Paper T06-2008

A Tutorial on Reduced Error Logistic Regression

Daniel M. Rice, Ph.D., Rice Analytics, St. Louis, MO

ABSTRACT

Reduced Error Logistic Regression (RELR) is a new form of regression that has significantly reduced error in its solutions. RELR is available as SAS macros that can also become an extension node to Enterprise Miner. This tutorial will give the SAS user an overview of RELR with comparisons to other predictive modeling methods.

INTRODUCTION

Reduced Error Logistic Regression (RELR) is a new form of regression that significantly reduces error related to multicollinearity, overfitting, unreliable regression coefficients, unreliable variable selections, small sample sizes, and response bias in unbalanced samples. While RELR does not completely remove error, it appears to perform significantly better than other predictive modeling methods in datasets where error is a significant problem.

RELR returns two types of solutions. These are “Full RELR” and “Parsed RELR” solutions. Full RELR solutions are variable selections that maximize the log likelihood across the training observations, but suffer from the disadvantage that they often contain a large number of variables. Parsed RELR results from a variable reduction method applied to the output of Full RELR models. In tests conducted to date, Parsed RELR has classification performance that is roughly equal to Full RELR. Both Full RELR and Parsed RELR return regression coefficients that are extremely reliable from sample to sample compared to other approaches. However, because of the small number of variables, Parsed RELR models are more interpretable as explanatory models.

RELR is implemented as SAS Macros that either are currently available or soon will be available as 1)as a stand-alone off-the-shelf program, 2)an extension node to Enterprise Miner, and 3)as a customized program. The stand-alone off-the-shelf program is limited in scalability, but the Enterprise Miner extension node and customized programs are scalable to problems with very large numbers of observations with very, very large numbers of independent variables. For example, RELR employs a variable importance screening process with a batch mode of processing that allows it to handle high dimensional variable sets and large numbers of observations significantly faster than standard logistic regression variable selection methods. RELR can be applied to all types of logistic regression problems including multinomial logit regression, ordinal logistic regression, and models with and without repeated measures or multi-levels.

RELR allows modeling of non-linear patterns in independent variables up to the 4^{th} order polynomial. RELR allows interactions between independent variables up to the three way interaction terms. All non-intercept independent variables are standardized in RELR to have a mean of zero and a standard deviation of 1. Hence, RELR imputes all missing values with a value of zero to reflect the mean of a variable. In addition, RELR models structural effects due to differences between missing and non-missing values in variables by creating new variables that dummy code each variable in terms of missing vs. non-missing values. RELR is able to handle such large numbers of variables because of its reduced error capabilities.

RELR solutions also minimize the classification response bias with unbalanced samples. This is possible because RELR uses a threshold approach to determine class membership. RELR employs an intercept correction approach to parameterize its thresholds. For this reason and because unbalanced samples often have a very small sample size of responses in the most important category for prediction, RELR often provides more useful results than other methods with unbalanced samples.

RELR is a completely automated method without any real arbitrary parameters. Because RELR is an automated process, this tutorial will not focus on how to run RELR as this should be self-evident to users. Instead, after a brief exposure to the statistical theory that underlies RELR, the focus will be on examples of the RELR results and a comparison of these results to other predictive modeling methods.

STATISTICAL THEORY

Reduced Error Logistic Regression (RELR) is related to Penalized Logistic Regression (PLR). In fact, RELR has been developed as a means to improve upon the problems inherent in PLR. Both RELR and PLR go beyond standard maximum likelihood logistic regression by introducing a new term into the objective function besides the standard log likelihood.

