL PERFORMANCE BY RACE/ETHNICITY

AA/H	ROC-AUC						KS					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	.8459		.8482		.0023	0.27%	.5347		.5466		.0119	2.23%
ML	.8709		.8744		.0035	0.40%	.5989		.6086		.0097	1.62%
Δ	.0250	2.96%	.0262	3.09%			.0642	12.01%	.0620	11.34%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	.8545		.8595		.0050	0.59%	.5515		.5586		.0071	1.29%
ML	NA		.8645		.0100	1.17%	NA		.5901		.0386	7.00%
Δ			.0050	0.58%					.0315	5.64%		

CAUCASIAN	ROC-AUC						KS					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	.8704		.8740		.0036	0.41%	.5881		.5918		.0037	0.63%
ML	.8851		.8863		.0012	0.14%	.6160		.6204		.0044	0.71%
Δ	.0147	1.69%	.0123	1.41%			.0279	4.74%	.0286	4.83%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	.8680		.8731		.0051	0.59%	.5836		.5974		.0138	2.36%
ML	NA		.8730		.0050	0.58%	NA		.6054		.0218	3.74%
Δ			-.0001	-0.01%					.0080	1.34%		

4.2.2 Credit access impacts

Improvements in model performance can potentially increase access to responsible credit by reducing false negatives (declines of creditworthy borrowers), reducing false positives (approvals of borrowers who ultimately default), or producing both effects at the same time. However, the specific impacts on overall approval rates and the other inclusion metrics depend on factors such as the lender's risk thresholds, the distribution of applicants, and the characteristics of the borrower population. To evaluate these effects, we conducted simulations using the methodology described in Section 4.1.2.

4.2.2.1 Overall impacts on credit access

Table 4.6 presents the access results for risk thresholds of 3%, 5%, and 7%, respectively. A 3% cutoff is most likely to be used by banks and other lenders focusing on the prime market, while the 7% cutoff is likely to be used by higher cost lenders that focus on the subprime market. It is helpful to read the chart by focusing first on changes in overall approval rates before shifting to the second set of columns to analyze how those changes impact the percent of defaulters that are approved under the model. This is often called the false positive rate.

Moving horizontally to compare the performance of **hybrid models relative to analogous credit only models,** the simulations indicate that all of the hybrid models would increase overall approval rates at a 3% cutoff compared to credit scores or credit only models. At higher cutoffs, most of the hybrid models still increase approval rates but by smaller percentages, though overall approval rates decline under the ML hybrid built with credit scores and cash flow data.

Shifting to the right side of the table provides additional insight into how false positive rates change under the hybrid models. Approvals of consumers who go on to default decline under the ML models, particularly the ML credit score hybrid. While this does not expand the credit box overall, it reduces the risk that consumers are extended credit that they are unlikely to succeed with.

Moving vertically to compare the **machine learning models relative to analogous logistic regression models** shows that the ML models built with credit bureau data (with or without cash flow information) both increase overall approval rates and reduce false positive rates compared to LR versions. The impacts on approval rates get somewhat smaller at higher thresholds, while reductions to the percentage of defaulters who are approved tend to increase. Again, the ML credit score hybrid has a different pattern, with both lower approval rates overall and lower false positive rates compared the LR version.

Overall, the ML credit + cash model and ML credit only model produce the most consistent results across the various thresholds and metrics, with the former slightly ahead as to its overall credit approval rate and the latter having a slightly lower false positive rate. Among the credit score hybrids, the LR version has a higher credit approval rate at every threshold, while the ML version approves fewer defaulters.

TABLE 4.6 CREDIT ACCESS ANALYSIS—OVERALL POPULATION

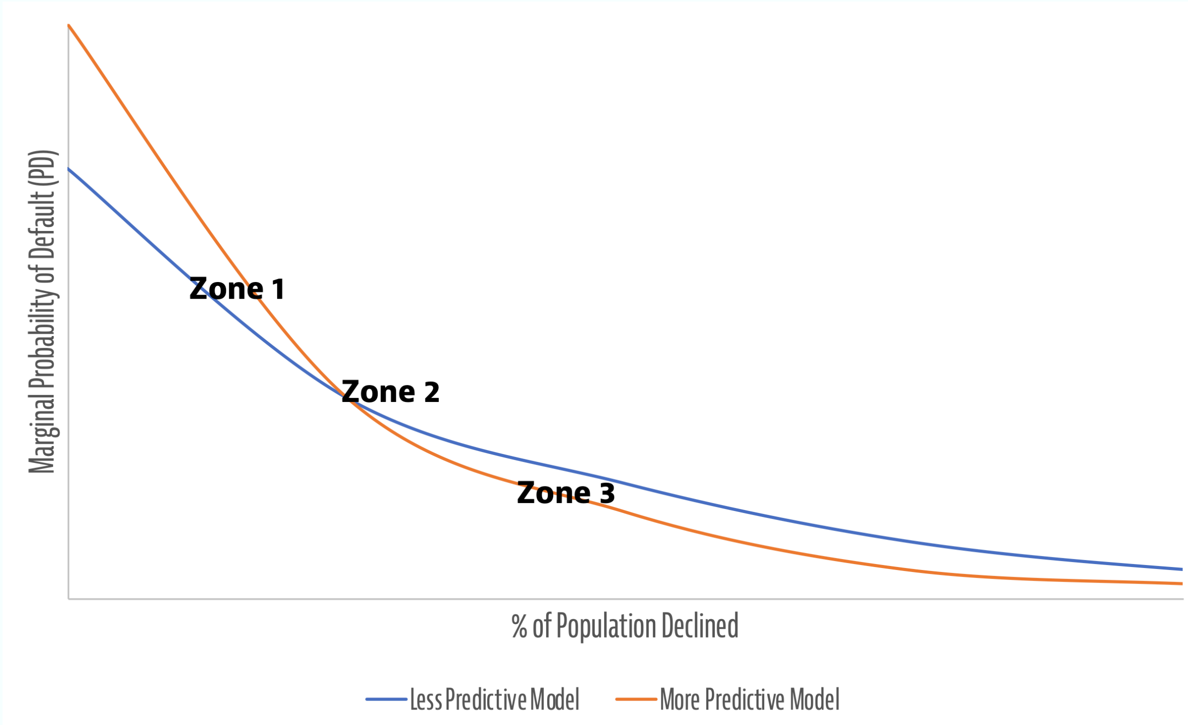
3% CUTOFF	APPROVAL RATES				% OF DEFAULTERS APPROVED			
MODEL	CREDIT ONLY	CREDIT + CASH	Δ		CREDIT ONLY	CREDIT + CASH	Δ	
LR	65.18%	65.71%	.5pp	0.81%	12.40%	13.10%	.7pp	5.65%
ML	67.66%	68.21%	.6pp	0.81%	11.23%	11.79%	.6pp	4.99%
Δ	2.5pp / 3.80%	2.5pp / 3.80%			-1.2pp / -9.44%	-1.3pp / -10.00%		
MODEL	CS ONLY	CS + CASH	Δ		CS ONLY	CS + CASH	Δ	
LR	65.24%	66.25%	1.0pp	1.55%	12.01%	12.51%	.5pp	4.16%
ML	NA	65.61%	.4pp	0.57%	NA	11.74%	-.3pp	-2.25%
Δ		-.6pp / -0.97%				-.8pp / -6.16%		

5% CUTOFF	APPROVAL RATES				% OF DEFAULTERS APPROVED			
MODEL	CREDIT ONLY	CREDIT + CASH	Δ		CREDIT ONLY	CREDIT + CASH	Δ	
LR	74.88%	75.20%	.3pp	0.43%	22.03%	22.21%	.2pp	0.82%
ML	76.00%	76.49%	.5pp	0.64%	18.88%	19.22%	.3pp	1.80%
Δ	1.1pp / 1.50%	1.3pp / 1.72%			-3.2pp / -14.30%	-3.0pp / -13.46%		
MODEL	CS ONLY	CS + CASH	Δ		CS ONLY	CS + CASH	Δ	
LR	75.47%	75.79%	.3pp	0.42%	21.75%	22.45%	.7pp	3.22%
ML	NA	73.93%	-1.5pp	-2.04%	NA	19.35%	-2.4pp	-11.03%
Δ		-1.9pp / -2.45%				-3.1pp / -13.81%		

7% CUTOFF	APPROVAL RATES				% OF DEFAULTERS APPROVED			
MODEL	CREDIT ONLY	CREDIT + CASH	Δ		CREDIT ONLY	CREDIT + CASH	Δ	
LR	81.01%	81.32%	.3pp	0.38%	31.84%	31.59%	-.3pp	-0.79%
ML	81.18%	81.65%	.5pp	0.58%	27.15%	27.44%	-.3pp	-1.07%
Δ	.2pp / 0.21%	.3pp / 0.41%			-4.7pp / -14.73%	-4.2pp / -13.14%		
MODEL	CS ONLY	CS + CASH	Δ		CS ONLY	CS + CASH	Δ	
LR	81.49%	81.72%	.2pp	0.28%	31.56%	30.95%	-.6pp	-1.93%
ML	NA	79.37%	-2.1pp	-2.60%	NA	27.14%	-4.4pp	-14.01%
Δ		-2.4pp / -2.88%				-3.8pp / -12.31%		

In interpreting these results, it can be helpful to use a visualization that maps the percent of population that would be declined at different predicted default rates to understand how the impacts of adopting a more accurate underwriting model can vary depending on the threshold used by the lender relative to where the model "intersects" with a less predictive model. A more predictive model is better at separating individuals with different levels of credit risk, therefore manifests as a steeper curve. It will also intersect the less predictive model in this view because better separation moves defaulters from a lower risk band to a higher one. An individual lender's risk threshold can be visualized as a horizontal line that may extend from the y-axis in zones 1, 2, or 3, depending on the lender's risk appetite.

