

POINT OF VIEW

ALTERNATIVE DATA AND THE UNBANKED

AUTHORS

Peter Carroll, Partner

Saba Rehmani, Engagement Manager

ALTERNATIVE DATA AND THE UNBANKED

‘Alternative data’ can improve access to credit for millions of Americans. It can do this by overcoming two important limitations of today’s best practices in lending, which rely heavily on credit scores from the three major credit reporting agencies. The two principal limitations of current best practices are that:

- *Many consumers remain ‘credit invisible’, meaning that no credit scores are available for them*
- *The accuracy of scores today, while very good, is still sufficiently limited that many potentially good borrowers must necessarily be denied access to credit because they cannot be statistically separated from poorer risks*

Alternative data show significant potential to improve the status quo by enhancing the accuracy of existing scores (achieving better risk separation), and by rendering visible many of today’s credit invisibles. Progress will come about through the private sector efforts of established lenders and credit bureaus, as well as the innovations of FinTechs, alternative data vendors and big data analytics firms, operating in the free market. There may also be a limited role for regulatory and/or legislative initiatives. The end result will be better and fairer access to credit for individuals, with macro benefits for the whole economy.

Consumers in the United States are heavy users of credit. Consumer debt, including personal loans, real-estate secured loans, auto loans, credit cards, and student loans, totals over \$12 trillion. Even excluding mortgages, this amounts to over \$30,000 of debt per household¹. Access to credit is an important and widespread benefit because it allows consumers to purchase productive assets such as a car, pick-up truck or work tools, to accelerate consumption from future earnings, invest in education for future earning power, enjoy tax advantages (through a home mortgage), or to build wealth, among other opportunities.

On the other hand, because consumer lending relies heavily on the use of credit scores, millions of Americans currently do not have beneficial access to credit either because their credit score is too low or because they have no credit score at all.

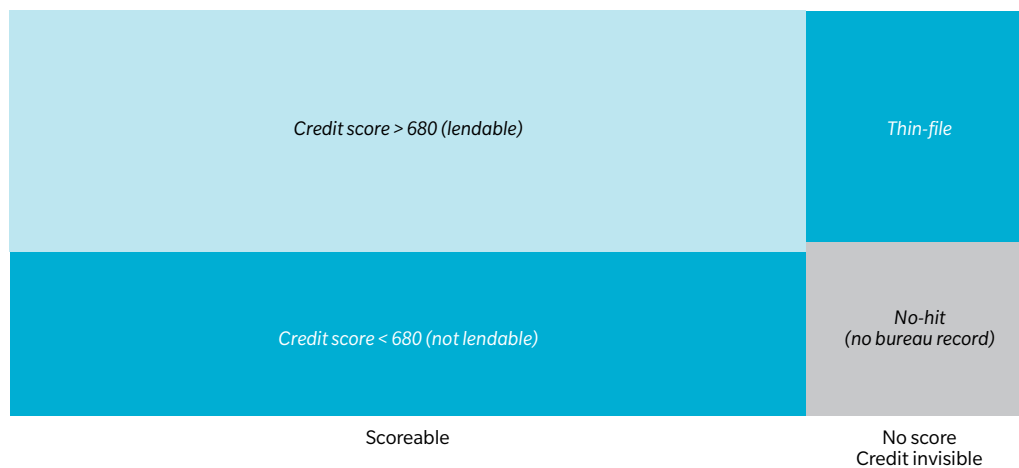
The people who have no credit score are sometimes referred to as ‘credit invisibles’; there are two ways in which they may come to have no score:

- They have no credit file at any of the three major credit bureaus—such people are referred to in the trade as ‘no-hits’
- Despite having a credit file there is insufficient recent information in that file to produce a score—these people are referred to as ‘thin-files’

¹ Source: Federal Reserve, Q3 2016

Exhibit 1 is a schematic that shows all US adults classified as either 'scoreable', thin-file or no-hit. The scoreable population is further separated into those with higher scores ('lendable') or lower scores ('not lendable'). Different lenders set different boundaries between lendable and not lendable; here the proportions are consistent with a FICO score or VantageScore of 680 as the approximate dividing line.

Exhibit 1: Credit scoreability and access for U.S. adult population



There are several underlying reasons why the no-hits and thin-files may be credit invisible, including:

- They never used credit: for example, students, recent legal (or undocumented) immigrants, or millennials seeking credit for the first time may all have zero recorded credit history
- They no longer use credit: for example, older people may have paid off all debt, or avoided using credit; some people are cash-rich and do not require credit; others may have lost access to credit due to previous economic difficulties
- They have made very limited use of credit

The credit invisible population is varied and contains a high proportion of minorities. The Financial Industry Regulatory Authority estimates that only 51% of African-Americans and 58% of Latinos have a credit score, compared to the 75% of all American adults who have a credit score². Lacking access to mainstream credit, credit invisibles are vulnerable to high-priced credit such as payday loans, buy-here-pay-here auto loans, lease-to-own, and other 'informal' high-rate lending products. If they can access credit at all, it is expensive for them to borrow short term and almost impossible to obtain a mortgage to buy a home.

