Understanding the Dun & Bradstreet SBFE Score

This document is intended to address the following questions:

- What is the D&B SBFE Score and what does it predict?
- What is the availability of the D&B SBFE Score?
- How was the D&B SBFE Score built?
- Adherence to regulatory statutes
- How does the D&B SBFE Score perform?
- What are the model's assumptions, limitations and oversight?
- Using the D&B SBFE Score
- What types of data elements are used as predictors in the D&B SBFE Score?
- What commentaries can be delivered with the score?

I. INTRODUCTION

On January 16, 2015, Dun & Bradstreet became the first SBFE Certified Vendor™ of the Small Business Financial Exchange (SBFE®) data. SBFE payment data provides state of the art clarity on how a financial institution's customers pay their obligations. As an SBFE Certified Vendor, Dun & Bradstreet gains access to the SBFE Data™ on more than 24 million small businesses. SBFE Data combined with Dun & Bradstreet's proprietary data sources and our world-class analytic capabilities provides our customers who are SBFE Members™ with more predictive power for their small business risk assessment. This enables improved transparency and profitability from a lender's small businesses portfolio.

Harnessing this power, Dun & Bradstreet's Advanced Analytical Services group built a new score that gives users unprecedented power to monitor and manage existing portfolios.

II. WHAT THE D&B SBFE SCORE PREDICTS?

The D&B SBFE Score predicts a business's likelihood of:

- Becoming Severely delinquent (91+ days past due) on any financial obligation in the 12 months subsequent to scoring
- Producing a charge-off on any of its financial accounts
- Filing for bankruptcy

The underlying models for the D&B SBFE Score are based upon the observed characteristics of more than a million business records in the Small Business Financial Exchange repository supplemented by additional elements in Dun & Bradstreet's trade, Firmographic and behavioral archives.

III. WHAT IS PRODUCED BY THE D&B SBFE SCORE?

A Score in the range of 706 – 999 is assigned to each business by the model where a higher score is associated with a lower risk of severe delinquency, charge-off or bankruptcy.

Up to 3 commentary messages are provided to give rationale for the score. When included, these messages will be listed in the order of their importance in the score calculation. A complete listing of these commentaries is included in appendix B of this document.

IV. AVAILABILITY OF THE D&B SBFE SCORE?

SBFE members have minimal data requirements to generate a score on a given set of accounts. You must supply basic business information about the small business; including business name, and address, and D-U-N-S® Number, if available. A D&B SBFE Score is generated for all D-U-N-S Numbered businesses, not just for those that match in the SBFE database, with only minimal exceptions as indicated below. As with most risk scores, if there is no archived information or supporting input data, then no score can be generated. For example, this can occur during a retro-score analysis, when the D-U-N-S Number did not exist as of the archive period. In this rare event, an exclusion code will be returned indicating why the record could not be scored.

In summary, a D&B SBFE Score is available on the vast majority of the 72 million U.S.-based businesses reported in the Dun & Bradstreet Data Cloud with the following exceptions and notations:

- D&B SBFE core will not be calculated if the D-U-N-S Number:
    - Is on "Stop-Distribution"
    - For which no data is available
    - Is a branch of a foreign-headquartered business
- A D&B SBFE Score of 0 indicates the business has been flagged as "High Risk"

V. MODEL DEVELOPMENT PROCESS

CONSTRUCTING THE MODEL DEVELOPMENT SAMPLE

This new SBFE-based Model utilizes the power of data from SBFE - in conjunction with attributes from D&B's CSAD (Commercial Score Archive Data), the highly granular DTRI (Detailed Trade Risk Insight) database and Dun & Bradstreet's specialized pool of Commercial Spending data. All the information contained within our database has passed through our DUNSRight Quality Process.

In the model development process, data is collected from two time periods designated as an observation window and a performance window. The observation window defines the sample used in the model and all identification and characteristic data are collected from this time period. The predictive variables and segmentation schemes are defined from this snapshot. The performance window defines the length of time the businesses in the sample are tracked to examine their performance.

