Empirical Credit Scoring Model: User's Guide

November 2009

Prepared by:

PORTFOLIO DEFENSE

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The purpose of this document is to provide you with key information about your Custom Empirical Credit Scoring Model. In it, we review how the models were developed, how they can best be implemented and used, certain regulatory requirements associated with the use of scoring models, as well as recommended actions your organization will likely want to take for portfolio and model tracking.

While we attempt to anticipate and answer questions that will likely arise during your use of these models, this document is not intended to be a substitute for necessary risk management and credit scoring training, nor is it a substitute for your institution's own risk management experience, oversight, and policies. It is intended to be both a refresher and a supplementary resource for your own efforts. Of course, Portfolio Defense Consulting Group is also pleased to offer training specifically related to Custom Credit Scoring Models as well as custom-designed risk management training courses tailored specifically for your firm's credit or underwriting staff.

This document is intended to serve as a supplement to your Portfolio Defense project deliverables, also known as the Technical Review and Summary document. This document should be retained by your risk management team for the life of your model use, and contains all of the relevant detailed information about how your models were developed, the actual variables and score weights, and performance estimates.

The key sections of this User's Guide cover the following topics:

- What is a Custom Credit Scoring Model?
- How was it developed?
- How can I verify that it will work on my applicant population?
- What are the initial implementation steps?
- What are the best strategies for initial model implementation?
- Does my organization need training?
- How do I address compliance issues?
- What more advanced strategies are possible (tiered pricing, policy testing, etc.)?
- How do I know the models work, and how can I use scoring to help manage my portfolio (the benefits of tracking)?

Custom Credit Scoring Models:

Custom Empirical Credit Scoring Models (or Scorecards) are origination (applicant evaluation) credit risk scoring models that are developed and sold by the Portfolio Defense Consulting Group. These types of models are also known as "empirical" credit risk models, because they are based upon the lenders' own data.

Empirical credit risk models are typically developed for a specific lender, using that lender's own loan account performance data. It is this data, which is then added to the predictive information available on applicants (such as application data, credit bureau data, borrower capacity ratios, etc.), to develop such empirical predictive models. That payment performance information from the lender's own Master billing file system is what "tunes" the predictive variables to that lender's specific applicant population, and these types of models are generally the most predictive risk management tools available for making initial lending decisions.

Custom empirical models are created using Portfolio Defense's substantial amount of modeling experience. Members of our team have developed literally thousands of models for the leading lenders in a variety of industries, over the past 20+ years. Any type of lending product, from credit card, to home equity lines and loans, to direct and indirect automotive loans, mortgages, small business products, large retailers, manufactured housing, prime and sub-prime populations, is covered by different modeling experience.

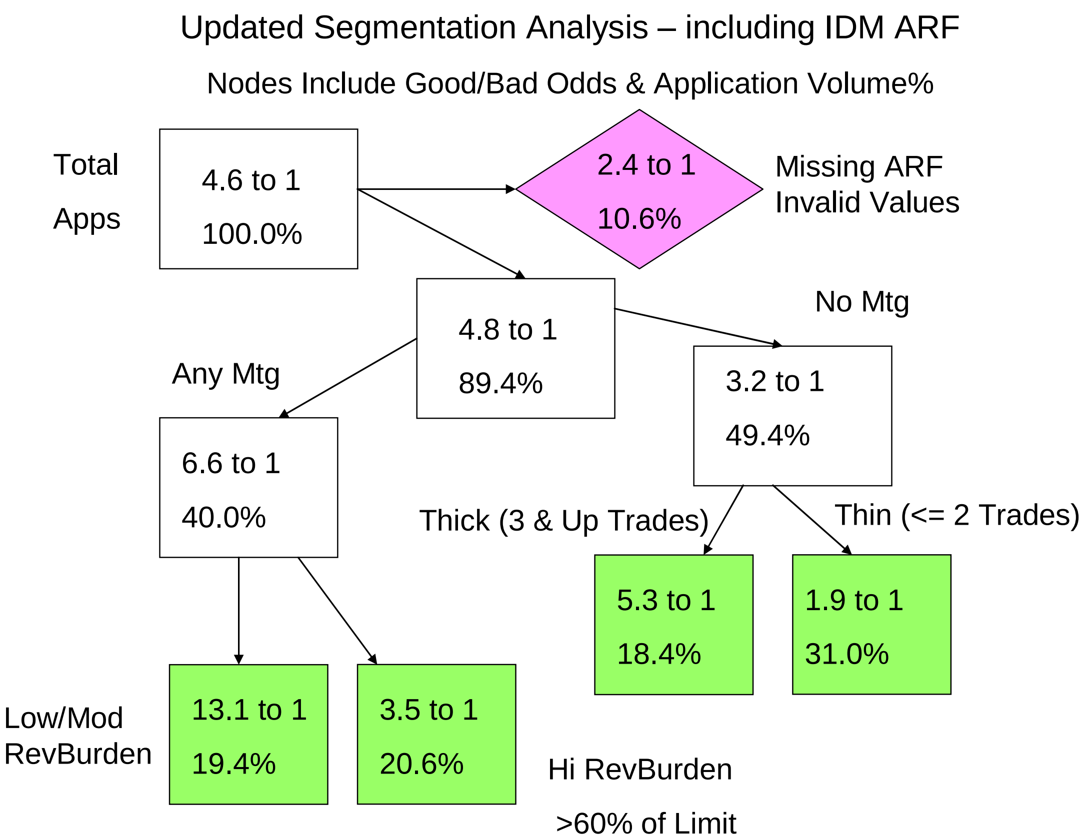
Model Development Process:

It is useful for you to know how your model was developed. This knowledge will help give you more confidence in using your new custom scores. But it will also let you be more informed, and understand what the limits of the score are, when to best use it, when to rely upon it heavily, and when not to.

There are two key components to the development of a custom empirical model. The first is the identification and capture of performance data within your systems. The second is determining what data is available to you at the time the credit decision is made. We work with you to construct a database containing a set of sample records of Goods, Bads, and Indeterminates (based upon actual payment behavior during a sample period) and link that data to copies of credit bureau data (and possibly other data sources). We then have a model development database which links the relationships between what an applicant looked like, in terms of predictive variables, at the time of application, and their subsequent performance behavior.

We produce a set of Known Good/Bad models, generating scoring weights for each predictive variable. Then, using a process of "reject inference", we generate performance data for a set of unbooked records, incorporate that data into the database, and generate a set of All Good/Bad or final models.

