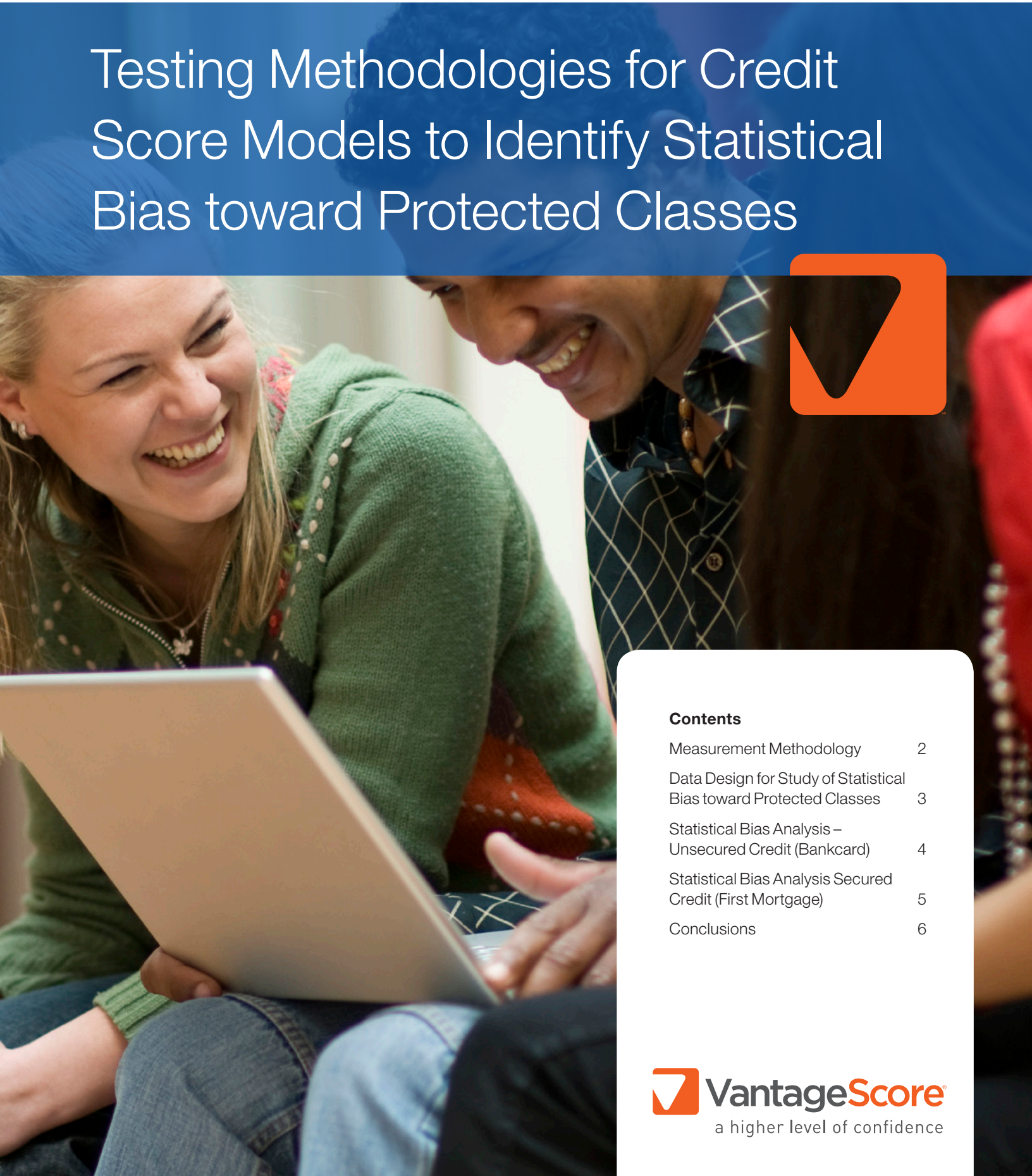
MAY 2014

Testing Methodologies for Credit Score Models to Identify Statistical Bias toward Protected Classes



Contents

VantageScore®
a higher level of confidence

Testing Methodologies for Credit Score Models to Identify Statistical Bias toward Protected Classes

