LexisNexis®
RISK SOLUTIONS

WHITE PAPER

Statistical Analysis of RiskView and Race

Tests of Differential Effect within LexisNexis® RiskView™ Score and Attributes

Introduction

The use of non-traditional, or "Alternative" Data in Fair Credit Reporting Act (FCRA) regulated credit decisions is a growing practice. Alternative Data refers to data that is not found on credit bureau reports but has bearing on a consumer's creditworthiness, credit standing, credit capacity, character, general reputation, personal characteristics or mode of living. Public records, such as property records, rent payments, household bill payments, education, and licensing information are common examples of Alternative Data.

LexisNexis® RiskView™ Solutions is a suite of FCRA-compliant credit risk scores and attributes based on Alternative Data that leverage hundreds of data sources, including property records, court records and education history. RiskView scores are empirically derived, demonstrably and statistically sound credit scores consistent with Equal Credit Opportunity Act (ECOA) requirements.

The US adult population, defined as 18 years or older, is over 245 million people[1], and RiskView can score over 229 million of these consumers, including 40M of the roughly 50M consumers that cannot be scored using traditional credit bureau data.[2]

RiskView offers comprehensive coverage that helps lenders assess consumers who are disenfranchised by the current system, including members of traditionally underserved minority groups.[3]

As lenders explore adopting RiskView, they may have questions about the nature of the RiskView solution, its impact on racial minorities and other protected classes, the extent to which it complies with existing regulations and the potential for consumer harm. To address these questions, LexisNexis® Risk Solutions conducted an analysis based on methods defined in the "Report to the Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit" (hereafter referred to as the Federal Reserve Study).[4] These methods tested statistical bias towards protected groups in traditional credit scores using credit bureau based tradeline-level data and found that credit scores did not contain prohibited bias towards protected classes.

1 https://www.census.gov/quickfacts/

2 For a more detailed summary of the Alternative Data housed by LexisNexis Risk Solutions and their use in RiskView Score, please see a LexisNexis Risk Solutions white paper entitled, "The Emergence of Alternative Data: Innovative Credit Risk Strategies to Drive Business Growth." This paper and others can be found at: http://insights.lexisnexis.com/creditrisk/resources/

3 Browdie, Brian (2015). *Can Alternative Data Determine a Borrower's Ability to Repay?* American Banker Online (http://www.americanbanker.com/news/consumer-finance/can-alternative-data-determine-a-borrowers-ability-to-repay-1072785-1.html?)

4 Robert Avery (2007), "Report to the Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit", Federal Reserve Board of Governors.

Statistical Analysis of RiskView and Race

There are other published methodologies for testing scoring models for statistical bias, however we found the Federal Reserve Study to be the most transparent and thorough. In addition, the Federal Reserve Study was published by a regulator and provided detailed methods and rationale for each analysis. Therefore, we adopted the Federal Reserve's approach and interpretations to conduct a similar and transparent analysis on the RiskView Score and variables, with slight adjustments to account for the differences between bureau data and Alternative Data.

In our research, we emphasize the same distinction that the Federal Reserve Study makes between two key concepts: *Disparate Impact and Differential Effect. Disparate Impact* is a legal term described in the Equal Credit Opportunity Act (ECOA) that states that credit applicants from protected classes should not be disproportionally adversely impacted by a lender's policies—in particular, approval decisions and product pricing.

Differential Effect, however, is a statistical term. It does not refer to the overall outcome of lender product assignment or pricing, but is more specific to the relationship between credit scores and the variables used in a strategy. According to the Federal Reserve Study, "A credit scoring model, or a credit characteristic used in the model is said to have a statistical differential effect based on a demographic characteristic – say, age – if the model's predictiveness or the credit characteristic's contribution stems, at least in part, from the fact that the score or credit characteristic serves as a proxy for age."[5] Credit tools lacking differential effect bias are necessary but not sufficient for a compliant credit campaign; ultimately the tools need to be fair and compliant to enable a lender to make a fair and compliant product decision using all the information at their disposal.

Our study's methodology is derived from tests used in the Federal Reserve Paper and adjusted to accommodate the differences between traditional credit bureau tradeline data and Alternative Data. Our research goal is not to replicate the Federal Reserve Study on LexisNexis Risk Solutions data, but rather to use the study as a guide and as a point of comparison for the results.

Our study tests differential effect in RiskView based on race but makes no legal or regulatory claims with regard to how a lender can use RiskView in a manner compliant with the FCRA and the ECOA. This research cannot function as a replacement for lender-initiated compliance tests, as studies related to ECOA compliance and Disparate Impact typically examine a more exhaustive set of criteria based on all factors and

5 Report to Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit, pg. O-8, footnote 4.

decisions implemented in credit underwriting campaigns.[6] In addition, this study does not constitute legal advice, nor does it endorse the implementation of any lender or issuer strategy with regard to protected classes.

Study Scope and Organization

The study examines RiskView's ability to evaluate a consumer's presence at a credit bureau, how RiskView scores and data correlate to consumer credit performance and race (as identified using a statistical approach) and to provide a case study that demonstrates how lenders can use both credit bureau data and RiskView data to improve outcomes for consumers and lenders.

LexisNexis Risk Solutions conducted four sets of analyses, based on the Federal Reserve Study methods:

ANALYSIS 1: Consumer Scorable Rates, Consumer Default Rates and Score Rank-Ordering by Racial Group

ANALYSIS 2: Test of Differential Effect in RiskView Variables

ANALYSIS 3: Test of Differential Effect in RiskView Scores

ANALYSIS 4: Case study in consumer outcomes using credit bureau data and Alternative Data

Overview of the Sample

We sampled from the entire consumer-level LexisNexis Risk Solutions database to create a sample of 2 million consumers stratified by Consumer Finance Protection Bureau (CFPB) race proxy and bureau experience proxy. A stratified sample means that conditions that are less common in the database (e.g., underbanked consumers) were oversampled to ensure a robust sample for analysis. The primary intention is to examine the effect of RiskView on thin-file, no-file and minority borrowers, and these groups are oversampled accordingly.

6 This study is a statistical examination of RiskView as it relates to race. It demonstrates that RiskView scores and attributes can be used responsibly as a decision-key to set strategy and/or include in other analytic decision-making tools. This analysis focuses on group-level differences, such as average score and average score shift. All tests were aggregated across all members of a class, versus examining impacts for each specific member of the class.

