A Generalized Qualitative-response Model and the Analysis of Management Fraud

J. V. Hansen • J. B. McDonald • W. F. Messier, Jr. • T. B. Bell
Marriott School of Management, Brigham Young University, Provo, Utah 84601
Fisher School of Accounting, University of Florida, Gainesville, Florida 32611
KPMG Peat Marwick, Montvale, New Jersey 07645

Management fraud has become a topic of increasing interest to the public accounting profession. Prior research indicates that management fraud is seldom experienced by audiGtors. As a result, it is doubtful that auditors have a well-developed cognitive model for making fraud risk assessments as part of the audit planning process. Early research studies attempted to identify factors that could be linked to the occurrence of management fraud, while more recent work has attempted to build models to predict the presence of management fraud. In this paper, we report on a study that uses a powerful generalized qualitative-response model, EGB2, to model and predict management fraud based on a set of data developed by an international public accounting firm. The EGB2 specification includes the probit and logit models and others as special cases. Moreover, EGB2 easily accommodates asymmetric costs of type I and type II errors. This is important for public accounting firms since failure to predict fraud when it is present (a type II error) is usually very costly to the firm in terms of litigation. The results demonstrate good predictive capability for both symmetric and asymmetric cost assumptions.
(*Auditing; Management Fraud; Decision-making; Generalized Qualitative-response Model*)

1. Introduction

Statement on Auditing Standards (SAS) No. 53 (AICPA 1993, AU 316) requires that auditors assess the risk that errors and irregularities may materially misstate a set of financial statements. Based on that assessment, the auditor designs the audit to provide reasonable assurance of detecting such errors and irregularities. SAS No. 53 and a related standard, SAS No. 47 (AICPA 1993, AU 312), provide a list of factors that should be considered by the auditor during this risk assessment process. However, the standards provide no evidence on the predictability of the individual factors or guidance on how to combine such factors into an overall judgment.

The problem of fraud risk assessment is compounded further by the fact that prior research (Loebbecke et al. 1989, Bell et al. 1993) indicates that auditors seldom experience management fraud. For example, Bell et al. (1993) report that 80 percent of the respondents to their survey indicated that they had encountered two or fewer instances of management fraud during their careers, which averaged 17 years of audit experience; 40 percent of the respondents had never worked on an engagement involving management fraud. Without significant experiences with management fraud, it is unlikely that auditors have sufficiently developed cognitive models for making fraud risk assessments as part of the audit planning process.

This absence of specific professional guidance and a lack of experience with management fraud has led practitioners and researchers to develop models or decision aids for predicting management fraud. Predicting management fraud is important because the presence of management fraud often leads to costly lawsuits against public accounting firms (Palmrose 1987). Thus, even moderate improvement in models that predict manage-

0025-1909/96/4207/1022$01.25

ment fraud may prove cost effective for public accounting firms.

Recently, Bell et al. (1993) developed and tested a model that assesses the likelihood of management fraud using bivariate and cascaded logistical analyses. Their model produced excellent results in predicting management fraud. The research reported in our paper has the same focus as Bell et al. (1993): to develop a model to predict management fraud. It extends their work in two important ways. First, a generalized qualitative-response model (EGB2) is applied to their data. The EGB2 specification is a very flexible functional form and includes the well-known probit and logit models as special cases. Within a family of possibilities, the EGB2 probability distribution is very robust, allowing parameter values to be determined from the data. Second, EGB2 (and other qualitative-response models) can readily incorporate asymmetric costs of type I and type II errors. This capability is particularly important because the costs of failing to detect management fraud (i.e., a type II error) are much higher than the costs of over-auditing an innocent client predicted to be fraudulent.[1]

The remainder of the paper is outlined as follows: In §2, prior research on management fraud is reviewed. Section 3 provides a discussion of the generalized qualitative-response model, EGB2. Section 4 describes the sample data and methodology. Section 5 reports the results of applying EGB2 to the management fraud problem assuming symmetric and asymmetric costs of misclassification. Section 6 offers a summary and concluding comments.

2. Prior Research on Management Fraud

SAS No. 53 distinguishes between errors and irregularities. The major difference between the two is the intention of the act. While errors are unintentional acts, irregularities are intentional. One specific type of irregularity is management fraud which is generally defined as fraudulent financial reporting undertaken to render financial statements misleading (AU 316.03). Fraudulent financial reporting has been the subject of examination over the years by various groups, such as the Treadway Commission (AICPA 1987).