INTRODUCTION

The **Equal Credit Opportunity Act** (ECOA), implemented by Federal Reserve Board's Regulation B (12 CFR 202), prohibits discrimination in any aspect of a credit transaction on the basis of specific population classifications. Protected classes are:

- Race
- Color
- National origin
- Marital status
- The applicant's exercise, in good faith, of any right under the Consumer Credit Protection Act
- Religion
- Sex
- Age (provided the applicant has the capacity to contract)
- The applicant's receipt of income derived from any public assistance program

Credit score models, such as VantageScore 3.0, are mathematical algorithms derived from information in a consumer's credit report to assess whether a consumer is likely to pay their debt obligations within the agreed upon terms. These credit reports, which use data from lenders, other creditors and public records, are primarily based on information regarding an individual's previous use of, and application for, credit. No additional information such as age of the consumer, marital status, employment history, ethnicity, etc., is used in the algorithm.

ECOA concerns would arise with respect to the credit score used in a credit extension transaction if the model unduly favors an outcome for a particular group of people over another outcome with a different group of people, even though they both receive the same score with the same model. Specifically, if a given credit score represents different levels of risk (probability of default) given the same model for two similarly situated populations of different ethnicity, then the credit score model is favoring one population over another.

It is the purpose of this paper to discuss how to appropriately analyze and measure evidence of statistical bias in a credit score that causes disparate impact in a lender's credit extension transaction and to demonstrate that VantageScore 3.0 reflects no bias on protected classes, specifically by analyzing ethnic classes.

MEASUREMENT METHODOLOGY

How to assess if a credit score model reflects statistical bias toward a protected class

The formal definition of a credit score is a measure of risk defined by the probability that a consumer will default on a loan. Default, in this instance, is defined as a consumer being 90 or more days past due (90+ DPD) on an account. Assessing if a credit score reflects statistical bias requires assessing the probability of default for each credit score (collectively known as "credit score default curves") for each subpopulation of consumers and comparing the credit score default curve to all other sub-populations. If there are measureable differences between the sub-populations, the corresponding curves will look decidedly different when compared amongst each other. In these cases, the credit score model is unduly biased (either positively or negatively) towards a particular sub-population and this suggests that there is potential bias or preferential treatment/mistreatment.

Graphically, such a comparison between biased and unbiased scores would look like the example charts (Figure 1 & 2).

Example of a Biased Model:
In this case, the grey sub-population is defaulting at a much lower rate than the orange population at each credit score value. Thus, the credit score is unfairly impacting the orange population. For example, at a score of 550, the grey population defaults at a 30% rate, whereas the orange population defaults at a 40% rate. A consumer in the grey population with a score of 550 behaves more similarly to a consumer in the orange population at a score of 600, and is being negatively impacted by the credit score.

Example of an Unbiased Model:
Here, given the same credit model, all the sub-populations have similar outcomes; thus, there is no statistical bias that favors any sub-population.

Although graphs show a nice visual explanation of bias, they are not conclusive in determining the existence of bias, since they do not formally compare default probabilities. A more rigorous process is required to conclusively determine the presence of bias. This requires statistical testing of individual default probabilities within each sub-population across the entire credit score range to determine if there are significant differences.

