

DOI: 10.7819/rbgn.v16i52.1555

SUBJECT AREA: ACCOUNTING AND CONTROLLING

Accounting Fraud: an estimation of detection probability

Fraudes Contábeis: uma estimativa da probabilidade de detecção

Fraudes contables: una estimación de la probabilidad de detección

Artur Filipe Ewald Wuerges¹
José Alonso Borba²

Received on February 20, 2013 / Approved on October 1, 2014

Responsible Editor: Ivam Ricardo Peleias, Dr.

Evaluation Process: *Double Blind Review*

ABSTRACT

Financial statement fraud (FSF) is costly for investors and can damage the credibility of the audit profession. To prevent and detect fraud, it is helpful to know its causes. The binary choice models (e.g. logit and probit) commonly used in the extant literature, however, fail to account for undetected cases of fraud and thus present unreliable hypotheses tests. Using a sample of 118 companies accused of fraud by the Securities and Exchange Commission (SEC), we estimated a logit model that corrects the problems arising from undetected frauds in U.S. companies. To avoid multicollinearity problems, we extracted seven factors from 28 variables using the principal factors method. Our results indicate that only 1.43 percent of the instances of FSF were publicized by the SEC. Of the six significant variables included in the traditional, uncorrected logit model, three were found to be actually non-significant in the corrected model. The likelihood of FSF is 5.12

times higher when the firm's auditor issues an adverse or qualified report.

Keywords: Accounting fraud. AAER. Misclassification. Logit. Factor analysis.

RESUMO

Fraudes nas demonstrações financeiras (FDF) custam caro para os investidores e podem prejudicar a credibilidade dos auditores. Para prevenir e detectar fraudes, é útil conhecer suas causas. Os modelos de escolha binária (por exemplo, logit e probit), frequentemente utilizados na literatura, porém, não levam em consideração os casos de fraudes não detectados e, portanto, apresentam testes de hipóteses pouco confiáveis. Usando uma amostra de 118 empresas acusadas de fraude pela Comissão de Valores Mobiliários dos Estados Unidos (Securities and Exchange Commission, SEC), estimamos um modelo logit que corrige os problemas oriundos de fraudes não detectadas em empresas dos Estados Unidos.

1. Master in Management from Federal University of Santa Catarina (UFSC). [artur.wuerges@uffs.edu.br]

2. Doctor in Accounting from University of São Paulo (USP). [jalonso@cse.ufsc.br]

Authors' address: Rua Major Antônio Cardoso, 590, CEP: 97900-000 – Cerro Largo – RS – Brazil

Para evitar problemas de multicolinearidade, extraímos sete fatores a partir de 28 variáveis, usando o método dos componentes principais. Nossos resultados indicam que apenas 1,43% dos casos de FDF foram divulgados pela SEC. Das sete variáveis significativas incluídas em um modelo logit tradicional e não corrigido, três na realidade não foram consideradas significativas em um modelo corrigido. A probabilidade de FDF é 5,12 vezes maior quando o auditor da empresa emite um parecer adverso ou com ressalvas.

Palavras-chave: Fraude contábil. AAER. Erros de classificação. Logit. Análise fatorial.

RESUMEN

El fraude en los estados financieros (FEF) es costoso para los inversionistas y pueden minar la credibilidad de los auditores. A fin de prevenir y detectar el fraude, es útil conocer sus causas. Sin embargo, los modelos de elección binaria (logit y probit, por ejemplo) a menudo utilizados en la literatura, no tienen en cuenta los casos de fraudes detectados y consecuentemente presentan pruebas de hipótesis poco fiables. Utilizando una muestra de 118 compañías acusadas de fraude por la Comisión de Bolsa y Valores de EE.UU. (Securities and Exchange Commission, SEC), hemos estimado un modelo logit que corrige los problemas derivados de los fraudes no detectados en las compañías estadounidenses. Para evitar problemas de multicolinealidad, hemos extraído siete factores de 28 variables, utilizando el método de componentes principales. Nuestros resultados indican que sólo el 1,43% de los casos de FEF se han dado a conocer por la SEC. De las siete variables significativas incluídas en un modelo logit tradicional y no corregido, tres en efecto no fueron consideradas significativas en un modelo corregido. La probabilidad de FEF es 5,12 veces mayor cuando el auditor de la compañía emite una opinión adversa o con reservas.

Palabras clave: Fraude contable. AAER. Errores de clasificación. Logit. Análisis factorial.

1 INTRODUCTION

In modern corporations, it is common to adopt a structure that separates ownership from management. Although the shareholders have voting rights, the daily management is actually done by professional executives appointed by the board of directors (that are themselves elected by the shareholders). This situation opens the way for conflicts of interest: The decisions taken by executives are not always the ones the shareholders would take if they were managing the company (JENSEN; MECKLING, 1976). For example, the CEO might buy a first class flight ticket, whereas a shareholder would probably prefer to pay a cheaper airfare.

This misalignment of interests will usually incur in agency costs for the company. Some of these agency costs are monitoring costs: the shareholders are willing to invest in the development of a monitoring system capable of avoiding abuse by managers. The financial statements, produced in accordance with some financial reporting standards (e.g. IFRS), are part of this monitoring system. Even though they are a cost (ultimately paid by the shareholders), they allow investors access to information about the quality of the management of the company. In order to assure investors that adequate reporting standards were actually followed, the company will usually engage an independent auditor. In fact, financial statement users commonly rely on auditors to prevent and detect financial statement fraud and errors (BEST; BUCKBY; TAN, 2001; LEE; GLOECK; PALANIAPPAN, 2007), although “it is management’s responsibility to design and implement programs and controls to prevent, deter, and detect fraud.” (AICPA, 2002, p. 169)

Even with the complex monitoring systems adopted in publicly traded companies, financial statement fraud (FSF) still haunts auditors and users of financial statements. Some high-profile cases of accounting fraud became known at the beginning of the first decade of the 21st century, shattering the trust of the investors, inasmuch as the costs of frauds are in

last instance paid by the stockholders when the stock prices plummet after the announcement of fraud charges by the Securities and Exchange Commission (GERETY; LEHN, 1997). Some big companies now associated to fraud in the U.S. are Adelphia, Enron, Tyco and WorldCom. This surge of frauds led to a legislative reform known as Sarbanes-Oxley Act (SOX), which established new standards of transparency for public companies. In spite of these measures, frauds still occur (even though their number apparently decreased after 2002), and therefore are still a relevant theme for academic research.

Failure to detect or prevent financial statement fraud can damage the reputation and the credibility of the audit profession (CHUI; PIKE, 2011). In fact, some argue that fraud is the Achilles heel of the audit profession, at least in part because risk-based auditing is predictable (JAMAL, 2008). Using past cases of fraud to build models to estimate the risk of fraud does not help auditors to become unpredictable, but may help them to understand the causes of financial statement fraud. Regression analysis (i.e. logit and probit models) allows the test of hypotheses about the influence of certain variables on the occurrence of fraud.

