

# A New Pathway to Financial Inclusion: Alternative Data, Credit Building, and Responsible Lending in the Wake of the Great Recession

Michael A. Turner, Ph.D., Patrick D. Walker, M.A.,  
Sukanya Chaudhuri, Ph.D., Robin Varghese, Ph.D.



June 2012

**PERC**  
RESULTS AND SOLUTIONS

Copyright: © 2012 PERC Press.

All rights to the contents of this paper are held by the Policy & Economic Research Council (PERC).

No reproduction of this report is permitted without prior express written consent of PERC. To request hardcopies, or rights of reproduction, please call: +1 (919) 338-2798.

**A New Pathway to Financial Inclusion:  
Alternative Data, Credit Building, and Responsible Lending  
in the Wake of the Great Recession**

Michael A. Turner, Ph.D., Patrick D. Walker, M.A.,  
Sukanya Chaudhuri, Ph.D., Robin Varghese, Ph.D.

**June 2012**

## Acknowledgments

PERC wishes to thank the Annie E. Casey Foundation, the Ashoka Foundation, Experian and TransUnion for their financial support for this analysis. PERC also extends our gratitude to TransUnion, Acxiom, and Experian for supplying the underlying data used in this analysis, as well as programming and run time, and access to analytical services experts.

In addition, we would like to thank a number of people who shaped our thinking and influenced our approach to examining this topic—directly and indirectly—including Carol Wayman, Michael Nathans, Joseph Duncan, and Birny Birnbaum.

Helpful comments and edits were also received from Whytney Pickens.

Ultimately, however, despite the contributions of those referenced above, the views and opinions presented in this study are exclusively those of the authors.

# Table of Contents

Acknowledgments 4

Key Findings 6

1. Introduction 7

2. Methodology and Data 8

3. Results 9

3.1 Change in Credit Scores with Alternative Data 10

3.2 Impact on Credit Score Distributions 13

3.3 Impact on Credit Scoring Models 14

3.4 Impact on Access to Credit 15

3.5 Socio-Demographic Impacts 16

Household Income 17

Age and Home Ownership 18

Race/Ethnicity 18

3.6 Impacts on Consumers with Prior Derogatories 19

3.7 Impacts from Negative-Only Alternative Data 20

3.8 Changes in Scores Over Time 20

3.9 Scores 22

4. Conclusion 23

## Key Findings

Currently, most non-financial services, such as energy utilities and telecommunications services, only report negative data, such as very late payments, charge-offs and collections, to Consumer Reporting Agencies (CRAs), also known as credit bureaus. This occurs either directly or indirectly through collection agencies. On the other hand, *fully* reported accounts are those in which positive data, such as on-time payments, mildly late payments, and balances, are also reported. This report assesses the impacts of including fully reported non-financial payment data—specifically energy utility and telephone payment data—in consumer credit files. These non-financial payment data are referred to as *alternative data* since they are not typically *fully* reported to CRAs. This report uses data during and after the Great Recession and compares results to a similar analysis using data from 2005 and 2006, prior to the economic downturn. Key findings from this research include:

### ► **Massive Material Impacts for the Financially Excluded:**

Among the so-called “thin-file population” including fully reported alternative payment data dramatically increases credit standings. In this group, 25% experienced an upward score tier migration (moved from a higher risk tier to a lower risk tier) as a direct consequence of fully reported alternative data being in their credit file. By contrast, 6% of the thin-file population experienced a downward score tier migration (from a lower risk tier to a higher risk tier). Including in this group those who become scoreable when alternative data is added, assuming that not having a score is viewed as very high risk, then 64% experience a score tier rise and 1% experience a score tier fall.

► **Score Impacts Stable Over Time:** Comparing the 2005 (pre-Great Recession) results with the 2009 (post-Great Recession) results is telling. Those whose scores improved with the inclusion of alternative payment

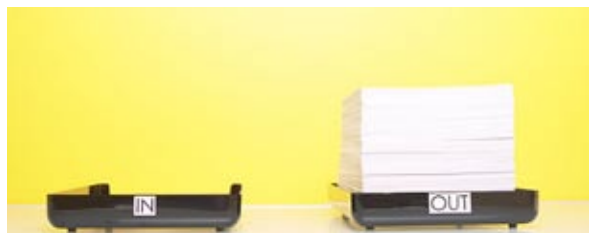
data increased by 4% (from 28% to 29%), those whose scores were unchanged increased by 10% (from 44% to 48%), and those whose scores lowered declined by 19% (from 16% to 13%). This strongly suggests that the general impacts from adding alternative data to credit scores are robust to macroeconomic conditions.

### ► **Credit Underserved Primary**

**Beneficiaries of Alternative Data:** As was the case when PERC examined credit reports from 2005, the largest net beneficiaries in terms of improved credit access are lower income Americans, members of minority communities, and younger and elderly Americans. For instance, those earning less than \$20k annually saw a 21% increase in acceptance rates, those earning between \$20k and \$30k saw a 14% increase, and those earning between \$30k and \$50k saw a 10% increase. The rate of increase was 14% for Blacks; 15% for those 18-25; and 11% for those above 66 years of age. We also see dramatic differences between renters and homeowners, with renters experiencing a 17% increase in credit access when alternative data are fully reported versus just a 7% increase for homeowners.

### ► **Those with past Serious Delinquencies**

**Benefit from Alternative Data:** Consumers with a public record including a bankruptcy and/or very late payments (90+ days late) among the traditional accounts reported to CRAs, witnessed more score increases than decreases (55% versus 30%) when alternative data were included in their credit files. This suggests that those with blemished credit files also benefit from the addition of alternative data. This may be particularly useful since those with blemished credit may otherwise find it difficult to access mainstream credit to improve their credit standing.



## 1. Introduction

In 2006, PERC published its first empirical study quantifying the credit impacts from including fully reported non-financial payment data in consumer credit reports. These fully reported payment data were referred to as alternative data since payments for non-financial services, such as for energy utility or telecommunication services are not typically fully reported to CRAs (also commonly referred to as credit bureaus). That is, on-time payments are not reported but severe delinquencies, such as very late payments, charge-offs, or collections are reported, either directly or indirectly via collection agencies. This is what is referred to as negative-only reporting. And it is still the case that non-financial data remains *alternative* data, that is, it is still not typically fully reported to CRAs. The initial PERC report was published during a much rosier economic period. Since then, the real estate market in the U.S. (and in many other large, advanced countries) has collapsed. The financial services sector was on the brink of implosion and likely would have also collapsed if not for coordinated government bailouts. A mild recession that began in December 2007 became a Great Recession and unemployment hit double digits in the U.S. The ensuing recovery has been noted for the historically slow rebound in the job market, and uncertainty continues to pervade throughout most advanced economies.

These macro-economic developments have had two primary impacts on the use of alternative data in consumer credit underwriting. First, and most importantly, it crowded the issue off of the agenda for many lenders, supportive lawmakers, and regulators. During

the period spanning 2005 to the end of 2007, alternative data was a hot topic in both industry and policy circles. Congress held hearings on the topic and draft legislation was circulated. PERC was invited to present to many oversight agencies including to then FDIC Chair, Sheila Bair, and hardly a financial services or energy utility industry event occurred without a panel discussion on the topic—PERC was presenting at least once a week on average during this time. Promising new solutions—including FICO’s Expansion Score, Link-2-Credit (L2C), Payment Reporting Builds Credit (PRBC), and VantageScore Solutions with its VantageScore® model—were addressing the growing interest in alternative data.

