FICO Decisions

Can Alternative Data Expand Credit Access?

New FICO research shows how to score millions more creditworthy consumers

Using alternative data, millions more consumers qualify for credit — and go on to improve their credit standing

Widespread adoption of credit scoring by financial institutions over the past 25 years has made credit available and affordable to a majority of US consumers. Is there an opportunity to go further, opening onramps to credit for a much broader population? Can scoring help lenders safely and responsibly extend credit to consumers who traditionally don't receive credit scores because of insufficient or nonexistent credit bureau information?

New FICO research says yes — provided data used for scoring is not limited to traditional credit bureau files. In fact, we found that with the addition of data sources currently residing outside credit bureau files, we can generate reliable, predictive risk scores for more than half of previously unscorable credit applicants. It's an approach that reveals significant differences in risk among these consumers — enabling lenders to recognize creditworthy individuals who would otherwise be difficult to identify.

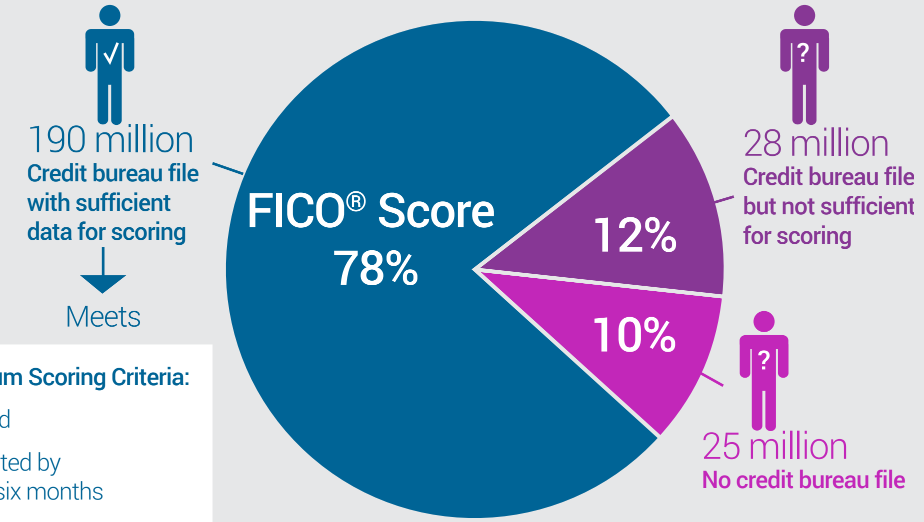
This white paper shares FICO research demonstrating that with the right alternative data approach, millions more consumers will score high enough to qualify for credit — with most who obtain credit going on to improve their credit status. We present key research findings showing:

- Why scoring more people without more data harms consumers and creditors
- How alternative data scoring releases millions stuck in credit "catch-22"
- How the newly scorable differ from other consumers and each other

FICO recently conducted research to determine how to generate reliable, predictive risk scores for the more than 50 million US adults who don't currently have FICO® Scores.

Roughly 190 million US consumers have credit bureau files that meet the minimum criteria for calculating a FICO® Score (Figure 1). But 28 million consumers have files with insufficient data to meet these criteria. And more than 25 million consumers have no bureau file at all.

Figure 1: Extending scoring to more US consumers



FICO® Score Minimum Scoring Criteria:

- Cannot be deceased
- One trade line reported by creditor within last six months
- One trade line at least six months old

These two unscorable populations include many creditworthy individuals — people many financial institutions would welcome as customers. Given scant or nonexistent bureau data on these people, can scoring be predictive and reliable enough to separate out the good risks so lenders can confidently extend credit to them?

The answer is yes — but only when bureau data is supplemented with alternative data that fills in these consumers' financial picture. Below are our key research findings supporting this conclusion.