FIGURE 4 EFFECT OF SWITCHING TO A MORE PREDICTIVE MODEL—AN ILLUSTRATIVE EXAMPLE



- » **Zone 1** is the area substantially above the point where the two models intersect. If the lender's credit threshold crosses zone 1 (visualize a horizontal line parallel to the X axis), the lender will typically reduce the size of its credit box when adopting the more predictive model because the model will facilitate more focused approvals. In this zone, the more predictive model will typically have the following access metrics relative to the less predictive model:
    - › Lower overall approval rate
    - › Likely higher rate of declines among non-defaulters (false negative rate)
    - › Lower rate of approving consumers who go on to default (false positive rate)
- » **Zone 2** is the area close to where the models intersect. If the lender's credit threshold crosses zone 2, the lender's credit box will typically remain close to the same size in adopting the more predictive model. In this zone, the more predictive model will typically have the following access metrics relative to the less predictive model:
    - › Little or no difference in overall approval rate
    - › Likely lower rate of declines among non-defaulters (false negative rate)
    - › Likely lower rate of approving consumers who go on to default (false positive rate)
- » **Zone 3** is the area substantially below the point where the two models intersect. If the lender's threshold crosses zone 3, the lender will typically expand its credit box to take advantage of more profitable growth. In this zone, the more predictive model will typically have the following access metrics relative to the less predictive model:
    - › Higher approval rates
    - › Lower rate of declines among non-defaulters (false negative rate)
    - › Rate of approving consumers who go on to default may be higher or lower depending on the shape of the curves and location within the zone (false positive rate)

Using these concepts and visualizations helps to understand the results reflected in Table 4.2. Starting with the ML credit score hybrid model, if it is compared to either the LR credit score hybrid or the credit score alone, a 3% risk cut off crosses the two models in Zone 2, while the higher cut offs cross the models in Zone 1. While the results yield lower overall approval rates at the higher thresholds, the reductions in false positive rates still have positive outcomes for both lenders and consumers by helping to reduce the number of applicants who are offered credit that they are not likely to succeed in repaying, particularly at high prices that may themselves present challenges to the borrowers' finances. By improving access to responsible credit, the ML credit score hybrid would reduce the number of consumers who are likely to experience negative consequences from collections, foreclosures, and credit score damage.

With regard to the other models, when the ML models built with full credit bureau data (with or without cash flow information) are compared to LR versions, the 3% risk cut off intersects the two models in Zone 3, while the higher risk cut offs move toward Zone 2. This means that there are still some increases in overall approval rates at the higher thresholds, though they get progressively smaller.

4.2.2.2 Access impacts by subgroup

The hybrid models' impacts on access to credit produced some variations across different consumer subgroups depending on credit history and income levels as described below.

By credit history

The three subgroups divided by credit history have highly divergent patterns with regard to credit access under traditional models such as credit scores or models built with traditional data. The simulations of impacts on credit access from adopting hybrid models or ML models also showed different patterns.

For example, consumers with recent derogatory history have the lowest approval rates of any subgroup analyzed in the study. Hybrid models and ML models both produced relatively large increases in approval rates relative to credit only and LR analogues at every risk threshold. The one exception was the ML credit score hybrid, which increased approval rates relative to the LR credit score hybrid and credit scores alone at the 3% risk level but not at higher thresholds. It did reduce false positives, however.

For consumers with older derogatory history, hybrid models and ML models generally increased approval rates at the 3% and 5% thresholds relative to credit only and LR analogues. At the 7% threshold, their primary impact was to reduce approvals of defaulters, often accompanied by declines in overall approval rates as well. For consumers with no derogatory history, hybrid models and ML models slightly increased approval rates at the 3% threshold but their main impact was to reduce false positives.

Overall, the ML credit + cash model and ML credit only model produced the most consistent results on credit access, as they generally had the highest credit approval rates for consumers with derogatory history (recent or older), while also reducing false positives for most subgroups. Among the credit score hybrids, the LR version had a higher credit approval rate at every threshold, while the ML version approved fewer defaulters.

TABLE 4.7 CREDIT ACCESS ANALYSIS—RECENT DEROGATORY

3% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	8.54%		9.56%		1.0pp	11.94%	1.41%		1.97%		.6pp	39.72%
ML	13.10%		15.00%		1.9pp	14.50%	1.61%		2.74%		1.1pp	70.19%
Δ	4.6pp	53.40%	5.4pp	56.90%			.2pp	14.18%	.8pp	39.09%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	9.55%		11.07%		1.5pp	15.92%	1.67%		1.83%		.2pp	9.58%
ML	NA		11.84%		2.3pp	23.98%	NA		1.95%		.3pp	16.77%
Δ			.8pp	6.96%					.1pp	6.56%		

5% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	20.41%		22.09%		1.7pp	8.23%	5.36%		5.85%		.5pp	9.14%
ML	24.60%		26.47%		1.9pp	7.60%	4.90%		6.31%		1.4pp	28.78%
Δ	4.2pp	20.53%	4.4pp	19.83%			-.5pp	-8.58%	.5pp	7.86%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	20.37%		22.45%		2.1pp	10.21%	4.53%		5.53%		1.0pp	22.08%
ML	NA		20.72%		.4pp	1.72%	NA		4.70%		.2pp	3.75%
Δ			-1.7pp	-7.71%					-.8pp	-15.01%		

7% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	31.71%		34.08%		2.4pp	7.47%	11.80%		11.97%		.2pp	1.44%
ML	34.94%		37.08%		2.1pp	6.12%	10.94%		12.35%		1.4pp	12.89%
Δ	3.2pp	10.19%	3.0pp	8.80%			-.9pp	-7.29%	.4pp	3.17%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	31.15%		33.62%		2.5pp	7.93%	10.25%		11.33%		1.1pp	10.54%
ML	NA		30.14%		-1.0pp	-3.24%	NA		9.58%		-.7pp	-6.54%
Δ			-3.5pp	-10.35%					-1.8pp	-15.45%		

TABLE 4.8 CREDIT ACCESS ANALYSIS—OLDER DEROGATORY

3% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	48.15%		48.76%		.6pp	1.27%	13.90%		15.31%		1.4pp	10.14%
ML	54.59%		55.53%		.9pp	1.72%	13.37%		14.48%		1.1pp	8.30%
Δ	6.4pp	13.37%	6.8pp	13.88%			-.5pp	-3.81%	-.8pp	-5.42%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	50.04%		51.61%		1.6pp	3.14%	14.73%		15.94%		1.2pp	8.21%
ML	NA		50.85%		.8pp	1.62%	NA		15.27%		.5pp	3.67%
Δ			-.8pp	-1.47%					-.7pp	-4.20%		

5% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	66.93%		67.23%		.3pp	0.45%	30.67%		30.70%		.0pp	0.10%
ML	68.95%		69.56%		.6pp	0.88%	25.08%		24.40%		-.7pp	-2.71%
Δ	2.0pp	3.02%	2.3pp	3.47%			-5.6pp	-18.23%	-6.3pp	-20.52%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	68.47%		68.64%		.2pp	0.25%	30.83%		31.91%		1.1pp	3.50%
ML	NA		65.29%		-3.2pp	-4.64%	NA		26.75%		-4.1pp	-13.23%
Δ			-3.4pp	-4.88%					-5.2pp	-16.17%		

7% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	78.34%		78.16%		-.2pp	-0.23%	45.25%		45.12%		-.1pp	-0.29%
ML	77.45%		77.80%		.4pp	0.45%	36.40%		35.92%		-.5pp	-1.32%
Δ	-.9pp	-1.14%	-.4pp	-0.46%			-8.9pp	-19.56%	-9.2pp	-20.39%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	79.38%		79.25%		-.1pp	-0.16%	46.33%		44.40%		-1.9pp	-4.17%
ML	NA		74.71%		-4.7pp	-5.88%	NA		37.26%		-9.1pp	-19.58%
Δ			-4.5pp	-5.73%					-7.1pp	-16.08%		

TABLE 4.9 CREDIT ACCESS ANALYSIS—NEVER DEROGATORY

3% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	90.80%		91.13%		*.3pp*	*0.36%*	48.42%		48.09%		-.3pp	-0.68%
ML	90.69%		90.64%		-.1pp	-0.06%	40.99%		38.31%		*-2.7pp*	*-6.54%*
Δ	-.1pp	-0.12%	*-.5pp*	*-0.54%*			*-7.4pp*	*-15.34%*	*-9.8pp*	*-20.34%*		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	89.66%		90.24%		*.6pp*	*0.65%*	43.13%		43.32%		.2pp	0.44%
ML	NA		89.22%		*-.4pp*	*-0.49%*	NA		39.10%		*-4.0pp*	*-9.34%*
Δ			*-1.0pp*	*-1.13%*					*-4.2pp*	*-9.74%*		

5% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	95.35%		95.27%		-.1pp	-0.08%	62.98%		62.41%		-.6pp	-0.91%
ML	95.11%		95.11%		.0pp	0.00%	55.49%		54.20%		-1.3pp	-2.32%
Δ	*-.2pp*	*-0.25%*	*-.2pp*	*-0.17%*			*-7.5pp*	*-11.89%*	*-8.2pp*	*-13.15%*		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	95.67%		95.54%		*-.1pp*	*-0.14%*	63.73%		62.55%		-1.2pp	-1.85%
ML	NA		94.35%		*-1.3pp*	*-1.38%*	NA		55.73%		*-8.0pp*	*-12.55%*
Δ			*-1.2pp*	*-1.25%*					*-6.8pp*	*-10.90%*		

7% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	97.32%		97.23%		-.1pp	-0.09%	74.60%		72.63%		-2.0pp	-2.64%
ML	97.07%		97.09%		.0pp	0.02%	65.13%		63.12%		-2.0pp	-3.09%
Δ	*-.2pp*	*-0.26%*	*-.1pp*	*-0.14%*			*-9.5pp*	*-12.69%*	*-9.5pp*	*-13.09%*		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	97.84%		97.56%		*-.3pp*	*-0.29%*	75.96%		72.15%		*-3.8pp*	*-5.02%*
ML	NA		96.64%		*-1.2pp*	*-1.23%*	NA		68.12%		*-7.8pp*	*-10.32%*
Δ			*-.9pp*	*-0.94%*					*-4.0pp*	*-5.59%*		

By income levels

The two income subgroups had the second largest divergence in patterns with regard to credit access under traditional models such as credit scores or models built with traditional data. The simulations of impacts on credit access from adopting hybrid models or ML models also showed different patterns.

