PERC

RESULTS AND SOLUTIONS

Alternative Data in the US: Progress, Promise, and Paralysis

WHITE PAPER

August 2019

SUMMARY

Despite being used in over 90 countries globally, the use of proven payment data (such as utility and telecom payment data) to improve access to credit remains mostly an aspiration in the US. While regulatory uncertainty continues to impede policy solutions to Credit Invisibility in the US, the market has responded. In little more than a decade, a surge of solutions has emerged ranging from alternative data scores, to consumer-permissioned data where individuals provide access to payment data in their online bank account or populate their credit file with proven payment data. While this represents progress, the resulting array of databases and solutions may be confusing to consumers. What consumers know best is their traditional credit report and how their payment behavior affects their credit worthiness. Arguably, American consumers and lenders are best served by flowing proven payment data through FCRA-regulated databases maintained by nationwide consumer reporting agencies. In so doing, consumers are unburdened and automatically benefit, while lenders benefit from a comprehensive view of a prospective borrower. Continuing along the current path may result in undue fragmentation of the data market and increased consumer confusion.

RAPIDLY CHANGING LANDSCAPE

In 2005, when PERC's first report on alternative data was published, the landscape of solutions providers was sparsely populated, and included a handful of firms including Pay Rent Build Credit (later "PRBC"), Lexis-Nexis, and a smattering of fully reported energy utility accounts at Experian and TransUnion. By 2019, there had been an explosion in new and alternative credit scoring and risk assessment solutions, utilizing a mix of traditional and nontraditional data sources. Newer entrants offered both new data assets (cash flow data from bank accounts, rental payment) and new methods to accessing data (consumer permissioned data).

PERC's *Research Consensus Confirms Benefits of Alternative Data*, published in 2015, presented findings from FICO, TransUnion, Equifax, VantageScore, Experian, and LexisNexis.[1] The findings showed clear benefits from the then-newer solutions in terms of improving underwriting *and* reaching more consumers (inclusion). Of course, the proliferation of alternative solutions providers was a direct response to the pent-up demand among creditors and others for new and more predictive data. For those firms with scale solutions, the revenue growth was impressive. For example, TALX's "Work Number" increased its income and employment verification revenue from $131.9 million in 2008 to $567 million in 2018.[2]

In addition to those large market players, there were also a number of smaller players and startups at the time with innovative solutions. Some of those solutions focused on particular market segments, such as immigrants with little credit history in the US or other persons with little to no prior credit history—the so-called Credit Invisibles.[3] Since that paper was published in 2015, growth in newer solutions has continued with the release of the FICO XD model and the emerging segment of so-called "consumer-permissioned" data solutions, where, for instance, consumers authorize a third-party to access their bank account or telecom/utility statement data. Examples of these are Experian's Boost, Urjanet, Envestnet/Yodlee, and Finicity solutions.[4]

Typically, consumer-permissioned data becomes available to a user such as a lender directly or, with Experian Boost for instance, is added to a consumer reporting agency (CRA) credit report where the data can be used by some credit scores. Unlike the traditional credit bureau model where data are "automatically" and directly reported to the nationwide CRAs (TransUnion, Experian, and Equifax), consumer-permissioned data requires the consumer to authorize the sharing of data from their accounts (bank, utility, assets, etc.) each time it is accessed. IT advances and online consumer access to account information has made this model possible (if not inevitable, given the growing ease of sharing data).

---

[1] Turner, Michael A., Robin Varghese and Patrick Walker. *Research Consensus Confirms Benefits of Alternative Data*. Durham, NC. PERC Press. March, 2015. Downloaded at http://www.perc.net/wp-content/uploads/2015/03/ResearchConsensus.pdf

[2] See Equifax, *2009 Annual Report*, at https://investor.equifax.com/~/media/Files/E/Equifax-IR/Annual%Reports/2009-annual-report.pdf; and Equifax, *2018 Annual Report*, at https://investor.equifax.com/~/media/Files/E/Equifax-IR/Annual%Reports/2018-annual-report.pdf;