Statistical Analysis of RiskView and Race

Sample Design and Proxy Methodologies

Race is not collected or stored by LexisNexis Risk Solutions in the natural course of product development or validation for RiskView. For purposes of this study, race was specially appended using logic provided by the CFPB in, "a Bayesian Improved Surname Geocoding (BISG) proxy method, which combines geography- and surname-based information into a single proxy probability for race and ethnicity."[7] BISG has well documented limitations, however the CFPB identifies this method as a best practice for conducting analysis as well as regulatory exams. Therefore, we are adopting this method for our tests but do provide some analysis of BISG compared to known race below.

The sample was first created from LexisNexis Risk Solutions' records and then matched to credit bureau trade-line files at two time points: April 2013 (Time1) and October 2014 (Time2). Several groups were sampled: (1) a No-Hit/Bureau Unscorable group that was not on the Credit Bureau file in April 2013, but subsequently was on the Credit Bureau file in October 2014; (2) a Thin-File group that was on the Credit Bureau file in April 2013 and October 2014, but had only been present at the Credit Bureau for less than a year in April 2013; and (3) a Thick-File group was present at the Credit Bureau for over 1 year by April 2013.

Racial Groups were also factored into the sampling scheme by appending the CFPB's BISG logic. The breakdown by race was 25% Hispanic, 25% Black, 25% White, 10% Asian-Pacific Islander, 10% American Indian/Alaskan Native, and 5% Unknown.[8] We did not have visibility to credit bureau trade-line data at the time we generated the sample, however within each race group we attempted to include consumers with thick credit files (60%), thin credit files (20%), and no-hits at Time1 (20%).

Sample Data & Coverage

The following data was appended to the sample:

- Credit Bureau Scores and Attributes from one of the three national credit reporting agencies.

7 CFPB (2014): Using publicly available information to proxy for unidentified race and ethnicity

8 This number can be contrasted to 2015 census numbers of 17% Hispanic and 12% Black. However, a more appropriate comparison would be against credit populations, where the Federal Reserve Study found 9% of a credit population were Black and 7% were Hispanic in a 2007 study. The Federal Reserve Study's numbers were based on Social Security Administration data for an active credit population.

Sources: www.census.gov and Avery. R. (2007). *Report to Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit.* Washington, DC: Board of Governors of the Federal Reserve System.

- Credit Performance was defined based on credit attributes at Time2 with the default definition defined as 90+ days past due over an 18-month performance window on any account.[9]
- LexisNexis RiskView Scores and Attributes.
- US Census Population Estimates for benchmarking purposes.

Comparison to the 2015 US Census and Federal Reserve Racial Distributions

Our sample statistics were compared to 2015 Census demographic estimates as a sanity check to ensure that the test sample represented the US population. There is a fundamental difference between the sample used in this analysis versus the US Census population. The consumers in the test sample are all active users of credit at Time2, whereas the US Census population includes a non-trivial percentage of individuals who are not active at any Credit Bureau.

When our stratified research sample is weighted back to the US population, the sample resembles the US population according to race and ethnicity estimates from the US Census Bureau more so than the Federal Reserve Study. Distributions by race for all three samples can be seen in Figure 1.

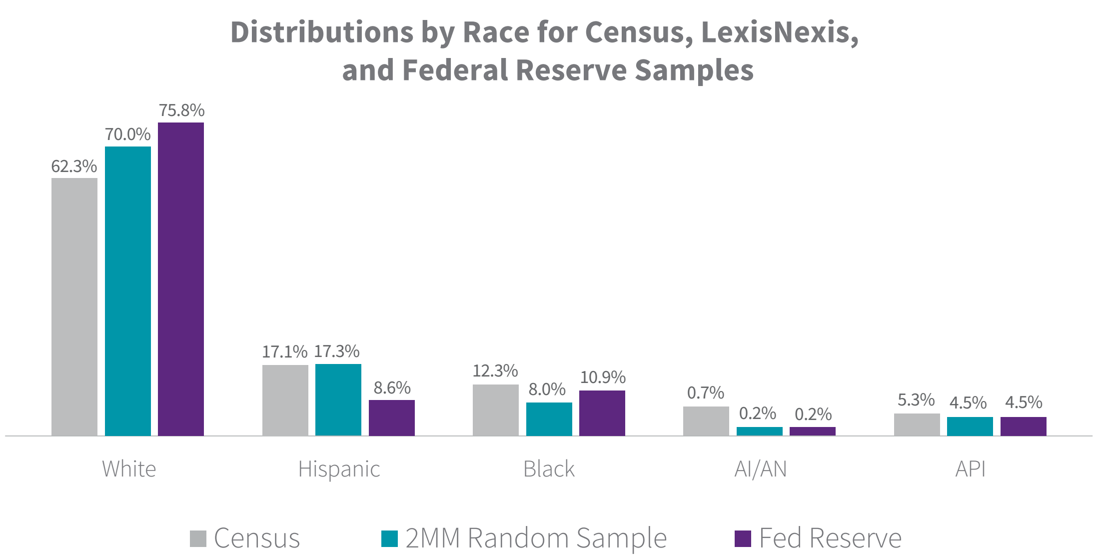

Figure 1: Comparison of RiskView Sample Distributions by Race and US Census (2015)

9 18 Months was used because credit risk scores typically use a period between 12 and 24 months.

Correlations between the CFPB BISG Race Proxy and Reported Race

In order to make inferences about statistical biases based on predicted race, we examined the relationship between the CFPB's race proxy logic and consumer reported race on driver's license and voter registration records collected by LexisNexis Risk Solutions through our government sources. This analysis is similar to an analysis done by the CFPB when they compared their BISG results to consumer *self-reported* race on credit-active mortgage applications. Their results showed moderate to high correlation between estimated race and actual race for Black, Hispanic, and White and Asian-Pacific Islander groups.

The CFPB found very low correlations for American Indian/Alaska Native group. According to the CFPB report, "...for non-Hispanic American Indian/Alaska Native and Multiracial, while generally improved by the use of the BISG proxy probabilities, [performance] is weak overall regardless of proxy choice, with only an 18% improvement in sorting over a random guess. These results suggest that proxies based on census geography and surname data are not particularly powerful in their ability to sort individuals into these two race and ethnicity categories."[10]

Our study noted similar trends. However, we reported only moderate correlations between proxy and actual race for Asian Pacific Islanders and American Indian/Alaska Native. Compared to the CFPB results, our study correlations were significantly greater for the American Indian/Alaska Native population and significantly lower for the Asian Pacific Islander group. This difference could be the result of differences in the type of records used in the benchmark populations.