PLR maximizes the log-likelihood subject to a size penalty on the L_2 norm of all coefficients except intercepts. This use of a quadratic cost function to diminish overfitting and multicollinearity error is a commonly used method in maximum likelihood/entropy logistic regression (Le Cessie & Van Houwelingen, 1992; Khozla, Singh & Rice, 1995, Park & Hastie, 2006). For binomial logistic regression, this amounts to maximizing the following function:

$$(1)\quad LL(\beta_0,\boldsymbol{\beta},\lambda) = l(\beta_0,\boldsymbol{\beta}) - \lambda/2 \sum_{r=1}^{m} \beta_r$$

where $l(\beta_0,\boldsymbol{\beta})$ is the standard log likelihood that is a function of the intercept β_0 and the vector of logit coefficients $\boldsymbol{\beta}$ across *m* input variables and the right hand summation results from the quadratic cost function where λ is the penalty parameter (Park & Hastie, 2006). The effect of PLR is to generate a smoother set of logit coefficients in a solution. However, PLR does not always provide a large reduction in error with highly abundant error in completely independent validations. This is because λ must either be set to an arbitrary value (Khozla, Singh & Rice, 1995), or it must be estimated based upon validation sample data (Park & Hastie, 2006). In addition, the extent to which PLR diminishes error in the logit coefficients and in some cases might even increase error is an open question. Hence, there exist some significant questions and problems related to PLR.

In contrast to PLR, RELR directly reduces error rather than an arbitrary quadratic cost estimate and RELR does not employ an arbitrary scaling constant. RELR has been gradually proposed through a series of refinements presented over the past few years (Rice, 2005; Rice, 2006; Rice, 2007; Rice, 2008). RELR is related to the maximum entropy subject to linear constraints formulation that Soofi (1992) showed to be equivalent to the McFadden (1974) maximum likelihood logit formulation and Golan, Judge and Miller (1996) further developed to model error probabilities. However, Golan et al. (1996) did not attempt to model Extreme Value error, but instead modeled arbitrary error values not recognized as extreme values. The essence of RELR is that it attempts to reduce logistic regression error by modeling extreme error values that are known to underlie the logit error. This error is described through the Extreme Value Type I or Gumbel distribution. This famous Extreme Value Theory result was initially given in the work of mathematical psychologists (Luce and Suppes, 1965) and was confirmed in Nobel Prize winning work in economics (McFadden, 1974).

The linear constraints in the Soofi-Golan et al. maximum entropy formulation are cross product sums that reflect the relationship between the target variable and each of the input variables in standard logistic regression. For each of these r=1 to M non-intercept cross product sums, RELR assumes that the expected extreme value error is proportional to $1/t_r$, where t_r is the t value that arises in determining whether a correlation between a target variable and an independent variable is significantly different from zero. Hence, a large t_r effect is associated with a small magnitude expected extreme error value, whereas a small t_r effect is associated with a large magnitude expected extreme error value. Please note that this t value is not used to make inferences to a t distribution in RELR; it is simply used as a measure that is inversely proportional to the expected extreme error on the cross product sums. This formulation provides estimates of probability of response on each observation and target variable category comparable to standard logistic regression, but it also provides estimates of the probability of positive and negative error and their log odds or logit error for each independent variable. Most interestingly, when the logit error for an independent variable is zero, the algebra shows that this use of a t value gives a logit coefficient for an independent variable that is proportional to the corresponding t value for the correlation between that independent variable and the target variable. This is analogous to least squares regression with orthogonal independent variables and to the Naïve Bayes algorithm with or without orthogonal variables..

