Paper 078-2007

IMPROVING CREDIT SCORING BY GENERALIZED ADDITIVE MODEL

Wensui Liu, Cincinnati Children's Hospital Medical Center, Cincinnati, OH
Jimmy Cela, Equifax Inc., Atlanta, GA

ABSTRACT

Logistic Regression has been widely used in the financial service industry for credit scoring models. Despite its advantages in easy interpretation and low computing cost, Logistic Regression is under the criticism of failure to model the nonlinear features of the predictors effect on the dependent variable and therefore might lead to unsatisfactory results. Modern statistical techniques such as Neural Network and Projection Pursuit Regression have been proven successful in the nonlinear modeling. However, this success comes with the price of interpretability. Introduced by Hastie and Tibshirani, Generalized Additive Model provides the ability to detect the nonlinear patterns without sacrificing interpretability. The purpose of this paper is to introduce Generalize Additive Model as a promising alternative to Logistic Regression for credit scoring. The modeling strategies for both Logistic Regression and Generalized Additive Model with SAS/STAT are investigated to guide readers from Logistic Regression through Generalized Additive Model. The comparison of predictive performance of the two modeling techniques shows the superiority of the Generalized Additive Model. We hope that statistical analysts will be able to implement Generalize Additive Model with SAS/STAT in real business world by following the practical guidance outlined in this paper.

INTRODUCTION

How to assess the credit risk has become crucial for credit card companies and commercial banks due to the explosive growth of credit market in recent years. Statistical models, such as Linear Discriminant Analysis, Logistic Regression, Classification and Regression Tree, and Neural Network, are widely used to evaluate the credit worthiness of potential borrowers in order to reduce the default risk (Franke, Hardle, and Stahl 2000, Shao 2004).

Logistic Regression, which is a special case of Generalized Linear Models (McCullagh and Nelder 1989), is the most widely used statistical model in the credit scoring industry. Introduced by McCullagh and Nelder, Generalized Linear Models provide a unified framework to model response from any member of the exponential family distributions, such as Gaussian, Binomial, or Poisson. In Generalized Linear Model, the dependent variable **Y** is related to the linear combination of predictors $\mathbf{B_1{*}X_1{+}...{+}B_m{*}X_m}$ in the form of $\mathbf{G(E(Y|X)) = G(u) = B_0{+}B_1{*}X_1{+}...{+}B_m{*}X_m}$, where **u** is the mean of dependent variable **Y** and **G(.)** is a monotonic differentiable function known as Link Function. For Logistic Regression, Generalized Linear Model can be expressed as $\mathbf{Logit(p) = Log(p/(1{-}p)) = B_0{+}B_1{*}X_1{+}...{+}B_m{*}X_m}$, where **u** is **p = Prob(Y=1|X)** and **G(.)** is **Logit(.)** function in this case. In SAS/STAT, both LOGISTIC and GENMOD procedures can be used to build Logistic Regression. While LOGISTIC procedure is specifically designed for Logistic Regression and provides a variety of features for model selection and diagnosis, GENMOD procedure can be used to fit a wide range of Generalized Linear Models with response from any exponential distribution. Considering the popularity of Logistic Regression in current business applications, we will use it as a benchmark model and compare it with Generalized Additive Model.

In Generalized Linear Model, the relationship between response and predictors is assumed to be linear. However, a potential risk of such assumption is model misspecification. While the effects of predictors are often nonlinear in real world, it is always challenging to find an appropriate functional form of the partial effect of predictors on the response variable. As a result, Generalized Linear Model might not always be able to provide an appropriate fit for the data of complex structures. Proposed by Hastie and Tibshirani, Generalized Additive Model (Hastie and Tibshirani 1990) relaxes the linearity assumption of Generalized Linear Model and assumes that the dependent variable Y is dependent on the univariate smooth terms of predictors rather than predictors themselves. Therefore, the functional form of Generalized Linear Model $\mathbf{G(E(Y|X)) = B_0{+}B_1{*}X_1{+}...{+}B_m{*}X_m}$ could be further extended to become $\mathbf{G(E(Y|X)) = B_0{+}S_1(X_1){+}...{+}S_m(X_m)}$, where $\mathbf{B_i{*}X_i}$ is replaced by a nonparametric smooth function $\mathbf{S_i(X_i)}$ for predictor $\mathbf{X_i}$.

The function **S(.)** is estimated in a flexible manner. For nonlinear terms, any nonparametric smoothing method can be used. In SAS/STAT, GAM procedure is using B-spline and local regression for univariate smoothing and thin-plate for bivariate. However, the function **S(.)** doesn't have to be nonlinear for all predictors in Generalized Additive Model. In practice, it is more often to mix linear terms with nonlinear ones, which gives a semi-parametric variation of Generalized Additive Model $\mathbf{G(E(Y|X)) = B_0{+}B_1{*}X_1{+}...{+}B_m{*}X_m{+}S_{m+1}(X_{m+1}){+}...{+}S_{m+n}(X_{m+n})}$. Since each individual effect is estimated using univariate smoother, the Curse of Dimensionality is avoided. Moreover, the important feature of interpretability in Generalized Linear Model is retained in Generalized Additive Model. The function **S(.)** is the analogy of co-efficient in Generalized Linear Model. The estimate of $\mathbf{S_i(X_i)}$ explains how the response changes along with the corresponding predictor $\mathbf{X_i}$.