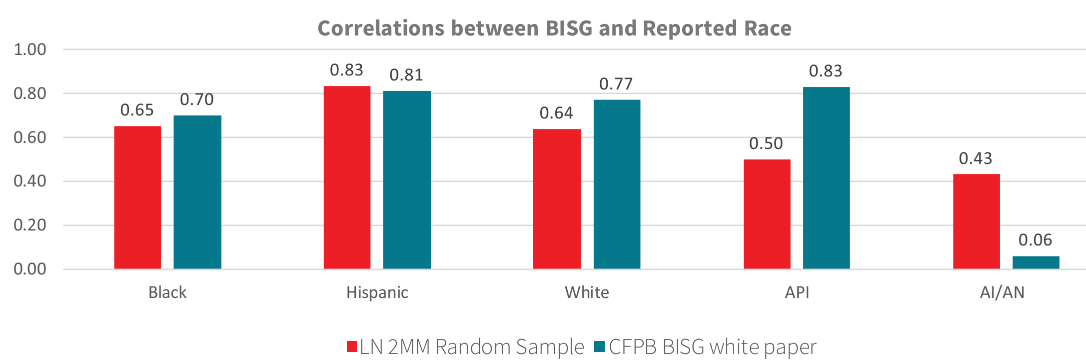

Figure 2: Correlations between Estimated and Reported Race

10 Using publicly available information to proxy for unidentified race and ethnicity: A methodology and assessment Page 18

Classifiable Performance

We found both the American Indian/Alaskan Native group and the Asian/Pacific Islander group to validate poorly, especially when considering consumers with valid credit performance. Despite our efforts to oversample from these groups, we could not append sufficient credit data with performance for robust statistical analysis, as seen in Figure 3. When we created the initial 2 million record sample from the LexisNexis Risk Solutions data, we had no visibility into credit bureau-based activity and we oversampled consumers with thin-files or no- files at Time1 that were less likely to have credit bureau based activity. As a result, a significant portion of our sample file did not have performance data at the credit bureau. The Asian/Pacific Islander group as well as the American Indian/Alaska Native group were smaller samples at the outset and also had low rates of credit bureau sourced credit performance. As a result of this finding, and the relatively low correlation between the CFPB proxy and actual race, these two groups were omitted from analysis.

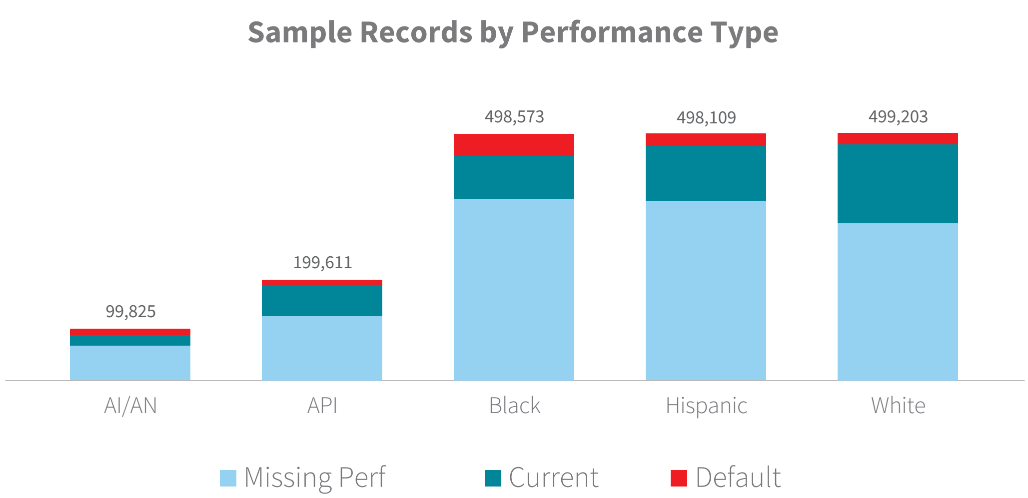

Figure 3: Sample Records by Performance Type

Analysis 1: Consumer Scorable Rates, Consumer Default Rates and Score Rank-Ordering by Racial Group

Analysis 1.1 Scorable Rates and Access to Credit

This analysis examined the relationship between RiskView Score and access to credit across minority populations. Throughout the remainder of this paper, we'll define "minority" as non-white consumers. The scorable rates of minority and non-minority consumers focused on when the consumers were scored using traditional Credit Bureau Scores and by RiskView at Time1. RiskView increases scorable rates of all consumers, especially minority consumers that apply for credit. A *Credit Invisible* is a consumer who is not scorable by a traditional Credit Bureau Score. LexisNexis Risk Solutions observed that 17.5% of the total population were *credit invisibles* using credit bureau scores, with 27.2% of Black consumers and 34.1% Hispanic consumers being *credit invisible*, much higher rates than the total population. 95% of the total *credit invisibles* and more than 90% of underserved minority *credit invisibles* can scored by RiskView.

The use of Alternative Data levels the playing field for consumers who have insufficient credit history to generate a traditional score, and allows them to participate in the credit system. Lenders who use Alternative Data can also benefit by finding and approving lower-risk consumers, many of whom are entering adulthood and just starting their financial lives.

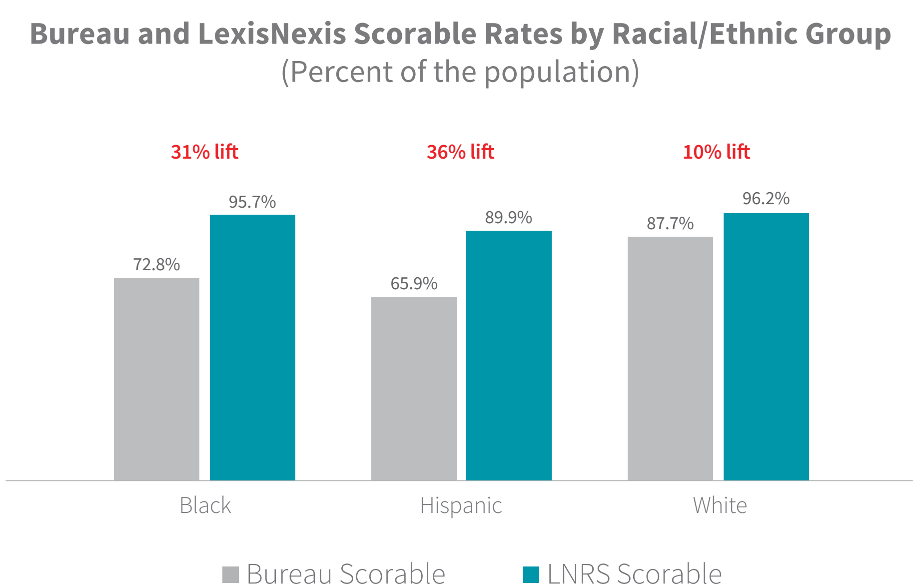

Figure 4: Scorable Rates by Racial / Ethnic Group

Analysis 1.2: RiskView Score and Consumer Default Rates across Racial Groups

We also examined the statistical relationship between RiskView Score and payment performance for different populations. As a reminder, credit performance was calculated as 90+ days past due or worse on the credit report 18 months after the scores were calculated on any new or existing account.