The model development sample consisted of a random sample of 1.5 million business entities comprising over 2.6 million financial accounts and 83.2 million total trade lines. The records were selected from 4 snapshots in 2011 (Jan, Apr, Jul and Oct) and 1 in 2012 (Jan) and were monitored for the corresponding 12-month period subsequent to their selection.

For each business, the 12-month performance of all of the accounts in the combined data pools were reviewed and categorized as follows:

- Any account becoming 4 or more cycles past due (with dollar amounts above a chosen minimum threshold) during this period; including chargeoffs and bankruptcy were defined as "Bad"
- Any account not exhibiting any delinquency were defined as "Good"
- Accounts not meeting either the "Good" or "Bad" definition were categorized as "indeterminate"

For the purposes of model development, all "indeterminate" accounts were removed. Next, the roll up to the Obligor was accomplished by labeling an Obligor as "Bad" if any of the (remaining) accounts for that obligor was categorized as "Bad"; and "Good" if all the (remaining) accounts associated with that Obligor were "Good". An Obligor was not included in the development sample if all its accounts were "indeterminate."

DATA CLEANSING/VARIABLE TRANSFORMATION

The purpose of data cleansing is to identify incomplete, incorrect, or inaccurate records. The data transformation process builds the raw input data into meaningful attributes. Examples of meaningful attributes include Percent of satisfactory trades, Percent of slow and negative trades, and Paydex Score variance (over the most recent 12 months). Two-stages of the classing process are also part of the transformation, including both fine and coarse classing.

For "missing value imputation", we treat records with missing values as a separate group, as we create the bins for those records. The "weight-of-evidence" (WOE) is calculated for its own group, and the risk level depends on the WOE.

MODEL DEVELOPMENT APPROACH

Variables Reduction/Selection:

- Apply variables clustering method by different data sources to reduce the dimensions of predictive attributes (over 2500), to minimize the multi-collinearity effects
- Check the attributes' data coverage
- Assess the attributes' predictive power within each cluster
- Assess the attributes' stability index over time

Model-Build Methodology:

- Create and use weight of evidence (WOE) variables as predictive variables
- Build logistic regression models (scorecard-based[1])

MODEL DEVELOPMENT VARIABLE REDUCTION/SELECTION

From the observation window data, Dun & Bradstreet performed extensive data analysis to determine those variables that are statistically the most significant factors for predicting severe delinquency, charge-off and bankruptcy and calculate the appropriate weights for each. In performing this exploratory analysis, Dun & Bradstreet leveraged the rich SBFE Data as well as our commercial trade, firmographics and behavioral databases. Dun & Bradstreet identified and tested thousands of predictive variables from evaluating a combination of both "good" and "bad" performing businesses in the Dun & Bradstreet database.

1. The D&B SBFE model was developed using the credit risk scorecard approach that is standard for building credit risk models in both the consumer and commercial sectors. This development process is an adaptation of the process outlined by Naeem Siddiqi in his Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. This process addresses the fundamental drivers:
    - Do unique risk segments exist within the portfolio?
    - What are the important obligor risk characteristics within the portfolio and how are these characteristics measured?
    - Which characteristics are most critical to the overall obligor risk and what is the relative importance of each characteristic?
    - How do the critical risk characteristics relate to the obligors' likelihood of default?

The variable reduction/selection process for the SBFE Score included:

Single-Variable Logistic Regression: To assess the predictive power of each variable individually against the dependent variable. The GINI index and other statistics (K- S/Divergence) are typically used for assessment.

Clustering Analysis: Each cluster represents a different dimension of the available attributes from different data sources. The IV (information Value) of the variables in each cluster is examined.

Step-Wise (Forward and Backward) Linear Regressions: With focus on the Variance Inflation and Condition Index associated with each variable (the goal being to minimize multi- collinearity) within modeling process.