With flexibility and interpretability discussed above, Generalized Additive Model is able to serve dual purposes. On one hand, Generalized Additive Model can help visualize nonlinear effects and be considered a powerful tool for data exploration and feature discovery. On the other, Generalized Additive Model can be used directly as a predictive modeling tool when over-fitting is carefully guarded, i.e. by limiting the degrees of freedom in smooth terms. The over-fitting problem due to over-complex smooth terms is a main issue in applications of Generalized Additive Model. A conservative choice of Degrees of Freedom in smooth terms is preferred.

DATA DESCRIPTION

The data analyzed in this paper has come from a French bank and been used by Muller (Muller, 2000) to demonstrate an application of Generalized Partial Linear Model. The data set consists of 6,180 cases and 24 variables. The response variable Y reflects the status of loan and it has been coded as 1 for "bad" and 0 for "good" loans. Predictors include eight numeric variables (X2 - X9) and fifteen categorical variables (X10 – X24) with levels ranging from 2 to 11. However, all predictors have been unlabelled and no prior information about the data is known. Outliers in the data set have been removed such that X2 – X9 are all within the range between -3 and 3. In order to compare the performance of out-of-sample prediction between Logistic Regression and Generalized Additive Model, we randomly divided the data into 2 parts, one for model training and the other for model testing, as shown in **Table 2.1**.

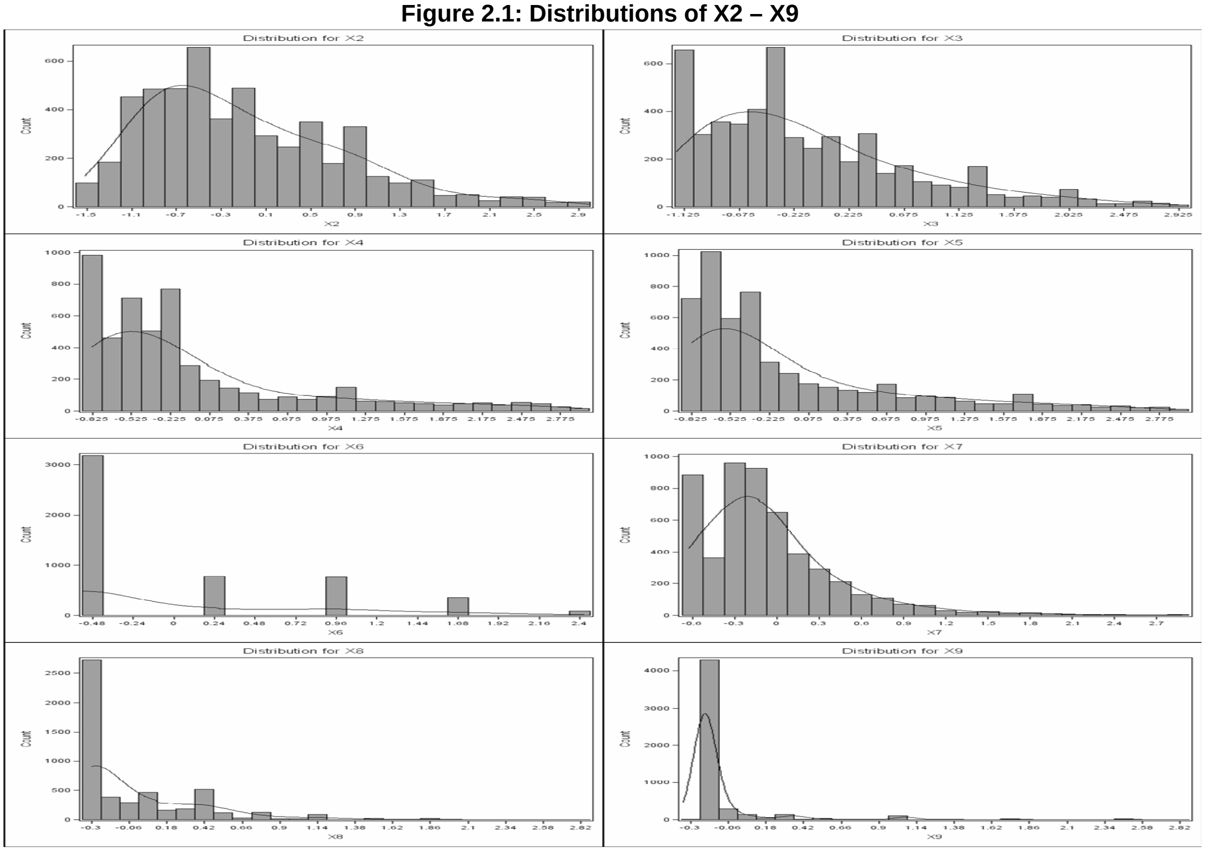
Table 2.1: Responses in Data Sets

	Full Data	Training	Testing
1	372 (6.02%)	312 (6.02%)	60 (6%)
0	5,808 (93.98%)	4,868 (93.98%)	940 (94%)

