

# Preparing for Basel II Modeling Requirements

## Part 2: Model Validation

by Jeffrey S. Morrison

**T**his article, the second in a four-part series, discusses some approaches to the validation of statistical models as required by the new Capital Accord.

**A**s mentioned in the first article in this series, the current school of thought surrounding the probability of default (PD) and loss given default (LGD) models mentioned in Basel Consultative Papers is that banks should have separate models for the obligor and the facility. The obligor model should predict the PD—usually 90-plus days delinquent or in foreclosure, bankruptcy, charge-off, repossession, or restructuring. Models on the facility side should predict the loss given default (LGD) or 1 minus the recovery rate.

In the initial article, *logistic regression* was the approach recommended for building PD models. This statistical technique uses a set of explanatory variables whose values today would hopefully predict a loan's probability of default sometime over the next 12 months. On the LGD side, the approach recommended was to

use either *linear regression* or *tobit regression* to estimate the model.

### Directives from Basel II

Paramount to using the advanced approach as specified in the Basel II Capital Accord is a focus on model validation. *The New Basel Capital Accord*, published January 2001, includes the following:

302. Banks must have a robust system in place to validate the accuracy and consistency of rating systems, processes, and the estimation of PDs. A bank must demonstrate to its supervisor that the internal validation process enables it to assess the performance of internal rating and risk quantification systems consistently and meaningfully.

305. The process cycle of model validation must also include:

- ongoing periodic monitoring of model performance, including evaluation and rigorous statistical

testing of the dynamic stability of the model and its key coefficients;

- identifying and documenting individual fixed relationships in the model that are no longer appropriate;
- periodic testing of model outputs against outcomes on an annual basis, at a minimum; and
- a rigorous change control process, which stipulates the procedures that must be followed prior to making changes in the model in response to validation outcomes.

As of yet, *The New Basel Capital Accord* does not give specifics or standards related to the validation process.

### Introduction to Validations

Validation includes issues of data quality, documentation, sensitivity analysis, model specification, sample design, the performance of statistical tests, and the development of measures for

© 2003 by RMA. Jeff Morrison is vice president, Credit Metrics—PRISM Team, at SunTrust Banks Inc., Atlanta, Georgia.

model accuracy. Although not minimizing the importance of these other areas, for brevity's sake the remainder of this article will focus on quantifying accuracy measures. In this light, model validation simply refers to checking the accuracy of your model over some specific period of time. How many loans actually went into default during the year and what did their predicted default probabilities look like? If most of your defaults had a predicted probability of default near 10%, then your model may be doing a poor job.

Not only is the validation process part of Basel requirements, it is central to any model development process, regardless of its application. Even econometric forecasting models—models developed using aggregated data with economic time series—are validated for accuracy. Credit-scoring models are also validated for accuracy. Because modeling is indeed an art, statistical algorithms are developed and redeveloped until a formulation is found that reflects the most accurate results and makes the most business sense.

Validations can be done in a variety of ways, ranging from the simple to the complex:

1. Performing the validation only on your model development sample.
2. Performing the validation on a sample of accounts that were not used to develop the model, but were taken from the same period of time.
3. Performing the validation on a single holdout sample from time periods outside your model development window.
4. Performing a *step-through* sim-

ulation process across multiple time periods while recalibrating the model.

The first approach is the most straightforward and is typically performed as the model is developed. Here, the same data that was used for estimating the model is used for validation. Although this type of validation tends to overstate the model's predictive ability, it may be necessary if there are a limited number of defaults available for model building purposes.

If there are sufficient defaults available, the second method is preferred. A random sample of data is held out from the model estimation; the second method runs the holdout data against the model to compute its predicted values for validation purposes. This method is widely used for validating a variety of different models and serves as an aid to the statistician in selecting the best model.