One of the problems of this approach is that there are cases of fraud that still have not been discovered, and some of them probably will never be. It is also possible that the SEC chooses to focus on high-profile fraud cases, and therefore some cases are actually detected to some degree, but not fully investigated and reported in AAERs. Finally, it is also possible that the commission suspects some financial statements are intentionally misstated but is unable to present evidence to support its suspicion. It would not be surprising, then, if just a small part of accounting frauds is discovered, fully investigated and disclosed to the public. In a regression analysis, this generates misclassification problems in the dependent variable. There are companies incorrectly excluded from the list of cases of fraud. These errors harm the model, biasing the estimates of the parameters. It is therefore possible that the hypotheses tests are affected, making the conclusions misleading.

Hausman, Abrevaya, and Scott-Morton (1998) developed a method capable of minimizing this problem and also of estimating the probability that a case of fraud is not discovered. This study examines the results of such method applied to financial statement fraud, to assess whether its results are different from the ones obtained with an uncorrected logit model.

This paper has three objectives. Our first objective was to estimate the percentage of unreported frauds in U.S. companies. Doing so, we can analyze how effective the SEC is in detecting and repressing financial statement fraud. Our second objective was to check if the results obtained using a traditional logit model are qualitatively different from those obtained using a logit model with the misclassification parameter. Our last objective was to obtain reliable hypotheses tests regarding some variables that might be related to financial statement fraud.

2 THEORETICAL BACKGROUND AND VARIABLES

Fraud, in a broad sense, is defined as the conscious distortion of truth or concealment of material fact with the objective of inducing other people to act to the detriment of their own interests (PEDNEAULT, 2009). When an executive intentionally distort financial statements, he or she is also committing fraud. There is a conscious act that induces other people to act to the detriment of their own interests, like paying a high price for common stock that is actually worth much less. These cases are known as financial statement fraud (ALBRECHT, W. S. et al., 2009).

It is important to add that financial statement frauds are perpetrated by violating Generally Accepted Accounting Principles (GAAP). If there is no violation of GAAP, then it is a case of earnings management, which is also harmful to investors but resorts only to legal means (DECHOW; SKINNER, 2000).

To estimate a logit model for fraud detection, it is necessary to identify a set of

variables that explain the occurrence of fraud. There are concepts and theories that explain white-collar crime and some of them can also be employed as a starting point to understand the causes of accounting fraud, thereby serving as a guide for the selection and organization of independent variables. It is common to use the fraud triangle, originally proposed by Donald R. Cressey to explain the occurrence of white-collar crime. This author believes that these crimes occur when three factors are present: pressure (financial need), opportunity and rationalization. In the early 1980's, Steve Albrecht adapted this theory to the study of financial statement fraud (CHOO; TAN, 2007). We classify the variables used in this study in these three groups.

It must be noted that, when estimating a logit model, an observation is dropped if there is a value missing for any variable. To avoid working with a small sample, we excluded all the variables with fewer than 20,000 observations. These variables were collected, but are not used or mentioned in this paper.

2.1 Pressure

The pressure (financial need) is related to the situation of the company and its managers. It is expected that companies that have been unable to attain the expectations of the market are more susceptible to fraud, since the managers are under pressure to perform well. For this reason, indicators of performance and financial security can be used for fraud detection, in association with other evidence. Enron, for example, was a highly leveraged company that required its executives to obtain (and report) large profits in order to be able to pay interests and fulfill contractual bonds (CHOO; TAN, 2007).

To capture the elements of pressure in the model, we used several different indicators. One of them is the Altman's Z score (**zscore**), a proxy for risk of financial distress often used in financial statement fraud research (ALTMAN, 1968). We expect companies with higher Z scores to be in better financial position and therefore less likely to resort to fraud, although previous research

shows it is negatively related to the likelihood of fraud in Greek companies (KIRKOS; SPATHIS; MANOLOPOULOS, 2007; SPATHIS, 2002), but statistically unrelated to fraud in U.S. firms (ERICKSON; HANLON; MAYDEW, 2006; FANNING; COGGER, 1998; SUMMERS; SWEENEY, 1998). We also used several other variables that attempt to measure the companies' ability to cover their liabilities, since companies that have trouble paying their debts (and therefore have financial problems) may recourse to FSF in order to obtain financial resources, either in the capital markets or through loans. One of these variables is the cash scaled by total assets (**CASHTA**), proposed by Gaganis (2009), who found it to be negatively related to fraud. We also adopted the cash change scaled by total assets (Δ CASHTA). Another variable related to liquidity is the current assets divided by current liabilities (**CACL**), also known as current ratio. A **CACL** lower than 1 suggests that the company does not have enough assets readily available to pay its short term debt and therefore signals possible insolvency problems. We expect this variable and its change (Δ CACL) to be negatively related to fraud, although there are previous studies that found this variable to be statistically unrelated to FSF (BENEISH, 1999a; KIRKOS; SPATHIS; MANOLOPOULOS, 2007). Another variable is the working capital (i.e. current assets minus current liabilities) divided by total assets (**WCIA**), used by Beneish (1997, 1999a, 1999b), Fanning and Cogger (1998), Spathis (2002), Kirkos, Spathis, and Manolopoulos (2007), and Gaganis (2009). We expect this variable and its change (Δ WCIA) to be negatively related to fraud.

Fanning and Cogger (1998) and Kaminski, Wetzel, and Guan (2004) used the ratio formed by the accounts receivables divided by the total assets (**ARTA**). We believe that companies with high **ARTA** are less financially secure, and therefore present a higher likelihood of fraud. A variable often used to measure financial security is **TLTA**, obtained by total liabilities scaled by total assets (BENEISH, 1997, 1999a, 1999b; CRUTCHLEY; JENSEN, M. R. H.; MARSHALL, 2007; ERICKSON; HANLON;

MAYDEW, 2006; ETTREDGE et al., 2008; JOHNSON; RYAN; TIAN, 2008; WANG, 2011). We also used the change in this variable ($\Delta TLTA$). Very similar measures of leverage, used by Fanning and Cogger (1998), can be obtained by substituting total equity for total assets (**TLTE and $\Delta TLTE$**). Another indicator of financial security is the ratio of sales to accounts receivables (**SALAR**), used by Summers and Sweeney (1998), Spathis (2002), Kirkos, Spathis, and Manolopoulos (2007), and Skousen and Wright (2008). We expect it to be negatively related to the likelihood of fraud. Kaminski, Wetzel, and Guan (2004) used yet another ratio for financial security, given by interest expenses divided by total liabilities (**IETL**).

Another variable that can be used is **FATA**, the ratio of fixed assets to total assets (KIRKOS; SPATHIS; MANOLOPOULOS, 2007). It is the fixed assets (e.g. equipment and factories) that generate revenues for the company, and therefore we expect that companies with a large value invested in fixed assets are less likely to engage in fraud, since they are in better position to generate real revenue in the future. According to Kirkos, Spathis, and Manolopoulos, **FATA** is unrelated to fraud in Greek companies. We also used the change in this variable ($\Delta FATA$). Another variable, proposed by Gaganis (2009) and negatively correlated to **FATA**, is obtained dividing the current assets by the total assets (**CATA**). Its change was also used in this study ($\Delta CATA$). Gaganis found **CATA** to be significantly lower in companies involved in FSF.