Auditing standards point out that the auditor should design the audit to provide reasonable assurance of detecting material errors or irregularities. However, an audit conducted in accordance with generally accepted auditing standards may not detect management fraud because audit procedures that are good at detecting errors may not be as effective at detecting irregularities such as management fraud. Yet, certain conditions or circumstances may come to the auditor's attention during audit planning, or during the conduct of audit procedures, that may indicate the likelihood of fraudulent reporting. Prior research has examined the effectiveness of these factors (sometimes referred to as "red flags") to predict management fraud.[2] We do not provide a review of the literature on management fraud. Our focus is on the specific studies that have attempted to build models or decision aids for predicting management fraud.

Pincus (1989) conducted a field experiment that required auditors to predict the presence of management fraud with or without a decision aid (i.e., a questionnaire). In an apparent paradox, she found that subjects without the questionnaire outperformed those who were given the questionnaire. Additional analysis suggested that the questionnaire contained potentially useful information, but that auditors were unable to determine how to combine the results or what relative weights to assign to the various factors.

Loebbecke et al. (LEW) (1989) refined an earlier model developed by Loebbecke and Willingham (1988) using 77 fraud cases and tested the mappings of various red flags resulting in the classifications depicted in Figure 1. Primary factors (Panel A) were observed more frequently than secondary factors (Panel B) in the actual fraud cases. A binary rule based on the presence or absence of primary and secondary factors used by LEW predicted fraud with 88 percent accuracy.

Bell et al. (1993) extended LEW's work in two ways. First, they tested the LEW model using *both* fraud and

[1] When an auditor commits a type I error (i.e., predicting fraud when it is not present), the costs are the additional audit work to investigate the possible fraud that cannot be billed to the client and the intangible costs that may be incurred in the way of negative client relations.

[2] See Elliott and Willingham (1980) for a summary of the literature on management fraud.

Figure 1 Risk Factors in LEW Model Components

Panel A: Primary Factors

Conditions

Dominated decisions:
- Management decisions dominated by a single person, or a few persons who act in concert.

Major transactions:
- Company entered into one or an aggregate of material transactions.
- Company involved in purchase, sale, or merger of/with another company.
- Company recently entered into a significant number of acquisition transactions.

Related party:
- Company entered into a significant transaction or transactions with one or more related parties.

Weak internal control:
- Company has a weak internal control environment.*
- There are inadequacies in the company's accounting system.
- Accounting personnel exhibit inexperience or laxity in performing their duties.

Difficult-to-audit transactions:
- Company has a significant number of difficult-to-audit transactions.

Motivation

Industry decline:
- Client's industry is declining with many business failures.
- There are adverse conditions in the client's industry.

Inadequate profits:
- Profitability relative to the client's industry is inadequate or inconsistent.
- Company is having solvency problems.

Emphasis on earnings projections:
- Management places undue emphasis on meeting earnings projections.*

Significant contractual commitments:
- Company is subject to significant contractual commitments.

Attitude

Dishonest management:
- Officers of the company have entered into collusion with outsiders.
- There is a need to cover up an illegal act.
- Auditor's experience with management indicates a degree of dishonesty.

Emphasis on earnings projections:
- Management places undue emphasis on meeting earnings projections.*

Personality anomalies:
- There is undue concern with the need to maintain or improve the reputation/image of the entity.
- Management displays a propensity to take undue risks.
- Management personnel engage in an inappropriate lifestyle.
- Top management is considered to be highly unreasonable.
- Client management displays a significant lack of moral fiber.
- Client personnel exhibit strong personality anomalies.

Prior year irregularities:
- There have been instances of irregularities in prior years.

Lies or evasiveness:
- Management has lied to the auditor or has been overly evasive.

Aggressive attitude toward reporting:
- Management displays an overly aggressive attitude toward financial reporting.

Panel B: Secondary Factors

Conditions

Significant judgments:
- Significant judgments required for material account balances.

High management turnover:
- Management turnover is high.

Decentralized organization:
- Organization is decentralized without adequate monitoring.

New client:
- New client with no audit history, or sufficient information not available from predecessor auditor.

Rapid growth:
- Company is in a period of rapid growth.*

Inexperienced management:
- Company has inexperienced management.

Conflict of interest:
- A conflict of interest exists within the company and/or its personnel.*

Motivation

Rapid growth:
- Company is in a period of rapid growth.

Rapid industry change:
- The rate of change in the client's industry is rapid.

