

# Score Consistency Index

A METHODOLOGY FOR CALCULATING A “SCORE CONSISTENCY INDEX”  
AS A MEANS OF MEASURING THE CONSISTENCY OF MULTIPLE  
GENERIC RISK SCORE ALGORITHMS ACROSS MULTIPLE CRCS.  
APRIL 2008

VANTAGESCORE<sup>SM</sup>

The New Standard in Credit Scoring.

# Executive Summary

## Overview

Typical real estate underwriting procedures require three credit scores for assessing a consumer's creditworthiness – one score from each of the three national credit reporting companies (CRCs). Lenders require that these scores are accurate in predicting credit risk and also **highly consistent in their absolute value across the CRCs**. Scoring algorithms that provide inconsistent scores can increase the risk exposure that a lender takes on, resulting in less attractive products and pricing offered to the borrower.

Inconsistent scores occur largely due to different score algorithms in place at each CRC (TransUnion, Equifax and Experian) as well as variations in data reported by creditors and the timing of that reporting. A credit score for a consumer can vary by more than 60 points between the CRCs.

Measuring score predictiveness is well understood using tests such as Kolmogorov-Smirnoff (KS) statistics, however measuring score consistency is challenging for the same reasons stated previously. Additionally different scores often use different numerical ranges, further confusing the understanding of risk. For example, it's possible that one algorithm has a range of 300 to 700 where 650 indicates low risk and a different algorithm has a range of 600 to 900 where 650 indicates high risk. Thus a consumer may score 650 using two different algorithms yet have very different risk profiles.

As lenders look to improve the quality of their underwriting processes, a framework is clearly necessary for evaluating the consistency of generic credit score algorithms. This paper presents a patent-pending methodology for calculating a “Score Consistency Index” as a means of measuring the consistency of multiple generic risk score algorithms across multiple CRCs.

## Methodology

We calculate consistency of consumer credit scores across multiple CRCs or across multiple algorithms by utilizing a simple ranking technique. We first obtain credit scores for a portfolio of consumers using two or more algorithms at each of the CRCs. Each consumer is ranked and then placed in tiers for each algorithm based on their score from that algorithm. For example, if a consumer receives two different scores from two different algorithms, but

both scores rank them in the top 10 percent of their respective scored populations, then for this consumer, those two algorithms are highly consistent in risk assessment. Conversely, if a consumer receives a score from one algorithm that ranks them in the top 10 percent of the scored population for that algorithm, and then receives a score from a second algorithm that ranks them in the bottom 10 percent of its unique scored population, then those two algorithms are highly inconsistent.

### Application

This methodology was applied to a pool of consumers who were scored with VantageScore and comparable CRC proprietary generic risk scores. The two scores for each consumer were obtained from the three CRCs. Consumers were first ranked by VantageScore according to four tiers. The first tier (the top 15 percent of the pool), defined as super-prime, the second defined as prime (approximately 50 percent of the pool), the third defined as near-prime and the final tier defined as sub-prime.

Tier	% of population	Risk profile
1	15	Super prime
2	50	Prime
3	25	Near prime
4	10	Sub prime
Total	100	

The percentage of consumers whose scores ranked them in the super-prime tier across all three CRCs was calculated. Similar percentages were calculated for prime, near-prime and sub-prime. The combined percentage gives the Score Consistency Index, that is, the percentage of consumers who were ranked consistently at the same risk level across multiple CRCs. The approach was repeated for the second generic risk score. Comparing these two Score Consistency Indices allows the lender to assess which score provides greater consistency in risk assessment.

### Results Summary

The tests conducted for this paper demonstrate Score Consistency Index values are consistently in the 70 percent range for VantageScore and in the 50 percent range for the comparative generic credit scores from each of the three CRCs. VantageScore is typically 30 percent more consistent than these other generic risk scores.

In this scenario, using VantageScore allows business users and consumers alike to have a more consistent prediction of consumers' credit payment behavior. As a result, creditors can plan more consistent lending strategies and make better credit decisions, while consumers see more consistent scores, reducing confusion

The Score Consistency Index approach provides a robust transparent assessment methodology for evaluating generic risk score consistency. The methodology enables lenders to quantitatively compare consistency performance of score algorithms and to factor this information in their overall assessment of the score algorithm's accuracy.

