Creating Interval Target Scorecards with Credit Scoring for SAS® Enterprise Miner™

Miguel Maldonado, Susan Haller, Wendy Czika, and Naeem Siddiqi

SAS Institute Inc.

ABSTRACT

Credit Scoring for SAS® Enterprise Miner™ has been widely used to develop binary target probability of default scorecards, which include scorecards for application and behavior scoring. Extending the weight-of-evidence binned scorecard methodology to interval-based targets such as loss given default and exposure at default is a natural step in the evolution of scorecard development methodologies. This paper discusses several new methods available in Credit Scoring for SAS Enterprise Miner that help build scorecards that are based on interval targets. These include cutoffs that convert the interval target to a binary form and an algorithm that maintains the continuous nature of the interval target. Each method is covered in detail, and all methods are compared to determine how they perform against known target values.

INTRODUCTION

In credit scoring, the scorecard format of a model has become the standard for explaining the probability of default or risk because it is easy to understand and interpret, and because it enables analysts to incorporate business expertise during the binning process. However, the nature of a scorecard requires a binary target. This requirement can be easily extended to support certain interval targets by using known transformation methods.

This paper discusses four methods for building a binary target scorecard for continuous targets: three methods that use a cutoff to bin interval targets and one method that creates a new binary target by using the original interval values as weights. These transformation methods, all available in Credit Scoring for SAS Enterprise Miner, are compared across various assessment measures that look at both the transformed binary target and the original interval target. This paper uses a loss given default (LGD) modeling scenario as an example. It also looks at the ability of each method to produce models that can isolate actual dollar losses most effectively—the overriding business goal for any banker who develops LGD models.

SCORECARD DEVELOPMENT PROCESS

In a typical flow for Credit Scoring for SAS Enterprise Miner, the process includes visualizing and exploring the data, binning or grouping the variables, running a logistic regression, and finally transforming the regression results into a points-based, scaled scorecard.

The Interactive Grouping node (IGN) enables you to do the following:

1. Transform an interval target if necessary.

2. Manage the number of levels of each input variable through grouping.

3. Improve the predictive relationships of the input variables through adjustments that are based on business rules.

4. Select predictive variables for consideration in modeling algorithms.

One of the important metrics that are calculated in the IGN is weight of evidence (WOE), which indicates how well each group within a variable separates the events from the non-events; a higher WOE corresponds to better separation. Because this metric is based on event and non-event counts, a binary target is required.

The Scorecard node then performs a logistic regression on the WOE values; the logistic regression also requires a binary target. Each value of the grouped variable is assigned a score by applying a linear transformation of its predicted probability. After you have a scorecard, you can easily calculate a credit score by adding the score points from all variables.

To facilitate the building of binary target models to predict continuous events, such as LGD, some new features have been introduced to the Interactive Grouping node in Credit Scoring for SAS Enterprise Miner 12.3. These features include three cutoff methods and one weighting method for binary transformations. All four methods are discussed in more detail in the next section. Display 1 shows the relevant options, which include scaling options to convert any interval variable to the range [0,1] and additional options specific to the transformation method chosen.
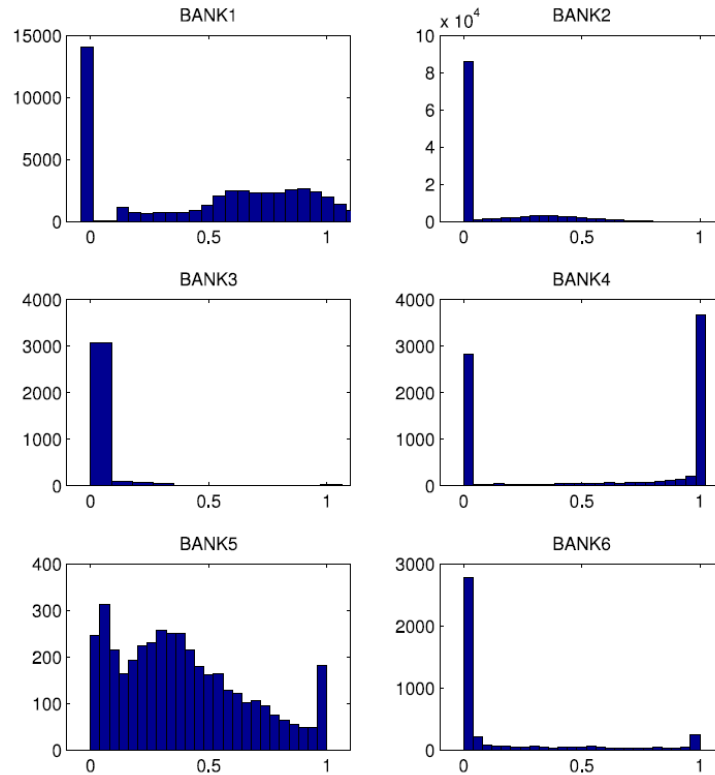
Interval Target Options	
Binary Transformation	
Use Target Min/Max for Scaling	No
Minimum for Scaling	0
Maximum for Scaling	1
Allow Out-of-Range Values	No
Random Seed	12345
Cutoff Value	0.20

2 Scaling Options
3 Additional Options
1 Binary Transformation Methods

Display 1. Interactive Grouping Node Property Sheet

TRANSFORMATION METHODS

Cutoff Methods

Figure 1 shows LGD distributions for portfolios from six actual banks, which are presented by Loterman et al. (2012). For most of these distributions, where bimodality is clear, you can make a case for creating a binary target with two groups by determining a cutoff value for observations. To be considered an event, the interval target must be greater than the cutoff; to be classified as a non-event, the interval target must be less than or equal to the cutoff. Cutoff methods are commonly used to bin modal or bimodal interval targets such as LGD. The distributions in Figure 1 illustrate situations where this method can be applied, except for BANK5. Because BANK5 has no clear mode, its binning and the quality of its model are highly sensitive to the cutoff value. Therefore, BANK5 is not suitable for a cutoff method.

