Crowdsourcing Enhancement to the FRISK® Score

EXECUTIVE SUMMARY

CreditRiskMonitor continues to improve the information provided to our subscribers. This document describes our use of a new source of data – "crowdsourced" information – to improve the accuracy of the FRISK® score. As far as we know, CreditRiskMonitor is the first and only provider of financial risk analysis to employ this innovative method.

The new crowdsourced information is derived from the aggregate behavior of our subscribers as they use the CreditRiskMonitor website. This information is unique to CreditRiskMonitor. Our subscribers are highly influential in the daily commerce of some of the world's largest corporations, making either Credit or Procurement decisions affecting billions of dollars of purchase and sale transactions every month. This is also a sophisticated group of subscribers: 35% of the Fortune 1000 and many more of the largest public and private firms use CreditRiskMonitor services.

It is worth noting that Trade Credit is a far more significant source of business financing than many appreciate: according to the Federal Reserve, Trade Credit was the third-largest liability of U.S. non-financial corporations, approaching $2 Trillion in early 2016, after bonds (approximately $5 T) and loans (approximately $2.6 T). The importance of procurement departments for all business-to-business sales does not need explanation.

Our research shows that the usage pattern a typical subscriber exhibits when evaluating a business on our website is related to the level of concern which the subscriber has about that business. By analyzing this data across many subscribers, we find that the aggregate usage pattern provides additional information about the failure risk of that business. Including this unique crowdsourced information in the FRISK® score significantly enhances the performance of the score. Overall accuracy is back-tested at 96%, with important increases in scoring accuracy for the most risky companies.

Data in the new version of the FRISK® score

Lots of data, available for many years, is needed to build crowdsourcing based models. This is how the computer is trained to detect patterns which are indicative of good/bad financial performance for a business. CreditRiskMonitor is uniquely suited to leverage this type of crowdsourcing information for a number of reasons. Since the beginning, our website has been highly structured. Subscribers interested in a "deeper dive" on particular information go to specific webpages on the site. The usage crowdsourcing is not using the specific information contained on those pages about a particular company (some of this is used by other parts of the FRISK® score). The crowdsourcing part of the model only cares that users are going to particular web-pages. This structure enables our crowdsourcing method. Also, as part of our operational policy, CreditRiskMonitor has maintained detailed records of user sessions for many years, mainly to support our efforts at product improvement. Our record retention policy has made a great deal of data available. Most importantly, consider whose sessions we are tracking. It is our subscriber base of sophisticated users that makes the record of their actions on the website highly informative regarding the risk of companies. Because of these factors, CreditRiskMonitor is uniquely positioned in this industry to leverage “the wisdom of the crowd” for evaluating the financial risk of businesses.

Why would the activity of our users on our website yield information about the credit risk of a business?

The subscribers to the CreditRiskMonitor service are a special group. We have thousands of users, many working at the world’s largest corporations, who are responsible for protecting their companies from credit risk. Other subscribers, also at large corporations, are responsible for monitoring the ability of their companies’ vendors to meet their supply obligations. Both of these functions require an understanding of the financial health of the businesses they are examining. These sophisticated professionals rely on CreditRiskMonitor when evaluating the financial risk of publicly listed companies. They also make use of other sources of information, including some sources not available to the general public. Input from their industrial relationships, sales and service teams, operations units, and others will often influence their decisions. These subscribers can be viewed as “gatekeepers,” with significant influence on the amount of business their companies do with a particular customer or vendor.

To reach their decisions and support their recommendations for action, subscribers must view a variety of information shown on different pages of the CreditRiskMonitor website. Unlike most other credit data providers, the CreditRiskMonitor website displays a great deal of information in a highly structured presentation. Evaluating a particular company usually starts with the “snapshot” page for that business, which provides a high-level overview of the business and its level of risk. If a subscriber wants to understand the condition of a business in more detail, to find out why it has a particular FRISK® score or a particular rating, they can view other pages that provide more detailed information.

In the aggregate, we observe that subscribers exhibit a different set of click patterns depending on whether or not they are concerned about the financial strength of a business. For example, if they are not concerned, they may be satisfied with just the information on the snapshot page. In contrast, for businesses of concern, they may feel the need to gather significantly more information to satisfy themselves and their colleagues of the subject company's financial strength, to reach the right decision on that business. These two situations lead to very different activity patterns on the CreditRiskMonitor website, and it is this difference that the model cares about.

How does the FRISK® score achieve great performance?