# Details

## 2. Score Consistency Index Introduction

Traditional generic risk scores are subject to large variations across CRCs. These variations are driven from three sources: 1) differences in data submission by lenders and other entities; 2) differences in data classification by CRCs; and 3) differences in the score algorithms in place at each CRC. Further, different scores use different ranges to measure risk.

A key strength of VantageScore is that the resulting score is highly consistent across data provided by any of the three national CRCs. A consistent predictive score enables lenders to implement optimal credit decision strategy, reduces confusion for the consumers when evaluating their own credit profile and helps regulators gauge lending exposure more precisely.

VantageScore utilizes sophisticated data standardization, known as characteristic leveling (1), and segmentation modeling (1) to minimize the impact of the primary drivers of score variability. As is shown in this paper, VantageScore has significantly improved score consistency over traditional CRC generic risk scores.

Similar to comparing the power of risk models using an industry standard measure of KS values, an objective framework must be developed to measure score consistency. This framework can be used to calculate the consistency index for VantageScore and any other credit score.

In the most rigorous sense, score consistency means to ask the following question: For any given consumer, if they score 800 at one CRC, do they also score 800 at the other two CRCs? (With the understanding that a score of 800 reflects the same risk profile for each score range at each CRC). For the overall population, this question can be re-phrased as: What percentage of the population receives the same score at all three CRCs?

While this question can be asked for a consumer's VantageScore because it uses the same range and algorithm regardless of CRC, the same question cannot be easily answered with other CRC generic risk scores, due to reasons described previously of data content and process, algorithm design and range. In order to have a fair comparison of the relative consistency for different risk scores a measurement must be developed that can be equally and objectively applied to all risk scores.

### 3. Score Consistency Index Formulation

Let GR\_Score be three of the respective CRC's proprietary generic risk scores. Let GR\_Score\_CRC1 denote the GR\_Score calculated and pulled from CRC1, GR\_Score\_CRC2 denote the GR\_Score calculated and pulled from CRC2, and GR\_Score\_CRC3 denote the GR\_Score calculated and pulled from CRC3.

Score a random sample with the condition that GR\_Scores are available for each and every record in the sample from all three CRCs. Rank order the population from high score to low score using GR\_Score\_CRC1. Assign the top scored X<sub>1</sub> percent of the population into a category labeled "Low Risk", put the next X<sub>2</sub> percent of population into category "Medium Risk", and the next X<sub>3</sub> percent of population into category "High Risk", and the rest X<sub>4</sub> percent (the lowest scored) population into category "Very High Risk", as shown in the table below.

Population Groups	Label	Population Breaks
Low Risk	L	X <sub>1</sub> %
Medium Risk	M	X <sub>2</sub> %
High Risk	H	X <sub>3</sub> %
Very High Risk	V	X <sub>4</sub> %
Total	-	X <sub>1</sub> % + X <sub>2</sub> % + X <sub>3</sub> % + X <sub>4</sub> % = 100%

Similarly, rank order the same population using GR\_Score\_CRC2, and assign them into the same risk categories using the SAME percentage breaks (i.e. X<sub>1</sub>%, X<sub>2</sub>%, X<sub>3</sub>%, X<sub>4</sub>%). Repeat the process using GR\_Score\_CRC3.

Next, we check the number of consumers who are categorized as ‘Low Risk’ in CRC 1 and are also categorized ‘Low Risk’ in CRC 2 and ‘Low Risk’ in CRC 3. Similarly the same check applies for the Medium Risk, High Risk and Very High Risk groups.

The score consistency index (SCI hereafter) is constructed using the following notations:

N: the total number of consumers in the sample

N<sub>1</sub>: the number of consumers who are categorized into “Low Risk” in all three CRCs

N<sub>2</sub>: the number of consumers who are categorized into “Medium Risk” in all three CRCs

N<sub>3</sub>: the number of consumers who are categorized into “High Risk” in all three CRCs

N<sub>4</sub>: the number of consumers who are categorized into “Very High Risk” in all three CRCs

**SCI (Score Consistency Index) = (N<sub>1</sub> + N<sub>2</sub> + N<sub>3</sub> + N<sub>4</sub>) / N**

#### 4. A Simple Example of SCI Calculation

For easy illustration, this section provides a simple example of SCI calculation using 20 consumers (so  $N=20$ ), and the population is broken into 4 equal sized risk groups (so  $X_1\% = X_2\% = X_3\% = X_4\% = 25\%$ ). We will calculate SCI for hypothesized generic risk scores, named GR1, 2 and 3, which are respectively available from the 3 CRCs, with a hypothetical score range of 1 to 1000. For each consumer, the GR score from CRC 1 is denoted by GR\_CRC1, from CRC 2 denoted by GR\_CRC2, and so on. All score values are arbitrary and for illustration purpose only.