Some authors suggest using the cost of goods sold divided by the sales (KAMINSKI; WETZEL; GUAN, 2004). We expect this variable (**COGSAL**) and its change ($\Delta COGSAL$) to be positively related to the likelihood of fraud, since an increase in the cost of goods sold can signal a decrease in the company's ability to compete.

Spathis (2002) and Kirkos, Spathis, and Manolopoulos (2007) used in their models a variable obtained by dividing the net profit by the sales. We used EBIT (earnings before interest and taxes) divided by sales (**EBITSAL**)

as a predictor of FSF. The rationale behind this variable is that a decrease in profitability creates pressure that might cause the company to recourse to fraud. A similar variable is EBIT scaled by total assets (**EBITTA**). We also used the change in this variable ($\Delta EBITTA$). Another measure of profitability is the net profit scaled by total assets (**ROA**) and its change (ΔROA). While building variables such as **EBITTA** and **ROA**, it should be noted that variables obtained from the income statement (such as net profit or EBIT) are attributed to the entire period, whereas variables obtained from the balance sheet (such as total assets) are attributed to a given point in time. To make the denominator comparable to the numerator, we used the average total assets in the relevant period of time to generate the variables **EBITTA** and **ROA**, as well as their changes.

The extant literature also presents variables based in the inventory value. The presence of too much inventory might signal that the company is unable to sustain the expected sales levels or is simply inefficient. Fanning and Cogger (1998) and Gaganis (2009), using inventory divided by sales (**INVSAL**), found evidence supporting this hypothesis. On the other hand, in Greek companies, this variable seems to be statistically not significant (KIRKOS; SPATHIS; MANOLOPOULOS, 2007; SPATHIS, 2002). Summers and Sweeney (1998) argue that financial statement fraud is related to the growth in inventory ($\Delta INVSAL$). We use this variable because some industries might have usually high inventory levels. Finally, some authors used inventory divided by total assets (**INVTA**), but found it to be statistically not significant (FANNING; COGGER, 1998; GAGANIS, 2009; KIRKOS; SPATHIS; MANOLOPOULOS, 2007).

2.2 Opportunity

The second element of the fraud triangle is the opportunity, which should not be understood only as lack of regulation or bad corporate governance. According to Choo and Tan (2007), the intense emphasis on monetary success by the American society may cause managers to

actively seek means to bypass the institutional mechanism created to repress fraud. This means that even when opportunity is not present, in some cases managers will create them. Pressure for success generates opportunities for criminal behavior. The managers of Cendant, a company charged with fraud, kept a yearly spreadsheet with “opportunities” available to inflate operational revenue, and also listed the values that should be obtained in each of these opportunities.

In our model, the variable that represents the opportunity to fraud is a dummy for auditor change (ETTREDGE et al., 2008; FANNING; COGGER, 1998; SKOUSEN; WRIGHT, 2008), here abbreviated as **AUDCH**. According to Summers and Sweeney (1998), it is possible that a company that engages in fraud will change its auditor to diminish the likelihood of getting caught.

2.3 Rationalization

Managers involved in fraud (and white-collar criminals in general) “adjust’ the symbolic construction of their behavior to conform to generalized social expectations,” in a process often referred to as rationalization (COLEMAN, 1987, p. 410). They might, for instance, claim that their actions were necessary for the company to survive and avoid layoffs. Or, in highly competitive industries, managers might contend that fraudulent acts are widespread, effectively constructing them as something natural and expected.

For Skousen and Wright (2008), the accruals level is representative of the management’s way of making decisions about financial statements. When earnings management is seen as normal, more serious actions might be the next step. Since excessive use of accruals often motivate modified audit reports (FRANCIS; KRISHNAN, 1999), we used a dummy variable (**UQUAL**) that is equal to 1 if the audit report presents an unqualified (clean) opinion. The opinion is unqualified when the financial statements reflect no unresolvable restrictions and the auditor has no significant exceptions. Unqualified opinions

with additional language are coded as 0, since they might reflect excessive use of accruals. In any other case (e.g. qualified or adverse opinion), the variable is equal to 0.

2.4 Similar studies

There are several papers that used probit and logit models to estimate the likelihood of FSF or identify the variables that influence this likelihood. Some of them also used AAERs to build the dependent variable (ABBOTT; PARK; PARKER, 2000; BRAZEL; JONES; ZIMBELMAN, 2009; CRUTCHLEY; JENSEN, M. R. H.; MARSHALL, 2007; ERICKSON; HANLON; MAYDEW, 2006; ETTREDGE et al., 2008; JOHNSON; RYAN; TIAN, 2008; LENNOX; PITTMAN, 2010; MILLER, 2006; SKOUSEN; WRIGHT, 2008; PERSONS, 1995). Other authors used the cases of fraud publicized by the press, either alone or jointly with AAERs (BEASLEY, 1996; BENEISH, 1997; LEE; INGRAM; HOWARD, 1999; SUMMERS; SWEENEY, 1998).

Articles using probit and logit often aim at checking the impact of a given variable on the occurrence of fraud. Erickson, Hanlon, and Maydew (2006), for example, studied the effect of stock option on fraud. The results of this article indicate that, contrary to what is often believed, there is no consistent evidence that stock-based incentives contribute to the occurrence of accounting fraud. These results were confirmed by the papers of Gerety and Lehn (1997) and Crutchley, Jensen, and Marshall (2007), but not by Erickson, Hanlon, and Maydew (2004). More recently, Feng et al. (2011) found evidence supporting the hypothesis that CFOs get involved in fraud because they succumb to pressure from CEOs, and not because they want to increase their income from equity incentives.

There are also papers that employed methodologies to estimate the probability that a case of fraud is not detected (WANG, 2011). This author used a method proposed by Lin (1980), in

which the probability of fraud and the probability of fraud detection are estimated separately. The model used by Wang requires the identification of a set of variables that influence the likelihood of fraud detection. Our paper, on the other hand, estimated the unconditional likelihood of fraud detection (without the use of an additional set of independent variables). Some papers used artificial intelligence techniques to detect frauds. It is common to use neural networks for this purpose (CHOI; GREEN, 1997; FANNING; COGGER, 1998; OGUT et al., 2009). Some studies employed fuzzy logic (DESHMUKH; TALLURU, 1998), whereas others adopted hybrid methods like fuzzy neural networks (LIN; HWANG; BECKER, 2003). At last, there are articles that used several different fraud detection methods with the intent of comparing them (KIRKOS; SPATHIS; MANOLOPOULOS, 2007; GAGANIS, 2009).

One of the shortcomings of studies using artificial intelligence is the difficulty in testing hypotheses. It is not possible, for example, to test whether the likelihood of fraud is related to executive compensation. On the other hand, several studies concluded that artificial intelligence methods can identify the occurrence of fraud with higher accuracy when compared to probit and logit models (LIN; HWANG; BECKER, 2003; OGUT et al., 2009). The choice of tools depends on the objectives of the researcher. In this paper, since our objective is to estimate the likelihood of a fraud being undetected and test the hypothesis that this likelihood is different from zero, we have chosen a logit model.