As part of the development, if we build multiple models, we perform what is known as a model segmentation analysis. This analysis allows us to build the optimal number and type of models and ensure the maximum predictive power of the system of models. The following chart shows a sample set of segmented models, as described within a Technical Review Summary.



This chart shows the different groups and their performance (Good/Bad Odds). The way to read this chart is that the average Odds of the entire population is 4.6 to 1, or that 4.6 out of 5.6 applicants will be Good (perform at the defined Good performance level) and 1 out of 5.6 applicants will be Bad. After removing the sample points with no valid CB data, the Odds shift slightly to 4.8 to 1. The first split is applicants with a mortgage and those without, which changes the performance from 6.6 to 1 versus 3.2 to 1. For those without a mortgage, they are split further to thick versus thin CB files (3 or more trades versus 2 or fewer trades). For applicants with a mortgage, they are split into high utilization versus low or moderate utilization. As you can see, in this example, just the use of these splits allows us to identify applicants as risky as 1.9 to 1 versus applicants as good as 13.1 to 1, compared to that average of 4.8 to 1.

We then produce detailed pseudo-code specifications, one set for each type of consumer credit report data format, deliver them to you or your IT vendor, and the models are implemented within your applicant processing, loan origination, and decision systems. As part of this process, we review the specifications with the programming staff, answer any questions, and ensure that we provide enough information to enable the accurate programming of the model.

MODEL IMPLEMENTATION

Model Validation:

An important regulatory requirement, as well as a good business practice, is the periodic validation of your credit scoring models. What this means is that, because you are using a scoring model to make credit risk decisions, you need to measure the performance of the model, or the ability of that model and the resulting score to separate between Good and Bad performing accounts.

Pre-Installation Validation:

There are several different ways to accomplish this. They depend upon the type of data that you have available and when it is available, and we will describe each of these in detail. One type of validation is a pre-installation validation, which can be performed if you already have some booked accounts with payment performance, and you will also need their credit bureau and application data from the time of application, as well as the ability to run this data through your systems and generate the score you want to validate. As part of a standard model development, we generate the new score and retain the original performance data, and so produce score distributions which demonstrate the performance of the new models during the development timeframe.

The first and primary group of performance statistics measures how well the new scoring models identify risk. The tables and charts are derived from distributions of the databases scores. An example score distribution table is provided below. These tables include estimated scorecard performance for various cutoff scores. The cumulative score distributions are created by calculating the new final score for the development database. These statistics provide the data for estimated scorecard performance. By selecting a specific score for a cutoff, we can estimate the impact on approval rates and booked loan quality. The tables include the following information:

- **Score -** The selected cutoff score for strategy purposes
- **Number of Goods –** Number of Goods at that score and above

- **% of Goods –** % of Goods at that score and above

- **Number of Bads –** Number of Bads at that score and above

- **% of Bads –** % of Bads at that score and above

- **Total Counts** – Total number of accounts at that score and above

- **%Total –** Total % of accounts at that score and above

- **Bad Rate -** The % of accounts that score at or above the score that is Bad (or the cumulative percent of Bads at or above that score). There is overlap between the score distributions of the Goods and the Bads. The goal is to develop a model that creates the greatest separation between the groups.

- **%Good - % Bad –** The difference between the two cumulative distributions. The max value is the K-S statistic.

- **Portfolio Odds** - The ratio of Goods / Bads for accounts that score at or above the score. These Cumulative Odds are greater than the Marginal/Interval Odds, because we are not only looking at the accounts that score at the specific cutoff score. We are also looking at the accounts that score above the score as well - these higher scoring accounts are of higher quality as well.

- **Fitted/Interval Odds –** The ratio of Goods / Bads for accounts that score at the specific score or interval. These Odds are calculated based on the scaling parameters. These parameters depend on the product. The scorecards in this example have been scaled at 1:1 Odds at a score of 240 with 40 points to double the odds; this is highly flexible based upon your needs.

Segment3 - No Mortgage & Thick File

Cumulative Custom Score Distribution - All Good vs. All Bad

Score	#Good	%Good	#Bad	%Bad	Total	%Total	Bad Rate	%G-%B	Portfolio Odds	Fitted Odds
540	49	0.22	0	0.00	49	0.18	0.1%	0.22	1155.5	181.0
500	720	3.27	5	0.09	725	2.58	0.7%	3.18	139.9	90.5
480	1,367	6.20	10	0.16	1,377	4.91	0.7%	6.04	140.7	64.0
460	2,144	9.73	18	0.30	2,162	7.71	0.8%	9.42	117.7	45.3
440	3,191	14.48	44	0.74	3,236	11.53	1.4%	13.74	71.9	32.0
430	3,883	17.62	70	1.17	3,953	14.09	1.8%	16.45	55.4	26.9
420	4,563	20.71	96	1.60	4,659	16.61	2.1%	19.11	47.5	22.6
410	5,350	24.28	124	2.06	5,474	19.51	2.3%	22.22	43.2	19.0
400	6,208	28.17	191	3.17	6,398	22.81	3.0%	24.99	32.5	16.0
390	7,144	32.42	245	4.06	7,389	26.34	3.3%	28.35	29.2	13.5
380	8,058	36.56	329	5.46	8,387	29.90	3.9%	31.10	24.5	11.3
370	9,189	41.70	457	7.59	9,646	34.38	4.7%	34.11	20.1	9.5
360	10,382	47.11	595	9.89	10,977	39.13	5.4%	37.22	17.5	8.0
355	10,948	49.68	665	11.05	11,613	41.40	5.7%	38.62	16.5	7.3
350	11,504	52.20	756	12.57	12,260	43.70	6.2%	39.63	15.2	6.7
345	11,996	54.43	848	14.09	12,843	45.78	6.6%	40.34	14.2	6.2
340	12,544	56.92	946	15.72	13,490	48.09	7.0%	41.20	13.3	5.7
335	13,170	59.76	1,067	17.75	14,237	50.75	7.5%	42.01	12.3	5.2
330	13,744	62.36	1,184	19.68	14,927	53.21	7.9%	42.68	11.6	4.8
325	14,324	64.99	1,297	21.57	15,621	55.68	8.3%	43.43	11.0	4.4
320	14,821	67.25	1,434	23.84	16,255	57.94	8.8%	43.41	10.3	4.0
315	**15,298**	**69.42**	**1,543**	**25.65**	**16,841**	**60.03**	**9.2%**	**43.77**	**9.9**	**3.7**
310	15,832	71.84	1,722	28.62	17,554	62.57	9.8%	43.22	9.2	3.4
305	16,327	74.08	1,873	31.14	18,200	64.87	10.3%	42.95	8.7	3.1
300	16,759	76.04	2,044	33.98	18,803	67.03	10.9%	42.06	8.2	2.8
290	17,764	80.61	2,418	40.19	20,182	71.94	12.0%	40.42	7.3	2.4
280	18,525	84.06	2,765	45.96	21,290	75.89	13.0%	38.10	6.7	2.0
270	19,355	87.83	3,186	52.97	22,542	80.35	14.1%	34.86	6.1	1.7
260	19,996	90.73	3,603	59.89	23,599	84.12	15.3%	30.84	5.6	1.4
250	20,584	93.40	4,036	67.09	24,619	87.76	16.4%	26.31	5.1	1.2
240	21,061	95.57	4,482	74.51	25,543	91.05	17.5%	21.06	4.7	1.0
220	21,624	98.12	5,185	86.19	26,809	95.56	19.3%	11.93	4.2	0.7
200	21,881	99.28	5,621	93.45	27,502	98.03	20.4%	5.84	3.9	0.5
180	22,009	99.87	5,903	98.13	27,912	99.50	21.1%	1.74	3.7	0.4
131	22,038	100.00	6,015	100.00	28,054	100.00	21.4%	43.77	3.7	0.2