The second consequence of the subprime meltdown and Great Recession for alternative data was that skeptics now could argue—and have argued—that prior studies no longer mattered as the world had changed in important ways. Those opposed to using alternative data to build credit histories have asserted that the positive results from earlier studies were either skewed owing to distortions in the financial services markets during the period from which the samples were drawn or not relevant today, following a downturn in the business cycle, with increased delinquencies.

PERC agrees that the macroeconomy is an important variable worthy of consideration, and agrees with the hypothesis that the financial sector meltdown and the Great Recession could have consequences on the value—to the individual consumer, to the financial services sector, and to the economy—of including non-financial payment data in consumer credit reports. As such, this report reexamines the impact of alternative data by using data from 2009 and 2010, well after the beginning of the Great Recession and the financial crisis, during a period when delinquencies and unemployment spiked. The general results of this study relative to those of the similar 2006 study enable us to determine whether there have been qualitative and meaningful changes to the impacts of alternative data.





## 2. Methodology and Data

Similar to the analysis in the 2006 PERC and Brookings Institution report *Give Credit Where Credit is Due*, all of the credit files with one or more energy utility and telecommunications accounts fully reported are used from the participating CRAs<sup>1</sup>. These energy utility and telecommunications payment data are referred to as alternative data since such payments are not typically fully reported to CRAs. Unlike the 2006 analysis, we do not break out findings by whether the alternative data are utility or telecommunications accounts. This was done so as not to isolate and report information from a single data furnisher (at least one of the CRAs participating had a single telecommunications provider supplying alternative data). As such, this report shows results for all alternative data as opposed to specific utility or telecom account results. And, unlike the 2006 analysis in which only TransUnion provided data, two CRAs are participating in the current study, TransUnion and Experian.

Each CRA provided data on over four million credit files that had at least one alternative account fully reported. By fully reported we mean that both on-time payments and late payments are reported. The purpose of PERC's Alternative Data Initiative is not simply to advocate that utilities and telecommunications companies just report data, but that they *fully* report data.

Currently, as it was in 2006, the majority of energy utilities and telecommunications companies do not fully report customer payment data to the CRAs. But a PERC survey of utilities, telecommunications companies, and other potential alternative data providers found that 89% referred delinquencies and defaults to collections agencies, with most respondents aware that these accounts would then be reported to the CRAs<sup>2</sup>. So, negative information, such as very late payments, is typically reported, but not positive information, such as on-time payments. Not all alternative data furnishers that fully report to a CRA report to both TransUnion and Experian. While some do, some do not. So, there is overlap in and differences between the customers. To maintain privacy, only anonymous credit records were shared with PERC, so we were not able to see which customers were shared between the two CRAs. Before the credit file data was sent to PERC, it was first appended with socio-demographic data from the Acxiom Corporation, such as race and household income, since this information is not included in credit files. This enabled segmentation analysis.

Given slight differences in system architectures and a modified data request, not all calculations contain data from both CRAs. For instance, one CRA provided data on negative-only reported alternative accounts (in addition to the accounts with fully reported alternative data) and the other provided data on separate counts of traditional accounts and alternative accounts. However, most calculations contain data from both CRAs. For these results, the average of the individual results obtained from the two CRAs are presented. So, if one CRA shows an unscorable rate of 7% without alternative data and the other CRA shows a rate of 9%, we will report the average of 8%. Individual level CRA data are not presented so as not to make this a comparison between the two CRAs; although, qualitatively, there

<sup>1</sup> Michael Turner et al., *Give Credit Where Credit Is Due* (Washington, DC: Brookings Institution, December 2006).

<sup>2</sup> Michael Turner et al., *Credit Reporting Customer Payment Data* (Chapel Hill, NC: PERC, March 2009).



was little difference between the two. Throughout this report, we tend to compare the results from the 2009/2010 data with those from the utility sample from the 2005/2006 data; this is done since most of the alternative data examined in the 2009/2010 contain alternative accounts that are utility payment data. We also show results for the thin-file population, this is defined as consumers with credit files with fewer than three accounts, also known as tradelines, reported.

Comparing the total number of tradelines between the 2005 utility sample and the 2009 sample shows that the 2009 files are thicker (have more tradelines). This is shown in table 1. This could be due either to the thickening of credit files over this period or due to differences between the samples. Neither sample is a representative sample of the total CRA databases; both are samples of credit files that have alternative data. Over this period some alternative data providers have stopped reporting and some have begun reporting. As such, small changes between the results from the two samples should not be over-interpreted, as they may be the result of changing populations with alternative data.

**Table 1: Total Number of Tradelines, 2005 and 2009 Samples**

Number of Tradelines	2009 Alternative Data Sample	2005 Utility Data Sample
1	3.9	7.7
2	2.5	4.1
3	2.4	3.5
4	2.5	3.2
5	2.7	3.1
6	2.8	3.1
7+	83.3	75.2

Tradelines includes both alternative and traditional tradelines. Every record has at least one alternative data tradeline.

We instead aim to focus on the basic, qualitative findings, and the changes therein. We are most interested in the very basic aspects of the updated data such as the following: Does alternative data still result in increased credit access? Does it still improve credit model performance? Does it still have a larger and more positive impact on members of lower-income households?

The timing across the two CRAs was not identical. TransUnion provided files covering the period of July 2009 to July 2010 and Experian covered the period of September 2009 to September 2010. Both CRAs were able to provide VantageScore® credit scores calculated with and without the alternative data in 2009 and then included a performance measure in 2010 (number of 90+ DPD delinquent accounts over the previous 12 months).<sup>3</sup>

In addition to the alternative data samples, both CRAs provided validation samples of about 1 million records from credit files that did not contain alternative data.

### 3. Results

In the initial 2006 analysis it was found that the inclusion of the alternative data directly impacts consumers in three ways. First, the addition of the data can change the credit scores of consumers who were already scoreable (without the alternative data). Second, it allows some consumers who were unscorable to become scoreable (either by gaining a credit file or by adding a needed account). Third, the additional data improves the risk rank ordering of consumers who were already scoreable without the alternative data. All three of these components resulted in increased access to credit. In what follows, the first two impacts, changes in credit scores and becoming scoreable, are examined.

<sup>3</sup>The same version of the credit score was used in the 2005/2006 analysis and the current analysis, namely the VantageScore 1.0.

