ASSESSING AUDIT QUALITY:
A VALUE RELEVANCE PERSPECTIVE

Li Dang*
Department of Accounting
College of Business
200 Bexell, Oregon State University
Corvallis, OR 97331-2603
Tel.: 541-737-6049
e-mail: dangl@bus.oregonstate.edu

Kevin F. Brown
Department of Accountancy
Wright State University

B D McCullough
Department of Decision Sciences
Drexel University

November 30, 2004

* Corresponding author

We would like to thank Chris Agoglia, Joe Brazel, Dennis Caplan, Seong-Yeon Cho, Roger Graham, Henry Jaenicke for their helpful comments.

ASSESSING AUDIT QUALITY:
A VALUE RELEVANCE PERSPECTIVE

ABSTRACT

This study examines whether the market assesses audit quality accurately in cases of apparent audit failures and whether auditor reputation influences the market's ability to assess audit quality. Value-relevance of accounting information serves as our proxy for market-perceived audit quality. We compare the value-relevance of accounting information across samples of firms experiencing apparent audit failures and matched non-audit failure samples using bootstrapping to test for significant differences in R^2s. We document that accounting information is less value-relevant for firms experiencing apparent audit failures. This result provides support for the assertion that the market appears to accurately assess audit quality *ex ante*, even though formal allegations of audit failure may not appear for years after its occurrence.

Keywords: Audit Quality, Value Relevance, Audit Failures, Bootstrapping

Data Availability: Data are available from sources cited in the paper.

**ASSESSING AUDIT QUALITY:
A VALUE RELEVANCE PERSPECTIVE**

I. INTRODUCTION

The purpose of this paper is to document market differentiation of audit quality by applying a value relevance perspective. DeAngelo (1981, 186) defines audit quality as "the market-assessed joint probability that a given auditor will both (a) discover a breach in the client's accounting system, and (b) report the breach." However, due to the nature of the audit process and the reporting of audit outcomes, assessment of quality for particular audit engagements remains somewhat murky, with users typically responding to an auditor's reputation (e.g., Teoh and Wong 1993).

However, whether auditor reputation serves as a reliable proxy for audit quality may be in doubt due to the high profile audit failures exposed during the latest downturn in the financial markets. Many now notorious firms (e.g., Enron, WorldCom, and Tyco) were audited by Big 5 auditors with reputations superior to those of the non-Big 5. One might conclude that an auditor's reputation may actually hinder the market's ability to assess the reliability of accounting information. Thus, value relevance, the market's perception of audit quality, may not reflect actual audit quality.

In order to determine whether the market assesses audit quality accurately and whether auditor reputation might have an impact on market assessment, this study compares the ex ante value relevance of accounting information of publicly-held U.S. firms experiencing apparent audit failures to a matched group of firms not experiencing such failures. Ex ante value relevance refers to the value relevance of accounting information prior to the discovery of the audit failure. Matching audit failure firms with non-audit failure firms allows audit quality to be measured on a “service-by-service” basis, consistent with the suggestion of Lam and Chang (1994). As such, the approach employed in this study offers an opportunity to investigate directly whether the market assesses audit quality accurately.

Several sources are used to identify apparent audit failures, which are, for the purpose of this study, instances of an auditor issuing an unqualified opinion on materially misstated financial statements. These sources include the SEC’s Accounting and Auditing Enforcement Releases (AAERs), restatements of financial statements because of past misstatements, and audit failures revealed by litigation against auditors.

Consistent with prior research (Collins et al. 1997; Rees 1999; Rajgopal 2002), value relevance is measured by the explanatory power of earnings and book values for stock prices, i.e., the regression of stock price on earnings and book values. The data are tested by applying a bootstrapping analysis. The results of this testing indicate that the accounting information of firms experiencing apparent audit failures is generally less value-relevant than that of a matched group of firms not experiencing such failures. Assuming that value relevance of accounting information reflects the market’s assessment of audit quality, these results indicate that the market appears to correctly discount less reliable accounting information regardless of auditor reputation.

The remainder of this paper is organized as follows. Section II describes prior research relevant to market perceptions of audit quality. Section III contains the development of the hypothesis. Section IV discusses data collection and the methodology followed in this study. Section V presents our analysis and results. Section VI provides concluding remarks.