However, RELR's logit coefficients are only proportional to corresponding t values when the logit error is zero, but can deviate significantly when the logit error is non-zero. That is, unlike least squares regression with orthogonal variables and the Naïve Bayes algorithm, RELR's regression coefficients are not rigidly tied to a proportionality relationship with corresponding t values across all independent variables, so RELR can effectively model multivariate relationships. In fact, RELR's coefficients will deviate significantly depending upon interrelationships between input variables that include interaction variables. Yet, like the Naïve Bayes algorithm, the very simple linear relationship between the expected logit coefficient and a corresponding t value across all input variables represents an enormous savings in processing time. This is because the t value magnitudes are highly predictive of corresponding RELR regression coefficient magnitudes across input variables for those most difficult problems when many important variables are highly collinear. In the easy cases where there are few important variables with large magnitude t values, t values are not proportional to the regression coefficients in RELR. However, the discovery of the most important variables is not an issue in these easy cases with few large t values. Likewise, the discovery of the most important variables is not an issue in the easy cases involving orthogonal independent variables, where it is also linked to the magnitude of t values. In general, if a RELR regression coefficient magnitude for an input or interaction variable is relatively large, then the t value magnitude for that variable also appears to be relatively large even when the proportionality between t values and regression coefficients breaks down (Rice, 2008). For this reason, employing t values to order importance in variable selection is an extremely effective method to capture the most important variables. Moreover, the computation of t values can be run easily in smaller batches in serial or parallel. Thus, this t value screening to get the small list of candidates for most important variables effectively overcomes the “curse of dimensionality”.

RELR further employs symmetrical constraints that force the probability of positive and negative extreme value error to be equal across all cross product sums. When one maximizes the entropy subject to the cross product sums and symmetrical constraints, RELR gives a solution that can also be expressed in terms of a maximum likelihood solution; this is consistent with the well established equivalence between maximum entropy and maximum likelihood models (Soofi, 1992, Golan et al., 1996). Hence, RELR amounts to maximizing the following function:

$$(2) \quad LL(p,w) = \sum_{i=1}^{N} \sum_{j=1}^{C} y_{ij} \ln(p_{ij}) \quad + \quad \sum_{l=1}^{2} \sum_{r=1}^{M} \sum_{j=1}^{C} y_{jlr} \ln(w_{jlr})$$

where j=1 to C indexes the number of levels in the target variable, l=1 to 2 indexes the positive and negative extreme error conditions, i=1 to N indexes the observations in the training phase, and r=1 to M indexes the number of non-intercept independent variables. In the left hand summation involving **p**, y_{ij} =1 if the ith observation and jth target level is associated with a jth target level response and 0 otherwise. Note that the values of y_{jlr} that are multiplied by $\ln(w_{jlr})$ are set to equal 1 if the jth condition is the reference condition and 0 otherwise.

The left hand summation set in Equation (2) involving **p** is the standard log likelihood function in logit regression; the right hand summation involving **w** is introduced to model error probabilities in analogy with Golan et al. (1996). The essential difference between RELR and PLR revolves around this second set of summations. When the probability of positive error equals the probability of negative error across all M independent variables, the effect of error is minimal in a RELR solution. Thus, RELR will tend toward solutions that yield equal probability of positive and negative error across all M independent variables given everything else being equal, as the sum of $w(j,1,r)$ and $w(j,2,r)$ must equal 1 for all M independent variables.

RELR MODELS WITH COMPARISONS TO OTHER METHODS

The examples of RELR models and comparisons to other methods that are presented in the following sections should not be taken as anything like benchmarking studies for RELR. For example, these comparisons were conducted by Rice Analytics and therefore are not independent validation studies. In addition, the samples and datasets were selected in some cases to have very small sample sizes and very large numbers of multicollinear variables. These are cases where RELR would be expected to perform extremely well. Independent, benchmarking studies are still needed for the RELR method across a wide array of datasets; readers should not take the following comparisons to be such benchmarking studies.

PREDICTIVE MODELS USING 2004 PEW ELECTION SURVEY

Data were obtained from the Pew Research Center’s 2004 Election Weekend survey. The survey consisted of a representative sample of US households. Respondents were asked about their voting preference on the weekend before the 2004 election. 2358 respondents met the criteria of this example model which only employed those respondents who indicated that they would vote for Bush or Kerry without regard to whether Nader was in the race. In addition to voting patterns, this survey collected demographic and attitude data. These demographic and attitude responses were employed as input variables in the predictive models. Examples of attitude questions were whether respondents agreed with the Iraq war and whether respondents thought that the US was winning the war against terrorism.

