© *The Journal of Risk and Insurance*, 2013, Vol. 80, No. 1, 121-143
DOI: 10.1111/j.1539-6975.2012.01467.x

A Robust Unsupervised Method for Fraud Rate Estimation

Jing Ai
Patrick L. Brockett
Linda L. Golden
Montserrat Guillén

Abstract

If one is interested in managing fraud, one must measure the fraud rate to be able to assess the degree of the problem and the effectiveness of the fraud management technique. This article offers a robust new method for estimating fraud rate, PRIDIT-FRE (PRIDIT-based Fraud Rate Estimation), developed based on PRIDIT, an unsupervised fraud detection method to assess individual claim fraud suspiciousness. PRIDIT-FRE presents the first nonparametric unsupervised estimator of the actual rate of fraud in a population of claims, robust to the bias contained in an audited sample (arising from the quality or individual hubris of an auditor or investigator, or the natural data-gathering process through claims adjusting). PRIDIT-FRE exploits the internal consistency of fraud predictors and makes use of a small audited sample or an unaudited sample only. Using two insurance fraud data sets with different characteristics, we illustrate the effectiveness of PRIDIT-FRE and examine its robustness in varying scenarios.

Introduction

Background and Motivation

Fraud is a fact of social behavior having increasingly important consequences including loss of revenues to businesses, government, and society. Fraud is also expensive, driving up cost for detection and fraud risk reduction. Boyer (2000) quotes a study by

Jing Ai is an Assistant Professor of Risk Management and Insurance at the Department of Financial Economics and Institutions, Shidler College of Business, The University of Hawaii at Manoa, Honolulu, Hawaii. Patrick L. Brockett is the Gus S. Wortham Chaired Professor in Risk Management and Insurance at the Department of Information, Risk, and Operations Management, Red McCombs School of Business, The University of Texas at Austin, Austin, Texas. Linda L. Golden is the Marlene and Morton Meyerson Centennial Professor in Business at Department of Marketing, Red McCombs School of Business, The University of Texas at Austin, Austin, Texas. Montserrat Guillén is a Chaired Professor in Economics at Department of Econometrics, University of Barcelona, Barcelona, Spain. The authors can be contacted via e-mail: jing.ai@hawaii.edu, brockett@mail.utexas.edu, utlindagolden@gmail.com, and mguillen@ub.edu, respectively. Jing Ai is the contact author.

the Rand Corporation Institute for Civil Justice estimating that in automobile insurance claims, questionable medical claims added between $13 and $18 billion to the nation's total automobile insurance bill in 1993. More recently, it is estimated that in 2007 claims fraud added between U.S. $4.8 billion to $6.8 billion to automobile injury insurance claims paid, or around 15 percent of total claim payment (Insurance Research Council [IRC], 2008). Viaene and Dedene (2004) detail results of an Insurance Research Council (IRC) and Insurance Services Office (ISO) 2001 study of property and casualty insurers where it was found that half of the respondents felt insurance fraud was "a serious problem."

As a result, active fraud control has gradually become an integrated part of business decision-making processes. Insurance companies must deal with fraud perpetrated by consumers on the firm and spend money on fraud detection and monitoring. A lot of research has focused on the fraud detection efforts, that is, assessing and ranking the fraud suspiciousness of individual claims, currently most of which are parametric and supervised.[1] For example, see Derrig and Ostaszewski (1995), Artis, Ayuso, and Guillén (2002), Viaene et al. (2002), and Caudill, Ayuso, and Guillen (2005) for insurance fraud detection. Also see Bolton and Hand (2002) for a review of statistical fraud detection methods. There are also limited unsupervised fraud detection methods available in the literature (cf. Brockett et al., 2002; Ai, Brockett, and Golden, 2009). However, in order to design an appropriate fraud management activity and assess its success, one must have a measure of fraud rate prior to intervention. As it is commonly said, you cannot manage something that you cannot measure. This article is devoted to this critical step in fraud risk management.

The present article investigates the fraud problem from a perspective distinct from detection of individual cases of fraud. It examines the rate of fraud in a population of claim files. Firms and governments need to obtain a fraud rate estimate to assess where and how much they are at risk for fraud (e.g., Caron and Dionne, 1999; Heron and Lie, 2009). In addition, some fraud detection techniques rely on fraud rate as an input to the analysis (e.g., see Durtschi, Hillison, and Pacini, 2004). Ultimately, an automated and continuous fraud monitoring system (Viaene et al., 2007) needs to include fraud rate estimation so that companies and regulators can ascertain if their fraud mitigation efforts are effective for internal control or peer comparison purposes.

However, obtaining a good estimate of fraud rate is very difficult (Caron and Dionne, 1999). First, as people and businesses always try to conceal fraudulent activities, getting a direct reported measure is virtually impossible (cf. Viaene and Dedene, 2004). A strand of theoretical literature has been developed in the context of insurance to study fraudulent behavior and fraud deterrence through the optimal design of insurance contracts and auditing policy (see, e.g., Townsend, 1979; Mookherjee and Png, 1989; Picard, 1996, 2000; Dionne and Gagne, 2001; Crocker and Tennyson, 2002; Dionne and Gagne, 2002; Schiller, 2006; Boyer, 2007).

Second, because a direct measure is difficult to obtain, a statistical estimation needs to be used. Sometimes a "count the number of fraud cases" type of estimation is

[1] Supervised learning (e.g., logistic regression) requires a training sample for parameter estimation. Unsupervised learning, on the other hand, does not require a training sample and relies on the data structure itself without knowledge of the dependent variable.

implemented inferring the population fraud rate from an audited sample. Alternatively, fraud rate estimation can be developed in the context of available fraud detection techniques, most of which are supervised and parametric. In addition to a large cost associated with frequent updates of samples and models necessary because fraudsters learn to avoid detection overtime, the above two approaches both rely on an audited sample that is potentially biased (cf. Boyer, 2000; Pinquet, Ayuso, and Guillen, 2007; Dionne, Giuliano, and Picard, 2009). The bias in the audited sample can arise from human assessor's bias and hubris and more importantly, from the auditing procedures used in many industry settings to obtain this sample, selecting the most suspicious cases to audit.

Possibly due to these difficulties, currently available estimates of fraud rate are largely based on conventional wisdom and they do not usually form a consensus (Tennyson, 2008). Some other sparsely available fraud rate estimates are based on a one-time targeted study and were subsequently applied to different contexts in general terms. For example, the often-cited 10 percent estimate of fraud rate was based on a 1992 study on the property and casualty insurance industry (Baker and Herbert, 1992). These targeted studies were specifically dedicated to understanding the scope of fraud in a given context and at the time of the study. It is more desirable to have a fraud rate estimation technique that can continuously monitor fraud activities in an automated manner, making use of a naturally generated sample in the claims-adjusting process and with relatively low cost. This is what we want to achieve in this article.

Purpose of the Article and Organization

Using the underlying mathematical formulation of an unsupervised fraud detection model PRIDIT (Principal Component Analysis of RIDITs) (Brockett et al., 2002; Ai, Brockett, and Golden, 2009), this article proposes a new methodology for fraud rate estimation called PRIDIT-FRE (PRIDIT-based Fraud Rate Estimation). As a fraud detection model, the PRIDIT method makes use of a consistent ensemble of fraud predictor variables within an individual claim file to obtain an overall fraud suspicion score for each claim. It enables one to rank order these claim files according to their level of fraud suspiciousness without access to a set of audited or classified cases. PRIDIT-FRE, on the other hand, provides an estimation of the aggregate measure of overall fraud rate in a population of claim files.