Setting Initial Score Cutoffs or Credit Tiers:

Now that you we have looked at the performance distributions, we can use them to set our initial cutoff scores or credit tiers. Recognizing that the Odds estimates are just that, estimates based upon previous performance, we tend to recommend setting cutoff scores that would allow you to book at the same approval rate as in the past, knowing that you are booking better quality by using the new models. Then, as you monitor performance over time, you can gradually relax your cutoffs and increase loan volumes.

You will want to consider a number of different factors. What credit quality are you seeking in your booked accounts? One way to answer this question is to determine your minimum acceptable credit standards, or what is the relative quality of the worst applicant that you would be willing to put on your books. This can be measured by Good/Bad Odds, or the proportion of Goods and Bads within a given group. For example, 10:1 Odds means that out of 11 total applicants, 10 will pay at an acceptable level, while one will become seriously delinquent or worse. If you have good profitability data on your portfolio, you can even eventually set a cutoff score which will balance the profit of a set of Good accounts versus the loss of a Bad account, and ensure that you do not lower your cutoff score to the point where you book unprofitable accounts. These Odds projections also go out over a considerable time period, so they are best used as relative or comparative measures. *So you should be conservative in how you use them to set cutoffs and credit tiers; they are NOT exact measures.*

Another factor you might want to consider is applicant volume, and the rank-ordering effect of the custom scores. For example, you might want to set a credit tier of A quality, which represents the highest scoring percentage of your applicants. In the example in figure above, if you set your cutoff at 410 for example, you might expect to book approximately the best 20% of your applicants, and their average Odds might be about 43:1. You could set your next tiers in the same manner. *But it is extremely important to track the accounts as they mature and determine the actual, observed performance of the new cohorts of booked accounts.*

Once you have these Odds estimates, you can further refine and adjust your cutoff strategy. Portfolio Defense is always ready to assist in these measurement and strategy issues, as you gather data over time.

Getting Started and Forming Your Team:

One of the first steps that you should take in an organization that is new to scoring is to appoint a scoring manager. This employee would be responsible for all aspects of the use and management of your custom model(s). This person's responsibilities would include everything from initial implementation issues, setting cutoffs, establishing policies, ensuring legal compliance, performing and interpreting tracking, and educating underwriters and the rest of the operation about scoring. It would be helpful if this employee had somewhat of a quantitative or analytical background; people with experience in programming, mathematics, finance, or statistics often perform well in this role.

It would also be valuable to create a team or committee that is responsible for getting scoring started at your firm. It could be interdisciplinary, and include a member from risk, operations, IT, credit policy, possibly training or HR, and compliance. The team should review or set measurable goals for your portfolio and how to achieve them. Approval rates, Bad rates, auto decision rates, permissible override rates, expected loan volume, pricing policies, and compliance can all be affected by the use of credit scoring.

A useful first step would be to perform an assessment of the organization's understanding of credit scoring. This can include senior management as well as mid-level managers. It is important to recognize that the introduction of scoring is in many organizations, the introduction of a new way of managing the lending process. Applying a quantitative and scientific approach to lending may be new to many, and this involves organizational change. The best way to accomplish this change is through an education process, so that employees can learn how they will benefit from scoring; the benefits can include improved efficiency, opportunities to develop and enhance their skills, the ability to help the organization grow and prosper financially.

A key point to recognize is the issue of organizational change – scoring will change the way that lending is done in your firm. Some may fear that it will cost jobs or responsibility. Others may be concerned that different decisions about credit will lead to losses. Most staff will naturally wonder how this new technology will affect them, and you need to anticipate this and be prepared to respond to it. The use of scoring models involves a disciplined, measured approach. If employees do not trust the models or believe the score works, they can override the score and often make poorer decisions. In these cases, you don't receive the benefits of scoring, which are increased volume, reduced risk, better understanding of your best customers, and improved productivity.

Scoring education programs can come from different sources; they can be either internally or externally developed. Regulatory bodies of all kinds are much more knowledgeable about scoring issues these days, and they expect you to have knowledgeable staff as well. Portfolio Defense has many years of detailed experience helping a great variety of companies successfully adopt and use scoring technology. We can tailor a variety of custom training programs to help your use of scoring be successful and compliant with regulatory expectations.

Important Compliance Issues:

Whether you are a bank, a thrift, a credit union, a retailer, a finance company, a leasing company, a utility, or some other type of business that is using credit scoring, the one thing you have in common is that you are making a decision to extend credit to a customer. When you do this, you are going to be impacted by a variety of laws. Laws governing consumer privacy, data protection, fair credit, and protection from discrimination are among the issues that you deal with on a daily basis. Your business may be overseen by the Federal Reserve, the Comptroller of the Currency, the Office of Thrift Supervision, the National Credit Union Administration, Public Utility Commissions, the Federal Trade Commission, or some other regulatory or law enforcement agency.

It should be noted that we are not attorneys, nor do we offer legal advice. You should always consult your own legal advisors before making decisions involving compliance or other legal issues. However, we can discuss common compliance issues that arise in the use of scoring, and review common practices that we have seen in the lending industry.

There are some policies that you should set. These include which accounts to score or not, setting cutoff scores, treatment of co-applicants, determining adverse action procedures, choosing permissible types and proportions of score overrides, and designing common credit policies (in harmony with scoring). These issues are discussed in detail in Portfolio Defense's training seminars.