3 METHOD

This article presents an empirical model. The data regarding financial statements were obtained in Compustat. The following sections describe how the AAERs published by the SEC were used to identify the companies charged with fraud in the U.S. and explain the novel method used to estimate the likelihood of a fraud being not detected.

3.1 Data

The companies accused of FSF were identified by an analysis of the Accounting and Auditing Enforcement Releases published by the SEC from 1 January 1998 to 23 November 2010. All the reports were read in search of material misstatements.

If the company was accused of fraud, we carefully read the release to identify the first fraud year – that is, the first year in which a materially false statement was published. There was care in checking the fiscal years of the accused companies. Compustat adopts a simple rule: If the fiscal year ends between 1 January and 31 May, then it is considered as being the previous calendar year. For example, a statement closed as of 31 March 2000 is stored in the database as being of the year 1999. Therefore, if a fraud was first identified in this statement, then the first fraud year is 1999. In other words, we made an effort to match the list of frauds and their years to the data provided by Compustat.

We removed all the occurrences of violation of the Foreign Corrupt Practices Act (FCPA) from the sample. The FCPA prohibits the practice of bribing foreign officials. Since bribery is illegal, the money used for such purposes does not pass through official accounting; the company might recourse to some kind of financial report violation in order to hide the destination of these resources. This can materially affect the financial statements of the company. Unfortunately, not every AAER issued because of FCPA violations states which financial statements were affected; sometimes, the amount channeled to bribes is negligible. Because of these problems, they were all removed from the sample.

We also ignored the occurrences of stock options backdating. The main reason was the difficulty of finding explanatory variables related to this kind of fraud, which is much more designed to benefit certain executives than to mislead the general public about the financial health of the

company. Including stock options backdating in this study would just mix two different phenomena in the same dependent variable. Nevertheless, the practice of backdating stock options might also distort financial statements.

Due to technical limitations, we kept in the sample only companies listed in three stock exchanges (NYSE, NASDAQ and AMEX). This means that companies traded over the counter were excluded, as well as those with no data available in Compustat.

A financial statement fraud is usually discovered only several years after it has begun. It is therefore very difficult to have a meaningful sample of companies that issued falsified financial statements in recent years. Any attempt to do so would increase the number of cases of misclassification, since many more companies would be incorrectly labeled as honest. Therefore,

we also opted to remove from the sample those instances of fraud that occurred before 1998 and after 2002. As one can see in Figure 1, the number of frauds discovered after 2002 apparently fell abruptly. Instead of signaling a true reduction in the number of fraud cases, this fall more likely shows the SEC's inability to report recent cases.

Financial companies – those with SIC (Standard Industrial Classification) codes from 6000 to 6999 – were also removed from the sample. The final sample contains 118 companies that fulfill all the requirements.

The models were not estimated only with the companies charged with fraud. The population of this study comprises all the companies with financial statements available in Compustat for the period ranging from the years 1998 to 2002. Companies with data for just some of these five years were included.

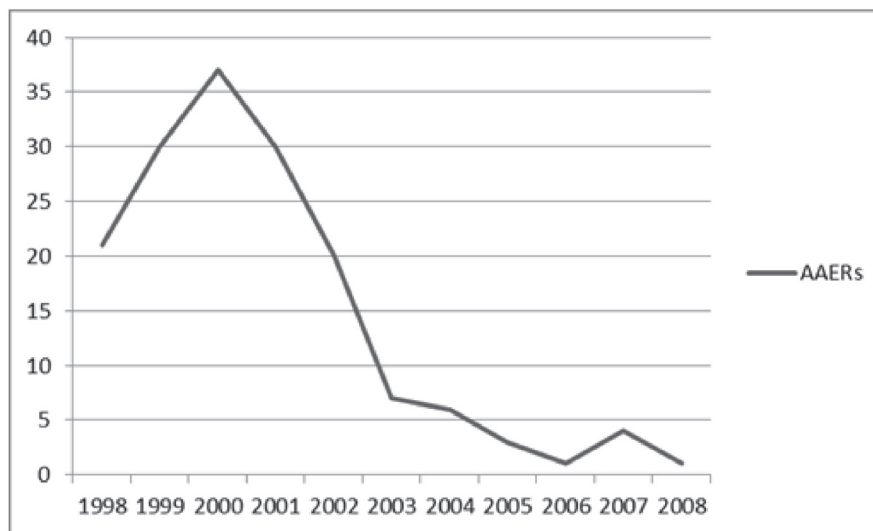


FIGURE 1 – AAERS issued between 1998 and 2008

Source: The authors

3.2 Factor analysis

Factor analysis is a multivariate technique that synthesizes the relationships observed in a set of interrelated variables, allowing the

identification of common factors. According to Jöreskog (1969), in a factor analysis with p variables and k factors, the basic model is:

$$\mathbf{z}_i = \Gamma \mathbf{f}_i + \boldsymbol{\varepsilon}_i \tag{1}$$

where \mathbf{z}_i is a vector of order p of observed variables for observation i , \mathbf{f}_i is a vector of order $k < p$ of latent common factor scores for the observation i , and $\boldsymbol{\varepsilon}_i$ is a vector of order p of unique scores for observation i . These unique scores can be seen as “error terms”: They represent the portion of \mathbf{z}_i that cannot be explained by \mathbf{f}_i . The capital Greek letter gamma (Γ) is a $p \times k$ matrix of factor loadings.

As the model suggests, factor analysis assumes it is possible to represent a set of p variables using a smaller number of k intrinsic factors. But, for factor analysis to be successful there must be reasonably strong correlations among the variables used in the analysis. In other words, the correlation matrix of these variables must be different from the identity matrix. To check for this assumption, we used the Bartlett’s test for sphericity, which rejected the null hypotheses of sphericity ($p = .000$). We concluded that factor analysis is adequate to our data set.

There are several factor-extraction procedures that can be used. The maximum-likelihood (ML) method has the advantage of permitting statistical significance testing of factor loadings. This method, however, requires the assumption of multivariate normality (FABRIGAR et al., 1999). Using the Doornik-Hansen test (DOORNIK; HANSEN, 2008), we found that the variables used in this paper rejected the null hypothesis of multivariate normality ($p = .000$). To avoid distorted results, we have chosen to use the principal factors (PF) method for factor extraction. This method has no distributional assumptions.

