An Adaptive Learning Model Which Accommodates Asymmetric Error Costs and Choice-Based Samples

James v. Hansen
James B. McDonald

See next page for additional authors

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub
Part of the Accounting Commons, and the Management Information Systems Commons

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/facpub/3234

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
AN ADAPTIVE LEARNING MODEL WHICH ACCOMMODATES ASYMMETRIC ERROR COSTS AND CHOICE-BASED SAMPLES

James V. Hansen
Marriott School of Management
Brigham Young University

James B. McDonald
Department of Economics and
Marriott School of Management
Brigham Young University

Rayman D. Meservy
Marriott School of Management
Brigham Young University

October 1995
Abstract

This paper introduces an adaptive-learning model, EGB2, which optimizes over a parameter space to fit data to a family of models based on maximum-likelihood criteria. We also show how EGB2 can be modified to handle asymmetric costs of Type I and Type II errors, thereby minimizing misclassification costs.

It has been shown that standard methods of computing maximum-likelihood estimators of qualitative-response models are generally inconsistent when applied to sample data with different proportions than found in the universe from which the sample is drawn. We investigate how a choice estimator, based on weighting each observation's contribution to the log-likelihood function, can contribute to estimator consistency and how this feature can be implemented in EGB2.
1. INTRODUCTION

This paper offers three contributions to the study of optimization in machine learning. First, it introduces an adaptive learning classifier, EGB2, which optimizes over a parameter space to fit data to a family of models based on maximum-likelihood criteria. While not completely model free, as with a neural network, EGB2 offers rich capabilities. Its adaptive-parameter algorithm is flexible, yielding predicted probabilities and a functional form from which the relative importance of problem attributes can easily be interpreted. These capabilities facilitate analysis, prediction, and theory construction.

A second contribution derives from the fact that many classification problems do not exhibit symmetric costs of misclassification. For example, it is important for public accounting firms to be able to identify fraud when it exists [Bell, et al., 1993]. A number of factors may help to predict the presence of fraud. A desirable model for decision makers would predict fraud when it is present, and predict no fraud when it is absent. The cost of predicting fraud when it is not present may result in unnecessary extra resources being expended.
conducting additional audit tests. The cost of failing to predict fraud when it *is* present is considerably greater, however, since that failure often leads to lawsuits and large-scale damage awards. We show how such asymmetric costs can be incorporated into our adaptive-learning model.

Third, standard methods of computing maximum-likelihood estimators of qualitative-response models have been shown to be generally inconsistent when applied to supervised learning data that are not representative of the population [Manski and Lerman, 1977]. For example, a recent study by Stice [1994] used a training set of 98 cases of audit litigation that was split 50-50 between cases that resulted in litigation and cases that did not result in litigation. The universe from which this sample was taken is proportioned approximately 98 percent nonlitigation and two percent litigation [Stice, 1994]. Such sampling proportion disparities are not uncommon [Manski and Lerman, 1977; Hansen, et al., 1992]. We show how a modified estimator, based on weighting each observation’s contribution to the log-likelihood function (thereby guaranteeing estimator consistency) can be implemented in EGB2.

Our methods apply to supervised learning as described by Shavlik and Dietterich [1990], where the learning program is given observations of the form \((X_i, Y_i)\), and the program attempts to learn or identify a function \(f\), such that \(f(X_i) = Y_i\) for all
i. The learning should be sufficiently robust that \( f \) can be applied to predict \( Y \) values for new and previously unseen, values of \( X \).

Our paper proceeds as follows: Section 2 outlines the structure of EGB2. Section 3 addresses the issue of asymmetric error costs. Section 4 outlines the method of estimators for choice-based samples. Section 5 discusses some important computational methods that apply to EGB2. Section 6 provides an example application, and Section 7 offers a summary and concluding remarks.