One of the key reasons why the CreditRiskMonitor FRISK® score achieves a high degree of accuracy in predicting failure is that the score incorporates multiple data sources to estimate risk. Since its inception in 2007, the FRISK® score model has used three components to estimate the risk: stock market information, financial data and agency ratings. The model is designed to operate with whatever data is available among the three components to produce a risk estimate. Each of these components can be used to build a credit model by itself. Since they all contain information about failure risk, they tend to move together, but not exactly. The way that the model handles this divergence between the components is a key to the superior performance of the FRISK® score. Stock market data, financial statement data, and agency ratings each can bring a slightly different perspective on the risk of bankruptcy of a business. So, by combining these varied sources, the model performs better than any single source, while mitigating the shortcomings of each.

How does CreditRiskMonitor incorporate crowdsourced website usage information?

Although the crowdsourcing data component is very different than the other types of data used in the FRISK® score, incorporating this type of information into the model follows a very similar set of steps. The primary requirement is that we need a great deal of historical data to train the model. The structure of the CreditRiskMonitor website has been hierarchical since the beginning, giving us many years of subscriber activity data.

In building these types of models there is always a trade-off between using the most recent information vs. using a longer historical time horizon which includes many more examples of bankruptcy under more varied economic conditions. Another consideration is that CreditRiskMonitor is always working to improve the service by adding new sources of information, adding and changing the information on specific webpages. For all of these reasons, we limited the historical data for this new component of information to activity from 2010 onward. (The other components are back-tested using data from 2003 onward.)

Aggregate behavior is the key to success with crowdsourced information. It is only when a consistent behavior pattern is shown by many users that the pattern becomes valuable. So the availability of this type of information is inherently limited to those businesses that are of high interest to a significant fraction of our subscriber base.

With this enhancement to the FRISK® score, it would be accurate to say that our subscribers are now helping each other make better risk decisions – anonymously!

Like the previous versions of the FRISK® score, this model was trained using our data set of historical bankruptcies. The model's parameters were adjusted using standard statistical techniques to optimize its predictive ability. We employed the technique known as "static pool analysis," in which the full historical time period used to build the model is segmented into a set of 12-month pools. All relevant information that would have been available at the start of a "pool" is collected on a set of active companies (those that have not filed for bankruptcy previously), and those companies in the pool that failed (filed for bankruptcy) during the 12-month period of the pool are flagged. Using the data from all the pools, the training process identifies the website usage patterns that are most predictive of bankruptcy.

In this research we have found that it is not the number of "clicks," but rather the patterns exhibited that are important. What is important is the relative number of "clicks" by subscribers on some pages versus other pages. We found important information indicating financial stress in these aggregate patterns.

Once the model was built we used static pool analysis to evaluate its performance. Rather than gathering all relevant information at the beginning of each static pool, in this testing step we began with the model's initial score for each active company and then looked to see which of those companies failed during the pool period. Like the previous FRISK® score, two tests were done to validate the enhanced FRISK® score. One involved generating a set of ROC[1] curves. The other tested the model's ability to correctly classify businesses that do file for bankruptcy.

In their simplest form, credit models can be used in a binary fashion, where you choose a cut-off score threshold and classify all businesses as either "low-risk" or "high-risk." For any given score threshold level you can then measure two quantities known as a "false positive[2] rate" and a "true positive rate." In our case the false positive rate is the fraction of all scored companies that do not file for bankruptcy in the corresponding static pool, but were classified as high-risk. In any model, you want to keep this false positive rate as low as possible, because these are companies you want to do business with and not reject as "too risky." You may also have operational reasons to minimize the false positive rate, because after seeing a bad score you may delay decisions and spend additional resources evaluating an apparently risky company in detail.

[1] ROC stands for "Receiver Operating Characteristics" and the name comes from the age of analog communications where this technique was used to evaluate a receiver's ability to discriminate signal from noise.

[2] Ironically, a "positive" here is a bankruptcy; positive just means that the model correctly predicted the event (i.e., bankruptcy) you are looking for.

The true positive rate is the fraction of all companies that do file for bankruptcy in the corresponding static pool and which were correctly classified as high-risk at the beginning of the pool. In any model, this ratio should be maximized, for obvious reasons.

Varying the level of the score threshold for a particular credit scoring model, you will find different values for the false positive and true positive rates. If you choose a very low score threshold, you will find a low false positive rate (a good thing), but you will also find a low true positive rate (a very, very bad thing): you will miss warnings on most of the bankruptcies. At the opposite extreme, if you choose a very high score threshold you will find a high true positive rate, because if most companies in the group are classified as being high-risk you may correctly tag all the companies who file for bankruptcy. However, you will also find a very high false positive rate, because you are classifying most of the companies as high-risk: you will reject a lot of good business.