Research Insight #1:

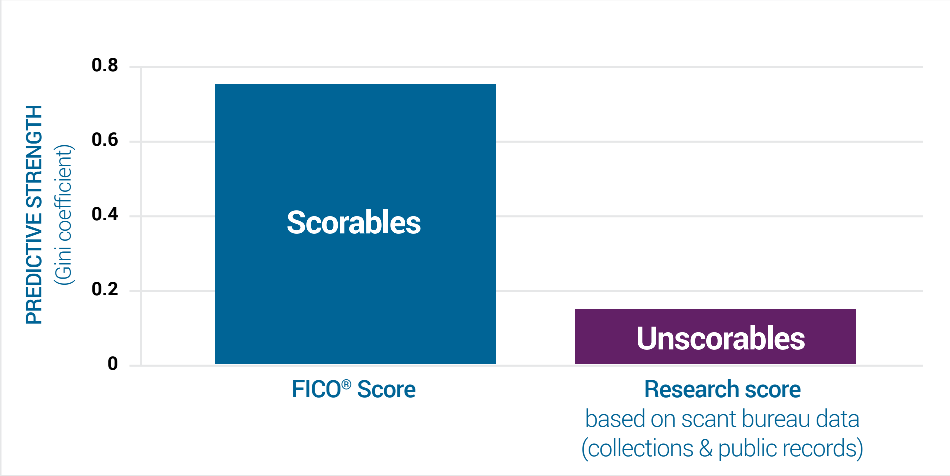
Bureau data alone is insufficient for scoring more consumers

The first step in our research was to find out if it's possible to calculate a meaningful, reliable score for more consumers using credit bureau data alone. Our focus was, therefore, the 28 million traditionally unscorable consumers with bureau files — in other words, those with sparse or old bureau data.

Research results consistently show that predictive scoring models relying solely on sparse or old credit data are weak and do a poor job forecasting future performance. For instance, we developed a research score for the approximately 7 million consumers (about 25% of the unscorable population with credit files) who have one or more collections or adverse public records but no other credit account information. We then calculated several standard predictive measures to evaluate the performance of this score on these consumers. For these scant-file consumers, the Gini index[1] of the score was 0.147, significantly less than the 0.600 to 0.800 Gini indices for scorable consumers (Figure 2). A lower Gini index means the score is less predictive of future behavior and less able to separate good credit risk from bad credit risk.

Figure 2: Sparse data results in weak predictiveness

Risk predictiveness analyzing bureau data only



Next, we looked at scoring consumers with older bureau data. Using a research model with a recent national credit bureau sample, we scored consumers with no credit account updated in the last six months. We compared the odds-to-score alignment of this group against a baseline of traditionally scorable consumers — those with at least one credit account updated in the last six months.

The results showed that the older the data, the less reliable the implied odds of the score.[2] Thus a risk level associated with a particular score, such as 700, will not be the same across successively more stale segments of the population.

To understand the implications, think about auto loans. Lenders setting an underwriting strategy among borrowers at a given score cutoff could be accepting consumers with markedly different repayment risk, depending on how long a lapse occurred since the bureau file was updated. For example, our research showed that a 640 score based on files that have not been updated in 21 months or more exhibits repayment risk roughly in line with a 590 score for the traditionally scorable population — an odds misalignment of about 50 points. This lack of reliability in odds-to-score relationship can undermine a lender's ability to precisely manage risk and lead to consumers being mispriced on loans relative to their true level of risk.

[1] A Gini index or coefficient is a statistic used to measure the effectiveness of a predictive model. Gini indices range from 0 to 1; the higher the number, the stronger the model. A Gini index of 1 represents perfect risk discrimination. A Gini index of 0 is equivalent to a random decision or completely imperfect risk discrimination.