3.3 Model estimation

In probit and logit models, the dependent variable \tilde{y}_i is usually expressed as a function of a latent and observable variable I_i and an unknown threshold I_i^* (GUJARATI, 1995; POWERS; XIE, 2000):

$$\begin{aligned} \tilde{y}_i &= 1 \text{ se } I_i > I_i^* \\ \tilde{y}_i &= 0 \text{ se } I_i < I_i^* \end{aligned} \tag{2}$$

When there is misclassification in the dependent variable, the conditions above are not respected. The observed variable Y_i is not the same as the true variable \tilde{y}_i . The model now has two new parameters α_0 and α_1 :

$$\begin{aligned} \alpha_0 &= \Pr(y_{it} = 1 | \tilde{y}_{it} = 0) \\ \alpha_1 &= \Pr(y_{it} = 0 | \tilde{y}_{it} = 1) \end{aligned} \tag{3}$$

In studies about financial statement fraud, the most important error is the one shown as the parameter α_1 in the second equation. This error is the probability that a company that issued false financial statements is not classified as such (type II error). The other error (type I) is expected to be very rare and close to zero. Therefore, to make model convergence easier, this paper adopts the assumption that $\alpha_0 = 0$. Without this restriction, the expected value of y_i in a logit model is (HAUSMAN; ABREVAYA; SCOTT-MORTON, 1998):

$$E(y_i | \mathbf{x}_i) = \Pr(y_i = 1 | \mathbf{x}_i) = \alpha_0 + (1 - \alpha_0 - \alpha_1)\Lambda(\mathbf{x}_i' \boldsymbol{\beta}) \tag{4}$$

The log likelihood function that was effectively maximized in this paper, based in equation 4 but with the restriction $\alpha_0 = 0$, is:

$$\ln L = \sum_{y_i=0} \ln[1 - (1 - \alpha_1)\Lambda(\mathbf{x}'\boldsymbol{\beta})] + \sum_{y_i=1} \ln[(1 - \alpha_1)\Lambda(\mathbf{x}'\boldsymbol{\beta})] \tag{5}$$

Function 5 was maximized using the Newton-Raphson method.

4 RESULTS

We extracted 7 factors from 28 variables using the principal factors method. These 7 factors account for 89 percent of the variance of the original variables (Table 1). This means there is not much loss of information when we substitute

these 7 factors for the 28 original variables in the logit model.

Table 2 shows the factor loadings for each variable used in factor analysis (the matrix Γ in Eq. 1). Each variable is equal to the sum of the

products of the factors by their respective factor loadings, plus its unique score. Analyzing the highest absolute factors for each variable, we can understand the composition of each factor.

TABLE 1 – Variance explained by factor analysis (principal factors method)

Factor	Eigenvalue	Proportion	Cumulative
Factor 1	3,39808	,2197	,2197
Factor 2	2,79139	,1804	,4001
Factor 3	2,22932	,1441	,5442
Factor 4	1,63002	,1054	,6496
Factor 5	1,41903	,0917	,7413
Factor 6	1,28765	,0832	,8245
Factor 7	1,06663	,0689	,8935

Source: The authors

The first factor (that is able to capture almost 22 percent of the variance in the original variables) seems to be positively related to the financial health of the companies. It is positively related to the variables **EBITTA**, **ROA**, **zscore**, **WCTA**, Δ WCTA. It is also negatively related to the variables **TLTA** and Δ TLTA. Companies with less debt and high **ROA** usually present higher Factor 1 values.

Factor 2 captures 18 percent of the variance of the variables used. It is positively related to **COGSAL**, **CASHTA** and **INVSAL**, and is negatively related to variables Δ COGSAL and **EBITSAL**. These mixed results suggest that companies with more resources invested in some kinds of current assets (cash and inventory) and decreasing (but still high) costs of sales tend to have higher values for Factor 2. This factor is also negatively related to a profitability measure (**EBITSAL**).

Factor 3 is positively related to **CATA** and negatively related to **FATA**. Companies with more current assets tend to present a higher Factor 3. The variable Δ TLTE, although mostly captured by Factor 3, has a factor loading too small to be meaningful.

Factor 4 is positively related to Δ **CASHTA**, Δ **CATA** and **IETL**. It is also negatively related to Δ **FATA**. This factor captures increases in cash and current assets. This factor captures well the increases and decreases of the variables included in Factor 3. Companies with high Factor 4 are changing the composition of their assets and also incurring in more interest expenses.

Factor 5 is positively related to Δ **INVSAL** and negatively related to Δ **EBITTA** and Δ **ROA**. It is a variable that captures decreases in profitability and increases in inventory. This factor captures increases and decreases in some of the variables included in Factor 1.

There are no variables that have their highest absolute factor loading in Factor 7. The highest factor loadings associated with Factor 7 are the ones of Δ TLTA (0.406), **WCTA** (0.368) and Δ ROA (0.377). It seems that companies that increased their liabilities (but still have enough current assets to cover their current liabilities) and also improved their profitability tend to have a higher Factor 7. Companies with high Factor 7 are becoming more profitable possibly due to financial leverage.

We performed a stepwise logit regression, starting with all the factors and three dummy

variables that were not included in factor analysis. The least significant variable (the one with higher p-value) was eliminated, and the model was estimated again. This elimination procedure was repeated until all the remaining variables had p-values lower than 10 percent. The resulting model is shown in Table 3. It should be noted that the sample in each intermediary

model (and also in the final model) contained only the observations with data available for all the variables used in the initial model. Adding observations after dropping a variable, although possible, could generate biased estimates. Due to the large number of missing observations, the sample contained just 41 of the 118 cases of fraud originally identified.

TABLE 2 – Factor loadings and unique variances

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Uniqueness
COGSAL _{t-1}	-0,434	0,673	-0,504	-0,121	0,156	0,019	0,024	0,065
ΔCOGSAL _{t-1}	0,273	-0,460	0,420	0,139	-0,164	-0,088	-0,075	0,479
EBITSAL _{t-1}	0,439	-0,676	0,482	0,114	-0,154	0,011	0,015	0,080
EBITTA _{t-1}	0,572	-0,249	-0,127	-0,182	0,271	0,023	0,283	0,408
ΔEBITTA _{t-1}	-0,218	0,172	0,006	0,223	-0,466	0,235	0,261	0,533
ROA _{t-1}	0,726	0,188	-0,261	-0,187	0,329	0,018	0,021	0,226
ΔROA _{t-1}	-0,447	0,132	0,191	0,255	-0,468	0,180	0,377	0,287
zscore _{t-1}	0,299	0,145	-0,080	0,003	-0,179	-0,203	-0,002	0,810
CACL _{t-1}	0,181	0,324	0,129	0,047	-0,093	-0,356	-0,104	0,698
ΔCACL _{t-1}	-0,050	-0,087	-0,018	-0,034	0,022	0,159	0,041	0,961
CASHTA _{t-1}	0,175	0,556	0,299	0,186	-0,077	-0,479	-0,284	0,220
ΔCASHTA _{t-1}	0,197	0,269	0,175	0,541	0,345	-0,017	0,211	0,403
WCTA _{t-1}	0,660	0,310	-0,169	-0,156	-0,328	-0,153	0,368	0,148
ΔWCTA _{t-1}	0,516	0,197	-0,281	0,404	-0,182	0,429	-0,275	0,160
TLTA _{t-1}	-0,680	-0,210	0,317	0,149	0,303	0,184	-0,297	0,157
ΔTLTA _{t-1}	-0,561	-0,156	0,349	-0,290	0,147	-0,413	0,406	0,099
TLTE _{t-1}	0,001	-0,003	0,003	-0,009	0,003	0,012	0,004	1,000
ΔTLTE _{t-1}	-0,003	-0,008	-0,009	0,007	0,000	-0,005	0,000	1,000
ARTA _{t-1}	0,095	0,141	0,385	-0,379	0,069	0,461	0,058	0,459
INVSAL _{t-1}	-0,109	0,248	-0,163	-0,132	0,137	0,076	0,053	0,855
ΔINVSAL _{t-1}	-0,014	0,120	-0,137	-0,070	0,142	0,039	0,062	0,936
INVTA _{t-1}	0,117	0,144	0,251	-0,293	0,055	0,328	0,151	0,683
CATA _{t-1}	0,252	0,618	0,635	-0,255	0,020	0,093	-0,110	0,066
ΔCATA _{t-1}	0,153	0,216	0,232	0,505	0,423	0,026	0,281	0,363
FATA _{t-1}	-0,178	-0,478	-0,488	0,250	0,009	-0,120	0,113	0,413
ΔFATA _{t-1}	-0,105	-0,181	-0,135	-0,364	-0,281	-0,059	-0,173	0,694
SALAR _{t-1}	-0,005	-0,018	-0,022	0,031	0,010	-0,014	0,019	0,998
IETL _{t-1}	-0,065	-0,030	-0,058	0,098	0,036	0,026	0,043	0,978