Advanced Strategies and Credit Policy Testing:

These types of strategies include using credit bureau scores in conjunction with Savant scores, risk-based product pricing approaches, varying loan terms with score, policy testing, and automated decisioning (auto decisioning rates and "grey areas"). Typically these strategies requite that you have more empirical performance data before you can implement them, so you would not likely try to implement them immediately upon starting to use scoring, but after you have had a chance to implement models and gather data.

It is possible to use credit bureau scores in conjunction with custom scores, typically with a "matrix" scoring approach. In a model development, we typically produce these matrices with both scores for you, and include the data in the Technical Review and Summary. You generate the custom score and also pull the credit bureau score, and populate a score matrix. You now have a more complex (and useful) score cutoff strategy. You see, in some cases, the two scores will disagree, in other words, perhaps 10-20% of the time one score will be relatively high and the other relatively low. But rather than this being a problem, it is actually an opportunity, because it allows you to take advantage of the additional information in the other score to make better decisions.

When you have more detailed performance information which includes financial results for booked accounts, you can actually estimate the relative profitability of accounts with different scores. Some of your accounts will perform better and generate more profit than others. Identifying this information would permit you to make use of “risk based pricing”, or adjusting your loan pricing to account for the different levels of loss and profitability. This is permissible, but is advisable only when you have actual performance data to back up your decisions.

When you have a score that represents account quality, like your custom score, and you have the performance data behind the score, you can vary your loan terms by score. You can offer longer loan terms to high quality applicants, accept higher levels of LTV percentages, or accept lower down payments. This can make your lending products more desirable to customers and help you be more competitive in the market, by offering the most favorable terms to the best quality applicants – applicants who have the greatest number of options, and where competition is the greatest.

With empirical data, it is also possible to put your credit policies to the test. In some lending organizations, credit policies have been handed down for many years, almost as a tradition, and often these policies do not have an empirical basis, or that basis was established a long time in the past, with old data. You are all familiar with long-established debt ratios rules, advance percentages, LTV percentages, etc. With data, it is possible to see how well these policies work, and whether they can be relaxed (giving you more volume without more risk), or whether they need to be tightened. Typically, credit policies which are based upon information that is *not* considered as part of the scoring process, and which are based upon historical performance data, can be worthwhile and work in harmony with scoring.

As you work more and more with scoring, and see that it is a reliable tool, you will seek to use it to improve your operational efficiency. One way to do this is to reduce the involvement of underwriting staff for decisions that are obvious. Some obvious decisions are very low scoring applicants. Others are high scoring applicants with high credit bureau scores, and who meet all of your credit policies. In this way, you can establish automated approval and decline thresholds (“auto decisioning”), where a certain percentage of applicants does not need to have a significant manual review. You will also want to consider loan amount in deciding which applicants to auto decision as well (the higher the loan amount, the less likely you will want to make an automated decision – the costs of a bad decision would be too high).

Portfolio Defense can work with you, both in training, and in analytic support to help you make use of these advanced strategies at your firm.

Tracking Reports and Portfolio Management:

There are a number of different types of tracking reports related to credit scoring, and they are typically divided into front-end reports and back-end reports. Front-end reports focus upon applicant quality, changes in applicant populations, and the use of scores (overrides and cutoff discipline). Back-end reports focus on the performance of the applicants, once you have put them on your books and the accounts have seasoned, and the performance of the models. More detailed descriptions and examples of these reports are in the Appendix.

Here are common names of the different types of reports, and their purposes (names can sometimes vary, depending upon which vendor you work with, but these are widely accepted industry terms):

- Portfolio Chronology Log
  This report permits you to summarize actions (as they occur) that might affect the portfolio and the use of the scoring system, either immediately or later on. These would include obvious items such as cutoff score changes, or credit policy changes, but can also include more subtle items such as changes in collections practices, pricing, or marketing efforts.

- Population Stability Report
  This report tells you whether your recent applicants are scoring higher or lower than your previous applicants (baseline measure), and where in the score distribution this is happening.

- Characteristic Analysis Report
  This report allows you to determine why applicants are scoring higher or lower over time, and which individual variables are most affecting the applicant final scores.

- Final Score Report
  This report looks at your use of the Savant models, and how often, and where in the score distribution you are overriding the recommended decision of the score.

- Dynamic Delinquency Report
  This report looks at the relationship between score and subsequent account performance.

APPENDIX – Example Portfolio Chronology Log

Purpose of Report:

Here is an example Portfolio Chronology Log. This report is documenting changes that have taken place over time which could affect scoring outcomes. It will help you diagnose, when you see population shifts or changes in model performance, what might be the underlying causes of such changes.

Portfolio: Direct Auto Loans

Date	Action Taken
January ’03	Implemented Custom Scoring Models
February ’03	Cutoff score lowered from 220 to 200
June ’03	Implemented CB score matrix with Savant scores
November ’03	Implemented credit grades A-D and risk based pricing strategy
January ’04	Collections call policy shifted for early delinquent accounts
February ’04	Pricing adjusted upwards 20 bp across all credit grades
April ’04	Branch expansion continued into Northwestern states
June ’04	Implemented cross-selling plan in branch network
July ’04	Started Used Car lending program
September ’04	Changed permissible advance criteria
October ’04	Cutoff score raised from 200 to 210

APPENDIX – Example Population Stability Report

Purpose of Report:
Here is an example Population Stability Report. This report is comparing the score distribution of a recent applicant population with a historical baseline. That baseline could be based upon a model development sample (this is included in your Technical Review and Summary document) if that is not too old (more than 2 years old), or it could be based upon the creation of an updated baseline which is more reflective of current and future business.

Population Stability (score comparison between two different applicant populations)

A *Score Range*	**B** *Baseline % (of total)*	**C** *Recent % (of total)*	**D** *Change (C-B)*	**E** *Ratio (C/B)*	**F** *Weight of Evidence ln(E)*	**G** *Contribution WOE x Change (FxD)*
Under 175	12.2%	9.5%	-.027	0.778	-.250	.007
175-184	7.7%	9.5%	.018	1.233	.210	.004
185-194	10.6%	6.3%	-.043	0.594	-.520	.023
195-204	14.5%	8.0%	-.065	0.550	-.598	.039
205-214	14.3%	8.1%	-.062	0.586	-.566	.035
215-224	8.8%	10.0%	.012	1.136	.128	.002
225-234	13.6%	17.8%	.042	1.308	.268	.011
235-244	11.1%	14.0%	.029	1.261	.232	.007
245 and Up	7.1%	16.8%	.097	2.366	.861	.084

POPULATION STABILITY INDEX (SUM of G): .212

This report is telling you whether there has been a significant change in your applicant population over time. If applicants are scoring significantly higher or lower than in the past, it is important to be aware of this from a business management standpoint (as loan volumes change), from a risk standpoint (since quality could be increasing or declining), and from a strategy standpoint (do you need to adjust your score cutoff, and what is causing the changes). The main result of this report is the Population Stability Index (PSI). More important than any single PSI number is the trend over time; are PSI's increasing, decreasing, or remaining relatively stable over time? Also, a high PSI is not an indicator that there might be a problem with your scoring models, you need to perform a validation to determine the model effectiveness. Although it is true that if there have been significant shifts, over an extended period, that indicates your new population is different from your historical populations, and you would likely benefit from building new models to better reflect this new population.