Logic to bin an interval target into a binary target with a cutoff method:

$$if \; Continuous\_Target \; > \; cutoff\_value \quad \rightarrow \quad Event$$
$$else \quad \rightarrow \quad Non\text{-}Event$$

Figure 1. Example LGD Distributions and Logic to Bin Them into Two Groups

The following cutoff methods are available in the Interactive Grouping node:

- The user-defined cutoff method allows any value to be selected as the cutoff. You can use this method to treat an extreme value that is also a mode as one separate bin, or to test any given value as a cutoff.

- The mean cutoff method uses the mean of the interval target as the cutoff. The mean is very sensitive to the skewness of a variable, making it a particularly useful cutoff for interval targets that have bimodal distributions.

- The random cutoff method generates a different random number for each observation to use as the cutoff. Because any observation has the potential to be classified as an event, this cutoff is especially useful as an unbiased alternative baseline to the other two cutoff methods.

Weighting Method

Cutoff methods are not always suitable for interval targets—for example, when you are looking at non-bimodal distributions such as BANK5 in Figure 1. In such cases, a weighting method might be more appropriate. A weighting method enables the continuous nature of the interval target to be maintained. An easy way to apply this method is to duplicate each observation and apply to each duplicate a weight that is based on the LGD. For example, an observation with an LGD of 0.8 can be considered as two separate observations, weighted 0.8 in the event group and weighted 0.2 in the non-event group. Figure 2 illustrates the creation of two weighted cases out of each original case.

Original observations:

ID	Predictors					Interval Target
customer 1	predictor 1	predictor 2	predictor 3	...	predictor *n*	0.8
customer 2	predictor 1	predictor 2	predictor 3	...	predictor *n*	0.3

Transformed into:

ID	Predictors					Weight	Binary Target
customer 1	predictor 1	predictor 2	predictor 3	...	predictor *n*	0.8	1
customer 1	predictor 1	predictor 2	predictor 3	...	predictor *n*	0.2	0
customer 2	predictor 1	predictor 2	predictor 3	...	predictor *n*	0.3	1
customer 2	predictor 1	predictor 2	predictor 3	...	predictor *n*	0.7	0

Figure 2. Two Observations Transformed Using the Weighting Method

Several methods are computationally and mathematically equivalent to the weighting method. One of the most prevalent methods in the credit scoring literature is the events/trials approach that uses the LOGISTIC procedure in SAS/STAT[®].

The LOGISTIC procedure supports a specific syntax for modeling weighted observations. Statistics define an *event* as the number of positive responses of an experiment and a *trial* as the number of times the experiment is performed. To use this syntax with an interval target, you simply specify the target as the event and specify a dummy variable as the trial in order to put the target in context. Because LGD is a continuous variable in the range [0,1], a dummy variable of 1 is created for each observation. The following statements show how to code this for LGD:

```
data LGD_data_with_dummy;
   set LGD_original_data;
   dummy_var = 1;    */ An appropriate dummy variable for LGD in the range [0,1] /*
run;

proc logistic data=LGD_data_with_dummy;
    model LGD / dummy_var = predictor1 – predictor5;
    …
run;
```

An equivalent process is to use the WEIGHT statement in PROC LOGISTIC with transformed data like the data described in Figure 2. The first row of duplicated inputs has a weight based on LGD, and the second row has a weight based on 1 – LGD.

```
proc logistic data=LGD_weighted_data;
    model BinaryTarget = predictor1 – predictor5;
    weight Weight;
run;
```

ASSESSMENT STATISTICS

The methods in the previous section imply either transformation of the target or weighting of duplicate observations. The scorecard generates posterior probabilities that correspond to the original interval target, along with assessment statistics that are based on the transformed binary target. Examples of these assessment statistics are misclassification, asymptotic standard error (ASE), and receiver operating characteristic (ROC). The assessment needs to be complemented with other measures that are based on the original interval target to ensure that the model addresses the true business need of capturing loss. This paper uses the following assessment statistics for the interval target:

- Spearman correlation: This nonparametric correlation measure is calculated between the posterior probabilities from the logistic regression (which represents the predicted interval target) and the actual interval target values.

- Cumulative accuracy profile (CAP) curve for captured loss: For the specific case of LGD, Van Berkel and Siddiqi (2012) introduced a CAP plot to address a business concern—captured loss of a predictive model. The data are sorted by descending predicted LGD, and cumulative proportions for loss and balance at default are plotted as shown in Display 2. Better models have more area under the curve. Display 2 illustrates two models; both models perform better than baseline, which is represented by the dashed purple line. Model 1 (top line, in blue) captures more loss than model 2 (middle line, in red), so model 1 represents the better model.