In a first “smaller training sample” model, the training sample consisted of a segmented sample of 8% or 188 observations. The remainder of the overall sample defined the validation sample. In a second “larger training sample” model, the training sample consisted of a segmented sample of roughly 50% or 1180; the remainder of this sample defined the validation sample. This segmented sampling method is the default sampling method in the SAS Enterprise Miner 5.2 software that was also used to build the models.

The target variable was Presidential Election Choice (Bush vs. Kerry). Kerry was the target condition. The 2004 election was very close, as roughly 50% of the respondents opted for Kerry in all of these sub-samples. Hence, these are extremely balanced samples. Target condition response numbers are in Figure 1 and 2.

There were 11 interval variables and 44 nominal input variables originally, but the nominal variables were recoded into binary variables for input into RELR. RELR also produced new input variables from these original input variables corresponding to two-way interactions, polynomial terms, and missing data. Over 2500 variables resulted in total. Both Full RELR and Parsed RELR models were run. In addition, PLR models were run on the identical set of variables that Full RELR selected in both the smaller training sample and larger training sample models. Full RELR employs a backward selection process based upon the magnitude of the t values. Full RELR chooses as its best model the model corresponding to the variable set associated with the maximal log likelihood across the training observations. 176 variables were selected in the Full RELR best larger training sample model, whereas 24 variables were selected in the Full RELR best smaller training sample model. Once again, these identical variables were input into PLR to have an "apples-to-apples" comparison to Full RELR. For this same "apples-to-apples" reason, the same intercept estimation procedure was used with both RELR and PLR- including the use of thresholds for classification and corresponding intercept correction. The λ values in these PLR models were systematically varied between .5 and 150; values above 150 had floating point error. In both samples, the λ=1 was either associated with the best validation misclassification rate as in the small sample, or very close to the best, as a λ of 150 was slightly better in the larger sample. However, we report the results based upon λ=1 for both samples here because this is in the range that is normally employed for λ.

Bush vs. Kerry models were also run within Enterprise Miner 5.2 using the Support Vector Machine, Partial Least Squares, Decision Tree, Standard Logistic Regression and Neural Network methods. The default conditions were employed in all cases except Standard Logistic Regression where two way interactions and polynomial terms up to the 3rd degree were specified and Support Vector Machines where the polynomial kernel function was requested. Also, Stepwise Selection was performed with the Logistic Regression in the Pew "smaller sample", but no selection was performed with the larger sample due to the long time that it took to run the stepwise process in this sample. In addition, the Imputation Node within Enterprise Miner was employed with the Regression and Neural Network methods and was run using its default parameters. The identical variables and samples were employed as inputs in all cases. Misclassification Rate was employed as the measure of accuracy in all cases.

Like most political polling datasets, there was a wide range in the correlation between the original input variables that went from roughly -.6 to about 0.81. These correlation magnitudes were even larger for many of the interactions and nonlinear variables produced by RELR being over .9 in many cases, so this dataset clearly exhibited multicollinearity. In addition, there was significant correlation to the target variable in a number of variables. The largest correlations to the target variable were in the .7-.8 range. Hence, this dataset contained a number of fairly informative variables that were well correlated to the target variable.

Figure 1 compares the rate of misclassification error across these predictive modeling methods in the Pew "smaller training sample" dataset. The significance levels in that figure are from chi-square tests that compare these misclassification proportions using McNemar's test for dependent proportions in the validation sample. Figure 2 shows these same comparisons for the Pew "larger training sample" models. In the smaller training sample, these results suggest that Full and Parsed RELR had better validation sample classification accuracy than a set of commonly used predictive modeling methods that include Stepwise Logistic Regression, Support Vector Machines,

Figure 1: Comparison of "Smaller Training Sample" Models.