For LMI consumers, the hybrid models reduced approvals of consumers who went on to default relative to credit only analogues, accompanied by reductions in overall approval rates. Again, similar to the overall results for the ML credit score hybrid model at higher thresholds, this indicates that all of the risk cut offs are intersecting the hybrid models in Zone 1 for LMI consumers. In contrast, moving from LR to ML models that used credit bureau data generally both increased approval rates and reduced false positives, indicating that the lower risk threshold intersects the models in Zone 3 and higher thresholds move toward Zone 2.

For MUI consumers, the hybrid models increased approval rates relative to credit only analogues, with the exception of the ML credit score hybrid where the 5% and 7% cut offs intersect the model in Zone 1. Moving from LR to ML models that used credit bureau data generally increased approval rates and reduced false positives.

Overall, the ML credit + cash model and ML credit only model generally ranked highest on approval rates, with the ML credit only tending to rank first for LMI consumers and the ML credit + cash model first for MUI consumers. Both models also had among the lowest false positive rates. Among the credit score hybrids, the LR version had a higher credit approval rate at most thresholds for both groups, while the ML version approved fewer defaulters.

TABLE 4.10 CREDIT ACCESS ANALYSIS—LMI

3% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	40.28%		38.51%		-1.8pp	-4.39%	6.72%		5.59%		-1.1pp	-16.82%
ML	42.95%		42.02%		-.9pp	-2.17%	5.94%		5.40%		-.5pp	-9.09%
Δ	2.7pp	6.63%	3.5pp	9.11%			-.8pp	-11.61%	-.2pp	-3.40%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	41.35%		38.18%		-3.2pp	-7.67%	6.24%		5.19%		-1.1pp	-16.83%
ML	NA		38.24%		-3.1pp	-7.52%	NA		4.78%		-1.5pp	-23.40%
Δ			.1pp	0.16%					-.4pp	-7.90%		

5% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	52.84%		50.56%		-2.3pp	-4.31%	13.24%		11.61%		-1.6pp	-12.31%
ML	53.84%		53.32%		-.5pp	-0.97%	11.81%		10.43%		-1.4pp	-11.69%
Δ	1.0pp	1.89%	2.8pp	5.46%			-1.4pp	-10.80%	-1.2pp	-10.16%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	54.38%		50.11%		-4.3pp	-7.85%	12.72%		10.61%		-2.1pp	-16.59%
ML	NA		49.27%		-5.1pp	-9.40%	NA		9.18%		-3.5pp	-27.83%
Δ			-.8pp	-1.68%					-1.4pp	-13.48%		

7% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	62.32%		59.88%		-2.4pp	-3.92%	22.93%		19.42%		-3.5pp	-15.31%
ML	61.39%		61.35%		.0pp	-0.07%	17.60%		16.53%		-1.1pp	-6.08%
Δ	-.9pp	-1.49%	1.5pp	2.45%			-5.3pp	-23.24%	-2.9pp	-14.88%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	63.53%		59.36%		-4.2pp	-6.56%	22.69%		16.45%		-6.2pp	-27.50%
ML	NA		57.14%		-6.4pp	-10.06%	NA		14.41%		-8.3pp	-36.49%
Δ			-2.2pp	-3.74%					-2.0pp	-12.40%		

TABLE 4.11 CREDIT ACCESS ANALYSIS—MUI

3% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	69.85%		70.80%		.9pp	1.36%	14.74%		16.19%		1.5pp	9.84%
ML	72.29%		73.11%		.8pp	1.13%	13.41%		14.42%		1.0pp	7.53%
Δ	2.4pp	3.49%	2.3pp	3.26%			-1.3pp	-9.02%	-1.8pp	-10.93%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	69.71%		71.51%		1.8pp	2.58%	14.40%		15.53%		1.1pp	7.85%
ML	NA		70.73%		1.0pp	1.46%	NA		14.61%		.2pp	1.46%
Δ			-.8pp	-1.09%					-.9pp	-5.92%		

5% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	79.00%		79.81%		.8pp	1.03%	25.65%		26.58%		.9pp	3.63%
ML	80.15%		80.82%		.7pp	0.84%	21.80%		22.84%		1.0pp	4.77%
Δ	1.2pp	1.46%	1.0pp	1.27%			-3.9pp	-15.01%	-3.7pp	-14.07%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	79.41%		80.60%		1.2pp	1.50%	25.47%		27.33%		1.9pp	7.30%
ML	NA		78.54%		-.9pp	-1.10%	NA		23.54%		-1.99pp	-7.58%
Δ			-2.2pp	-2.56%					-3.8pp	-13.87%		

7% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	84.51%		85.33%		.8pp	0.97%	35.51%		36.61%		1.1pp	3.10%
ML	84.88%		85.45%		.6pp	0.67%	31.08%		31.94%		.9pp	2.77%
Δ	.4pp	0.44%	.1pp	0.14%			-4.4pp	-12.48%	-4.7pp	-12.76%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	84.85%		85.90%		1.1pp	1.24%	35.22%		36.94%		1.7pp	4.88%
ML	NA		83.54%		-1.3pp	-1.54%	NA		32.39%		-2.8pp	-8.04%
Δ			-2.4pp	-2.75%					-4.6pp	-12.32%		

By race/ethnicity

The impact of hybrid and ML models relative to credit only and LR analogues on African-Americans/Hispanics and Caucasians was generally similar to each other and to trends for the sample as a whole.

Impacts of the hybrid and ML models on approval rates were generally most positive at the 3% cut off compared to credit only and LR analogues, but shrunk at higher thresholds. Reductions of false positives tended to be high at 7%. The ML credit score hybrid tended to make large reductions in approvals of defaulters, often accompanied by reductions in approval rates relative to the LR credit score hybrid.

Overall, the ML credit + cash model and ML credit only model produce the most consistent results across the various thresholds and metrics, with the highest approval rates and among the lowest false positive rates. Among the credit score hybrids, the LR version has a higher credit approval rate at every threshold for both subgroups, while the ML version approves fewer defaulters.

TABLE 4.12 CREDIT ACCESS ANALYSIS—AFRICAN-AMERICAN/HISPANIC

3% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	56.68%		56.73%		.1pp	0.09%	11.80%		11.61%		-.2pp	-1.61%
ML	60.21%		60.74%		*.5pp*	*0.88%*	10.70%		9.91%		-.8pp	-7.38%
Δ	*3.5pp*	*6.23%*	*4.0pp*	*7.07%*			-1.1pp	-9.32%	-1.7pp	-14.64%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	56.75%		57.51%		*.8pp*	*1.34%*	10.06%		10.30%		.2pp	2.39%
ML	NA		56.85%		.1pp	0.18%	NA		7.84%		*-2.2pp*	*-22.07%*
Δ			*-.7pp*	*-1.15%*					*-2.5pp*	*-23.88%*		

5% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	68.59%		68.76%		.2pp	0.25%	19.79%		18.87%		-.9pp	-4.65%
ML	70.13%		70.66%		.2pp	0.25%	16.81%		17.14%		.3pp	1.96%
Δ	*1.5pp*	*2.25%*	*1.9pp*	*2.76%*			-3.0pp	-15.06%	-1.7pp	-9.17%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	69.17%		69.21%		.0pp	0.06%	19.06%		18.69%		-.4pp	-1.94%
ML	69.17%		69.21%		*-2.0pp*	*-2.95%*	NA		16.43%		*-2.6pp*	*-13.80%*
Δ			*-2.1pp*	*-3.01%*					*-2.3pp*	*-12.09%*		

7% CUTOFF	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	CREDIT ONLY		CREDIT + CASH		Δ		CREDIT ONLY		CREDIT + CASH		Δ	
LR	76.35%		76.78%		.4pp	0.56%	30.17%		29.31%		-.9pp	-2.85%
ML	76.26%		76.87%		*.6pp*	*0.80%*	22.90%		25.10%		*2.2pp*	*9.61%*
Δ	-.1pp	-0.12%	.1pp	0.12%			*-7.3pp*	*-24.10%*	-4.2pp	-14.36%		
MODEL	CS ONLY		CS + CASH		Δ		CS ONLY		CS + CASH		Δ	
LR	76.57%		76.54%		.0pp	-0.04%	30.06%		27.20%		*-2.9pp*	*-9.51%*
ML	NA		73.90%		*-2.7pp*	*-3.49%*	NA		24.38%		*-5.7pp*	*-18.90%*
Δ			*-2.6pp*	*-3.45%*					*-2.8pp*	*-10.37%*		

TABLE 4.13 CREDIT ACCESS ANALYSIS—CAUCASIAN

3% CUTOFF

	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	**CREDIT ONLY**		**CREDIT + CASH**		**Δ**		**CREDIT ONLY**		**CREDIT + CASH**		**Δ**	
LR	68.97%		69.63%		.7pp	0.96%	13.74%		14.83%		1.1pp	7.93%
ML	71.48%		71.90%		.4pp	0.59%	12.71%		13.62%		.9pp	7.16%
Δ	2.5pp	3.64%	2.3pp	3.26%			-1.0pp	-7.50%	-1.2pp	-8.16%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**		**CS ONLY**		**CS + CASH**		**Δ**	
LR	68.89%		69.98%		1.1pp	1.58%	13.46%		14.03%		.6pp	4.23%
ML	NA		69.29%		.4pp	0.58%	NA		13.82%		.4pp	2.67%
Δ			-.7pp	-0.99%					-.2pp	-1.50%		

5% CUTOFF

	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	**CREDIT ONLY**		**CREDIT + CASH**		**Δ**		**CREDIT ONLY**		**CREDIT + CASH**		**Δ**	
LR	78.22%		78.48%		.3pp	0.33%	24.62%		24.63%		.0pp	0.04%
ML	79.26%		79.67%		.4pp	0.52%	21.39%		21.90%		.5pp	2.38%
Δ	1.0pp	1.33%	1.2pp	1.52%			-3.2pp	-13.12%	-2.7pp	-11.08%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**		**CS ONLY**		**CS + CASH**		**Δ**	
LR	78.47%		78.87%		.4pp	0.51%	24.07%		25.28%		1.2pp	5.03%
ML	NA		77.17%		-1.3pp	-1.66%	NA		21.45%		-2.6pp	-10.88%
Δ			-1.7pp	-2.16%					-3.8pp	-15.15%		

7% CUTOFF

	APPROVAL RATES						% OF DEFAULTERS APPROVED					
MODEL	**CREDIT ONLY**		**CREDIT + CASH**		**Δ**		**CREDIT ONLY**		**CREDIT + CASH**		**Δ**	
LR	83.75%		84.07%		.3pp	0.38%	34.95%		34.51%		-.4pp	-1.26%
ML	84.01%		84.32%		.3pp	0.37%	30.63%		30.83%		.2pp	0.65%
Δ	.3pp	0.31%	.2pp	0.30%			-4.3pp	-12.36%	-3.7pp	-10.66%		
MODEL	**CS ONLY**		**CS + CASH**		**Δ**		**CS ONLY**		**CS + CASH**		**Δ**	
LR	83.97%		84.36%		.4pp	0.46%	34.20%		34.17%		.0pp	-0.09%
ML	NA		82.20%		-1.8pp	-2.11%	NA		29.72%		-4.5pp	-13.10%
Δ			-2.2pp	-2.56%					-4.5pp	-13.02%		

5. DISCUSSION AND SUPPLEMENTAL ANALYSES

To supplement the core results presented in Section 4.2, this section provides supplemental information and discussions with regard to the potential use of cash only models for evaluating consumers with little or no traditional credit history, considerations regarding managing models for explainability and fairness, and the implications for lenders in considering staggered adoption of machine learning models and new data sources.