FORMULATION

Logistic regression is a type of modeling technique designed to model the relationship between a binary dependent variable and explanatory (or independent)

$$logit(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 * X_1 + \ldots + \beta_k * X_k$$

variables. It is a form of the generalized linear (GLM) and is given by:

Where 'P' is the probability of an observation taking on a particular value, and are the parameters associated with each explanatory variable. The logistic regression model uses the explanatory variables to predict the probability that the dependent variable takes on a given value. The logistic regression model assumes neither normally distributed error terms nor homoscedasticity, and it produces predictive probabilities that lie between 0 and 1.

Logistic regression was used to estimate parameters (that explain the relationship between the independent variables (X_i) and the binary dependent variable. Using the parameters estimated from the model and the values of the independent variables[2], the logistic regression model was used to calculate a score for each of the D-U-N-S predicting the likelihood of the D-U-N-S being bad. As a result, the lower the score, the more likely a client will become severely delinquent in the next 12 months.

2. Appendix A contains a categorization of the data elements used in the SBFE Scoring Model.

SEGMENTATION

The ability to accurately assess risk is dependent on the availability of robust underlying data elements, so Dun & Bradstreet has developed a scoring system that accounts for the correlation between depth of predictive data and future viability.

The result is a suite of models consisting of four unique scorecards. The scorecard selected for a given business is driven by the presence, utilization and turnover of credit card balances on a business. Each algorithm was developed and optimized on a more homogenous subpopulation to account for the amount of information contained in our database on the business and the difference is bad rates. The four models are:

- No Card/Missing Payment-to-Balance Ratio For Card(s): Bad Rate 1.99%
- "Transactors": Bad Rate 0.83%
- Low-Utilization Revolver (<40%) ["Revolvers"]: Bad Rate 1.75%
- High-Utilization Revolver (>40%) ["Parkers"]: Bad Rate 7.93%

Having a system of models allows for better separation of "goods" and "bads" by focusing on unique populations. It also provides for the most predictive score possible, optimized on the data available. The D&B SBFE Scoring Model, therefore, provides maximum risk discriminatory power with segmented scorecards for improved risk management decisions.

REGULATORY REQUIREMENT ADHERENCE

All variables used as either segmentation or predictor variables in the D&B SBFE Scoring model strictly adhere to the requirements of the Equal Credit Opportunity Act. Specifically, this includes the handling of all protected classes (as it applies to commercial entities) defined by the ECOA; including

- Gender
- Age
- Race
- Color
- Religion
- National Origin
- Marital Status
- Ethnic Group
- With/Without Children

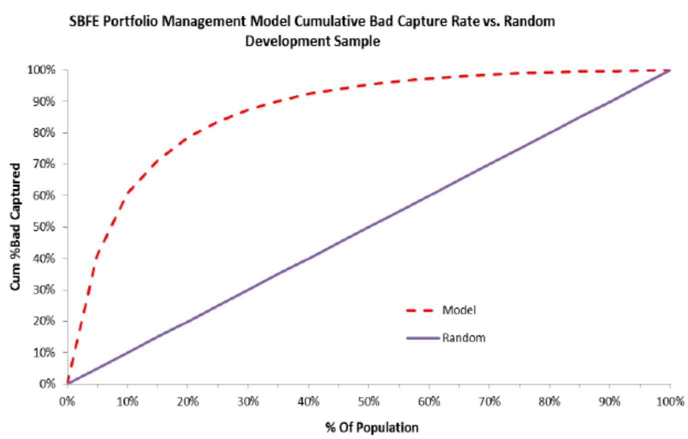
VI. MODEL PERFORMANCE?

PERFORMANCE MEASURED ON THE DEVELOPMENT SAMPLE

One way to measure model performance is by examining a trade-off curve. A trade-off curve is a plot of ascending accumulation of good accounts vs. bad accounts. It is useful for illustrating model performance both at a particular score and across the spectrum of score distribution.