Sensitive operating results:
- Operating results are highly sensitive to economic factors.

Incentive compensation:
- Compensation arrangements are based on recorded performance.

Adverse legal circumstances:
- Company is confronted with adverse legal circumstances.

Significant portion of management's wealth:
- Company holdings represent a significant portion of management's personal wealth.

Management's job threatened:
- Management personnel perceive their job is threatened by poor performance.

Attitude

Weak internal control:
- Company has a weak internal control environment.*

Conflict of interest:
- A conflict of interest exists within the company and/or its personnel.*

Poor reputation:
- Management's reputation in the business community is poor.

Frequent disputes with auditor:
- Management has engaged in frequent disputes with the auditors.

Undue pressure on auditor:
- Client places undue pressure on the auditors.
- Client has engaged in opinion shopping.

Disrespectful attitude:
- Client personnel display a hostile attitude toward the auditors.
- Client management displays significant disrespect for regulatory bodies.
- Client personnel display significant resentment of authority.

* Indicates the factors appearing in two categories.

nonfraud cases. Second, they used a cascaded logit approach to provide different weights for individual factors and individual components, with a likelihood assessment as model output. A test set of 382 cases was subdivided into an estimation sample and a holdout sample. The estimation sample included 180 cases: 37 fraud and 143 nonfraud. Using a cutoff value of 0.03, the model achieved within-sample correct classification rates of 97 percent on the fraud cases and 75 percent on the nonfraud cases. The model achieved holdout-sample correct classification rates of 95 percent on the fraud cases and 67 percent on the nonfraud cases using the cutoff value of 0.03.

With these prior studies of models to predict management fraud as a foundation, we turn to an outline of a generalized qualitative-response model (EGB2).

3. A Generalized Qualitative-response Model (EGB2)

Qualitative-response models are 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, these models have been used to predict the probability that an individual will default on a loan or that a corporation will declare bankruptcy (e.g., Bar Niv and McDonald 1992). Important to our research study is the fact that qualitative-response models share a common structure that facilitates development of a generalized form, which is presented here.

We first outline the common structure of qualitative-response models and the particular structures of the popular probit and logit models are derived. Generalizations of the probit and logit models that allow for the possibility of increased predictive capability are then presented.

The general form for qualitative response models is

$$\Pr(Y = 1 | X) = F(X'\beta) = \int_{-\infty}^{X'\beta} f(z|\theta)dz, \qquad (1)$$

where F and f are the cumulative distribution and probability density functions, respectively, $\boldsymbol{\theta}$ represents distributional parameters, Y represents the binary dependent variable being predicted, **X** specifies a $k \times 1$ vector of exogenous variables useful in predicting Y, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of unknown parameters that generate scores ($Z = \mathbf{X}'\beta$). 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, $f(z|\boldsymbol{\theta})$, to estimate the vectors $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ 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 $f(z|\boldsymbol{\theta})$ 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

$$\mathrm{EGB2}(z; a, b, p, q) = e^{zap}/(b^{ap}B(p, q)(1 + (e^{az}/b^{ap}))^{p+q}), \qquad (2)$$

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

Note that a, b, p, and 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., Logistics(z) $= \mathrm{EGB2}(-z; a = 1, b = 1, p = 1, q = 1) = e^{-z}/(1 + e^{-z})^2$).

The probit results from the limiting case of (2) when $f(z{:}\boldsymbol{\theta})$ is selected to be the standard normal

$$N(z{:}\ 0, 1) = \lim_{\substack{a\to 0\\ q\to\infty}}[\mathrm{EGB2}(z; a, b = (a^2q)^{1/a}, p = 1/a^2, q)] = e^{-z^2/2}/\sqrt{2\pi}.$$

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 a approaches zero is a lognormal (see Kalbfeisch and Prentice 1980 or McDonald 1984).