[3] "About Us." Nova Credit. Available at: https://www.novacredit.com/about-us

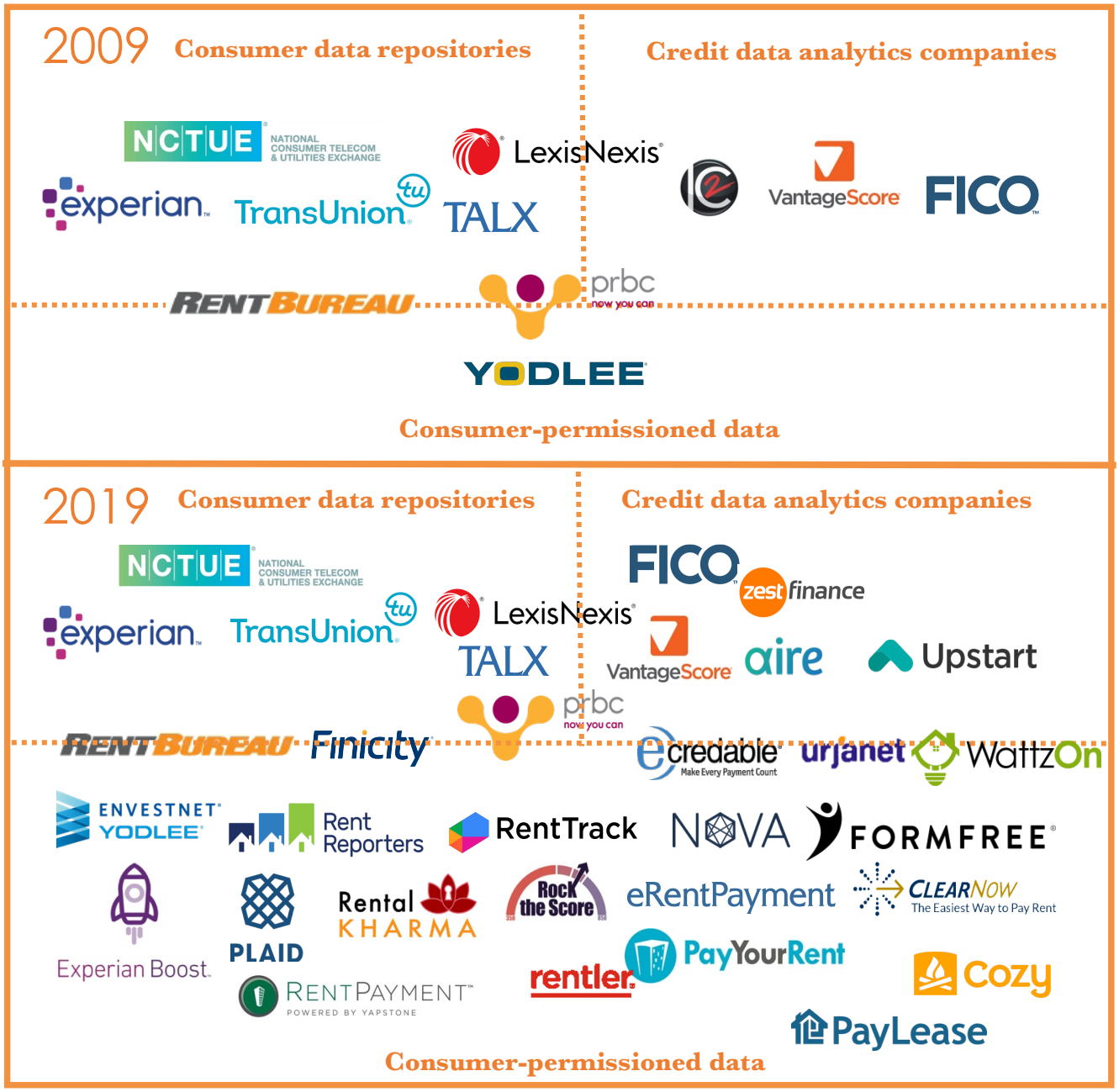
[4] See Experian Boost at https://www.experian.com/consumer-products/credit-score.html; Urjanet at https://urjanet.com/; and Finicity at https://www.finicity.com/.

As Figure 1 shows, the growth in the number of firms with proven payment data and alternative data assets and solutions has been impressive over a relatively short time span. It also illustrates a trend toward market fragmentation. It is important to note that Figure 1 is just an example to illustrate the growth in the alternative data space and, no doubt, is incomplete. It is also difficult to perfectly categorize these companies as some may fall in multiple categories (those straddling the category lines in Figure 1), such as Experian having analytics solutions, a data repository, and consumer permission data solutions. In addition there are also other examples of consumer permission information, such as consumers authorizing the IRS to share return data via the 4506T form, but these sources have been a mainstay of the underwriting process for many years and are not the newer variety of consumer-permissioned data solutions of interest here.

