Moving from Rankings to Ratings

Joseph L Breeden
Strategic Analytics Inc.
Santa Fe, NM 87505, USA
breeden@strategicanalytics.com
+1 505-995-4747

Robert Parker
Strategic Analytics Inc.
Santa Fe, NM 87505, USA
rparker@strategicanalytics.com
+1 505-995-4732

July 22, 2009

Abstract

Most standard logistic regression credit scores have continued to rank-order correctly through recent crises, but they have generally failed to predict the odds of default. Leveraging Dual-time Dynamics, we describe an application and behavior scoring methodology that rank-orders at least as well as logistic regression scores, but can also be used to predict the probability of default or odds of default given a scenario for the future macroeconomic environment. DtD scores can be used to predict the cumulative probabilities over a fixed horizon, as is commonly done, or can provide monthly default forecasts to aid in the timing of account acquisition and management.

DtD scores have several advantageous properties such as the ability to integrate with vintage-level modeling. In addition, the distribution of the scores is stable through the economic cycle, but the probability of default at each score value is predictable with the economy. Basel II stress testing is easier with this approach, as the impact of the macroeconomic environment is concentrated in a single model component.

Keywords: Dual-time Dynamics, Credit Scores, Behavior Scores, Retail Lending, Macroeconomic Scenarios, Basel II, Stress Testing

1 Introduction

The US Mortgage Crisis has brought widespread investigation into why credit scores employed by loan originators, portfolio managers, and traders of securitized pools failed to predict the unprecedented losses that ensued. Most credit score vendors have offered evidence that the scores properly rank-ordered accounts by credit risk. The subsequent step of calibrating credit scores to default odds is where the dramatic failures occurred.

The failure of both origination and behavior scores can be traced to an inability to incorporate scenarios for the future macroeconomic environment or recent trends in adverse selection into the prediction of default odds. Although many attempts have been made to create models that predict default odds given scores and macroeconomic factors as inputs, we believe the core problem stems from the method by which the scores are constructed.

Traditional logistic regression does not distinguish between an account that defaults in a bad environment versus an account that defaults in a good environment. In the US in 2009, when unemployment was rising rapidly and banks were reducing credit available to consumers, many consumers will default through no apparent fault of their own. Conversely, at the peak of the Dot-com boom, when unemployment rates were low and credit was widely available, consumers had fewer excuses to default on their debt.

Although default remains a binary variable, when creating a credit score we can estimate the weights while normalizing for the environment through which the consumers were living. Importantly for behavior scores, the model will also normalize for lifecycle effects according the age of the account. From the resulting model, we will be able to quantify adverse selection within a four to six months of loan origination and incorporate this information into loan pricing.

2 Methodology

The approach employed is an account-level version of Dual-time Dynamics (DtD) [2], and shares much in common with Proportional Hazards Behavior Scores [1, 7].

DtD as originally developed for portfolio modeling operates on vintage data to create scenario-based forecasting models for retail loan portfolios. Vintage performance is measured at regular intervals from the origination date. Default rate performance data of a vintage may be modeled as

$$DR(a,t,v) = e^{f_m(a)} e^{f_Q(v)} e^{f_g(t)} \tag{1}$$

where DR is the vintage default rate, $f_m(a)$ is the maturation function of months-on-book, a, $f_g(t)$ is the exogenous function of calendar date, t, and $f_q(v)$ is the quality function of vintage. We have the further relationship that $a = t - v$. The functions $f_m(a$, $f_g(t)$, and $f_q(v)$ are estimated from a set of vintages by assuming the relationship in equation 1 and solving for the unknown functions non-parametrically.

For any retail lending model, we need to address the intrinsic uncertainties that arise in the levels and trends of the three functions in Equation 1 [5]. To obtain unique solutions, we estimate these functions with the constraint that the mean of the exogenous and quality functions are zero and the combined trends are minimized.

For account-level modeling, the form of Equation 1 is largely unchanged, except that instead of a quality function of vintage (or marketing campaign), specific account-level attributes are included.

$$PD(a,t,i) = e^{f_m(a)} e^{B \cdot X(i)} e^{f_g(t)} \tag{2}$$

Instead of an aggregate default rate, we are now modeling an account-level probability of default, or more generically, probability of "bad", using whatever definition of bad is appropriate to the problem. The vector of attributes $X(i)$ are information about account i at time of origination, such as bureau attributes, application attributes, or attributes of the product offered. The vector of coefficients B scale the attributes such that $B \cdot X(i)$ becomes a time-independent, age-independent credit score. We carry forward the estimation constraint from Equation 1 that the average "score" be zero,

$$< B \cdot X(i) >_i = 0. \tag{3}$$

This score then takes a range from $-\infty$ to $+\infty$ centered on zero, but can be scaled to match the range of any typical score.

The maturation and exogenous functions in Equation 2 are identical to those in Equation 1 if both models are estimated over the same data set. In practice, longer vintage-level data is usually available, so these curves can be measured at the vintage-level first, and then included as constant normalization functions when estimating B at the account level. This is a significant advantage, because the longer maturation and exogenous curves allow us to better employ the score in support of forecasting, stress testing, and economic capital applications.