Note. The factor with the highest absolute factor loading in each line is bolded.

Source: The authors

In the final model, five factors and one dummy variable (**UQUAL**) were significant at the 10 percent level. The model as a whole was also significant, with the likelihood ratio test

rejecting the null hypothesis at the 1 percent level ($p = .0024$); this means that the final model fits the data better than a model with only the intercept.

TABLE 3 – Logit model without misclassification parameter

Variable	Odds Ratio	Std. Error	z	P> z
Factor 1	1,5426	0,3470	1,9300	,054
Factor 2	0,5671	0,1769	-1,8200	,069
Factor 4	0,3444	0,1166	-3,1500	,002
Factor 6	0,4695	0,1561	-2,2700	,023
Factor 7	3,7244	0,7088	2,8700	,004
UQUAL	0,3488	0,1135	-3,2400	,001

Dependent variable: AAER
Observations: 10858
Period: 1998-2002

LR χ^2 : 20,39
Prob > χ^2 : ,0024
Log-likelihood: -259,47

Source: The authors

Taking into account the possibility of misclassification, we used the model from Table 3 as the initial model and performed a stepwise procedure. This procedure dropped three factors (1, 2 and 6), and the resulting model is shown in Table 4. Both remaining factors (4 and 7), while still rejecting the null hypothesis of the z-test, presented higher p-values in the model with a misclassification parameter. On the other hand, the variable **UQUAL** presented a lower p-value. The signs of the ratios of these variables remained unchanged. Since the Wald test rejected the null hypothesis, we conclude that at least one of the ratios of the variables used is statistically different from zero. The Wald test is asymptotically equivalent to the LR test used in the model without the misclassification parameter (ENGLE, 1984).

Companies with low Factor 4 and high Factor 7 are more likely to resort to fraud. *Ceteris paribus*, each increase by 1 in Factor 7 makes a company 2.6 times more likely to be involved in FSF. Since Factor 4 and **UQUAL** present odds ratios lower than 1, we decided to raise their odds ratios to the -1 power to better elucidate their influence on the likelihood of fraud. This way, each decrease by 1 in Factor 4 makes a company 2.49 times () more likely to be involved in fraud. Likewise, a company that did not receive an unqualified audit report is 5.12 times () more likely to be involved in fraud.

The misclassification parameter (α_1) estimated was 98.57 percent and is statistically different from zero ($p = .000$). This means that just a small part (1.43 percent) of the instances of fraud is discovered and charged by the SEC.

TABLE 4 – Logit model with misclassification parameter

Dependent variable: AAER				
Observations: 10858				
Period: 1998-2002				
Variable	Odds Ratio	Std. Error	z	P> z
Factor 4	0,4014		-2,27	0,023
Factor 7	2,6047		1,69	0,091
UQUAL	0,1955		-3,97	0,000
Misc. Parameter	Coefficient	Std. Error	z	P> z
α_1	0,9855	0,0030	297,72	0,000

Wald χ^2 : 18,58
 Prob > χ^2 : ,0003
 Log-likelihood: -261,99

Source: The authors

5 DISCUSSION AND CONCLUSION

Our first objective was to estimate the percentage of undetected cases of fraud in U.S. companies. The misclassification parameter α_1 was estimated as 98.57 percent and found to be statistically different from zero ($p = 0.000$). This result confirms what common sense suggests: The SEC is unable to report every case of fraud through AAERs. Companies involved in FSF have (*a priori*) a 1.43 percent probability of getting caught. The 95 percent confidence interval of this parameter is between 97.90 percent and 99.22 percent.

It must be noted that these results are no statement about what happens inside the commission or its willingness to punish accounting fraud. It is possible that the SEC is completely unaware of all the unreported cases of fraud, or perhaps it is aware of some, but chooses to focus in high-profile cases. In fact, leaving some (or many) crimes unpunished can be economically acceptable, since there are costs associated with investigation and prosecution. As Becker (1968) makes it clear in his economic approach to crime, attempting to punish every crime, from murder to petty larceny, is unlikely to be optimal from society's point of view. The same reasoning can be applied to white-collar crime.

Our second objective was to check if the logit model with misclassification parameter

would produce different results when compared to the traditional model. No test was run in order to verify whether these parameter estimates are statistically different, but the results are indeed qualitatively different: Three variables (factors) were dropped in the misclassification model. These results suggest that conclusions obtained using uncorrected fraud detection models might be simply not valid.

Our last objective was to obtain reliable hypotheses tests regarding some variables that might be related to financial statement fraud. We found three variables related to fraud ($p < .10$). Factor 4 is negatively related to fraud, indicating that companies with increasing cash scaled by total assets (Δ CASHTA) are less likely to be involved in FSF. Since companies with more cash are usually under less financial stress, this result matches our expectations. Gaganis (2009), on the other hand, found that the ratio of cash to total assets (**CASHTA**) is negatively related to fraud. It must be noted, however, that this author did not use the increase in cash scaled by total assets as a variable, and therefore these results are not directly comparable. The same caution applies to other comparative analyses presented here.

Factor 4 also indicates that companies with increasing current assets scaled by total assets (Δ CATA) are less likely to be involved in FSF. We expected different results, since companies with higher fixed assets (and fewer current assets) are

more capable of generating real revenue from their operations. These results are nevertheless consistent with Gaganis (2009), who found **CATA** to be negatively related to FSF.

Factor 7 is positively related to fraud, revealing that companies with increasing return on assets (Δ ROA) and leverage (measured by total liabilities scaled by total assets - Δ TLTA) are more likely to be involved in fraud. Although **ROA** is widely used in the extant literature, it is still not clear how it affects the likelihood of FSF. According to Summers and Sweeney (1998), it is positively related to fraud, whereas Crutchley, Jensen and Marshall (2007) and Gaganis (2009) hold that it is negatively related to fraud. We expected it to be negatively related to fraud, since decreasing profitability should pressure managers to get involved in FSF in order to please stockholders.