² Source: LexisNexis Risk Solutions (Alternative Data and Fair Lending, 2013)

For a subset of the credit invisibles population, as well as many of the low-score/non-lendables, of course, ‘access to credit’ would mean getting into debt and might not necessarily be a good thing, for either the borrower or the lender. While we should encourage *qualified* access to credit we should not regard ‘credit to everyone’ as a correct policy goal. One of the few undisputed errors leading up to the financial crisis was the appeal to “roll the dice a little bit more” with respect to subsidized home-ownership³. The proper goal is surely to make credit available at a fair price to those who are most likely to benefit from the loans *and also repay them*. Interestingly, there is evidence to suggest that quite a high percentage of the credit invisibles could meet these requirements.

A CLOSER LOOK AT LENDING AND CREDIT SCORES

In recent decades, consumer lending has moved towards the practice of making lending decisions guided by statistical models. Lenders evaluate potential borrowers through an underwriting process that relies heavily on credit scores derived from data in applicants’ credit files. These credit files are maintained by the major credit bureaus (Experian, Equifax and TransUnion). In effect, a credit score provides a lender with a guide to the probability of being repaid, in the event they decide to approve a loan application.

Despite their widespread adoption there are some limitations to credit scores and many aspects of their creation and use that are not widely understood. We have included a brief explanatory section later in this article which may be useful to anyone unfamiliar with credit scores, or for those needing a refresher. One point we will make here, however, is that credit scores, though enormously helpful, are still far from perfect.

Let’s take it as given that consumer scores provide a reasonably accurate rank-ordering of relative credit risk among potential borrowers for whom such scores are available. Lenders primarily use scores to determine whether or not they will lend to a given borrower; they typically set a score below which they will generally not lend, often referred to as a ‘cut-off score’. Scores are also used to help set the terms of a loan, like the loan (or line) amount and the interest rate. Applicants with higher credit scores are less likely to default and can therefore secure larger loan amounts and typically are asked to pay lower interest rates.

A credit score, by itself, does not directly predict the probability of default. Scores are calibrated, using real-world data, to show so-called ‘bad odds’ rates. ‘Bad odds’ are basically the observed default rates within groups of borrowers with the same credit score. Individual defaults are, of course, binary—people either default or they do not. Credit scores cannot predict individual defaults, but they can position individuals among pools of people where the pool can be expected to exhibit a measureable default rate.

³ Rep. Barney Frank (D.-Mass.), speaking in 2003 in support of looser underwriting standards for FNMA and FHLMC, which led to a huge expansion of the sub-prime mortgage market

Exhibit 2 shows a fairly typical set of data on bad rates (90+ days past due) for a generic credit score.

Exhibit 2: Illustrative bad rate by credit score interval

SCORE INTERVAL	90+ DPD
811-850	0.1%
791-810	0.2%
771-790	0.4%
751-770	0.7%
731-750	1.1%
711-730	1.7%
691-710	2.6%
671-690	3.7%
651-670	4.9%
631-650	7.0%
611-630	9.2%
591-610	12.2%
571-590	15.5%
551-570	19.6%
531-550	23.7%
511-530	29.0%
491-510	33.4%
471-490	37.4%
451-470	39.9%
300-450	44.8%

As the table shows, people with the best scores have a bad rate of just 0.1%. Those with the worst scores have a bad rate of nearly 50%. And the bad rate rises uniformly as the scores go lower. This means that the score is doing a very good job, at a macro level, of separating low-risk from high-risk borrower pools: a consumer in the worst band is nearly 500 times more likely to default than one in the best score-band. But notice that if a lender uses a cut-off score of 670, the group being narrowly rejected by the cut-off (i.e. those with scores between 651 and 670) has a bad rate of 4.9% which means that of all the applicants in the 651–670 pool being declined by this lender 95% would most likely not default. More accurate scores would allow clearer separation of the 5% ‘bads’ in this score-band from the likely 95% ‘goods’.

RETURNING TO THE CREDIT INVISIBLES

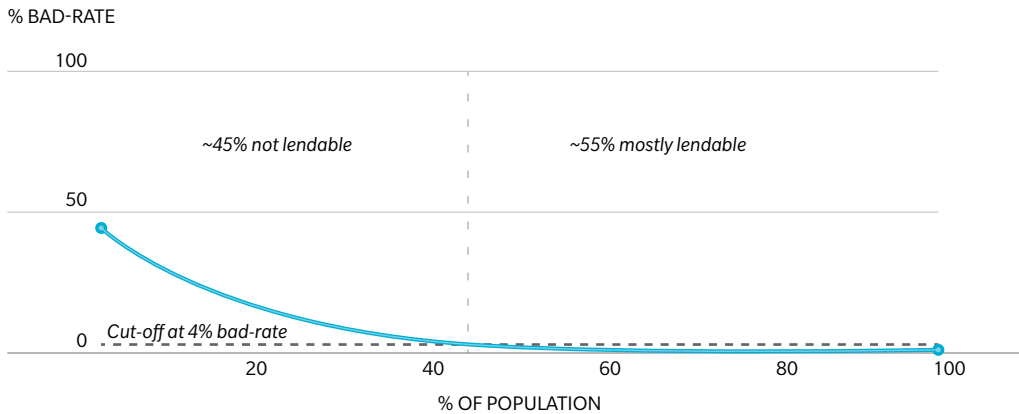
The catch-22 of modern credit is that in order to borrow, you need a score; but to generate a score, you need to have borrowed before. This leaves many Americans without a score and without access to credit. The numbers are not trivial: of the approximately 240 million adults in the U.S., nearly 25%, or 60 million, are credit-invisible⁴:

- **Full-file (180–190 million):** Borrower has a credit file with sufficient recent tradeline data to generate a traditional credit score
- **Thin-file (25–35 million):** Borrower has credit file but with insufficient and/or outdated tradeline data to generate a traditional credit score
- **No-hit (20–25 million):** Credit bureau has no information/file on the person at all

In addition to the problems of no-hits and thin-files, merely having a credit score does not—as we have seen—automatically translate to being lendable. Why is this? A bank decides upon the score it will use as a cut-off by the entirely logical process of asking: “At what score-band am I letting in a sufficiently high percentage of new borrowers who *will* default that the losses I will suffer on those loans wipes out all the profit I stand to make from the ones in that same score-band who will *not default*?” This is an economic judgment that balances the costs and benefits of admitting a marginal pool of people with a rising bad rate. And since the costs of a borrower who defaults are much greater than the profit from a borrower who does not, the lender has to set a cut-off that necessarily excludes many potentially good borrowers. Exhibit 3 shows schematically how the non-lendable population consists, in part, of full-file consumers with low credit scores.

⁴ Source: Experian, FDIC, FICO, LexisNexis Risk Solutions, NCRA, PERC, VantageScore, Oliver Wyman analysis

Exhibit 3: Illustrative depiction of lendable full-file population



The diagram shows how the scoreable population tracks against the bad-odds rate and includes a representative cut-off rate of ~4% (the horizontal dashed line). As the diagram shows clearly, the bad-odds line intersects the cut-off at a point where about ~55% of the full-file population is thereby deemed to be above the cut-off, and therefore—broadly—lendable. The rest of this full-file population resides in pools of consumers whose bad-rates are above the cut-off even though the overall bad-rate is between 4% and 45%.

With traditional credit scores it is clear that you have to throw out a lot of likely goods in order to limit the number of bads you let in. Among the marginal turn-downs there are many more goods than bads—they just can't be distinguished using current scores based on current data sources and methodologies.

The non-lendable population also includes the thin-files—those people who can be found in credit bureau files but who have insufficient data to generate a traditional score at all.

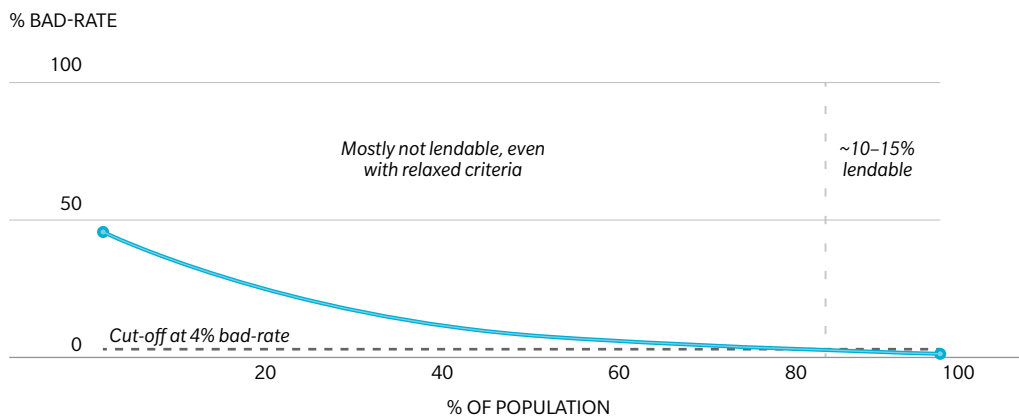
In recent years, both FICO and VantageScore have made changes to their analytical methodologies in order to increase the scoreable population. Specifically, they have relaxed the previous requirement that to be admitted into the score-development process (and therefore to be viewed as 'scoreable' by the resulting model) the consumer's file needed to have at least one tradeline that had been updated within the prior six months, and at least one tradeline that was six months old, or older. According to VantageScore Solutions, the methodology they have adopted for version 3.0 can score 30–35 million more consumers than the previous version, for a total of 215–220 million scoreable adults. For inclusion in the 3.0 version, a consumer needs to have just one month of payment history and at least one tradeline reported to the bureau in the last two years.

FICO's traditional model can score ~190 million adults. But Fair Isaac has a relatively new score developed jointly with LexisNexis Risk Solutions and Equifax, called FICO Score XD. This score is currently developed for credit card applicants only; it is said to be able to score millions more consumers than classic FICO scores. It achieves its broader coverage, in part, through the inclusion of alternative data provided by LexisNexis Risk Solutions and Equifax—data that can substitute for elements of a traditional full-file.

An obvious drawback of loosening the filter for scoreability using credit bureau data alone is that the incremental scoreable population only has limited credit file information from which to generate the new score. Thus the resulting credit score is inherently less reliable than the conventionally scored full-file population—at least for the ‘incremental’ scores.

Exhibit 4 shows a hypothetical (but directionally correct) illustration of what happens when you loosen the criteria for generating a score on a consumer so that many former thin-files become scoreable.