Consistent with practice, PRIDIT-FRE can be implemented with only a small amount of audited or classified cases (relative to the large set of cases with case characteristics only) naturally generated in the claims-adjusting process. Since highly suspicious claims are more likely to be selected for audit or further investigation, the naturally arisen data set will be biased in the represented percentage of fraudulent claims. PRIDIT-FRE is robust to this data set bias. PRIDIT-FRE can even be implemented with a sample of unaudited cases only. Other estimation methods (e.g., a "count" estimation method that uses the count of fraudulent claims in an audited sample as an estimate of population fraud rate) may fail to perform in light of this bias.

The remainder of the article is organized as follows. The "Overview of the PRIDIT Method Notations" section provides a brief introduction and gives the necessary notations from Brockett et al. (2002) and Ai, Brockett, and Golden (2009) for the unsupervised PRIDIT fraud detection methodology, from which PRIDIT-FRE is derived. The "A Simple Fraud Rate Estimation Method Based on the PRIDIT Method (PRIDIT-FRE)" section proposes the fraud rate estimator in PRIDIT-FRE and discusses the key challenges as well as the proposed solutions in implementing the estimation methodology.

The "An Empirical Illustration Using Insurance Fraud Data Sets" section presents an empirical illustration of the PRIDIT-FRE method and assesses its performance. We first briefly introduce the two insurance claims fraud data sets used in the article and describe the design of the empirical analyses. Next, we provide baseline estimation results and discuss a variety of robustness checks. In particular, we highlight the robustness of the proposed method to a biased audited sample that may invalidate a plausible "count" estimation approach.

In the "Extension: Estimating Fraud Rate Using PRIDIT-FRE Without a Sample of Audited Claims" section, we develop an extension of PRIDIT-FRE where only a sample of unaudited cases is needed. Less information is required and savings on the auditing expenses are realized. The "Conclusions and Implications" section discusses the conclusions and implications that can be drawn from this research.

OVERVIEW OF THE PRIDIT METHOD NOTATIONS

The fraud rate estimation method PRIDIT-FRE builds on the mathematical foundations underlying the fraud detection method PRIDIT (Brockett et al., 2002; Ai, Brockett, and Golden, 2009). The PRIDIT method was designed to rank order claim files according to their level of fraud suspicion without *a priori* knowledge of the actual fraud classification for even a sample of claim files. In the PRIDIT model, each claim file is assumed to have a relative position on an underlying latent "fraud suspicion" dimension, and the set of fraud predictor variables (or "red flags" for fraud) yields information concerning this position. The set of fraud predictors is selected and constructed by domain experts so that they bear a monotonic relationship with the suspicion level. PRIDIT concatenates the "hints" contained in these predictor variables to obtain a measure of claim file fraud suspicion. Details on the PRIDIT method along with proofs and empirical evaluations in the discrete predictor and continuous predictor cases can be found in Brockett et al. (2002) and Ai, Brockett, and Golden (2009). We provide below only the mathematical background and notations pertaining to constructing the PRIDIT-FRE estimator of fraud rate.

By assumption, each predictor variable, t, contained within a claim file has a monotonic relationship with the level of fraud suspiciousness of the claim file in that the lower the observed value x is on variable t the more suspicious it is of fraud. To exploit this relationship, a "variable score" $B(x)$ is calculated for each fraud predictor variable. We define the score $B(x)$ for response value x to variable t (i.e., $B_t = B(x)$) as the proportion of claims in the sample with response value less than x on variable t minus the proportion of cases with response value larger than x. Thus,

$$B(x) = \hat{F}(x^-) - [1 - \hat{F}(x)] = [\hat{F}(x) - \hat{P}(x)] - [1 - \hat{F}(x)] = 2\hat{F}(x) - 1 - \hat{P}(x),$$

where $\hat{F}(x)$ is the empirical distribution of the fraud predictor variable t, and $\hat{P}(x)$ is the sample proportion of responses to variable t, that equal x.

A discriminatory power measure for assessing the importance of each predictor in identifying fraudulent claim files is defined by calculating the expected predictor variable score for the fraudulent and non-fraudulent claims. More specifically, let $F_1(x)$ be the distribution function describing the random variation in responses to variable t among the class of fraudulent claims. Let $G_2(x)$ denote the distribution function of responses to variable t for non-fraudulent claims. F_1 does not equal G_2 since the predictor variables were constructed to help discriminate fraudulent claims from non-fraudulent claims. For all x, $F_1(x) \geq G_2(x)$; that is, the fraudulent claims are expected to have lower response values than non-fraudulent claims.

Define $\Delta(x) = F_1(x) - G_2(x) \geq 0$ and let θ, the fraud rate, denote the proportion of cases that belong to the fraud class. The expected variable score of a predictor variable t for a member of the fraud class can be calculated as (cf. Ai, Brockett, and Golden, 2009),

$$\mathrm{E}\left[B_t|\text{Fraud Class}\right] = 2(\theta - 1)\int_{-\infty}^{\infty} \Delta(x)dF_1(x).$$

The discriminatory power measure A_t for predictor variable t is defined by

$$A_t = 2\int_{-\infty}^{\infty} \Delta(x)dF_1(x) = 2\int_{-\infty}^{\infty} \Delta(x)dG_2(x).$$

Note that the more F_1 differs from G_2, the larger is A_t, indicating stronger discriminant ability of variable t. Therefore, we have

$$E\left[B_t|\text{Fraud Class}\right] = (\theta - 1)A_t \tag{1}$$

and similarly

$$E\left[B_t|\text{Non-Fraud Class}\right] = \theta A_t. \tag{2}$$

The relations between the expected variable scores for each class, the discriminatory power measure and the fraud rate as shown in Equations (1) and (2) are exploited to design the PRIDIT-FRE method to estimate the fraud rate in the next section.

A Simple Fraud Rate Estimation Method Based on the Pridit Method (Pridit-Fre)

An estimate of the fraud rate θ can be obtained based on the relationship between the discriminatory power measure A_t and the expected predictor variable t scores for the fraud and non-fraud classes given in Equations (1) and (2). We have, for each predictor variable t, E[B_t | Fraud Class] = $(\theta-1)A_t$ and

$E[B_t \mid \text{Non-Fraud Class}] = \theta A_t$, yielding $\theta = \frac{\sum_t E[B_t|Non\text{-}Fraud]}{\sum_t (E[B_t|Non\text{-}Fraud] - E[B_t|Fraud])}$, where the summation is over all predictor variables t. Therefore, we obtain an estimate $\hat{\theta}$ of θ by calculating the average variable score for each class,

$$\hat{\theta} = \frac{\sum_t \bar{B}_{2t}}{\sum_t (\bar{B}_{2t} - \bar{B}_{1t})}, \tag{3}$$

where $\bar{B}_{1t}$ and $\bar{B}_{2t}$ are the average variable t scores for fraud class and non-fraud class claims, respectively.[2] $\hat{\theta}$ given by Equation (3) is the proposed fraud rate estimator in the PRIDIT-FRE methodology and the consistency of this estimator is proven in the Appendix.

Using the fraud rate estimator $\hat{\theta}$, we design an approach to implement the PRIDIT-FRE estimation for the examined population of claims. Notice that the PRIDIT variable scores $B(x)$ are first calculated using the entire set of claims (audited and unaudited) where claim characteristics are known. However, in order to calculate for each class the average variable scores $\bar{B}_{1t}$ and $\bar{B}_{2t}$, we need to know the class membership (fraud or non-fraud) for at least a subset of claims. This can be done by using only a small sample where the fraud classifications are known (e.g., obtained via an audit which may be costly). We also propose an extension not to rely on any audited sample but to use a PRIDIT-generated fraud classification to implement the estimator.