The first assessment was the score distribution and default rates of minority and majority consumers in our sample to examine different risk dynamics of the populations. As a subgroup, Blacks were more likely to default on credit obligations with a total default rate of 33.3% followed by Hispanics at 18.6% and Whites at 10.6%. Average RiskView Scores for Blacks were lowest at 665, Hispanics at 672 and Whites at 678.[11] In short, we observed different population odds in credit behavior for the three groups, and that these differences are reflected in the score. As we will see in the next section, these results are very consistent with those found in the Federal Reserve Study.

Comparison to Federal Reserve Default Rates

The Federal Reserve sample was based on a random sample of credit consumers that are active users of credit. Our sample included non-credit active consumers at the time of scoring (Time1), although only a small percentage of those individuals resulted in a booked trade line. Our approach, while a slight departure from the Federal Reserve Study, provides a more complete picture of how both credit bureau Scorable and Unscorable consumers are impacted when they apply for credit.[12]

In the Federal Reserve Study, performance was defined over an 18-month period. A subject's performance was calculated based on all new and existing accounts during the performance window, with default defined as any account 90+ days past due, involved in terminal delinquency, such as bankruptcy, repossession, collection or charged-off over that time period. Our performance definition followed the Federal Reserve Study. LexisNexis Risk Solutions was provided with aggregated performance fields in the sample so this was matched with the Random Account Performance

11 Default rate is defined as the percentage of consumers with a 90+ day delinquency in 18 months following the scoring date. The overall default rate is 14%. The Federal Reserve Study has a 12.2% overall default rate.

12 The Federal Reserve Study utilized "…demographics collected by the Social Security Administration (SSA) with a large, nationally representative sample of the credit records of individuals. The sample comprised the full credit records of 301,536 anonymous individuals drawn in June 2003 and updated in December 2004 by TransUnion LLC (TransUnion)."

classifications outlined in the Federal Reserve Study as closely as possible. This led to similar default rates as shown in Table 1 below.

Race	Federal Reserve Study Default Rate	LexisNexis Risk Solutions 2MM Random Sample Default Rate	Federal Reserve Study Scaled Average Score	LexisNexis Risk Solutions 2MM Random Sample Scaled RiskView Average Score
White	9.4%	10.6%	54	64
Hispanic	18.4%	18.6%	38	47
Black	33.4%	33.3%	26	45

Table 1: Default Rates and Average Scaled Scores for Minority and Non-Minority Populations

Analysis 1.3: RiskView Score Rank Ordering across Racial Groups

RiskView and other credit risk scores predict the probability of default; default rates decrease as scores increase. Default rate results for the study are in Figure 5.

- The y-axis denotes the overall default rate
- The x-axis denotes the RiskView Score band
- The lines represent the relationship between the RiskView Score and the Default rate for each racial subgroup

For each score band and racial group on the x-axis, as the score increases, the default rates decrease. This is true for all minority groups and indicates that the score differentiates risk well within all subgroups. RiskView Score is effectively rank-ordering consumer performance for the whole population as well as each sub-population. The curves are directionally consistent across all groups; however, the default rates differ substantially for each population. Given similar consumer profiles, Black consumers tend to have a higher probability of default than Hispanic or White consumers.

Our results are consistent with the Federal Reserve Study's results using credit bureau scores for minority consumers.[13] The Federal Reserve Study notes:

"A performance curve that is uniformly above (below) means that the group consistently underperforms (overperforms), that is, on average performs worse (better) on its loans than would be predicted by the performance of individuals in the

13 The Federal Reserve Study does not report specific numbers directly corresponding to score at specific default rates. Table O-3 shows a graph of default rates by score. There is a demonstrably higher score for underserved minorities than whites at the 10% default rate. For more information, see Figure O-3 on page O-27 of the report; and Figure 6 on pages 244-248 of the *Report to Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit*. Washington, DC: Board of Governors of the Federal Reserve System.

overall population with similar credit scores. Blacks, single individuals and individuals residing in lower-income or predominantly minority census tracts show higher incidences of bad performance than would be predicted by their credit scores."[14]

This under performance[15] by minority consumers is not an indication that the score has a differential effect on a protected class; rather it is an expected result, given the wide difference in population odds listed in Table 1. In short, when consumers from different groups with different overall default rates have the same profile, credit performance may differ in a way consistent with that group difference. To counter the under performance effect, the protected class information would need to be embedded into the score directly or by proxy, in order to adjust the score lower or higher. Doing so would bring these lines into much closer alignment. However, including these dimensions in the score algorithm would be a direct violation of ECOA. We demonstrate this effect for research purposes only in Analysis 3.2 below.