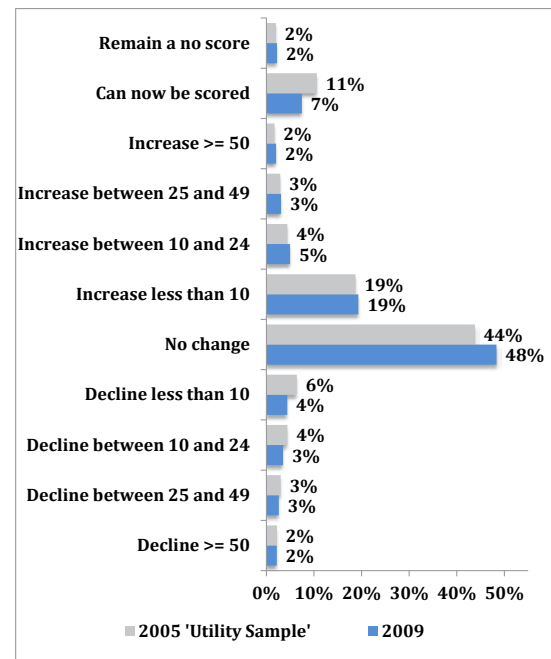
### 3.1 Change in Credit Scores with Alternative Data

The credit score impacts of the inclusion of alternative data are determined by simply obtaining a VantageScore credit score of a consumer’s credit file with the alternative data and then obtaining a VantageScore credit score of a consumer’s credit file with the alternative data removed. The two VantageScore credit scores are then compared and the change in score is recorded.

In figure 1 changes in the VantageScore credit score are shown for the 2005 TransUnion sample of files with energy utility alternative data and for the 2009 alternative data sample. The results reported for the 2009 sample are simple averages of the corresponding individual distributions from TransUnion and Experian. It is worth noting that comparing TransUnion’s and Experian’s distributions separately revealed that they were virtually identical. The largest percentage point difference found among the categories shown in figure 1 was under 3%. This is reassuring for a number of reasons. First it suggests that findings from the initial PERC research were not just some result of a particular way one CRA handled alternative data. Second, while both CRAs do have some overlap among their alternative data providers, they are not identical. As such, the results of adding alternative data do not appear to be unduly vulnerable to different population samples. In fact, there is a bigger difference between TransUnion 2005 and TransUnion 2009 than *between* the CRA’s in 2009. This may be a result of changes in the levels and depth of coverage of “traditional” data, other changes in TransUnion’s database, changes in the types of files with alternative data, and dramatically changed macroeconomic conditions. Essentially, the differences between 2005 and 2009 may be a result of everything that may have changed between March 2005 and July 2009 for consumers and TransUnion’s data.

What is perhaps most remarkable about the results presented in figure 1 is how stable the distributions of score changes remained over time. The three most noticeable differences are 1) a decline in the share that become scoreable with alternative data, 2) an increase in the share that have no score change, and 3) a rise in the share that have score increases and a decline in the share that have score decreases.

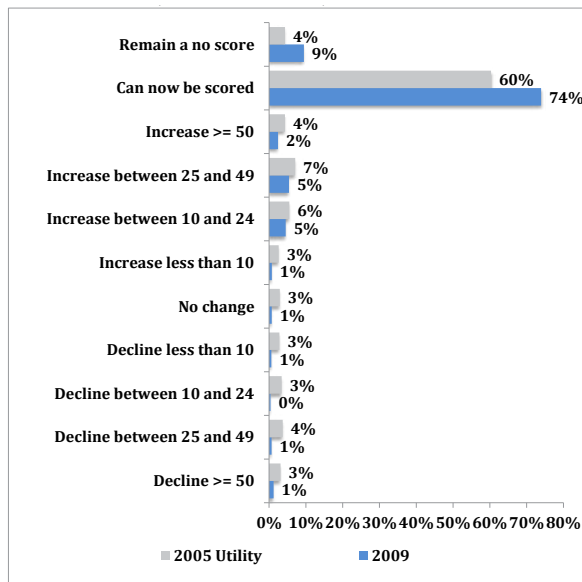
**Figure 1: Change in VantageScore Credit Score with Inclusion of Alternative Data**



As was found with the initial results, most consumers would see little to no change in their credit scores with the inclusion of alternative data; of those that do see a change, more would see increases than decreases. That so many see little or no change is not surprising since most consumers are scoreable and have many accounts reported to the CRAs. So, for most, including a utility or telecom account in their credit files would have little to no impact. Nevertheless, some consumers do become scoreable when alternative data are added.

It is among those with no accounts or only one or two accounts reported to the CRAs that the inclusion of alternative data such as utility and telecom account data should have the greatest impact. Figure 2 presents score changes for these thin-file consumers and does show the increased impact.

**Figure 2: Change in VantageScore Credit Score with Inclusion of Alternative Data, Thin File Consumers (Fewer than 3 tradelines)**



The results in Figure 2 reflect and amplify, in some ways, the findings among the broader population. Compared with the 2005 data, fewer consumers are becoming scoreable, more are seeing score increases, and fewer are seeing score decreases. The differences seen between the TransUnion distribution and the Experian distribution (in the thin-file population) were greater than seen in the broader population. The biggest difference between the same score change categories between the two CRAs is six percentage points. Qualitatively, however, the two are very similar, with both showing the same general

pattern shown in the average of the two (such as far more score increase than decreases). As before, the differences between the CRAs in 2009 are less than the differences between 2005 and 2009. Again, this suggests that one should not expect a particularly unique TransUnion or Experian impact from adding alternative data.

Thin-file consumers appear to be much more likely to have score increases than decreases relative to the findings from 2005. Specifically, on average, there were over three times as many consumers with score increases as there were with score decreases. On the other hand, in 2005 there were only one-and-a-half times as many score increases as decreases among the thin-file population.

What makes these results and their relative stability so interesting is that the 2009 data is for July from TransUnion and for September from Experian. So, these data are over a year and a half after the beginning of the recession and, on average, about a year after the beginning of the financial crisis.<sup>4</sup>

Furthermore, from Experian we were also able to obtain the score changes with and without alternative data for September 2010 for the same group of consumer in 2009 data. The score change distributions for 2009 and 2010 were virtually indistinguishable. The largest change was that the no change category rose by 0.65%. The score decline categories fell a little in 2010 and the score rise categories increased somewhat. So, if anything, the score change distribution became a little more beneficial for consumers.

Taken together, the March 2005 and the July 2009 data from TransUnion, the September 2009 and September 2010 data from Experian, and the individual case studies presented by the CRAs and PERC, suggests that the pattern seen in figure 1 is generally what one should expect from the inclusion of alternative data, regardless of macroeconomic conditions.<sup>5</sup>

<sup>4</sup> According to the NBER, the official arbiter of these things, the recession began in December 2007 and lasted until June of 2009. The recovery since the end of the recession has been relatively slow. By financial crisis we mean the sharp economic downturn and the freezing of the credit markets that occurred following the September 15th, 2008 announcement that Lehman Brothers would file for bankruptcy. The financial crisis in the mortgage markets (particularly for subprime mortgages) began prior to December 2007.

**A New Pathway to Financial Inclusion: Alternative Data, Credit Building, and Responsible Lending in the Wake of the Great Recession**

By itself, these score changes should result in increased access to credit. But in addition to score changes, consumers are also becoming scoreable and the performance is improving in the credit scoring models when alternative data is added. These last two changes should also increase credit access.

Score changes alone overlook a key reality of credit markets that access to credit and the terms and price of credit do not typically change for every point change in credit scores. They typically change as consumers migrate from one credit tier or credit band to another. Given the large number of credit scores in the market place and the various tiers used by individual lenders, determining exactly how individuals would be impacted is not possible. On the other hand, by using a standard set of credit tiers reported for the VantageScore credit score, we will likely well approximate the immediate credit market impact of alternative data. This tier change analysis was not performed in *Give Credit where Credit is Due*, and so comparisons are not possible.