An empirical evaluation of the PRIDIT-FRE methodology implemented with a small audited sample is presented in the "An Empirical Illustration Using Insurance Fraud Data Sets" section along with a discussion of different variations of this method and robustness checks. It is worth noting that although a "count the fraudulent claims" estimation method for fraud rate determination can be easily implemented using the same audited sample, the "count" estimation will fail if this sample is biased, as can happen with a naturally generated audited sample in the current practice where an initial screening procedure is used to select the most suspicious claims to audit (Pinquet, Ayuso, and Guillen, 2007; Dionne, Giuliano, and Picard, 2009). Our method, on the other hand, is robust to this sample bias, making it possible to directly use the naturally generated data and to incorporate fraud rate estimation as part of the automated fraud management process without incurring much additional cost. We design an experiment to empirically demonstrate this strength.

AN EMPIRICAL ILLUSTRATION USING INSURANCE FRAUD DATA SETS

Data Description and Fraud Definition Operationalization

We use two insurance fraud data sets to empirically illustrate the proposed method. The two data sets are taken from two different developed insurance markets (Spain

[2] An alternative estimation method is to obtain an estimate $\hat{\theta}_t$ of θ for each variable t by calculating the average variable scores for each class: $\hat{\theta}_t = \frac{\bar{B}_{2t}}{\bar{B}_{2t} - \bar{B}_{1t}}$, where $\bar{B}_{1t}$ and $\bar{B}_{2t}$ are the average variable t scores for fraud class and non-fraud class, respectively. A final estimate of θ can then be obtained from the individual estimates $\hat{\theta}_t$'s by taking the average: $\hat{\theta} = \frac{1}{T}\sum_{t=1}^{T} \hat{\theta}_t$. However, it can be shown theoretically by a double central limit theorem argument (cf. Brockett and Witt, 1982) that the method presented in the text should perform better. We conducted empirical experiments (not reported here) that confirm this.

and United States) containing two different major types of automobile insurance claims (physical damage and personal injury protection) with different fraud rates. By using both for the empirical illustration, we provide evidence on the robustness of our estimation method.

The first data set is a Spanish insurance fraud data set containing 1995 physical damage claims in automobile insurance. Claims in this data set have been classified by the insurer as fraudulent or legitimate, coded as a binary variable. This data set contains a set of fraud predictor variables which are used for the fraud rate estimation. The fraud rate for this data set is 49.97 percent, which was intentionally structured to have roughly equal fraud and non-fraud representation. See Artís, Ayuso, and Guillén (2002) for a variable description and more details of this data set.

The second data set is a U.S.-based personal injury protection claims data set produced by the Automobile Insurance Bureau (AIB) in Massachusetts. It consists of 1,399 claims, each of which was rated for likelihood of fraud (i.e., a suspicion score on a 0- to 10- point scale) by experts in the special investigation unit of AIB.

We ultimately use two different breakpoints in the human-assessed fraud suspicion scores to determine fraud classification and fraud rate for the U.S. data set. Consistent with Viaene et al. (2002), we first operationalize the notion of "fraud" to classify any claim with a suspicion score no less than 4 to be fraudulent, resulting in 396 identified fraud cases, or a fraud rate of 28.31 percent. We also conduct a subsequent robustness analysis with an alternative operational definition using a suspicion score of 7 or above as the breakpoint. For this definition, there are 123 claims resulting in a fraud rate of 8.79 percent.

This data set also contains a set of fraud predictor variables selected by the AIB claims experts (the "red-flag" fraud indicators). These predictors are then used to estimate the fraud rate. See Viaene et al. (2002) for a description of the variables and more details of the data set.

Experiment Design

In the empirical evaluations of PRIDIT-FRE, we treat the entire data set as the population of cases and estimate its fraud rate using samples taken from it. To obtain individual variable scores needed in the PRIDIT-FRE estimation, we first run the PRIDIT analysis on the *entire* data set using the fraud predictors. In order to calculate average variable scores for each class, we select a sample of claims from the data set and make use of the known class membership of this sample (i.e., the fraud classification variable). Since it is costly to obtain the fraud classification of the claims, the larger the sample size is, the higher is the cost incurred. However, using a larger sample is likely to reduce the estimation variance, thereby increasing the estimation accuracy. We compare estimation results from samples of various sizes to evaluate this trade-off.

To implement our fraud estimator, we select nine different sizes of random samples that contain 10 percent to 90 percent of the data at increments of 10 percent. For each sample of claims, the PRIDIT-FRE estimate is obtained using Equation (3) by calculating the average variable scores for each class and for each predictor variable. To make sure that our experimental results are not due to sample idiosyncrasy,

TABLE 1
PRIDIT-FRE Results Using Random Samples (Spanish Data Set)

Sample Size (% of Data Set)	Estimate (Mean)	MSE	Variance	Median	Interquartile Range
10	0.4981	0.0096	0.0096	0.4906	0.1458
20	0.4954	0.0016	0.0016	0.4900	0.0611
30	0.4908	0.0006	0.0005	0.4912	0.0379
40	0.5006	0.0006	0.0006	0.4971	0.0325
50	0.4978	0.0003	0.0003	0.4951	0.0282
60	0.5011	0.0000	0.0000	0.4997	0.0086
70	0.4972	0.0002	0.0002	0.4962	0.0202
80	0.4943	0.0001	0.0001	0.4947	0.0125
90	0.5010	0.0001	0.0001	0.5007	0.0095

Note: This data set has 1,995 claims and the true fraud rate equals 49.97 percent.

TABLE 2
PRIDIT-FRE Results Using Random Samples (U.S. Data Set)

Sample Size (% of Data Set)	Estimate (Mean)	MSE	Variance	Median	Interquartile Range
10	0.2809	0.0034	0.0034	0.2769	0.0520
20	0.2601	0.0058	0.0053	0.2616	0.0665
30	0.3069	0.0034	0.0028	0.2978	0.0612
40	0.2935	0.0012	0.0011	0.2857	0.0397
50	0.2770	0.0014	0.0014	0.2823	0.0486
60	0.2847	0.0007	0.0007	0.2855	0.0302
70	0.2824	0.0007	0.0007	0.2842	0.0346
80	0.2809	0.0003	0.0003	0.2814	0.0141
90	0.2831	0.0000	0.0000	0.2837	0.0056

Note: This data set has 1,399 claims and the true fraud rate equals 28.31 percent.

10 repetitions of random selections are done for each of the nine sample sizes (resulting in 90 estimation analyses) and the average of the 10 repetitions is the final estimate by each sample size.

Baseline Results

The estimated fraud rates by each sample size along with other summary statistics describing variations among the 10 repetitions are presented in Tables 1 and 2 for the Spanish and the U.S. insurance fraud data sets, respectively. In Table 1, we can see the fraud rate estimates are close to the true fraud rate value (49.97 percent) statistically and economically as indicated by the small mean square error (MSE). Also note that as the size of the sample increases, the MSE, the variance, and the interquartile range of the fraud rate estimates from different draws of random samples decrease quickly, indicating that even with a small audited sample, an accurate and reliable fraud rate estimate can be secured.

Table 2 presents the estimates for the U.S. data set. This data set has a much lower fraud rate 28.31 percent. Here we also find a pattern similar to the Spanish data set: good accuracy with even small audited samples.

Estimation by Incorporating Information in the Predictor Variable Importance Weights

As discussed in the "Overview of the PRIDIT Method Notations" section, the fraud detection method PRIDIT makes use of a set of predictor variables to calculate individual predictor variable scores and an overall suspicion score for each claim. Different weights are derived and assigned to these predictor variables to reflect their respective importance in predicting fraud.

To further investigate the accuracy of the fraud rate estimator, we operationalized simultaneously two variations of the original PRIDIT-FRE estimation taking into account information in the predictor variable importance weights.[3] First, we adjust for negative weights by reverse scoring the scales of these predictor variables (e.g., $0 \to 1$ and $1 \to 0$).[4] Second, we conduct the analysis with only those variables that are of high predictive importance,[5] that is, we are only summing across those variables in Equation (3).