The first step of data exploration is to visualize the distributions of numeric predictors. Empirical distributions of X2 – X9 in training data set are estimated with kernel density estimation using KDE procedure and shown in **Figure 2.1**. The statement ODS GRAPHICS ON is used to create histogram and density estimate curve for the variable.

```
ods graphics on;
proc kde data = train;
  univar X2 X3 X4 X5 X6 X7 X8 X9;
run;
ods graphics off;
```

Figure 2.1: Distributions of X2 – X9



Clearly, all numeric predictors are highly skewed to the left, as shown in **Figure 2.1**. For X6, X8, and X9, the observation that one value covers the majority of cases is consistent with Muller's finding of quasi-discrete structure in these 3 variables. However, there is no compelling evidence of such structure in X7, which is more similar to X2 – X5 in nature (see **Figure 2.1**).

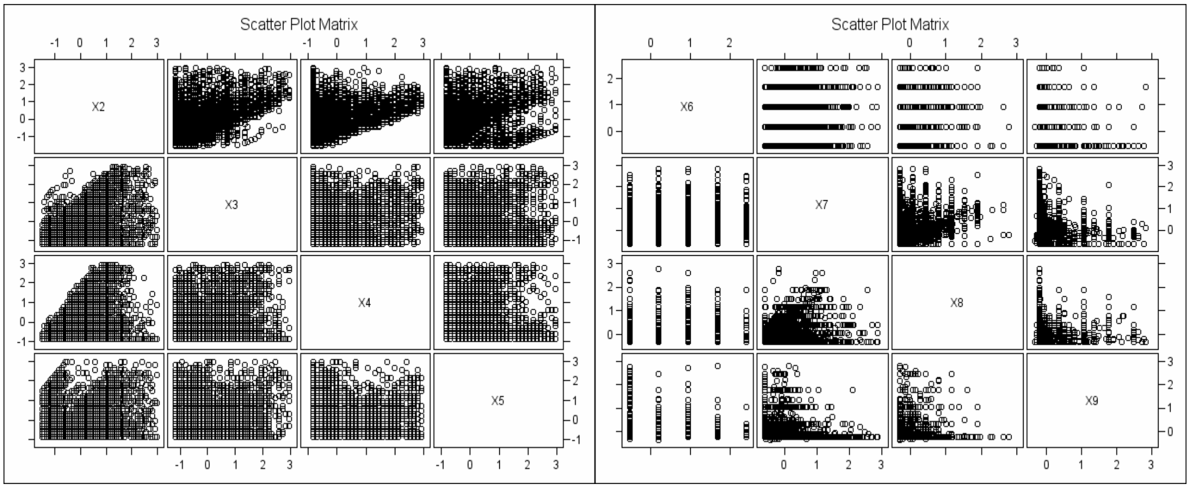
The correlations between numeric predictors X2 – X9 are also investigated using CORR procedure and the result is presented in Table 2. The option PLOTS = MATRIX in PROC CORR statement creates a scatter plot matrix to visualize the relationship between every pair of variables, which is appended to **Table 2.2**.

ods graphics on;

```
proc corr data = train plots = matrix;
  var X2 - X9;
run;
ods graphics off;
```

Table 2.2: Correlation Matrix of X2 – X9

	X2	X3	X4	X5	X6	X7	X8	X9
X2	1.00000	0.43836	0.36896	0.25764	0.28118	0.33288	0.08880	0.04168
X3	0.43836	1.00000	0.35755	0.23702	0.20237	0.19693	0.03667	0.04759
X4	0.36896	0.35755	1.00000	0.16029	0.30705	0.28439	0.03967	-0.01036
X5	0.25764	0.23702	0.16029	1.00000	0.04856	0.00224	-0.02308	0.04368
X6	0.28118	0.20237	0.30705	0.04856	1.00000	0.35182	0.05854	-0.03547
X7	0.33288	0.19693	0.28439	0.00224	0.35182	1.00000	0.22544	-0.02687
X8	0.08880	0.03667	0.03967	-0.02308	0.05854	0.22544	1.00000	0.05766
X9	0.04168	0.04759	-0.01036	0.04368	-0.03547	-0.02687	0.05766	1.00000



Only weak correlation coefficients exist between predictors based on Guilford's suggested interpretation. Therefore, the impact of multicollinearity on parameter estimates is not a major concern.