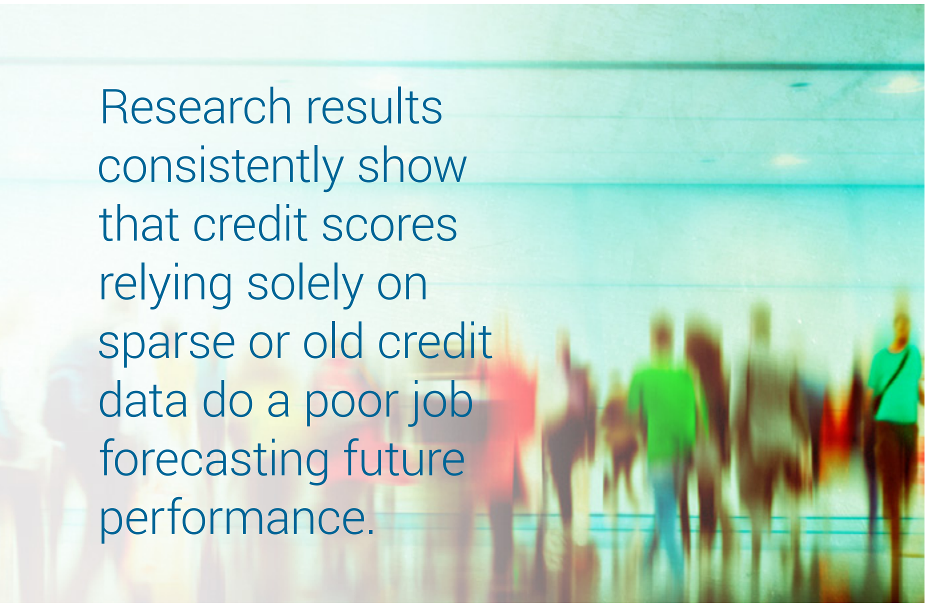
[2] Credit scores are designed to rank-order risk — that is, sort accounts so that higher scores indicate less risk (lower odds of serious delinquency). For example, a credit score of 720 indicates less risk than one of 680. The different levels of risk associated with scores is called the "odds-to-score relationship."

Ultimately, risk discrimination is weak when scoring on sparse or old bureau data. Such data is not sufficient to accurately identify the good risks creditors will accept — and, therefore, not helpful for expanding access to credit.

For lenders, use of a weak score could mean declining applicants they should have accepted, and vice versa — producing higher levels of delinquency and lower lending volume than necessary. For consumers, it could mean receiving lower credit lines/loans than requested and needed or higher than they can handle.

Figure 3: Scoring without good risk discrimination does not benefit consumers

Moreover, for the majority of the 28 million consumers with scant or stale bureau data, scoring would not make it easier for them to establish credit. About 65% of these consumers have a negative item and no active account. With no positive data flowing into their files to offset the negative, they would likely score too low to obtain credit.



Take a consumer who has recovered from a negative financial event occurring three years ago: Without current information flowing into the credit file, no amount of analytic segmentation or other innovation can generate a score reflecting that consumer's current risk profile. To accurately score this consumer, the credit file must contain up-to-date information on the consumer's current behaviors and risk markers.

Thus, scoring based on credit bureau data alone won't help consumers with inactive credit and a need to rebuild their credit standing. These consumers are stuck in a catch-22: To obtain credit, they have to be using credit — but without a reliable way to assess current creditworthiness, lenders may not take a chance on them.

Similarly, scoring from bureau-only data won't help the 25 million with no credit files. They're stuck in the same catch-22.

Bottom line: Bureau data must be supplemented to score more consumers in a manner that reliably reflects their true level of risk.

Research Insight #2:

Unscorable credit applicants aren't all alike

Next, we examined the credit behaviors of consumers within the unscorable population. Since our goal was to help expand credit access, we focused on those within this population who actually apply for credit. These are the consumers for whom extending scoring capabilities can make a difference, and we wanted to better understand how to more accurately assess their credit risk.