Table 3 shows that accounting for negative weights and summing across only predictor variables with high importance weights do not generally improve the already accurate results of PRIDIT-FRE presented in Table 1. Likewise, estimation results in Table 4 are similar to what was presented in Table 2. Thus, the PRIDIT-FRE results are robust for both the Spanish and U.S. data sets, even when not adjusting for factors that might inhibit predictability such as including low importance predictor variables. However, by eliminating variables of lower importance, we use a smaller set of predictor variables to obtain fraud rate estimates that are equally accurate. This is important for the parsimony of the model and may result in cost savings as well.

[3] We also conduct analyses with each variation separately: adjusting for negative weights or summing across only variables with high importance weights. The results showed that PRIDIT-FRE is robust. Neither analysis showed material differences from the baseline results using the original set of variables. Therefore, to conserve space, we only present the results for the combined variation in all experiments presented here and later in the article.

[4] In the PRIDIT method, the predictor variables were selected and constructed by experts to monotonically assess positioning of the claim file on an underlying latent dimension of "fraud suspicion." If a situation occurs in which the actual relation of a predictor variable with being non-fraud is negative instead of positive (due to expert error or incorrect construction of the variable), it will result in this variable having a negative correlation with the overall summative score and hence a negative value for predictor variable importance weight, due to the correlative structure in the iterative process to determine weights. See Ai, Brockett, and Golden (2009) for more detailed comments. One limitation of our model is the assumption that domain experts will be able to select at least *some* useful predictor variables that correlate well with fraud suspicion. However, this assumption is needed for any estimation method to work.

[5] In our experiments, the threshold for "high predictive importance" is selected to be an importance weight value of 0.1.

TABLE 3
PRIDIT-FRE Results Adjusting for Negative Weights and Using Only Variables of High Importance (Spanish Data Set)

Sample Size (% of Data Set)	Estimate (Mean)	MSE	Variance	Median	Interquartile Range
10	0.4690	0.0103	0.0094	0.4708	0.1167
20	0.4846	0.0025	0.0023	0.5023	0.0357
30	0.5064	0.0006	0.0006	0.5023	0.0316
40	0.5000	0.0007	0.0007	0.4888	0.0251
50	0.4966	0.0005	0.0005	0.4958	0.0330
60	0.5014	0.0001	0.0001	0.4991	0.0094
70	0.4993	0.0001	0.0001	0.4960	0.0153
80	0.4980	0.0001	0.0001	0.4943	0.0148
90	0.5023	0.0000	0.0000	0.5027	0.0084

Note: This data set has 1,995 claims and the true fraud rate equals 49.97 percent. Negative weights are adjusted by reverse scoring the variable scales and a cutoff importance weight of 0.1 is used for selecting high importance predictors.

TABLE 4
PRIDIT-FRE Results Adjusting for Negative Weights and Using Only Variables of High Importance (U.S. Data Set)

Sample Size (% of Data Set)	Estimate (Mean)	MSE	Variance	Median	Interquartile Range
10	0.2828	0.0020	0.0020	0.2781	0.0451
20	0.2672	0.0036	0.0033	0.2677	0.0937
30	0.3019	0.0027	0.0023	0.3073	0.0480
40	0.2929	0.0011	0.0010	0.2993	0.0198
50	0.2810	0.0009	0.0009	0.2852	0.0426
60	0.2859	0.0004	0.0004	0.2908	0.0120
70	0.2857	0.0004	0.0004	0.2860	0.0220
80	0.2837	0.0002	0.0002	0.2797	0.0112
90	0.2810	0.0000	0.0000	0.2819	0.0131

Note: This data set has 1,399 claims and the true fraud rate equals 28.31 percent. Negative weights are adjusted by reverse scoring the variable scales and a cutoff importance weight of 0.1 is used for selecting high importance predictors.

Biased Audited Sample: When the "Count" Method for Estimation Does Not Work

The PRIDIT-FRE implementation presented previously requires a fraud classification for at least a sample of audited cases. Often with such an audited sample, fraud rate is estimated by simply counting the number of identified fraud cases in the sample (divided by the sample size). However, this "count" estimation will only work well in predicting the population fraud rate when the sample of cases examined has the same expected fraud rate as that of the population. This condition is not always satisfied in practice, as the sample selected for obtaining the fraud rate is usually not randomly chosen.

A major goal for fraud management is to identify in the claims-adjusting process those cases warranting further investigation or disciplinary actions through auditing. As it would be cost prohibitive to audit every case for suspiciousness, an optimal "red-flag" type of auditing strategy is often employed (cf. Dionne, Giuliano, and Picard, 2009; Yim, 2009; Tennyson and Salsas-Fom, 2002) necessitating an "initial-screening" sampling process as part of standard business practices. This auditing practice is at high risk of generating a biased sample overrepresenting suspicious cases. The resulting fraud rate estimate by the count method (or other supervised methods, for that matter) will be higher than what actually is in the population; that is, the count estimation method will only accurately predict the fraud rate of the *biased sample*, not the population.

We could, of course, select and audit random samples (a "suboptimal" auditing strategy) to guarantee effective count estimation, but this will lead to unnecessary additional expenses not incurred by PRIDIT-FRE, which produces accurate estimates of the population fraud rate even when the naturally generated audited sample *is* biased and can thus be easily incorporated into an automated fraud management system without adding much cost. The unsupervised nature of PRIDIT-FRE allows it to rely on calculated predictor variable scores for the entire population that can be obtained at relatively little cost, rather than relying directly and solely on the count of fraud cases in the sample. Under PRIDIT-FRE, the sample is used only to identify fraud classifications of a small amount of cases so that *average* variable scores can be obtained for each class.

Biased Sample: Baseline Results. We design an empirical experiment to evaluate and compare the estimation performance of PRIDIT-FRE and the count estimation in the presence of a biased sample using the Spanish and U.S. data sets as populations. The Spanish data set population has a fraud rate of 49.97 percent and the U.S. data set population has a fraud rate of 28.31 percent. To emulate a biased audited sample found in practice, we create two samples from these two data sets respectively such that the fraud rates of the (biased) samples and the original data sets are significantly different.[6]

Claims in the Spanish data set are randomly selected from each class to create a biased sample of 1,000 claims with a target sample fraud rate of 25 percent (versus 49.97 percent in the original Spanish data set). Similarly, the biased sample created from the U.S. data set has 800 claims and the fraud rate is 12.5 percent (versus 28.31 percent in the original U.S. data set). The sample sizes and fraud rates are chosen for illustrative purposes to demonstrate the effects of having a largely biased sample fraud rate.

The same estimation procedure as was previously described in the "Baseline Results" section is implemented to estimate the fraud rates of the original data sets. More

[6] Since the fraud rates of the original data sets are already very large (because they are actually audited claims samples which are used as populations here), our biased samples are constructed so that the sample fraud rates are much lower than those of the original data sets. This construction is actually opposite to the real world situation, but the logic works in the same way.

TABLE 5
PRIDIT-FRE and Count Method Results for a Population Having a Fraud Rate of 49.97 Percent Using Random Subsamples Having a Fraud Rate of 25 Percent (Spanish Data Set)

Subsample Size (% of the Biased Sample)	PRIDIT-FRE Estimation					Count Estimate
	Estimate (Mean)	MSE	Variance	Median	Interquartile Range	
10	0.5072	0.0119	0.0118	0.5088	0.1284	0.2509
20	0.4989	0.0051	0.0051	0.4837	0.1034	0.2535
30	0.5027	0.0021	0.0021	0.5021	0.0609	0.2532
40	0.4985	0.0016	0.0016	0.5003	0.0505	0.2497
50	0.4994	0.0011	0.0011	0.4976	0.0434	0.2524
60	0.4983	0.0008	0.0008	0.4955	0.0371	0.2499
70	0.4980	0.0004	0.0004	0.4981	0.0238	0.2506
80	0.5004	0.0003	0.0003	0.5008	0.0246	0.2505
90	0.4990	0.0001	0.0001	0.4984	0.0168	0.2499
100	0.4997	0.0000	0.0000	0.4997	0.0000	0.2500

Note: This data set ("population") has 1,995 claims and the biased sample has 1,000 claims.