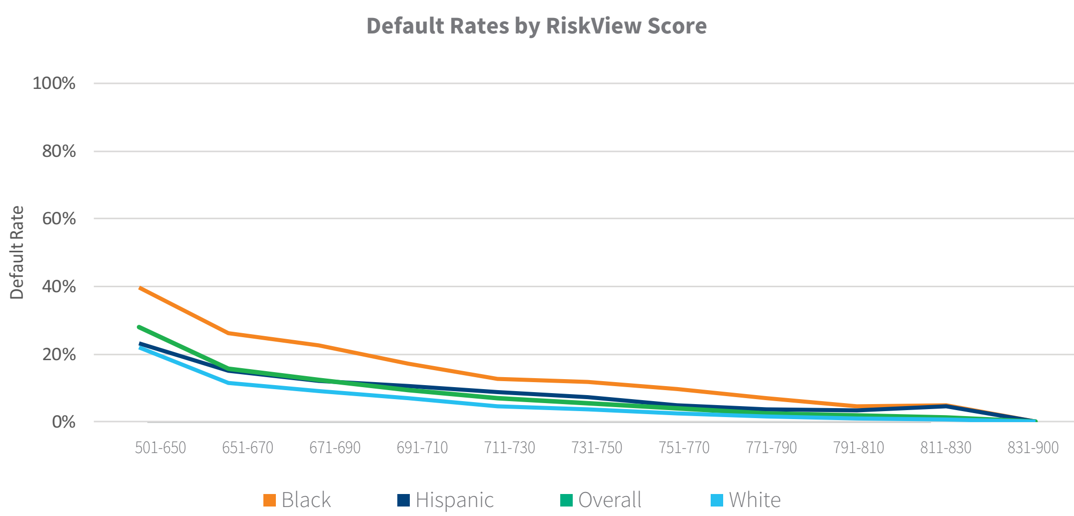

Figure 5: Default Rates by RiskView Score

We observed that score distributions for Black and Hispanic consumers were very similar, which is further corroborated by similarities between the two groups when

14 *Report to Congress on Credit Scoring and Its Effects on the Availability and Affordability of Credit* Page O-14

15 We observed underperformance, using the Federal Reserve definitions of over and under performance. By Avery's definition, under performance for minorities, does not constitute prohibited differential effect. Over performance, where a protected group has a lower default rate than would be indicated by score, could constitute differential effect.

we looked at RiskView variables. For example, Blacks and Hispanics tend to have similar values for address stability, homeownership, college attendance and other factors, therefore their RiskView scores tend to be similar. Figures 6 and 7 display the distribution of scores for each group.

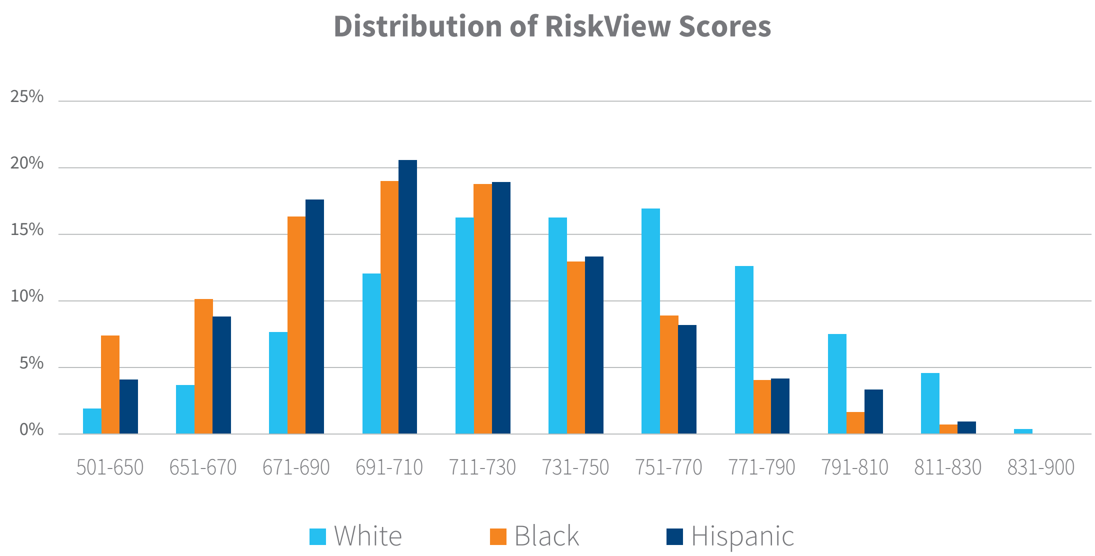

Figure 6: Distribution of RiskView Scores

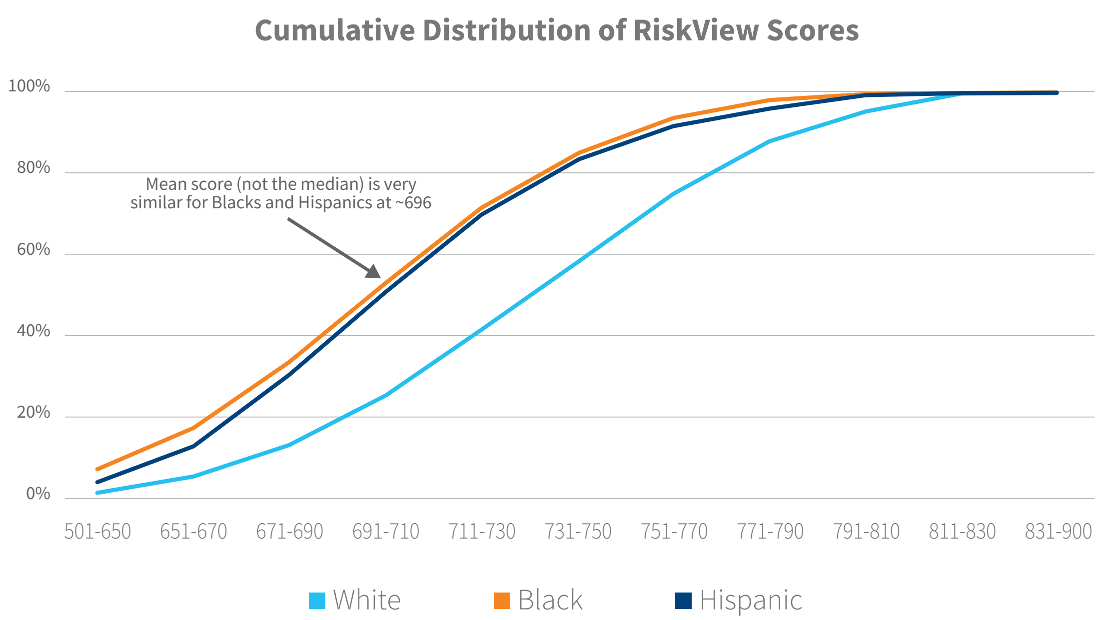

Figure 7: Cumulative Distribution of RiskView Scores

Analysis 2: Analysis of the relationship between Race and Variables in RiskView Score

In the prior analyses, RiskView Score was examined in its entirety. In analysis 2, RiskView's modeling *variables* were examined individually to gauge whether they are subject to statistical bias. The following are some of the key variables in the RiskView 5.0 Score.