In many applications of credit scoring, important variables are often selected before used in predictive models. In SAS/STAT, LOGISTIC procedure offers backward, forward, and stepwise methods for variable selection. However, none of them are statistically sound and they all tend to give biased results with the presence of noise. In our study, an importance rank of predictors in logistic regression is assessed using a self-written SAS macro based on bootstrapping.

```
%macro LogitBoot(data = , dv = , iv = , class = , n = );
proc sql noprint;
  create table logit_result
    (iv   char(10),        prob num format = 6.4,
     sig1 num format = 4., sig2 num format = 4.,
     sig3 num format = 4., sig4 num format = 4.);
  select count(*) into :sample from &data;
quit;
%do i = 1 %to &n;
proc surveyselect data = &data method = urs out = &data._tmp n = &sample
noprint;
```

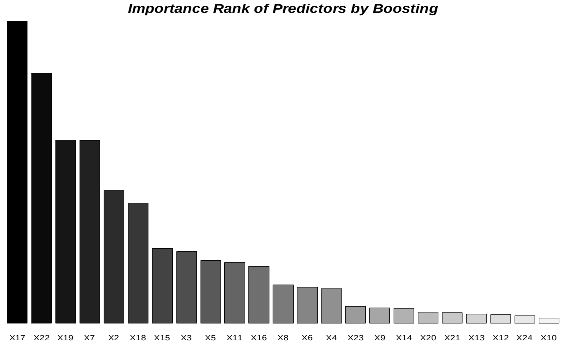
```
run;

ods output type3 = model_tmp;
proc logistic data = &data._tmp desc;
  freq numberhits;
  class &class;
  model &dv = &iv;
run;
proc sql;
  insert into logit_result
  select
    upcase(effect) as iv, ProbChiSq as prob,
    case when ProbChiSq <= 0.01 then 1 else 0 end as sig1,
    case when ProbChiSq > 0.01 and ProbChiSq <= 0.05 then 1 else 0 end as sig2,
    case when ProbChiSq > 0.05 and ProbChiSq <= 0.1 then 1 else 0 end as sig3,
    case when ProbChiSq > 0.1 then 1 else 0 end as sig4
  from model_tmp;
quit;
%end;
proc summary data = logit_result nway;
  class iv;
  output out = out_table (drop = _type_ rename = (_freq_ = count))
  sum(sig1) = sum(sig2) = sum(sig3) = sum(sig4) = ;
run;
%mend LogitBoot;
```

Table 2.3: Importance Rank of Predictors X2 – X24

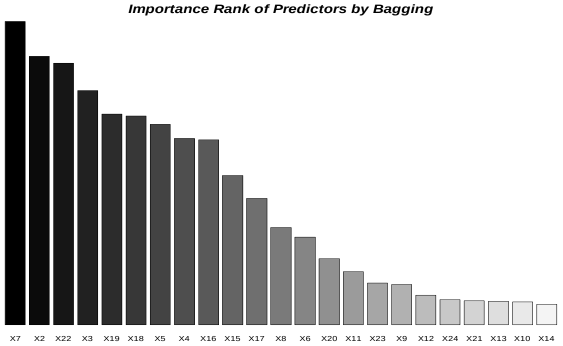
Importance Rank of Predictors

Predictors	Number of 1% significant	Number of 5% significant	Number of 10% significant	Number of Insignificant	Total Models
X11	1000	0	0	0	1000
X17	1000	0	0	0	1000
X8	958	34	6	2	1000
X3	870	88	22	20	1000
X18	791	136	39	34	1000
X15	630	197	71	102	1000
X19	556	255	85	104	1000
X9	516	334	97	53	1000
X22	442	241	95	222	1000
X16	275	230	132	363	1000
X13	267	215	126	392	1000
X12	243	181	126	450	1000
X2	238	227	130	405	1000
X21	164	182	138	516	1000
X14	162	169	103	566	1000
X6	131	171	108	590	1000
X23	79	149	127	645	1000
X24	76	117	103	704	1000
X10	48	80	90	782	1000
X20	42	103	112	743	1000
X7	38	102	93	767	1000
X5	23	71	67	839	1000
X4	15	68	68	849	1000

Importance Rank of Predictors by Boosting



Importance Rank of Predictors by Bagging



The Logistic Regression with all predictors is run 1000 times on random samples drawn with replacement from training data and the frequency of significance levels of 1%, 5%, and 10% from all runs is reported for each predictor, as shown in **Table 2.3**. For instance, X8 is 958 times at 1% level, 34 times at 5% level and 6 times at 10% level of significance out of 1000 runs and therefore is the 3rd most important predictor next to variables X17 and X11. Note that X4, X5, and X7 are the least important predictors. However, one problem with the method described here is that logistic regression is evaluated within the linear framework of the mean structure and nonlinear patterns might be still missed. In order to have a better understanding of predictors, data mining techniques not available in SAS/STAT are also used to help discover important nonlinear features, namely boosting (Friedman 1999) and bagging (Briemen 1994). The R language (www.r-project.org) provides packages "gbm" (developed by Greg Ridgeway) and "randomForest" (developed by Andy Liaw) to implement boosting and bagging respectively. Details in R language, boosting, and bagging are out of the scope of this paper and therefore won't be further discussed. The importance ranks of predictors from both boosting and bagging are reported as histograms attached in **Table 2.3**. Interestingly, the two models produce similar results. While 80% of top ten drivers are consistent in both models, 9 out of ten least important predictors are shown the same. More surprisingly, X5 and X7 are shown as key drivers in both bagging and boosting, indicating a conflicting result with the one from logistic regression. Given the above contradictory results and the lack of prior information, all explanatory variables shall be included in both Logistic Regression and Generalized Additive Model such that no information will be excluded intentionally and the same basis for comparison will be provided to both models.