Figure 1: Biased Score

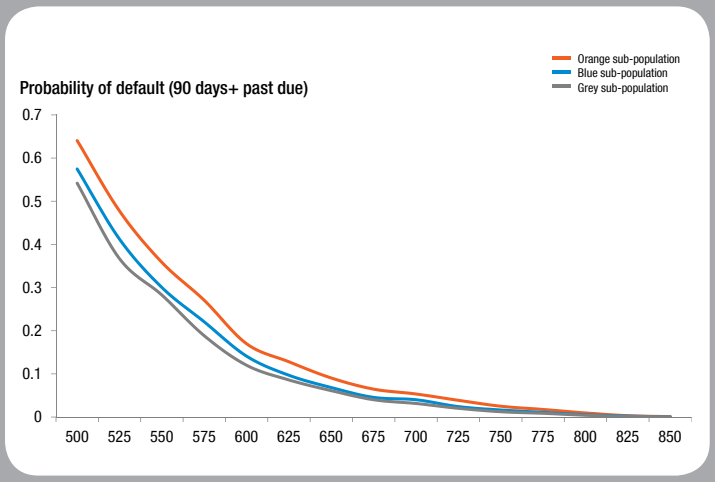


Figure 2: Unbiased Score

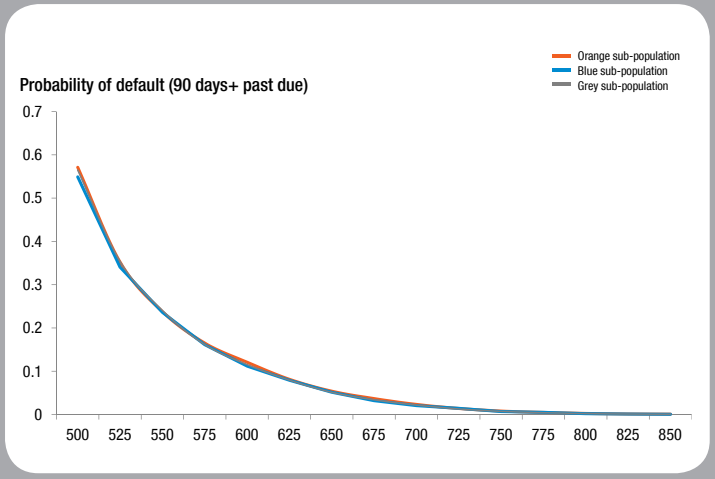


Figure 3: Chi-Square Comparison Test

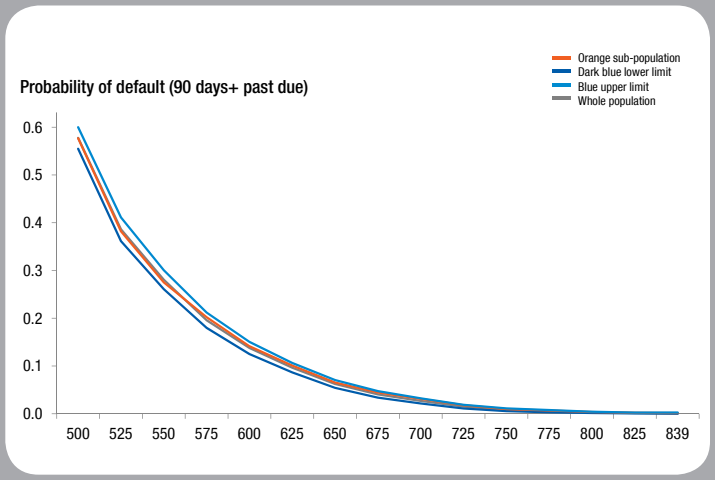


Figure 4: Disparate Impact: Bankcard default profiles by ethnicity with confidence intervals

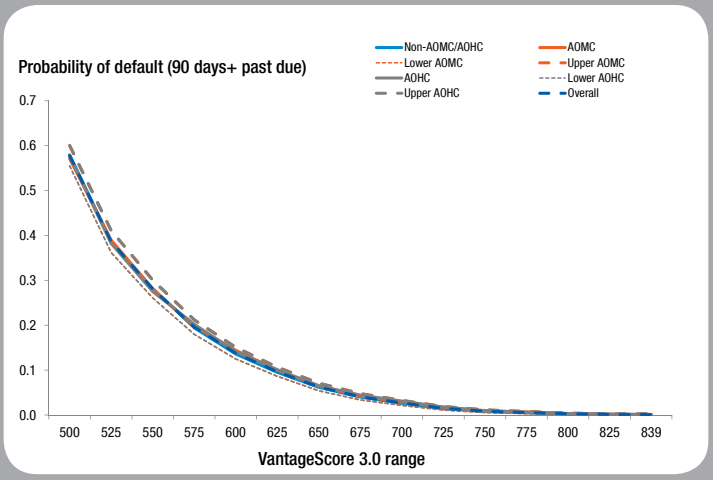
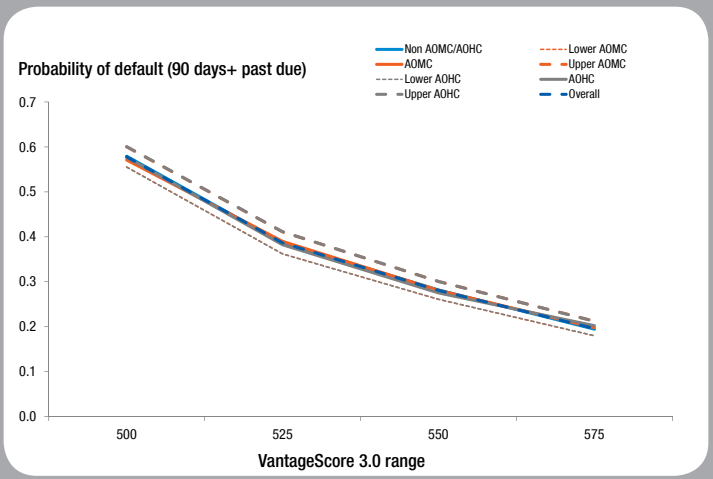