The probit and logit models are very similar; however, the logit model has thicker tails than the probit model.[3] The probit model does not have a closed

[3] Amemiya (1992) and Maddala (1983) provide excellent discussions of probit and logit models.

form for the cumulative distribution and (1) must be evaluated numerically. The EGB2 can be obtained from the GB2 in McDonald (1984) by making the transformation $y = e^z$ and multiplying the resultant expression by the Jacobian of the transformation (e^z).[4]

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) which permit the distribution of Z to be asymmetric are based on the Burr3 and Burr12 distributions labeled here, EBurr3 and EBurr12:

$$\text{EBurr3}(z; a, b, p) = \text{EGB2}(z; a, b, p, q = 1),$$

$$\text{EBurr12}(z; a, b, q) = \text{EGB2}(z; a, b, p = 1, q).$$

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

$$F_{\text{EBurr3}}(z; a, b, p) = ((e^z/b)^a/(1 + (e^z/b)^a))^p,$$

$$F_{\text{EBurr12}}(z; a, b, q) = 1 - 1/(1 + (e^z/b)^a)^q,$$

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

$$\max \sum_t [Y_t \ln F(X_t'\beta; a, b, p, q) + (1 - Y_t) \ln(1 - F(X_t'\beta; a, b, p, q))]$$

over the parameters $\boldsymbol{\beta}$ and the relevant distributional parameters.[5] Except for certain limiting cases, the 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.

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

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

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.

In our study we apply the flexibility (not having to specify a particular functional form) of EGB2 with promising results. EGB2 was also found valuable in its facility for incorporating asymmetric costs of type I and type II errors.

We note in passing 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 is necessarily maximized.

[4] The GB2 has been referred to as a generalized F by Kalbfeisch and Prentice (1980) and as the beta prime distribution by Patil et al. (1984). The EGB2 can be obtained as a generalization of a Pearson type VI distribution and has also been referred to as a generalized logistic distribution by Johnson and Kotz (1970). The cumulative distribution $F(z|\boldsymbol{\theta})$ maps the scores ($Z = \mathbf{X}'\boldsymbol{\beta}$) into the unit interval [0, 1]. The density of $f(z|\boldsymbol{\theta})$ is symmetric for the logistic and normal density functions and for the EGB2, if $p = q$.

[5] Since absolute run times vary with machines and hardware configurations, we report EGB2 run time average, *relative* to probit and logit. For the data set considered, the EGB2 model takes twice as long to converge as the probit model, which in turn takes twice as long as the logit model. The EGB2 requires estimating two more parameters (distributional) than probit or logit. The cdf for the logit model can be written in closed form; whereas, the cdf for the probit or EGB2 models involves an infinite series, which must be approximated.

4. An Application to Management Fraud

4.1 Sample Data

The current study uses the same data set as Bell et al. (1993). It includes the 77 fraud cases collected by LEW (1989) and the 305 nonfraud cases collected by Bell et al. (1993). The dependent variable is the presence or absence of fraud. The independent variables are the risk factors shown in Figure 1 and were evaluated as present (1) or absent (0) on each of the 382 cases by members of the audit team.

The 77 fraud cases are described in detail in LEW (1989) and were each judged by responding partners to be material to the financial statements. Although the accuracy of information about each fraud was dependent on the engagement partner's ability to recall the surrounding events, efforts were made by the researchers to mitigate the effects of hindsight bias. For example, the instructions to the survey instrument asked respondents to "select an instance about which you have a good recollection and therefore, your response will be accurate and complete." When asking respondents to recollect whether risk factors were present, the researchers were careful to sequentially ask first whether the given factor applied to the engagement, and second to indicate whether the factor was apparent to the auditor during audit planning. For the 305 nonfraud engagements, partners' responses were "real time" judgments about the presence or absence of given risk factors and are described in Bell et al. (1993). Table 1, Panel A provides sample demographics by client industry and ownership characteristics. Panel B provides information based on the number of years the entity has been a client of the firm.

4.2 Methodology

The simplest method for estimating model error rate is the single estimation and test as done by Bell et al. (1993). The sample cases are divided into two groups: a model-building group and a test group. The model is independently estimated from the first group and the error estimate is computed from the performance of the resulting model on the test cases.

While simple in concept, the single estimation-and-test experiment can produce results that may not accurately reflect expected performance on the universe of problem instances (true error rate). In particular, it has been shown that for classification problems, 1000 or more cases are necessary to ensure that (for the single model-and-test experiment) the error rate on the test cases is very close to the true error rate (Efron 1983). Most studies do not have the luxury of such a large set of cases with which to experiment.

For smaller sample sizes, an effective alternative is random resampling via methods such as "leaving-one-out" (sometimes called the jackknife method). Given the availability of n cases, a model is developed using $n - 1$ of those cases and that model is then tested on the single holdout case. This is repeated n times. Thus, each case is used as a test case and each case influences the structure of nearly every model. While the "leaving-one-out" method is effective and reliable (cf., Efron 1983), it can be computationally prohibitive depending on the sample size.