II. PRIOR RESEARCH

DeAngelo (1981) analytically demonstrates how auditor size might be positively related to audit quality. In her study, auditor size is measured by number of clients. She argues that since auditors earn client-specific quasi-rents, auditors with more clients have more to lose by failing to report discovered misstatements in financial statements. Based on DeAngelo's (1981) analytical results, many prior studies have used auditor size, specifically a Big 8/4 vs. non-Big 8/4 dichotomy, to differentiate audit quality levels (Krishnan 2003; Zhou and Elder 2001; Bauwhede et al. 2000; Becker et al. 1998; Clarkson and Simunic 1994; Copley 1991). Research employing auditor size/reputation as a proxy for audit quality is extensive. For example, Hogan (1997) documents that the perception of higher audit quality is associated with less underpricing in the IPO market.

The results of studies that test the relationship between audit quality and auditor size usually support the hypothesis that audit quality and audit firm size are positively associated. For instance, Palmrose (1988) tests auditor size and audit quality using litigation activity as the proxy for audit quality. Davidson and Neu (1993) argue that higher auditor quality should be associated with larger management earnings forecast errors. Specifically, they find that Big 8 accounting firms are associated with larger management earnings forecast errors. Teoh and Wong (1993) test this relation using earnings response coefficients to measure audit quality. However, the proxy used in Teoh and Wong's study may capture only market perceptions of audit quality, rather than an auditor's actual ability to detect and report accounting misstatements. Because audit quality is unobservable when audit services are provided, and some period of time usually elapses before audit failures become apparent, the market may be misled by an auditor's reputation. Moreland (1995) demonstrates how SEC enforcement actions against Big 8/6 accounting firms affect market perceptions of their audit quality. Whether the market properly discounts less reliable accounting information in spite of auditor reputation is the focus of the current study.

In contrast to the studies documenting that Big 8/4 auditors provide higher audit quality than non-Big 8/4 auditors (e.g., Davidson and Neu 1993), other studies indicate that Big 8/4 auditors might not always provide higher quality audit service than do non-Big 8/4 auditors. Kim et al. (2003) find that Big 6 auditors are more (less) effective in preventing earnings management than non-Big 6 auditors in the presence (absence) of reporting incentive conflicts between managers and auditors.

Nichols and Smith (1983) test whether the market distinguishes audit quality levels between Big 8 and non-Big 8 auditors. Their results do not show a difference between these two groups of auditors. Lam and Chang (1994) investigate the relationship between audit quality and auditor size using the mean error in earnings forecasts as an audit quality measure in prospectuses of IPOs in Singapore. They find that, on average, use of Big 6 auditors does not result in smaller prediction errors in earnings forecasts than does use of non-Big 6 auditors. Petroni and Beasley (1996) find no systematic difference in claim loss reserve accuracy or bias between clients of Big 8 auditors and clients of other auditors. Tate (2001) examines the results of financial statement and compliance audits performed in accordance with the U.S. Office of Management and Budget's A-133 requirements. She finds that, although Big 5 auditors report more questioned costs and more findings than non-Big 5 auditors, Big 5 auditors are less likely to issue qualified opinions given their clients' deficiencies in internal control than non-Big 5 auditors. She also finds that Big 5 auditors are less likely than non-Big 5 auditors to report significant deficiencies in internal control.

In summary, although many studies support the argument that Big 8/4 auditors provide higher-quality audit service than non-Big 8/4 auditors, other studies have found conflicting results. Reputation might not be an effective proxy for audit quality in certain circumstances. Therefore, the primary objective of this study is to test how the market assesses audit quality given the auditor reputation effect.

III. DEVELOPMENT OF HYPOTHESIS

Investigating how well the market assesses audit quality might provide insights for audit quality research, specifically, evidence of whether market perception reflects audit quality. DeAngelo's (1981) argument for a positive relationship between auditor size and audit quality assumes market transparency. However, whether financial statement users assess audit quality accurately is unclear. It is possible that the reputation effect is so significant that the market assesses audit quality mainly on the basis of auditor reputation rather than on the auditor's ability to detect and report material misstatements of financial statements for specific audits.