Cumulative proportion of Balance at Default
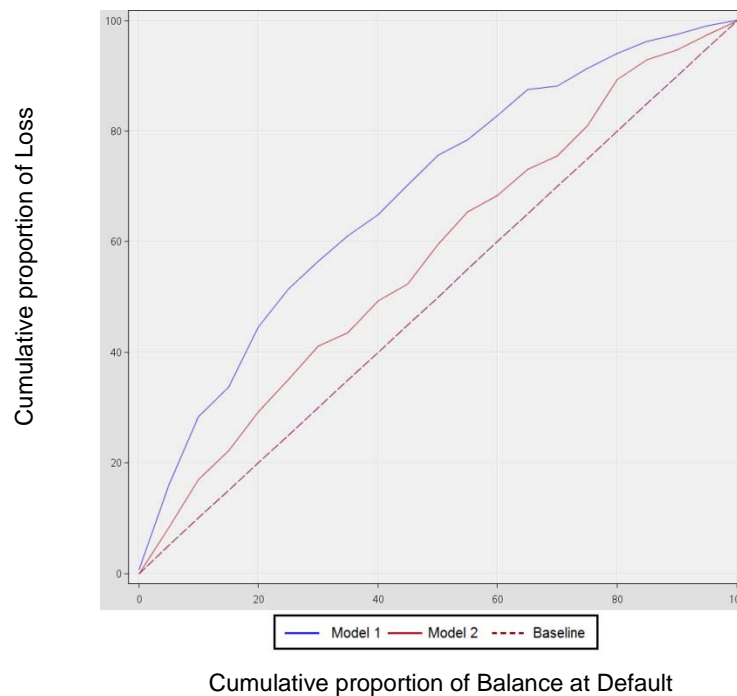
Display 2. CAP Curve for Captured Loss

REAL DATA EXAMPLE: MODELING LGD OF A BANK

This example is based on the same data and groupings that Van Berkel and Siddiqi (2012) use to explain LGD from an overdraft portfolio of a major Canadian bank.

They observed accounts that were in good standing on October 31, 2009, and defaulted for the first time at some point during the following 12 months. LGD was defined as the ultimate loss two years later (October 31, 2011) divided by the balance at the time of the first default in the observation period. The selected input variables in the final scorecard included demographic information from applications, credit bureau information, and performance both on the overdraft account and on other accounts at the same bank.

EXAMPLE 1: ORIGINAL LGD DISTRIBUTION

The first steps in this example are to explore the interval target to find any modes or distribution skewness that could affect binary transformations and to determine good user-defined cutoffs to test for binning. The histogram of the LGD, in Display 3, shows two modes at LGD=0 and LGD=1, a very common characteristic in LGD data.
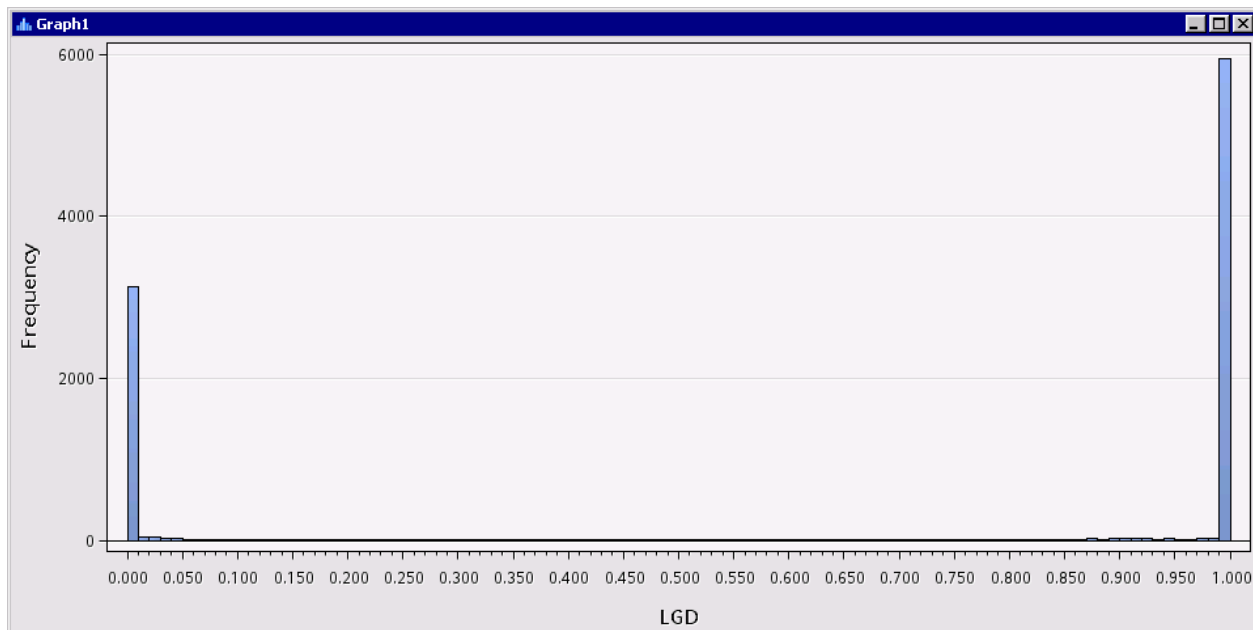
Display 3. Histogram of the Interval Target

The next step is to run a Data Partition node to divide the data into training and validation subsets, followed by four Interactive Grouping nodes, one for each method described in the section "Transformation Methods." The user-defined cutoff is set at 0.8. The mean value for these data is 0.64. For each transformation method, predefined groupings from Van Berkel and Siddiqi (2012) are used in the IGN to ensure consistency across the scorecards and to make the comparisons easier. These definitions are based on statistical considerations and business rules.