5.1 Cash only models for consumers with little or no traditional history

A significant portion of the U.S. population remains underserved by traditional credit scoring systems. According to a Consumer Financial Protection Bureau report using 2010 data, approximately 11% of U.S. consumers had no credit files (often called credit invisibles) and 8.3% had such limited files that they could not be scored under widely used third-party models (often called unscorables).[34] Applying these percentages to the 2023 U.S. adult population yields an estimated 50 million individuals without credit scores under the third party models that are in broadest use today. This represents a substantial barrier to financial inclusion, as these individuals may struggle to access mainstream credit products and services. More broadly, consumers whose credit reports reflect two or fewer "tradeline" accounts may have credit scores but may be subject to additional underwriting requirements or different models because their default risk is harder to predict that consumers with more extensive traditional credit experience.[35]

Cash only models, which rely on bank account data rather than traditional credit bureau information, offer a potentially transformative solution to this challenge by enabling lenders to assess the creditworthiness of these underserved populations. Estimates derived from the 2023 Survey of Household Economics and Decisionmaking (SHED) suggest that between 18 to 40 million U.S. adults who lack robust credit histories could be scored using bank account data.[36] This range reflects varying assumptions about the overlap between credit invisibility/unscoreability and access to bank accounts:

- **Conservative Estimate:** SHED data indicate that approximately 6.4% of U.S. adults do not have a bank account. If we assume that all of these individuals cannot be scored under the most widely used third-party models, the remaining 36% of the credit invisible/unscorable population (18 million adults) could be scored using cash only models.

- **Optimistic Estimate:** A more generous estimate suggests that roughly 80% (40 million adults) of the population who cannot be scored under the most widely used third-party models could be underwritten using cash only models. According to the 2023 SHED report, 17.9% of those who were not confident about being approved for a credit card did not have a bank account. The remainder, however, would likely have sufficient bank account activity to be assessed using cash only models.

Estimates based on the Federal Deposit Insurance Corporation's periodic survey of unbanked and underbanked households similarly yield an estimate of between 29 and 32 million adults who have not used mainstream credit in the past year but have bank statement data that could be used for underwriting.[37]

These results highlight the potential of cash only models to significantly expand the scorable population, particularly for individuals who are excluded from traditional credit systems due to thin or nonexistent credit files. By leveraging bank account data, these models provide a more inclusive approach to credit risk assessment, enabling lenders to serve previously untapped markets and promote responsible credit access.

5.2 Managing models for explainability and fairness

While this research was designed to probe the potential predictiveness benefits of integrating cash flow data and machine learning techniques into credit risk models, it is important to acknowledge that lenders make a range of development decisions based on additional considerations, such as the need for explainability to help manage models for both business and compliance purposes, fair lending compliance, and practical considerations based on their technology infrastructure, data availability considerations, expertise, and other topics. This section briefly outlines some of the considerations and processes that lenders use.[38]

5.2.1 Explainability

Explainability is both required at an individual consumer level when providing adverse action notices under the Equal Credit Opportunity Act and Fair Credit Reporting Act and helps to facilitate compliance with other regulatory regimes such as fair lending and model risk management as developers analyze how and why models generate their predictions. Lenders and regulators have become accustomed to working with logistic regression models over several decades to produce various types of analyses that assess which features are driving models' operations for these various purposes. While there can be complexities in working with logistic regression models and critics point out that common processes have some limitations and drawbacks, such models are considered the easiest to manage because of their long track record.

However, depending on their configuration, machine learning models are often substantially more complex than traditional approaches because they can be built to account for more nuanced relationships in the data and to incorporate substantially larger numbers of features (whether traditional or non-traditional). During the development process, decisions about the type of machine learning model used, the complexity of architecture (such as number of layers or nodes), and the number of inputs can substantially impact the explainability of the underwriting model. Many lenders also apply what are called post hoc explainability tools to analyze feature importance and other aspects of model operations without changing the structure or nature of those operations. These are complex and computationally intensive statistical techniques in their own right, and past

FinRegLab research emphasizes the importance of making thoughtful choices about tool selection, deployment, and interpretation of results to answer different explainability questions for different models. More broadly, some critics argue that the techniques are not sufficient for fully understanding highly complex models and that relying heavily on structural constraints to build what are sometimes called inherently interpretable models is the most responsible option, despite concerns about potential tradeoffs on performance.[39]

In practice, while some lenders may rely nearly entirely on structural constraints to manage for explainability and others nearly entirely on post hoc tools, the most common approach appears to be somewhere in the middle by deploying both strategies. Nevertheless, there is substantial variation among different types of lenders. For instance, based on interviews and limited published reports it appears substantially more common for non-banks to deploy models that incorporate thousands of features, while banks may evaluate large numbers of potential variables at the outset but frequently winnow their final models down to a much more constrained number of inputs.

In this study, we focused on building models to maximize predictive performance as measured by metrics such as ROC-AUC and KS statistics, without diving deeply into feature importance analyses or imposing constraints on the machine learning models solely to improve explainability. For example, we did not constrain the models for monotonicity.[40]

5.2.2 Fairness

In light of fair lending laws, lenders may deploy a range of analyses at certain stages of the model development process to consider whether the input features used in a particular model could be considered a proxy for a protected characteristic (which raises potential concerns about disparate treatment liability) or are causing disproportionate adverse effects on the basis of such characteristics (which raises potential disparate impact concerns). In the latter case, lenders may also analyze whether there are alternate versions of the inputs or other data that could achieve reasonably comparable levels of predictiveness with fewer adverse effects. This is often called searching for a less discriminatory alternative.

Understanding the impacts of the individual features on model operation is a basic component of these analyses, and for the reasons described above can become more complicated in the context of machine learning models. Lenders often deploy post hoc explainability techniques to assist in this process, as well as monitoring other approaches to measuring and mitigating fair lending risk that have been spurred by broader data science advances in working with machine learning and artificial intelligence across a range of fields.

5.3 Implications for lenders: Staggered strategies for adoption

While our findings suggest that the integration of both cash flow data and machine learning methodologies represents the most robust strategy for enhancing credit risk assessment, we recognize that many institutions may be hesitant to adopt both innovations at the same time. Smaller lenders, for example, may lack the internal data or computational resources required either to develop and deploy full-scale ML models for themselves or to manage vendors in doing so. Lenders of various sizes may also face challenges related to compliance, system limitations, or organizational readiness that make them decide to prioritize one innovation over another. Small lenders may also see substantial value in incorporating third-party credit scores into their underwriting systems, since such scores are calibrated on datasets that are much larger than the individual lenders may be able to access on their own, and are widely used by investors and other secondary market actors facilitate comparisons across loan portfolios.

The performance and credit access results discussed in Section 4 suggest that there could be meaningful benefits to focusing on whichever innovation is most manageable for the lender to prioritize in adjusting their underwriting systems on a staged basis. For example:

» **Leveraging Hybrid Logistic Regression Models.** For institutions that are unwilling to adopt full-scale ML models, the findings suggest that moving to a hybrid logistic regression model that combines a traditional credit score with cash flow data may offer benefits over continuing solely with traditional data inputs. These models provided significant incremental benefits over a traditional credit score alone, particularly for subprime and near-prime borrowers, without requiring the computational complexity or interpretability challenges of ML. For example, the credit score + cash flow LR model achieved meaningful improvements in predictive accuracy and credit access, particularly for consumers with derogatory credit histories. This approach potentially allows lenders to harness the benefits of cash flow data while maintaining the simplicity and transparency of traditional LR models, making it easier to ensure compliance with regulatory requirements.

» **Adopting ML Models Built on Traditional Credit Data.** Institutions constrained by their data platforms to credit bureau inputs can still potentially benefit from machine learning by developing credit only ML models. These models, while not incorporating cash flow data, substantially outperformed LR credit only models in predictive accuracy and risk separation. For example, the credit only ML model demonstrated higher separation and predictive power compared to the credit only LR model, particularly for consumers with complex credit profiles. This strategy is particularly suitable for lenders with limited access to alternative data sources but who wish to modernize their underwriting processes and improve risk assessment.

Although the process of updating underwriting models can be daunting particularly for smaller lenders, the range of support services is continuing to expand over time. For example, a broad range of credit scoring development companies, credit bureaus, and data aggregators now offer cash flow attributes, scoring models, and other support services to lenders who are seeking to use the data for credit underwriting purposes.[41] Vendors also provide a broad spectrum of support services for lenders that are interested in adopting machine learning models, including providing platforms and tools for lenders to use in developing their own models as well as providing tailored advice and testing services, working with lenders to develop customized models, and in some cases providing more "off the shelf" proprietary models.[42]

Materials to help lenders think through key implementation considerations are also increasing. FinRegLab recently published a qualitative paper focusing on lessons learned by community development financial institutions and minority depository institutions in incorporating electronic cash flow data and adopting more sophisticated loan origination platforms that may be helpful to smaller lenders that are considering similar changes. Although the paper is focused on small business lending, many of the basic considerations in working with applicants to obtain the data, deciding how to deploy it within different types of underwriting systems, and managing vendor selection and internal change management are equally applicable in the consumer credit context.[43] Other implementation resources and white papers are also available.[44]

6. CONCLUSION

This study probes the potential impacts of integrating cash flow data and machine learning techniques into consumer lending underwriting models. The findings suggest that combining these two innovations produces the most substantial improvements in both predictive accuracy and credit access.