Source Record Activity

Time on file

Time since last data update

Number of unique sources

Time since reported by a credit header source

Identity Confirmation

Address confirmation

Social Security Number (SSN) confirmation

Name confirmation

Date-of-birth confirmation

Identity High Risk Conditions

SSN validity

SSN reported deceased

SSN issued prior to date-of-birth

Residential Address History

Length of residence

Number of address moves

Dwelling type

Economic trajectory of last move

Current address ownership status

Statistical Analysis of RiskView and Race

Educational and Occupational History

Evidence of college attendance

Characteristics of college attendance

Assets and Evidence of Wealth

Real property ownership

Tax and market value of real property

Mortgage type

Watercraft registrations

Aircraft registrations

Derogatory Public Records

Bankruptcy filings

Tax liens

Civil judgments

Eviction judgments

Criminal convictions

Credit Shopping and Credit Inquiry Activity

Recent credit inquiries

Recent collections activity

Short-term loan offer request

LexisNexis® RiskView™ Attributes are based on extensive public records and alternative credit data sources. These variables contain information about positive and negative life events related to a consumer's stability, ability and willingness to repay debt obligations. Lenders may ask questions about these data types because they intuitively believe they may be highly correlated with race. Subsequent analyses show that this relationship does not exist at a statistically meaningful level and that use of this data actually increases approval rates for historically disadvantaged minority consumers.

Analysis 2.1: Drop a Variable Analysis

In this test, we examined whether a specific variable was having a large effect on lowering minority scores. This possibility was tested by systematically dropping individual variables from RiskView Score. If scores increase when a variable is dropped, it could mean that the variable's role in the score was penalizing minority candidates. Importantly, when a variable is dropped, the score was re-optimized without allowing new variables to enter the model to add additional "lost" information from the dropped variable. As a result, any changes in the score of the model are the result of losing only the dropped variable from the model.

RiskView scores were normalized on a rank-order scale ranging from 1 to 100 as described in the Federal Reserve Study. This allows direct comparisons between scores with different score distributions because of dropped variables or across different scores and studies. Each score is normalized to represent its place in the score's full distribution; a score of 50 places that individual at the median of the distribution. This rescaling allows for easier comparison between our study and the results found in the Federal Reserve's Study using credit bureau scores.

If there were prohibited differential effects due to a particular variable, minority scores would increase substantially when a variable highly correlated with race is dropped from the model. In this case, across 90+ variables used in RiskView Score's calculation, the average minority consumer score increased no more than 1.3 points when any specific variable was dropped. This is consistent with findings in the Federal Reserve Study for variables that exhibited no evidence of differential effect.

Analysis 2.2: Correlation Analysis

Next, we reviewed correlations between each of the RiskView variables and racial group. Sizable correlations between the variables in RiskView and race may indicate that a variable is serving as a proxy for race. The Federal Reserve Study cites variables with correlations to race that are greater than 0.2 as problematic. We found that no correlation exceeded this level, with the highest correlations being .17. Figure 8 shows the variables with the highest correlations with any of the race variables. No variable in RiskView is a proxy for race in this correlational analysis.

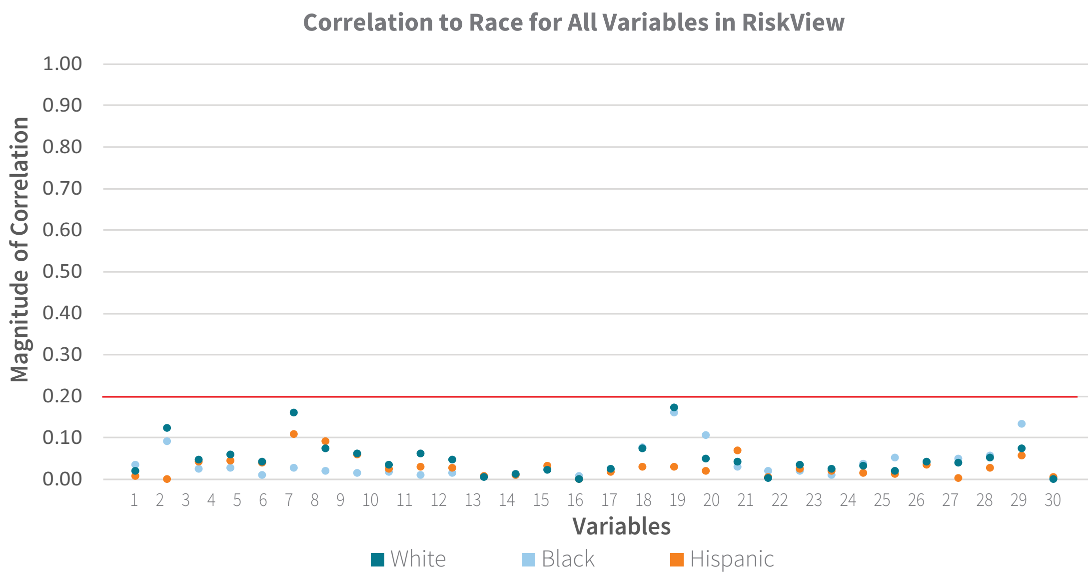

Figure 8: Correlation to Race for All Variables in RiskView

Analysis 3: Test of Differential Effect in RiskView Score

The final series of tests directly measured the relationship between racial group and RiskView Score by (1) statistically controlling race (analysis 3.1); and (2) explicitly including race in the scoring algorithm (analysis 3.2). These tests identify if race is a driving factor in RiskView when either explicitly neutralized or explicitly included in the model.

Analysis 3.1: Race-Neutral Model

Analysis 3.1 examined how minority consumers score when race is statistically controlled (neutralized) in the model. To conduct this test, a race-neutral environment was created by limiting the development sample to only White consumers. By limiting the racial groups comprising the sample, we can create a test environment where there are no differences between consumers as the result of racial differences. We can build a race-neutral model and compare scores as a function of race when race is statistically controlled (race-neutral model) or not (standard model). Each minority group was rescored with the race-neutral score and we compared the rank ordering of the race-neutral score with that of the original score. If minority consumer scores rise substantially when scored with the race-neutral model (where race was statistically controlled), factors in the original score may have been suppressing scores of minority consumers. The comparison showed that minorities scored within 1.6 points in each modeling environment across all minorities.[16]

Another check analyzed whether the race-neutralized model was as predictive as the original model. If the predictive value—as measured by Kolmogorov-Smirnov statistic (K-S)[17]—goes down for race-neutral models, it could be that the original model was predictive because of the presence of race proxies on a sample. In the final test, the loss of K-S was less than one point in each minority group using the scaled model. This result suggests that race is not a driver of the RiskView Score.