Apparent audit failures provide a unique way of testing the accuracy of the market's assessment of audit quality because they provide less ambiguous, *post hoc* determinations of quality. In order to provide insights into audit quality, this study examines whether the value relevance of accounting information is different for firms experiencing apparent audit failures than for firms not experiencing apparent audit failures. Value relevance indicates the extent to which accounting information explains the variation of market prices. When financial statement users perceive higher audit quality and therefore more reliable financial statements, accounting information such as earnings and book values should explain more of the variation in stock prices. If auditing is a monitoring device that adds assurance to financial statements, then, *ceteris paribus*, financial statement users should view financial statements resulting from high-quality audits as more reliable than financial statements resulting from low-quality audits. In other words, the value relevance of accounting data resulting from high-quality audits and low-quality audits should be different. If an audit is perceived to be of higher quality, the financial statements associated with that audit should be perceived to be more reliable. Investors will rely more heavily on accounting data when perceived audit quality is higher. Therefore, the value relevance of accounting data should reflect perceived audit quality.

Following prior studies (Amir 1996; Amir and Lev 1996; Collins et al. 1997; Rees 1999; Rajgopal et al. 2002), the Ohlson Model (Ohlson 1995) is used to test the value relevance of accounting information. The Ohlson Model relates a firm's market value to its contemporaneous accounting information. Specifically, this model provides a structure for empirical study of the relationship between equity values and both earnings and book values (Stober 1999). The degree of value relevance is measured by the R^2 of the Ohlson Model (Collins et al. 1997; Rees 1999; Rajgopal 2002). The R^2 measures how much of the variation in the dependent variable is explained by independent variables. If financial statement users assess audit quality accurately, the association between stock price and accounting information, which is measured by the R^2 of the Ohlson Model, should be weak when audit quality is poor because accounting information would be viewed as less reliable for investment decision-making.

This discussion leads to the following hypothesis:

> ***Hypothesis:*** *The explanatory power of earnings and book values for stock prices of firms experiencing apparent audit failures will be lower than the explanatory power for a matched group of firms that have not experienced apparent audit failures.*

IV. DATA COLLECTION AND METHODOLOGY

Cases of Apparent Audit Failures and the Matched Control Group

Firms included in this study are selected from two primary sources: the SEC's AAERs and restatements of financial statements found in the Wall Street Journal Index and Lexis-Nexis news library for fiscal years ending between 1980 and 2000, in which an apparent audit failure occurred. The initial sample is the combination of these two data sources. AAERs indicate firms whose financial statements contain misstatements documented in SEC sanctions against firms or their auditors. We exclude AAER cases where quarterly misstatements are the focus of the sanction because quarterly statements are only reviewed but not audited by auditors. Restatements consist of firms that restated prior years' financial statements because of significant misstatements. These represent apparent audit failures because initially auditors did not detect and/or report those material misstatements. We searched keywords with the root "restat-" in Wall Street Journal Index and Lexis-Nexis News Library. Quarterly restatements are also excluded from our sample. In order to ensure that the search of the above two sources did not omit any audit failures, we also searched, as a secondary source, news accounts of auditor litigation including the auditor litigation database compiled by Palmrose (1999). Cases of litigation against auditors contain allegations that auditors failed to detect and report material misstatements. Such cases are included in this analysis due to possible lags in SEC scrutiny which would ordinarily result in enforcement activity and/or restatements of prior financial statements. Accordingly, litigation cases that provide substantial evidence of audit failure are included as apparent audit failures. While lawsuits seeking damages from an auditor are not always indicative of the auditor's failure to adhere to professional standards, such suits may reveal audit failures before subsequent issuance of an AAER. A careful review of the litigation news accounts helped to ensure that only apparent audit failures are captured in the sample.

In order to perform hypothesis testing, all accounting data for the years of the alleged audit failures that is necessary for testing the proposed relation is extracted from the COMPUSTAT database. All stock price information is obtained from Center for Research in Security Prices (CRSP). The matched control group is selected based on established matching criteria. These criteria include the year of financial statements, auditor size (Big 8/4 vs. non-Big 8/4), client industry, and client size. The required accounting data and price information is compiled for the matched control group from COMPUSTAT and CRSP, respectively.