Report Detail:
You typically choose score ranges that are 10 points or so, and try to have about 9 or 10 cells. You might expect to see between 10-15% of the population fall into each of these cells. You calculate the percentage of applicants which fall into these score ranges or cells for both groups. You then calculate the difference (subtract baseline from recent or column C-B) and express that result as a decimal value. If a higher percentage of recent applicants are falling into the higher score ranges, then you can conclude that your recent applicants are tending to score higher (and would likely be approved at a higher rate, if the cutoff is unchanged); if a higher percentage of recent applicants are falling into the lower score ranges, then your recent applicants are scoring lower (and would likely be approved at a lower rate, if the cutoff is unchanged). It would be good to have at least several thousand counts, with a few hundred in each cell.

You then calculate the ratio of recent to baseline applicants or column C/B, again expressed as a decimal value. Using the natural logarithm, you calculate the logarithm of that ratio to create what we call the Weight of Evidence (WOE). This is done because the natural log of 1 equals 0; if the two cell proportions are the same, then the value of the WOE will be 0; if there are more recent applicants the WOE will have a positive value; if there are fewer baseline applicants then the WOE will have a negative value). Finally, you calculate the product of the WOE and the decimal change value to determine each cell's contribution to the index, or the Population Stability Index (or PSI). Lastly, you total all of the contributions to calculate the PSI. Usually a PSI of .1 or less is not felt to be a substantial change, between .1 and .250 somewhat of a change, and .250 or more is a significant population shift, worthy of further analysis (creating the Characteristic Analysis Report).

Data Needed:
You need your original applicant score, as well as a baseline score distribution from your application processing system.

Production Frequency:
Produce no more frequently than once a quarter. Minimum counts of 1000, although several thousand are better.

Management Issues:
Update baseline distribution every 2 years, with more recent data covering at least 6 months of applications and 1000 counts. Produce separate reports by product type and by scorecard. Keep track of quarterly PSI values over time.

In cases where there was a significant shift, you would also want to try to determine what might be causing the shift, by producing a Characteristic Analysis Report as described below.

APPENDIX – Example Characteristic Analysis Report

Purpose of Report:
Here is an example Characteristic Analysis Report for one imaginary variable (in this case, it is Revolving Utilization). You would need to know all of the variables in the scorecard that is being analyzed, as well as the individual score weights, the percentage of the baseline applicant population that receive each different potential point value for each variable, and the results of a more recent sample of applicants, again broken out by variable attribute value.

Revolving Utilization (Percentage)

A *Variable Value*	**B** *Baseline % (of total)*	**C** *Recent % (of total)*	**D** *Change (recent-baseline) (C-B)*	**E** *Attribute Score Weight*	**F** *Score Difference (DxE)*
Over 69	12.7%	15.0%	.023	10	.23
55-69	15.0%	22.3%	.073	13	.95
30-54	18.8%	19.7%	.009	17	.15
12-29	19.1%	21.0%	.019	21	.40
5-16	16.7%	15.3%	-.014	27	-.38
Under 5	17.0%	5.7%	-.113	33	-3.73
Missing	.7%	1.0%	.003	18	.05
			TOTAL (NET) SCORE DIFFERENCE:		-2.33

Report Detail:
You would have ranges for this report broken out by the attribute values for each variable that you wish to analyze. Also be sure to include special and missing attributes (unable to calculate, missing, etc.). You compare the percentages of recent applicants who fell into a specific variable attribute value range to a previous baseline of applicants. You calculate the change for each attribute (the percentage of recent applicants minus the percentage in the baseline). This tells you whether there are more or less applicants within an attribute. You then multiply this change column by the score weight for each attribute. This tells you what the average score difference was by attribute. You then sum the differences for all the attributes to get a net score difference number.

This report is telling you the average score change by variable, comparing recent applicants to a previous baseline. In this case, applicants are scoring on average 2.33 points lower than in the past, for the Revolving Utilization variable (due to applicants having higher average revolving utilization, in this case). A *net* shift of 5 points or more indicates a significant change has taken place in a variable, and you might want to consider a change in your scorecard cutoff strategy. A positive score difference for an attribute means that the average applicant is scoring higher for that attribute value, a negative value means the applicant is scoring lower. The individual values are not important – the net or total value for a variable is. Also, you need to sum up the values across all of the variables, and look at the net or total score change for the entire scorecard – if there are shifts within the variables, but they tend to cancel each other out and the total for the scorecard is small, then that is not really a cause for concern. It would be good to have at least 300 counts for each variable.

In cases where there was a significant shift, you would also want to try to determine what might be causing the shift. In the case of credit bureau variables, it is not likely that anything under your control is causing a shift (although if you changed application processing software, or recently re-coded certain variables, you should check the logic for consistency and accuracy if there is a large shift – however, economic and consumer demographic shifts, such as typical applicants carrying higher revolving balances, is also a possible explanation). For application variable shifts you should check whether there have been changes in how information is taken (in the branches, for example), or if the applications themselves might have been changed, affecting the possible responses.

If you have changed policy rules or operational procedures (such as opening a new checking account, within your branch, for every applicant for a credit card, for example), that could cause shifts within variables. In that example, if you used the variable, Checking and Saving Account Reference, you would find a shift to more applicants with checking accounts than in the past.

Data Needed:
You need a recent applicant data including counts at the variable and attribute level, from your application processing system. You also need baseline standards by each variable and attribute level, taken from a baseline sample of recent applicants (the same as that used for your PSI baseline). You need copies of all scorecards that will be analyzed, at the attribute level, including all score weights.

Production Frequency:
Produce no more frequently than once per quarter, or when the PSI value exceeds a value of .100. Minimum counts of 1000, although several thousand are better.