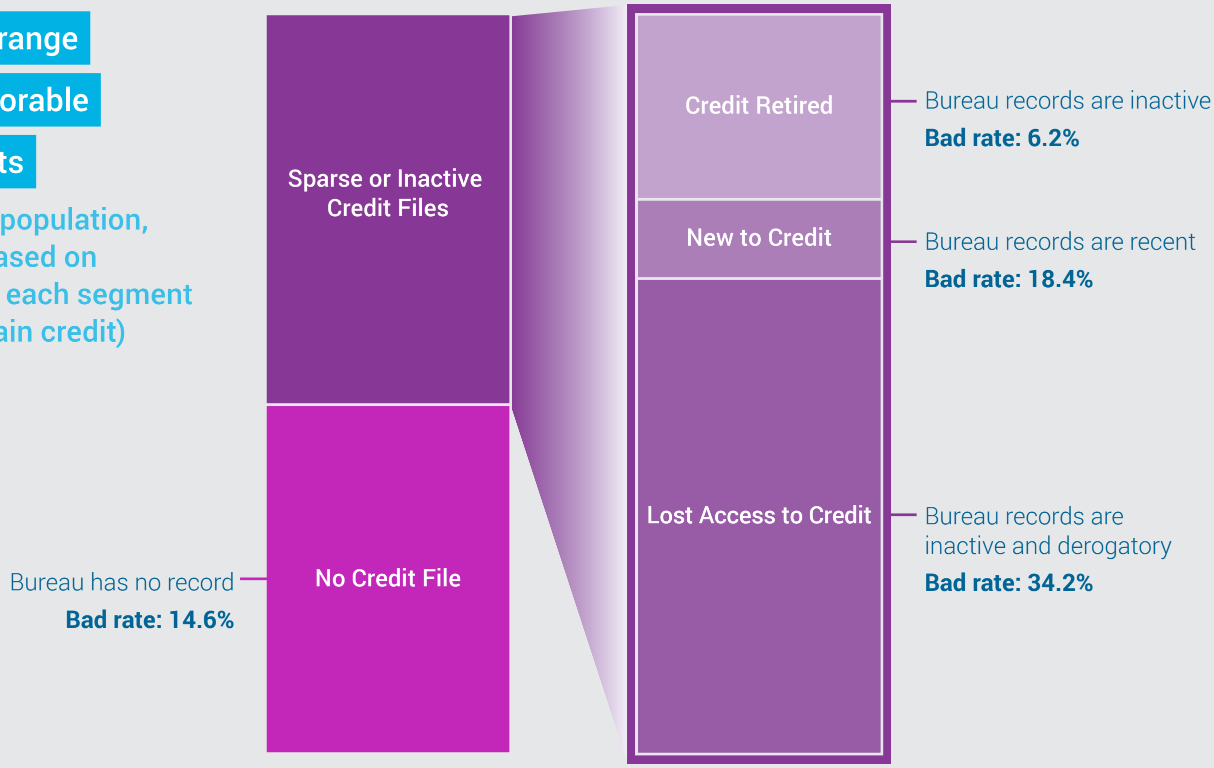
We found that these consumers differ from the mainstream credit population — and from each other. As a whole, unscorable applicants are more risky. Their overall default rate is almost three times higher than for scorable consumers.

Yet risk levels vary considerably within this population. Figure 4 shows unscorable applicants separated into risk bands, using a very simple segmentation system based on the quantity and quality of information in bureau files. Bad rates (based on those within each segment who go on to obtain credit) range from 6.2% to 34.2%.[3]

That's still very coarse separation. To differentiate the risk in greater detail, we need additional data.

Figure 4: Wide range of risk for unscorable credit applicants

Segmentation of population, with bad rates (based on applicants within each segment who go on to obtain credit)



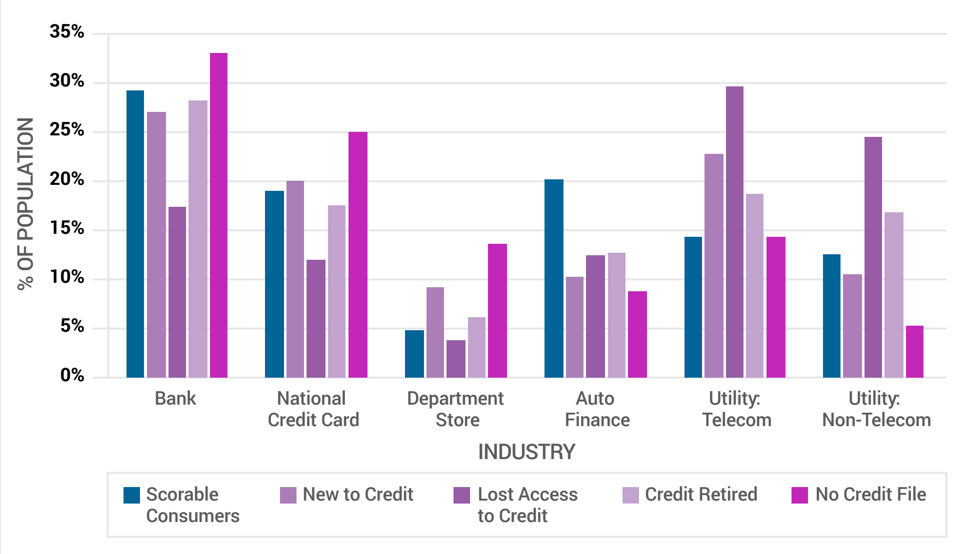
[3] Accounts were classified as of May 2013 based on prior 24 months of repayment behavior as good (no missed payments), bad (90+ days past due or worse) or indeterminate (all else). Only accounts opened prior to the observation date (May 2011) or opened within the following six months were classified.