TABLE 6
PRIDIT-FRE and Count Method Results for a Population Having a Fraud Rate of 28.31 Percent Using Random Subsamples Having a Fraud Rate of 12.5 Percent (U.S. Data Set)

Subsample Size (% of the Biased Sample)	PRIDIT-FRE Estimation					Count Estimate
	Estimate (Mean)	MSE	Variance	Median	Interquartile Range	
10	0.2994	0.0441	0.0438	0.2907	0.1978	0.1273
20	0.2971	0.0118	0.0116	0.3027	0.1149	0.1274
30	0.3100	0.0161	0.0154	0.2896	0.1093	0.1243
40	0.2991	0.0047	0.0044	0.2947	0.0818	0.1260
50	0.3022	0.0030	0.0026	0.2985	0.0640	0.1246
60	0.2904	0.0019	0.0018	0.2914	0.0523	0.1271
70	0.2989	0.0015	0.0013	0.3007	0.0402	0.1252
80	0.2900	0.0112	0.0006	0.2909	0.0336	0.1248
90	0.2929	0.0003	0.0002	0.2940	0.0196	0.1253
100	0.2938	0.0001	0.0000	0.2938	0.0000	0.1250

Note: This data set ("population") has 1,399 claims and the biased sample has 800 claims.

specifically, variable scores are first obtained by running PRIDIT on the entire original data sets. After the two biased samples are created, subsamples of different sizes are randomly taken from the *biased samples* and fraud classifications in these subsamples are then used in Equation (3) to estimate the population fraud rate. The sizes of subsamples range from 10 percent to 100 percent of the biased samples at increments of 10 percent. Count estimates of the fraud rate are also obtained from these subsamples.

TABLE 7
PRIDIT-FRE Results for a Population Having a Fraud Rate of 49.97 Percent Using Random Subsamples Having a Fraud Rate of 25 Percent, Adjusting for Negative Weights and Using Only Variables of High Importance (Spanish Data Set)

Subsample Size (% of the Biased Sample)	PRIDIT-FRE Estimation Estimate (Mean)	MSE	Variance	Median	Interquartile Range	Count Estimate
10	0.5283	0.0122	0.0114	0.5300	0.1480	0.2509
20	0.5231	0.0058	0.0053	0.5174	0.0907	0.2535
30	0.5140	0.0026	0.0024	0.5170	0.0652	0.2532
40	0.5104	0.0026	0.0025	0.5026	0.0639	0.2497
50	0.5128	0.0015	0.0013	0.5129	0.0473	0.2524
60	0.5086	0.0008	0.0007	0.5089	0.0358	0.2499
70	0.5098	0.0005	0.0004	0.5104	0.0280	0.2506
80	0.5088	0.0005	0.0004	0.5087	0.0222	0.2505
90	0.5111	0.0003	0.0002	0.5103	0.0182	0.2499
100	0.5111	0.0001	0.0000	0.5111	0.0000	0.2500

Note: This data set ("population") has 1,995 claims and the biased subsample has 1,000 claims. Negative weights are adjusted by reverse scoring the variable scales and a cutoff importance weight of 0.1 is used for selecting high importance predictors.

To reduce the undue influence that might arise from a particular randomly generated biased sample or from a particular random subsample used in our experiments, for each data set, we create 10 such biased samples and within each biased sample use 10 repetitions of random selections of subsamples of different sizes. The estimation statistics for a random subsample of a certain size are averaged across the 10 biased samples, as are presented in Tables 5 and 6. Again, PRIDIT-FRE produces fraud rate estimates that accurately and reliably approximate the true population represented by the *original* data sets. The count method, on the other hand, only predicts the fraud rates of the *biased samples*, which are very different from those of the population. As the goal is to predict the population fraud rate which the enterprise needs to be tracking over time, not the fraud rate for any particular sampling, PRIDIT-FRE is a more managerially relevant predictor. The count method estimates generated from a biased audited sample, in practice, are likely misleading and of little use.

Biased Sample: Estimation by Incorporating Information in the Predictor Variable Importance Weights. For PRIDIT-FRE, we also repeat the analyses done in the "Estimation by Incorporating Information in the Predictor Variable Importance Weights" section adjusting for negative weights and using only variables of high importance to test the robustness. The results are shown in Tables 7 and 8. The count estimates remain the same.

In Tables 7 and 8, we can see again that PRIDIT-FRE estimates accurately the population fraud rates with small MSE and the count method estimates the biased sample fraud rates. PRIDIT-FRE is also robust whether or not the set of predictor variables is culled for importance and whether or not negative weights are adjusted.

TABLE 8
PRIDIT-FRE Results for a Population Having a Fraud Rate of 28.31 Percent Using Random Subsamples Having a Fraud Rate of 12.5 Percent, Adjusting for Negative Weights and Using Only Variables of High Importance (U.S. Data Set)

Subsample Size (% of the Biased Sample)	PRIDIT-FRE Estimation Estimate (Mean)	MSE	Variance	Median	Interquartile Range	Count Estimate
10	0.2809	0.0196	0.0196	0.2733	0.1723	0.1273
20	0.2845	0.0077	0.0077	0.2919	0.1127	0.1274
30	0.2956	0.0074	0.0072	0.2830	0.1068	0.1243
40	0.2919	0.0033	0.0032	0.2866	0.0661	0.1260
50	0.2919	0.0019	0.0018	0.2926	0.0464	0.1246
60	0.2816	0.0014	0.0014	0.2781	0.0500	0.1271
70	0.2910	0.0010	0.0009	0.2882	0.0399	0.1252
80	0.2833	0.0004	0.0004	0.2833	0.0303	0.1248
90	0.2874	0.0002	0.0002	0.2880	0.0174	0.1253
100	0.2867	0.0000	0.0000	0.2867	0.0000	0.1250

Note: This data set ("population") has 1,399 claims and the biased subsample has 800 claims. Negative weights are adjusted by reverse scoring the variable scales and a cutoff importance weight of 0.1 is used for selecting high importance predictors.

An Alternative Operational Definition of Fraud ("Hard Fraud") With Different Classification Heuristics

Due to the subjective nature of fraud assessment, it is very difficult to have a definitive classification of fraud unless court action is sought (cf. Brockett, Derrig, and Xia, 1998). Fraudulent activities vary in their severity, for example, exploiting ambiguity in the tax code versus "hard fraud" of taking a fabricated deduction. Because there is a continuum of fraud, an operational cutoff for fraud classification is often used to assess suspicious cases. In insurance, these operational definitions are usually based on claims experts' judgments, which can differ. Only the U.S. data set provides information for the robustness analysis presented below on different fraud definitions.[7]

Alternative Definition: Baseline Results and Estimation by Incorporating Information in Variable Importance Weights. In the U.S. data set, expert opinion for claim suspicion level is recorded on a 0- to 10-point scale. Different cutoff values may be chosen to operationalize the fraud definition. Following Viaene et al. (2002), a cutoff value of suspicion level of four was used previously in this article to identify fraudulent claims, which is a proxy for both "soft" and "hard" fraud. As a robustness check, an alternative operational definition is used here to classify any claim with a suspicion level of 7 and above as fraud, a proxy for "hard" fraud only. The resulting fraud

[7] In the Spanish data set, only a dichotomous classification of fraud is provided. As a result, we cannot use this data set to test for alternative definitions of fraud as what we did for the U.S. data set.

