Cross-Tab Weighting for Retail and Small-Business Scorecards in Developing Markets

Dean Caire (DAI Europe) and Mark Schreiner (Microfinance Risk Management L.L.C.)

August 24, 2011

Abstract
This paper presents "cross-tab weighting", a simple technique for building credit-scoring models. Since 2008, we have been using this method to build application credit-scoring models for retail and small business segments with predictive power that compares well with more popular—and more complex—methods such as Logistic Regression. More importantly to our work in developing markets, cross-tab weighting facilitates the transfer of skills for model development and maintenance to bankers with limited experience and knowledge of credit scoring and/or statistics.

We discuss empirical results for three portfolios:
1. "Group loans" to micro-borrowers in Tajikistan
2. Overdraft loans to small businesses in Bulgaria
3. Microloans in Latin America

As long as the relevant population is stable, cross-tab weights can be refreshed using recent performance data with little additional work.

Introduction

This paper presents a simple technique for building credit-scoring models that we call "cross-tab weighting". The point weights in the scorecards come directly from cross tabulations (or "cross tabs") of each individual risk factor with loan status ("good" or "bad"). Most importantly, cross-tabulation is the only concept required to build and understand the models.

We have used this method in our consulting practice since 2008 to build application credit-scoring models for retail and small business segments with predictive power (measured by "AUC") that compares well with more popular—and more complex—methods such as Logistic Regression. More importantly to our work in developing markets, cross-tab weighting facilitates the transfer of skills for model development and maintenance to bankers with limited experience and knowledge of credit scoring and/or statistics.

In this paper, we explain how to build and evaluate cross-tab weighted scorecards and present examples of results for three portfolios from developing markets. We purposely use as little math as possible to facilitate communication of the key concepts to professionals with a wide range of specialties and educational backgrounds.

Dean Caire is a Risk Management Consultant for DAI Europe (http://www.dai-europe.com)
Mark Schreiner is the Director of Microfinance Risk Management L.L.C. (www.microfinance.com)

Cross-Tab Analysis

A cross tab (also known as a contingency-table analysis) is for our purposes a two-dimensional table that shows the number of good and bad contracts for clients with different risk characteristics.

As for all credit-risk modeling, the first step in preparing cross tabs is to define a "bad" loan. The definition varies by financial institution, but it should describe a loan the financial institution would prefer to avoid making in the future. For example, a "bad" loan for a bank is often one that reaches 90 days of arrears. In contrast, a microlender may define "bad" more conservatively as a loan that reaches 30 days of arrears. For any given set of portfolio data, we define "bad" and then assign each loan a good/bad status.

Table 1: Cross Tab: Gender of Borrower

A		Women	Men	TOTAL
B	# Goods	52	40	92
C	# Bads	3	5	8
D	% Bad	5.5%	11.1%	8%
E	# Total	55	45	100
F	**% Total**	**55%**	**45%**	**100%**

Table 1 is a cross-tab for the risk characteristic "Gender of Borrower" in which:

Row A labels the risk characteristics, which we will refer to from here onward as "bins" – for this factor, the two bins are "Women" and "Men".

Row B is the number of "good" contracts in each bin.

R**ow C** is the number of "bad" contracts in each bin.

Row D is the "bad rate", the focus of our analysis. For a given bin, it is defined as the number of bads (row C) divided by the number of total contracts (rows B and C). A higher bad rate indicates higher risk.

R**ow E** is the total number of contracts (good + bad) in each bin.

Row F is the number of total contracts in each bin as a percentage of all contracts. This is used to understand the distribution of contracts across/among bins.

Cross-Tab Weighting

Cross-tab weighting simply means using the bad rate for a given bin as its points in a scorecard. For the example in Table 1, a male borrower would receive 11.1 points and a female borrower would receive 5.5 points, because in the model development data, 11.1 percent of contracts with men were bad and 5.5 percent of contracts with women were bad.

We could also use the data in the cross-tab to express the odds ratio per bin, but to keep things simple, in practice we focus only on bad rates and the differences in bad rates across bins.

Building a Model with Cross-Tab Weighting

Single-Factor Analysis

The first step is to create and analyze cross-tabs of good/bad status with each potential risk factor.

For categorical factors such as gender, industry, legal form, etc., this process has been illustrated in Table 1. For numeric factors such as age, years in business, or financial factors such as income, indebtedness, etc., we sort the values in our data and put them into five bins so that 20% of observations are in each bin.[1] We examine these "quintiles" for common-sense patterns of change in the bad rate, such as non-reversing increases or decreases. Then we use judgment, usually the combined knowledge and experience of a working group, to adjust the bin cut-off points so that the trends make sense and fit with the expectations of local bankers. In this way, the models are both intuitively appealing and avoid over-fitting the development data.