Figure 5: Disparate Impact: Bankcard default profiles by ethnicity with confidence intervals (score range 500-575)



A formal test to determine if there are differences between subpopulation default probabilities is a statistical comparison test called the "Chi-Square" test for multiple probabilities. To perform this test on VantageScore 3.0, the score range is divided into buckets of 25-point bands, to ensure that sufficient samples exist for the testing procedure. The initial score band is set to be any score less than 500 since the population in the distribution tail is very sparse. All intervals above 500 are in 25 point bands.

In each score band the Chi-Square comparison tests to see if there are differences in default probabilities amongst sub-populations. The test calculates the actual proportions within each sub-population in the score interval and compares them to the whole population in the same interval. If the differences between sub-populations and whole population proportions are large (i.e., statistically significant, as measured by comparing to a critical value), then there is a demonstrated measureable bias. If not, then there is no measureable impact. The test is performed across all score bands; if one band fails the test then there is a bias implication for the model as a whole.

This test can be represented graphically. The comparison test produces thresholds (lower and upper) to determine where each sub-population is considered within normal population boundaries. In Figure 3, if the orange sub-population breaches the lower and upper thresholds (grey dashed lines) then there is statistically significant evidence to suggest bias.

DATA DESIGN FOR STUDY OF STATISTICAL BIAS TOWARD PROTECTED CLASSES

To assess if VantageScore 3.0 exhibits statistical bias toward protected classes, two products are considered: an unsecured credit product (Bankcard) and a secured credit product (First Mortgage). This study is assessed on ethnic protected class sub-populations, namely African American (AOMC) and Hispanic-American (AOHC) populations. One million consumers owning the product in question are randomly selected from data spanning the 2010 to 2012 time frame.

To measure ethnicity, a consumer's ZIP Code also was appended to the file for look-up purposes based on the US Census Bureau's database.

Ethnicity Weighting

Since a consumer's ethnicity cannot be directly determined, "weights" are applied to each of the randomly selected consumer credit reports in the sample. This is done by using corresponding ZIP Codes matched to the US Census

Bureau's 2011 "American Community Survey"[1] which help identify the ethnic demographic; specifically:

- AOMC—proportion of African-American households in ZIP Code
- AOHC—proportion of Hispanic-American household in ZIP Code
- Non-AOMC/Non-AOHC—proportion neither African-American nor Hispanic-American in ZIP Code

For each consumer, these proportions are attached to their credit file and the corresponding credit score and account information are summed to produce population credit score default curves. For example, if a particular ZIP Code has the following weights: 30% African-American, 20% Hispanic-American and 50% Non–African-American/Hispanic-American, then each consumer is weighted 0.3 AOMC, 0.2 AOHC and 0.50 Non-AOMC/NonAOHC. If the ZIP Code had 100 consumers sampled then the weights would sum to 30 AOMC, 20 AOHC and 50 Non AOMC/Non AOHC, respectively.

STATISTICAL BIAS ANALYSIS – UNSECURED CREDIT (BANKCARD)

Protected Classes –Ethnicity

Using VantageScore 3.0, a graphical comparison shows all three ethnic classes essentially establishing the same probability of default as the credit default curves are on top of each other (Figure 4). Moreover, all default curves based on VantageScore 3.0 are well contained within the upper and lower acceptable thresholds. Although the graphs "look" to align, the exact results from the Chi-Square test are needed to provide conclusive evidence that no disparate impacts exist.