TABLE 9
PRIDIT-FRE Results Using Random Samples (U.S. Data Set Using Suspicion Score of 7 and Above as the Cutoff for Fraud Definition)

Sample Size (% of Data Set)	Estimate (Mean)	MSE	Variance	Median	Interquartile Range
10	0.0806	0.0077	0.0076	0.0811	0.1117
20	0.0936	0.0022	0.0022	0.0781	0.0492
30	0.0948	0.0013	0.0013	0.1068	0.0379
40	0.0879	0.0012	0.0012	0.1010	0.0674
50	0.0965	0.0009	0.0008	0.0952	0.0506
60	0.0810	0.0008	0.0008	0.0853	0.0362
70	0.0938	0.0002	0.0002	0.0937	0.0183
80	0.0918	0.0001	0.0001	0.0911	0.0119
90	0.0913	0.0001	0.0001	0.0877	0.0145

Note: This data set has 1,399 claims and the true fraud rate equals 8.79 percent.

TABLE 10
PRIDIT-FRE Results Using Random Samples Adjusting for Negative Weights and Using Only Variables of High Importance (U.S. Data Set Using Suspicion Score of 7 and Above as the Cutoff for Fraud Definition)

Sample Size (% of Data Set)	Estimate (Mean)	MSE	Variance	Median	Interquartile Range
10	0.0721	0.0069	0.0067	0.0859	0.0939
20	0.0852	0.0021	0.0021	0.0802	0.0332
30	0.0942	0.0005	0.0005	0.0915	0.0263
40	0.0824	0.0014	0.0014	0.0866	0.0425
50	0.0961	0.0006	0.0005	0.0906	0.0404
60	0.0826	0.0005	0.0005	0.0810	0.0239
70	0.0961	0.0002	0.0001	0.0967	0.0129
80	0.0904	0.0001	0.0001	0.0880	0.0109
90	0.0936	0.0002	0.0002	0.0893	0.0147

Note: This data set has 1,399 claims and the true fraud rate equals 8.79 percent. Negative weights are adjusted by reverse scoring the variable scales and a cutoff importance weight of 0.1 is used for selecting high importance predictors.

rate is 8.79 percent. As was done previously, we use 10 repetitions of the random sample selection for each sample size (10 percent to 90 percent of population with 10 percent increments), and present the estimation results in Tables 9 and 10, respectively for the baseline analyses (across all predictor variables) and analyses incorporating information in the variable importance weights (adjusting for negative weights and summing across only high importance variables).

As is evident from Tables 9 and 10, PRIDIT-FRE provides an accurate estimate of this newly defined fraud rate when compared to the true value, suggesting that the predictive accuracy of PRIDIT-FRE is robust to different operational definitions

TABLE 11
PRIDIT-FRE Results for a Population Having a Fraud Rate of 8.79 Percent Using Random Subsamples Having a Fraud Rate of 15 Percent (U.S. Data Set Using the Cutoff of Suspicion Score of 7 and Above for Fraud Definition)

Subsample Size (% of the Biased Sample)	PRIDIT-FRE Estimation Estimate (Mean)	MSE	Variance	Median	Interquartile Range	Count Estimate Count Estimate
10	0.0928	0.0150	0.0150	0.1099	0.1749	0.1451
20	0.0724	0.0087	0.0085	0.0802	0.0793	0.1484
30	0.0947	0.0019	0.0019	0.0970	0.0599	0.1502
40	0.0982	0.0023	0.0022	0.0996	0.0651	0.1523
50	0.0942	0.0011	0.0011	0.0911	0.0371	0.1488
60	0.0885	0.0012	0.0012	0.0893	0.0472	0.1493
70	0.0911	0.0005	0.0005	0.0918	0.0291	0.1505
80	0.0911	0.0003	0.0003	0.0901	0.0238	0.1493
90	0.0915	0.0001	0.0001	0.0904	0.0155	0.1501
100	0.0921	0.0000	0.0000	0.0921	0.0000	0.1500

Note: This data set ("population") has 1,399 claims and the biased sample has 800 claims.

TABLE 12
PRIDIT-FRE Results for a Population Having a Fraud Rate of 8.79 Percent Using Random Subsamples Having a Fraud Rate of 15 Percent, Adjusting for Negative Weights and Using Only Variables of High Importance (U.S. Data Set Using the Cutoff of Suspicion Score of 7 and Above for Fraud Definition)

Subsample Size (% of the Biased Sample)	PRIDIT-FRE Estimation Estimate (Mean)	MSE	Variance	Median	Interquartile Range	Count Estimate
10	0.0908	0.0196	0.0196	0.1078	0.1768	0.1451
20	0.0749	0.0082	0.0080	0.0873	0.0975	0.1484
30	0.0976	0.0025	0.0024	0.0990	0.0642	0.1502
40	0.0994	0.0025	0.0024	0.1015	0.0634	0.1523
50	0.0974	0.0013	0.0012	0.0966	0.0464	0.1488
60	0.0915	0.0013	0.0013	0.0925	0.0460	0.1493
70	0.0942	0.0006	0.0006	0.0938	0.0335	0.1505
80	0.0947	0.0004	0.0004	0.0938	0.0290	0.1493
90	0.0959	0.0002	0.0001	0.0955	0.0168	0.1501
100	0.0960	0.0001	0.0000	0.0960	0.0000	0.1500

Note: This data set ("population") has 1,399 claims and the biased sample has 800 claims. Negative weights are adjusted by reverse scoring the variable scales and a cutoff importance weight of 0.1 is used for selecting high-importance predictors.

of fraud. Again, PRIDIT-FRE retains its predictive accuracy whether or not the data are culled for predictor variable importance or adjusted for negative variable weights.

Alternative Definition: Estimation Results for the Biased Audited Sample Experiments. Similar to our biases sample experiments before, we create a biased sample of 800 claims with a fraud rate of 15 percent (versus an 8.79 percent fraud rate in the original data set) to test the performance of PRIDIT-FRE and the count estimation under this alternative fraud definition. Again, 10 such biased samples are created and 10 repetitions are used to select random subsamples of different sizes from each biased sample. The average estimation statistics for subsamples of different sizes are reported in Tables 11 and 12, respectively, with an unadjusted and adjusted set of predictor variables and weights as before.

Again, PRIDIT-FRE provides an accurate predictor for the population fraud rate in the presence of a biased audited sample whereas the count method does not. These results hold when using the entire predictor variable set or when using a smaller collection adjusting for negative weights and including only high importance predictors.

In summary, all the estimation results from our experiments illustrate the robustness of PRIDIT-FRE under various sources of inaccurate information, including a biased audited sample, an alternative fraud definition, an adjusted set of predictor variables, and a different application domain.[8] Additionally, by taking advantage of the current industry auditing practice that might result in a small and/or biased audited sample, PRIDIT-FRE can be easily incorporated into various industries' automated fraud-monitoring routine and can lead to more efficient resource allocation in fraud detection activities.

EXTENSION: ESTIMATING FRAUD RATE USING PRIDIT-FRE WITHOUT A SAMPLE OF AUDITED CLAIMS

Deriving an Alternative to Random Sampling

All the prior analyses have had the benefit of an audited sample of cases. This is not always available in practice (e.g., in emerging markets or new areas of fraud

[8] Analyses using another data set confirmed the ability of PRIDIT-FRE to accurately estimate the event rate in a different application domain. For this additional robustness check, we used the publicly available ADULT database (a census income classification data set) from UCI Machine Learning Repository (http://www.ics.uci.edu/~mlearn/MLSummary.html). The exact same estimation method is used with 10 repetitions of random selections for the baseline analyses. The results are consistent with analyses discussed in this article for insurance fraud data sets. Likewise, for this ADULT database, PRIDIT-FRE was robust when adjusting for negative weights and including only high importance predictor variables. We also did experiment on the ADULT database implementing PRIDIT-FRE using a biased audited sample, as done previously. Again, the results are consistently robust; that is, PRIDIT-FRE predicts accurately the population event rate rather than the biased sample event rate. Results are not reported to conserve space and are available upon request. All the robustness checks show that PRIDIT-FRE performs well under different variations and across completely different domain areas.