As shown in Figure 3, a Scorecard node is connected to each of the IGNs to produce a scorecard for each transformation method. The final step is to compute assessment statistics for both the binary and the interval targets. Misclassification and an ROC index that are based on the binary target are available in the Model Comparison node (labeled Binary Target Assessment); captured loss and Spearman correlation that are based on the original, interval target are calculated with a SAS Code node (labeled Interval Target Assessment).
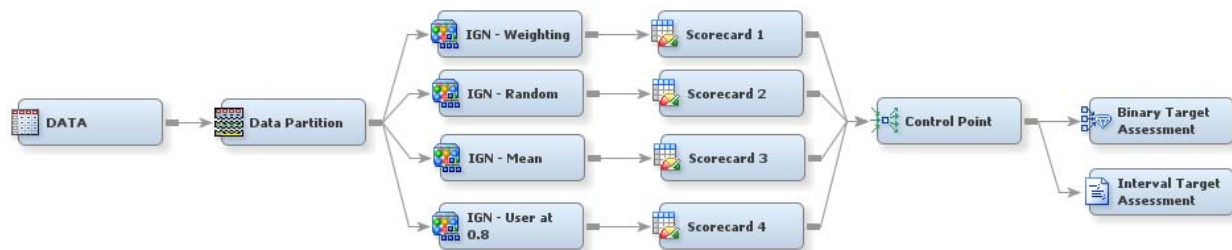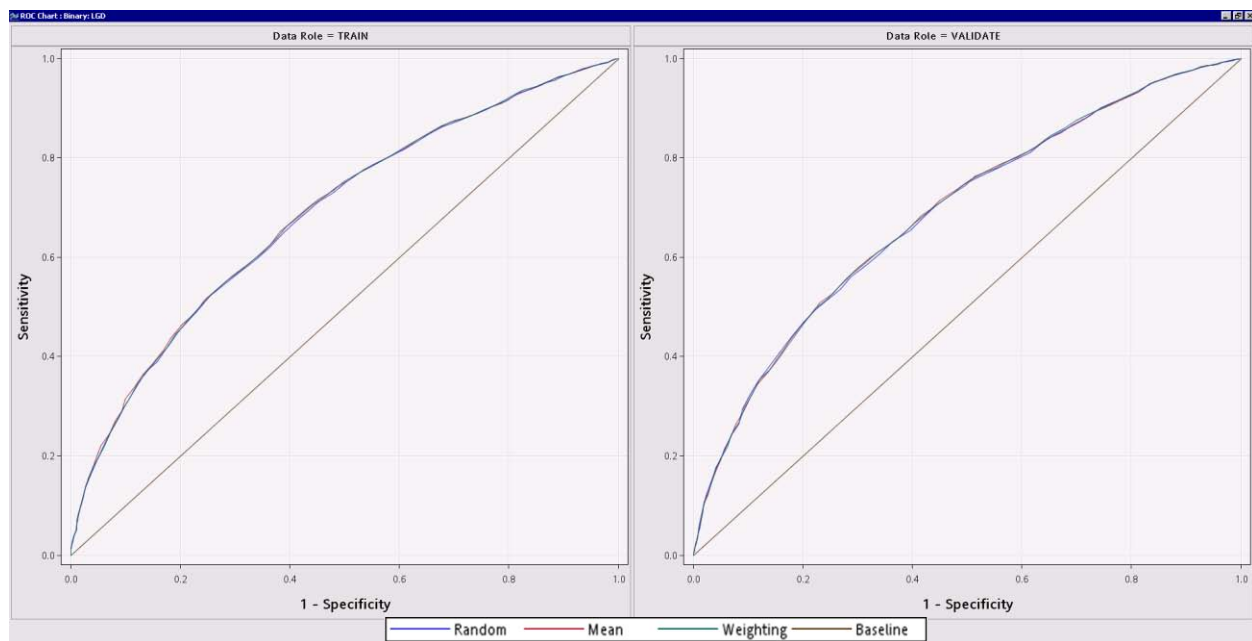
Figure 3. Project Diagram to Compare Binary Transformation Methods

All methods perform equally well based on misclassification and the ROC index, as shown in Display 4 and Table 1.

Display 4. ROC Curve for Binary Target

Transformation Method	Binary Target			
	Misclassification		ROC	
	Valid	Train	Valid	Train
Mean	0.3347	0.3316	0.69	0.69
Random	0.3333	0.3328	0.69	0.68
User at 0.8	0.3426	0.3415	0.69	0.68
Weighting	0.3282	0.3296	0.69	0.69

**Table 1. Binary Target Statistics Summary**

Although misclassification and the ROC index help determine which method best predicts the binary target, it is also necessary to assess the original, interval-based LGD. This is done using the SAS code in Appendix A to calculate the Spearman correlation for both training and validation data. Appendix A also includes the code to generate an LGD-specific CAP curve for captured loss to address a business need of having a model that is predictive and also captures loss.

7

All four target transformation methods have approximately the same correlation between the actual and predicted LGD, as shown in Table 2. The user-defined cutoff method at 0.8 slightly, but not significantly, outperforms the others.