The research measures modest but meaningful improvements in predictiveness from hybrid models—those blending cash flow data with traditional credit bureau information—compared to traditional approaches. While we cannot fully measure the impacts of this innovation across the entire spectrum of borrowers due to sample limitations, the results are encouraging even for a sample population that is skewed toward borrowers with relatively strong credit scores and greater wealth.

We also find substantial improvements in underwriting performance by machine learning models compared to logistic regression models across most data sources. The model that combines both of these innovations—the ML hybrid model that incorporates both cash flow and credit bureau data—produced the largest and most widespread increases in general predictiveness overall and across a range of subgroups, followed by the ML model that used only credit bureau data.

The research also traces the impacts on credit access of both hybrid models and machine learning models, finding that they both tend to increase credit approvals relative to traditional models using traditional data particularly at lower risk/higher score cutoffs used by most mainstream lenders. Positive impacts on credit approval rates tend to get somewhat smaller at higher risk thresholds, but the ML models in particular tend to reduce approvals of defaulters at these higher cutoffs which helps to reduce the number of consumers who are offered loans that they are not likely to be able to repay and avoid negative consequences from failed loans.

By tackling these issues, we hope to advance the responsible adoption of data and analytical innovations in consumer lending. The ultimate goal is to better understand how data and technology can be used to increase access to responsible credit while meeting regulatory expectations and business needs concerning reliability, explainability, fairness, and trust. This research represents a step toward that vision, offering valuable insights for lenders, technologists, researchers, advocates and policymakers as they consider how to facilitate customer friendly innovation going forward.

APPENDIX A

Literature Survey

Academic and other public research on machine learning models for credit underwriting has expanded substantially in the last several years both in the United States and internationally, with dozens of papers in the past decade. However, much of the research focuses on the benefits of different model types on general predictiveness without looking in equal detail at potential impacts for credit access. Available research on the latter topic also shows a wide range of results when using only traditional data sources. In contrast, research on the impacts of using electronic bank account records or other forms of cash flow data is much more limited in scale but suggests that there are positive impacts for both predictiveness and access, particularly for applicants who likely would otherwise be rejected because their traditional credit bureau data is limited.

This section provides an overview of this literature, as well as a brief discussion of research on methodological and compliance issues in credit underwriting and machine learning models more generally. The current paper contributes to this literature both because our analysis has worked to take account of industry practices in building the underwriting models and estimating access impacts and because we provide a systematic comparison of moving from logistic regression to machine learning models using both traditional credit bureau data and cash flow information.

A.1 Machine learning literature

Recent academic surveys have tallied more than 100 papers in the past decade focusing on the use of machine learning models for consumer and commercial credit underwriting, using a wide variety of techniques ranging from XGBoost and random forest models, neural networks, and ensemble models that incorporate multiple types of algorithms.[45] The academic literature generally finds substantial performance gains from adopting these more sophisticated models. Some papers also explore impacts on access to credit and the potential tradeoffs between predictive power, explainability, and fairness considerations in credit underwriting contexts, but these topics have not been covered in as much depth. (For a brief discussion of data science literature on managing expainability and fairness concerns in machine learning models more generally, see Appendix A.3.)

Where impacts on credit access have been a primary focus of analysis, papers that evaluate machine learning underwriting models built using credit bureau and other traditional data attributes have reached substantially different conclusions. In Fuster et al., the authors built a range of models varying from logistic regression through random forest machine learning models using traditional mortgage underwriting attributes including income, credit score, and loan-to-value ratios. They found acceptance rates increased by modest amounts across all groups under the more predictive machine learning models, with more pronounced gains for Black consumers. However, the more accurate models also produced a broader range in the predicted probability of default than the more traditional models, and Black and Hispanic borrowers were less likely to be scored as having substantially lower default propensities than White borrowers under the new models. This pattern resulted in a broader range of interest rates under risk-based pricing models, and increased net pricing disparities between demographic groups.[46]

Blattner and Nelson analyzed variations in the predictiveness of traditional credit scoring models for low-income and minority populations using traditional credit bureau records merged with data regarding mortgage and real estate transactions. They concluded that data sparsity and "noise" in traditional credit bureau records for those populations accounted for up to 70% of the variation depending on the subgroup analyzed. They also built machine learning models based on the credit bureau files to test whether more powerful analytics could help to overcome the data issues. They found that the XGBoost model produced "modest" improvements in accuracy and approval rates, but was not able to close the gaps between the underlying populations particularly when compared to simulations that more directly addressed the underlying data gaps. The paper also found limited impacts from incorporating data reflecting payday loan borrowing that is not included in traditional credit bureau files but noted that sources such as cash flow information from bank accounts may be more informative.[47]

In contrast, Albanesi and Vamossy found that their machine learning model built with traditional credit bureau attributes substantially improved upon predictiveness compared to a commercially available scoring model, both overall and when applied to data from the 2008 financial crisis. Compared to traditional credit scores, the machine learning model put greater weight on overall debt levels and less weight on mix of credit types, the incidence of new credit, and credit history. This mix produced greater accuracy and more favorable risk assessments for young, low income, and in most cases minority borrowers.[48]

A.2 Cash flow data literature

The primary source of public research on the impact of cash flow data on credit underwriting in the United States consists of FinRegLab's prior reports, including a 2019 independent analysis of loan origination and performance data from several consumer and small business lenders and a 2025 independent analysis of loan origination and performance data from two fintech small business lenders.[49] The collective findings include:

- The cash flow metrics studied had substantial predictive value that was not fully replicated by the predictive insights from traditional credit bureau data and scores, such that adding the two sources of data together generally provides additional lift over either source in isolation. This suggests that cash flow data both provide a useful basis for underwriting applicants who have limited or no traditional data and can improve underwriting for substantial numbers of applicants who do have more robust credit histories.
- The lenders in question were able to serve populations who historically were likely to struggle to access mainstream credit as measured using a variety of benchmarks such as income, credit score, and residence or business location in economically disadvantaged neighborhoods.
- In the first study, available data suggested that the degree to which the cash flow data was predictive of credit risk appeared to be relatively consistent across borrowers who likely belong to different demographic groups.
- In the second study, increases in predictive performance were higher for small business owners with lower credit scores and particularly large for low-score owners with younger

businesses. Those businesses often face particular challenges in accessing credit because they are considered higher risk, affecting their ability to bridge short-term needs and leverage expansion opportunities to create more jobs and community impact.

Other available sources also suggest that cash flow data can improve predictiveness and access, but are more limited in scope or focus on international settings. A 2023 study by the Consumer Financial Protection Bureau found that proxies for cashflow data reported by participants in a large-scale consumer survey added predictive value to credit bureau data, but did not have direct access to the underlying bank account information.[50] Some providers of cash flow scores and attributes have also released information measuring predictiveness and access impacts.[51] International research has also found positive impacts from using bank account data and other sources of information regarding applicants' non-credit financial transactions, both in underwriting consumers who lack traditional credit history and in combining with traditional credit history where available.[52]

A.3 Literature on methodological and compliance issues

A number of academic surveys also tally recent work on data science techniques for managing fairness and explainability concerns in working with machine learning models.[53] While much of this work is not grounded specifically in financial services, some studies focus on the credit underwriting context in part because these issues are so directly implicated by regulatory requirements as discussed above.[54] A number of papers are also focusing on specific methodological or compliance issues in the lending context such as reject inference[55] and disparate impact analyses and alternative approaches to fair lending testing.[56]

APPENDIX B

Additional Cash Only Model Results

TABLE B.1 MODEL PERFORMANCE BY PAST CREDIT HISTORY

RECENT DEROG	ROC-AUC		KS	
MODEL	CASH ONLY		CASH ONLY	
LR	.6427		.2379	
ML	.6496		.2486	
Δ	.0069	1.07%	.0107	4.50%

OLDER DEROG	ROC-AUC		KS	
MODEL	CASH ONLY		CASH ONLY	
LR	.6743		.2632	
ML	.6945		.3036	
Δ	*.0202*	*3.00%*	*.0404*	*15.35%*

NEVER DEROG	ROC-AUC		KS	
MODEL	CASH ONLY		CASH ONLY	
LR	.7731		.4219	
ML	.7922		.4604	
Δ	*.0191*	*2.47%*	*.0385*	*9.13%*

TABLE B.2 MODEL PERFORMANCE BY INCOME LEVEL

LMI	ROC-AUC		KS	
MODEL	CASH ONLY		CASH ONLY	
LR	.7484		.3708	
ML	.7627		.4032	
Δ	*.0143*	*1.91%*	*.0324*	*8.74%*

MUI	ROC-AUC		KS	
MODEL	CASH ONLY		CASH ONLY	
LR	.7717		.4138	
ML	.7897		.4499	
Δ	*.0180*	*2.33%*	*.0361*	*8.72%*

TABLE B.3 MODEL PERFORMANCE BY RACE/ETHNICITY

AA/H	ROC-AUC		KS	
MODEL	CASH ONLY		CASH ONLY	
LR	.7606		.4105	
ML	.7811		.4488	
Δ	*.0205*	*2.70%*	*.0383*	*9.33%*

CAUCASIAN	ROC-AUC		KS	
MODEL	CASH ONLY		CASH ONLY	
LR	.7904		.4504	
ML	.8037		.4732	
Δ	*.0133*	*1.68%*	*.0228*	*5.06%*

TABLE B.4 CREDIT ACCESS ANALYSIS—OVERALL POPULATION

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	68.39%		12.51%	
ML	68.72%		11.74%	
Δ	*.3pp*	*0.48%*	*-.8pp*	*-6.16%*

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	81.19%		22.45%	
ML	80.20%		19.35%	
Δ	*-1.0pp*	*-1.22%*	*-3.1pp*	*-13.81%*

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	87.95%		30.95%	
ML	86.37%		27.14%	
Δ	*-1.6pp*	*-1.80%*	*-3.8pp*	*-12.31%*