Most previous studies in the U.S. found no significant association between leverage (**TLTA**) and FSF (CRUTCHLEY; JENSEN; MARSHALL, 2007; ERICKSON; HANLON; MAYDEW, 2006; ETTREDGE et al., 2008), although Beneish (1999a) found a positive association using an older sample. In emerging markets such as Greece and China, TLTA is actually positively related to fraud (CHEN et al., 2006; KIRKOS; SPATHIS; MANOLOPOULOS, 2007; SPATHIS, 2002). Our results contribute to the conclusion that leverage seems to exert pressure to engage in fraud only in emerging markets.

Interestingly, first difference (change), variables played a considerable role in the extraction of both significant factors (4 and 7). These results suggest that fraud may be related to performance variations, especially decreases in net assets and profitability improvements due to increases in leverage. It may also explain why our results are often different from previously published articles, since they rarely used changes in variables. We suggest that future studies might benefit from a shift to models that focus on changes in the financial position of companies.

Audit opinion reports are also related to fraud, since an adverse audit opinion indicates

that the financial statements are materially misstated, and a qualified opinion indicates there might be deviations from the adopted accounting standards. In our study, we used a dummy variable (**UQUAL**) that equals one whenever the audit opinion report is unqualified, and zero otherwise. This dummy variable was used by Skousen and Wright (2008) as a proxy for rationalization, and in their analysis this variable failed to reject the null hypothesis of the Wilcoxon test ($z = -0.814$ and $p = .208$). We repeated the Wilcoxon test and found quite consistent results ($z = -1.30$, $p = .1921$). Since Skousen and Wright adopted a strict criterion for inclusion of a variable in their logit regression ($p < .15$ in the Wilcoxon test), they did not test whether this variable was relevant in a multivariate model. We found that, when controlling for the effect of other variables (included in factors 4 and 7), **UQUAL** is in fact negatively related to the likelihood of financial statement fraud. Our findings show that companies with clean audit reports are less likely to be involved in fraud, as expected. This result also suggests that using univariate tests to decide which variables should be used in a multivariate model might lead to the omission of relevant variables.

We recommend caution in the use of uncorrected binary choice models as a decision-support tool in fraud detection or fraud risk assessment. It is more adequate to use a specification that considers the possibility of misclassification problems, and doing so yields a more reliable model. It is also important to avoid the downward-biased fraud likelihood estimates generated by traditional models. These measures can contribute to more precise assessments of fraud likelihood when using econometric models. Adopting the misclassification model is quite straightforward—although not yet integrated in statistical and econometric software—and could become standard procedure in financial statement fraud research.

It is important to stress that this is not the only method that can be employed to avoid the problems that arise from misclassification. Wang (2011) was able to estimate the conditional

misclassification probability using a different method. It is important to note, however, that the method proposed by Wang requires the identification of variables related to the likelihood of fraud detection. While this might allow the testing of interesting research hypotheses, it is an unnecessary burden for researchers interested just in variables related to the likelihood of fraud.

As research advances, it will be possible to use the method adopted here to study problems related to the efficiency of regulatory agencies in repressing fraud. Testing for changes in the misclassification parameter along time might allow conclusions about the efficiency of certain policies concerning fraud. When more data becomes available, it will be possible to test, for instance, whether the Sarbanes-Oxley Act increased the SEC's capability of identifying and punishing financial statement fraud.

The method employed here, however, is not able to correct specification problems in the model. If the model is underspecified (i.e. there are missing variables), then the results will nevertheless be inconsistent. There is also need for theoretical development on the causes of fraud that allows us to identify better models.

REFERENCES

ABBOTT, L. J.; PARK, Y.; PARKER, S. The effects of audit committee activity and independence on corporate fraud. **Managerial Finance**, Bradford v. 26, n. 11, p. 55-68, 2000.

ALBRECHT, W. S. et al. **Fraud examination**. 3rd. ed. Mason: South-Western Cengage, 2009.

ALTMAN, E. I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. **The Journal of Finance**, Malden, v. 23, n. 4, p. 589-609, Sept. 1968.

AMERICAN INSTITUTE OF CERTIFIED PUBLIC ACCOUNTANTS (AICPA). **Consideration of fraud in a financial statement audit**. AU Section 316. New York: AICPA, 2002.

BEASLEY, M. S. An empirical analysis of the relation between the board of director composition and financial statement fraud. **The Accounting Review**, Sarasota, v. 71, n. 4, p. 443-465, Oct. 1996.

BECKER, G. S. Crime and punishment: an economic approach. **Journal of Political Economy**, Chicago, v. 76, n. 2, p. 169-217, Mar./Apr. 1968.

BENEISH, M. D. Detecting GAAP violation: implications for assessing earnings management among firms with extreme financial performance. **Journal of Accounting and Public Policy**, New York, v. 16, n. 3, p. 271-309, Autumn 1997.

_____. The detection of earnings manipulation. **Financial Analysts Journal**, Charlottesville, v. 55, n. 5, p. 24-36, Sept./Oct. 1999a.

_____. Incentives and penalties related to earnings overstatements that violate GAAP. **The Accounting Review**, Sarasota, v. 74, n. 4, p. 425-457, Oct. 1999b.

BEST, P. J.; BUCKBY, S.; TAN, C. Evidence of the audit expectation gap in Singapore. **Managerial Auditing Journal**, Bradford, v. 16, n. 3, p. 134-144, 2001.

BRAZEL, J. F.; JONES, K. L.; ZIMBELMAN, M. F. Using nonfinancial measures to assess fraud risk. **Journal of Accounting Research**, Chicago, v. 47, n. 5, p. 1135-1166, July 2009.

CHEN, G.; FIRTH, M.; GAO, D.; RUI, O. Ownership structure, corporate governance, and fraud: evidence from China. **Journal of Corporate Finance**, London, v. 12, n. 3, p. 424-448, June 2006.

CHOI, J. H.; GREEN, B. P. Assessing the risk of management fraud through neural network technology. **Auditing: A Journal of Practice & Theory**, Sarasota, v. 16, n. 1, p. 14, Mar. 1997.

CHOO, F.; TAN, K. An “American dream” theory of corporate executive fraud. **Accounting Forum**, [S.l.], v. 31, n. 2, p. 203-215, June 2007.

CHUI, L.; PIKE, B. Auditors’ responsibility for fraud detection: new wine in old bottles? In: AMERICAN ACCOUNTING ASSOCIATION - ANNUAL MEETING, 2011, Denver. **Proceedings...** Sarasota: American Accounting Association – AAA, 2011.

COLEMAN, J. W. Toward an integrated theory of white-collar crime. **American Journal of Sociology**, Chicago, v. 93, n. 2, p. 406-439, Sept. 1987.

CRUTCHLEY, C. E.; JENSEN, M. R. H.; MARSHALL, B. B. Climate for scandal: corporate environments that contribute to accounting fraud. **The Financial Review**, Knoxville, v. 42, n. 1, p. 53-73, Feb. 2007.

DECHOW, P. M.; SKINNER, D. J. Earnings management: reconciling the views of accounting academics, practitioners, and regulators. **Accounting Horizons**, Sarasota, v. 14, n. 2, p. 235-250, June 2000.

