Won't Get Fooled Again

by Joseph Breeden

In late 2006, when mortgage delinquency rates started to climb, many bankers said, "...but my credit scores haven't changed..." The idea was that if the scores were unchanged, the lender had done nothing wrong. This sentiment was repeated so frequently in the industry that FICO made a significant effort in research and conference presentations to promote the case that its scores still rank-ordered properly throughout the crisis.

FICO may be right about rankings, but if the scores were working, why were lenders so easily and frequently fooled? What was wrong? The answer is that the problem was in predicting probabilities, not in predicting rankings. Yes, the scores were working, but not the way lenders thought they were, and perhaps not to solve the problems lenders really needed solved.

IT'S PROBABLY ME

First of all, scores do not predict financial stress. Scores project future stress based upon observed past stress. To allow lenders to focus on the accounts currently exhibiting the most stress, they risk-rank the observed stress across many consumers. Consequently, consumer scores are "pro-cyclical"—they rise and fall on average through the economic cycle. As an indicator, they significantly lag the economic cycle. At the beginning of an economic downturn, the average consumer credit score will be better than the long-run average because consumers just lived through an easy economic period. At the beginning of a recovery, the average score will be worse than the long-run average because consumers just lived through a stressful period.

As a tracking and prediction tool, this is misleading. Good scores at the onset of a recession are not predicting that the consumers will perform well—simply that they have performed well so far. Poor scores at the start of a recovery are not predicting that the consumers will perform badly—only that they have performed badly previously.

The separate ranking of risk between consumers in these two periods may be accurate, but comparing scores across periods is misleading. To set policies, lenders need more than just a number for ranking. Changing credit lines, planning collections staffing and pricing future originations all require knowing the actual probability that the consumer will default on a loan, or generate the required level of revenue. Only probabilities can be ranked across time periods, product segments, etc.

In 2009, as consumer lending performance was deteriorating across all product types, credit risk managers often commented that "my 720s are performing like 660s!" This is not a comment about rank ordering, but, rather, the industry's failure to predict the probability of default based on a given score.

Lender Requirement 1: Predicted probabilities of consumer default, not just risk rankings.

SOMETIMES, "BAD" IS BAD

The limitations of traditional credit scores begin with the way they are built. When the problem is set up, accounts are separated into "bads" and "goods," based upon criteria such as whether they were seriously delinquent during an observation period. The missing element in this formulation is the environment. Is someone who defaults in a severe recession really the same as someone who defaults in an economic boom? When unemployment is at 10 percent and house prices have fallen 30 percent year-over-year, many people who were financially fragile will be pushed over the edge. However, for someone to default when unemployment is at 5 percent and house prices are rising at 10 percent year-over-year suggests much deeper problems. In the traditional approach, both situations would be labeled "bads" and the resulting model would be muddled.

The usual work-around for this problem is to frequently rebuild or recalibrate the scores. Although solving the short-term problem of being calibrated to the wrong environment, this solution fails on two points. The scores always miss turning points and will not be up-to-date until data is available for the new environment. Also, the training data is always short, so the scores never remember how consumers behaved in previous recessions or expansions.

Lender Requirement 2: Stable scores that do not need to be rebuilt every time the economy shifts.

YOU CAN'T ALWAYS GET WHAT YOU WANT...

When a score is built, the biggest assumption is that the distribution of consumers in the future will be the same as that upon which the score was built. If you said, "During the next phase of the economic cycle, few of the good consumers from your training set will want a loan," an analyst might rightly throw up his hands in discouragement. However, this is exactly what happens. Studies by Strategic Analytics have shown that, even when one normalizes for product type, risk band, loan-to-value, geography, etc., significant shifts in consumer appetite for credit occur through the economic cycle.[1]

Almost all scores in use today are based upon a specific set of explanatory factors provided during training. If the factor was not included in the training set, it will not be in the score. If the analyst did not know it was important before a change in consumer behavior, it will not be in the score to detect that change. The studies by Strategic Analytics have found that, when interest rates are rising or house prices are rising excessively, the fiscally conservative consumers pull out of the market, leaving the risk takers. Factors like these are not included in scores. In fact, they represent a self-selection on the part of the borrower that is not observable until the loan fails. This is called Macroeconomic Adverse Selection.

The lender is therefore always in the situation of using a score that sees only what it was built to see, but consumer dynamics are more complex. Much of the story of the U.S. mortgage crisis involves lenders booking loans that were risky in ways their scores could not see.

Lender Requirement 3: The ability to monitor score performance and quantify adverse selection quickly enough to adapt.

THERE MUST BE A WAY

Build scores that predict probabilities from the start

Tradition can be difficult to break, but industry experts today know how to factor-in problems. Survival models have been available for decades and are designed to predict monthly probabilities of an event. For the last decade, Strategic Analytics has been using an approach called Dual-time Dynamics (DtD) to predict the performance of retail loan portfolios. DtD scoring applies technology to loan-level forecasting or probabilities that include lifecycle, environmental and credit-quality drivers of performance.

These methods are part of a class of models called Nonlinear Decomposition. While it is more complicated to build than logistic regression scores, the return on investment is that one has a score capable of predicting the monthly probabilities of default under scenarios for economic changes and account management changes. As a score, they still rank-order as well or better than a traditional score, but the probabilities are available right from the initial design.

It is by changing the means of creating credit scores that we will be able to look beyond the numbers.

...consumer scores are "pro-cyclical"—they rise and fall on average through the economic cycle.

DtD scoring applies technology to loan-level forecasting or probabilities that include lifecycle, environmental and credit-quality drivers of performance.

Build scores that know the difference between a "bad" in a bad economy and a "bad" in a good economy.

When we employ a nonlinear decomposition approach for score building, such as DtD Scoring, we automatically normalize the score weights for shifts in the economy by incorporating the environment in the overall model. This has a useful side effect of allowing the score to be built on much longer data sets, and therefore to be stable through turning points in the economic environment. We switch from a backward-looking risk ranking to a forward-looking prediction.

Build scores that can identify Macroeconomic Adverse Selection

Scores do not have to be blind to adverse selection. In additional to the usual credit score factors, we can allow for vintage-specific factors that quantify whether the originations in a specific month, quarter or year were riskier than normal for unobserved reasons. This is easily incorporated in behavior scores. For origination scores, simply assuming that any recent adverse selection will continue into the near-future is much better than ignoring the effect.

YOU SAY YOU WANT A REVOLUTION

A revolution is underway in scoring. The mortgage crisis highlighted the failings inherent in traditional scoring approaches. Researchers at many universities are exploring ways to address the lender requirements listed above. All of the solutions being explored fall into the category of Nonlinear Decomposition, meaning simply that they are attempting to capture lifecycle, environment and credit quality within the score, where traditional approaches have ignored the first two.

The Dual-time Dynamics Scoring technology developed by Strategic Analytics is the first to market this new way of doing scoring. It can be applied as widely as traditional scoring, but with the added advantages of predicting probabilities, being robust through economic turning points and avoiding falling victim to adverse selection.

It is by changing the means of creating credit scores that we will be able to look beyond the numbers. TFL

[1] This research was published in Breeden, JL, L. Thomas, and JW MacDonald III, "Stress testing retail loan portfolios with dual-time dynamics", *Journal of Risk Model Validation*, Summer 2008, pp. 1–20, and summarized in *National Mortgage News* and Verisk Online.