2. AN ADAPTIVE LEARNING MODEL--EGB2

The EGB2 qualitative-response model is typically used to predict the probability that an object with a certain set of characteristics (\( X \)) will be a member of a particular class of interest. For example, such models have been used to predict the probability that an individual will default on a loan or that a corporation will declare bankruptcy [cf., Bar Niv and McDonald, 1992].

We first outline the common structure of qualitative-response models which include the probit and logit models as special cases. Generalizations of the probit and logit models allow for the possibility of increased predictive capability.

The general form for qualitative-response models is
\[ Pr(Y=1|X) = F(X'\beta) = \int f(z|\theta)dz \]  

where \( F \) and \( f \) are the cumulative distribution and probability density functions, respectively; \( \theta \) represents possible distributional parameters, \( Y \) represents the binary dependent variable being predicted, \( X \) denotes a \( k \times 1 \) vector of exogenous variables useful in predicting \( Y \), and \( \beta \) is a \( k \times 1 \) vector of unknown parameters that generate scores \((Z=X'\beta)\). For example, \( Y = 1 \) could correspond to an entity (with economic and demographic characteristics denoted by \( X \)) that defaults on a loan, and \( Y = 0 \) otherwise. The empirical problem becomes that of, given observations on \( Y \) and \( X \) and a selected density, \( f(z|\theta) \), to estimate the vectors \( \beta \) and \( \theta \) in order to obtain predicted probabilities of loan default given by (1). The form of the model defined in (1) clearly yields positive predicted probabilities that are less than one.

The two most common qualitative response models are the probit and logit models, which correspond to selecting \( f(z|\theta) \) to be the logistic and standard normal density functions, respectively. We specify the density in (1) to be the
exponential generalized beta of the second kind (EGB2) defined by

$$EGB2(z; a, b, p, q) = e^{z p} / \left( b^p B(p, q) \left( 1 + \left( e^{z^2 / b^2} \right)^{p+q} \right) \right),$$

where

$$B(p, q) = \int_0^1 t^{p-1} (1-t)^{q-1} dt = \Gamma(p) \Gamma(q) / \Gamma(p+q).$$

Note that $a, b, p, q$ denote positive parameters, and both the logit and probit models are included as special or limiting cases. The logit model corresponds to (1) and (2) with $a=b=p=q=1$ (i.e., $\text{Logistics}(z) = EGB2(-z; a=1, b=1, p=1, q=1) = e^{-z^2} / (1 + e^{-z})^2$).

The probit model results from the limiting case of (2) when $f(z; \theta)$ is selected to be the standard normal

$$N(z; 0, 1) = \text{limit}_{a \to 0, \ q \to \infty} \left[ EGB2(z; a, b = (a^2 q)^{1/a}, \ p=1/a^2, \ q) \right]$$

$$= \frac{e^{-z^2/2}}{\sqrt{2\pi}}.$$

This result follows from the corresponding limit of a GB2 being equal to a standard lognormal. The indicated limit of a GB2 as $q$ grows indefinitely large is a generalized gamma. The limit of the corresponding generalized gamma as the parameter approaches zero is a lognormal (Kalbfleisch and Prentice [1980] and McDonald and Xu, 1995).
The probit and logit models are very similar; however, the
logit model has thicker tails than the probit model. The
cumulative distribution function (cdf) for the logit model has a
closed form; however, the cdf for the probit model does not have
a closed form and must be evaluated numerically.

The EGB2 allows for, but does not impose, symmetry in
applications (unless \( p = q \)). The importance of the additional
flexibility associated with the possible asymmetric densities can
be tested within the EGB2 family. The EGB2 (generalized
logistic) family was selected because it includes the logit and
probit models as special cases and allows for departures from
these popular models, including the possibility of asymmetry.