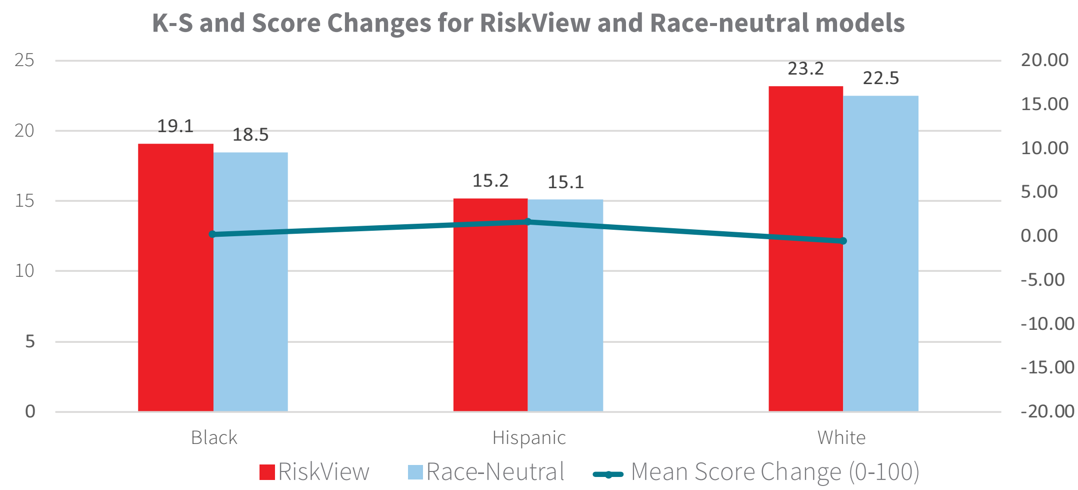

Figure 9: Score Change Before and After Race-Neutralized Model Development

Analysis 3.2: Race-Included Model

Finally, we measured the impact of racial group when race was included as a predictive variable candidate in the model. A score is likely to have differential effects if race does not add predictive value to the model. When race is included in the model, it should change the scores by race in line with the group differences in default

16 Because the base model and racially neutralized modeling population each have different population odds, we used the Federal Reserve Study's method of comparing the two scores based on percentile ranking changes instead of absolute point differences. While larger than The Federal Reserve Study's finding of .1, both values are statistically insignificant changes in score. In both cases, these would equate to a less than one-point difference.

17 K-S is measured as the maximum difference between the cumulative distribution of delinquent consumers and the cumulative distribution of payers. Higher K-S generally means that a score identified a certain number of percentage point more future delinquent consumers than future payers. The higher the K-S, the more effective the score.

rate (see Table 1). Note that while this is a useful test of statistical differential effect to measure the impact of race in the model, this practice is strongly discouraged in production models because the resulting model is very likely to have differential effect when implemented. For example, since Blacks have the highest default rate (see Table 1) and Whites have the lowest default rates, it would be expected that Black scores would decrease the most and Whites scores to increase the most when racial group is included in the model.

When included in the model, race does capture the default rate of each population and shows significant predictive power. Therefore, Black's scores decreased the most because their default rate is the highest, and White's scores increased the most because their default rates are lowest. We see that the overall K-S increases 8.6% (from 23.2 to 25.2) when race is added to the model.

If any variables in the model were acting as a proxy for race, then we would expect that the direct addition of a race indicator to cause the proxy variable(s) to drop out of the model or become significantly less predictive; and no variables showed this pattern. In combination, these results indicate that racial group is not captured by the model.

Figure 10 shows evidence that RiskView does not exhibit differential effects based on race. As hypothesized, when race is included in the model, all scores change.

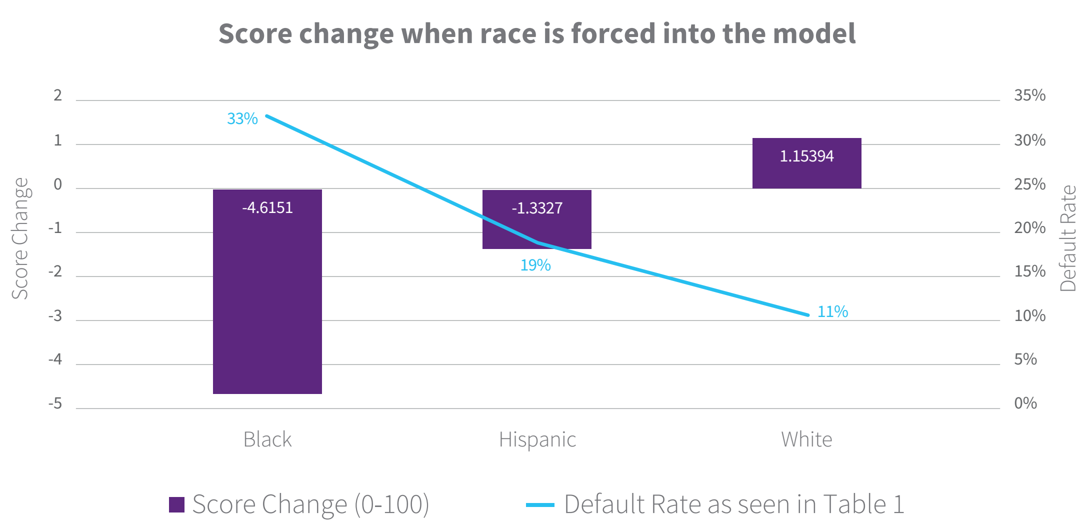

Figure 10: Mean Score Change with Race Included

Analysis 4: Case study in consumer outcomes using credit bureau data and Alternative Data

Our final analysis is an example of how a lender could create a strategy that leverages Alternative Data and credit bureau data to optimize bookings and risk while simultaneously benefitting historically disadvantaged minorities. We designed a simplified strategy in which a credit bureau score was the main criteria in determining approvals, and then examined the swap-in potential of the RiskView Score on margin. To simulate an approval strategy, we applied an approval cutoff score of 660; if a consumer found on the file had a credit score of 660 or above, they would be "approved" under our example strategy.