Consumers	GR_CRC1	GR_CRC2	GR_CRC3
Consumer 1	739	750	630
Consumer 2	890	981	730
Consumer 3	150	366	233
Consumer 4	460	761	638
Consumer 5	890	996	988
Consumer 6	874	379	569
Consumer 7	762	475	485
Consumer 8	569	345	651
Consumer 9	68	98	123
Consumer 10	256	569	432
Consumer 11	334	442	365
Consumer 12	786	835	998
Consumer 13	589	489	543
Consumer 14	489	478	467
Consumer 15	109	308	508
Consumer 16	982	820	880
Consumer 17	590	585	620
Consumer 18	680	589	591
Consumer 19	368	490	461
Consumer 20	678	873	690

Step 1:

Sort the population by GR\_CRC1, GR\_CRC2, GR\_CRC3, respectively, and assign them to four risk groups (i.e. 25 percent of the population per risk group); the results are shown by the following table:

25% For Each Risk Group Low Risk	Sorted by GR_CRC1		Sorted by GR_CRC2		Sorted by GR_CRC3
Consumer 16	982	Consumer 5	996	Consumer 12	998
Consumer 2	890	Consumer 2	981	Consumer 5	988
Consumer 5	890	Consumer 20	873	Consumer 16	880
Consumer 6	874	Consumer 12	835	Consumer 2	730
Consumer 12	786	Consumer 16	820	Consumer 20	690
<b>Medium Risk</b>					
Consumer 7	762	Consumer 4	761	Consumer 8	651
Consumer 1	739	Consumer 1	750	Consumer 4	638
Consumer 18	680	Consumer 18	589	Consumer 1	630
Consumer 20	678	Consumer 17	585	Consumer 17	620
Consumer 17	590	Consumer 10	569	Consumer 18	591
<b>High Risk</b>					
Consumer 13	589	Consumer 19	490	Consumer 6	569
Consumer 8	569	Consumer 13	489	Consumer 13	543
Consumer 14	489	Consumer 14	478	Consumer 15	508
Consumer 4	460	Consumer 7	475	Consumer 7	485
Consumer 19	368	Consumer 11	442	Consumer 14	467
<b>Very High Risk</b>					
Consumer 11	334	Consumer 6	379	Consumer 19	461
Consumer 10	256	Consumer 3	366	Consumer 10	432
Consumer 3	150	Consumer 8	345	Consumer 11	365
Consumer 15	109	Consumer 15	308	Consumer 3	233
Consumer 9	68	Consumer 9	98	Consumer 9	123



Step 2:

Simply count the number of consumers who are in the same risk group across all 3 CRCs.

For Low Risk, consumers numbered 2, 5, 12, 16 are in the low risk group for all 3 CRCs, so  $N_1=4$ ;

For Medium Risk, consumers numbered 1, 17, 18 are in the medium risk group for all 3 CRCs, so  $N_2=3$ ;

For High Risk, consumers numbered 13, 14 are in the high risk group for all 3 CRCs, so  $N_3=2$ ;

For Very High Risk, consumers numbered 3, 9 are in the very high risk group for all 3 CRCs, so  $N_4=2$ ;

Step 3:

Calculate the SCI by taking the ratio as percentage

$$SCI = (N_1 + N_2 + N_3 + N_4) / N = (4+3+2+2)/20=11/20=55\%.$$

**SCI Interpretation:** 55 percent of the population is consistently ranked in the same risk tier across the three CRCs.

This methodology provides a simple yet logical framework to assess the consistency of any score and consequently the exposure for a lender of inconsistent scores in their decision strategies.

## 5. Application

This methodology provides several valuable business frameworks for the lending industry.

**Product Assignment Consistency:** Utilizing a simple ‘4 primary tier’ framework, a score can be evaluated for its ability to consistently place a consumer in the appropriate product range given their credit risk profile. Tiers can be defined such that they reflect super prime, prime, near and sub-prime behavior. For example, the super-prime tier could be defined as the top 15 percent of the population, prime as the next 50 percent, near-prime as the next 15 percent and sub-prime as the final 10 percent.