Transformation Method	Interval Target	
	Spearman Correlation	
	Valid	Train
Mean	0.3081	0.3058
Random	0.3084	0.3057
User at 0.8	0.3123	0.3109
Weighting	0.3088	0.3064

Table 2. Fit Statistics for Interval Target

The CAP curve comparison for captured loss in Display 5 shows identical captured loss curves for all four methods. Each has an area under the curve of 0.21 for the training data and 0.24 for the validation data.

Cumulative Proportion of Balance at Default
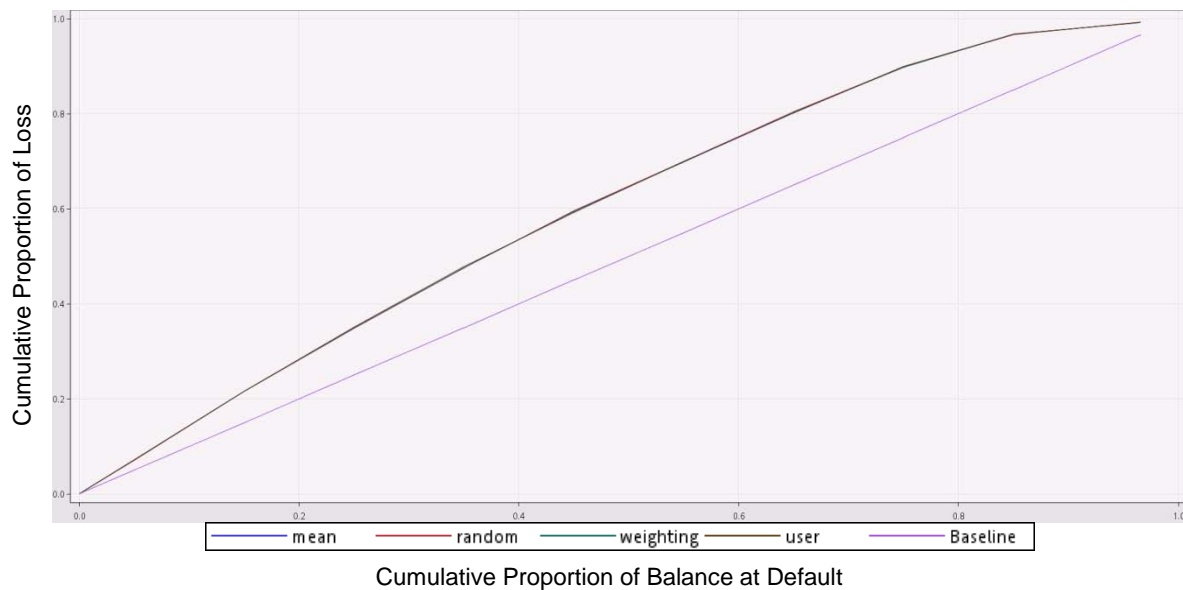
Display 5. CAP Curve for Captured Loss

No method outperforms another across both binary and interval criteria, and all the methods produce equal results for the captured loss plot shown in Display 5. These results are anticipated because of the extreme separation of the modes at 0 and 1. You would expect more differentiation as the modes move away from the extremes.

From a business perspective, it is also useful to examine a segment with a high LGD. Based on the LGD distribution in these data, the segment with LGD greater than 0.8 is analyzed, where 97% of the loss occurs for these data.

The same process that was performed on all the original data is repeated for this filtered segment to compare the weighting-based, random cutoff, and mean cutoff transformation methods, as shown in Figure 4.

Figure 4. Diagram for LGD > 0.8

For this subset of the data, the mean cutoff method (with a mean value of 0.995) has a more consistent ROC curve between training and validation, but its misclassification is considerably higher than that of the other methods, as shown in Display 6.

Display 6. ROC Plot for Binary Target

The weighting-based method has low misclassification across both training and validation and also has a consistent ROC index for both partitions, as shown in Table 3.

| Transformation Method	Binary Target			
	Misclassification		ROC	
	Valid	Train	Valid	Train
Mean	0.0473	0.0536	0.64	0.77
Random	0.0043	0.0060	0.63	0.94
Weighting	0.0044	0.0047	0.61	0.77

Table 3. Binary Target Assessment for the High-LGD Segment

The weighting-based and mean cutoff methods have similar Spearman correlations, as shown in Table 4.

Transformation Method	Interval Target	
	Spearman Correlation	
	Valid	Train
Mean	0.1038	0.2144
Random	0.0375	0.0885
Weighting	0.0914	0.2142

**Table 4. Interval Target Statistics for the High-LGD Segment**

The CAP curves for captured loss for this subset of the data, shown in Display 7, are lower than the curves for the complete data shown in Display 5. The mean cutoff and weighting-based methods have an area under the curve of 0.01, and the random cutoff method has an area under the curve of 0. When such a high proportion of the data have LGD=1, all the binary transformation methods produce extremely rare "non-events," which is causing the poor performance of these models. Because of this poor performance, you might want to next explore how the different transformation methods perform with data where the large modes at 0 and 1 are not as extreme. The results are shown in the next section.