DESHMUKH, A.; TALLURU, L. A rule-based fuzzy reasoning system for assessing the risk of management fraud. **Intelligent Systems in Accounting, Finance & Management**, Chichester, v. 7, n. 4, p. 223-241, Dec. 1998.

DOORNIK, J. A.; HANSEN, H. An omnibus test for univariate and multivariate normality. **Oxford Bulletin of Economics and Statistics**, Oxford, v. 70, p. 927-939, Dec. 2008.

ENGLE, R. F. Wald, likelihood ratio, and Lagrange multiplier tests in econometrics. In: GRILICHES, Z.; INTRILIGATOR, M. D. (Eds). **Handbook of econometrics**. Amsterdam: North-Holland, 1984. v. 2, p. 775-826.

ERICKSON, M.; HANLON, M.; MAYDEW, E. L. How much will firms pay for earnings that

do not exist? Evidence of taxes paid on allegedly fraudulent earnings. **The Accounting Review**, Sarasota, v. 79, n. 2, p. 387-408, Apr. 2004.

_____; _____. Is there a link between executive equity incentives and accounting fraud? **Journal of Accounting Research**, Chicago, v. 44, n. 1, p. 113-143, Mar. 2006.

ETTREDGE, M. L. et al. Is earnings fraud associated with high deferred tax and/or book minus tax levels? **Auditing: A Journal of Practice & Theory**, Sarasota, v. 27, n. 1, p. 1-33, May 2008.

FABRIGAR, L. R. et al. Evaluating the use of exploratory factor analysis in psychological research. **Psychological Methods**, Washington, D.C., v. 4, n. 3, p. 272-299, Sept. 1999.

FANNING, K.; COGGER, K. O. Neural network detection of management fraud using published financial data. **Intelligent Systems in Accounting, Finance & Management**, Chichester, v. 7, n. 1, p. 21-41, Mar. 1998.

FENG, M. et al. Why do CFOs become involved in material accounting manipulations? **Journal of Accounting and Economics**, Amsterdam, v. 51, n. 1-2, p. 21-36, Feb. 2011.

FRANCIS, J. R.; KRISHNAN, J. Accounting accruals and auditor reporting conservatism. **Contemporary Accounting Research**, Toronto, v. 16, n. 1, p. 135-165, Mar. 1999.

GAGANIS, C. Classification techniques for the identification of falsified financial statements: a comparative analysis. **Intelligent Systems in Accounting, Finance & Management**, Chichester, v. 16, n. 3, p. 207-229, July 2009.

GERETY, M.; LEHN, K. The causes and consequences of accounting fraud. **Managerial and Decision Economics**, Chichester, v. 18, n. 7/8, p. 587-599, Nov./Dec. 1997.

GUJARATI, D. N. **Basic econometrics**. 3rd ed. New York: McGraw-Hill, 1995.

HAUSMAN, J. A.; ABREVAYA, J.; SCOTT-MORTON, F. M. Misclassification of the dependent variable in a discrete-response setting. **Journal of Econometrics**, Amsterdam, v. 87, n. 2, p. 239-269, Sept. 1998.

JAMAL, K. Mandatory audit of financial reporting: a failed strategy for dealing with fraud. **Accounting Perspectives**, Toronto, v. 7, n. 2, p. 97-110, May 2008.

JENSEN, M. C.; MECKLING, W. H. Theory of the firm: managerial behavior, agency costs and ownership structure. **Journal of Financial Economics**, Amsterdam, v. 3, n. 4, p. 305-360, Oct. 1976.

JOHNSON, S. A.; RYAN, H. E.; TIAN, Y. S. Managerial incentives and corporate fraud: the sources of incentives matter. **Review of Finance**, Dordrecht, Oxford, v. 13, n. 1, p. 115-145, May 2008.

JÖRESKOG, K. G. A general approach to confirmatory maximum likelihood factor analysis. **Psychometrika**, New York, v. 34, n. 2, p. 183-202, June 1969.

KAMINSKI, K. A.; WETZEL, T. S.; GUAN, L. Can financial ratios detect fraudulent financial reporting? **Managerial Auditing Journal**, Bradford, v. 19, n. 1, p. 15-28, 2004.

KIRKOS, E.; SPATHIS, C.; MANOLOPOULOS, Y. Data mining techniques for the detection of fraudulent financial statements. **Expert Systems with Applications**, [S.l.], v. 32, n. 4, p. 995-1003, May 2007.

LEE, T. A.; INGRAM, R. W.; HOWARD, T. P. The difference between earnings and operating cash flow as an indicator of financial reporting fraud. **Contemporary Accounting Research**, Toronto, v. 16, n. 4, p. 749-786, Winter 1999.

LEE, T. H.; GLOECK, J. D.; PALANIAPPAN, A. K. The audit expectation gap: an empirical study in Malaysia. **Southern African Journal of Accountability and Auditing Research**, [S.l.], v. 7, p. 1-15, 2007.

LENNOX, C.; PITTMAN, J. A. Big five audits and accounting fraud. **Contemporary Accounting Research**, Toronto, v. 27, n. 1, p. 209-247, Spring 2010.

LIN, J. W.; HWANG, M. I.; BECKER, J. D. A fuzzy neural network for assessing the risk of fraudulent financial reporting. **Managerial Auditing Journal**, Bradford, v. 18, n. 8, p. 657-665, 2003.

MILLER, G. S. The press as a watchdog for accounting fraud. **Journal of Accounting Research**, Chicago, v. 44, n. 5, p. 1001-1033, Dec. 2006.

OGUT, H. et al. Prediction of financial information manipulation by using support vector machine and probabilistic neural network. **Expert Systems With Applications**, Tarrytown, NY, v. 36, n. 3, p. 5419-5423, Apr. 2009.

PEDNEAULT, S. **Fraud 101**. 3rd ed. Hoboken: John Wiley & Sons, 2009.

PERSONS, O. S. Using financial statement data to identify factors associated with fraudulent financial reporting. **Journal of Applied Business Research**, Littleton, v. 11, n. 3, p. 38-46, June 1995.

POIRIER, D. J. Partial observability in bivariate probit models. **Journal of Econometrics**, Amsterdam, v. 12, n. 2, p. 209-217, Feb. 1980.

POWERS, D. A.; XIE, Y. **Statistical methods for categorical data analysis**. San Diego: Academic Press, 2000.

SKOUSEN, C. J.; WRIGHT, C. J. Contemporaneous risk factors and the prediction

of financial statement fraud. **Journal of Forensic Accounting**, Tampa, Florida, v. 9, n. 1, p. 37-62, 2008.

SPATHIS, C. Detecting false financial statements using published data: some evidence from Greece. **Managerial Auditing Journal**, Bradford, v. 17, n. 4, p. 179-191, 2002.

SUMMERS, S. L.; SWEENEY, J. T. Fraudulently misstated financial statements and insider trading: an empirical analysis. **The Accounting Review**, Sarasota, v. 73, n. 1, p. 131-146, Jan. 1998.

WANG, T. Y. **Corporate securities fraud: insights from a new empirical framework**. Oxford, 2011. (Journal of Law, Economics, and Organization). No prelo.