MODELING

A vanilla version of Logistic Regression is fitted using LOGISTIC procedure and part of the result is given in **Table 3.1**, in which the independent variables significant at 10% level are flagged by start(*).

```
proc logistic data = train desc;
  class X10 - X24;
  model Y = X2 - X24 / rsquare lackfit;
  score data = train out = predict_train;
  score data = test  out = predict_test;
run;
```

Table 3.1: Partial Output of Logistic Regression

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	349.0102	61	<.0001
Score	364.0240	61	<.0001
Wald	296.2755	61	<.0001

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq	
X2	1	3.3899	0.0656	*
X3	1	13.5527	0.0002	*
X4	1	0.2230	0.6367	
X5	1	0.4617	0.4968	
X6	1	2.0582	0.1514	
X7	1	0.6616	0.4160	
X8	1	16.7184	<.0001	*
X9	1	6.6856	0.0097	*
X10	1	0.6563	0.4179	
X11	1	37.1469	<.0001	*
X12	1	3.2497	0.0714	*
X13	1	3.9366	0.0472	*
X14	1	2.1064	0.1467	
X15	5	13.8757	0.0164	*
X16	5	7.5567	0.1824	
X17	5	79.4644	<.0001	*
X18	6	18.1702	0.0058	*
X19	9	17.0507	0.0479	*
X20	3	1.6409	0.6502	
X21	2	3.6510	0.1611	
X22	10	16.6599	0.0822	*
X23	2	2.2715	0.3212	
X24	1	1.0629	0.3026	

Hosmer and Lemeshow Goodness-of-Fit Test

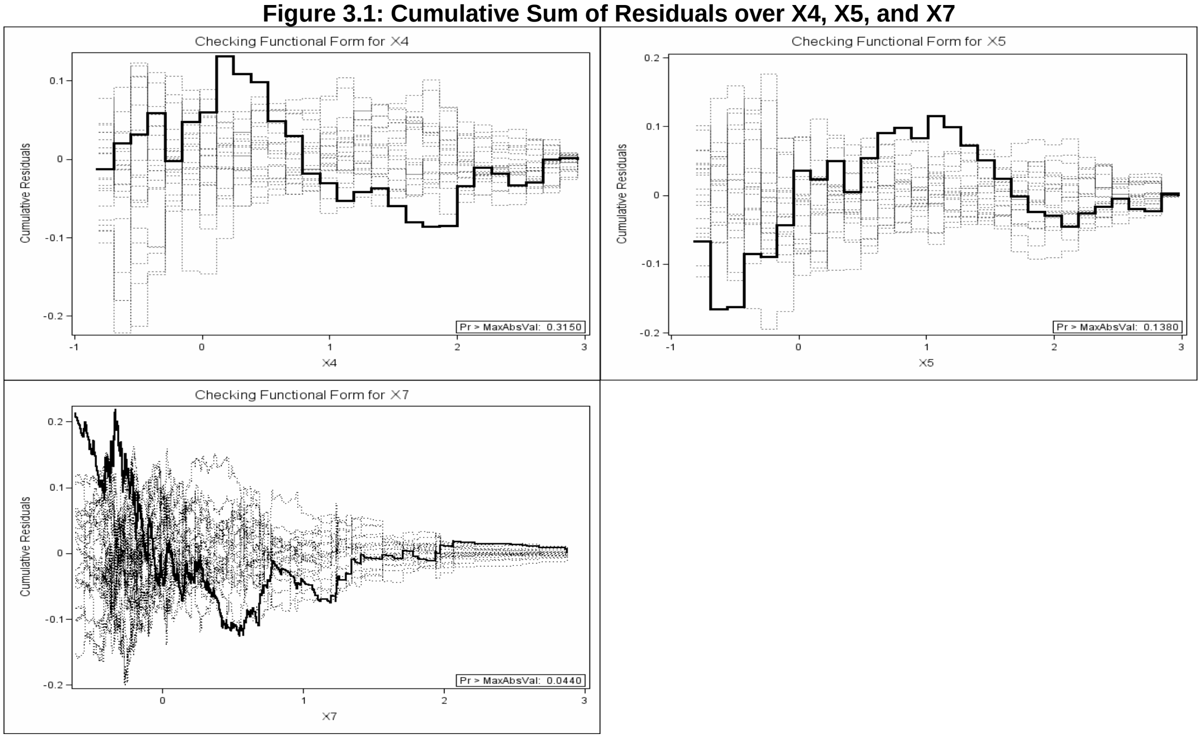
Chi-Square	DF	Pr > ChiSq
3.7731	8	0.8770

Type 3 analysis of the effects is in line with the previous result of bootstrapping. As expected, X4, X5, and X7 are the least significant numeric predictors. Results of Global Null Hypothesis and Goodness-of-fit test indicate that logistic regression gives a reasonable fit for the training data. However, whether the functional form of our model is correctly specified or not remains still questionable and needs to be addressed.