The coefficients B and maturation and exogenous functions are estimated via a maximum likelihood framework with a binomial probability distribution for both the vintage-level and account-level models. Although the vintage-level aggregated data requires a different estimation function from the account-level model, the resulting curves are equivalent.

We know that maturation curves will change from peaking high and early for subprime to peaking lower and later for prime accounts, meaning that maturation and score are not truly independent. However, the maturation curve changes slowly with score, so a segmented approach is sufficient to avoid the complication of combining the maturation and score functions. The segmented modeling approach is equivalent to what is done in vintage-level analysis, where three to five credit risk segments are usually employed to capture changes in the maturation function.

The exogenous function can vary with geography, so if enough data is present, segmentation along metropolitan or regional lines is appropriate.

The approach described here is very similar to proportional hazards scores. The primary difference is in the use of a nonparametric exogenous function rather than a matrix of specific macroeconomic factors. When creating a score from portfolio data, we assume that some of the calendar-based impacts are actually caused by management actions rather than macroeconomic factors [4]. A nonparametric function allows us to account for those factors without requiring a management log to specify the impacts. Such logs rarely exist in a form useful for modeling.

The nonparametric function also allows us to avoid the problem of trying to create a stress test model from a short time interval. Although such models can be successfully built [6], even a five-year data set is typically insufficient for creating anything better than an order-of-magnitude econometric model.

3 Implementation

Regardless of the form used to capture the calendar-based environmental impacts, both proportional hazards scores and the Dual-time Dynamics Scores (DtD Scores) share some important properties. The most important is that the score component of the model, $s_i = B \cdot X(i)$ can be implemented within existing account management systems without significant modification. S_i is the rank-order component of Equation 2 and will differ from a standard score only in the weights assigned.

For logistic regression scores, the score distribution for a portfolio will drift as the economy changes. The distribution of the s_i component of DtD Scores or proportional hazards scores does not drift with changes in the economy, because any systematic drift is captured in the exogenous curve. This will be an extremely useful property when using these scores for stress testing or economic capital calculations.

Although the distribution of s_i will not drift, the probability of default will change with the economy, but in a predictable way given either the current environment or a scenario for the future environment. To properly price a loan, we want to know the lifetime cumulative probability of default. This can be written simply as

$$CDP(s,t) = e^s \sum_{a=1}^{N} e^{f_m(a)} e^{f_g(t_0+a)} \tag{4}$$

where N is the number of months we want to consider as the lifetime of the loan. $e_g^f(t_0 + a)$ is the needed scenario for the future of the exogenous curve, which could be the most recent environment, but is best created as the same macroeconomic scenario being used for forecasting through the rest of the bank, relaxing back to a long-run average environment by 18 months forward and continuing flat from there. t_0 is the starting month of the new origination campaign.

The cumulative default probability, $CDP(s)$ can be used as an input to a pricing model.

Many institutions seek to set a cutoff score that maintains a maximum acceptable level for the default odds. Within the current framework, the cumulative default odds are simply

$$CDO(s,t) = \frac{CDP(s,t)}{1 - CDP(s,t)} \tag{5}$$

The desired cutoff score this month $cs(t)$ to maintain a maximum CDO of z is

$$cs(t) = CDO^{-1}(z,t) \tag{6}$$

As the scenario for the future of the exogenous curve changes, so will the desired cutoff score. Therefore, the cutoff score is a function of time and thus different for each campaign through an economic cycle.

4 Adverse selection

No score is perfect, and, as of yet, no score has been able to capture the changes in consumer appetite through the economic cycle. The structure of Equation 2 is not the problem, but few scores are built with attributes than can measure the consumers propensity to take risks if they have been successful in the past. In other words, a successful investor could be smart or lucky, but their investment track record cannot distinguish the two. The same problem exists in scoring.

Thus, we can expect that any score will under-predict or over-predict the intrinsic consumer risk at times. Further, we have seen from the US Mortgage Crisis that periods of adverse selection have a significant autocorrelation from campaign to campaign. The errors in predicting the true credit risk are referred to as adverse selection. The autocorrelation in adverse selection means that we can incorporate this into $CDO(s,t)$, even if we can only explain it intuitively.

To measure adverse selection, it can be viewed as the different between the credit quality that was predicting by the score and the credit quality that occurred. We do not want to confuse this with picking the wrong macroeconomic scenario. Scenario errors are in the control of the portfolio manager. Adverse selection is when the observed credit quality diverges from expected, even when normalizing for the environment.

We can compute the aggregate adverse selection for the originations in month v from Equations 1 and 2 as

$$AS(v) = f_Q(v) - \frac{1}{N}\sum_{i=1}^{N} B \cdot X(i) \tag{7}$$

If we assume that it takes four to six months to obtain a usable estimate of the adverse selection $AS(v)$, and if the autocorrelation of $AS(v)$ is positive for lags at least thet long, then we can augment the cumulative default probability model as

$$CDP_AS(s,t) = e^s e^{\overline{AS}} \sum_{a=1}^{N} e^{f_m(a)} e^{f_g(t_0+a)} \tag{8}$$

where $\overline{AS}$ could be computed as the average adverse selection for campaigns four to seven months old, or similar calculation.