**Table 2: Change in Credit Tiers with Inclusion of Alternative Data**

	Rise one or more Tiers	No Change	Fall one or more Tiers
<b>Entire Sample</b>			
Including Unscoreables	9%	88%	3%
Excluding Unscoreables	4%	93%	3%
<b>Thin-file</b>			
Including Unscoreables	64%	35%	1%
Excluding Unscoreables	25%	69%	6%

The standard tiers that we will use are sometimes referred to as the ABC tiers. Just like with a school grading scale, the tiers are labeled with grades. The highest tier (lowest risk) is

labeled A and is between 900 and 990, a B is 800 to 899, a C is 700 to 799, and D is 600 to 699 and the lowest grade and highest risk is F, and ranges from 501 to 599.

To include the unscorable population we will assume that being unscorable is viewed by lenders as being in the lowest score tier, therefore, becoming scoreable but going in to the lowest score tier would be considered no change. Only movement to a tier above the lowest would be considered upward movement. It should be noted that consumers who are unscorable are, in reality, not monolithically high risk. This, in fact, is the point of PERC’s Alternative Data Initiative. But, if due to a lack of information they are *viewed* as high risk, then they will have reduced access to mainstream, affordable credit.

The tier change results are shown in table 2. Examining only those who were scoreable with *and* without the alternative data, that is excluding all of those who were unscorable with or without alternative data, we see four percent rise one or more tiers and three percent fall one or more tiers. This is broadly consistent with findings on score changes for the entire population. Most saw very little or no score change with slightly more seeing a sizable score rise than those that saw sizable score falls. Looking at the entire sample and including the unscorable population, eight percent of the sample would rise one or more credit tiers and three percent would fall one or more credit tiers when alternative data are added. These figures include one of the main impacts of adding alternative data, consumers moving from being unscorable to scoreable.

Focusing on the thin-file population, much larger changes are evident. Excluding the unscorable population, twenty-five percent see a one or more score tier rise and six percent see a fall when alternative data is added. Including those unscorable without alternative data results in 64% seeing a one or more score tier rise and 1% seeing a fall.

If instead of assuming that no score was equivalent to the lowest score tier, we instead assumed that it was equivalent to the second lowest score tier then going from unscorable

<sup>5</sup> See Michael Turner et al., Credit Reporting Customer Payment Data (Chapel Hill, NC: PERC, March 2009) for a case study example from DTE.

to the lowest score tier would actually be a score tier decline. This is likely an overly cautious assumption to make since feedback from lenders has been that those borrowers with no information (no credit scores or credit files) are typically considered to be of the highest risk. This is not necessarily because this group, overall, is high risk but because lenders need information to underwrite and without basic information the risk associated with the consumer is not known well. From the perspective of risk management, it is safer to classify an unknown risk as a high risk. Nonetheless, assuming the unscorable are in the D tier, for the entire sample, seven percent rise one or more tiers and five percent fall one or more tiers when alternative data are added. For the thin-file population, 44% rise tiers and 16% fall tiers.

For all those that became scoreable, about one-third scored in the F category, 22% scored in the D category and 45% scored in the C or higher category.

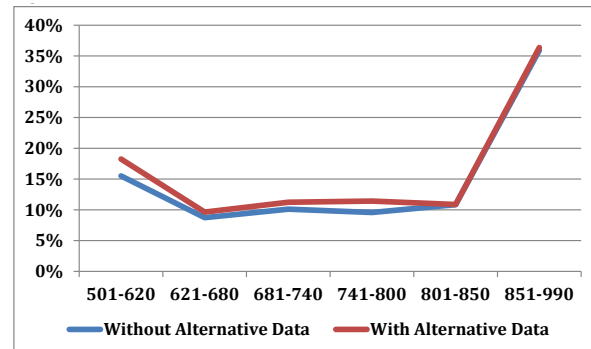
These results show that in the entire population and more particularly in the thin-file population there are more score tier increases when alternative data are added than score tier decreases. This is whether or not the unscorable population is included in the calculations and whether unscorable is classified and equivalent to a D tier or an F tier.

As mentioned previously, score tier changes (and score changes) shown thus far do not include the impact of improved model performance when the alternative data are added, it just indicates the score changes and score tier changes with the inclusion of alternative data.

### 3.2 Impact on Credit Score Distributions

Figure 3 shows the distribution of credit scores with and without the alternative data. The results shown are simple averages of the individual distributions from TransUnion and Experian. As with the 2005 results, it is not the case that the inclusion of alternative data only adds scores at the bottom of the distribution.

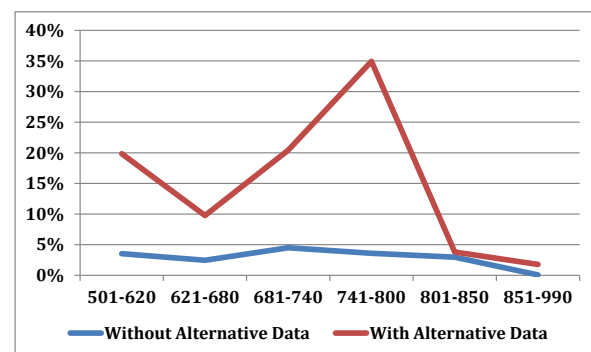
**Figure 3: Score Distribution with Inclusion of Alternative Data**



In fact, there is an increase in each category, including the highest score category (851-990) when the alternative data are added. Most of the additional scores, however, do occur below the highest category, with a good deal occurring in the middle of the distribution. This is qualitatively what was seen with the 2005 data. As before, this suggests that additional borrowers would be able to be granted mainstream credit as a result of the use of alternative data.

Figure 4 shows the corresponding results for the thin-file consumers. First, these results are also broadly consistent with those obtained in 2005. Second, they reflect the greater impact of alternative data on the thin file population.

**Figure 4: Score Distribution with Inclusion of Alternative Data, Thin-file Consumers**



It should be noted that a little over eighty percent of the thin-file population would not be scoreable without the alternative data, and so the blue line shown in figure 4 covers less than twenty percent of that population.<sup>6</sup> With the alternative data, around ninety percent have a score, thus the red line (with alternative data) is well above the blue line. Again, much of the ‘new’ scores are in the middle of the distribution, suggesting that many of the thin file population could gain access to mainstream credit with the inclusion of the alternative data.

### 3.3 Impact on Credit Scoring Models

In the previous sections it was shown how scores change, score distributions change, and the share of the sample that became scoreable when alternative data were included in credit files. In some ways, score changes alone can be a simplistic view of how including alternative data in lending decisions would ultimately impact borrowers. The reason for this is that if all scores were *arbitrarily* raised 30 points, so it was declared that everyone’s VantageScore or FICO credit scores were to be raised 30 point then ultimately, nothing has changed. That is, the risk rank ordering of individuals has not changed and, as a result, no increased access to credit would occur. Credit access is ultimately related to how well lenders can predict who is likely pay back a loan and who in likely not to pay it back. The actual

number of a credit score is only useful to the extent it helps lenders predict who are good and bad risks. This section, then, will help determine whether the score changes seen with the inclusion of alternative data were arbitrary or whether they improved risk assessment (and ultimately credit access).