A new model-checking technique of Generalized Linear Models is introduced in GENMOD procedure based on cumulative sum of residuals (Lin, Wei, and Ying 2002). If the functional form of a linear predictor is correctly specified, cumulative sum of residuals over the predictor should be centered around zero and demonstrate a pattern without systematic trend.

```
ods graphics on;
proc genmod data = train desc;
  class X10 - X2;
  model Y = X2 - X24 / link = logit dist = binomial;
  assess var = (X4) / resample = 1000; /* to be replaced by X5 & X7 */
run;
ods graphics off;
```

Figure 3.1: Cumulative Sum of Residuals over X4, X5, and X7



Above SAS code checks the cumulative sum of residuals over X4, X5, and X7 respectively via 1000 simulation paths and the result is shown in **Figure 3.1**, where the black solid lines depict the cumulative residuals and light dashed lines show simulation paths following Gaussian process. The P-value of null hypothesis that functional form of the predictor is correctly specified is reported at the lower right-hand corner of each plot. As indicated by p-values, more appropriate functional forms might be needed for X5 and X7. While there is no strong evidence of misspecification for X4, the cumulative residuals plot does imply a potential improvement using quadratic term of X4.

The result of model assessment indicates that a better fit might be achieved by incorporating X4, X5, and X7 into more appropriate functional forms. The semi-parametric variation of Generalized Additive Models is used to estimate X4, X5, and X7 with univariate spline smoothers and to model the rest of predictors linearly.

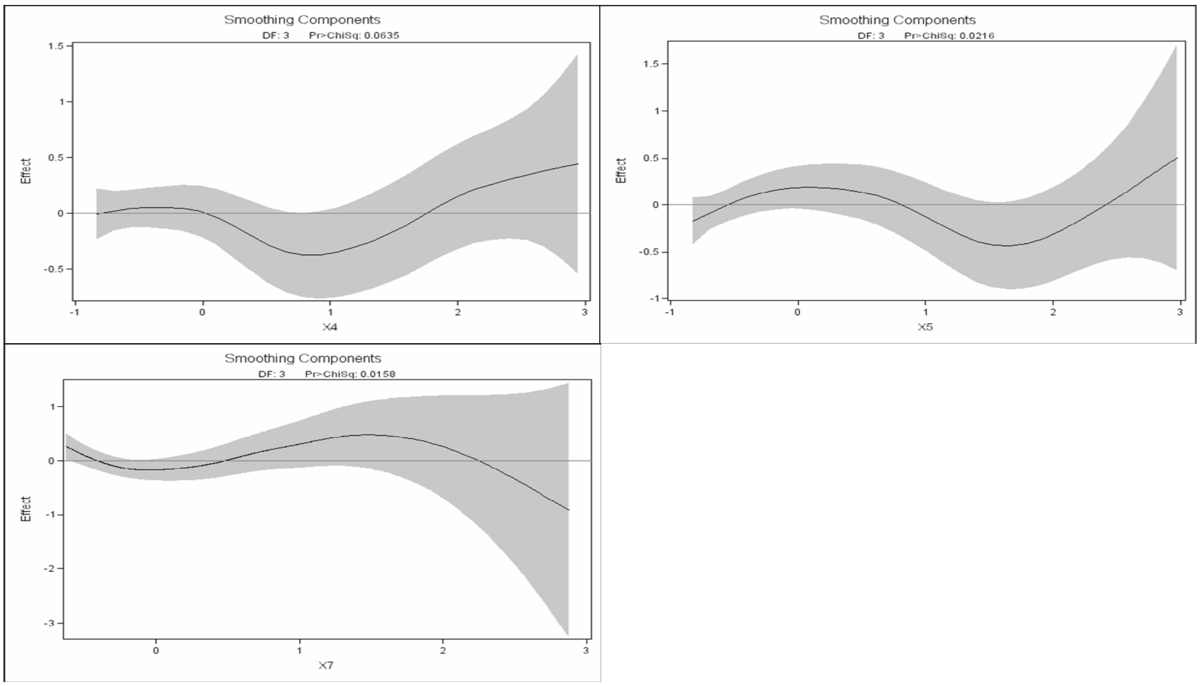
```
ods graphics on;
proc gam data = train plots(clm);
  class X10 - X24;
  model Y = param(X2 X3 X6 X8 - X24) spline(X4) spline(X5) spline(X7) / dist =
binomial;
  output out = predict_train p;
  score data = test out = predict_test;
run;
```

```
ods graphics off;
```