Cumulative Proportion of Balance at Default
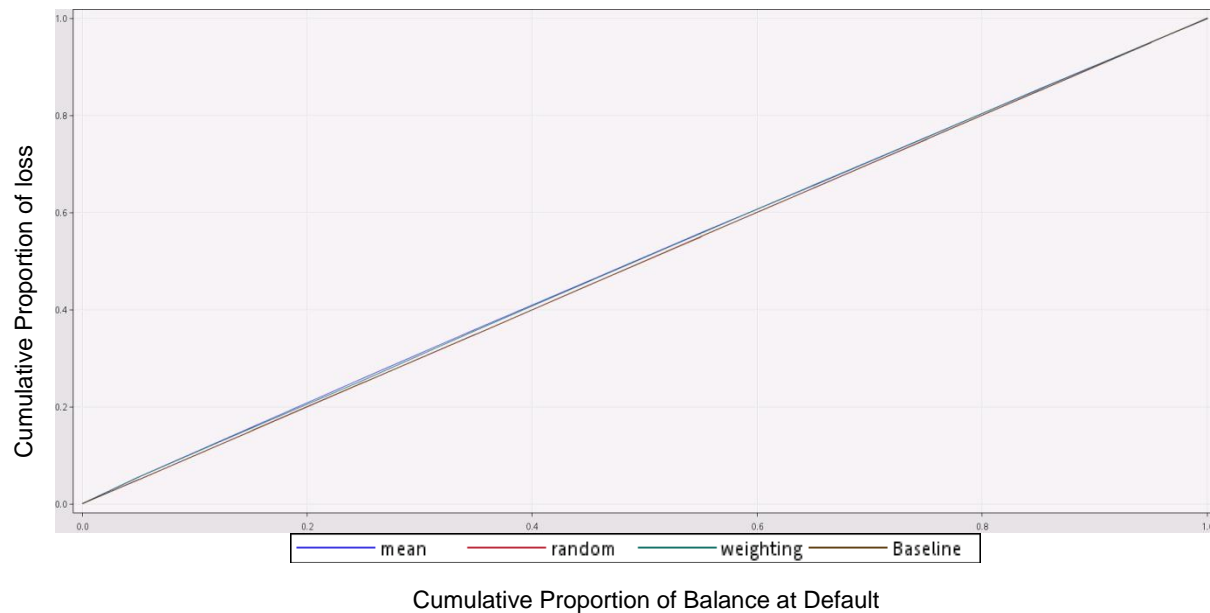
Display 7. CAP Curve for Captured Loss for the High-LGD Segment
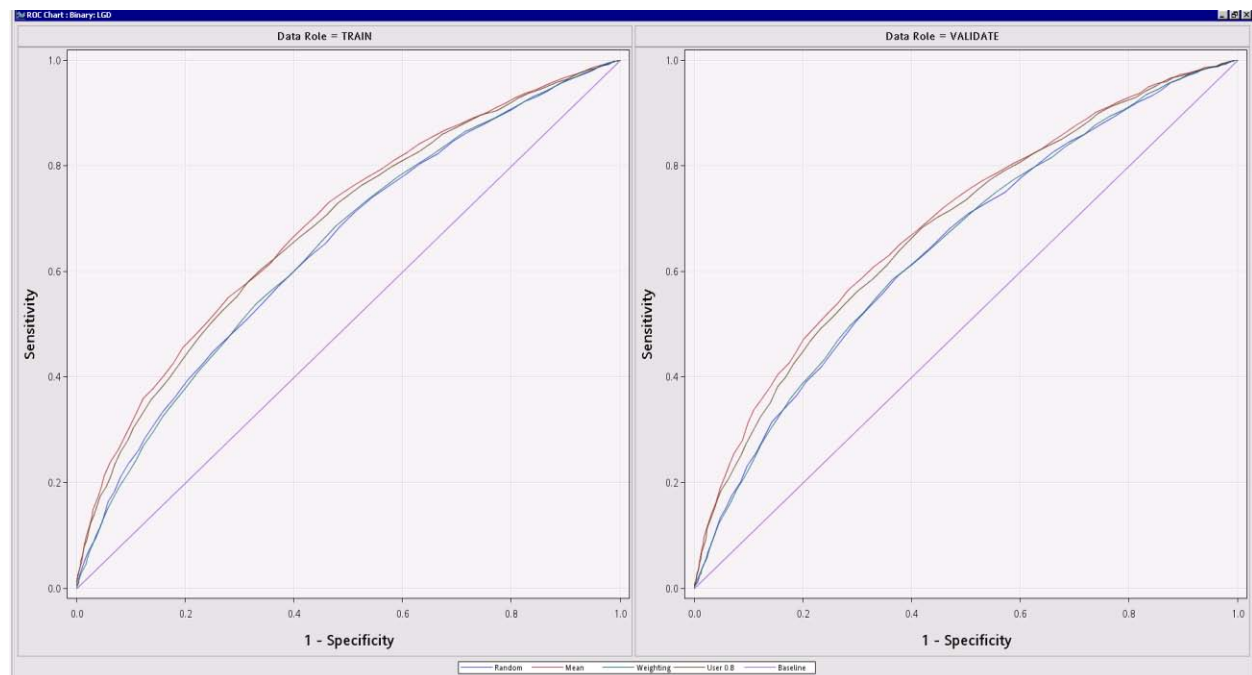
EXAMPLE 2: MODIFIED LGD DISTRIBUTION