The remaining methods are more advanced—not because their techniques are necessarily more complicated but because they require a greater depth of default history. The third approach holds out data for validation from prior periods to see if the level of accuracy remains the same from year to year. This is an indication of how stable your model may be over time. The fourth approach is a combination of validations and model recalibrations. The idea is to simulate model development and its predictiveness over time given that model revisions are done annually as new defaults are accumulated and added to the process.

### Validating the Obligor Model: Probability of Default

Let's assume you use the second validation method to evaluate the accuracy of a PD model and hold out a sample of accounts that were not used in model development. So how many defaults do you need? Generally speaking, hundreds of defaults are necessary to properly test the model—the more, the better. Since more defaults are available in credit card portfolios, thousands of defaults are typically used in validations.

To perform a validation, you need predicted default probabilities from your model and a default status indicator showing whether the account defaulted or not. With this information handy, you can easily calculate two measures of model accuracy where all you need to be able to do is sort and add. Following is a step-by-step guide for PD validation.

**Step 1: Create your holdout sample, if available.**

**Step 2: Code your default indicator.** Since you built your model with a default indicator of “1” if the loan defaulted and “0” otherwise, make sure your default indicator in your holdout sample is coded the same way.

**Step 3: Sort your holdout sample.** Sort the data from highest to lowest, based on the probability of default. If your data set is small enough, you could even do this in Excel.

**Step 4: Record minimum and maximum probabilities in each 5% bucket.**

**Step 5: Add some numbers together.** Now start totaling the data into buckets at 5% intervals from the top down. Produce the following columns by bucket:

- Number of defaults.
- Number of nondefaults.

**Step 6: Calculate cumulative percentages by bucket:**

- Cumulative number of defaults.
- Cumulative number of nondefaults.

Figure 1 shows this process for a fictional model in which the number of defaults was totaled into 20 buckets, each representing about 5% of the accounts. When this procedure is applied to an accurate model, the majority of the defaulters

should be accumulated in the earlier buckets. Likewise, the nondefaulters should be found toward the bottom buckets. This shows us the power the model has in distinguishing between defaulters and nondefaulters.

**Step 7: Identify percentages of defaulters and nondefaulters in the top two buckets.** Note that the column F labeled

“Cumulative % Defaults” indicates that 20.98% of the total defaulters in the holdout sample were identified in the top 10% (two buckets) of the sorted list. The bigger these numbers, the better. Note only 1.64% of the nondefaulters were found, as shown in column G labeled “Cumulative % Nondefaults.” These measures, reflecting the accuracy of the model for the top 10% of the data, can serve as an excellent way to validate competing models.

**Step 8: Calculate the K-S value.** Another measure of accuracy can be computed from columns F and G. This value, called K-S, is simply the maximum difference

between these two columns of numbers. The K-S value can range between 0 and 100, with 100 implying the model does a perfect job in predicting defaults or separating the two populations. In general, the higher the K-S, the better the model. The place where that maximum occurs is that point of maximum separation. As shown in Column H, the K-S value in this example is 74.1 and occurs at the 9th bucket.

**Step 9: Produce a graph.** This is done by simply graphing columns F and G. A graphical depiction of this table, as seen in Figure 2, goes by a variety of names, such as power curve or lift chart.