TABLE B.5 CREDIT ACCESS ANALYSIS—RECENT DEROGATORY

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	33.60%		1.83%	
ML	31.59%		1.95%	
Δ	-2.0pp	-5.98%	.1pp	6.56%

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	52.46%		5.53%	
ML	48.04%		4.70%	
Δ	-4.4pp	-8.43%	-.8pp	-15.01%

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	65.73%		11.33%	
ML	60.54%		9.58%	
Δ	-5.2pp	-7.90%	-1.8pp	-15.45%

TABLE B.6 CREDIT ACCESS ANALYSIS—OLDER DEROGATORY

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	56.80%		15.94%	
ML	56.62%		15.27%	
Δ	-.2pp	-0.32%	-.7pp	-4.20%

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	74.98%		31.91%	
ML	73.35%		26.75%	
Δ	-1.6pp	-2.17%	-5.2pp	-16.17%

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	84.39%		44.40%	
ML	82.23%		37.26%	
Δ	-2.2pp	-2.56%	-7.1pp	-16.08%

TABLE B.7 CREDIT ACCESS ANALYSIS—NEVER DEROGATORY

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	84.68%		43.32%	
ML	85.98%		39.10%	
Δ	*1.3pp*	*1.54%*	-4.2pp	-9.74%

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	92.99%		62.55%	
ML	93.35%		55.73%	
Δ	*.4pp*	*0.39%*	-6.8pp	-10.90%

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	96.46%		72.15%	
ML	96.27%		68.12%	
Δ	*-.2pp*	*-0.20%*	*-4.0pp*	*-5.59%*

TABLE B.8 CREDIT ACCESS ANALYSIS—LMI

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	34.88%		5.19%	
ML	36.30%		4.78%	
Δ	*1.4pp*	*4.07%*	-.4pp	-7.90%

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	53.60%		10.61%	
ML	52.73%		9.18%	
Δ	*-.9pp*	*-1.62%*	*-1.4pp*	*-13.48%*

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	66.65%		16.45%	
ML	63.39%		14.41%	
Δ	*-3.3pp*	*-4.89%*	*-2.0pp*	*-12.40%*

TABLE B.9 CREDIT ACCESS ANALYSIS—MUI

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	74.66%		15.53%	
ML	74.79%		14.61%	
Δ	.1pp	0.17%	*-.9pp*	*-5.92%*

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	86.36%		27.33%	
ML	85.34%		23.54%	
Δ	*-1.0pp*	*-1.18%*	*-3.8pp*	*-13.87%*

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	91.94%		36.94%	
ML	90.67%		32.39%	
Δ	*-1.3pp*	*-1.38%*	*-4.6pp*	*-12.32%*

TABLE B.10 CREDIT ACCESS ANALYSIS—AFRICAN-AMERICAN/HISPANIC

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	61.22%		10.30%	
ML	61.35%		7.84%	
Δ	.1pp	0.21%	-2.5pp	-23.88%

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	76.89%		18.69%	
ML	75.38%		16.43%	
Δ	*-1.5pp*	*-1.96%*	*-2.3pp*	*-12.09%*

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	85.11%		27.20%	
ML	82.87%		24.38%	
Δ	*-2.2pp*	*-2.63%*	*-2.8pp*	*-10.37%*

TABLE B.11 CREDIT ACCESS ANALYSIS—CAUCASIAN

3% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CASH ONLY		CASH ONLY	
LR	71.35%		14.03%	
ML	71.84%		13.82%	
Δ	*.5pp*	*0.69%*	*-.2pp*	*-1.50%*

5% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	83.43%		25.28%	
ML	82.69%		21.45%	
Δ	*-.7pp*	*-0.89%*	*-3.8pp*	*-15.15%*

7% CUTOFF	APPROVAL RATES		% OF APPROVALS THAT DEFAULT	
MODEL	CS + CASH		CS + CASH	
LR	89.66%		34.17%	
ML	88.35%		29.72%	
Δ	*-1.3pp*	*-1.46%*	*-4.5pp*	*-13.02%*

Endnotes

1 This analysis builds most directly on two rounds of prior FinRegLab empirical research. *See* FinRegLab, "The Use of Cash-Flow Data in Underwriting Credit: Empirical Research Findings" (analyzing loan performance data from both consumer and small business lenders that incorporated cash flow data in their underwriting models); FinRegLab et al., "Machine Learning Explainability & Fairness: Insights from Consumer Lending" (analyzing the use of secondary tools for managing explainability and fairness concerns about machine learning underwriting models).

2 Calculations based on origination data published in Consumer Financial Protection Bureau, "Consumer Credit Trends," and score distributions in Federal Reserve Bank of New York, "Quarterly Report on Household Debt and Credit 2025: Q1."

3 Consumer Financial Protection Bureau, "Technical Correction and Update to the CFPB's Credit Invisibles Estimate."; Hepinstall et al., "Financial Inclusion and Access to Credit."; Toh, "Addressing Traditional Credit Scores as a Barrier to Accessing Affordable Credit." Equifax, "Achieve More Inclusive Credit: Expanding Your Credit Reach to Marginalized Populations Can Help Them and You."

4 For background, *see* FinRegLab, "The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis," §§ 2.1, 2.3, 3.

5 For background, *see* FinRegLab, "The Use of Machine Learning for Credit Underwriting: Market & Data Science Context," §§ 2.1, 2.4.

6 The CFPB did not named the credit score it used but indicated that it was widely used and relatively conservative. Kenneth P. Brevoort et al., "Credit Invisibles and the Unscored."

7 Consumer Financial Protection Bureau, "Technical Correction and Update to the CFPB's Credit Invisibles Estimate."

8 Equifax, "Achieve More Inclusive Credit: Expanding Your Credit Reach to Marginalized Populations Can Help Them and You."

9 Equifax, "Achieve More Inclusive Credit: Expanding Your Credit Reach to Marginalized Populations Can Help Them and You."; Hepinstall et al., "Financial Inclusion and Access to Credit."

10 For background, *see* FinRegLab, "Explainability & Fairness in Machine Learning for Credit Underwriting: Policy Analysis," §§ 4-6; FinRegLab, "The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis," §§ 5-7.

11 15 U.S.C. § 1691(d); 12 CFR § 1002.9.

12 15 U.S.C. § 1681m(a)(1), (b), (h); 12 CFR § 1022.72.

13 15 U.S.C. § 1961(a) (prohibiting discrimination on the basis of race, color, national origin, religion, sex, marital status, or age or because of the receipt of public assistance or the good faith exercise of certain rights under federal consumer financial law); 42 U.S.C. § 3605 (prohibiting discrimination on the basis of race, color, national origin, religion, sex, familial status or disability).

14 Regulations under ECOA also set forth standards for certain "empirically derived, demonstrably and statistically sound, credit scoring system(s)." In developing underwriting models, lenders and model developers may follow those standards even if they are not legally required to do so. 12 C.F.R. 1002.2(p), 1002.6(b)(ii).

15 Executive Office of the President, "Restoring Equality of Opportunity and Meritocracy."

16 Board of Governors of the Federal Reserve System, Supervisory & Regulation Letter 11-7, "Guidance on Model Risk Management."; Office of the Comptroller of the Currency, Bulletin 2011-12, "Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Management."; Federal Deposit Insurance Corporation, Financial Institution Letter 22-2017 "Adoption of Supervisory Guidance on Model Risk Management."

17 12 U.S.C. §§ 1861–1867, 5516(e).

18 FinRegLab, "The Use of Cash-Flow Data in Underwriting Credit: Empirical Research Findings."; FinReg Lab et al., "Sharpening the Focus: Using Cash-Flow Data to Underwrite Financially Constrained Businesses."; Alexandrov et al., "Looking at Credit Scores Only Tells Part of the Story – Cashflow Data May Tell Another Part."; Johnson, "Everything You Ever Wanted to Know About Cash Flow Underwriting But Were Afraid to Ask."

19 *See, e.g.*, Albanesi and Vamossy, "Credit Scores: Performance and Equity."; Chang et al., Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers."; de Lange et al., "Explainable AI for Credit Assessment of Banks."; Bussmann et al., "Explainable Machine Learning in Credit Risk Management."; Butaru et al., "Risk and Risk Management in the Credit Card Industry."; Khandani et al., "Consumer Credit-Risk Models via Machine-Learning Algorithms."

20 FinRegLab et al., "Machine Learning Explainability & Fairness: Insights from Consumer Lending." *See also* Alonso and Carbó, "Accuracy of Explanations of Machine Learning Models for Credit Decisions."

21 *See, e.g.*, Albanesi and Vamossy, "Credit Scores & Equity."; Blattner and Nelson, "How Costly Is Noise? Data and Disparities in Consumer Credit."; Fuster et al., "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *See also* Meursault et al., "One Threshold Doesn't Fit All: Tailoring Machine Learning Predictions of Consumer Default for Lower-Income Areas."

22 Note that the same consumer could appear in the dataset multiple times and be counted as separate observations if he/she had new originations in different months with the sampling period. "Consumer" therefore designates an individual consumer at a specific time. About 58% of the consumers (263,327 out of 453,804 used in the final sample as described in Section 3.1.2) appeared only once in the data.

23 The credit bureau offers the dataset for non-underwriting purposes. The inferred demographics are based on an analysis of first name, last name, and zip code information to infer race, ethnicity, and various other affiliations. The credit bureau did not provide names or geographic information to FinRegLab.

24 Lenders and model developers vary as to the minimum amount of cash flow history they require for underwriting, with three months and six months as common general thresholds. In this study we chose to focus on six months to ensure a more robust baseline. We estimate that shrunk the sample size by about 7 percent.

25 Not every account had all daily balance records. We imputed end-of-day balances for dates when such balance records were missing. The imputation process is described in Appendix C.3.1.

26 Banks particularly emphasize parsimonious models as being easier to understand and manage, while often reducing the risk of overfitting training data so that they remain robust even when external conditions start to change.

27 The credit only LR model does not generate a prediction for consumers without sufficient credit bureau information, namely, those whose credit profile contains no credit tradelines, who only have tradelines that are less than six months old, or have no reported tradeline activity within the last six months. This aligns with certain commercially available credit scoring models.