Exhibit 4: Illustrative depiction of lendable thin-file population



The score for these thin-file consumers successfully rank-orders people from low-to-high bad rates. The bad-odds rate can be expected to be slightly different than we saw in *Exhibit 3*, with the highest scores being associated with a slightly higher bad-rate (bottom-right corner). The bad rate rises gradually to approximately the same end point (lowest scores have a bad-rate of ~50%) but, crucially, the bad-rate intersects the cut-off line at a point where only ~10–15% of this group is rendered lendable. The remainder are still not lendable despite being scoreable and despite having far more goods than bads in the group.

Thus, if consumers with insufficient credit data to be scored by conventional models are nevertheless scored using only that thin-file data, it is unlikely that a prime score will result for very many, i.e., these consumers may be scored but most will still not be deemed lendable.

A recent FICO study showed that consumers with lower recent credit activity generate a flatter (and thus weaker) score-odds rank-ordered relationship⁵. Similarly, VantageScore reports that among the ~30–35 million consumers considered newly scoreable under Version 3.0 only ~3% had a score of above 680. Meanwhile, the estimated bad rate for all 35 million newly scoreable consumers is ~30%. There are therefore still thought to be many millions of potentially good borrowers in this newly scoreable sub-population—around 25 million—but the new score is not yet strong enough to separate goods and bads to the point that more than a few become lendable.

⁵ Source: FICO (To Score Or Not To Score, 2013)

This is partly because of the inherent limitations of the data contained in credit bureau files—and especially thin-files. Current credit bureau data are only part of the solution to increase access to credit-invisible consumers who are in fact credit-worthy, i.e., willing and able to pay back loans.

ALTERNATIVE DATA TO THE RESCUE?

Since the passing of the Fair Credit Reporting Act in 1970, bureau data have been transformative in enabling widespread lending that is reliable and fair. These data have been essential for generating the \$12.3 trillion in outstanding consumer debt in the U.S.⁶ Over the last several years, however, industry participants have searched for additional reliable data sources that can provide information on a consumer's ability to honor financial commitments.

Alternative data may provide additional financial payment information on consumers or otherwise provide information with predictive power; some of the sources of such data are:

- Utilities (gas, water, electricity)
- Telecom (TV, mobile, broadband)
- Rent
- Property/asset record: including value of owned assets
- Public records: beyond the limited public records information already found in standard credit reports
- Alternative lending payments (e.g., payday, instalment loan, rent-to-own, buy-here-pay-here auto loans, auto title loans): including both on-time and derogatory payment data
- Demand deposit account (DDA) information: including recurring payroll deposits and payments, average balance, etc.

Several of these alternative data sources involve records of whether a consumer makes payments that they have committed to make (rent, utilities). In credit bureau jargon, when someone fails to make such payments and that is reported to a bureau it is referred to as 'negative data'; conversely when on-time payments are reported to a bureau it is referred to as 'positive data'. The traditional credit bureau market in the US is based on both types of data, but mostly just for payments made on loans. If alternative data were also to be reported, there is some debate about whether positive, or negative, or both types of data should be reported. This is a point we will return to later.

The value of alternative data varies by source. Data like rent payments have been shown to be predictive and may be available on many consumers with no credit file (although many landlords now demand credit scores for new tenants!) But the rental market is very fragmented and data are not uniformly reported. Therefore coverage is low. Public records data are available on even more people than those with files at the credit bureaus, so their coverage is very wide. However, since the information contained in public records is not explicitly focused on payments, it is not as predictive as credit bureau data; nevertheless,

⁶ Source: Federal Reserve, Q3 2016

it also proves to be additive in scores developed from combined data sources. The main characteristics of a good source of alternative data are these:

- **Coverage:** a new data source will ideally have broad and consistent coverage (e.g., over 90% of U.S. adults use a cell phone, and the market is concentrated so data collection would be easy to achieve; approximately 40% of U.S. adults pay rent, but this is a low-concentration market and so the data are expensive to collect)
- **Specificity:** a data source should ideally contain detailed data elements about an individual—data elements that provide part of a full picture of the borrower (e.g., on-time and late payments over a significant time series, or specific asset or income data); some data sources are based on ‘segment data’ or ‘modeled data’ and are typically less predictive than consumer-specific sources
- **Accuracy and timeliness:** data should be accurate and frequently updated; a data source should have a system for ongoing data verification and management
- **Predictive power** (‘signal’): most important, data should contain information relevant to the behavior that you’re trying to predict
- **Orthogonality:** ideally, the data source should be additive to traditional bureau data; this means that using it will improve the predictive accuracy of any new score by improving the signal-to-noise ratio
- **Regulatory compliance:** data sources must comply with existing regulations for consumer credit (i.e., Fair Credit Reporting Act, Equal Credit Opportunity Act, Gramm-Leach-Bliley Act)

One important development of the last few years has been the realization by big banks that internal data is also ‘alternative data’ and can play a role in predicting default behavior. Checking account data, in particular, has been shown to predict repayment behavior and to enhance the predictive power of traditional credit scores. This seems so intuitively obvious that it is surprising the idea took so long to emerge. But patterns of transaction and balance behavior in consumer—and small business—checking accounts can be used to predict default behaviour quite well, and to enhance the predictive power of established credit scores. We are aware of a few banks that have developed and deployed such enhancements or who are in the process of doing so—indeed some of them are our clients. We expect this to become the norm over the next 5–10 years. There is even a possibility that consumer and small business checking account data could become available to third-parties and not just to the banks who manage the accounts. If the US adopts its own equivalent of the EU’s Payment Services Directive and related ‘Open API’ movement, then it could lead to the availability of such data in enhanced credit score development. These moves would empower consumers to make their data available to credit grantors, via data intermediaries (who could be the existing credit bureaus or new players) but only with the permission of the consumer.

