Sharpening the Focus:
Using Cash-Flow Data to Underwrite Financially Constrained Businesses

Research Shows Underwriting Improvements from Adding Cash-Flow Data to Credit Scores, Particularly for Early-Stage Businesses and Financially Constrained Entrepreneurs

FinRegLab and researchers from the New York University Stern School of Business have analyzed fintech lenders' data to evaluate impacts on predictive accuracy and credit access from incorporating electronic cash-flow data into small business underwriting models, finding particular benefits for early-stage businesses and financially constrained entrepreneurs who often struggle to access loans because they are perceived to be higher risk.

The study examines performance data from about 38,000 small business loans originated by two fintech lenders between February 2015 and January 2024, excluding the Paycheck Protection Program and other specialty initiatives. It compares the effect of using roughly two dozen cash-flow metrics derived from bank account data in combination with the business owner's personal credit score to relying on models that only incorporate credit scores, industry, age of firm, and other basic metrics.

The models that incorporate cash-flow information added substantial predictive signal compared to the baseline models that relied primarily on credit scores. For example, a commonly used metric for evaluating the performance of predictive models called the Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) increased by almost 2% (from .652 to .663) for a machine learning model that incorporated the cash-flow data over the baseline model that did not.[1] As expected, performance gains were somewhat larger among borrowers with lower credit scores (under 700) than for borrowers with higher scores. But even accounting for that baseline differential, performance improvements were particularly large for low-score owners whose businesses were young (less than five years old), suggesting that adopting cash-flow data could be particularly important in increasing lenders' confidence when underwriting business owners that face both financial constraints on both their personal credit and the length of their business history.

Data and Technology Adoption Among Mission and Community Based Lenders

Community development financial institutions, small banks, and credit unions are increasingly looking to electronic cash-flow data to help increase the speed and scale of their small business lending programs, but often face resource and technology constraints in changing their systems. FinRegLab has released an updated qualitative study of lessons learned based on five implementation pilots and other data and technology initiatives, as well as related considerations for smaller lenders who are working to expand the scale of their small business programs.

[1] The ROC-AUC is one method of comparing a model's correct predictions against its errors that 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. Its scale varies from .5 to 1.0, with 1.0 representing model results that perfectly separate between defaulters and non-defaulters.

Analyzing results for the borrowers whose risk predictions changed most (either positively or negatively) with the addition of cash-flow data suggests that cash-flow models are likely to increase access to credit among financially constrained business owners. For example, the cash-flow model reclassified 3.6 times as many low-score borrowers as substantially lower risk than substantially higher risk. Impacts were even stronger for low-score borrowers with early stage firms, where the ratio of positive reclassifications was 7.6.

A supplemental analysis focused on entrepreneurs with low credit scores whose businesses are located in zip codes that have relatively low median incomes and higher percentages of Black and Hispanic residents. There were stronger impacts for some categories, but many of the results appeared to be driven primarily by the overall benefit to business owners with low credit scores rather than location-specific effects. Nevertheless, to the degree that certain geographies have higher numbers of residents with lower scores, incorporating cash-flow data in underwriting models could be important for increasing overall credit access in economically disadvantaged areas.

The cash-flow metrics used in the study included log transformations of deposits, withdrawals, and account balances; standard deviation calculations to measure the amount of volatility in deposits and balances; various ratios between core metrics; and features relating to low or negative balance events, insufficient funds incidents, and loans that require daily payments (such as merchant cash advances). In the machine learning model, the amount and volatility of deposits and balances made the greatest contributions to predictiveness among the cash-flow variables, followed by withdrawals and the number of low or negative balance events. Insufficient funds incidents and the presence of daily pay loans made a smaller contribution but still contributed to predictive performance.