Two special cases of (2) that permit the distribution of \( Z \)
to be asymmetric are based on the Burr3 and Burr12 distributions,
labeled here EBurr3 and Eburr12:

\[
\begin{align*}
\text{EBurr3}(z; a, b, p) &= \text{EGB2}(z; a, b, p, q=1), \\
\text{EBurr12}(z; a, b, q) &= \text{EGB2}(z; a, b, p=1, q).
\end{align*}
\]

These have closed forms for the cumulative distributions, which
facilitate estimation:

\[
\begin{align*}
F_{\text{EBurr3}}(z; a, b, p) &= ((e^z/b)^a/(1+(e^z/b)^a))^p, \\
F_{\text{EBurr12}}(z; a, b, q) &= 1-1/(1+(e^z/b)^a)^q,
\end{align*}
\]
respectively.

Note that qualitative-response models based on the EGB2,
EBurr3, or EBurr12 distributions involve unknown distributional
parameters $\Theta = (a, b, p, q)$; whereas, the probit and logit models do not. The unknown parameters, distributional ($\Theta$) and scoring parameters ($\beta$), can be simultaneously estimated using maximum likelihood procedures; that is,

$$\max \sum_i [Y_i \ln F(x_i' \beta; a, b, p, q) + (1 - Y_i) \ln (1 - F(x_i' \beta; a, b, p, q))]$$

over the parameters $\beta$ and the relevant distributional parameters $\Theta$. Except for certain limiting cases, the distributional parameters $a$ and $b$ can, without loss of generality, be assumed to be unity. If either the probit or logit model is the correct specification, the EGB2 estimators would not be efficient since they involve estimating two additional parameters. We have not investigated the magnitude of this loss of efficiency for qualitative-response models. However, for regression models, some Monte Carlo simulations suggest that there is little efficiency loss in estimating the two extra distributional parameters for samples as small as fifty. Furthermore, the researcher can test for statistically significant improvements in the log-likelihood values. [McDonald and White, 1993]

Given parameter estimates for $\beta$, $a$, $b$, $p$, and $q$, the predicted probabilities

$$p(Y_i = 1|X_i) = F(X_i' \hat{\beta}; \hat{a}, \hat{b}, \hat{p}, \hat{q})$$

(6)
can be used in conjunction with a decision rule to classify individual cases. Clarke and McDonald [1992] used the EBurr3 and EBurr12 to predict consumer default on credit cards. In addition, Bar Niv and McDonald [1992] used the EGB2 and special cases to predict corporate bankruptcy.

We emphasize that EGB2 offers the flexibility of not having to specify a particular functional form, easily incorporating consistent choice estimators over a wide variety of non-normal or non-logistic distributional types, as well as asymmetric costs of type I and type II errors. We note that EGB2 is not always guaranteed to exhibit superior predictive performance over its special cases. The fundamental reason is that EGB2 and its family of models are estimated by maximizing the log-likelihood function, which does not imply that the accuracy of the prediction on either the training sets or holdout sets is necessarily maximized. An alternative estimation procedure would be to select estimators of unknown parameters to maximize predictive performance. The corresponding estimators might be thought of as being "extremum" estimators, [Amemiya, 1985].

3. ASYMMETRIC MISCLASSIFICATION COSTS

It is often the case that in practical applications the costs of Type I and Type II errors are different [cf. Stice [1991], Bell, et al. [1993]. Qualitative response models such as
the EGB2, probit, and logit models can easily incorporate the consideration of asymmetric cost assumptions as described in the remainder of this section.

Consider the payoff matrix illustrated in Figure 1. Here \( \pi_{ij} \) is the return (positive or negative) of predicting outcome \( i \) for actual outcome \( j \), where \( i = 1 \) denotes the presence of the concept of interest and \( i = 0 \) denotes absence of that concept. \( \pi_{00} \) is the return from correctly predicting the absence of the concept, \( \pi_{01} \) is the return associated with a Type II error, \( \pi_{10} \) is the return associated with a Type I error, and \( \pi_{11} \) is the return from correctly predicting presence of the concept.