28 Unlike credit bureau features/attributes, the cash flow features we developed do not have apparent sparsity patterns in terms of being missing for a large segment of the consumers in the sample. One example of such a sparsity pattern on the credit bureau side is that the credit features/attributes related to the severity or timing of delinquency events are all missing for consumers without a delinquency event. Having one separate model segment for these consumers obviate the need to include features related to the severity or timing of delinquency in the equation. The cash flow features we developed, on the other hand, tend to be densely populated for all consumers in the sample.

29 In the context of statistical modeling, generalization performance refers to a model's ability to make accurate predictions on unseen data, reflecting how well it captures underlying patterns rather than memorizing what is effectively noise in the training dataset. Strong generalization indicates robustness across diverse datasets, while poor generalization suggests overfitting, where a model performs well on training data but fails on new observations.

30 The separate data set that included certain demographic data for purposes of subgroup analysis was not used for model building either.

31 If a difference is negative, it is considered significant if the 95th percentile of the difference is negative; if a difference is positive, it is considered significant if the 5th percentile of the difference is positive.

32 Specifically, the recalibration was performed by estimating a logistic regression model with an intercept and the log-odds output of the model to be calibrated as inputs. The resulting logistic regression uses a shift and a scale parameter to translate the model output into probabilities of default (PD) as observed in the test sample. In the case of the traditional credit score benchmark, we used the score as the input, as it is designed to have a linear relationship with the log-odds.

33 The reasons for this preference can be seen in Table 4.6 in Section 4.2.2. The rate at which the cash only ML model would approve consumers who went on to default was roughly double the rates for the ML models that incorporate credit bureau data in some form. Accordingly, lenders are likely to want to use models that incorporate the traditional data where it is available. *See* Section 5.1 for further discussion.

34 Consumer Financial Protection Bureau, "Data Point: Credit Invisibles."; *see also* Brevoort et al., "Credit Invisibles and the Unscored."

35 *See* Endnote 2 for additional discussion and more recent sources that update portions of the CFPB's estimates.

36 Board of Governors of the Federal Reserve System, "Economic Well-Being of U.S. Households in 2023." We chose to use the SHED results to estimate the potential impact because the survey uses individual weights, which make the ratios directly applicable to the US adult population.

37 Federal Deposit Insurance Corporation, "2023 FDIC National Survey of Unbanked and Underbanked Households." The 2023 FDIC study, which uses household weights, estimates that about 15.7% of households had no mainstream credit in the past 12 months. The average of number of adults of these households was 1.7, slightly lower than the 1.9 for the population with access to mainstream credit. For these households, about 21% were unbanked, with nearly an identical average number of adults for the unbanked and banked households. If we ignore the differences in the average number of adults within households, this suggests that, out of the 262 million U.S. adult population, about 32 million adults who had no mainstream credit could be underwritten using models based on their bank statement data. If we account for the smaller household size, the number reduces to 29 million.

38 For more detailed discussions of these issues, *see* FinRegLab, "Explainability & Fairness in Machine Learning for Credit Underwriting: Policy Analysis."

39 FinRegLab, "Explainability & Fairness in Machine Learning for Credit Underwriting: Policy Analysis;" FinRegLab, "The Use of Machine Learning for Credit Underwriting: Market & Data Science Context."

40 Monotonicity assumes that the relationship between a given input variable and the predicted outcome is consistent in direction no matter what the specific values. This makes a model easier to explain but may not be consistent with real world patterns. For example, it is possible that medium rates of credit utilization might be associated with lower rates of default risk than very low rates (since they do not provide much signal on how the applicant would handle higher levels of debt) or very high rates (which may suggest that the applicant would have a difficult time managing additional debt payments).

41 Examples from credit scoring developers and credit bureau companies include EdgeScore, Experian's Cashflow Score and Cashflow Attributes, FICO's UltraFico, Nova Credit's Cash Atlas, Prism's CashScore, and VantageScore's 4Plus model. In addition, data aggregators that act as intermediaries in transferring bank account data to other financial service providers on behalf of consumers (such as Finicity, Plaid, and Yodlee) frequently offer services to help categorize transactions and in some cases provide attributes or metrics for potential use in credit underwriting. They increasingly are following requirements for credit bureaus under the Fair Credit Reporting Act. *See* FinRegLab, "The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis."; Johnson, "Everyone Wants to Be the FICO of Cash Flow Data."; Experian, "New Experian Tool Empowers Financial Inclusion Through Open Banking Insights."; Experian, "Launch of Experian's Cashflow Score Signals New Era of Open Banking-Powered Lending."; VantageScore, "Credit Bureaus + Bank Data = VantageScore 4 Plus."; Taylor and Sriram, "Introducing Transactions for Business: Powering Real-Time Business Tools." Plaid Check, "Consumer Report."

42 *See* generally FinRegLab, "The Use of Machine Learning for Credit Underwriting: Market & Data Science Context."

43 FinRegLab, "Transforming Small Business Credit: Technology and Data Adoption in Mission-Based Lenders."

44 *See, e.g.*, Bates, "Cash Flow Underwriting: A Practical Credit Risk Implementation Guide for Lenders."; Gote and Mendhe, "Building a Cash Flow Underwriting System: Insights from Implementation."

45 Shi et al., "Machine Learning Driven Credit Risk: A Systemic Review."; Bari et al, "A Systematic Literature Review of Predictive Models and Analytics in AI Driven Credit Scoring." *See also* Wang and Perkins, "How Magic a Bullet Is Machine Learning for Credit Analysis? An Exploration with FinTech Lending Data"; Meursault et al., "One Threshold Doesn't Fit All: Tailoring Machine Learning Predictions of Consumer Default for Lower-Income Areas."

46 Fuster et al., "Predictably Unequal? The Effects of Machine Learning on Credit Markets."

47 Blattner and Nelson, "How Costly Is Noise? Data and Disparities in Consumer Credit."

48 Albanesi and Vamossy, "Credit Scores & Equity."

49 FinRegLab, "The Use of Cash-Flow Data in Underwriting Credit: Empirical Research Findings."; FinReg Lab et al., "Sharpening the Focus: Using Cash-Flow Data to Underwrite Financially Constrained Businesses."

50 Alexandrov et al., "Looking at Credit Scores Only Tells Part of the Story – Cashflow Data May Tell Another Part."

51 Johnson, "Everything You Ever Wanted to Know About Cash Flow Underwriting But Were Afraid to Ask."

52 The World Bank Group, "The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges."; Lee et al., "Who Benefits from Alternative Data for Credit Scoring? Evidence from Peru."; Chioda et al., "FinTech Lending to Borrowers with No Credit History."

53 For explainability resources *see* Yang et al., "Survey on Explainable AI: From Approaches, Limitations and Applications Aspects."; Abusitta et al., "Survey on Explainable AI: Techniques, Challenges, and Open Issues."; Burkart and Huber, "A Survey on the Explainability of Supervised Machine Learning." For fairness and bias research *see* Siddique et al., "Survey on Machine Learning Biases and Mitigation Techniques."; Bogiatzis-Gibbons et al., "A Literature Review on Bias in Supervised Machine Learning."; Mavrogiorgos et al., "Bias in Machine Learning: A Literature Review."

54 *See, e.g.*, FinRegLab et al., "Machine Learning Explainability & Fairness: Insights from Consumer Lending."; Bicharra Garcia et al., "Algorithmic Discrimination in the Credit Domain: What Do We Know About It?"; Braak et al., "How Can Consumers Without Credit History Benefit from the Use of Information Processing and Machine Learning Tools by Fnancial Institutions?"; Dessain et al., "Cost of Explainability in AI: An Example with Credit Scoring Models."; Nallakaruppan et al., "Credit Risk Assessment and Financial Decision Support Using Explainable Artificial Intelligence."; Davis et al., "Explainable Machine Learning Models of Consumer Credit Risk."

55 Liao et al., "Data Augmentation Methods for Reject Inference in Credit Risk Models."; Wu et al., "A 'Divide and Conquer' Reject Inference Approach Leveraging Graph-Based Semi-Supervised Learning."; Anderson, "Using Bayesian Networks to Perform Reject Inference."

56 Caro et al., "Differential Validity in Fair Lending."; Caro et al., "Modernizing Fair Lending." *See also* Aigner et al., "Statistical Approaches for Assessing Disparate Impact in Fair Housing Cases."

Bibliography

Abusitta, Adel, Miles Q. Li, and Benjamin C.M. Fung. "Survey on Explainable AI: Techniques, Challenges, and Open Issues." *Expert Systems with Applications*, December 2024.

Aigner, Dennis J., Marco del Ángel, and Joel Wiles. "Statistical Approaches for Assessing Disparate Impact in Fair Housing Cases." *Statistics and Public Policy*, September 2023.

Albanesi, Stefania and Domonkos F. Vamossy. "Credit Scores: Performance and Equity." National Bureau of Economic Research, September 2024.

Alexandrov, Alexei, Alyssa Brown, and Samyak Jain. "Looking at Credit Scores Only Tells Part of the Story – Cashflow Data May Tell Another Part." Consumer Financial Protection Bureau, July 2023.

Alonso, Andrés and José Manuel Carbó. "Accuracy of Explanations of Machine Learning Models for Credit Decisions." Banco de España, June 2022.

Anderson, Billie. "Using Bayesian Networks to Perform Reject Inference." December 2019.

Bari, Md Hasanujamman, Shaharima Juthi, Asha Moni Mistry, and Md Kamrujjaman. "A Systematic Literature Review of Predictive Models and Analytics in AI Driven Credit Scoring." *Journal of Machine Learning, Data Engineering and Data Science*, October 2024.

Bates, Tim. "Cash Flow Underwriting: A Practical Credit Risk Implementation Guide for Lenders." *Fintech Takes*, 2025.

Bicharra Garcia, Ana Cristina, MarcioGomes Pinto Garcia, and Roberto Rigobon. "Algorithmic Discrimination in the Credit Domain: What Do We Know About It?" *AI & Society*, May 2023.

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

Board of Governors of the Federal Reserve System. "Economic Well-Being of U.S. Households in 2023." May 2024.

Board of Governors of the Federal Reserve System. "Guidance on Model Risk Management." Supervisory & Regulation Letter 11-7, April 2011.