Besides risk level, we discovered that other credit behaviors also vary for unscorable applicants. For instance, the type of credit sought varies significantly by segment (Figure 5). Consumers in the Lost Access to Credit segment are most likely to seek out telecommunications-related credit. Those in Credit Retired and No Credit File segments tend toward credit cards or bank products.

These subtle but important distinctions in credit behavior aren't visible when the unscorable population is lumped together and scored using a single scorecard. For more insights into these consumers, we need a segmented system of scorecards coupled with additional data.

Figure 5: Differences in types of credit sought

Distribution of credit bureau inquiries by industry type (as % of top six inquiry industry types)



Research Insight #3:

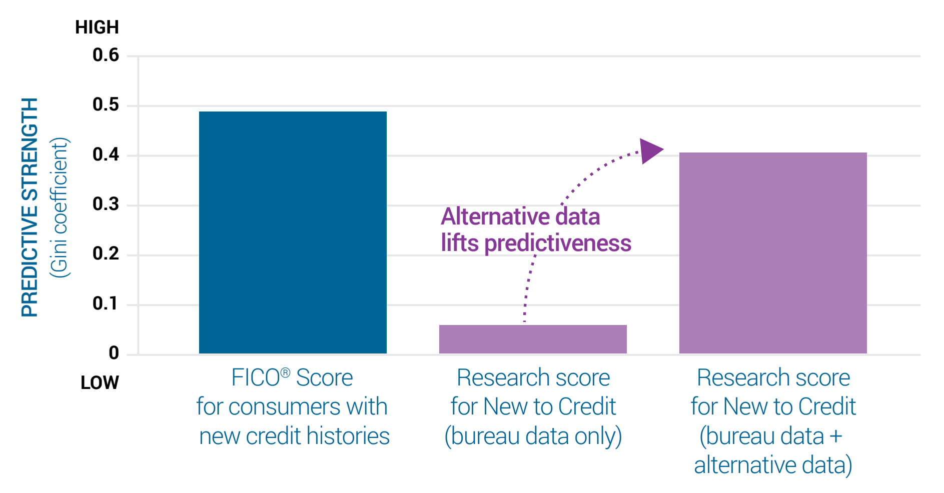
Alternative data is essential for accurate risk separation

To achieve more risk differentiation for traditionally unscorable credit applicants, we need to fill in the partial or missing picture of current financial behavior available from credit bureau files. This stage of our research sought to answer: By complementing bureau data with alternative data, could we generate scores that are strongly predictive of risk within segments?

To find out, we built a research model and scored New to Credit consumers using bureau data only (which generally consisted of only one or more credit inquiries). We then compared the model's performance when bureau data was complemented with alternative data. As a reference, we also included a closely comparable group of consumers with traditional FICO® Score — those with new credit histories (five years or less).

Results, shown in Figure 6, demonstrate that alternative data indeed offered a performance lift. While the Gini index for the research score based on bureau data alone was very low, the Gini index for the score based on both bureau and alternative data increased substantially, bringing the model's predictive strength near the range of a traditional FICO® Score for consumers with new credit histories.

Figure 6: Increasing predictive power with alternative data



Our conclusion: A model combining alternative data with bureau data sufficiently differentiates risk within traditionally unscorable segments of consumers, enabling responsible credit decisions.

Research Insight #4:

Not all alternative data drives a more predictive score

A key reason there is an opportunity to score more consumers today is the growing number of alternative data providers that have entered the market in recent years. But not all provide equal value for scoring.

Consider telecommunications payment data. It has many similar qualities as data reported in traditional credit files. In fact, telecom companies occasionally report customer account status to credit bureaus. Yet this information is present in less than 10% of bureau files — and it tends to be negative.

More complete telecom data is available from alternative sources — *and it includes positive as well as negative information*. That's important for expanding credit access since it may provide current evidence of good financial behavior where that's missing from bureau files. For consumers with no credit history and others emerging from financial problems, opening a telecom account can be a first step in establishing or re-establishing creditworthiness.