We also expect the availability of third-party alternative data to broaden and deepen, possibly even through legislative initiatives to require reporting to the established bureaus. There have been periodic moves in the US towards mandating the reporting of some alternative data to the credit bureaus.

BENEFITS OF ALTERNATIVE DATA

Having more data is only valuable if it results in real incremental benefits; in this case, the benefits of using alternative data in addition to traditional bureau data, beyond just technical improvements to the credit score, should flow to both consumers and lenders.

As discussed earlier, many newly scoreable consumers should ideally now be part of a sufficiently accurate risk-separation pattern that a fraction of them also become lendable. That is, they have to have a score above the cut offs used by most lenders.

For consumers, the use of alternative data provides two distinct benefits: first, more potential borrowers will be able to secure a loan, including many from among today's credit invisibles. In a recent Experian study where energy utility tradeline data were added to the file, almost half of the subprime participants moved up to either the nonprime or prime category.

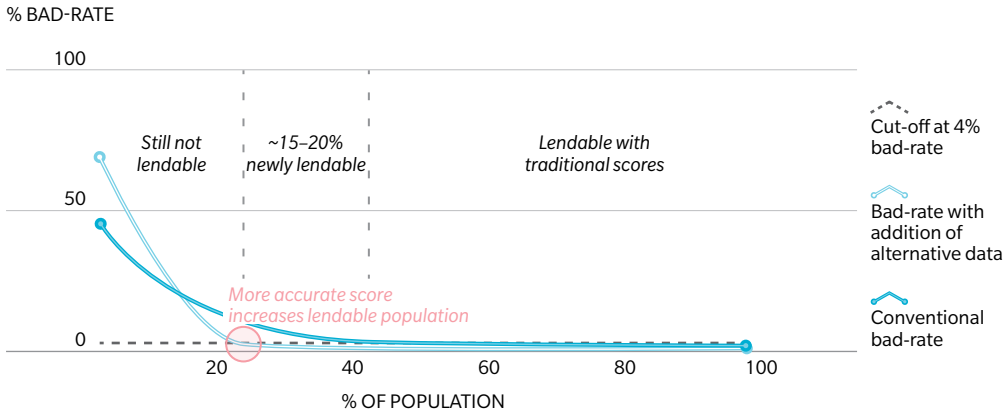
Second, most existing borrowers will have access to lower interest rates; that is, they will receive more favorable pricing because the small percentage of bad risks will be better separated from the good risks than is possible with traditional scores. The same Experian study showed that even for customers that remain in a given risk segment (subprime, nonprime, prime), moving from thin-file to full-file yields some improvement in pricing. As an example, the average credit card interest rate fell from 23% to 21%, a 10 percent reduction, for those (previously thin-file) subprime borrowers who had sufficient utility payment history to qualify them as full-file⁷. This is an important and often-overlooked benefit in the push to add alternative data.

As we have seen, the objective of a credit score is to separate likely good borrowers from likely bad ones; effective sources of alternative data must therefore provide a finer separation within and across credit score bands, improving the accuracy of credit scores. In other words, better credit scores will be better precisely because they can cluster bads in greater concentration at the low-score end of the spectrum. This means that the combination of alternative data and traditional bureau data should improve the score-versus-bad-odds relationship for the new score such that there will be an observable *decrease* in default rates associated with higher and mid-level credit score bands and an observable *increase* in associated default rates in lower credit score bands. In other words, 'the curve will become steeper'. *Exhibit 5* illustrates the effect of adding new data that improve the accuracy of existing scores and allow credit invisibles to be meaningfully scored.

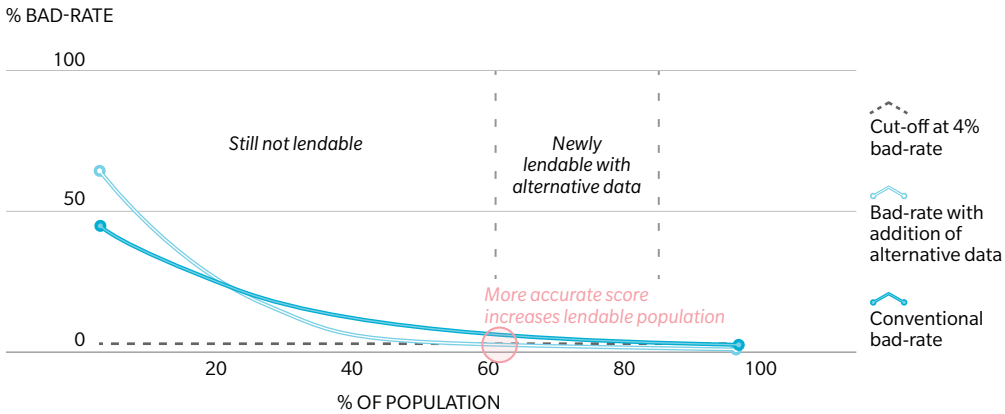
⁷ Source: Experian (Let There Be Light – The impact of positive energy-utility reporting on consumers, 2015)