A closer inspection of the graph (Figure 5) shows the lower score range default rates (500 to 575) are slightly lower for AOHC (Hispanic grey line) population versus the other groups, whereas AOMC (African-American orange line) default rates are on top of overall population default rates. Both ethnic groups are well within their confidence intervals, indicating there are no measureable differences between the groups at each credit score value and the overall population default rates.

Within each VantageScore 3.0 score interval, the Chi-Square test statistics for comparisons of probability of default are well below the critical value, 11.408, established as the threshold to determine if statistical bias is evident. In other words, since no calculation is larger than 11.408, there is no evidence of statistical bias amongst the three sub-populations' default rates across the entire VantageScore 3.0 score range.

1 US Census Bureau's "American Community Survey", 2011: http://www.census.gov/acs/www/

Figure 6: Multiple Comparison Test of Probability to Default for Identifying Statistical Bias in the credit score model toward Ethnic Classes on Unsecured Credit

VantageScore 3.0 Interval																
	Start point	350	501	526	551	576	601	626	651	676	701	726	751	776	801	826
	End point	500	525	550	575	600	625	650	675	700	725	750	775	800	825	839
Test Chi-Square		1.048	0.424	0.821	1.879	3.581	3.744	2.265	3.543	7.545	9.682	3.239	6.821	2.932	3.729	0.808
Critical Value		11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408
Is Test -> Critical Value (if "Yes" then Disparate Impact)		NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO

Figure 7: Measurable Bias: First Mortage Default Profiles by Ethnicity with Confidence Intervals

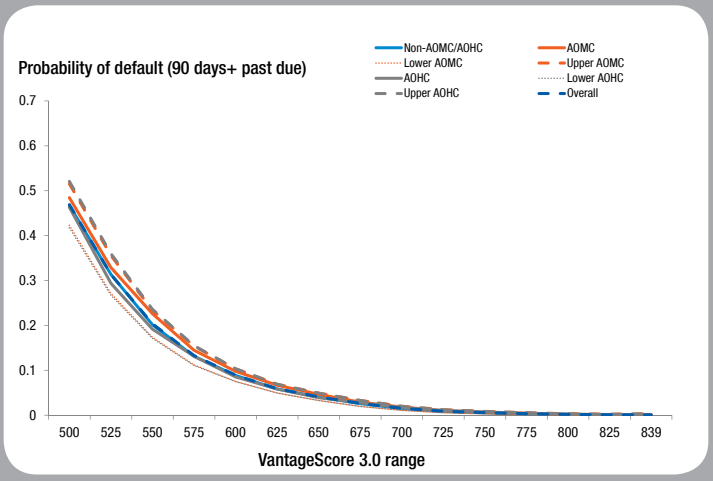
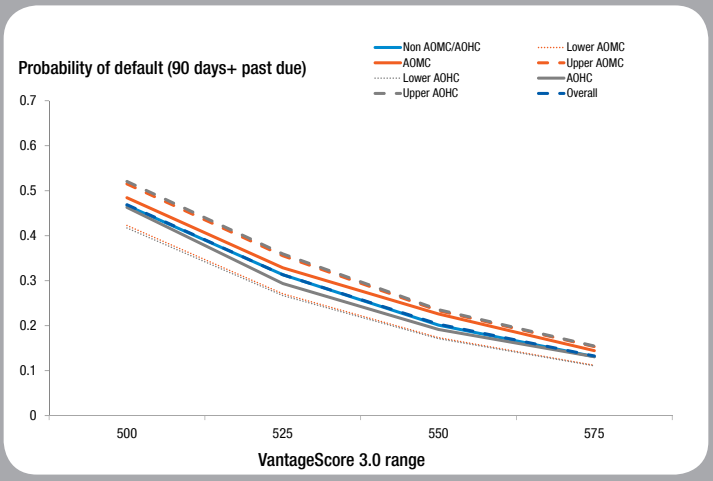


Figure 8: Measurable Bias: First Mortage Default Profiles by Ethnicity with Confidence Intervals (Score Range 500-575)



STATISTICAL BIAS ANALYSIS SECURED CREDIT (FIRST MORTGAGE)

Unlike the unsecured product, the secured product (namely a first mortgage) has an "asset value" attached to the loan. As a result of the 2008 economic housing crisis, these asset values have been under extreme stress. Hence, most mortgages prior to 2009 have had unprecedented stresses applied to them and have produced overwhelming factors, unrelated to credit scores, that contribute to default behaviors.