TABLE 13
PRIDIT Classifications for Extreme and Random Samples (Spanish Dataset)

Sample Size as a Percentage of the Data Set (x %)	PRIDIT Classification Accuracy for the Extreme Sample (of Top $x/2$ % and Bottom $x/2$ %)	PRIDIT Classification Accuracy for a Random Sample (of x %)[a]
$x = 10$	0.9347	0.6350
$x = 20$	0.8697	0.6667
$x = 30$	0.8581	0.6578
$x = 40$	0.7729	0.6579
$x = 50$	0.7462	0.6483

Note: Extreme samples are created by selecting a certain equal number of cases with the highest and lowest PRIDIT suspicion scores. For example, an extreme sample of size $x = 10$ percent of the data set consists of claims ranked as the highest 5 percent and lowest 5 percent by the PRIDIT suspicion score. Classification accuracy is obtained by comparing PRIDIT classification of fraud with the true fraud classification. This data set has 1995 claims.
[a]Note that variation of accuracies seen for random samples here is because that random sampling is only done once for comparison purposes as PRIDIT analysis cannot be repeated multiple times.

there may be no databases of audited samples available). Thus, our final analyses presented in this section extend PRIDIT-FRE to an unsupervised application that does not use an audited sample. Since a classification into fraud/non-fraud classes can be obtained from the PRIDIT claim suspicion scores, we can use the PRIDIT-based class assignment of claims as an alternative to the audited true fraud classification to calculate average variable scores for each class in Equation (3). As the U.S. data set does not contain a binary fraud classification variable but rather requires the selection of a certain cutoff value to define the fraud classification, it will be difficult to evaluate the performance of PRIDIT-FRE in the unsupervised application (as no supervised information would be available to determine the cutoff). Therefore, we use only the Spanish insurance claims data set to implement this extension.

Table 13 presents the accuracy of PRIDIT classification relative to the true fraud classification for "extreme" and random samples of claims for the Spanish data set. To conduct the classification analyses, we first run PRIDIT to derive an overall suspicion score for each claim. All claims are then ranked by the suspicion scores from the highest to the lowest to form samples of claims with "extreme" scores. Column 2 in Table 13 presents the classification accuracies for such characterized subsets, from highest and lowest 5 percent to highest and lowest 25 percent. For comparison purposes, classification accuracies are also assessed for random subsets of 10 percent, 20 percent, 30 percent, 40 percent, and 50 percent of the data set as shown in column 3. PRIDIT classifications are obtained as in Brockett et al. (2002) and Ai, Brockett, and Golden (2009). A brief overview of the PRIDIT method was also given in the "Overview of the PRIDIT Method Notations" section.

Table 13 shows that the classification accuracy of PRIDIT, obtained by comparing to the true fraud classification is strong. It is best for cases with the highest and lowest PRIDIT suspicion scores, that is, at the extremes of the PRIDIT score range. For example, accuracy is 93.47 percent when using the sample of cases with highest and lowest 5 percent PRIDIT scores. This result is very strong especially given that no supervised information (the true fraud classification in an audited sample) is used. Moreover, classification accuracy of PRIDIT is general better for extreme samples of the data than random samples across different sample sizes. These results are used next to design a PRIDIT score-based sampling approach to implement the PRIDIT-FRE estimation in the unsupervised application.

Determining Fraud Rate Without an Audited Sample

We now turn to the application of PRIDIT-FRE to determine fraud rate when an audited sample is not available. Motivated by the results shown in Table 13 that predictability is best for cases with the highest and lowest 5 percent PRIDIT scores, our analysis focuses on this extreme sample. After obtaining predictor variables, calculating PRIDIT variable scores and overall claim suspicion scores, the claims are ranked and the *unaudited* extreme sample is selected. In applying Equation (3), the fraud class is defined as the lowest 5 percent extreme subset (in lieu of a definition using the true fraud classification available from an audited sample) and the non-fraud class is defined as the highest 5 percent extreme subset. Average predictor variable scores for each class are thus calculated and a fraud rate estimate is obtained by Equation (3).

We implement this extension on the Spanish insurance data. The fraud rate estimate using the 10 percent extreme sample (highest and lowest 5 percent subset) is 50.79 percent, where the true fraud rate is 49.97 percent. We can see that PRIDIT-FRE's predictive accuracy is very strong for this data set. This is very encouraging since the estimation only uses information from the fraud predictor variables, which can be selected and gathered from the insurance application and claims-adjusting routines already in place. No prior information on the fraud classification from an audited sample is required to estimate the fraud rate. This implies potentially large savings in the auditing cost (e.g., expertise, time, and money) for the purpose of fraud rate estimation normally incurred by the original PRIDIT-FRE method or by any other supervised learning based estimation methods, and much more flexibility in determining the frequency and breadth of fraud monitoring activities. As the normal claims-adjusting process develops, more legitimate and fraudulent claims will be uncovered. They can be combined with the current PRIDIT-based extreme subsets to update and further improve the fraud rate estimate.

CONCLUSIONS AND IMPLICATIONS

Increasingly, fraud risk assessment has become an important ongoing, standard business practice for firms and government agencies. In insurance, paying fraudulent claims drives up costs, impacting market competitiveness. Fraud rate estimation serves both as a precursor to and an assessment of mitigation and is becoming part of routine record keeping. However, currently there is a lack of rigorously developed fraud rate estimation method. The ostensibly "standard" practice of using a "count the fraud cases" method with an audited sample is costly and more importantly not reliable when the sample contains a bias. This article offers an effective and less costly alternative approach.

We introduce a new fraud rate estimation method, PRIDIT-FRE, developed based on an unsupervised fraud detection method PRIDIT. Different from a fraud detection method that assesses individual claims for fraud suspicion, PRIDIT-FRE estimates the fraud rate of the population of claim files. In achieving this, PRIDIT-FRE relies on easily obtained fraud predictor variables and a small audited sample. PRIDIT-FRE can even be implemented with only an unaudited sample when there is a new fraud definition or an emerging market situation where fraud detection is a new practice. In either implementation, PRIDIT-FRE is relatively free of the likely upward bias introduced by an audited sample naturally obtained from current industry practice (e.g., bias can result from the fact that claims are often audited and included in the sample because of some fraud suspicion in the first place when using the popular "red-flag" auditing policy). This makes it possible to include the fraud rate estimation in the companies' routine business practices of fraud management and also entails cost savings.

Using a Spanish physical damage automobile insurance claims data set and a U.S. personal injury protection automobile insurance data set as illustrations, we show that PRIDIT-FRE produces consistent and accurate fraud rate estimates. Further, we demonstrate that PRIDIT-FRE produces very reliable estimates of the population fraud rate even when only a very small and largely biased sample is available for use.

PRIDIT-FRE's robustness was checked across multiple dimensions: (1) robustness to different populations (countries, types of claims, and fraud rates), (2) robustness when there is a strong possibility of a biased fraud rate in the sample, (3) robustness across a variety of predictor variables (different for Spain and U.S. applications, also a set adjusted for predictor variable importance weights), (4) robustness for different definitions of fraud, and (5) robustness across different application domains (insurance claims and census income classification). The predictive accuracy of PRIDIT-FRE was strong for all of these robustness tests.

In summary, the fraud rate estimation methodology PRIDIT-FRE is relatively straightforward to implement, can be routinized for ongoing fraud monitoring and is cost efficient. Further, PRIDIT-FRE is a valuable robust alternative allowing for current industry practices of auditing more suspicious cases, when a supervised learning based "count the fraud cases" method estimating within such a naturally generated biased audited sample will fail completely. Future research should focus on further developing the PRIDIT-FRE method to exploit the full potential of its unsupervised nature and cost-related benefits.