Management Issues:
Update baseline distribution every 2 years, with more recent data covering at least 6 months of applications and 1000 counts. Produce separate reports by scorecard.

APPENDIX – Example Final Score Report

Purpose of Report:
Here is an example Final Score Report. It is generally used to determine the extent of overrides to scoring decisions as well as to determine where the greatest number and percentage of overrides are taking place. If the data is available, it is also possible to determine the number of specific overrides that are occurring by override reason (if you use and track override reason codes).

Final Score Report (looking at cutoff discipline/override behavior)

A *Score Range*	**B** *Number of Apps)*	**C** *Number Approved)*	**D** *Number Declined*	**E** *Interval Approval Rate (C/B)*	**F** *Interval Override Rate (C/B below cutoff, D/B at or above cutoff)*
Under 175	631	5	626	.7%	.7%
175-184	709	12	697	1.6%	1.6%
185-194	295	36	259	12.2%	12.2%
195-199	359	55	304	15.3%	15.3%
Sub Total	1994	108	1886	5.4%	5.4%
200-209	440	420	20	95.4%	4.6%
210-219	175	165	10	94.2%	5.8%
220-229	511	500	11	97.8%	2.2%
230-229	655	650	5	99.2%	.8%
240 and Up	462	460	2	99.6%	.4%
Sub Total	2243	2195	48	97.9%	2.1%
Grand Total	4237	2303	1934	54.4%	

Report Detail:

Key statistics in this report are: the overall approval rate was 54.4% (total approved divided by total applications). The low side override rate (or reversals of recommended decline decisions) was 5.4% (total approved below cutoff divided by total below cutoff). The high side override rate (or reversals of recommended approval decisions) was 2.1% (total declined at and above cutoff divided by total at and above cutoff). Low side overrides are generally of more concern since they represent potential added risk, while high side overrides are lost opportunity. Acceptable percentages vary by type of lending product but generally 5-15% on the low side and 10-20% on the high side are not uncommon. You should be sure that you do not count policy decisions (eg. decisions that are consistently applied, regardless of score, such as declining apps with major derogs) as overrides.

Data Needed:

You need score and recent applicant decision information (whether applicant was approved or declined, and reason code, if available).

Production Frequency:

Produce quarterly, preferably with minimum counts of 1000. Produce separate reports by product type and by scorecard.

Management Issues:

Keep track of approval rates and override rates over time to see how they might be changing.

APPENDIX – Example Dynamic Delinquency Report (DDR or Validation report)

Purpose of Report:
Here is an example Dynamic Delinquency Report. This report is comparing the relationship between the proportion of Goods and Bads within a score range or cell, to that score. The proportions of Goods and Bads are converted to Good/Bad Odds, and you are looking for a pattern of payment performance that varies by score.

Score vs. Delinquency - Accounts Booked 7/1/03-6/30/04 (Maximum Delinquency as of September 2004)

Score Range	Number Accounts	% of Accounts	Never Delq	Ever 30+	Bad Rate	G/B Odds
No Score	134	4.6%	121	13	9.7%	9.3
Below 180	68	2.3%	49	19	27.9%	2.6
180 - 199	179	6.1%	152	27	15.1%	5.6
200 - 219	479	16.4%	442	37	7.7%	11.9
220 - 239	553	19.0%	523	30	5.4%	17.4
240 - 259	572	19.6%	551	21	3.7%	26.2
260 And Up	929	31.9%	904	25	2.7%	36.2
Total	2,914		2,742	172	5.9%	15.9

Note : Bad Rate = %(Bads / Number of Accounts)
G/B Odds = Goods/Bads

Report Detail:
A more specific definition of the calculations are: the Goods are considered Never Delinquent, the Bads are Ever 30+, the Bad Rate is the percentage of Ever 30+/Total Number of Accounts, and Good/Bad Odds are the ratio of Never Delinquent/Ever 30+.

From this table you can see the general pattern of increasing loan performance (lower Bad Rate and higher Odds) as the score ranges increase. From this distribution of scores by performance, we calculate the odds to score relationship to test the validity of the scores.

Data Needed:
You need the original score from your application processing system. You also need to be able to measure maximum level of delinquency over a period of time (preferably 1 year) for all accounts from your masterfile or billing file, looking at Never Delinquent, 30+ DPD, 60+ DPD, etc.

Production Frequency:
You may produce the DDR quarterly, but at minimum annually. A lender with small volumes or with a clean portfolio will need to wait longer to accumulate delinquent accounts, in order to have sufficient Bads with which to measure the model's performance. You should strive to have at least 300 total Bads, and 20-30 Bads within a score range in order to create a meaningful report.

Management Issues:
Produce separate reports by scorecard, and if possible, by lending product within a scorecard (if you use one model for multiple products). Only produce the reports when you have sufficient counts as described above; producing the report with insufficient data will only lead to high variation and misleading results. Track the PDO (points to double the Odds) over time. If PDO is increasing significantly over time, that is an indication of degrading model performance.

APPENDIX – GLOSSARY

Acceptance Rate
Percent of all applicants which pass cutoff score and all credit policies. Raising cutoff score generally reduces acceptance rate, as does creating tighter credit policies, and a reduction in applicant quality usually reduces acceptance rate also.

Account Attrition
The propensity of a booked account to close, to transfer balances, to pre-pay, or to refinance. Attrition and pre-payment models seek to predict this behavior, which if unchecked, dramatically reduces account profitability.

Account Management
The process of managing different account features and processes to optimize profitability and risk; such as managing delinquent collections; managing credit limits; authorizing transactions; and supporting cross-selling efforts. Usually a system such as a "decision engine" or an "adaptive control" process which uses different decision variables and strategies is the method used.

Adaptive Control
The process used within account management software which permits the development, testing, and measurement of alternative strategies. Also known as "champion - challenger"strategies, which allow different approaches to compete, and measure the results on randomly selected groups of accounts.

Adverse Action
Taking an action, usually some form of credit decision, on an applicant or accountholder which results in less favorable treatment. Such as declining an applicant for a loan, or reducing the credit limit of an account which is in good standing. Every creditor should have the ability to provide adverse action reasons to the applicant, based upon the scoring model used for the decision. These reasons may be based upon the applicant's distance from the neutral value for a variable, the average value for a variable, or the maximum value for a variable (read Reg B for more rules about adverse action reasons).

Adverse Selection
In consumer credit, this generally refers to a process which results in the unintended consequence of attracting lower-qualified applicants. An example might be a high-priced loan offer to a high quality applicant group. The higher scoring applicants, because they have more options, choose a better offer, while the lower scoring applicants accept the offer. The net result is a poorer quality population of booked accounts. The opposite of this is called "virtuous selection".