Exhibit 5: Illustrative depiction of lendable population with addition of alternative data

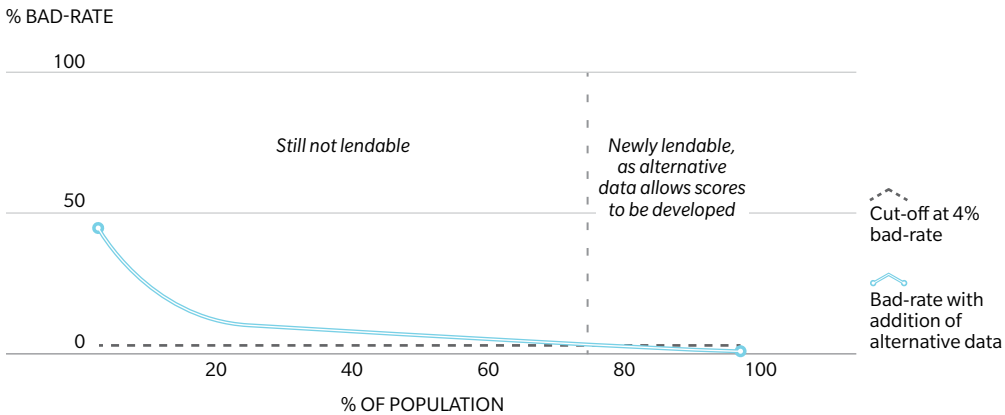
FULL-FILE



THIN-FILE



NO-HIT



As *Exhibit 5* illustrates, a more accurate credit score can significantly expand the lendable fraction among the full-file population. The diagram repeats the earlier bad-rate curve and superimposes a hypothetical curve, based on the addition of alternative data, that can more effectively separate goods and bads. From a mathematical point of view, the area under the curve is the same (these are the bads). But they are more clearly concentrated in the ‘bottom left-hand corner’. This has the effect of shifting the place where the new bad-rate curve intersects the cut-off line much farther to the left: far more people are now lendable. The diagram also shows the improvement in scoreability and lendability among the thin-file population, relative to the use of traditional data alone. Finally, the no-hit population is represented as now being, at least to some degree, scoreable. The power of the score for the formerly no-hit population is likely to be relatively weak (the bad-rate curve intersects the cut-off rate at a point representing perhaps only 20–30% of this population) but this will still represent a big improvement.

The Corporation for Enterprise Development (CFED) estimates that by adding utility payment data alone, the no-hit population would decline from 20–25 million adults to approximately five million adults. Mike Mondelli, SVP of TransUnion Alternative Data Services, has provided an even more optimistic estimate:

“Alternative data sources (such as property, tax, deed records, checking and debit account management, payday lending information, address stability and club subscriptions) have proven to accurately score more than 90% of applicants who otherwise would be returned as no-hit or thin-file by traditional models.”

For lenders, the primary benefit of alternative data is the increased number of profitable loans to be made consistent with a given risk appetite. Additional benefits may include reduced transaction costs (additional borrower information could reduce the need for manual processes) and lower aggregate credit losses for any cut-off defined by a marginal loss rate. Also, by having a more complete picture of prospective borrowers, lenders will be better positioned to offer competitive interest rates, which many lenders cite as a challenge today.⁸

Some people have expressed concern that alternative data could further marginalize vulnerable credit-excluded populations. They argue that alternative data reporting should include only positive data (meaning on-time payments) and not negative data (payments made late). This may seem humane in that it targets the inclusion of data on deserving people, while discreetly ignoring late-payers, but it doesn’t actually help the late-payers even as it subtly undermines the overall ability to provide accurate scores and default probabilities on timely payers. Besides, if a landlord was only reporting a renter’s timely payments and then suddenly one month filed no report, the statistical inference will soon be made that this missing report represents a late payment. To the extent that is not actually true (the renter may have moved out) it would be better simply to report all payments, timely or not. It certainly would support more accurate risk assessment across the entire population. This is an area where many policy proposals to date have been less-than-entirely logical.

⁸ Source: TransUnion (The State of Alternative Data; survey of 317 lenders, 2015)

Overall, research indicates that far more people who were previously unscorable and unlendable *benefit* from the inclusion of alternative data than are further disadvantaged by it. In one PERC study, adding alternative data increased credit scores for 64% of thin-file borrowers and reduced the credit score for only 1% of the same population⁹. For full-file borrowers, an Experian study demonstrated that inclusion of energy utility tradelines improved or had no impact on credit scores for 97% of the participants and reduced credit scores for just 3% of the participants¹⁰. The idea that alternative data can prove very helpful to low-score and thin-file populations is perhaps easier to understand in light of the numbers provided earlier showing how high the good rates have been estimated to be in these non-lendable populations.