The CDP formula adjusted for adverse selection, CDP_AS, would be used in the same way shown before for computing odds and setting pricing or cutoff scores.

5 Behavior Scores

Behavior scores differ from origination scores only in the allowance of time varying account attributes,

$$PD(a,t,i) = e^{f_m(a)} e^{B \cdot X(i,t)} e^{f_g(t)}. \tag{9}$$

In theory, we could now have a model specification error between $B \cdot X(i,t)$ and $f_g(t)$, since both are time varying quantities, but whereas earlier we observed that the origination score distribution would be stable with time, now we impose as a constraint that the mean of the score distribution be stable at zero through time. Practically, this constraint is easily maintained, because under this constraint, the original DtD model, Equation 1, the DtD Origination Score, Equation 2, and the DtD Behavior Score, Equation 9, all have exactly the same maturation and exogenous functions. We could choose to estimate these curves on vintage-level portfolio performance data at the start, and then hold those curves constant when creating either the originations scores or behavior scores. This has the effect of maintaining the constraints described at the start without requiring any complexity in during construction of the scores.

Creating a behavior score whose distribution does not vary with time has important practical implications. Through a recession, economic impacts will be observed to rise via the exogenous curve, and the odds prediction via Equation 5 will take this into account. The score component, $s_i = B \cdot X(i,t)$, will continue to rank-order accounts, so the score does not need to be rebuilt or recalibrated every time the economy enters a new phase. Importantly, during the 2009 global recession, retail loan portfolios globally are curtailing lending and cutting credit lines. Scores typically punish consumers for having high utilization rates on revolving credit products, so when lenders cut credit lines universally as is being done in 2009, consumer credit scores deteriorate even though consumers have not changed their behavior. With DtD Behavior Scoring, this change in management policy and its implications for odds of default are fully captured in the exogenous curve as an overall environmental impact. Individual consumer scores do not deteriorate unless the lender selects specific segments of consumers. Thus, even in the face of such dramatic changes, the scores will continue to rank-order and the default odds prediction will maintain fidelity.

6 Validation

DtD Origination and Behavior Scores, s_i, can be validated via all the usual methods for testing a score's ability to rank-order. K-S, Gini, and ROC statistics are a few of the standard measures for measuring a score's strength in rank-ordering.

Because of the temporal aspect of Equations 2 and 9, we can also apply validation measures appropriate for portfolio forecasting models [3]. Since our ultimate goal is to predict the probability or odds of default, we can conduct ideal scenario validation (ISV) to measure model accuracy and old account / new account tests to measure model stability.

7 Conclusions

The primary goal of creating DtD scores was to go beyond scores that rank-order to a system that can also predict the probability or odds of default. Alternate approaches have struggled in the past, because the data sets tested were too short to calibrate sensitivities to macroeconomic factors. By using a nonparametric environmental function in DtD scoring, we can normalize for changes in the macroeconomic environment even when the available data is too short to robustly include macroeconomic factors.

DtD Scoring also shows several advantages in being able to integrate with vintage-level modeling in order to augment the analysis with longer performance histories. The short account-level data sets that are common in scoring can be augmented with portfolio performance information to better estimate the maturation and exogenous curves, thereby allowing the scores to be better normalized for changes in the environment.

For stress testing and Basel II, having a score distribution that does not drift with time is extremely powerful. With standard behavior scores, as the portfolio lives through a recession, the score distribution will drift and the probability of default versus score will deteriorate. The implication is that to stress test properly, one must predict both the shift in the distribution and the deterioration in probability of default at each score. With DtD scores, individual accounts will change their ranking in the population, but the distribution overall is stable, meaning that stress testing, forecasting, or capital calculation can all focus on creating scenarios for the future of the exogenous curve, for which we have extensive experience. Macroeconomic impacts on the portfolio are concentrated in a single function, so sensitivities to macroeconomic factors are easier to calibrate.

References

[1] T. Bellotti and J.N. Crook. Credit scoring with macroeconomic variables using survival analysis. online, 2008.

[2] Joseph L. Breeden. Modeling data with multiple time dimensions. *Computational Statistics & Data Analysis*, 51:4761 – 4785, May 17 2007.

[3] Joseph L. Breeden and David Ingram. Monte carlo scenario generation for retail loan portfolios, 2009.

[4] Joseph L. Breeden and Lyn C. Thomas. The relationship between default and economic cycle for retail portfolios across countries: identifying the drivers of economic downturn. *Journal of Risk Model Validation*, 2(3):11 – 44, 2008.

[5] Joseph L. Breeden and Lyn C. Thomas. Model specification errors in stress testing models. In *Credit Scoring and Credit Control XI Conference August 2009, Edinburgh*, 2009.

[6] Joseph L. Breeden, Lyn C. Thomas, and John McDonald III. Stress testing retail loan portfolios with dual-time dynamics. *Journal of Risk Model Validation*, 2(2):43 – 62, 2008.

[7] M. Malik mand L.C.Thomas. Modelling credit risk of portfolios of consumer loans. online, 2008.