By looking at the differences in the bad rates and the distribution of contracts across bins in each cross tab, we can get a good idea of a factor's predictive power. All else constant, greater differences in the bad rates and a more-even distribution of contracts across bins indicate greater power. If the bad rates are nearly identical for most bins, or if the vast majority of contracts are clumped in a single bin, then the factor will be less powerful. We also evaluate the predictive power of each one-factor "cross-tab model" with the Area Under the Curve (AUC) statistic[2]. This statistic is the area under the Receiving Operating Characteristic (ROC) curve, and it is also a simple function of the Gini coefficient.

When using cross-tabs to model a factor, we balance the goals of maximizing predictive power with ensuring that the relationships in the scorecard make sense to business users.

Building the Multi-Factor Model

Once we have a set of one-factor cross-tabs, we can begin to build the multi-factor model. Our two overriding goals in this stage are:

1. To choose a set of factors that form a comprehensive risk profile covering what banker's call "the Five Cs of Credit" (Cash Flow, Collateral, Capital, Character and Conditions) that make up traditional subjective/manual underwriting
2. To maximize predictive power

We start by choosing the factor that looks the best based on the two goals above. With only one factor, the model's AUC will be the first factor's AUC. Then we add a second factor, again based on the goals above, and look at the AUC of the two-factor model. It should be higher, and it normally will be considerably higher unless the two factors contain similar information. We continue

[1] We use the cut2() function in the library "Hmisc" and the CrossTable() function in the library "gmodels" of the open source software "R" – but such cross tables can be generated in any statistical software or even in Excel.

[2] An explanation of AUC can be found on http://en.wikipedia.org/wiki/Receiver_operating_characteristic. We calculate it using the colAUC function in the library "caTools" in "R" software.

choosing factors in this way to assemble a group of factors that create a comprehensive and powerful risk profile. Usually, the combined factor model AUC will rise quickly as we add the first 5-10 variables and then begin to taper off as we add additional factors due to the common scoring phenomenon known as the "flat maximum" We do not recommend any numeric rules for how many factors to include, only adherence to our two goals stated above.

Given the cross-tab weights—that is, weights set equal to the bad rates in each bin for each risk factor—the total points (the "score") for a borrower is the sum of the points received for each risk factor in the scorecard. We can then rank the borrowers by score, where lower scores indicate lower risk. For risk-management purposes, we create a number of risk groups (usually 7 or more) by dividing the scores into equal intervals. Given the risk groups and total scores, we can evaluate model performance with a "Good/Bad Table". This is just one more cross tab between loan status (good/bad) and risk groups based on scores. We present such a cross tab in Table 2, the results for a "group loan" scorecard for a microlender in Tadjikistan (discussed further below).

Table 2: Cross Tabulation: Multi-Factor Model

A	Risk Group	1	2	3	4	5	6	7	Total
	Points in Range	9.2 - 11.7	11.8 - 14.2	14.3 - 16.8	16.9 - 19.3	19.4 - 21.9	22.0 - 24.4	24.5 – 27.0	
B	# Goods	1,694	11,850	4,937	1,906	1,216	161	10	21,774
C	# Bads	1	35	32	37	96	25	3	229
D	Bad Rate	0.1%	0.3%	0.6%	1.9%	7.3%	13.4%	23.1%	1.1%
E	# Total	1,695	11,885	4,969	1,943	1,312	186	13	22,003
F	%Total	7.7%	54.0%	22.6%	8.8%	6.0%	0.8%	0.1%	100.0%

To sum up, we teach our clients how to use cross tabs to build complete scorecards. We also follow other standard modelling practices, such as breaking the data into development and validation sets. Our point here, however, is to emphasize that novice users from various backgrounds can easily learn to create, interpret, and manipulate cross tabs to create models that are nearly as powerful as more complex Logistic regression models. In the next section, we compare the power of these two techniques using three examples.

Cross-Tab Weighting Versus Logistic Regression

We have analyzed many sets of data using both cross tabs and logistic regression, and we consistently find that—given a set of factors—the logistic regression model is slightly more powerful. Given the overall similar performance, however, we have opted to promote cross-tab weighting to our clients purely for its practical advantages:

- Ease of interpretation of points
- Ability to "refresh" points using population bad rates from a more recent observation period (given a stable population)

Of course, we do not advocate cross-tab weights in all cases. As shown below, Logit models are generally more powerful and would not create any particular technical issues for a financial institution with an experienced modeller. We use cross-tab weights in our work in developing markets where limited resources place a premium on simple, powerful models that are easy to understand and maintain.

Empirical Examples

1. “Group loans” to micro-borrowers in Tajikistan

The first example was developed with data on 31,433 contracts that were issued and fully repaid (or defaulted) in the previous two years. The loans were issued to individuals in small groups who mutually provided each other with a joint-liability guarantee. Out of all contracts, 327, or just over 1%, were bad, defined as having reached 30 days in arrears. For model development, we randomly sampled (without replacement) 70% of goods and 70% of bads, using the remaining 30% for out-of-sample validation.