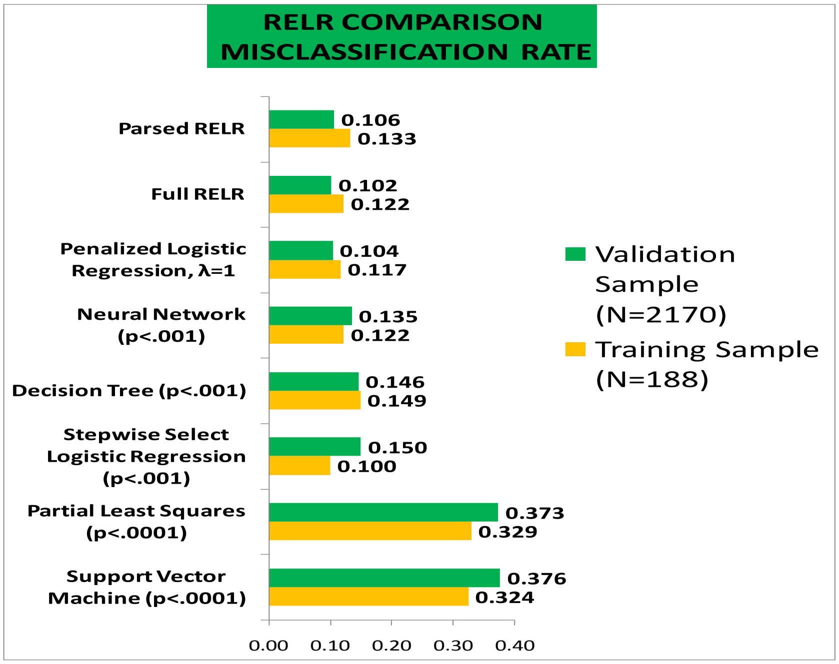
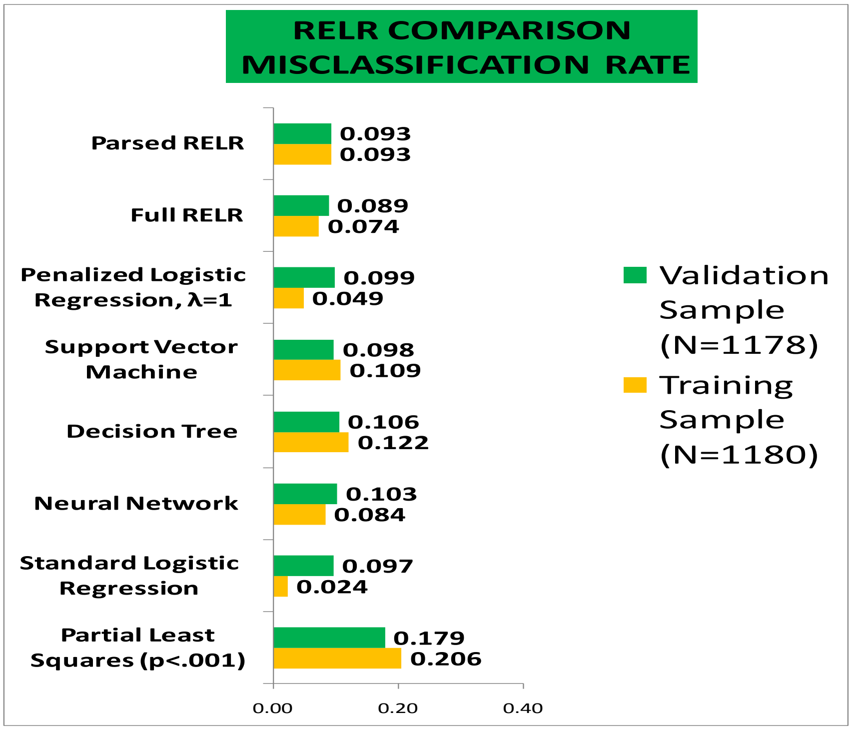


Figure 2: Comparison of "Larger Training Sample" Models.