Weiss and Kapouleas (1989) observe that the leaving-one-out method is a special case of k-fold classification methods with $k = 1$. If the sample size is over 100, Weiss and Kapouleas (1989) recommend the more general method (with $k > 1$). The cases are randomly divided into k mutually exclusive test partitions of approximately equal size. The union of the k partitions should comprise the entire sample. Given a partition p_i, all cases not found in p_i are used to estimate a model. The estimated model is then tested on the partition.

Our analysis utilized the latter approach. We used a 19-fold classification method (20 for two of the trials) repeated 20 times, for a total of 382 cases in the data set. This approach has previously been used by Hansen et al. (1992) for examining audit-opinion decisions and litigation.

5. Results

5.1 Symmetric Misclassification Costs

Our first analysis of the data uses the EGB2 model and assumes symmetric misclassification costs. A case was classified as being "fraud" if the predicted probability was greater than 0.5. Overall, EGB2 yielded an average 89.3 percent accuracy rate (standard error = 4.76) in classifying the cases across the 20 holdout samples. This

Table 1 Sample Demographics

Panel A: Client Industries and Ownership Characteristics

	Fraud						Nonfraud					
	Private		Public		Total		Private		Public		Total	
	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent	Count	Percent
Agribusiness	3	3.9%	0	0.0%	3	3.9%	7	2.3%	1	0.3%	8	2.6%
Banking	2	2.6%	8	10.4%	10	13.0%	14	4.6%	9	3.0%	23	7.5%
Education, Gov't, & Other Not-for-Profit	3	3.9%	0	0.0%	3	3.9%	73	23.9%	0	0.0%	73	23.9%
Financial Services	5	6.5%	0	0.0%	5	6.5%	9	3.0%	3	1.0%	12	3.9%
High Tech	2	2.6%	5	6.5%	7	9.1%	10	3.3%	4	1.3%	14	4.6%
Health Care	1	1.3%	1	1.3%	2	2.6%	13	4.3%	1	0.3%	14	4.6%
Insurance	2	2.6%	1	1.3%	3	3.9%	13	4.3%	1	0.3%	14	4.6%
Manufacturing	3	3.9%	11	14.3%	14	18.2%	47	15.4%	4	1.3%	51	16.7%
Merchandising	4	5.2%	3	3.9%	7	9.1%	28	9.2%	6	2.0%	34	11.1%
Real Estate	1	1.3%	2	2.6%	3	3.9%	11	3.6%	2	0.7%	13	4.3%
Savings & Loan	5	6.5%	3	3.9%	8	10.4%	11	3.6%	5	1.6%	16	5.2%
Other	7	9.1%	5	6.5%	12	15.6%	25	8.2%	8	2.6%	33	10.8%
Totals	38	49.4%	39	50.6%	77	100.0%	261	85.6%	44	14.4%	305	100.0%

Panel B: Number of Years Firm Has Been Auditor

	Fraud		Nonfraud	
	Count	Percent	Count	Percent
1st Year	17	22.1%	26	8.5%
2–5 Years	30	39.0%	121	39.7%
6–10 Years	25	32.5%	64	21.0%
>10 Years	5	6.5%	91	29.8%
No Response	0	0.0%	3	1.0%
Totals	77	100.0%	305	100.0%

is slightly higher than the 85.7 percent accuracy rate achieved by Bell et al. (1993) in their one holdout sample.

Figure 2 presents the average classification matrix for the 20 holdout samples. The type I error is 4.5 percent (0.7/15.5) while the type II error is 37.1 percent (1.3/3.5).

Table 2 contains the weights for the variables in a high-performance (95 percent accuracy) EGB2 model. The lowest performance of any of the models generated from the 20 trials was 81 percent accuracy. The standard errors for factor weights were all less than 15 percent, meaning that the weights shown in Table 2 varied only moderately across the 20 models. Meaningful comparisons to the weights generated by the Bell et al. model are not possible because, in their two-tiered model, tier-1 weights do not reflect the impact of a factor on the final probability from tier 2. Also, some factors are included in more than one tier-1 model, meaning that the overall factor effects are really a combination of effects on tier-1 probabilities transformed with tier-1 weights.

Figure 2 Classification Matrix of Average Prediction Performance* (Symmetric Cost Assumption)

Predicted	Actual Fraud	Nonfraud	Totals
Fraud	2.2	0.7	2.9
Nonfraud	1.3	14.8	16.1
Totals	3.5	15.5	19.0

* The values in the cells represent the average outcomes across the 20 holdout samples.