**Pricing Assignment Consistency:** A secondary framework can be deployed **within** any of the above primary tiers to further evaluate the scores’ ability to consistently rank the consumer within a specific risk tier (e.g. high, medium, low risk) such that the appropriate pricing can be assigned. The secondary framework is essentially nested within the primary tier.

## 6. Sensitivity Test of SCI

As previously referenced, a framework design using **four** risk categories logically aligns with business lending strategy, since the majority of the lenders categorize their portfolio or prospects into four risk groups and formalize business strategies around that framework. Commonly-used terminology for the four tiers is Super-Prime, Prime, Near-Prime, and Sub-Prime.

The absolute definition of these risk groups (in terms of score cuts or population percentage breaks) varies for different lenders, and for different products. For example, the definition of Sub-Prime for a mortgage lender may be quite different from that of a credit card lender.

Therefore, it is useful to vary the population percentage breaks for the four tiers to understand the stability of the index. It is crucial that SCI exhibits good consistency and stability. In section 8, we provide four different scenarios and examine the corresponding SCI values.

## 7. Data

The data used here is a randomly selected sample with equal number of consumers from each CRC, satisfying the following two requirements: 1) all records exist in all three CRCs; 2) all of the records are scoreable by VantageScore and each generic risk score used for comparison. Additionally, to examine robustness of results over time, a sample was pulled from each of the following observation points: June 2003 and June 2004.

## 8. Results

The following table summarizes the key results of this study, providing four scenarios by using different percentage population breaks for the four risk categories.

Scenario 1 is equal breaks of 25 percent, such that Low Risk is 25 percent of the population, Medium Risk is 25 percent, High Risk is 25 percent and Very high Risk is 25 percent of the population. Scenario 2 reflects the fact that most consumers have good credit, with extremely good and extremely poor credit profile consumers in the tails. Scenario 3 reflects a distribution where the population size decreases from low to high risk, and Scenario 4 reflects the reverse of scenario 3. Of the 4 scenarios, scenario 2 is generally recognized as reflecting the US population credit profile distribution.

	Sample	Scenario	SCI VantageScore	SCI Reference Score	% Difference	% Lift
1	June, 2003	L: 25%, M: 25%, H: 25%, V: 25%	72%	51%	21%	41%
	June, 2004	L: 25%, M: 25%, H: 25%, V: 25%	74%	51%	22%	44%
2	June, 2003	L: 20%, M: 50%, H: 15%, V: 15%	75%	55%	20%	37%
	June, 2004	L: 20%, M: 50%, H: 15%, V: 15%	77%	55%	22%	40%
3	June, 2003	L: 40%, M: 30%, H: 20%, V: 10%	77%	59%	18%	30%
	June, 2004	L: 40%, M: 30%, H: 20%, V: 10%	78%	59%	19%	33%
4	June, 2003	L: 10%, M: 20%, H: 30%, V: 40%	73%	52%	21%	40%
	June, 2004	L: 10%, M: 20%, H: 30%, V: 40%	75%	52%	23%	44%

%Lift = VantageScore Score Consistency Index (SCI) improvement over Reference Score SCI

The four scenarios are intended to reflect a wide range of variations in risk group breaks, and the associated variations in SCI values and lifts. From the above table, we see that all SCI values are in the 70 percent range for VantageScore and in the 50 percent range for each generic risk score. **The percentage difference is on average 20 percent, and the lift is consistently over 30 percent.** Clearly there is a strong lift in score consistency by VantageScore over the referenced generic risk scores. This result holds consistently across four scenarios and two observation points.

## 9. Conclusion

Score consistency is increasingly relevant for well-managed real estate underwriting processes. As demonstrated, this methodology provides a quantitative framework for assessing algorithm consistency. The approach is robust and transparent and easily applied to any scoring algorithm.

SCI values show that VantageScore delivers 30 percent more consistency in its assessment of consumer risk than the CRC generic risk scores used for this comparison

(1) “Characteristic Leveling Process White Paper”: May, 2006. Online. Internet.

<http://www.vantagescore.com/about/whitepapers/leveling>

“Segmentation for Credit Based Delinquency Models White Paper”: May, 2006. Online. Internet.

<http://www.vantagescore.com/about/whitepapers/segmentation>