Of course, more accurate credit scores will mean that there are a few people who will now have an even harder time accessing credit; high accuracy in separating bad risks from good ones is what confers benefits on the better risks but it makes things somewhat harder for the bads themselves. However, from a public policy perspective this should not be viewed as a drawback. The ability to borrow comes to those who can demonstrate that they are willing and able to repay their loans; access to credit is not a constitutional or human right. If the addition of alternative data accurately deems an individual unlendable, that person at least becomes protected from over-extending himself in debt. It also encourages him to change his behavior to access credit in the future. And there are both private enterprise initiatives (Credit Karma, LendUp and many banks' credit advisory initiatives) and public policy options available to help such people make the necessary changes in behavior. In the meantime, better risk separation helps improve access and reduce prices for a far-greater number of credit-worthy borrowers, while reducing credit losses for lenders. Thus, excluding these few people—the true bads—from credit access actually benefits them, all other borrowers, and lenders.

9 Source: PERC (A New Pathway to Financial Inclusion: Alternative Data, Credit Building, and Responsible Lending in the Wake of the Great Recession, 2012)

10 Source: Experian (Let There Be Light – The impact of positive energy-utility reporting on consumers, 2015)

LOOKING AHEAD: EMBRACING ALTERNATIVE DATA

According to a TransUnion survey, 34% of lenders already use some types of alternative data to evaluate both prime and nonprime borrowers. Use of alternative data is most prominent in credit card, auto loans and consumer finance, as well as in the FinTech world but there are also signs of early adoption in the mortgage industry. 66% of lenders surveyed reported that they were able to lend to additional borrowers in their current markets and 56% reported access to new markets by using alternative data¹¹.

Major credit bureaus (Experian, Equifax and TransUnion) are already starting to incorporate alternative data within their databases, through acquisitions and/or partnerships. For example, Equifax acquired TALX (employment and income verification) and IXI (consumer wealth and asset data), Experian acquired RentBureau (positive and negative rent payment history), and TransUnion acquired L2C (public records and selected alternative lender data sources). However, there is significant opportunity to increase the scope and coverage of alternative data much further at each of these credit bureaus.

Being a free market, lenders and data sources will surely work towards more widespread use of alternative data—after all, it opens new profit pools for both parties. However, the social benefits of using alternative data may be significant enough to warrant the use of legislation to accelerate the pace of adoption by mandating the reporting of selected alternative data.

A CLOSER LOOK AT LENDING AND CREDIT SCORES

Lenders evaluate potential borrowers through an underwriting process that relies heavily on credit scores derived from data in applicants' credit files. These credit files are maintained by the major credit bureaus (Equifax, Experian and TransUnion). A credit file contains the following four types of information:

- **Header data:** borrower personal information (i.e., name, address, social security number, date of birth—information that helps define *identity*)
- **Public records:** e.g. debt-related derogatory public records (bankruptcy, tax lien, civil judgements) that remain on the credit file for up to ten years
- **'Tradeline data':** data about each loan or line a consumer has taken out; this includes *descriptive* data, such as lender, loan type, loan amount, date credit was extended and closed, and *payment information*, such as payment status, date of first delinquency (late payment), late-payment status (30 days, 60 days, 90 days), for each loan. The types of loan that can be found in a credit file include both secured loans (first mortgages, home equity loans, home equity lines of credit, auto loans) and unsecured loans (credit cards, retail store cards, personal loans, and student loans)
- **Inquiries:** the bureaus count and log the number and details of credit checks that lenders make (on borrowers) before they conclude a lending decision

These credit file data are used to build a credit score. Credit scores first appeared in the US in the late 1960s and 1970s but their use has subsequently become ubiquitous.

¹¹ Source: TransUnion (The State of Alternative Data; survey of 317 lenders, 2015). Definition of alternative data is unclear.

Today, credit scores are used by lenders to determine whether or not they will lend to a given borrower, to help set the terms of a loan, like the loan (or line) amount and the interest rate, and for ongoing credit risk management and collection strategies.

Credit scores are developed by credit bureaus, data analytics companies, or by individual lenders. Scores may be 'generic', meaning that the score has been built using broad industry cross-sectional data, and such scores can be applied across lending products. Scores can also be product-specific (e.g., an auto loan score, a credit card score).

The most widely used generic credit scores are 'the FICO score' (from Fair Isaac Corporation¹²) and VantageScore¹³ with scores ranging from 300 to 850. Large lenders sometimes develop their own proprietary credit scores. For example, firms like American Express, Bank of America and Chase have built numerous in-house credit scores for specific products like credit cards or personal loans.

To build a credit score, companies take a very large sample of consumer loan data that will include many loans that are being repaid on time (are 'current') and some that are delinquent; they take the data from an historical starting point that precedes a future 'performance window' that is typically 18 or 24 months. A statistical technique such as logistic regression is used to determine which independent consumer credit variables in the credit file—available at the starting point—have a statistically significant correlation with the most important dependent variable: late payment during the performance window.

The resulting model rank-orders consumers from least likely to repay (those with low scores) to most likely to repay (those with high scores). The credit scores can then be calibrated against a pool of relevant data to show how the scores may translate into a so-called 'bad odds rate'. 'Bad odds' are basically the observed default rates within groups of borrowers with the same or similar credit scores. It is worth noting that a score can be calibrated against different populations and generate somewhat different bad odds for each score-band. For example, a generic score like VantageScore can be calibrated for new credit card borrowers, or it can be calibrated for new indirect auto loans.