APPENDIX: PROOF OF CONSISTENCY OF THE FRAUD RATE ESTIMATOR

Theorem: *The fraud rate estimator $\hat{\theta}$ in Equation (3) is a consistent estimator of the fraud rate θ.*

Proof: Let $F(x)$ be the distribution function of some predictor variable t and recall that $F_1(x)$ and $G_2(x)$ are, respectively, the distribution functions of variable t for the fraud class and the non-fraud class. Let $\hat{F}(x)$ and $\hat{F}_1(x)$ denote the corresponding empirical distributions. As in Ai, Brockett, and Golden (2009), for the fraud class and variable t,

$$\bar{B}_{1t} = \int_{-\infty}^{\infty} \left[2\hat{F}(x) - 1\right] d\hat{F}_1(x).$$

By Glivenko–Cantelli theorem, this is uniformly close to

$$\int_{-\infty}^{\infty} [2F(x) - 1]\, d\hat{F}_1(x).$$

By the Helly–Bray theorem and noting $F(x) = \theta F_1(x) + (1-\theta)G_2(x) = F_1(x) - (1-\theta)\Delta(x)$, we have

$$\int_{-\infty}^{\infty} [2F(x) - 1] d\hat{F}_1(x) \to \int_{-\infty}^{\infty} [2F(x) - 1] dF_1(x)$$
$$= 2(\theta - 1)\int_{-\infty}^{\infty} \Delta(x) dF_1(x) = (\theta - 1)\, A_t.$$

Similarly, $\bar{B}_{2t} \to 2\theta \int_{-\infty}^{\infty} \Delta(x) dF_1(x) = \theta A_t$. Therefore, by Slutsky's theorem, we have $\hat{\theta} = \frac{\sum_t \bar{B}_{2t}}{\sum_t (\bar{B}_{2t} - \bar{B}_{1t})} \to \frac{\sum_t \theta A_t}{\sum_t (\theta A_t - (\theta - 1) A_t)} = \theta.$ Q.E.D.

REFERENCES

Ai, J., P. L. Brockett, and L. L. Golden, 2009, Assessing Consumer Fraud Risk in Insurance Claims With Discrete and Continuous Data, *North American Actuarial Journal*, 13: 438-458.

Artís, M., M. Ayuso, and M. Guillén, 2002, Detection of Automobile Insurance Fraud With Discrete Choice Models and Misclassified Claims, *Journal of Risk and Insurance*, 69: 325-340.

Baker, K., and E. Herbert, 1992, *Fighting the Hidden Crime: A National Agenda to Combat Insurance Fraud* (New York: Insurance Information Institute).

Bolton, R. J., and D. J. Hand, 2002, Statistical Fraud Detection: A Review, *Statistical Science*, 17: 235-249.

Boyer, M., 2000, Centralizing Insurance Fraud Investigation, *Geneva Papers on Risk and Insurance Theory*, 25: 159-178.

Boyer, M., 2007, Resistance (to Fraud) Is Futile, *Journal of Risk and Insurance*, 74: 461-492.

Brockett, P. L., R. A. Derrig, L. L. Golden, A. Levine, and M. Alpert, 2002, Fraud Classification Using Principal Component Analysis of RIDITs, *Journal of Risk and Insurance*, 69: 341-372.

Brockett, P. L., R. A. Derrig, and X. Xia, 1998, Using Kohonen's Self-Organizing Feature Map to Uncover Automobile Bodily Injury Claims Fraud, *Journal of Risk and Insurance*, 65: 245-274.

Brockett, P. L., and R. C. Witt, 1982, The Underwriting Risk and Return Paradox Revisited, *Journal of Risk and Insurance*, 49: 621-627.

Caron, L., and G. Dionne, 1999, Insurance Fraud Estimation: More Evidence From Quebec Automobile Insurance Industry, in: G. Dionne and C. Laberge-Nadeau, eds., *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation* (Boston, MA: Kluwer).

Caudill, S., M. Ayuso, and M. Guillen, 2005, Fraud Detection Using a Multinomial Logit Model With Missing Information, *Journal of Risk and Insurance*, 72: 539-550.

Crocker, K., and S. Tennyson, 2002, Insurance Fraud and Optimal Claims Settlement Strategies, *Journal of Law and Economics*, 45: 469-508.

Derrig, R. A., and K. M. Ostaszewski, 1995, Fuzzy Techniques of Pattern Recognition in Risk and Claim Classification, *Journal of Risk and Insurance*, 62: 447-482.

Dionne, G., and R. Gagné, 2001, Deductible Contracts Against Fraudulent Claims: Evidence From Automobile Insurance, *Review of Economics and Statistics*, 83: 290-301.

Dionne, G., and R. Gagné, 2002, Replacement Cost Endorsement and Opportunistic Fraud in Automobile Insurance, *Journal of Risk and Uncertainty*, 24: 213-230.

Dionne, G., F. Giuliano, and P. Picard, 2009, Optimal Auditing With Scoring: Theory and Application to Insurance Fraud, *Management Science*, 55: 58-70.

Durtschi, C., W. Hillison, and C. Pacini, 2004, The Effective Use of Benford's Law to Assist Detecting Fraud in Accounting Data, *Journal of Forensic Accounting*, 5: 17-34.

Heron, R. A., and E. Lie, 2009, What Fraction of Stock Option Grants to Top Executives Have Been Backdated or Manipulated? *Management Science*, 55: 513-525.

Insurance Research Council (IRC), 2008, World Wide Web: http://www.ircweb.org/News/IRC_Fraud_NR.pdf.

Mookherjee, D., and I. Png, 1989, Optimal Auditing, Insurance and Redistribution, *Quarterly Journal of Economics*, 104: 399-415.

Picard, P., 1996, Auditing Claims in the Insurance Market With Fraud: The Credibility Issue, *Journal of Public Economics*, 63: 27-56.

Picard, P., 2000, Economic Analysis of Insurance Fraud, in: G. Dionne, ed., *Handbook of Insurance* (Norwell: Kluwer Academic Press), pp. 315-360.

Pinquet, J., M. Ayuso, and M. Guillen, 2007, Selection Bias and Auditing Policies for Insurance Claims, *Journal of Risk and Insurance*, 74: 425-440.

Schiller, J., 2006, The Impact of Insurance Fraud Detection Systems, *Journal of Risk and Insurance*, 73: 421-438.

Tennyson, S., 2008, Moral, Social, and Economic Dimensions of Insurance Claims Fraud, *Social Research*, 75: 1181-1204.

Tennyson, S., and P. Salsas-Fom, 2002, Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives, *Journal of Risk and Insurance*, 69: 289-308.

Townsend, R., 1979, Optimal Contracts and Competitive Markets With Costly State Verification, *Journal of Economic Theory*, 21: 265-293.

Viaene, S., M. Ayuso, M. Guillen, D. V. Gheel, and G. Dedene, 2007, Strategies for Detecting Fraudulent Claims in the Automobile Insurance Industry, *European Journal of Operational Research*, 176: 565-583.

Viaene, S., and G. Dedene, 2004, Insurance Fraud: Issues and Challenges, *The Geneva Papers on Risk and Insurance*, 29: 313-333.

Viaene, S., R. A. Derrig, B. Baesens, and G. Dedene, 2002, A Comparison of State-of-the-Art Classification Techniques for Expert Automobile Insurance Claim Fraud Detection, *Journal of Risk and Insurance*, 69: 373-421.

Yim, A., 2009, Efficient Committed Budget for Implementing Target Audit Probability for Many Inspectees, *Management Science*, 55: 2000-2018.