Decision Trees, Neural Networks, and Partial Least Squares. However, there was no significant advantage to Full RELR or Parsed RELR in the larger sample, except in comparison to Partial Least Squares. Full and Parsed RELR's classification performance was comparable in both smaller and larger training samples. Also, Full and Parsed RELR performed comparably to PLR in both samples, in spite of the fact that PLR needed to use validation sample information to generate its λ. Taken as whole, these results suggest that RELR's advantageous classification performance effects may become most pronounced as more error is present as with a smaller sample size of training observations.

Overfitting, as measured by a significant difference (p<.05) in classification performance between the training and validation samples, was not significant with either Full or Parsed RELR in either sample – although there was a strong non-significant trend (p<.10) with Full RELR in the larger training sample. On the other hand, PLR was associated with significant overfitting in the larger training sample; this also would have been observed with a value of λ=150 in the larger training sample. Significant overfitting was also observed in a number of the other methods as is obvious in Figures 1 and 2.

We have replicated this basic pattern of results a number of times with different samples from this same dataset. In the process, we have learned that RELR gives solutions that are reliably similar from sample to sample, whereas PLR does not show comparable efficiency. This is shown in Table 1 and 2 below for the larger training sample model. Table 1 shows the 5 largest magnitude logit coefficients from the larger training sample Full RELR model

Table 1. Full RELR's Largest Magnitude Coefficients (r=.976 between Training and Validation)

PARAMETER	ESTIMATE	STDERR	PROB
IRAQWASWRONG	0.364	0.047	<.0001
USLOSINGWARONTERROR	0.266	0.035	<.0001
NORISKINCHANGEDURINGWAR	0.263	0.034	<.0001
REPUBLICAN	-0.247	0.032	<.0001
BUSHWILLLOSEELECTION	0.244	0.032	<.0001

and r=.976 represents the Pearson correlation between training and validation samples across all logit coefficients of all magnitudes in this full model. Table 2 shows the 5 largest magnitude logit coefficients from the larger training sample PLR model and r=.238 represents the Pearson correlation between training and validation samples across all logit coefficients of all magnitudes. These are standardized variables, so an interaction between dummy coded variables has potential meaning and interpretation that relates to the relative frequency and importance of the various sub-groups even with nonlinear variables. However, notice that Full RELR's 5 highest magnitude logit coefficients are from simple variables that do not involve interactions and do not involve any nonlinear terms. Besides this simplicity of structure, the RELR model seems to have greater face validity than the PLR model. How voters rate variables like IRAQWASWRONG, USLOSINGWARONTERROR, NORISKINCHANGE-DURINGWAR, and REPUBLICAN are generally thought to be the most important predictors of presidential election voting patterns since 9/11 and the Iraq war. On the other hand, some of the variables associated with higher magnitude logit coefficients in the PLR model are not reported as important predictors of recent presidential election voting patterns. Indeed, the RELR model would be expected to be a reliable explanatory model given the very tight confidence band around its higher magnitude logit coefficients. On the other hand, some of these highest magnitude logit coefficients in the PLR model are not signicantly different from zero. This lower reliability of the PLR logit coefficients is reflected in the r=.238 between training and validation coefficients across all variables, whereas this same value is r=.976 for Full RELR. Taken as a whole, this comparison of Table 1 and Table 2 suggests that the RELR model has greater simplicity, face validity, reliability, and explanatory power than the PLR model.