Figure 1 also does not capture the number of solutions per firm or sales volumes. For instance, although FICO had the Expansion Score in 2009 it now has a more advanced line of solutions, such as the FICO Score XD and UltraFICO. The same can be said of Lexis-Nexis, and the other firms shown in 2009. There are also lenders that utilize consumer permissioned data more directly, such as Petal, to supplement underwriting or to be able to extend mainstream financial service products to consumers with little credit history.[5] We did not include this category since these were not strictly data/analytics companies. As such, figure 1 understates the true growth the alternative data space.

[5] See Petal's website: https://www.petalcard.com/

Figure 1: Example of Alternative Data Landscape in 2009 and 2019



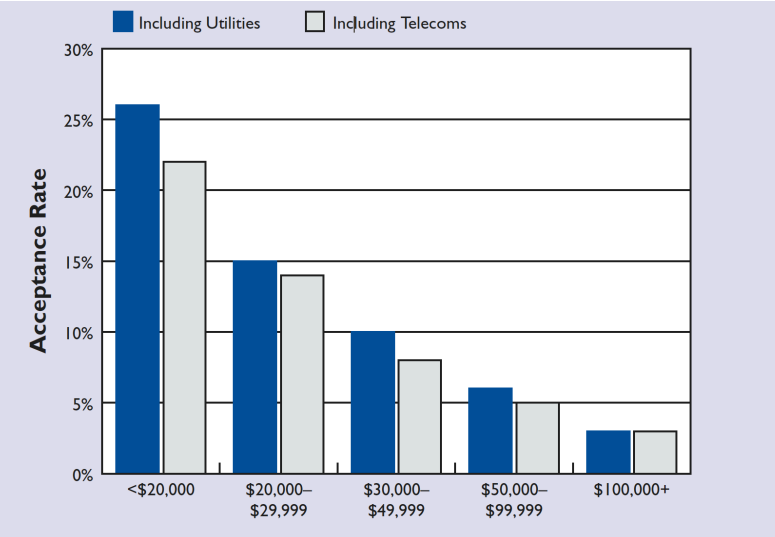
WHY THE INTEREST IN NEW SOLUTIONS?

The CFPB, in their "Credit Invisible Data Point," found that around 1 in 5 adult Americans (45 million) were unscoreable using traditional data and credit scores at the time of their study.[6] This rate jumps to 45% in the lowest income census tracts. Work by PERC and others have also

[6] Kenneth P. Brevoort, Philipp Grimm, and Michelle Kambara. "Data Point: Credit Invisibles." CFPB. May 2015. Available at: https://files.consumerfinance.gov/f/201505_cfpb_data-point-credit-invisibles.pdf

found that members of lower income households have much higher unscoreable and thin-file rates.[7] For example, *Give Credit Where Credit is Due* found that, in the "utility" sample, for individuals with household incomes under $20,000, the thin-file rate was 31%.[8] This compares to a rate of just 4% for those from households with incomes of $100,000 or over. When fully reported proven payment data (timely and late utility or telecom payments) was added to the credit files, simulated credit acceptance for those in the lowest income group increased over 20%, while it increased under 5% for those in the highest income group (see Figure 2).

Figure 2: Acceptance rate of consumers by income with alternative data included in tradelines (assumes 3% serious delinquency rate)



In addition to those who are unscoreable, there are a number of consumers who are thin file and are barely scoreable. For these consumers, adding additional data to be considered by the credit scoring models—referred to as "thickening" their credit reports—can also improve how well the credit scores perform and increase credit inclusion. In these cases, adding additional information to those with little information is helpful.

PERC's *Predicting Financial Account Delinquencies with Utility and Telecom Payment Data* examined whether adding utility and telecom payment data to credit scores could improve risk assessment for mortgages and among people *with* mortgages), benefiting consumers and lenders.[9]

[7] Thin-file consumers are typically defined as those with fewer than three traditional accounts reported in their credit file.

[8] Michael Turner & Alyssa Lee, *Give Credit Where Credit is Due: Increasing Access to Affordable Mainstream Credit Using Alternative Data*. Washington, DC: The Brookings Institution, December 2006, available at http://www.perc.net/wp-content/uploads/2013/09/alt_data.pdf

[9] Turner, Michael A. and Patrick Walker. *Predicting Financial Account Delinquencies with Utility and Telecom Payment Data*. Durham, NC. PERC Press. May, 2015. Available at: http://www.perc.net/wp-content/uploads/2015/05/Alt-Data-and-Traditional-Accounts.pdf

There are clear benefits in terms of credit inclusion and underwriting generally when credit scores can go beyond "traditional" credit data and also incorporate a broader array of data.