The trade-off curve in Graph 1 illustrates the screening effectiveness of the D&B SBFE Score. For example, in the worse scoring 15% of the cumulative population, the models identify approximately 71% of the cumulative "bads". This means that by eliminating the worst scoring 15%, you would expect to capture or eliminate 71% of the "bads" in your portfolio.

Graph 1: D&B SBFE Score Performance on the Development Sample



During model development, various statistics from the development sample are gathered similar to the trade-off curve shown above. Development statistics provide useful information that can be used to help management determine policy related to the use of the models. For several reasons, however, statistics from model development should not be construed as precise forecasts for individual portfolios.

Summary Stats for Graph 1

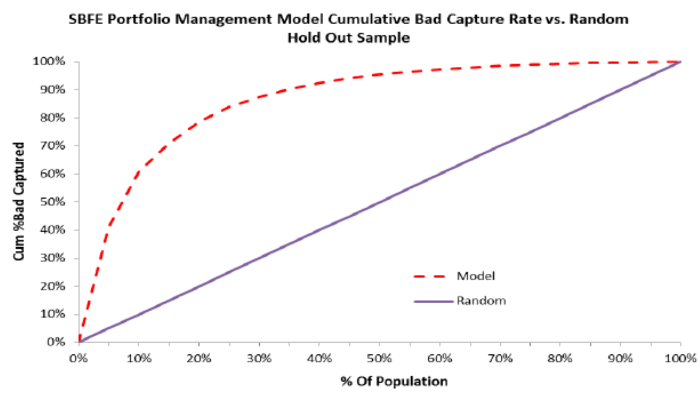
	SAMPLE DESCRIPTION
Number of Records	1,510,244
Number of Bads	37,062
Bad Rate	2.45%

	PERFORMANCE STATISTICS
KS	60.6
GINI	0.754
Area Under ROC	0.877

PERFORMANCE MEASURED ON THE HOLD-OUT SAMPLE

The trade-off curve in Graph 2 illustrates the screening effectiveness of the D&B SBFE Score as measured on the "Hold-Out" sample. For example, in the worse scoring 15% of the cumulative population, the models identify approximately 71% of the cumulative "bads". This means that by eliminating the worst scoring 15%, you would expect to capture or eliminate 71% of the "bads" in your portfolio.

Graph 2: D&B SBFE Score Performance on Hold-Out Sample



Summary Stats for Graph 2

	SAMPLE DESCRIPTION
Number of Records	1,221,844
Number of Bads	29,596
Bad Rate	2.42%

	PERFORMANCE STATISTICS
KS	60.4
GINI	0.754
Area Under ROC	0.877

Models are developed assuming that the relationships observed between past customers' characteristics and subsequent payment performance will hold true on future customers. Because of this assumption, development statistics should be viewed as estimates, and not precise forecasts of future performance at a given score.

Nevertheless, models are robust tools for rank-ordering risk in changing circumstances; higher scoring businesses perform better than lower scoring businesses. Tracking the score distributions and the actual performance of accounts provides the most accurate projections for individual portfolios.

USING THE D&B SBFE SCORE

The D&B SBFE Score is a transactional/batch score that is targeted for use within your portfolio monitoring and management strategies. Therefore, it facilitates more efficient and complete account management across the customer lifecycle. The D&B SBFE Score offers a competitive advantage because it is built upon the state of the art, high resolution SBFE payment database combined with power of the Data Cloud.

Incorporating the D&B SBFE Score in your portfolio management process allows you to retain/upsell the most profitable accounts. The D&B SBFE Score incorporates predictive data not previously available to existing portfolio management processes thereby adding invaluable insight into the process. Combining this invaluable insight with other predictive techniques strengthens your account management strategy. The D&B SBFE Score will provide the guidance you need to optimize your portfolio's profitability.