This example examines data that have the LGD distribution shown in Display 8.
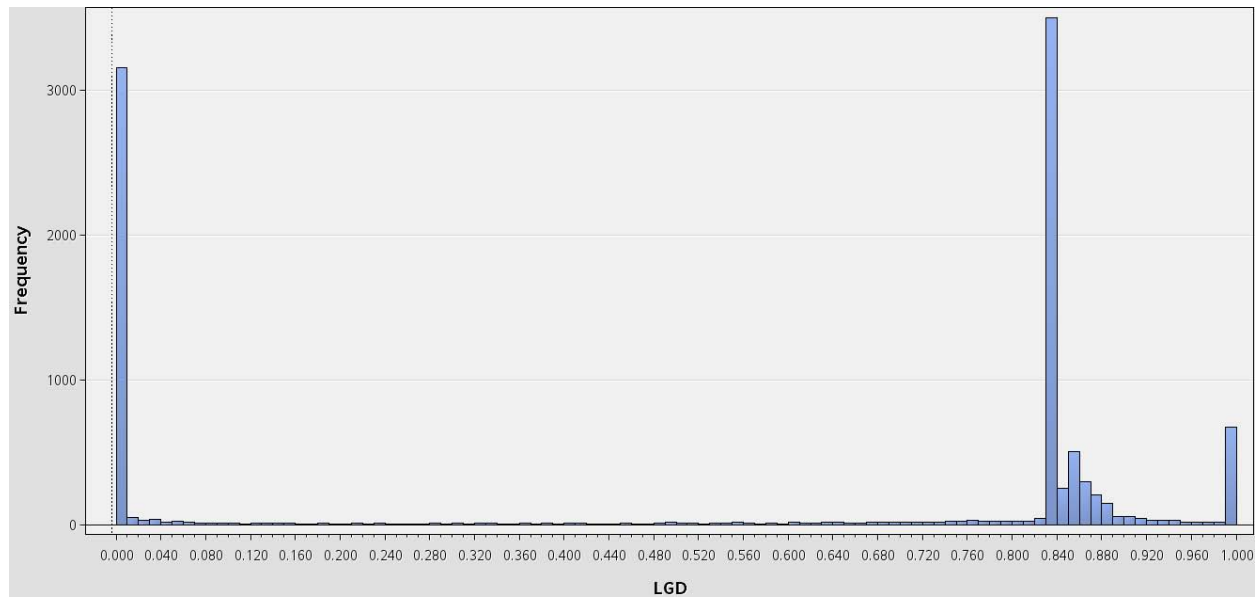
Display 8. Histogram of LGD

The same flow as in Figure 3 is run for these new data, with the following results.

- Based on the binary target assessment measures shown in Display 9 and Table 5, the mean and user-defined cutoff methods performed best.

Display 9. ROC Curve for Binary Target

| Transformation Method	Binary Target			
	Misclassification		ROC index	
	Valid	Train	Valid	Train
Mean	0.3395	0.3326	0.69	0.69
Random	0.3874	0.3902	0.65	0.65
User at LGD=0.8	0.3559	0.3559	0.68	0.68
Weighting	0.3873	0.3862	0.65	0.68

Table 5. Binary Target Statistics for Example Data

- All four target transformation methods have approximately the same Spearman correlation between the predicted and actual LGD, as shown in Table 6.

Transformation Method	Interval Target	
	Spearman Correlation	
	Valid	Train
Mean	0.2746	0.2629
Random	0.2749	0.2630
User at LGD=0.8	0.2795	0.2671
Weighting	0.2762	0.2655

Table 6. Fit Statistics for Interval Target

- The CAP curve comparison for captured loss in Display 10 shows that all four methods perform the same. Each has an area under the curve of 0.21 for the training data. Similar results are observed in the validation data, where all four methods have an area under the curve of 0.24.This is a noticeable improvement over the curves for the segmented data, which are shown in Display 7.

Cumulative Proportion of Balance at Default
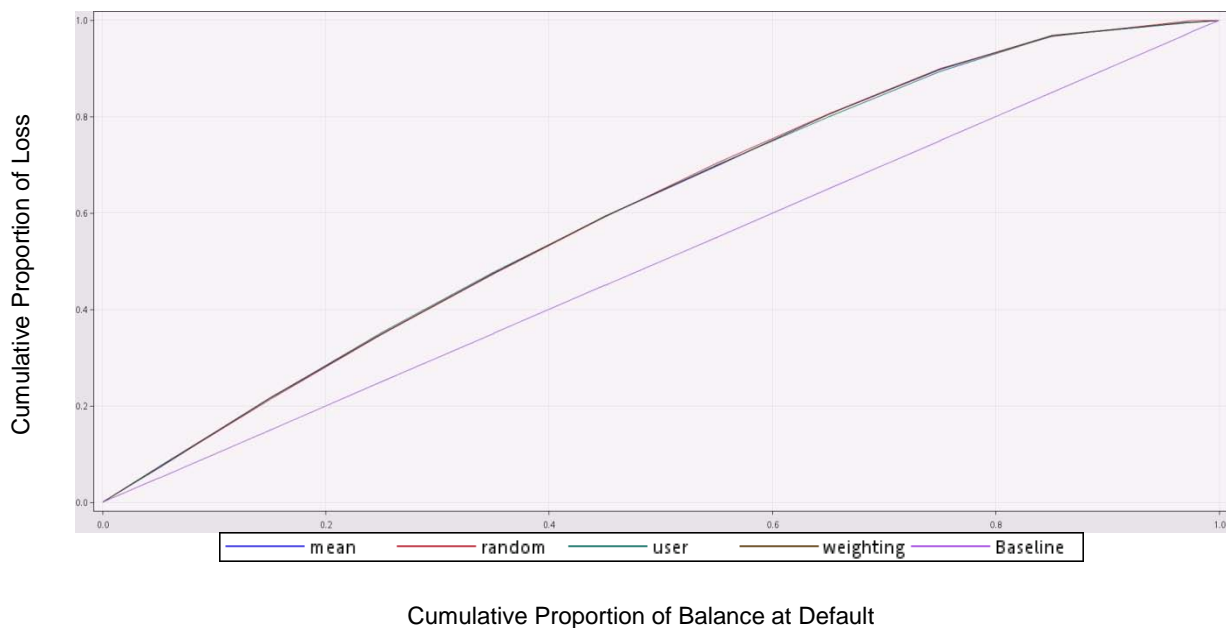
Display 10. CAP Curve for Captured Loss on the Validation Data