Based on our research and experience, alternative data sources must demonstrate that they make the grade across a number of important dimensions. All of the alternative data sources we used in the research from this paper passed these hurdles:

FICO Six-Point Test	
Regulatory compliance	Any data source must comply with all regulations governing consumer credit evaluation. To comply with the Fair Credit Reporting Act (FCRA), for example, a data provider must have a process in place for investigating and resolving consumer disputes in a timely manner. In addition, for the data to be useful in high-volume scoring, the vendor must have an infrastructure that supports compliance at a significant scale. In evaluating potentially useful data, it's also critical to think ahead about how creditors will communicate with consumers, for example, about adverse action decisions resulting from the use of the data. Will creditor decisions be palatable and defensible? Can the role the data plays in decisions be clearly explained to consumers and regulators?
Depth of information	The deeper and broader the data, the greater its value. Consider a repository of rental data: Does the data reflect both on-time and late payments? Is the account history captured from the beginning of a consumer's rental history or just for a recent period? If the consumer has moved, are there records from multiple addresses?
Scope and consistency of coverage	Since the objective is to score as many consumers as possible, useful databases must cover a broad percentage of the population. For instance, with over 90% of US adults using cell phones,[4] mobile companies are a potential data source with broad coverage. The data must also be consistent in nature — not undergoing significant change that would undercut its value for comparative analysis.
Accuracy	Inaccurate data compromises the predictiveness and, therefore, the value of the data. Data repositories must have a mature data management process in place to ensure data accuracy. It's important to ask questions like: How reliable is the data? How is it reported? Is it self-reported? Can the data be easily manipulated by applicants or others? Are there verification processes in place?
Predictiveness	The data should predict future consumer repayment behavior. For example, analysis of public record databases shows that in many cases, consumers who have been at their address for a longer period of time are more likely to pay their credit obligations than those more transient. Such a data source would add value for credit risk evaluation.
Additive value — aka "orthogonality"	Useful data sources should be supplemental or complementary to what's in credit bureau reports. For example, if a repository collects only foreclosure data from public record information, that data may add little value since it is already largely captured in bureau reports.

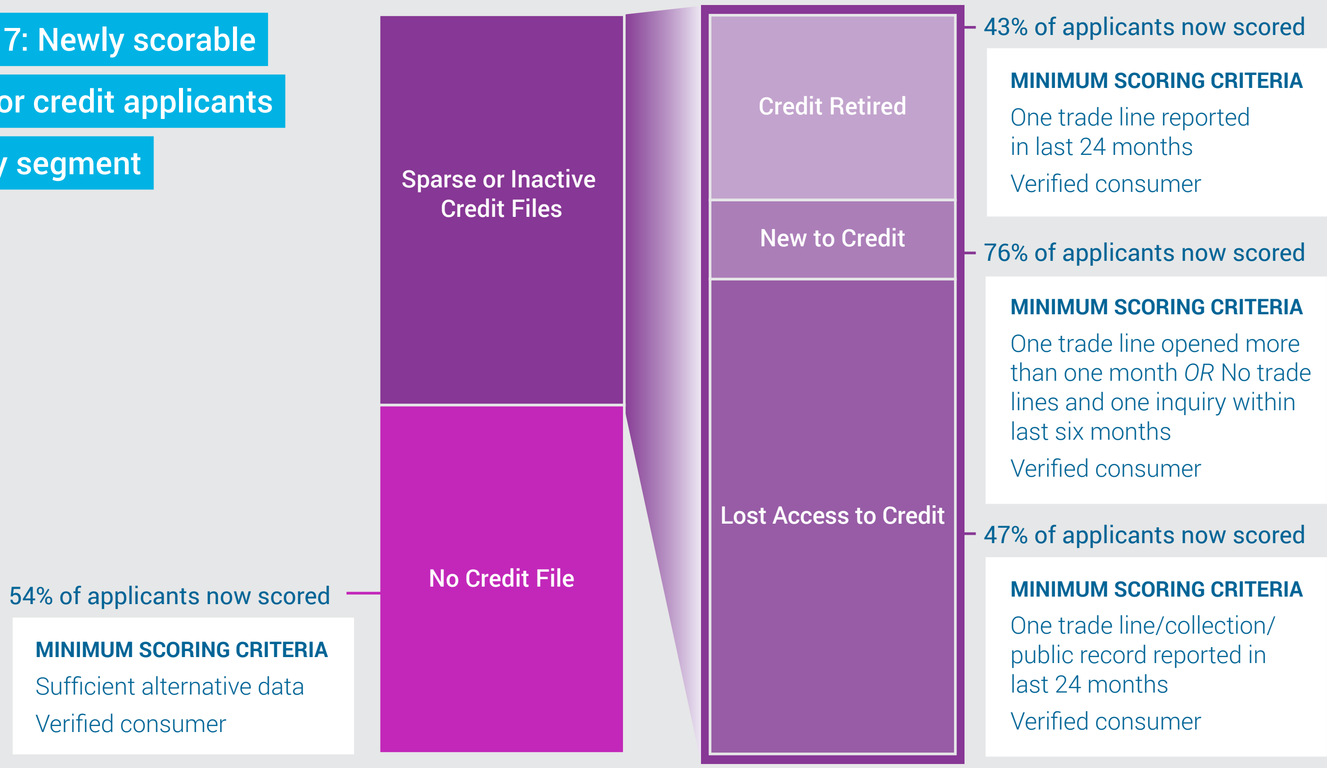
Research Insight #5:

> 50% of previously unscorable applicants can be accurately scored

FICO research showed that with the right alternative data, we can accurately score large numbers of previously unscorable credit applicants. In fact, more than 50% of these applicants can now be scored.

Scorable rates vary by segment, however. As shown in Figure 7, these rates reflect the percentage of applicants who, while unable to meet FICO® Score minimum criteria, are able to meet — with the addition of alternative data — new FICO minimum scoring criteria for their respective segment.

Figure 7: Newly scorable rates for credit applicants vary by segment



How did we arrive at this segment-based minimum scoring criteria? By using several well-established analytic techniques in an innovative way:

- **Bigger data — maximizing what we know about how unscorable applicants perform when granted credit.** We did our research on over 14 million consumers, which included one of the largest data samples of the traditionally unscorable ever analyzed. This oversampling was necessary because model development requires a sample of consumers who are representative of the population and who have observable credit behavior (what we call "classifiable performance") over a subsequent period.

  Obtaining an adequate sample on traditionally unscorable consumers is difficult because only a relatively small percentage — about 10% — are granted credit and open accounts resulting in observable behavior. To observe credit behavior on at least a million unscorable consumers, for example, we would need to sample from a random population of at least 10 million such unscorables. For this research, we utilized stratified sampling on the full 28 million unscorable population to arrive at an analysis dataset consisting of 7.5 million unscorable records. By analyzing such a large and stratified sample, we were able to capture classifiable performance on more consumers.

- **Reject inference — reducing bias in the development population by considering those who did not receive credit.** We can't observe payment performance on those unscorable applicants who did not receive credit. We have to infer this behavior analytically based on the data we do have.

  But performance data on those consumers that were able to open accounts may be biased because these applicants were likely "cherry picked." In lieu of a score, lenders may have granted credit to some individuals based on a special aspect of the borrower, such as verified income or assets. They may have offered only credit products with strict risk controls, such as secured cards.

  As a result, consumers granted credit are unlikely to be representative of their population segment as a whole. Ignoring this effect can lead to models that grossly understate the true credit risk of applicants. Applying the tried-and-true analytic technique of reject inference allows us to mitigate this bias and build reliable models.

- **Propensity modeling — determining how far to go in applying the model to the unscorable population.** After building a segmented research model based on our sample population, we used propensity modeling to ensure that the profile of consumers scored by the model is similar to the profile of the consumers on which the model was built. These similarities were key inputs in establishing segment-based minimum scoring criteria for our alternative data sources.

The resulting alternative data score provides a consistent measure of risk across all consumer populations. We see this in Figure 8, where the dotted line showing the odds-to-score relationship for the FICO® Score 9 population is indistinguishable from the solid lines representing populations scored with the addition of alternative data. Precise alignment shown here means a score of 700, for instance, represents the same risk odds, regardless of whether the consumer is in a traditionally scorable or newly scorable population.