The VantageScore model is used to create two sets of credit scores in 2009 (July for TransUnion and September for Exeprian), one set using all data including alternative data and one set with data that excludes the alternative data. The prediction of the model (the score) is then compared to actual credit outcomes over the following year (90+ day delinquencies). How well the scores predict these outcomes (rank orders the consumers from low risk to high risk) is the performance of the model. That is, the performance of the model is how well the model predicts reality.

One commonly used measure of credit score model performance is the Kolmogorov-Smirnov (K-S) test. This test returns a value between 0 and 1 that represents the maximum difference between the percent of total goods (low risks) and the percent of total bads (high risks) that are captured for different score cutoffs. Say the maximum occurs at a score of 700, where a score cutoff of 700 would mean that 85% of all the good risks would be accepted and only 20% of the bad risks. In this case the K-S would be  $.85 - .2 = .65$ . Essentially, this tells us how well a model separates goods from bads.

**Table 3:Realtive Change in VantageScore Performance with Alternative Data (Measured by K-S)**

	2009-2010		2005-2006	
	With Alternative Data	Without Alternative Data	With Alternative (Utility) Data	Without Alternative (Utility) Data
<b>Including Unscoreables</b>				
All Consumers	1.074	1.000	1.098	1.000
Thin File	3.489	1.000	3.294	1.000
<b>Excluding Unscoreables</b>				
All Consumers	1.024	1.000	1.022	1.000
Thin-file	1.157	1.000	1.078	1.000

<sup>6</sup> Specifically, from figure 2 we see that 74% become scoreable with alternative data and 9% remain unscorable, so the unscorable population shifts from 83% to 9% when alternative data is added.



In table 3, the actual K-S values are not given; instead the relative change with the inclusion of alternative data is provided. So, for all consumers (including unscorable) for 2009-1010, 1.074 means that if the K-S was .65 without alternative data then the score rose 7.4 percent with the alternative data to  $0.65 \times (1.074) = 0.698$ .

Table 3 shows values for relative K-S changes using the 2009/2010 data and the 2005/2006 utility data. It also shows results when the unscorable population is included and when it is not. When the unscorable population is included, those without scores are put at the bottom of the score distribution. That is they are viewed as very high risk and essentially excluded from credit that utilizes credit scores. This is useful since this tells us the impact of going from a no score to a score, where no score implies high risk. The second way of calculating the K-S is excluding all those that would be unscorable with or without the alternative data. This tells us specifically how much better scores predict outcomes when alternative data is added among those who already have a score.

The general magnitudes of the K-S and changes in K-S are consistent with the findings using the 2005/2006 data. For the entire sample, in which those without scores are included, the K-S rises, on average, 7.4% with the 2005/2006 data and rose 9.8% with the 2005/2006 data. This may be due to the fact that in the 2009/2010 data there are a smaller proportion of consumers without scores. On the other hand, the K-S rise for the entire sample excluding the unscorable population is slightly higher with the 2009/2010 data, 2.4% versus 2.2%.

When looking at the thin-file population there is an enormous change in the K-S when the unscorable population is included, this is due to the fact that the majority of the thin-file population is unscorable; so the addition of the scores (from adding alternative data) has an enormous impact on the K-S. That magnitude of the rise in the K-S for the thin-file population (excluding the unscorable population) is somewhat higher in

the 2009/2010 data, 15.7% versus 7.8%. The fact that this remains a healthy increase shows that there is a pretty substantial increase in model performance for the thin-file consumers who had scores without the alternative data. This increased model performance should translate to more accurate lending and increased access to lending for thin-file consumers who were scoreable without the alternative data.

This confirms that the changes in credit scores seen in the previous sections are reflective of improved model performance for the entire population and the thin-file population. So, the scores changed (rose and fell) to more accurately match risk and predict actual credit outcomes. The combination of net score increase and improved model performance should act to increase access to credit. This will be examined in the following section.



### 3.4 Impact on Access to Credit

Table 4 shows the average acceptance rate (across the two CRAs) for various delinquency rates (rates of occurrences of 90+ days late payments) when alternative data are included and excluded. So, to maintain a portfolio with a 3% delinquency rate, 58.7% of the population could be accepted if alternative data were used in the underwriting while only 53.7% could be accepted if alternative data were not used. This increase is due to two factors. First, with alternative data more people have scores and can be accepted. Second, with

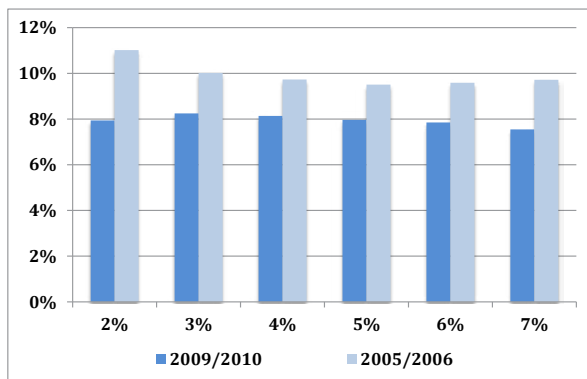
alternative data, the performance of the scoring model improves (for those who already had scores) and so better predictions of risk are made, resulting in more people being able to be accepted. So, some people are brought into the system and those already in are generally better rank ordered for risk.

**Table 4: Acceptance Rates with and without Alternative Data**

Average Acceptance Rate		
Target Portfolio Delinquency Rates	Including Alternative Data	Excluding Alternative Data
2%	48.8	45.2
3%	58.1	53.7
4%	63.7	58.9
5%	67.7	62.7
6%	71.0	65.9
7%	73.8	68.6

Figures shown are the averages for the two participating CRAs

**Figure 6: Percent Increase in Acceptance with Alternative Data Added**



The change from 53.7% to 58.1% represents an increase of a little over eight percent. So, portfolio size could increase by eight percent (at a 3% delinquency rate) when alternative data are used in underwriting. This eight percent is shown in figure 6 as well as figures for other target delinquency rates and the corresponding increase found in the 2005/2006 data. Generally, the percent increase in lending is fairly steady across different target delinquency rates. And while the typical increase in lending with alternative data were around 10% with the 2005/2006 it is around 8% with the 2009/2010 data.

Importantly, this increased potential lending does not result from easy lending or lowering of standards but from greater inclusion and improved risk assessment. Next, the increased lending (acceptance) will be broken out into various socio-demographic segments to determine which groups are benefiting from the increased access to credit.

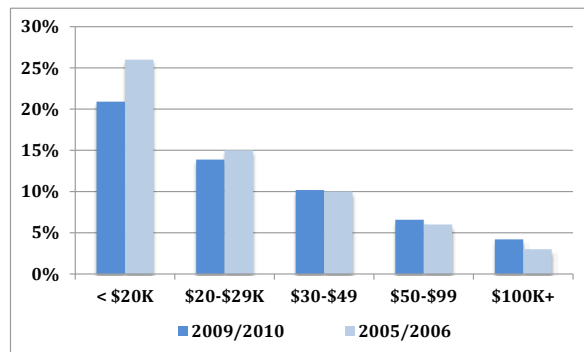
### 3.5 Socio-Demographic Impacts

In the last section it was shown that the potential overall increase in access to credit was an increase of portfolio size of about eight percent. So, overall, an eight percent increase in lending would be expected if alternative data are reported and lenders maintain the same target default rate. This change in acceptance with the inclusion of alternative data will now be segmented along a few key dimensions, such as household income and age. The segmentation analysis will assume a 3% target default rate and will simply show the percent change in the number of individuals accepted (above the respective score cutoff for the target default rate) with the inclusion of alternative data.