Data Design

To remove the inherent problems associated with these stress factors, two filters have been established to assess whether the model reflects bias in credit extension transactions used for secured products. Reviewing mortgages originated after 2009 removes some of the stresses induced by the crisis and the housing bubble. A second filter, price-to-income (PTI) scaled mortgages, represents "ability to pay off" based on an income factor. This scale determines if the consumer has enough resources to pay off the mortgage. Many of the mortgages that went into default during the housing crisis have been linked to a class of "low documentation" or "no documentation" loans which essentially ignored the "ability to repay" requirement, causing mortgage defaults to occur. To establish a 'sound' mortgage, the value of the mortgage is divided by the household's income to produce a PTI ratio on the mortgage. A mortgage is considered 'sound' if the mortgage has a PTI of 3 or less.

If no actual income information is available on the credit file, the US Census American Community Survey data can be used as a proxy. The survey has median homeowner household income according to zip code. A random sample of 860,000 mortgage consumers from 2009 onwards with 'sound' price-to-income (PTI ≤ 3) values was obtained for this analysis.

Protected Classes—Ethnicity

Again, applying VantageScore 3.0, the graphical comparison (Figure 7) shows some initial separation, in terms of default profiles, in the lower credit score range. Yet, all credit score default curves are contained within the upper and lower acceptable boundaries, providing evidence that VantageScore 3.0 does not exhibit statistical bias when used on credit extension transactions for 'sound' mortgages.

Again, a closer inspection of the graph (Figure 8) shows the lower score range default rates (500 to 575) are lower for AOHC (Hispanic solid grey line) population versus the overall

Figure 9: Multiple Comparison Test of Probability to Default for Identifying Statistical Bias in the credit score model toward Ethnic Classes on Unsecured Credit

VantageScore 3.0 Interval																
	Start point	350	501	526	551	576	601	626	651	676	701	726	751	776	801	826
	End point	500	525	550	575	600	625	650	675	700	725	750	775	800	825	839
Test Chi-Square		0.450	2.102	4.651	1.420	6.325	5.606	5.819	2.111	9.261	7.618	5.111	3.993	4.500	0.568	0.943
Critical Value		11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408	11.408
Is Test -> Critical Value (if "Yes" then Disparate Impact)		NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO

population and AOMC (African-American solid orange line) default rates are higher than the overall population. However, both ethnic groups are well within their confidence intervals indicating there are no measureable differences between the groups at each credit score value and the overall population default rates.

Examining the Chi-Square tests for proportions demonstrates that VantageScore 3.0 does not exhibit statistical bias in credit extension transactions for secured lending, since all intervals show no probability of default differences amongst the ethnic classes.

CONCLUSIONS

Credit score bias toward protected class analysis requires focusing only on the outcomes used in credit-making decisions. Doing so will ensure that all consumers are properly assessed by comparing default rates based on credit scores within any protected class. The probability of default given a credit score should be consistent amongst all protected classes, although there may be score distribution differences between different protected classes. VantageScore 3.0 exhibits no statistical bias amongst protected classes.

Ethnicity Study

In both instances, secured and unsecured credit products, there is no evidence of bias toward protected classes when using VantageScore 3.0. By comparing one million randomly selected consumers in each case, there were no discernible differences in the probability to default within each score band when these consumers were overlaid with demographic data. VantageScore 3.0 produces no favorable biases in assessing risk outcomes for either product for any impacted ethnic group.

The VantageScore credit score models are sold and marketed only through individual licensing arrangements with the three major credit reporting companies (CRCs): Equifax, Experian and TransUnion. Lenders and other commercial entities interested in learning more about the VantageScore credit score models, including the VantageScore 3.0 credit score model, may contact one of the following CRCs listed for additional assistance:

EQUIFAX®

Call 1-888-202-4025

www.equifax.com/vantagescore

Experian

Call 1-888-414-4025

www.experian.com/consumer-information/vantagescorelenders.html

TransUnion

Call 1-866-922-2100

www.transunion.com/corporatebusiness/solutions/financialservices/bank_acq_vantage-score.page

VantageScore
May 2014

www.vantagescore.com