5.2 Asymmetric Misclassification Costs

As we discussed earlier, failing to predict fraud when it occurs (type II) is much more costly to CPA firms than predicting fraud when it does not occur (type I) because type I errors lead to overauditing while type II errors lead to litigation. EGB2 allows straightforward consideration of asymmetric cost assumptions for the two types of errors in the following manner.

Figure 3 Payoff Matrix

Prediction	Actual Fraud	Nonfraud
Fraud	Π_{11}	Π_{10}
Nonfraud	Π_{01}	Π_{00}

Consider the payoff matrix illustrated in Figure 3. Here Π_{ij} is the payoff (loss) of predicting outcome I for actual outcome j, where $I = 1$ is fraud, and $I = 0$ is nonfraud. Π_{00} is the reward from correctly predicting the absence of management fraud, Π_{01} is the return

Table 2 Factor Weights for a High Performance EGB2 Model (95 percent accuracy)*

	Predictive Factor	Weight
Q12.	Is there a need to cover up an illegal act?	40.00
Q4.	Does your experience with management indicate a degree of dishonesty?	8.99
Q18.	Is the client's organization decentralized without adequate monitoring?	3.87
Q14.	Are there frequent and significant difficult-to-audit transactions or balances?	3.85
Q16.	Is a significant amount of judgement involved in determining the total of an account balance or class of transactions?	3.51
Q20.	Does the client have solvency problems?	3.24
Q5.	Does management display a propensity to take undue risks?	2.97
Q15.	Is the client a public company?	2.88
Q8.	Do client personnel display significant resentment of authority?	2.29
Q6.	Does management display significant disrespect for regulatory bodies?	1.96
Q11.	Is this a new client?	1.90
Q21.	Does a conflict of interest exist involving the client entity and/or its personnel?	1.82
Q10.	Do key managers exhibit strong personality anomalies?	1.71
Q3.	Have managers recently entered into collusion with outsiders?	1.42
Q22.	Do accounting personnel exhibit inexperience or laxity in performing their duties?	1.41
Q9.	Is Management's attitude toward financial reporting unduly aggressive?	1.38
Q2.	Does management place undue emphasis on meetings earnings projections or quantitative targets?	1.00
Q23.	Is the client confronted with adverse legal circumstances?	0.36
Q1.	Are management operating and financial decisions dominated by a single person or a few persons acting in concert?	−0.09
Q13.	Does the client have a weak control environment?	−0.51
Q7.	Have managers lied to the auditors or been overly evasive in responses to audit inquiries, or have they shown some other indications of dishonesty?	−1.06
Q24.	Is the client's profitability relative to its industry inadequate or inconsistent?	−1.20
Q19.	Is the client in a period of rapid growth?	−1.42
Q17.	Have there been instances of material management fraud in prior years?	−2.03

* The lowest performing model in the 20 trials yielded 81 percent accuracy. The standard errors of factor weights were all less than 15 percent.

associated with a type II error, Π_{10} is the return associated with a type I error, and Π_{11} represents the return from correctly predicting fraud.

Let p_{ij} denote the corresponding probabilities. Let z^* be the decision threshold, such that for larger values of z^* a client will be classified as fraud. The conditional probabilities will depend on the distribution of scores from fraud (F) and nonfraud (N) clients. The expected return as a function of z^* is implicitly defined by the threshold that maximizes the expected return given by

$$E(\text{return}) = \sum_{i,j} p(j \mid I) p_i \, \Pi_{ij} \, .$$

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^* implicitly defines the benefit-maximizing decision rule as the solution to

$$f_N(z^* \mid N)/f_F(z_* \mid F) = p_1(\Pi_{11} - \Pi_{10})/p_0(\Pi_{00} - \Pi_{01}).$$

If the prior probabilities of fraud and nonfraud are equal, and if $\Pi_{11} - \Pi_{10}$ is equal to $\Pi_{00} - \Pi_{01}$, 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_1 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.[6,7]

These procedures are readily accommodated by EGB2 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 weights used in calculating EGB2 scores are unaffected. These methods should be implementable in other models, as well, including that of Bell et al. (1993).