## Household Income

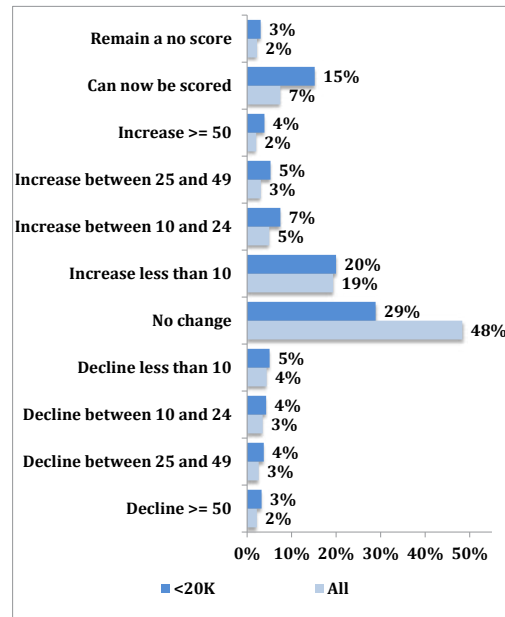
Perhaps the most important segmentation is household income. Figure 7 compares the segmentation results derived from the 2005/2006 data to those derived from the 2009/2010 data.

**Figure 7: Change in Acceptance by Household Income**



There appears to be a slight reduction in the gain for the lowest income groups and a slight rise in the gain for the highest income groups. But, fundamentally, the same pattern persists. Members of lower income households benefit much more from the use of alternative data than members of higher income households. This is not surprising since it is the case that members of lower income households make up a disproportionately large share of the credit underserved, specifically those consumers with no credit files or thin credit files. We see that while the inclusion of alternative data raises acceptance for members of the lowest income group by a little over 20% it raises it by less than 5% for members of the highest income group. The finding of the persistence of the pattern of outsized benefits to the lowest income consumers is an important finding for practical reasons.

**Figure 8: Change in Scores for the Lowest Income Group**



Since the release of *Give Credit where Credit is Due*, some consumer advocates have argued that lower income consumers are more likely to pay utility bills late and therefore more likely to be ‘hurt’ if utility payments are reported to CRAs. Setting aside the notion that hurt would mean that a person that is found to be paying bills late would have a reduced credit score and reduced access to credit, this overlooks two key points. First, the vast majority of lower income households pay their utility bills on time. Second, how the addition of alternative data impacts a credit score depends on the other information in a credit file. For instance, if late utility payments were added to a credit file that already contained other similar derogatories, then the late utility payments may have little impact on the credit score.

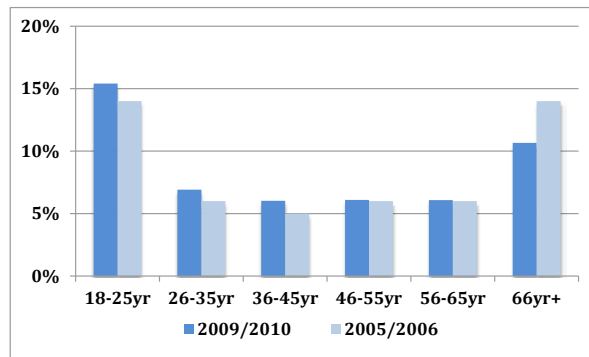
Figure 8 shows the change in credit scores when alternative data are added to the credit files for members of the lowest income households and for the entire sample. There are a few key differences. First, relative

to the entire sample, a greater share of members of the lowest household income group become scoreable and see credit score changes when the alternative data are added. Second, in this lowest income group, while 4% are more likely to see credit score declines relative to the whole sample, 7% are more likely to see credit score rises. For the lowest income group, 37% see credit score rises, 16% see credit score declines, 29% see no change, and 15% become scoreable.

### Age and Home Ownership

Credit access impact by age was found to have an unexpected pattern when the 2005/2006 data was initially analyzed. Specifically, while, as expected, the youngest consumers saw the greatest benefit from the data, surprisingly, those over 65 also saw a large impact. This is the pattern that is seen again. There are a number of possible explanations for this such as many older consumers may be thin-file since they may have their house and car paid off and do not use credit cards. Or, older couples may be more likely to have most of their bills and credit obligation in one person's name.

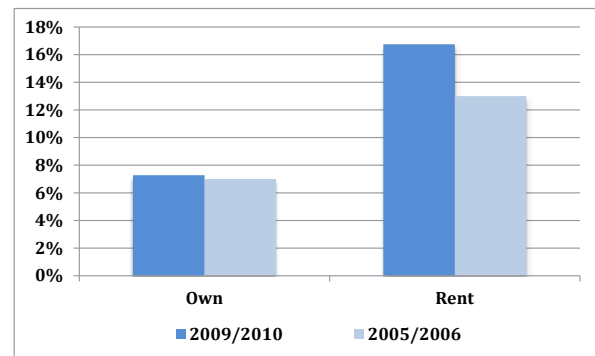
Figure 9: Change in Acceptance by Age



The fact that the youngest consumers witness a large increase in acceptance is unsurprising given the large share of this group that is thin file and new to credit. For this group, in particular, alternative data offers a way to establish a payment history without going in to debt.

Also seen in both the 2005/2006 and 2009/2010 data is that renters see a much larger increase in acceptance when alternative data are included in credit files than do homeowners. This may be of particular importance following the subprime mortgage meltdown and the financial crisis of 2008. Lending standards have been tightened, including minimum credit scores. As such, the use of alternative data has the potential to help individuals currently shut out of the more restrictive mortgage market via better risk assessment and not simply by easier lending.

Figure 10: Change in Acceptance by Home Ownership

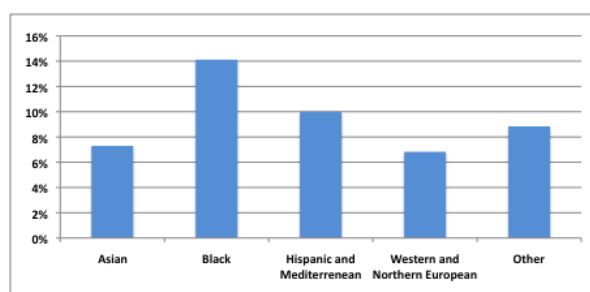


### Race/Ethnicity

Due to a different race/ethnicity variable provided with the 2009/2010 data then was provided with the 2005/2006 data, direct comparison is not possible. However, as with the other socio-demographic segmentation results, the general pattern remains the same.

Specifically, ethnic and racial minorities witness larger increases in acceptance relative to white consumers in the 2005/2006 data and consumer of Western and Northern European ancestry in the 2009/2010 data. In the 2005/2006 data, for instance, black consumers saw an increase in acceptance about two-and a half times that seen for white consumers. In the 2009/2010 data, black consumers see an increase in acceptance about twice that seen by consumers of Western and Northern European ancestry.