Table 3.2: Partial Output of Generalized Additive Model

Analysis of Deviance

Source	DF	Sum of Squares	Chi-Square	Pr > ChiSq
Spline(X4)	3.00000	7.280131	7.2801	0.0635
Spline(X5)	3.00000	9.670425	9.6704	0.0216
Spline(X7)	3.00000	10.351649	10.3516	0.0158



The GAM procedure, as given above, fits Generalized Additive Model with binary response by specifying the option DIST = BINOMIAL in the MODEL statement. The option PARAM(X2……) indicates that variables inside the parentheses are estimated parametrically in the linear form. The option P(PREDICTED) in OUTPUT statement gives predicted values for smoothing terms and the response, which are prefixed by 'P_' in the output data set. By default, each component is fitted with 4 degrees of freedom, 1 for linear term and 3 for smooth term. The Benefit of using conservative degrees of freedom in a Generalized Additive Model is twofold. It avoids over-fitting and lowers the computing cost. Alternatively, the smooth term's degrees of freedom can be determined automatically by Generalized Cross-Validation with the option METHOD = GCV in MODEL statement. However, the degrees of freedom selected by Generalized Cross-Validation often tend to over-fit the data based on our experience with GAM procedure. The model summary is given in **Table 3.2**.

As shown in Analysis of Deviance, smooth terms of X4, X5, and X7 are all significant at 10% level. The partial prediction plots for smooth terms attached in **Table 3.2** demonstrate the nonlinear pattern of the credit risk (Y) change across the range of the predictor. For instance, the plot of X5 shows that the relationship between X5 and Y is not linear and monotonic as assumed by the logistic regression. Instead, the credit risk goes up first as X5 increases, starts decreasing once X5 reaches 0, and then increases again after X5 becomes larger than 1.5. In addition, the increasing width of the confidence interval also indicates that the uncertainty of estimate increases as X5 increases, which is consistent with the density estimate of X5 shown previously. As shown in the partial prediction plot, there is a clear nonlinear pattern of X5 partial effect on Y in the high density region of X5.

Table 3.3: Summary Statistics Comparison Between GLM and GAM

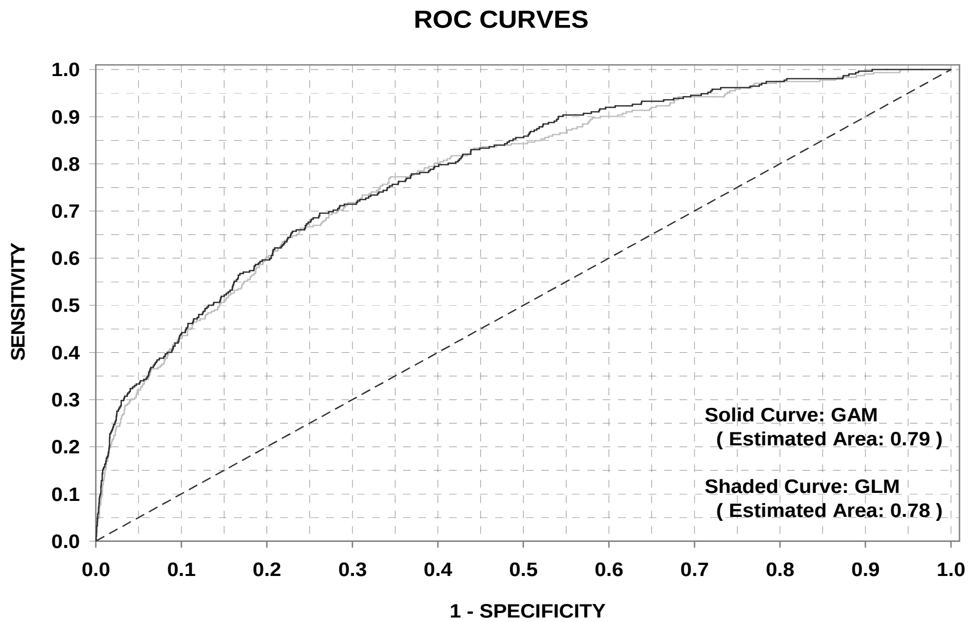
	Log-likelihood	Deviance	DF	R^2	AIC	BIC
GLM	-1004.49	2008.97	5118	18%	2132.97	**2539.23**
GAM	-990.83	1981.66	5109	**19%**	**2123.66**	2588.90

Table 3.3 shows a comparison of summary statistics between Logistic Regression and Generalized Additive Model, where **AIC = -2*Loglikelihood+p*2** and **BIC = -2*Loglikelihood+p*ln(N),** with p equal to the number of parameters

and N giving the number of cases. While higher pseudo R-square and lower AIC favor Generalized Additive Model, lower BIC favors more parsimonious Logistic Regression. However, since the model with fewer parameters tends to inflate Type-I error and it is well-known that it is more dangerous to misclassify a bad loan as a good one than to misclassify a good loan as a bad one in credit scoring, the Generalized Additive Model should be preferred to Logistic Regression.