THE BENEFITS OF CONSUMER-PERMISSIONED DATA & SPECIALTY DATABASES

Many data elements may be useful in underwriting and credit scores but also may not fit neatly in the typical credit bureau framework that includes full-file account payment histories, collections, inquiries, and basic public records. These data elements could include information on a consumer's income, employment history, assets, bank account cash flow, and expanded public records. Specialty CRAs have developed to collect this data or otherwise make it available as needed—including for verifying credit application data. The CFPB produces an annual list of the nation's CRAs, many of which are focused on areas other than credit origination, and the majority could be considered as specialty CRAs.[10] Lexis-Nexis Risk Solutions (LNRS), for example, collects a number of data elements that have been shown to improve underwriting, notably for those with little traditional credit history.[11] LNRS offers the Risk-View solution for lenders using this data and also collaborates with other firms create credit scores built on a combination of the LNRS data and other data elements.

Details of bank account cash flows and balances, tax return information, value of assets can all be made available in a consumer permissioned model. This type of model may in fact be ideal for many such data elements. Solutions such as Urjanet, ExperianBoost, Envestnet/Yodlee and FICO XD (which combines data from a nationwide CRA and a specialty CRA) are providing ways for consumers who have little or no traditional credit data to gain or improve their credit standing/scores. In other words, the market is incrementally moving to reduce Credit Invisiblity—thereby making the US financial services sector fairer and more inclusive.

The introduction of such solutions is a truly positive development.

THE NEED TO FULLY REPORT PROVEN PAYMENT DATA TO THE NATIONWIDE CRAs WHEN POSSIBLE

However, despite the new solutions using consumer-permissioned data and specialty CRA databases described above, it is still crucial that financial accounts, such as credit cards, auto loans, and mortgages, are fully reported to the main nationwide CRAs. PERC has studied the impact of fragmented and segmented CRAs with low lender and/or consumer coverage seen in

[10] Consumer Financial Protection Bureau, *List of Consumer Reporting Companies.* Washington D.C.: Consumer Financial Protection Bureau, 2019, accessed at: https://files.consumerfinance.gov/f/documents/cfpb_consumer-reporting-companies-list.pdf

[11] See Jeffrey Feinstein. "Alternative Data and Fair Lending." Lexis Nexis. 2013. Available at: https://insights.lexisnexis.com/creditrisk/wp-content/uploads/2013/09/alternative-data-and-fair-lending-wp.pdf

some other countries and found they substantially underperform relative to America's full-file and comprehensive credit reporting system.[12] Indeed, CRAs are key financial infrastructure.

Table 1: Predictive power of models in various credit data sharing scenarios

Data Sharing Arrangement		Relative K-S Statistic
Scenario 1 (Full-file comprehensive)	Positive & negative information from all reporting sectors are available; all furnishers participate in providing payment information.	100.00
Scenario 2 (Bank simulation)	Positive and negative information from banks is available; only negative payment information of 90+ days past due from non-banks is available.	97.93
Scenario 3 (Non-bank simulation)	Positive and negative information from non-banks, with the exception of 25% of non-bank revolving credit (or financial credit cards). No bank information is available.	92.46
Scenario 4 (Lower participation)	Only 50% of furnishers (bank and non-bank) provide positive and negative information, while the other 50% provide only negative information.	95.50

Source: Table 4 of Michael Turner, Robin Varghese, & Patrick Walker, On the Impact of Credit Payment Reporting on the Financial Sector and Overall Economic Performance in Japan. New York: The Information Policy Institute, March 2007, available at http://www.perc.net/wp-content/uploads/2013/09/Japan.pdf.

Table 1 shows how the "goodness-of-fit" or how well a credit scoring model's ability to assess credit risk as measured by the K-S statistic degrades as data available to the credit scoring model is removed. The K-S statistic ranges between 0 and 100, with 100 indicating a perfect fit where the model perfectly predicts which consumers would default. Here, Scenarios 2, 3 and 4, which utilize less information, are compared to Scenario 1, which uses the fullest and most complete set of information. So, if the actual K-S was, say, 60 in Scenario 1, then the K-S in Scenario 3 is only 92.5% of this or would be about 55.5. These (and other examples) clearly show that reducing information that are used as inputs in credit scores reduces credit score performance.

In addition to the core "financial" account data, other data, which we call "Proven Payment Data" that includes utility, telecom, and rental payment histories, should be fully reported to the main consumer databases of Nationwide CRAs to the extent possible. These payment histories fit well with other "financial" payment histories currently reported. In fact, *some* millions of utility, telecom and rental histories are already currently reported to these main databases and have been for many years. And as discussed previously, these data elements have the potential to increase credit inclusion, particular for members of lower-income households. This is well-established.