**Figure 11: Change in Acceptance by Race/Ethnicity**



As with income, age, and home ownership, the general pattern persists with ethnicity/race, but specific magnitudes and magnitudes of the differences between groups are different.

### 3.6 Impacts on Consumers with Prior Derogatories

For those with no credit score or thin credit files, the value of alternative data is clear. Furthermore, for those who do not want to have multiple lines of credit, including credit cards, the appeal of using non-financial payments as a way of demonstrating a good payment history is also clear—build credit without going into debt. But alternative data may also be helpful in another key way; it may be helpful (on average) to those who have had past credit problems. A person with a past bankruptcy or other serious derogatories may find it

difficult to open new lines of credit, however they likely do pay for many services that are not reported to credit bureaus, such as for electric, gas, water, mobile phone, internet, and other similar services. To test whether this group would be impacted by the inclusion of alternative data (and how), we examined consumers (from one of the CRAs that had the needed data) that had any public record, including bankruptcies or 90+ DPD reported on a traditional account. For this group we calculated the score changes when alternative data were added to their credit files. Table 5 shows that very few of those with past derogatories were unscorable and that most, 86%, had score changes. 55.8% saw score increases and 30.2% saw score decreases.

**Table 5: Score Changes with Alternative Data, Among Consumers with a Previous Bureau Derogatory Reported (includes public records and 90+ DPD on accounts)**

Score Changes		
Decrease		Increase
1.5%	≥ 50 pt	4.0%
3.7%	25-49 pt	6.7%
11.0%	10-24 pt	11.7%
14.1%	< 10 pt	33.5%
<b>30.2%</b>	<b>Total</b>	<b>55.8%</b>
No Change	11.8%	
Can Now be Scored	1.1%	
Remain a "No Score"	1.1%	

There are clearly some net benefits to this group, and since the majority of the sample contains only one alternative tradeline, it may be the case that adding two or three tradelines (gas, electric, and phone for instance) would extend these net benefits. On the other hand, alternative data is certainly not a magic bullet for rehabilitating credit for those with past credit problems. Since this group has much lower than average credit scores, a relatively small share would be accepted

for credit at a 3% target default rate with and without alternative data. Specifically, 3.8% would be accepted without alternative data and 4.9% would be accepted with alternative data. Thus, while the levels are low, acceptance rises by about 30% for this group. At a higher target default rate of 5%, which might be near prime but certainly above deep sub prime, 9.4% would be accepted without alternative data and 11.5% would be accepted with alternative data, which is a rise of 22.5%.

### 3.7 Impacts from Negative-Only Alternative Data

Perversely for consumers, some data furnishers only report negative or derogatory information to the CRAs. In some cases this may make sense. For instance, a very small lawn care company may not be able to report monthly on their customers to CRAs but they may decide to report customers that are over 90 days past due to a collections agency that would then report the delinquency to a CRA. Unfortunately, some utilities report negative-only data directly to CRAs.

To gauge the credit score impact of the reporting of negative-only data, we use a sample of credit files with negative data reported by utilities that do not report positive data. Hence, all of the files have some negative reported.

One interesting result is that some scores increase. This has been seen previously. A derogatory may enter a file, but if it is not too bad and there are other derogatories already present then the derogatory information alone may not lower the score. But, the record may increase the score perhaps by making the file thicker and giving the file a longer credit history. This being said, the results presented in table 6 clearly show that on average, the negative-only reporting of alternative data (such as utility data) lowers credit scores.

**Table 6: Score Changes with Negative Alternative Data**

Score Changes		
Decrease		Increase
6.5%	≥ 50 pt	0.3%
7.3%	25-49 pt	1.1%
9.0%	10-24 pt	2.0%
19.7%	< 10 pt	9.9%
<b>42.5%</b>	<b>Total</b>	<b>13.3%</b>
<b>No Change</b>	28.2%	
<b>Can Now be Scored</b>	10.9%	
<b>Remain a "No Score"</b>	5.1%	

All of those that did not have a derogatory reported would have been unchanged, this includes many who had paid on time and may have had a credit score increase. Hence, it can be concluded that not reporting alternative data at all leaves credit scores unchanged, negative-only reporting tends to lower scores, and full-file alternative data reporting tends to raise scores, and raise scores disproportionately for thin file consumers and member of low-income households.

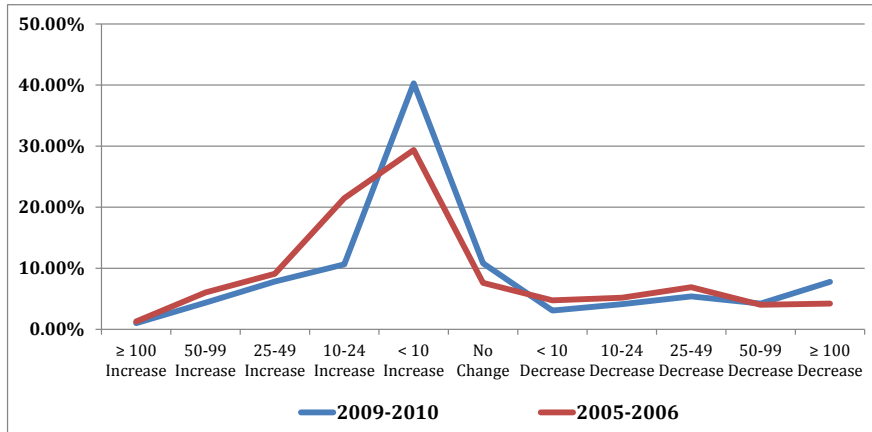
### 3.8 Changes in Scores Over Time

In a follow up to *Give Credit Where Credit is Due*, PERC examined changes in credit scores over time for those with only alternative data in their credit file.<sup>7</sup> This was to examine whether there were longer-term negative impacts from gaining access to credit from ‘only’ alternative data. No long-term score declines were found; on the contrary evidence of the opposite was found. One way this was examined was to compare score changes over a year for those with only alternative data to score change for those with thicker credit files.

<sup>7</sup> Michael Turner et al., *You Score You Win* (Chapel Hill, NC: PERC, 2008)



**Figure 12: Change in Scores over a Year for those with Only Alternative Accounts Reported**

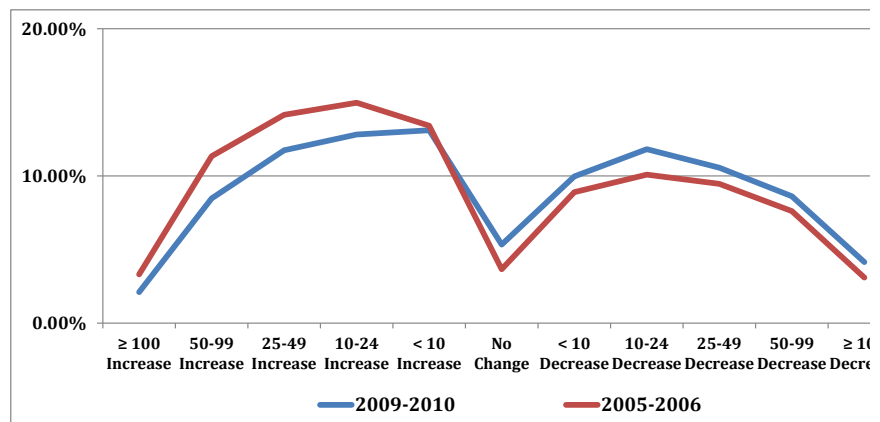


As can be seen from figure 12, the pattern of score changes over a year for those with only alternative data are very similar for the 2009-2010 period and the 2005-2006 period. There is no obvious large movement toward score declines. In both cases the majority see score increases.