For computational convenience, the bad rate usually does not define 'default' but 'serious delinquency'—often 90 days past due. When loans are seriously delinquent, some may 'cure' and revert to a current or fully-paid status, but a high (and relatively stable) percentage will go on to default and so the bad rate of a credit score can also be used to directly estimate a probability of full default.

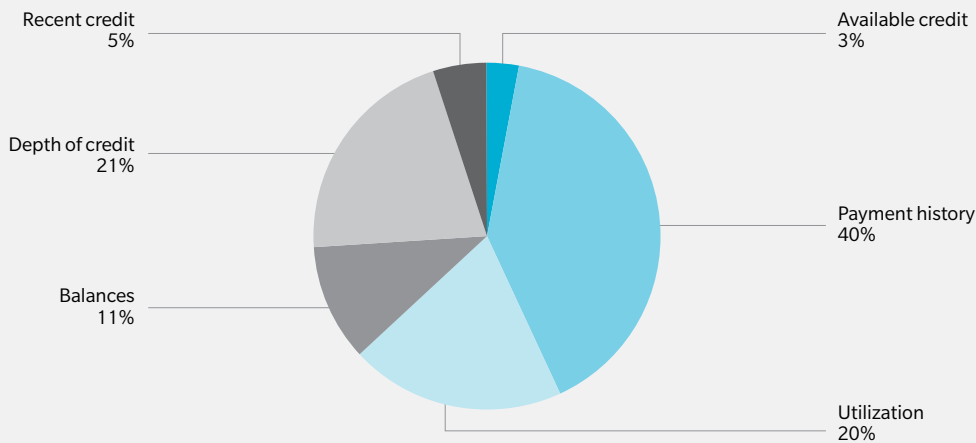
Although credit score models can vary, they all leverage similar types of information contained within a credit file. The categories of consumer information used to develop VantageScore 3.0¹⁴, with their approximate weighting, are shown below:

¹² There are actually nearly 30 different 'FICO' scores

¹³ VantageScore 3.0 is developed by VantageScore Solutions LLC, a joint venture between three major credit bureaus, Equifax, Experian and TransUnion

¹⁴ Source: VantageScore

Exhibit 6: Attributes used to develop VantageScore



- **Payment history:** The most significant predictor of future loan repayments is previous payment history. Notable factors within this category include payment history across all loan accounts, length of time without a missed payment, length of time of credit use, number and severity of prior delinquencies, presence/absence of severe unpaid debts such as bankruptcy or foreclosure
- **Utilization:** An individual in financial stress is more likely to use a greater portion of their available credit line for revolving balances (such as credit cards, lines of credit). Because of its 'weight' in the final score, VantageScore recommends consumers keep this utilization below 30%
- **Current balances:** The extent of indebtedness (total current debt outstanding) is an important factor in predicting future payment behavior. For term accounts (such as mortgages and instalment loans), a lower ratio of debt outstanding to total loan is more desirable
- **Depth of credit:** A greater variety of accounts (e.g., mortgage, credit cards, auto loan) over time is indicative of a mature healthy credit user and positively influences credit score
- **Recent credit:** Too many new accounts opened within a short time period could be a signal for financial duress, and statistical analysis suggests that it actually is! This portion of the score incorporates the elapsed time since the most recent account opening, number of new accounts, total new loan amount, and number of recent inquiries
- **Available credit:** Borrowers with too much unused credit are considered riskier, as in the event of financial distress, they will be able to run up a large amount of additional debt (and overextend themselves in the process). VantageScore suggests that prime consumers keep an average of \$20,000 to \$22,000 of available credit

In line with the Equal Credit Opportunity Act (ECOA), credit scores may not employ variables such as race, age, sex or marital status, even if these are statistically predictive of future repayment behavior.

Oliver Wyman is a global leader in management consulting that combines deep industry knowledge with specialized expertise in strategy, operations, risk management, and organization transformation.

For more information please contact the marketing department by email at info-FS@oliverwyman.com or by phone at one of the following locations:

AMERICAS

+1 212 541 8100

EMEA

+44 20 7333 8333

ASIA PACIFIC

+65 6510 9700

PETER CARROLL

Partner in the Retail and Business Banking Practice

peter.carroll@oliverwyman.com

SABA REHMANI

Engagement Manager

www.oliverwyman.com

Copyright © 2017 Oliver Wyman

All rights reserved. This report may not be reproduced or redistributed, in whole or in part, without the written permission of Oliver Wyman and Oliver Wyman accepts no liability whatsoever for the actions of third parties in this respect.

The information and opinions in this report were prepared by Oliver Wyman. This report is not investment advice and should not be relied on for such advice or as a substitute for consultation with professional accountants, tax, legal or financial advisors. Oliver Wyman has made every effort to use reliable, up-to-date and comprehensive information and analysis, but all information is provided without warranty of any kind, express or implied. Oliver Wyman disclaims any responsibility to update the information or conclusions in this report. Oliver Wyman accepts no liability for any loss arising from any action taken or refrained from as a result of information contained in this report or any reports or sources of information referred to herein, or for any consequential, special or similar damages even if advised of the possibility of such damages. The report is not an offer to buy or sell securities or a solicitation of an offer to buy or sell securities. This report may not be sold without the written consent of Oliver Wyman.