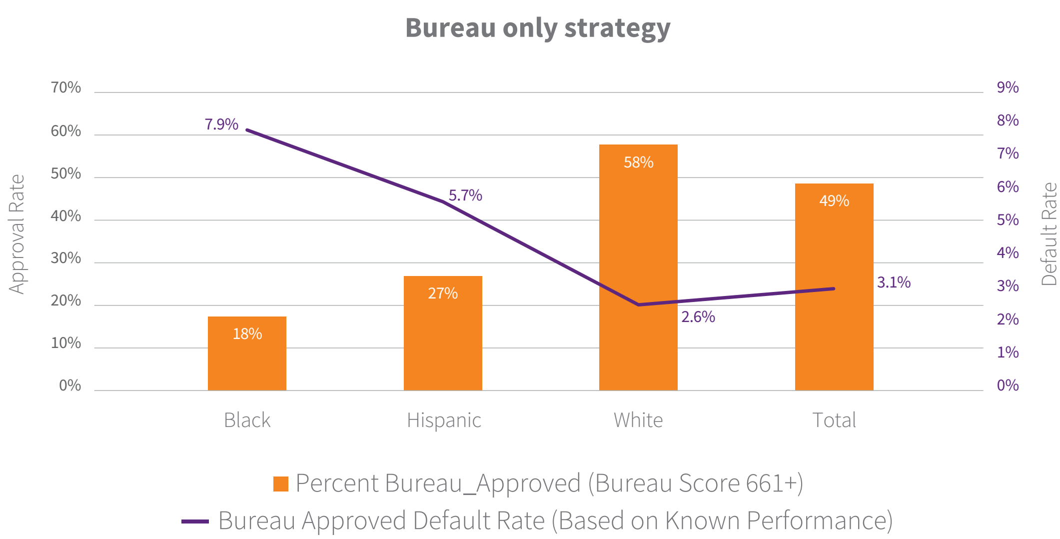

Figure 11: Approval and Default Rates by Race (Bureau Only Strategy)

The overall approval rate in the example strategy is 49% and the overall default rate is 3.1%, which is in line with the metrics for a typical prime/near-prime portfolio. However, the approval rates for Black and Hispanic consumers are far below acceptance rates for White consumers. This is reflective of the higher default rates in those populations.

We then added additional criteria using the RiskView data to the strategy:

- For consumers that had a traditional credit score between 630 and 660, those that were close to the approval cutoff, everyone with a RiskView Score above 750 was "approved"
- For consumers that had thin-files or were not present on the credit bureau sample at the time of application (but had performance data on file 18 months later), everyone with a RiskView Score above 700 was "approved"

This is how a lender might implement Alternative Data in a "2nd chance" strategy that increases approvals while controlling default rates.

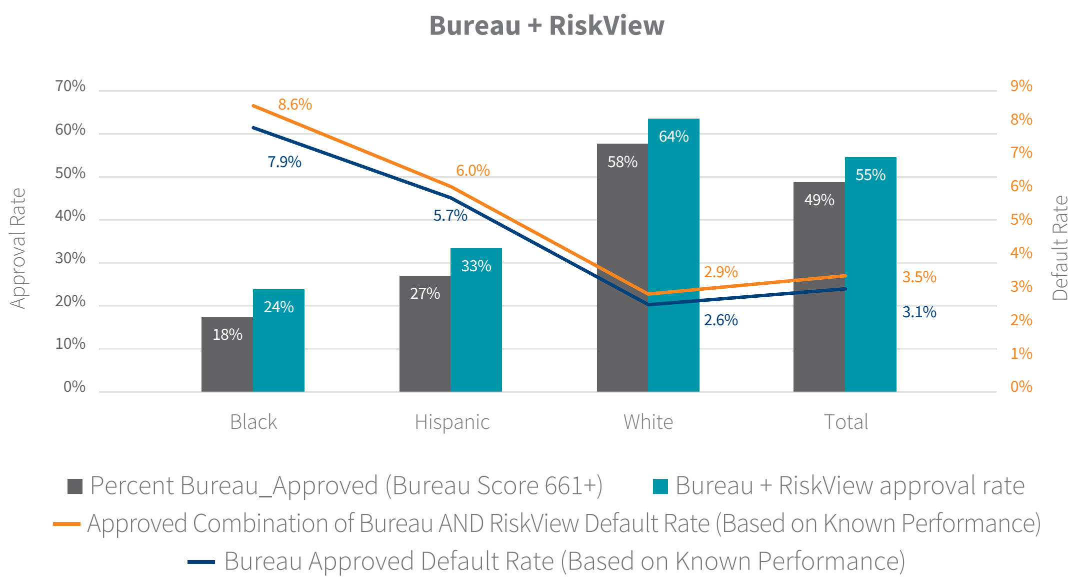

Figure 12: Approval and Default Rates by Race (Bureau and RiskView Strategy)

The overall approval rate increases by 6% and the overall default rate increases by 0.4%, which is a reasonable tradeoff that can be further optimized using a more sophisticated strategy than we could build using only the sample data.

The incremental approval rate for consumers with thin-files or no presence at the credit bureau was strikingly high for all three groups using this strategy. 86% of Black, 85% of Hispanic and 71% of the White new approvals were thin- or no-files. This study shows that lenders can use Alternative Data not just to improve underwriting and increase bookings, but also to extend credit to historically disadvantaged minorities and increase financial inclusion.

Statistical Analysis of RiskView and Race

Summary

Our analysis subjected RiskView Score to a battery of tests of differential effect against minorities that were based on a 2007 Federal Reserve Study of credit bureau scores. The Federal Reserve Study used trade-line level data to explore the relationships between race and credit scoring and found no evidence of differential effect by race. The results of the RiskView study were very similar. RiskView Score rank-orders consumers based on credit profile independent of their minority status. Specifically, our study found that RiskView Score:

- doesn't unfairly penalize consumers relative to their performance
- shows no bias when race is statistically removed from the model
- hows no bias when variables are dropped from the model
- helps Credit Invisibles get access to credit
- can be used responsibly in credit risk underwriting models to foster financial inclusion

Learn more at risk.lexisnexis.com

LexisNexis®
RISK SOLUTIONS

About LexisNexis Risk Solutions

LexisNexis Risk Solutions harnesses the power of data and advanced analytics to provide insights that help businesses and governmental entities reduce risk and improve decisions to benefit people around the globe. We provide data and technology solutions for a wide range of industries including insurance, financial services, healthcare and government. Headquartered in metro Atlanta, Georgia, we have offices throughout the world and are part of RELX Group (LSE: REL/NYSE: RELX), a global provider of information and analytics for professional and business customers across industries. RELX is a FTSE 100 company and is based in London. For more information, please visit www.risk.lexisnexis.com, and www.relx.com.