The chart on the next page illustrates how the D&B SBFE Score was distributed on the model development population as a function of score and the probability of a serious delinquency/charge-off/bankruptcy (a "bad") in each score range. Using the Good-to-Bad Odds and Bad Rate columns provides approximate metrics for accounts in each score range and facilitates the computation of approximate overall metrics for the portfolio. These computations can then be expanded to include strategic "what-if" scenarios that arise in managing the composition of the portfolio.

D&B SBFE Score Odds Chart & Population Distribution for Development Sample

SCORE RANGE	INTERVAL %	CUMULATIVE %	GOOD/BAD ODDS	BAD RATE
Below 740	0.04%	0.04%	0.57	63.70%
740-759	0.20%	0.24%	0.86	53.76%
760-779	0.52%	0.76%	1.56	39.12%
780-799	1.21%	1.97%	2.79	26.37%
800-819	2.57%	4.54%	5.65	15.05%
820-839	5.24%	9.78%	10.99	8.34%
840-859	8.82%	18.61%	20.75	4.60%
860-879	11.48%	30.09%	41.62	2.35%
880-899	13.13%	43.22%	92.34	1.07%
Above 900	56.78%	100.00%	357.16	0.28%

MODEL OVERSIGHT – ANNUAL VALIDATIONS

Annual validations of the D&B SBFE Score will be performed by Dun & Bradstreet's compliance team (a group that is independent of the model development team). These validations will verify the following:

- The model's predictive performance, as measured by the KS and GINI statistics, has not deteriorated significantly from time of development
- The distribution of model scores has not shifted significantly from the time of model development. (population stability)
- The distribution of the scored population on each of the model characteristics (predictor variables) has not shifted significantly. (characteristic analysis)

ASSUMPTIONS & LIMITATIONS

MODEL ASSUMPTIONS

In the D&B SBFE Score (as in all predictive models) the most important assumption is that the relationships that were found to exist between the set of predictor variables and the outcome (in this case – Severe Delinquency) at the time of model development continue to hold true during future periods when the model is called upon to deliver insight.

Moreover, a reasonable level of consistency between the population used to develop (or "train") the model and that which the model is used to score is assumed.

A more specific assumption: For business branch locations, the model will automatically produce results for the associated headquarter location.

MODEL LIMITATIONS

A D&B SBFE Score is available on more than 71 million of the 72 million U.S.-based companies. D&B SBFE Score is not available on businesses that fall into the following categories:

- Business records with a missing or invalid address.
- Branch records with a foreign headquarter location.
- Businesses labeled as "High-Risk"
- Businesses on "Stop Distribution" list

dun & bradstreet

Appendix A

Categorized Data Elements in the D&B SBFE Score

The variables underlying the suite of D&B SBFE Score risk models include data elements from the Small Business Financial Exchange repository, Dun & Bradstreet's vast trade archive, and Dun & Bradstreet's insight-rich firmographic and behavioral archives. Some of the categories that these data elements come from include:

- Total/Max Amount of Exposure
- Max Amount Past Due
- Payment-to-BalanceRatios
- Changes in Account Balances
- Usage and Status on Financial Accounts
- Recency of Payment Delinquency
- Firmographics
- Payment Experiences (Dun & Bradstreet trade archive)
- Spending Behavior Patterns
- Payment-to-Credit Limit Ratios
- Payment-to-BalanceRatios
- Guarantor Status On Term Loans