Bogiatzis-Gibbons, Daniel, et al. "A Literature Review on Bias in Supervised Machine Learning." Financial Conduct Authority, November 2024.

Brevoort, Kenneth, Philipp Grimm, and Michelle Kambara. "Credit Invisibles and the Unscored." Cityscape, Vol. 18, No. 2.

Brevoort, Kenneth, Philipp Grimm, and Michelle Kambara. "Data Point: Credit Invisibles." Consumer Financial Protection Bureau, Office of Research, May 2015.

Burkart, Nadia, and Marco F. Huber. "A Survey on the Explainability of Supervised Machine Learning." *Journal of Artificial Intelligence Research*, January 2021.

Bussmann, Niklaus, Paolo Giudici, Dimitri Marinelli, and Jochen Papenbrock. "Explainable Machine Learning in Credit Risk Management." *Computational Economics*, January 2021.

Butaru, Florentin, Qinqing Chen, Brian Clark, Sanmay Das, Andrew W. Lo, and Akhtar Siddique. "Risk and Risk Management in the Credit Card Industry." *Journal of Banking & Finance*, November 2016.

Caro, Spencer, Talia Gillis, and Scott Nelson, "Modernizing Fair Lending," forthcoming 2025.

Caro, Spencer, Talia Gillis, and Scott Nelson. "Differential Validity in Fair Lending," *Journal of Legal Analysis*, forthcoming 2025.

Chang, Victor, Sharuga Sivakulasingam, Hai Wang, and Siu Tung Wong. "Credit Risk Prediction Using Machine Learning and Deep Learning: A Study on Credit Card Customers." *Risks*, November 2024.

Chen, Tianqi and Guestrin, Carlos. "XGBoost: A Scalable Tree Boosting System." arXiv. June 2016.

Chioda, Laura, Paul Gertler, Sean Higgins, Paolina Medina. "FinTech Lending to Borrowers with No Credit History." NBER Working Paper 33208, June 1, 2025.

Consumer Financial Protection Bureau. "Consumer Credit Trends." Accessed June 30, 2025.

Davis, Randall, et al. "Explainable Machine Learning Models of Consumer Credit Risk." January 2022.

de Lange, Petter, Borger Melsom, Christian Bakke Vennerød, and Sjur Westgaard. "Explainable AI for Credit Assessment of Banks." *Journal of Risk and Financial Management*, November 2024.

Dessain, Jean, Nora Bentaleb, and Fabien Vinas, "Cost of Explainability in AI: An Example with Credit Scoring Models." Explainable Artificial Intelligence World Conference Paper, October 2023.

Equifax. "Achieve More Inclusive Credit: Expanding Your Credit Reach to Marginalized Populations Can Help Them and You." 2022.

Executive Office of the President. "Restoring Equality of Opportunity and Meritocracy." Federal Register, 90, no. 80 (April 28, 2025): 17537.

Experian. "New Experian Tool Empowers Financial Inclusion Through Open Banking Insights." May 2024.

Experian. "Launch of Experian's Cashflow Score Signals New Era of Open Banking-Powered Lending." March 2025.

Federal Deposit Insurance Corporation. "2023 FDIC National Survey of Unbanked and Underbanked Households." November 2024.

Federal Deposit Insurance Corporation. "Adoption of Supervisory Guidance on Model Risk Management." Financial Institution Letter 22-2017, June 2017.

Federal Reserve Bank of New York. "Quarterly Report on Household Debt and Credit 2025: Q1." May 2025.

FinRegLab. "Explainability & Fairness in Machine Learning for Credit Underwriting: Policy Analysis." July 2023.

FinRegLab. "The Use of Cash-Flow Data in Underwriting Credit: Empirical Research Findings." July 2019.

FinRegLab. "The Use of Cash-Flow Data in Underwriting Credit: Market Context & Policy Analysis." February 2020.

FinRegLab. "The Use of Machine Learning for Credit Underwriting: Market & Data Science Context." September 2021.

FinRegLab. "Transforming Small Business Credit: Technology and Data Adoption in Mission-Based Lenders." Updated June 2025.

FinRegLab, Laura Blatter, and Jann Spiess. "Machine Learning Explainability & Fairness: Insights from Consumer Lending." July 2023.

FinRegLab, Sabrina T. Howell, and Siena Matsumoto. "Sharpening the Focus: Using Cash-Flow Data to Underwrite Financially Constrained Businesses." June 2025.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. "Predictably Unequal? The Effects of Machine Learning on Credit Markets." *Journal of Finance*, November 2017.

Gote, Amol, and Vikas Mendhe. "Building a Cash Flow Underwriting System: Insights from Implementation." *International Journal of Computer Trends and Technology*, February 2024.

Hepinstall, Mike, Chaitra Chandrasekhar, Peter Carroll, Nick Dykstra, and Yigit Ulucay. "Financial Inclusion and Access to Credit." OliverWyman, 2022.

Johnson, Alex. "Everyone Wants to Be the FICO of Cash Flow Data," *Fintech Takes*, April 2025.

Johnson, Alex. "Everything You Ever Wanted to Know About Cash Flow Underwriting But Were Afraid to Ask." *Fintech Takes*, May 2022.

Kambara, Michelle, and Cooper Luce. "Technical Correction and Update to the CFPB's Credit Invisibles Estimate." Consumer Financial Protection Bureau, June 2025.

Khandani, Amir Ehsan, Adlar J. Kim, and Andrew W. Lo. "Consumer Credit-Risk Models via Machine-Learning Algorithms." *Journal of Banking & Finance*, November 2010.

Lee, Jung Youn and Yang, Joonhyuk and Anderson, Eric. "Who Benefits from Alternative Data for Credit Scoring? Evidence from Peru." SSRN, May 20, 2025.

Liao, Jingxian, Wei Wang, Jason Xue, and Anthony Lei. "Data Augmentation Methods for Reject Inference in Credit Risk Models." Association for the Advancement of Artificial Intelligence, 2021.

Mavrogiorgos, Konstantinos, Athanasios Kiourtis, Argyro Mavrogiorgou, Andreas Menychtas, and Dimosthenis Kyriazis. "Bias in Machine Learning: A Literature Review." *Applied Sciences*, October 2024.

Meursault, Vitaly, Daniel Moulton, Larry Santucci, and Nathan Schor. "One Threshold Doesn't Fit All: Tailoring Machine Learning Predictions of Consumer Default for Lower-Income Areas." Federal Reserve Bank of Philadelphia, November 2022.

Nallakaruppan, M.K., et al. "Credit Risk Assessment and Financial Decision Support Using Explainable Artificial Intelligence." October 2024

Office of the Comptroller of the Currency. "Sound Practices for Model Risk Management: Supervisory Guidance on Model Risk Management." Bulletin 2011-12, April 2011.

Plaid. "Cash Flow-Based Consumer Report by Plaid Check CRA | Plaid." Accessed June 13, 2025.

Shi, Si, Rita Tse, Wuman Luo, Stefano D'Addona and Giovanni Pau. "Machine Learning Driven Credit Risk: A Systemic Review." *Neural Computing and Applications*, July 2022.

Siddique, Sunzida, Mohd Ariful Haque, Roy George, Kishor Datta Gupta, Debashis Gupta, and Md Jobair Hossain Faruk. "Survey on Machine Learning Biases and Mitigation Techniques." *Digital*, December 2023.

Taylor, Niki, and Rohan Sriram. "Introducing Transactions for Business: Powering Real-Time Business Tools." Plaid, May 2025.

TransUnion. "How Often Do Credit Scores and Reports Update?" March 2023.

The World Bank Group. "The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges." World Bank Group, 2024.

van Braak, Bjorn, Joerg R. Osterrieder, and Macros R. Machado. "How Can Consumers Without Credit History Benefit from the Use of Information Processing and Machine Learning Tools by Financial Institutions?" *Information and Processing Management*, March 2025.

VantageScore. "VantageScore 4Plus." Accessed June 13, 2025.

Wang, J. Christina and Charles B. Perkins. "How Magic a Bullet Is Machine Learning for Credit Analysis? An Exploration with FinTech Lending Data." Federal Reserve Bank of Boston, October 2024.

Watanabe, Shuhei. "Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance." arXiv. May 2023.

Wu, Zongxiao, Yizhe Dong, Yaoyiran Li, and Yaorong Liu. "A 'Divide and Conquer' Reject Inference Approach Leveraging Graph-Based Semi-Supervised Learning." *Annals of Operations Research*, May 2025.

Yang, Wenli, Yuchen Wei, Hanyu Wei, and Yanyu Chen. "Survey on Explainable AI: From Approaches, Limitations and Applications Aspects." *Human Centric Intelligent Systems*, August 2023.

Ying Lei Toh. "Addressing Traditional Credit Scores as a Barrier to Accessing Affordable Credit." Federal Reserve Bank of Kansas City, June 2023.

With support from:

JPMORGAN CHASE & CO.

JPMorgan Chase & Co. (NYSE: JPM) is a leading financial services firm based in the United States of America ("U.S."), with operations worldwide. JPMorgan Chase had $4.4 trillion in assets and $351 billion in stockholders' equity as of March 31, 2025. The Firm is a leader in investment banking, financial services for consumers and small businesses, commercial banking, financial transaction processing and asset management. Under the J.P. Morgan and Chase brands, the Firm serves millions of customers in the U.S., and many of the world's most prominent corporate, institutional and government clients globally. Information about JPMorgan Chase & Co. is available at www.jpmorganchase.com.

Capital One

Capital One Financial Corporation (www.capitalone.com) is a financial holding company which, along with its subsidiaries, had $367.5 billion in deposits and $493.6 billion in total assets as of March 31, 2025. Headquartered in McLean, Virginia, Capital One offers a broad spectrum of financial products and services to consumers, small businesses and commercial clients through a variety of channels. Capital One, N.A. has branches located primarily in New York, Louisiana, Texas, Maryland, Virginia, New Jersey and the District of Columbia. A Fortune 500 company, Capital One trades on the New York Stock Exchange under the symbol "COF" and is included in the S&P 100 index.

Copyright 2025 © FinRegLab, Inc.

All Rights Reserved. No part of this report may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Digital version available at finreglab.org

Published by FinRegLab, Inc.

1701 K Street NW, Suite 1150
Washington, DC 20006
United States