Table 2. PLR's Largest Magnitude Coefficients, λ=1 (r=.238 between Training and Validation)

PARAMETER	ESTIMATE	STDERR	PROB
REPUBLICANxREGISTERED^3	-7.405	4.442	0.096
DEMOCRATxPARENT^3	-6.132	2.868	0.033
REPLUBICANxBORNAGAIN^2	3.823	2.011	0.057
REGISTEREDxIRAQWASRIGHT^3	3.632	1.930	0.060
NOTPARTYLINEVOTERxDEMOCRAT	-3.211	4.016	0.424

Table 3 below shows the Parsed RELR model corresponding to the same larger training sample N=1180 as the Table 1 model. All variables selected in Parsed RELR are reported, but a Wald Chi-Square probability is not reported on the intercept since it is not recovered during maximum likelihood estimation, but is instead recovered from intercept correction using the threshold. Table 4 shows the Parsed RELR model from the completely independent validation sample of N=1178 respondents. The identical very small set of 9 variables was selected by Parsed RELR in each case. In addition, the correlation between the non-intercept logit coefficients in these Parsed RELR Training and Validation models was r=.9925 and the intercepts were identical. This indicates a very high degree of reliability in both selected variables and in their logit coefficients in this Pew sample (N=1180).

The Parsed RELR procedure employs a variable reduction process designed to give maximal reliability. Parsed RELR does not necessarily return the same variables as the largest magnitude Full RELR variables, but there was complete overlap with the 5 largest magnitude Full RELR variables in this case. These Parsed RELR models cannot be an artifact of the sample selected, as these are completely independent samples. Because of this simplicity, reliability, and validity, Parsed RELR would appear to show promise as a tool to generate explanatory models.

Table 3. Parsed RELR Model – Same Training Sample as Table 1.

PARAMETER	ESTIMATE	STDERR	PROB
INTERCEPT	-0.040	0.000	
DEMOCRAT	1.081	0.116	<.0001
NOTPARTYLINEVOTER	-0.984	0.109	<.0001
NOTPARTYLINEVOTERxDEMOCRAT	-0.964	0.107	<.0001
IRAQWASWRONG	1.709	0.183	<.0001
USLOSINGWARONTERROR	1.264	0.144	<.0001
HIGHRISKINCHANGEDURINGWAR	-1.051	0.115	<.0001
NORISKINCHANGEDURINGWAR	1.064	0.114	<.0001
BUSHWILLLOSEELECTION	1.197	0.130	<.0001
REPUBLICAN	-1.244	0.131	<.0001

Table 4. Parsed RELR Model – Independent Validation Sample

PARAMETER	ESTIMATE	STDERR	PROB
INTERCEPT	-0.040	0.000	
DEMOCRAT	0.937	0.093	<.0001
NOTPARTYLINEVOTER	-0.937	0.102	<.0001
NOTPARTYLINEVOTERxDEMOCRAT	-0.908	0.100	<.0001
IRAQWASWRONG	1.255	0.121	<.0001
USLOSINGWARONTERROR	0.759	0.087	<.0001
HIGHRISKINCHANGEDURINGWAR	-0.748	0.083	<.0001
NORISKINCHANGEDURINGWAR	0.856	0.091	<.0001
BUSHWILLLOSEELECTION	0.788	0.084	<.0001
REPUBLICAN	-0.951	0.094	<.0001

PEW BLACK VOTERS: A PREDICTIVE MODEL WITH AN UNBALANCED TARGET VARIABLE

RELR also appears to show some promise in difficult problems involving unbalanced target variables. In this sample model, data were taken from the identical Pew survey used in the Bush vs. Kerry models. However, Black vs. Non-Black was now taken as the target variable. This target variable was relatively unbalanced with Blacks making up approximately 10% of this sample. All other variables were employed as independent variables just as in the previous models. Also, all other RELR methods related to imputation, interactions, and model build exactly paralleled the previous RELR models. For comparison, only Support Vector Machine was used. The sampling consisted of a segmented sample of 472 observations in the Training Sample and 1886 observations in the Validation Sample. The target condition was non-blacks, but there were only 46 blacks in the training sample.

The classification results are presented in Table 5. While SVM had a lower overall misclassification rate, it got to that by classifying all records as non-black. In fact, the score that SVM produced was very close to .55 for every single record in the validation sample. On the other hand, RELR showed a wide range of probability scores and was able to show very little bias in its classifications, as indicated by the relative distributions of hits, misses, correct rejections (crs), and false alarms (fas). The Parsed RELR and Full RELR models gave similar reliable predictions from sample to sample, but this sample size was too small for reliable variable importance ordering and coefficient weights in the Parsed RELR output.