Let \( p_{ij} \) denote the corresponding probabilities. Let \( z' \) be the decision threshold, such that for values of \( X'\beta \) larger than \( z' \) an observation will be classified as positive instances of the concept of interest. The conditional probabilities will depend on the distribution of scores from positive instances of the concept (P) and negative instances of the concept (N). The expected return as a function of \( z' \) is implicitly defined by the threshold that maximizes the expected return given by

\[
E(\text{return}) = \sum_{ij} p(j|i)p_i \pi_{ij}.
\]

(7)

This expression takes account of the prior probabilities of group size, conditional probabilities, and costs and benefits associated with correct and incorrect classification.

Using Leibniz's rule to maximize the expected return with
respect to $z'$. Clarke and McDonald [1992] demonstrate that
\[ f_N(z'|N)/f_P(z'|P) = p_1(\pi_{11} - \pi_{10})/p_0(\pi_{00} - \pi_{01}). \] (8)
implicitly defines the benefit-maximizing value of $z'$. If the prior probabilities of concept and nonconcept are equal, and if $\pi_{11} - \pi_{10}$ is equal to $\pi_{00} - \pi_{01}$, then $z'$ would correspond to the point where the distributions of scores for fraud and nonfraud clients have the same ordinate. The threshold $z'$ increases as either $p_1$ or $\pi_{11} - \pi_{10}$ decreases, or as $p_0$ or $\pi_{00} - \pi_{01}$ increases. That is, an increase in the cost of a misclassification results in adjustment of the "optimal" threshold to reduce the expected costs of this type of error.

These procedures are readily implemented with the EGB2 qualitative-response model to provide very flexible distributions of scores, as well as enabling investigation of the possibility of determining an optimal expected-benefit threshold. We note that asymmetric costs determine the decision threshold ($z'$); the estimated distributional parameters and weights ($\beta$) used in calculating EGB2 scores are unaffected. These methods are implementable in other qualitative-response models, as well.

4. ESTIMATION FROM CHOICE-BASED SAMPLES

Manski and Lerman [1977] show that standard methods of maximum likelihood computation are subject to error when the classification proportions are different in the training sample.
than in the problem domain universe. Their analysis applies particularly to the type of supervised training sets commonly used in machine learning.

Manski and Lerman [1977] demonstrate that when the proportions of positive and negative concept examples in the problem domain are different than that of the training set, the standard unweighted sampling maximum likelihood estimator is inconsistent. To deal with this problem, they develop a weighted sampling maximum likelihood estimator (WSMLE) and show that it is consistent. The fundamental WSMLE concept is this: define \( \omega_i = \frac{Q(Y_i)}{H(Y_i)} \) where \( Q(1) \) and \( Q(0) \) denote the fraction in the population corresponding to \( Y = 1 \) and \( 0 \) respectively. \( H(1) \) and \( H(0) \) denote the analogous fractions in the sample. Given the researcher's assumed knowledge of the population shares \( Q(\cdot) \), and given his ability to calculate the sample shares \( H(\cdot) \) directly from the data, the weights \( \omega_i \) are known non-negative constants. We note that \( Q(\cdot) \) is known for many problems of interest [cf. Hansen, et al., 1994; Manski and Lerman, 1977].

The weighted log-likelihood function can be written as

\[
\ell_w = \max \sum \omega_i [Y_i \ln F(X_i; \beta; a, b, p, q) + (1 - Y_i) \ln (1 - F(X_i; \beta; a, b, p, q))]
\]

Manski and Lerman [1977] prove that estimates obtained by maximizing (6) are strongly consistent and asymptotically normal. An additional term,
\[ \sum_{i=1}^{n} \omega_i \log g(X_i) \]

is included in the previous objective function, where \( g(X_i) \) denotes the pdf for the attributes. This can be neglected if we are considering a prediction problem as being conditional on the \( X \)'s.