Ascending Distribution
A table showing, for each score, what fraction of applicants can be expected to obtain that score or less.

Attribute Value
The value that a predictive variable can have. Such as number of months, for a variable such as Time on Job.

Bads
Booked accounts that prove to be unprofitable or otherwise exhibit payment behavior that is incompatible with the organization's credit standards. These could include accounts which were seriously delinquent, repossessions, bankrupts, charge-offs, and foreclosures.

Bad Rate
The percentage of booked applicants which exhibit the payment performance of Bads.

Behavior Score
The score generated by an account management risk model, to predict future serious delinquency. These models use as their predictive variables, the payment and delinquency behavior exhibited on the account. The scores are generated, usually on a monthly basis, and used as part of account management strategies.

Break-even Odds
The point within a score distribution where the proportion of Good paying applicants balances the proportion of Bad paying applicants, from a profitability perspective. For example, if a Good account earns $200 and a Bad account costs $2,000, then the point where on the margin, no losses would occur, would be 10:1 odds. Some aggressive risk management strategies attempt to set cutoff scores near this point. The risk is that your accounting data might be incomplete, or an odds shift might occur, causing you to lose money on a group of accepted applicants.

Characteristic
Another term for a predictive variable.

Continuous Performance Variable
A performance variable which has a continuous set of possible values, usually used to predict items such as Payments made in dollars (for a Collection score), or Recovered Dollars (for a Recovery Score).

Continuous Predictive Variable
A predictive variable which has a continuous set of possible values, such as Age, Time on Job, Income, Time in File, or Revolving Utilization.

Credit Bureau Score
This typically refers to generic risk scores vended by various credit bureaus. These can be used by themselves or in conjunction with other scores.

Custom Application Score (or Account Origination Score)
A risk score developed with the lender's own performance and predictive data, tailored to their applicant population. Typically the most powerful predictive score that is used during the applicant decisioning process. Also known as an "empirical" origination or applicant model.

Cutoff Score
The score which represents a lender's minimum acceptable credit standards. Anyone scoring at and above this score would typically be approved unless they violated a credit policy, or were a judgmental over ride.

Cutoff Odds
The odds of applicants scoring exactly the cutoff score. This represents the minimum acceptable credit standards of the lending institution using the model(s).

Descending Distribution
A table showing, for each score, what fraction of applicants can be expected to obtain that score or more. Most frequently used to determine the quality of applicants that would be accepted by a given risk model.

Dichotomous Performance Variable
A performance variable that is split into one of two mutually exclusive values, such as Good vs. Bad, Attritor vs. Non-Attritor, or Profitable vs. Unprofitable Account. As distinguished from a Continuous Perfomance Variable, which is used to predict a specific value in performance (such as Dollars Collected).

Discrete Predictive Variable
A predictive variable whose different values are not related to each other, such as with Occupation, Home Ownership Status, or Credit References.

Divergence
One method of measuring the predictive power of a model, by calculating the separation between the mean of the Goods and the mean of the Bads, adjusted by the shape of the two curves, or variance (standard deviation squared). It is a useful measure when examining the impact of changes being made to an individual model, but because it is sample specific (affected by performance definition used, sample window used, and any sampling bias), it is not a useful measure for comparing the predictive power of one model to another (built with different data samples).

ECOA (Equal Credit Opportunity Act)
Federal legislation addressing credit scoring with the intent of preventing discrimination. The Federal Reserve Governors had developed regulations to enforce the Act. The resulting regulation is know as Regulation B (Reg B), along with the Staff Commentary to Reg B, which amends and clarifies the regulations.

Empirical Model
A type of risk model that is based upon custom client performance data, where that data is "tuned" to the client's applicant population and based upon actual performance. This is the most powerful type of applicant risk model.

Exclusions
A type of account or applicant whose inclusion would introduce bias into an empirical model development sample, and thus, these are consciously excluded from that sample. These could include accounts whose performance is unknown (perhaps caused by data processing problems during system conversion), or accounts whose performance was not due to the actions of the account holder (such as employee accounts, deceased account holders), or accounts which were not representative of the general population (such as accounts with insufficient experience, large dollar or small dollar accounts), or accounts which would not generally be scored (such as VIP).

Expert Model
A type of risk model that is based upon the experience of a model developer, who knows the type of predictive models, variables and score weights that have been useful with similar products and/or similar populations. This technique can be quite useful, but is most often used for lenders who are starting new lending operations, or expanding their businesses into products or regions where they lack historical empirical data.

FACT (Fair and Accurate Credit Transactions) Act
Federal legislation amending FCRA, permitting consumers to receive a free copy of their individual credit report from each bureau every 12 months, giving consumers greater rights to dispute inaccurate credit bureau information directly with lenders, increased fraud alerts and identity theft protections.

FCRA (Fair Credit Reporting) Act
Federal legislation regulating who is permitted to access your consumer credit report data, how that data can be used, and what consumer notifications are required when such data has been used.

Goods
Booked accounts that prove to be profitable or otherwise exhibit payment behavior that meets the organization's credit standards. These could include accounts which were never delinquent or which were only mildly delinquent.

Information Odds
The degree to which the Population Odds (or Average Applicant Odds) are adjusted for, in an empirical model, by the inclusion of information relating to payment performance.

Inquiry
As reflected in a credit bureau file, a lender's (or potential lender's) purchase of an individual applicant's credit file, initiated by an applicant. The presence of excessive inquiries is normally indicative of added risk, due to an applicant's "credit appetite". This type of inquiry is distinguished from a promotional inquiry, which results from a lender's direct mail activities, is not initiated by an applicant, is not reported on a purchased credit report, and generally does not affect a credit score.

Neutral Score Value
Each predictive variable in an empirical model has a neutral value which represents the applicant that displays neither positive nor negative payment behavior. It is designed to provide a method of calculating Adverse Action reasons, or to provide a value to use in the case of an applicant's missing a rare piece of scored information. Neutral scores should not be used in the case of multiple missing pieces of scored information however (such as a "no-hit" applicant, who should not be given neutral values for all credit bureau-based variables in the model, because this would create a misleading and unreliable score).

No-Hit at the Bureau
A response to a credit bureau inquiry which indicates no file is present or the applicant has not been properly identified to the bureau and a match can not be made. Due to the substantial predictive power of credit bureau-based data, it is not recommended that "no-hit" applicants be approved for credit without some additional form of credit worthiness information from non-traditional sources, such as proof of positive payment behavior with landlords or utilities, or local stores that might not report to credit bureaus. "No-hit" applicants are typically much riskier than applicants who have credit bureau information available, but they should not be automatically declined since minorities often make up a higher percentage of these applicants than non-minorities, and automatic rejection could be interpreted as discriminatory behavior.