Figure 8: Consistent risk measure for traditionally and newly scorable populations

Odds-to-score chart: Account originations + account management performance

*Bads were defined as 90+ days past due

Research Insight #6:

Millions more consumers score high enough to qualify for credit

Our approach of analyzing alternative data with bureau data goes beyond making more consumers scorable. It reveals many creditworthy individuals who would otherwise still be stuck in the credit catch-22.

As shown in Figure 9, more than a third of the newly scorable — millions of consumers — achieve high enough scores to gain access to credit. With the addition of alternative data, creditworthy individuals stand out. Lenders can effectively identify good risks and safely expand access to credit.

And unlike a credit-bureau-only score built for this population, these scores are dynamic and will improve with continued on-time payments of everyday bills.

Figure 9: More than a third of newly scorables are at 620 or above

Interval score distribution: Newly scorable

Research Insight #7:

Many of the newly scorable rapidly enter the credit mainstream

FICO research also shows that with this approach, the majority of previously unscorable applicants granted credit go on to manage their credit obligations responsibly.

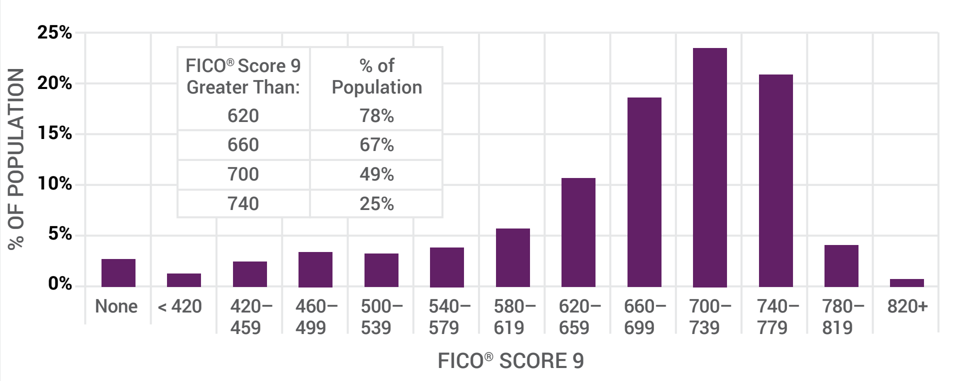
As shown in Figure 10, a majority of applicants with an alternative data score of 620 or higher at account origination have a FICO® Score 9 of 620 or higher 24 months later. Two thirds achieve a FICO® Score of at least 660, and nearly half rise above 700.

This data supports the premise that an alternative data score can be an effective tool in providing unbanked consumers a safe onramp to mainstream credit. Moreover, as Figure 10 demonstrates, consumers identified as good credit risks with this score are likely to maintain and improve their credit standing over time.

Figure 10: Within two years, many have FICO® Scores of 660 or higher

FICO® Score distribution two years after obtaining credit

For consumers with alternative data score greater than 620 at time of credit application



Conclusion

The financial services industry and the majority of US consumers have benefited immensely from the broad access to credit made possible by credit scoring. Now we have an opportunity to extend scoring to millions more consumers, thereby helping them establish or re-establish their creditworthiness.

The goal, however, is not to just generate more scores — but to generate scores that enable lenders to safely and responsibly extend credit to more people. Today, credit bureau data alone isn't enough to do that. Alternative data is essential for scoring to accurately reflect the financial behavior and risk of previously unscorable consumers seeking to join the credit mainstream.

The FICO research in this paper is providing the foundation for our work on a **score based on alternative credit data**. Lenders interested in learning more can contact us at **ficoscoreinfo@fico.com**. To keep tabs on the latest FICO research on scoring best practices and credit risk trends, visit the **FICO Blog**.

The Insights white paper series provides briefings on research findings and technology innovations from FICO.

FICO

FOR MORE INFORMATION
www.fico.com
www.fico.com/blogs

NORTH AMERICA
+1 888 342 6336
info@fico.com

LATIN AMERICA & CARIBBEAN
+55 11 5189 8267
LAC_info@fico.com

EUROPE, MIDDLE EAST & AFRICA
+44 (0) 207 940 8718
emeainfo@fico.com

ASIA PACIFIC
+65 6422 7700
infoasia@fico.com

4151WP_EN 10/15 PDF