5. COMPUTATIONAL METHODS

The application of the procedures outlined in this paper involves three steps: (1) estimating the coefficients in the score \((X\beta)\) and distributional parameters; (2) evaluating the scores and predicted probabilities of failure, and (3) the determination of the decision threshold.

The estimated coefficients of the variables in the score function are obtained by maximizing the, possibly weighted, log-likelihood function

\[ \ell_w = \max \sum_i \omega_i \left[ Y_i \ln F(X_i, \beta; a, b, p, q) + (1-Y_i) \ln (1-F(X_i, \beta; a, b, p, q)) \right] \tag{10} \]

over the \( \beta \) and the distributional parameters \( a, b, p, \text{and} q \) corresponding to the assumed cumulative distribution function. Closed form expressions for the cumulative distribution functions for the Logit and Burr types 3 and 12 are given in section 2. The cumulative distribution functions for the probit and EGB2
involve infinite series and are given by

\[ Pr_{\text{Probit}} (Y=1|X_1) = F_{\text{Normal}}(X_1\beta) \]

\[ = \frac{1 + \chi \beta}{\sqrt{2\pi}} e^{\frac{-(\chi \beta)^2}{2}} {}_1F_1[1; \frac{3}{2}; \frac{(\chi \beta)^2}{2}] \]  \hspace{1cm} (11) 

\[ Pr_{\text{EGB2}} (Y=1|X) = F_{\text{EGB2}} (X_1\beta) \]

\[ = \frac{z^p}{p B(p,q)} {}_2F_1[ p, 1-q, p+1 ; z] \]  \hspace{1cm} (12) 

where \( z = \frac{e^{\alpha \beta}}{b^a} \)

\[ = \frac{e^{\alpha \beta}}{1 + e^{\alpha \beta} \frac{b^a}{b^a}} \]

where \( {}_1F_1 \) and \( {}_2F_1 \) denote the confluent hypergeometric and hypergeometric series, both of which are infinite series. Maximum likelihood estimation involves iterative procedures to solve a nonlinear optimization problem. In each model (probit,
logit, EBurr, or EGB2) optimization was performed using the program GQOPT obtained from Princeton University. Various starting values were tried and the Simplex, David Fletcher Powell, Pattern, and Gradient optimization algorithms were used, with a convergence criterion of $10^{-8}$ to insure that an optimum was reached. Quandt [1983] discusses some issues related to computational methods and problems. The logit model converges most rapidly, followed by the Burrit models, then the probit, and finally the EGB2 models.

Given estimates of the $\beta$, $\beta$, and the distributional parameters, the estimated cumulative distribution yields predicted probabilities of failure. This function is evaluated for each data point

$$p(Y_i = 1|X_i) = F(X_i \beta; \hat{\alpha}, \hat{\beta}, \hat{p}, \hat{q})$$

in the hold-out sample to estimate the "likelihood" of failure.

The next step involves the determination of an "optimal" threshold ($Z'$) such that if $X\beta > Z'$ we predict failure. $Z'$ is determined by solving the equation

$$f_N(z'|N)/f_P(z'|P) = p_1(\pi_{11} - \pi_{10})/p_0(\pi_{00} - \pi_{01}).$$

This procedure is set up to maximize expected profits.

6. AN EXAMPLE

While the focus of this paper is not an empirical
study, we do wish to demonstrate an application of the choice-based weighted estimator methodology. In order to do so we either need a universe of instances, or a subset of that universe, which represents the same classification proportions found in the universe. With this set of instances in hand, we can choose a sample set with a different proportion in order to contrast results achieved with-and-without choice-based estimation methods. In order to provide an illustration, we developed a data set of some complexity, incorporating a mix of compensatory and conjunctive decision problems.