Figure 1

A 5% Bucket	B Min Probability	C Max Probability	D # Defaults	E # of Non Defaults	F Cumulative % Defaults	G Cumulative % Nondefaults	H Difference in % Cumulatives
1	0.987637	0.998625	33	4	10.1852	0.9390	9.2
2	0.961737	0.987637	35	3	20.9877	1.6432	19.3
3	0.932949	0.961737	36	1	32.0988	1.8779	30.2
4	0.897801	0.932949	36	2	43.2099	2.3474	40.9
5	0.821660	0.897801	30	7	52.4691	3.9906	48.5
6	0.813351	0.821660	38	0	64.1975	3.9906	60.2
7	0.545989	0.813351	28	9	72.8395	6.1033	66.7
8	0.530848	0.545989	24	14	80.2469	9.3897	70.9
9	0.398955	0.519828	22	15	87.0370	12.9108	74.1
10	0.314351	0.398955	8	30	89.5062	19.9531	69.6
11	0.192926	0.295930	5	33	91.0494	27.6995	63.3
12	0.132092	0.192926	3	34	91.9753	35.6808	56.3
13	0.117058	0.132092	5	33	93.5185	43.4272	50.1
14	0.099934	0.117058	0	37	93.5185	52.1127	41.4
15	0.099934	0.099934	0	38	93.5185	61.0329	32.5
16	0.090868	0.099934	7	30	95.6790	68.0751	27.6
17	0.090868	0.090868	0	38	95.6790	76.9953	18.7
18	0.069882	0.090868	12	25	99.3827	82.8638	16.5
19	0.045697	0.057391	2	36	100.0000	91.3146	8.7
20	0.036294	0.045697	0	37	100.0000	100.0000	0

Preparing for Basel II  
Modeling Requirements  
Part 2: Model Validation

The vertical axis is the cumulative percentage of defaulters or nondefaulters counted or identified. The horizontal axis reflects how far down the sorted list you are. In other words, a value of 30 on the horizontal axis means that you have examined the upper 30% of the validation data. The grey line at the top reflects the results of a theoretically perfect model that correctly predicts all the defaults. That's the best you can do. The dark blue line shows the cumulative percentage of defaulters based upon your estimated model (column F) while the lighter blue line shows the cumulative percentage of nondefaulters (column G).

The black line in the middle reflects a naïve model that identifies defaulters simply at random. In other words, this random “model” has no predictive information content. It is sometimes used as a benchmark when comparing competing models. For a statistical model to have any value at all, it must perform better than a random guess at who would default. The K-S value of 74.1 is the vertical distance between the

dark blue and lighter blue lines.

Although these approaches are commonly used in evaluating accuracy associated with default models, other approaches may also be helpful.

**Classification errors.**

Assuming some probability cutoff value, an account can be classified into one of two buckets—defaults or nondefaults. For example, if the estimated probability of default is greater than or equal to, say, 50%, the account might be assumed to default. If the account's estimated probability is less than 50%, then we have a nondefault. By comparing these classification results to our historical data, we can determine an overall classification error rate as well as the number of false positives (Type I error) and false negatives (Type II error). A false positive is mistakenly predicting an account will default when it actually was a nondefault. Likewise, a false negative is mistakenly predicting an account will not default when it, in fact, actually did. The distribution of false positives and false negatives produced by the

model can have substantial cost implications when applied to an entire portfolio.

**Information entropy ratios.**

These are accuracy measures where two states of uncertainty are compared. For further details, see the second footnote at the end of this article.

**Validating the Facility Model: LGD**

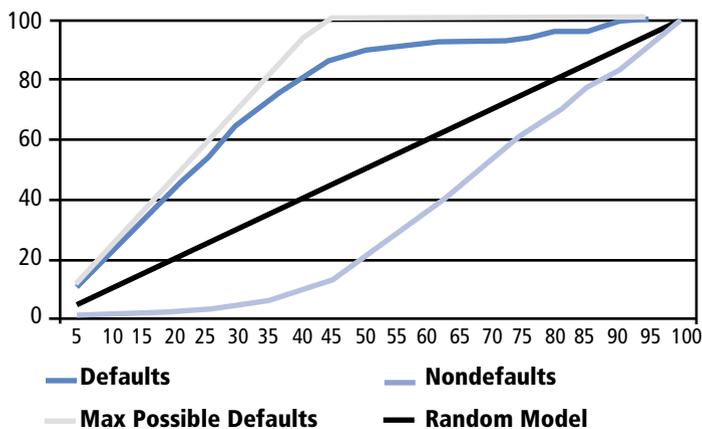
The validation process for the LGD model is a little different from that for the PD model. In the PD model, there are two sub-populations—defaulters and nondefaulters. In the LGD model, only defaulted information is used. For validating LGD models, three items are required: (1) actual dollars recovered, (2) defaulted dollars, and (3) predicted recovery rate. Here is a step-by-step guideline for validating your LGD model:

1. Select the period of time for validation.
2. Put together holdout data if available.
3. Compute actual recovery rate from defaulted dollars and total dollars recovered.
4. Compute Mean Squared Error (MSE). This is your measure of validation accuracy:

$$MSE = \sum(\text{Actual}\% - \text{Predicted}\%)^2 / N - 1$$

where N represents the number of observations in the validation sample. This is just the squared difference between the actual and predicted loss rates, summed up and then divided by the sample size, less one. The lower the MSE, the more accurate the model, all other things remaining equal. Additional measures of

Figure 2



accuracy can also be computed, such as *root mean squared error* (RMSE), *mean absolute deviation* (MAD), and *mean absolute percent error* (MAPE)—formulas for which can be found in any forecasting textbook. In the fictional example shown as Figure 3, LGD validation data is shown for 10 loans with a recovery rate ranging from 0 to 100% (Columns A, B, and C will have a minimum value of 0% and a maximum value of 100%). Column D is column C minus column B, then we square that answer for each loan. Then we sum column D and divide by 9. Loss given default (LGD, column B) is equal to one (100%) minus the recovery rate.

### Summary

From a validation perspective, how good is good enough? Unfortunately, there is no magic answer. On the obligor side, a validation for a model using payment history and delinquency information could produce K-S values of 50 or more. At a bare minimum, it should certainly perform better than predicting defaults randomly. As research has shown, it's hard to make LGD accurate because of the difficulty in obtaining good predictive explanatory variables and the wide variation in collection efforts across collaterals. In general, remember that on the PD side, the higher the K-S, the better, but as for LGD, the one with the lowest MSE wins.

So, how often should your models be redeveloped? That depends on availability of default information in your current model and the stability of your portfolio. If you had only a marginal number of defaults available for model development, then additional default information could improve validation accuracy enough to prompt redevelopment. On the other hand, if the underlying characteristics of your portfolio change from year to year, then

model redevelopment would be recommended. Since it is required by Basel to produce validation studies on a periodic basis anyway, perhaps a wise approach might be to evaluate newer competing models each year and compare their validation results to your existing model—winner takes all.

*The New Basel Capital Accord* references words like *process* or *systems* about 275 times in its 139 pages. This implies a special emphasis on the need to develop a systematic approach to model analytics—an approach that integrates programming requirements from a variety of sources with standardized methods and procedures. The third article in this series will present the development of an analytics platform by SunTrust that integrates these requirements into a Windows-like interface, allowing sophisticated statistical models to be developed, validated, and documented quickly and efficiently. □

Contact Morrison at at [Jeff.Morrison@suntrust.com](mailto:Jeff.Morrison@suntrust.com)

### Notes

1 *Credit Risk Modeling—Design and Applications*, edited by Elizabeth Mays, 1998, Glenlake Publishing Company, .

2 Sobehart, Keenan, and Stei, "Benchmarking Quantitative Default Risk Models: A Validation Methodology," Moody's Investor Services Global Credit Research, March 2000.

Figure 3

**Mean Squared Error**

Actual Recovery Rate	Actual LGD Rate	Predicted LGD Rate	MSE Calculations
A	B	C	D
32.4	67.6	55	158.76
54.5	45.5	33.3	148.84
87.3	12.7	66.4	2883.69
22.3	77.7	17.8	3588.01
43.4	56.6	33.2	547.56
0	100	76.9	533.61
0	100	65.8	1169.64
1.7	98.3	12.7	7327.36
3.8	96.2	95.1	1.21
0	100	87.7	151.29
		MSE =	1834.44