In both operating extremes any model has low economic value. However if you change your score threshold (or operating point: this is the "O" in ROC) for each value you get a different set of true and false positive rates. Each set of values define a point in the ROC graph, so as you sweep threshold values you generate an ROC curve. This process can be used to find the optimal operating point for a model. CreditRiskMonitor's use of a FRISK® = 5 as the threshold dividing high-/low-risk companies came from this kind of analysis.

In most cases, the enhanced FRISK® score has the same value as the previous FRISK® score version. There are several reasons for this; a major reason is that the businesses that are of keen interest to our users are a fraction of all the businesses scored by the FRISK® score, and these are the only ones which will have valuable crowdsourcing derived data. Another reason is that the performance of the FRISK® score just using the original three components (stock market data, financial information, and agency ratings) was already very good. Often, the website usage patterns picked up by the model reflects the same underlying information, so the addition of crowdsourced data does not change the FRISK® score value. Having said this, we find that, for a significant fraction of companies that are of keen interest to our subscribers, the crowdsourcing derived data does bring important additional information that is not fully reflected in the stock market, financial statements or ratings data.

The graph in Figure 1 shows the ROC curve for businesses which show a difference between the FRISK® scores derived from the enhanced FRISK® score model and the previous FRISK® score. The new model shows a significant performance increase for this population:

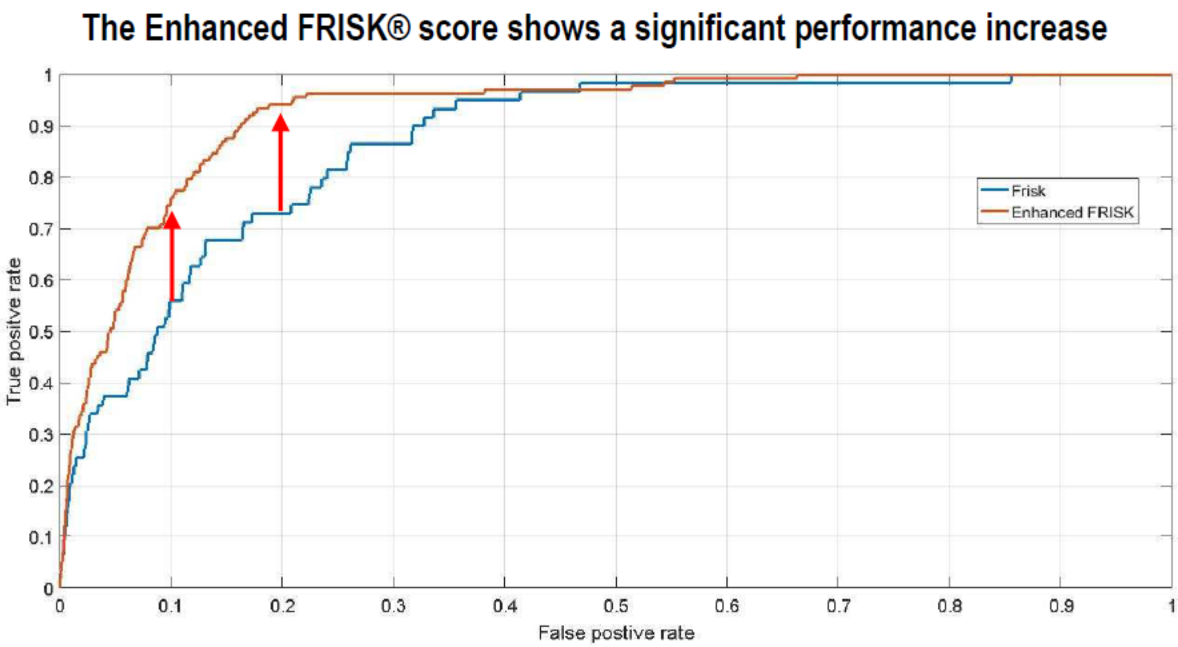

Figure 1 ROC curve for businesses for which the enhanced FRISK® score changes

A test that provides more clarity on how the crowdsourced data adds value to the FRISK® score can be created by plotting the bankrupt companies for each FRISK® score bucket. The graph in Figure 2 shows the FRISK® distribution for all bankrupt businesses in the validation data set. The graph is a cumulative plot; each bar represents the percentage of bankrupt businesses that received that FRISK® score, or a lower score, at the start of each static pool. The length of the bar for FRISK® = 5 is 96%, this means that 96% of all bankruptcies in the validation data received a FRISK® score of 5 or less. In other words, the model was able to provide early warning (flag as high-risk) in 96% of all public company bankruptcies in the United States. This may seem like a small improvement over the previous FRISK® model which achieved a value of 95%, but this one point improvement also means that 20% of the bankruptcies that were missed by the previous score are now being correctly classified as high-risk by the new score.