Selecting Decision Characteristics

Decision strategies are often represented in the form of compensatory or noncompensatory models. In the model representation, the decision is portrayed as a result of considering all available information in a single global judgement. Compensatory models are typical of those used in regression studies. Regression studies usually represent the decision process as a linear additive regression model of the form

\[ Y = a + b_1X_1 + b_2X_2 + \ldots + b_kX_k + e. \]

This represents a type of compensatory model, where the high score of one attribute can offset the low score of another. The degree of offset is determined by the relative weights \((b_i)\)
placed on the attributes \( (X_i) \). This process of trading off attributes has been found to be integral to many day-to-day decisions such as selecting a home or automobile [cf. Libby, 1981]. The strength of traditional statistical methods that use compensatory models is their robustness or effectiveness across a wide variety of problem conditions [Dawes, 1979].

However, most traditional statistical methods assume continuous tradeoffs among attributes, a distribution which frequently does not hold among real-world problems and often results in serious errors [cf. Johnson, et al. 1985].

In noncompensatory models, the high score of one attribute cannot compensate for the low score of another. One type of noncompensatory model often found in the literature is conjunctive [Kearns and Vazirani, 1995]. Conjunctive models involve multiple cutoffs, requiring that some minimal level of performance be achieved or exceeded by all variables

\[ X_i > X_e \text{ for all } i. \]

This results in a choice that is actually based on the level of the worst attribute. Many police academies use this model for minimum physical admittance standards. No matter how strong or able a person is, if he has a bad knee, that person will not be admitted. Previous research suggests that conjunctive models are used to prescreen alternatives [Libby, 1981]. Research also suggests that many decisions are made using a combination of
these models, such as conjunctive for prescreening and compensatory for final choices [Payne, 1976].

For this study we defined a universal set of instances having a decision form that mixed compensatory and conjunctive structures:

\[
\text{decision} = 1 \text{ if } (X_1 > t_1 \text{ and } (X_2 + X_3 + \ldots + X_n)/n > t_4) \\
0 \text{ otherwise.}
\]

The results are presented in the following section.

Sample Selection and EGB2 Results

The normal procedure for testing the predictive capability of a model generated from sample data typically commences with the collection of a set of sample instances. Rarely is the universe of instances available, although some characteristics of the universe may be known. The sample set is then (one time or repeatedly) partitioned into a training set (to train or devise a model) and a test set (for testing that model's predictive performance). The test set(s) acts as a surrogate for the population, in that it tests performance on data sets that were not seen by the algorithm that constructed the model.

For our application we generated a population of 400 from an assumed universe proportioned 87.5 percent classification '0' and 12.5 percent classification '1'. (For the reader who desires concreteness, '0' could be thought of as an audit engagement of a
public accounting firm that does not result in litigation, and 'l' could be thought of as an audit engagement that does result in litigation [Stice, 1991]."

Separately, we randomly selected a training set of 100 from the 87.5/12.5-universe that was proportioned 50 percent decision '0' and 50 percent decision 'l'. We earlier cited examples of experiments that have been characterized by this proportional anomaly. The EGB2 learning model was then applied to the training set in the conventional way without use of choice-based weighted estimators. The resulting models were tested on holdout sets and the outcomes recorded. This procedure was repeated for EGB2 incorporating choice-based weighted estimators. A summary of comparative results for three of the models generated by EGB2 are shown in Table 1. These results assumed that the costs of Type I and Type II errors were symmetric, meaning that the overall accuracy is the appropriate measure of performance. When applied to "unseen" part of the universe (400 cases), EGB2 with choice-estimator methods consistently yielded higher accuracy, as suggested by the theory of Manski and Lerman [1972]. The practical significance of the improved accuracy may vary with the problem domain. Evidence on the statistical significance of differences in accuracy awaits further study.