As expected, the performance of the models improves when the heavy concentration of LGD values at 1 is dispersed and more non-events are created. This distribution of data, and data even more uniformly distributed, might be more

typical of non-LGD values such as utilization, which can also be analyzed in this framework. It is also expected that the transformation methods will perform differently for non-bimodal data, where one method might significantly outperform the others. This possibility presents the need for the assessment statistics used in the analysis for both the transformed binary target and the original interval target.

CONCLUSION

For interval targets that have certain distributions, the transformation methods can work equally well to convert the values from interval to binary, enabling you to apply the typical and well-known scorecard process. In other cases, assessment statistics can provide guidance about the best transformation method to use. Which method is most appropriate depends on the distribution of your interval target.

Experienced credit scoring analysts will find value in transferring their expertise from binary target scorecards to interval target scorecards. New functionality in Credit Scoring for SAS Enterprise Miner makes this transfer easier and enables you to gain knowledge about all dimensions of expected loss. This functionality also enables you to better understand and plan strategies for loss given default and exposure at default.

Analysts from other industries will also benefit from the new support that Credit Scoring for SAS Enterprise Miner offers for creating scorecards for transformed interval targets. The scorecard framework summarizes the model in an easy-to-interpret way in which predictive power and accuracy are not compromised.

APPENDIX A

The following code plots the CAP curve for captured loss and calculates both the area under this curve and the Spearman correlation.

```
/* This sort is needed for the weighting-based method to remove the duplicated
observations */
proc sort data=&EM_IMPORT_DATA out=temp nodupkey;
   by credit_id;
run;

%EM_REGISTER(key=TOPLOT, type=DATA);
%EM_REGISTER(key=AUCTABLE, type=DATA);

proc sort data=temp out=&EM_USER_toplot;
by descending P_BIN_LGD1;
run;

proc corr data=temp spearman outs=spcorr noprint;
   var LGD; with p_BIN_LGD1;
run;

data inttgtstats;
   length STAT $8 Label $200;
   set spcorr(where=(_type_="CORR"));
   _type_="_SPCORR_";
   label="Spearman Correlation";
   keep _type_ label LGD;
run;

proc sql;
    select sum(a.defbal),sum(a.loss) into :total_defbal,
            :total_loss
     from temp as a;
quit;

data &EM_USER_toplot(keep=defbal loss P_BIN_LGD1 cum_defbal cum_loss roundedvar
method) &EM_USER_AUCTABLE(keep=AUC stat label);
   set &EM_USER_toplot end=eof;
   cum_defbal + defbal/&total_defbal;
   cum_loss + loss/&total_loss;
   roundedvar=round(cum_defbal,0.1);
```

```
        AUC + (lag(cum_loss)+cum_loss)*dif(cum_defbal);

        if eof then do;
         AUC=AUC-1;
         output &EM_USER_AUCTABLE;
         STAT="_Gini_";
         Label="Gini Index";
        end;

        if lag(roundedvar) ne roundedvar then do;
         output &EM_USER_toplot;
         method="Weighting";
        end;

    run;

    %em_report( key=TOPLOT, viewtype=Lineplot,  x=cum_defbal, y=cum_loss,
    autodisplay=Y, description=%bquote(CAP curve), block=%bquote(My Graphs) );


    %em_report( key=AUCTABLE, viewtype=DATA, autodisplay=Y,
    description=%bquote(AUC calculation), block=%bquote(My Graphs) );
```

REFERENCES

Baesens, B. 2011. *Credit Risk Modeling Using SAS® Course Notes.* Cary, NC: SAS Institute Inc.

Loterman, G., Brown, I., Martens, D., Mues, C., and Baesens, B. 2012. "Benchmarking Regression Algorithms for Loss Given Default Modeling." *International Journal of Forecasting* 28:161–170.

Van Berkel, A. and Siddiqi, N. 2012. "Building Loss Given Default Scorecard Using Weight of Evidence Bins in SAS Enterprise Miner." *Proceedings of the SAS Global Forum 2012 Conference.* Cary, NC: SAS Institute Inc. Available at http://support.sas.com/resources/papers/proceedings12/141-2012.pdf.

Yang, B. and Tkachenko, M. 2012. "Modeling Exposure at Default and Loss Given Default: Empirical Approaches and Technical Implementation." *Journal of Credit Risk* 8:81–102.

RECOMMENDED READING

- *Base SAS® Procedures Guide*

- *Credit Risk Modeling Using SAS®*

ACKNOWLEDGMENTS

The authors would like to thank Laura Ryan, Leslie Anderson, Anthony Van Berkel, and Jared Dean for their contributions to this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Miguel M. Maldonado	MiguelM.Maldonado@sas.com
Susan Haller	Susan.Haller@sas.com
Wendy Czika	Wendy.Czika@sas.com
Naeem Siddiqi	Naeem.Siddiqi@sas.com