We made cross-tabs for each potential risk factor together with a working group of the microlender’s business, credit-risk, and database managers. This group selected the eventual 14 factors in the scorecard by adding variables one at a time based on subjective preference (expertise) and based on each factor’s additional contribution to AUC.

The 14 factors[3] give a comprehensive view of risk across the same key dimensions as the lender used to make subjective/manual lending decisions:

- Loan terms (3 factors)
- Borrower demographics (6 factors)
- Borrower credit history (2 factors)
- Financial ability to repay (3 factors)

The Logit model that we compare with the cross-tab weighted model uses the same binned variables created for the cross-tab weighted model.

Out-of-sample, the Logit’s AUC of 0.85 is about 9 percent higher than the cross-tab weighted model’s AUC of 0.78. In this case, performance favors the Logit model *if predictive power were all that mattered.* As we continue to emphasize, however, the simplicity of the cross-tab approach helps the client to understand the model and to monitor its performance over time.

A caveat on this Tajik example is that with only 1 percent of the sample being “bad”, drawing a different 70% development sample can lead to significant changes in the bad rates for some bins and factors. Nevertheless, the general trends hold. Furthermore, the lender’s working group understands this because it is easy for them to study and comprehend “sampling variation” by setting cross-tab weights for one 70% sample and then comparing them to cross-tab weights in a second 70% sample.

[3] We do not describe the individual factors and weights in order to focus on our main point – that the performance of simple cross-tab models is similar to that of Logit models.

2. Overdraft loans to small businesses in Bulgaria

The second example uses data from an SME lender in Bulgaria. In 2006, we helped this bank to develop and test an "expert" scorecard to support decision-making for overdraft loans to small businesses.

In 2010, we reviewed this scorecard using performance data from January 2007 to November 2010. This data set has 1,434 contracts, of which 140 (9.7 percent) were bad, or reached 90 days of arrears.

Given the existence of an expert scorecard, the first step was to compare bad rates in the performance data with the relative risk predicted by the expert scorecard. In creating a new cross-tab weighted scorecard for this bank, we adjusted the bins based on patterns that the working group confirmed as realistic, even in cases where the patterns differed from those implied by the expert scorecard. Equally importantly, however, we did not adjust bins if the patterns in the data could not be reasonably explained. We then set points based either on the actual bad rates in the data (per the cross-tab method), or, where the data appeared "idiosyncratic", based on the risk relationships originally expected by expert judgment but with adjustments to the point scale to fit the 9.7-percent bad rate in the performance data.

This models' 17 factors form a comprehensive risk profile similar to the one the bank used in subjective/manual decision-making:

- Loan terms (2 factors)
- Non-financial factors (3 factors)
- Borrower credit history (2 factors)
- Financial ability to repay (5 factors)
- Risk mitigation (5 factors)

Applied to all the 1,434 contracts in the performance data set, the original expert scorecard had an out-of-sample AUC of 0.70. After using a 70-percent sample of goods and bads from the performance data to transform the expert scorecard into a cross-tab weighted scorecard, the AUC for the remaining 30% of contracts was 0.82. A Logit regression of these same factors had a slightly higher AUC of 0.85 on the hold-out sample.

3. A Microlender in Latin America

In our Latin American example, the data set had 14,039 contracts, of which 2,187 (15.5 percent), were bad because they reached 30 days of arrears. As usual, we divided the data 70/30 for development and validation.

The model profiles borrower risk with 37 factors including:

- Loan terms (6 factors)
- Borrower demographics (18 factors)

- Borrower credit history (8 factors)
- Financial ability to repay (5 factors)

The Logit model uses the same bins as the cross-tab model. The Logit model is slightly more powerful with an out-of-sample AUC of 0.71 compared to the cross-tab weighted AUC of 0.69.

Table 3: Summary of Empirical Examples

Sample	Model	Tajikistan	Bulgaria	Latin America
Development (70%)	Cross-tab AUC	0.83	0.82	0.73
	Logit AUC	0.87	0.85	0.75
	Difference	5%	4%	3%
Validation (30%)	Cross-tab AUC	0.78	0.82	0.69
	Logit AUC	0.85	0.83	0.71
	Difference	9%	1%	3%

As Table 3 summarizes, the cross-tab and Logit models in our three examples have similar decreases in AUC between development and validation samples and the Logit models are consistently more powerful, with AUC between 1 and 9 percent higher. However, these three lenders use the models for decision support (as opposed to “auto-decisioning”) and portfolio risk management, so the slightly lower AUC of the cross-tab models is unlikely to materially negatively impact the business. Given the similarity of performance, we prefer the simplicity and transparency of cross-tab weighting.

Conclusion

Cross-tab weighted scorecards for retail and small business segments have predictive power that compares well with Logit. The simplicity of the cross-tab method facilitates the transfer of skills for model development and maintenance to bankers with limited experience and knowledge of credit scoring and/or statistics. Given a stable population, cross-tab weights can be periodically refreshed as new data becomes available without the need to redevelop the models from scratch.