dun & bradstreet

Appendix B

Model Output Commentaries

Model Reason Code (Commentary) Associated With Score	Commentary Code
Amount past due	1
Average balance velocity on all revolving accounts in the last 36 months	2
Delinquent past or present credit obligation(s)	3
Highest amount 30 days past due on all financial accounts in the last 12 months	4
Highest available credit on all commercial credit card accounts in the current month	5
Highest available credit on all commercial credit card accounts in the last 12 months	6
Highest available credit on all commercial card accounts in the last 36 months	7
Highest available credit on all open revolving charge accounts in the current month	8
Highest balance velocity on all financial accounts in the current month	9
Highest credit on all financial accounts in the last 48 months	10
Highest amount 90 days past due on all revolving accounts in the last 48 months	11
Highest past due amount on all financial accounts in the current month	12
Highest past due amount on all financial accounts in the last 12 months	13
Highest past due amount on all revolving accounts in the last 48 months	14
Highest past due amount on all installment accounts in the last 36 months	15
Highest payment to credit limit ratio on all commercial credit card accounts in the current month	16
Highest payment to credit limit ratio on all commercial credit card accounts in the last 36 months	17
Highest payment to balance velocity ratio on all revolving accounts in the last 12 months	18
Highest proportion of accounts reported 31 or more days past due over the last four months	19
Highest proportion of payment to balance velocity on all revolving accounts in the current month	20
Highest ratio of past due accounts in the 4 months	21
Length of time in business under present management	22
Lowest balance velocity on all revolving accounts in the current month	23
Lowest balance velocity on all revolving accounts in the last 12 months	24
Lowest proportion of payment to balance on all commercial credit card accounts in the last 36 months	25
Lowest proportion of payment to balance on all commercial credit cards accounts in the last 48 months	26
Lowest proportion of payment to balance on all financial accounts in the current month	27
Lowest proportion of payment to balance on all financial accounts in the last 12 months	28
Lowest proportion of payment to balance on all financial accounts in the last 36 months	29
Weeks since the last purchase in the last 12 months	30
Minimum age of open commercial credit card account	31
Number of inquiries in the last 24 months	32

Model Reason Code (Commentary) Associated With Score	Commentary Code
Number of negative payment experiences	33
Number of satisfactory payment experiences	34
Percentage of satisfactory financial accounts in the current month	35
Presence of account financial payment that are 31 or more days past due	36
Proportion of amount that are 61 or more days past due	37
Proportion change of maximum balance over the last 12 months	38
Proportion of accounts 60 days past due	39
Proportion of all financial accounts with guarantor(s) in the last 12 months	40
Proportion of payments 60 days past due	41
Proportion of payment to balance on all commercial credit card accounts in the current month	42
Proportion of payment to balance on all financial accounts in the current month	43
Proportion of payment to balance on all commercial credit card accounts in the last 12 months	44
Proportion of satisfactory financial accounts in the last 12 months	45
Proportion of accounts that are 61 or more days past due	46
Proportion of slow outstanding balance	47
Proportion of total utilization on all financial accounts in the current month	48
Recency of accounts reported 31 or more days past due	49
Recency of delinquencies on all commercial credit card accounts	50
Recency of delinquencies on all financial accounts	51
Total amount owed in the most recent available month	52
Total available credit on all commercial credit card accounts in the current month	53
Total available credit on all financial accounts in the current month	54
Total number of accounts 31 or more days past due in the last quarter	55
Total number of financial accounts past due in the current month	56
Total number of past due accounts in the last 12 months	57
Total number of revolving accounts delinquent in the current month	58
Total past due amount on all financial accounts in the last 12 months	59
Total utilization on all financial accounts in the last 12 months	60
Total utilization on all revolving accounts in the current month	61
Actual employee figure not reported	62
Limited business activity signals reported in the past 12 months	63
Proportion of past due balances to total amount owing	64
Proportion of slow payment experiences to total number of payment experiences reported	65

Appendix C

Model Scorecard IDs

MODEL/ SCORECARD	SHORT DESCRITION	VALUE
No Card	Missing ratio payment balances for cards	001
Transactor	Transactor	002
Revolver	Revolver Utilization < 0.4	003
Parker	Revolver Utilization >= 0.4	004
No Fin Trade	No Finanical Obligations	005