TABLE 5 – Validation Sample – Pew Black vs. Non-Black Voters

Validation	misclassification rate	hits/(misses+hits)	crs/(crs+fas)
Full RELR	0.250	0.746	0.786
Parsed RELR	0.242	0.762	0.714
SVM	0.096	1.000	0.000

CONCLUSION

The available data suggest that RELR overcomes significant problems in regression related to error. The most striking feature of RELR is this ability to generate simple, reliable, and valid models that cannot be an artifact of the sample selected, as shown with the Parsed RELR 2004 presidential election weekend model. These results need to be expanded with a wider array of datasets. SAS users will undoubtedly play a critical role in this testing of RELR across a wide array of data.

REFERENCES

Golan, A., Judge, G., and Miller, D. (1996), A maximum entropy approach to recovering information from multinomial response data. *Journal of the American Statistical Association*, 91: 841-853.

Lecessie, S. Van Houwelingen, JC, Ridge estimators in logistic regression, Appl-Stat-J Roy STC 41 (1): 191-201 1992.

Luce, R.D. and Suppes, P. (1965). Preference, utility and subjective probability, in R.D. Luce, R.R. Bush and E. Galanter (eds), Handbook of Mathematical Psychology, Vol. 3, John Wiley and Sons, New York, NY, pp. 249-410.

Khosla, D., Singh, M., and Rice, D. (1995) Three dimensional EEG source imaging via the maximum entropy method. IEEE Nuclear Science Symposium and Medical Imaging Record. Vol. 3, pp. 1515-1519.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka (ed) Frontiers in Econometrics, New York, Academic Press, pp. 105-142.

McFadden, D. (2002). Logit. Online book chapter: http://elsa.berkeley.edu/choice2/ch3.pdf.
Peduzzi, P., J. Concato, E. Kemper, T. R. Holford, and A. Feinstein (1996). A simulation of the number of events per variable in logistic regression analysis. Journal of Clinical Epidemiology 99: 1373-1379.
Park, M.Y. and Hastie, T. (2008). Penalized logistic regression for detecting gene interactions. Online manuscript: http://citeseer.comp.nus.edu.sg/cache/papers/cs2/525/http:zSzzSzwww-stat.stanford.eduzSz~hastiezSzPaperszSzplr_interactions.pdf/penalized-logistic-regression-for.pdf
Rice, D.M. (2005). A new distribution-free hierarchical Bayes approach to Logit Regression. In Proceedings of the 70th Annual Psychometric Society Meeting in Tilburg, Netherlands.
Rice, D.M. (2006). Logit regression with a very small sample size and a very large number of correlated input variables. In Proceedings of the 71st Annual Psychometric Society Meeting in Montreal, Quebec.
Rice, D.M. (2007). An overview of reduced error logistic regression. Invited presentation at the SAS M2007 Conference at Caesar's Palace in Las Vegas, Nevada.
Rice, D.M. (2008). Generalized reduced error logistic regression machine (in press). Proceedings of the 2008 Joint Statistical Meetings: Statistical Computing Section.

ACKNOWLEDGMENTS

We thank the SAS Institute for their continued partnership with Rice Analytics.

Portions of this work have been presented at the M2007 Conference and JSM 2008 Conference; We thank the Pew Research Center for making their data available for this type of research.

PATENTS PENDING

Generlaized Reduced Error Logistic Regression Method is currently pending patent.

CONTACT INFORMATION

Daniel M. Rice Ph.D.
Rice Analytics
10805 Sunset Office Drive, Suite 300
St. Louis, MO 63127
314-968-8175
DRice@RiceAnalytics.com
www.RiceAnalytics.com