While we do not know the exact costs of misclassification for CPA firms, there is little doubt that the costs of type II errors are considerably greater than for type I errors. Recent information published by the Big 6 firms (Arthur Andersen & Co. et al. 1993) provides some insight into the extent of the auditor's litigation problem:

1. Total expenditures for 1991 for settling and defending lawsuits by the Big 6 were \$477 million. This represents 9 percent of their auditing and accounting revenues in the United States. More recent data suggests that litigation costs are close to 12 percent of revenues.
2. The average 10b-5 securities law claim against the Big 6 in 1991 was \$85 million with an average settlement of \$2.7 million and average legal costs of \$3.5 million.
3. The estimated damage claims against the entire accounting profession are approximately \$30 billion.

Whereas the previous analysis assumed that the costs of type I and type II errors were the same, we arbitrarily assume that the costs of a type II error are at least 10 times that of a type I error.[8] With these new parameters, we ran the same repeated experiments as before. While our assumption is not definitive in terms of the asymmetric costs for management fraud, the application does demonstrate the capabilities of EGB2 to incorporate asymmetric costs of misclassification.

[6] Without loss of generality, a and b can be assumed to be unity. In our analysis, the following values of p and q were determined: $p = 18.3$, $q = 0.097$. This provides a strong indication of asymmetry. We can test for closeness to the logistic distribution of Bell et al. as follows: Log likelihood values: $l_{EGB2} = -79.4$, $l_{Logit} = -83.6$; H_0: EGB2 = Logit; LR = 2(−79.4 + 83.6) = 8.4; Asymptotically $\chi^2_{(2)}$ with 95% critical value of 5.99. Therefore, reject the null hypothesis.

[7] The weights do not change when the EGB2 model is run assuming asymmetric costs. What changes is the point in the cdf at which a case is placed in the "fraud" or "nonfraud" category. In particular, the estimated weights or coefficients of the explanatory variable (β) are obtained by maximizing the log likelihood function,

$$\sum_t [Y_t \ln F(x_t\boldsymbol{\beta}; \boldsymbol{\theta}) + (1 - Y_t) \ln(1 - F(x_t\boldsymbol{\beta}; \boldsymbol{\theta}))],$$

over the regression ($\boldsymbol{\beta}$) and the distributional ($\boldsymbol{\theta}$) parameters for arbitrary (symmetric or asymmetric) cumulative distribution functions. Thus, the estimated coefficients are independent of the relative costs of Type I and Type II errors.

[8] Our use of a multiple of 10 is very conservative. Kiernan and Lewin (1994, p. 36) provide the following quote:

> . . . There is no correspondence, of course, between the amount of damages to which an accountant is exposed and the amount of audit fee. In the Standard Charter Bank action against Price Waterhouse, for example, the jury's verdict was more than 2,400 times larger than the audit fee for the 1986 audit.

In this case, the overall accuracy rate was 85.3 percent. This compares with 89.5 percent under the symmetric cost assumption. However, the overall average accuracy does not have the same meaning because, in this analysis, it is more important to classify a fraud client correctly than it is to classify a nonfraud client correctly.

Figure 4 shows the results of the analysis for type I and type II errors. The type I error rate is 15.5 percent (2.4/15.5). This is an increase from the 4.5 percent reported for the symmetric cost assumption. More importantly, the type II error rate is 11.4 percent (0.4/3.5) which is a substantial reduction from the 37.1 percent reported for the symmetric case.

As the fraud issue is studied in more depth, the actual misclassification costs should become better understood. The point of emphasis here is that those costs can easily be included in EGB2, thus focusing the analysis on the type of prediction that is most important.

Figure 4 **Classification Matrix of Average Prediction Performance* (Asymmetric Cost Assumption**)**

	Actual		
Predicted	Fraud	Nonfraud	Totals
Fraud	3.1	2.4	5.5
Nonfraud	0.4	13.1	13.5
Totals	3.5	15.5	19.0

* The values in the cells represent the average outcomes across the 20 holdout samples.

** This analysis assumes a multiple of 10 for the cost of type II error to type I error.

6. Summary and Conclusions

EGB2 provides a four-parameter generalized qualitative-response model whose flexible specification includes several useful models as special cases, with the "best" model being determined by the test data. This flexibility is an approximate statistical metaphor for current methods being developed by the artificial intelligence community. Specifically, White (1989) has shown that feedforward neural networks, which require no prespecification of functional form, perform the same stochastic approximation as nonlinear regression, of which EGB2 is a generalized qualitative-response model.