Receiver Operating Characteristic curve, also known as ROC curve, is a graphical representation of the tradeoff between Type-I (Sensitivity) error and Type-II (Specificity) error for different possible cutoffs and is often used to compare predictive performance between different classification models. In a ROC curve, Sensitivity is placed on the Y-axis and 1 – Specificity on the X-axis. A model with perfect performance has an area under ROC curve equal to 1, whereas this area is 0.5 in a model with random guess. In practice, the area under the ROC curve is between 0.5 and 1. A comparison of ROC curves obtained by the Logistic Regression and the Generalized Additive Model is displayed in **Figure 3.2**. The area under ROC curve of Generalized Additive Model (0.79) is slightly larger than that of Logistic Regression (0.78), which suggests a better predictive performance of Generalized Additive Model. A closer look at curves of both models reveals the solid line above the shaded line in the major part of X-axis, indicating that Generalized Additive Model generally performs better than Logistic Regression at various risk levels.

Figure 3.2: Comparison of ROC Curves between GLM and GAM



While all the statistical evidence so far is in favor of Generalized Additive Model, it should be also noted that model performance can be highly sensitive to the data sample used to train the model. As pointed out in the introduction, a major criticism on Generalized Additive Model is over-fitting the data, which brings about the failure of prediction in data sets other than the training data. Therefore, to avoid sample-dependent result of model comparison, the predictive performance of Generalized Additive Model and Generalized Linear Model is validated on a separate testing set, as shown in Table.

Table 3.4: Misclassification Rates Comparison between GLM and GAM

Misclassification Rate			
Training Set		Testing Set	
GLM	GAM	GLM	GAM
8.42%	8.14%	9.80%	9.20%

A comparison of misclassification rates between Generalized Additive Model and Logistic Regression using both training data and test data is reported in the above table. As expected, misclassification rate of Generalized Additive Model is lower than that of Logistic Regression by 0.28% for training data. But more importantly, Generalized Additive Model out-performs Logistic Regression in misclassification rate by 0.6% even for testing data.

CONCLUSION

In the paper, we have discussed and compared Logistic Regression and Generalized Additive Model with their applications in credit scoring and implementation in SAS/STAT. It is shown in our study that Generalized Additive

Model outperforms Logistic Regression at least in two aspects. First, Generalized Additive Model relaxes the assumption of linearity between the predictors and the response and avoids the problem of model misspecification often happened in Generalized Linear Model. While Generalized Linear Model is the special case of Generalized Additive Model, linear effects can always be captured. Secondly, by incorporating nonlinear effects, Generalized Additive Model helps discover the hidden pattern of predictors and therefore improves predictive performance as a result. In the situation where nonlinearity is not intuitively justified to business managers, an alternative named Binning can be used to approximate the nonlinear effect based on the discovery from Generalized Additive Model. For instance, a nonlinear effect in Generalized Additive Model can be categorized with incorporation of the business sense and then Generalized Linear Model is implemented with this categorical variable. By doing so, interpretability can be dramatically improved with little price of predictive performance.

While we have shown its successful application in credit scoring, Generalized Additive Model can also be used in a broad range of domains such as database marketing and medical research. Balancing between flexibility and interpretability, Generalized Additive Model provides a promising alternative to Generalized Linear Model in most situation when the latter is applicable.

REFERENCES

Breiman, L. (1994), Bagging Predictors, Technical Report, Department of Statistics, University of California.
Dong, X. (2001), Fitting Generalized Additive Models with the GAM Procedure, SUGI Proceedings.
Franke, J., Hardle, W., and Stahl, G. (2000), Measuring Risk in Complex Statistical Systems, Springer Verlag.
Friedman, J. H. (1999), Greedy Function Approximation: A Gradient Boosting Machine, Technical Report, Department of Statistics, Stanford University
Hastie, T. and Tibshirani, R. (1990), Generalized Additive Models, Chapman and Hall.
Johnson, G. and So, Y. (2003), Let the Data Speak: New Regression Diagnostics Based on Cumulative Residuals, SUGI Proceedings.
Lin, D. Y., Wei, L. J., and Ying, Z. (2002), Model-Checking Techniques Based on Cumulative Residuals, Biometrics
McCullagh, P. and Nelder, J. (1989), Generalized Linear Models, Chapman and Hall.
Muller, M. (2000), Semi-Parametric Extensions to Generalized Linear Models, Habilitationsschrift.
Shao, N. (2004), Semi-Parametric Estimation for Credit Scoring, Research Projects, Department of Quantitative Analysis, University of Cincinnati

CONTACT INFORMATION

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

WenSui Liu
Cincinnati Children's Hospital Medical Center
Cincinnati, OH
E-mail: wensui.liu@cchmc.org
Web: http://statcompute.spaces.live.com

Jimmy Cela
Equifax Inc.
Alpharetta, GA
E-mail: Jimmy.Cela@equifax.com