Taken together, these results suggest that there is no apparent change to the earlier findings that gaining access to mainstream credit from alternative data has longer term negative credit and credit score consequences.

On the other hand, for the more general population, shown in figure 13, there is a distinct movement to more score declines over the yearlong period than score increases. These would be for the more typical credit files for which the addition of alternative data would have little impact.

**Figure 13: Change in Scores over a Year for those with Four or More Accounts Reported**



## 3.9 Credit Scores and Alternative Data Reporting

In the 2006 analysis, in addition to the VantageScore credit score, other credit scores were also examined, two bankruptcy scores and the TransUnion New Account score. While we found that the inclusion of alternative data increased credit access across these scores, the magnitudes varied. This is not surprising since risk scores are typically correlated to a high degree but not identical. For instance, it is not likely that a person would be super prime with a FICO credit score and subprime with a VantageScore credit score. But it would also be surprising if *everyone* that was in the 85th percentile of one score distributions were in the same percentile of the other score distribution.

However, it is the case that the credit scores examined are not optimized for an environment where non-financial data is widely reported. What this means is that the improved risk assessment observed with alternative data should likely improve as more alternative data is reported. For instance, some scoring models may view a newly opened utility account the same way it views a newly opened credit account. This may not be optimal. In the current environment it may not make sense to optimize models for data that only 2% of consumers may have. As this share increases, models will begin to distinguish between the different types of accounts (to the extent that it improves model performance). Any such optimization, by definition, would only improve model performance when alternative data is added. This improved performance would result in even better risk assessment and increased credit access from what is found in this report. That being said, it is not clear whether this increased performance would be slight or substantial.

Whether or not such optimizations would result in very large improvements it is likely a sound idea for analytics firms to begin accounting for different types of data that will be entering consumer credit files. Interested

parties, such as lenders and consumer advocates, should also request that such firms take this data into specific account to make it a greater priority for them.

In addition to ensuring that lenders use alternative data and in ways that are most beneficial to consumers and the financially underserved, consumer friendly choices can also be made in the ways alternative payment data are reported to the CRAs. These include not reporting small unpaid balances on accounts that are closed; not indicating that a customer is subsidized, on a payment plan or in forbearance; only reporting payments over 60 days overdue as late; not reporting retrospective data when the furnisher first begins reporting; and clearly communicating with customers that their payments will be fully reported to CRAs.

When delinquencies have to be reported to CRAs is often misunderstood. Although industry-reporting standards and guidelines exist, for example Metro 2, and the CRAs encourage their customers to follow the industry reporting guidelines, data furnishers are necessarily able to exercise some discretion as to the thresholds they use in reporting delinquencies. For example, one of the participating CRAs noted that even a major card issuer reported to the CRA only after their customers were over 60 days late. This was seen as an acceptable practice. And among the utilities that reported to that CRA, 25% started reporting delinquencies at 30+ days, 67% started at 60+ days, and 9% reported at 90+ days. Clearly, utilities do not need to report to CRAs in a rigid, inflexible way.

This flexibility suggests that many of the benefits of full file reporting by utilities and other potential alternative data providers can accrue to consumers without reporting in anything close to a harsh or strict way. Full file reporting by utilities is now no longer new or novel for the participating CRAs. Each has many years of experience with utilities and can advise them on these flexibilities and other reporting details important for alternative data furnishers.

## 4. Conclusion

The preponderance of evidence establishing the value proposition of alternative data is overwhelming and incontrovertible. Borrowers benefit from improved access to mainstream credit, enabling them to create wealth and improve their life's chances. Lenders benefit from being able to better assess risk and thereby profitably and soundly extend credit to segments previously viewed as too risky. Society benefits from a fairer and more inclusive financial services system, while the economy benefits from greater lending to the private sector that is sustained over time.

Importantly, credit score changes resulting from the inclusion of alternative data are found to result in better predictions of who will have a future serious delinquency. So, the score changes are not arbitrarily harming or benefiting consumers in some unfair manner, the score changes improve risk assessment. Following the financial crisis and mortgage crisis it should be obvious to everyone that properly identifying a borrower's risk is not a harm, even if they are found to be high risk. The true harm to society and individuals results from lending mistakes, misidentifying low-risk borrowers and high-risk borrowers.

A key aspect to the reporting of alternative data is that individuals can establish payment histories without having to go into debt. If rental, energy utility, telecom, and other alternative payments were reported, an individual would not need to establish and maintain credit card accounts, for instance, to build their credit and payment record. Without such options, credit cards and other forms of credit will remain the primary way people build or rebuild their credit history prior to needing credit for a car or home.

This alternative data solution also has the added benefits of being costless to the government (a proverbial "pen stroke solution"), scalable, and rapidly implementable with immediate impacts (6 to 12 month after tradelines are reported, credit standing can change dramatically for the underserved). It doesn't require building a complex new system, and instead serves to enhance a well established and universally used network of credit information sharing.

Despite these demonstrated benefits and practical advantages, some skeptics will continue to oppose including fully reported non-financial payment data in consumer credit reports. They will oppose this solution as long as some consumers have their scores reduced, even if they are reduced because the consumer had delinquencies and was found to be a higher risk. This opposition seems more emotional than rational given the facts. It also ignores the following:

- ▶ Including alternative payment data is an effective method to build or restore/rebuild a good credit history. If these data are not reported, or are only reported when accounts are seriously late or in collections, then most thin-file/no-file persons will remain financially excluded as they are trapped in the Credit Catch 22—lacking a prior credit history lenders will not extend credit.
- ▶ Low-score is always better than no score—Unscoreable persons are almost always rejected by mainstream lenders when applying for credit. They must, overwhelmingly, have their credit needs fulfilled by high cost lenders like pawn shops and pay day lenders. By virtue of being scoreable, a person greatly improves their chances of entering the credit mainstream. By meeting non-financial payment obligations, credit standing can rapidly improve. Fully reported data makes for the fairest and most forgiving system.
- ▶ More consumers in general, and low-income and thin-file consumers, in particular, benefit when alternative data is fully reported than when it is not. This is true whether the metric is credit score changes, credit score tier changes, or changes in portfolio acceptance given a target default rate.

Given the wide ranging net benefits shown in this report from making credit files *more inclusive* of consumers' payment behaviors, we believe focus should squarely shift to bringing more alternative data online. That is, focus should turn to making the financial system *more inclusive* by making credit files *more inclusive*. Importantly, the data should be brought online in consumer friendly ways. And we would hope that groups who have in the past opposed the full file reporting of energy utility, telecom, or other types of alternative data outright over specific concerns would, instead, shift their focus to finding solutions to their specific concerns and work to make the reporting of alternative data more pervasive, consumer friendly, and of upmost benefit to the financially underserved.



**PERC**

302 E. Pettigrew St., Suite 130  
Durham, NC 27701

p: +1 (919) 338-2798

f: +1 (919) 640-8881

[www.perc.net](http://www.perc.net)