With respect to asymmetric costs of Type I and Type II errors, Weiss and Kulikowski [1991] affirm that when such error
costs can be agreed upon, those costs should be included in the learning model. When the costs of Type I (false negatives) and Type II (false positives) errors are not symmetric, the accuracy of one classification is more valuable than the other. Weiss and Kulikowski [1991] provide an example relating to a system of approval for credit card applicants, where the cost of a Type II error (approval of a credit card for a poor credit risk) is greater than the cost of a Type I error (disapproving a credit card to a good credit risk). Clarke and McDonald provide a similar application [1992]. Similar asymmetric error costs apply to our prior example. Suppose that "0" represents a client engagement that results in no litigation, and that "1" represents a client engagement that results in litigation. Stice [1994] showed that the costs of failing to correctly identify litigation were far greater than the additional costs of incorrectly classifying a non-litigant as a litigation client. A useful predictive model would sacrifice some accuracy in correctly predicting non-litigants in favor of increased accuracy in predicting litigants.

Letting \( z_\alpha \) be the threshold such that \( F(z_\alpha) = \alpha \), Table 2 shows the result of incorporating the asymmetric cost method of Section 3 in the EGB2 model, with \( \alpha = 0.1 \). The results show a corresponding drop in the percentage of correct classifications of non-litigants, but an increase in the correct classifications.
of actual litigants, reflecting the greater costs of failing to identify that classification. This is expected since $z_{0.1}$ represents a large differential in error costs. Larger values of $\alpha$ will result in less dramatic outcomes as $\alpha$ approaches 0.5 (equal error costs) from either side of the interval (0,1).

7. CONCLUDING REMARKS

We have presented an adaptive learning model, EGB2, which has several useful features. First, it uses a flexible parametric family of models to fit training data. Second, asymmetric error costs can easily be incorporated in EGB2. When error costs are symmetric, minimizing the overall error rate serves to minimize misclassification costs. But when costs are asymmetric, the objective of minimizing misclassification costs is not served by simply minimizing the overall error rate. We have shown how this problem can be handled.

Third, estimators which take account of choice-based samples are readily included in the EGB2 framework. The inclusion of choice-based weighted estimators is especially important when the classification proportions of the training set are different than the proportions of the universe from which the training set is drawn--at least when the resulting model is to be used to predict cases not used in the training set.

The methods we have shown for extending EGB2 to include
asymmetric costs should be applicable to other learning models, as well. Choice-based weighted estimator methods are applicable to learning models that use maximum-likelihood estimators [cf. Hassoun, 1995].
REFERENCES


McDonald, J. “Some Generalized Functions for the Size


Figure 1
Payoff Matrix

Predicted

\[
\begin{array}{c|c}
1 & 0 \\
\hline
1 & \Pi_{11} & \Pi_{01} \\
0 & \Pi_{10} & \Pi_{00} \\
\end{array}
\]
<table>
<thead>
<tr>
<th>Model</th>
<th>Choice-Based Weights</th>
<th>Equal Weights</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accuracy on Holdout Sample</td>
<td></td>
</tr>
<tr>
<td></td>
<td>92.5%</td>
<td>89.0%</td>
</tr>
<tr>
<td>EGB2</td>
<td>92.8%</td>
<td>89.8%</td>
</tr>
<tr>
<td>logit</td>
<td>92.8%</td>
<td>90.8%</td>
</tr>
<tr>
<td>probit</td>
<td>92.8%</td>
<td>90.8%</td>
</tr>
</tbody>
</table>

Table 1
Choice - Estimator Performance (Symmetric Costs)
Table 2
Comparative Results of Asymmetric Error Costs

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>EGB2</strong></td>
<td>1</td>
<td>84%</td>
</tr>
<tr>
<td>Actual</td>
<td>0</td>
<td>11%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>logit</strong></td>
<td>1</td>
<td>90%</td>
</tr>
<tr>
<td>Actual</td>
<td>0</td>
<td>12%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Predicted</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>probit</strong></td>
<td>1</td>
<td>88%</td>
</tr>
<tr>
<td>Actual</td>
<td>0</td>
<td>12%</td>
</tr>
</tbody>
</table>