[12] Michael Turner, Robin Varghese, & Patrick Walker, *On the Impact of Credit Payment Reporting on the Financial Sector and Overall Economic Performance in Japan*. New York: The Information Policy Institute, March 2007, available at http://www.perc.net/wp-content/uploads/2013/09/Japan.pdf

To capture the full benefits of the proven payment data, it is best if that data is reported to the main consumer databases of all three nationwide CRAs. Some lenders may pull reports and scores from just one CRA and some lenders only use data from the main consumer databases at the CRAs. So, having proven payment data available *solely* via specialty CRAs, or via a consumer permission method, will result in it being underutilized and many consumers being excluded or inaccurately and unfairly assessed when applying for credit and other permissible purposes (jobs, insurance). While this is a better outcome than having predominantly negative non-financial payment data reported to nationwide consumer reporting agencies (as is currently the case), it is socially and economically suboptimal to pervasive full-file non-financial credit reporting that would be the case, for example, under a reporting mandate as enforced in many countries around the world.

Having proven payment data in separate databases also acts to unduly fragment the credit reporting system, where some payment histories are in one database and some are in others. This is not ideal, as it reduces the effectiveness of the credit reporting system (as seen in other economies) and makes it more difficult for consumers to review their payment histories as they would be in different databases, something of which consumers may be unaware. Very few consumers—likely less than 1 percent—are aware of the NCTUE, which is by far the oldest and largest single repository of energy utility and telecoms payment data. By contrast, nearly everyone is familiar with Equifax—the owner/administrator of NCTUE—and understands where to look for a copy of their traditional credit report.

Having different data in different databases can also make it costlier for score developers (value-added service providers) to create credit scores, as they may need to produce and test several credit scores where they would otherwise produce one. This could also produce undue marketplace confusion if lenders, investors, and regulators need to adjust to credit scores that use different types of payment data.

We, by no means, are arguing that *all* data should be in the same databases. There are good reasons that specialty databases exist. In addition, data can be merged from these various databases and sources to create new solutions (for example FICO XD or Experian Extended View). We are simply making a very practical argument regarding proven payment data. Namely that they are able to be reported to the main nationwide consumer databases (a very small share of such potential data is currently fully reported), they fit well with the traditional credit data, and they would have greater benefits for consumers and lenders if they were reported to all three main nationwide consumer databases.

And while consumer permissioned data holds great promise, there are some limitations with that model. Firstly, its opt-in configuration means there is no universal coverage of the population and the potential uptake of the solutions by Credit Invisibles remains uncertain. Some consumers may not know of such services or find utilizing such services relatively difficult – studies have shown that the "Digital Divide" is wider than previously thought.[13] Secondly, consumers will no doubt cherry-pick, that is choose to report their best, on-time payment histories if given a choice. Data that have major gaps (such as negative-only or positive-only) will ultimately be less

[13] Lohr, Steve. "Digital Divide is Wider Than We Think, Study Says." *New York Times.* December 4, 2018. Accessed at: https://www.nytimes.com/2018/12/04/technology/digital-divide-us-fcc-microsoft.html

beneficial to consumers and lenders compared to full-file data. That said, PERC strongly supports the development of the new consumer-permissioned solutions that supplement traditional credit file data with proven payment data.

However, we believe the first-best solution for large utilities and telecoms that can directly full-file report to the nationwide CRAs is to do so.

This would benefit consumers and the Credit Invisibles to the greatest extent. These large utility and telecom firms, after all, often do use traditional credit file data for eligibility determination; report collections (late payments and unpaid balances) to the CRAs; but, perversely, tend *not* to report on-time payments. In the absence of full-file credit reporting to the CRAs, we do, however, welcome the multiple new solutions that aim to fill this important data gap.

PERC will continue to vigorously support the new solutions that aid consumers and the credit invisible by allowing a greater range of data to be considered in lending and other permissible purposes. But, at the same time, we will also advocate for the first-best solution for consumers of full-file credit reporting of Proven Payment Data (utility, telecom, and rental payment data) to the Nationwide CRAs main consumer databases.

PERC
RESULTS AND SOLUTIONS

Dr. Michael Turner
PERC
6409 Fayetteville Road, Suite 120-240
Durham, NC 27713 USA

+1 (919) 338-2798 x801 Tel.
+1 (919) 975-5565 Skype
+1 (919) 265-3466 Mobile

turner@perc.net
www.perc.net

The **Policy & Economic Research Council** (PERC) is a tax-exempt 501(c)(3) incorporated in the State of North Carolina. Our employer identification number is: E.I.N # 20-5179792.