Since Bell et al. (1993) performed only a single trial and test experiment (with 85.5 percent predictive accuracy), their results may not be as reliable as a replicated study. EGB2 averaged 89.3 percent predictive accuracy over 20 trials. When EGB2 was adjusted for asymmetric misclassification costs, its overall accuracy dropped to 85.5 percent, but the rate of costly type II errors decreased markedly.

From a theoretical standpoint, EGB2 provides the user with considerable flexibility and power. This study offers evidence that EGB2 can provide useful analysis for complex practical applications. Empirical studies from other problem domains will no doubt provide more insight and evidence concerning EGB2's capabilities. Current work is being directed toward extending the adaptive functions of EGB2.[9]

[9] The authors would like to thank the two anonymous reviewers for their helpful comments.

References

Amemiya, T., "Qualitative-Response Models: A Survey," *J. Economic Literature*, 19 (1992), 1483–1536.

American Institute of Certified Public Accountants, *AICPA Professional Standards*, AICPA, New York, 1992.

——, *Report of the National Commission on Fraudulent Financial Reporting*, AICPA, New York, 1987.

Arthur Andersen & Co., Coopers & Lybrand, Deloitte & Touche, Ernst & Young, KPMG Peat Marwick, and Price Waterhouse, *The Liability Crisis in the United States: Impact on the Accounting Profession*, A Statement of Position, August 6, 1992.

Bar Niv, R. and J. McDonald, "Identifying Financial Distress in the Insurance Industry: A Synthesis of Methodological and Empirical Issues," *J. Risk and Insurance*, 59 (1992), 543–574.

Bell, T., S. Szykowny, and J. Willingham, "Assessing the Likelihood of Fraudulent Financial Reporting: A Cascaded Logit Approach," Working Paper, KPMG Peat Marwick, Montvale, NJ, 1993.

Clarke, D. and J. McDonald, "Generalized Bankruptcy Models Applied to Predicting Consumer Credit Behavior," *J. of Economics and Business*, 44 (1992), 47–62.

Efron, B., "Estimating the Error Rate of a Prediction Rule," *J. American Statistical Association*, 78 (1983), 316–333.

Hansen, J., J. McDonald, and J. Stice, "Artificial Intelligence and Generalized Qualitative-Response Models: An Empirical Test on Two Audit Decision-Making Domains," *Decision Sci.*, 23 (1992), 708–723.

Johnson, N. and S. Kotz, *Continuous Univariate Distributions* (Vol. 2), Houghton Mifflin, New York, 1970.

Kalbfeisch, J. and R. Prentice, *The Statistical Analysis of Failure Times*, Wiley, New York, 1980.

Kiernan, J. S. and L. L. Lewin, "Those Monstrous Damage Awards Against Accountants," *CPA J.*, April 1994, 36–41.

Loebbecke, J. and J. Willingham, "Review of SEC Accounting and Auditing Enforcement Releases," Unpublished Working Paper, Montvale, NJ, 1988.

——, M. Eining, and J. Willingham, "Auditors' Experience with Material Irregularities: Frequency, Nature, and Detectability," *Auditing: J. Practice & Theory*, 9 (Fall 1989), 1–28.

Maddala, G., *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, England, 1983.

McDonald, J., "Some Generalized Functions for the Size Distribution of Income," *Econometrica*, 52 (1984), 647–663.

Palmrose, Z., "Litigation and Independent Auditors: The Role of Business Failures and Management Fraud," *Auditing: J. Practice & Theory*, 6 (Spring 1987), 90–103.

Patil, G., M. Boswell, and M. Ratnaparkhi, *Dictionary and Classified Bibliography of Statistical Distributions in Scientific Work: Continuous Univariate Models*, International Cooperative Clearinghouse, Burtonsville, MD, 1984.

Pincus, K., "The Efficacy of a Red Flags Questionnaire for Assessing the Possibility of Fraud," *Accounting Organizations and Society*, 14 (1989), 153–163.

Weiss, S. and I. Kapouleas, "An Empirical Comparison of Pattern Recognition, Neural Nets, and Machine-Learning Classification Methods," *Proc.* 1989 *International Joint Conf. on Artificial Intelligence*, Detroit, MI, 1989.

—— and C. Kulikowski, *Computer Systems That Learn*, Morgan Kaufmann Publishers, San Mateo, CA, 1991.

White, H., "Learning in Artificial Neural Networks: A Statistical Perspective," *Neural Computation*, 1 (1989), 425–464.

Accepted by Bala V. Balachandran; received October 1993. This paper has been with the authors 5 months for 3 revisions.