Overall Odds
The result obtained when Average Applicant Odds are adjusted by Information Odds through the use of a score-based model. These Overall Odds are reflected within a score distribution, and represent the estimated performance of each applicant in relation to the other. The odds used within a score distribution are intended to "rank-order" an applicant population (higher scoring

applicants perform substantially better, and a positively sloping “odds-to-score” relationship exists). These are also sometimes called Total Odds.

Odds are intended to help compare the quality of different applicants, but due to the nature of statistics, they are better predictors for large numbers of applicants than as predictors of specific individual borrower performance. For example, it is more accurate to say that Applicant A is in a population with higher odds than Applicant B, and is likely to perform better; rather than Applicant A has odds that are precisely 2.345 times better than Applicant B.

Override
A judgmental reversal of the recommended decision of the risk model, either on the high-side (turning down an applicant with a passing score - lost opportunity) or on the low-side (approving an applicant with a failing score - added risk). This is distinguished from the consistent application of a credit policy (such as decline all applicants with multiple major derogs, regardless of score), which is applied independently of score. Regulators pay particular attention to override behavior of lenders when examining potential discriminatory practices. See also, ECOA.

Pass Rate
The percentage of applicants which pass the cutoff score. Pass rates may differ from acceptance rates, which are the percentage of all applicants that are accepted, since overrides may affect the acceptance rate.

PDO (Points to Double the Odds)
This is a scaling factor that is sometimes used to establish an initial odds-to-score relationship for a model. For example if it takes 20 points to double the odds, an applicant scoring 220 would have twice the odds of an applicant scoring 200. It is purely cosmetic, and can be changed based on the desire of the client.

Performance Statistics (or Performance Forecast)
A set of score distributions produced by running the model development sample against the new models. This tells you, retrospectively, what results you might have obtained if you had been using the new model(s) during the model development sample period. However, because these statistics rely upon the results of a reject inference process, and because they do not typically account for the impact of previous credit policies which were in place, in many cases, they may tend to overstate the projected reduction in delinquency or increase in volume. Portfolio Defense recommends performing an empirical policy analysis upon the model development database, to more accurately reflect the expected results of the new model(s) when they are implemented.

Population Odds (or Average Applicant Odds)
The relative proportion of Goods to Bads within the applicant population, without taking any predictive information into account, within an empirical development.

Rank-Ordering
A score distribution can be said to “rank-order” an applicant population when higher scoring applicants perform substantially better, and a positively sloping “odds-to-score” relationship exists. See also, Overall Odds, and Validation.

Reject Inference
The process of incorporating unbooked applicants within risk model development to
represent the total “through the door” applicant population, within an empirical model development. Reg B states that for a modelto meet their definition of “empirically derived and statistically sound”, the entire applicant population must be represented in the development. Since the declined and uncashed (approved but unbooked) applicants do not have observe performance, their payment performance must be “inferred” through statistical means, to include them in the sample. This process is one of the most challenging and complex in the model development process.

Sample Size, Minimum
There is a certain, statistically significant minimum sample size for a model development that is preferred. Going below that number could result in a model which would be unstable, one which might over-fit a small amount of data, and which could lead to an unreliable model. In risk model development, the constraining factor is having a sufficient number of Bads, since the vast majority accounts are usually Good. Portfolio Defense can work with a sample as small as 450 or 500 Bads to build a robust model, although we would prefer not to go lower than that. If not enough Bads are available to build a custom model, it is still possible to use a smaller sample to validate a pooled model, or to adjust the weights of a borrowed model (as few as 250-300 Bads could suffice here). See also, Sampling.

Sampling
The process of statistically selecting a random portion from a larger pool of data. We seek a sample large enough to be unbiased and representative, but small enough to work with in a model development. Samples should also be relatively recent, so that they do not differ greatly from current populations, and they should be made up of data records that would be similar to those currently available. See also, Exclusions.

Scaling
In an empirical model development, in order to create scores based upon whole numbers, rather than probability estimates, a process known as scaling converts these “unscaled values” into more intuitive whole number score weights, which can be adjusted to meet the client’s needs. One way of thinking about scaling is to think about the SAT college entrance exam, where 800 is a perfect score. This is an arbitrary choice, and it could just as easily be a score of 1000 or 100 instead. Scaling changes are cosmetic, and do not affect the predictive power of the model.

Score Weight
A numerical value assigned to a given variable value. See also, Scaling.

Segmentation Analysis
How many empirical models do I need, and of what type should they be? Whenever a large portfolio is involved, it makes sense to conduct a segmentation analysis to ensure that the lender is not buying too many or too few models to provide the optimal predictive power.

The selection of potential splits should be based upon added predictive power, but also based upon the availability of the splitter variables, the legal sensitivity of potential splits, the proportion of the portfolio being scored by a given model (for example, it makes little sense to build 4 models which, when combined will be used on 15% of all applicants, and then build 1 model to be used on 50% of the applicants), where the lender perceives the risk coming from (riskier lending products or applicants), areas where the lender plans to grow the business (new states or products), and possibly under-served applicant populations (such as LMI, or low-moderate income models).

Swap Set (or Alternative Decision Set)
A measure of the extent to which decisions taken under the new model will differ from those made under the previous method. This includes applicants who had previously been accepted who will now be declined, and those applicants who had previously been declined who will now be accepted. One of the methods of "reasonableness testing" a new model is to see the extent of the swap set - if it is too large, the model may be unreliable, if it is too small, you may not have enough of an impact upon decisions to improve acceptance rate or delinquency.

One thing to bear in mind is that, when models are implemented, decisions will shift. Some scoring novices fear that these changing decisions mean the model is not working, when it fact it is a confirmation that the model is working as expected. See also, Performance Statistics.

Validation
A process to statistically demonstrate the predictive power of a model, using empirical data. Typical validations examine the relationship between score and subsequent observed payment (or other) performance.

Vintage Analysis (or Static Pool Analysis)
An analysis of observed levels of delinquency over time, comparing groups or cohorts of accounts with the same time on books (or account vintage) at different observation periods. Using this type of data, it is possible to graph out a Credit Life Cycle, observing the relationship between delinquency and time. Once this benchmark measure is known, it is possible to compare newly booked cohorts, and to compare their early delinquency trends to the benchmark measure, and determine whether aggressive account management techniques are needed to reduce the rising delinquency trend.