A more subtle, but equally important, improvement is that the new score moves companies correctly to new higher-risk buckets. As seen in Figure 2, the most dramatic changes between the enhanced FRISK® and its previous version shows itself in the lower (high-risk) buckets. This does not change significantly the total number of bankrupt businesses classified as high-risk, again it is a one point increase, but will move a number of high-risk business to higher-risk buckets. The number of bankruptcies being captured within the FRISK® = 1 bucket is 50% of all bankruptcies. This is a significant lift over the previous score which captured 35% of all bankruptcies in the highest-risk bucket.

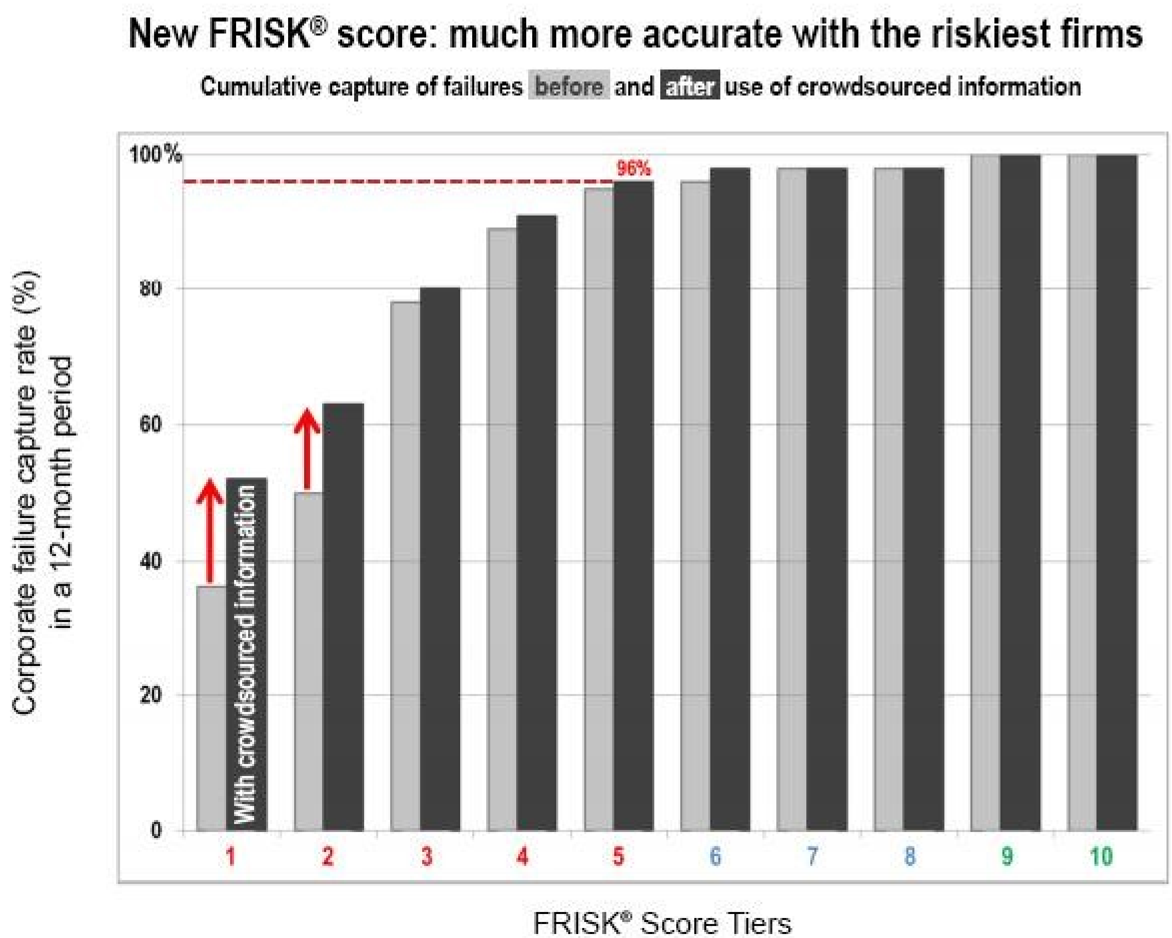

Figure 2 enhanced FRISK® score bucket distribution of bankrupt companies.

Conclusion

The new enhanced FRISK® score has been in operation since June 11, 2016. The new model leverages the aggregate website usage activity of our subscribers to provide a better estimate of bankruptcy risk for a small group of companies that are of keen interest to our subscribers. The enhancement only impacts the scores of U.S. businesses that are of interest to a significant fraction of our subscribers.

Note that the FRISK® scores produced by the new model will have the same relationship between each FRISK® score bucket and probability of bankruptcy as before. So, the meaning of a particular score is unchanged, and the score should be used exactly as it was before this enhancement was implemented.