Model Specification

The model used to test the hypothesis is Ohlson's (1995) valuation model. Ohlson analytically demonstrates that a firm's market value can be expressed by its contemporaneous abnormal earnings and book value. Following Amir (1996), Amir and

Lev (1996), and Collins et al. (1997), the current study applies the Ohlson Model expressed as follows:

$$P_{it} = \alpha_1 EPS_{it} + \alpha_2 BVPS_{it} + \varepsilon_{it} \qquad \text{(Equation 1)}$$

where:

P_{it} = closing stock price of firm i's equity three months after fiscal year end t;

EPS_{it} = firm i's reported earnings per share before extraordinary items for period t;

$BVPS_{it}$ = firm i's book value of equity per share at time t;

ε_{it} = random error term with mean 0 and variance 1;

α_1 and α_2 = the regression coefficients.

Testing of Hypothesis

In order to test the hypothesis, comparisons are made between the apparent audit failure group of firms and the matched control group of firms not experiencing audit failures. For each comparison, two OLS regressions of the Ohlson Model are performed: one regression for the audit failure group, and the other for the matched control group. If the R^2 for the audit failure group is statistically significantly lower than the R^2 for the matched control group, it suggests that the market appears to assess audit quality accurately. If the R^2 for the audit failure group is equal to or higher than the R^2 for the matched control group, it suggests that the market may not assess audit quality accurately. Moreover, we would like to see whether financial statement users assess audit quality differently in different situations. For example, users might assess audit quality differently for firms audited by Big 8/4 auditors than those audited by non-Big 8/4 auditors. Comparisons we use to test the hypothesis are summarized in Table 1.

Insert Table 1 here

Comparison 1 tests whether the market assesses audit quality accurately in general. Comparisons 2 and 3 test the market's assessment within Big 8/4 and non-Big 8/4 groups. Comparisons 4 and 5 directly test whether auditor reputation impacts the market's assessment of audit quality.

V. ANALYSIS AND RESULTS

Sample Selection and Characteristics

We searched both the SEC's online archives and the Lexis-Nexis News Library for AAERs. Between 1982 and 2000, AAERs No. 1 through No. 1357 were released. There are 559 unique firms identified from those AAERs.[1] Of these 559 firms, 383 unique firms are selected and the remaining 176 firms are deleted because the AAERs: (1) do not indicate the years misstated; (2) indicate the misstatement years are before 1980; (3) pertain to quarterly financial statements; or (4) indicate that the auditors are not responsible for the misstated financial statements. Our search for financial statement restatements in the Wall Street Journal Index and the Lexis-Nexis News Library yields 462 unique firms with restatements. Of those 462 firms, 273 unique firms meet the requirements for our analysis. The other 189 firms are deleted for similar reasons as those discussed in (1) through (4) above. We then combine the 383 firms sanctioned in the AAERs with the 273 firms disclosing subsequent restatements and find that 42 firms

[1] Several AAERs do not identify firm names. Further, some firms received multiple AAER sanctions. There are also some cases where the alleged impropriety involves a governmental entity.

exist in both data sources. Therefore, we identify 614 firms that experienced apparent audit failures from these two data sources.

Since our analysis requires accounting data from COMPUSTAT, we first searched for the ticker identification numbers for those 614 firms by their firm names in COMPUSTAT. One hundred and fifty-four firms were deleted because they do not have ticker identification numbers,[2] leaving 460 firms remaining for further analysis. An additional 297 firms that meet our data requirements are found in the auditor litigation database compiled by Palmrose (1999).[3] Since 74 of these firms are already been included in our 460-firm sample, our overall sample is augmented to 683 unique firms. All these firms have materially misstated annual financial statements within the period from 1980 to 2000, and their COMPUSTAT ticker identification numbers are available.

As this study defines apparent audit failures as cases in which auditors provide unqualified opinions on financial statements that contain material misstatements, a search of the COMPUSTAT database[4] for those firms receiving unqualified opinions was performed. Of the 683 firms, 442 firms (848 firm/years) have audit opinion information available for the specified financial statement years. An additional twenty-six firms (69 firm/years) were deleted because auditors issued qualified opinions to those firms. Therefore, 416 firms (779 firm/years) are eligible for inclusion as apparent audit failures.

The sample for this study was reduced further because of the absence of other data required to test the hypothesis. In order to test the hypothesis, earnings per share, book

---

[2] Having ticker identification numbers does not necessarily mean that these firms have all required data for specific years in question available in COMPUSTAT.

[3] Palmrose's (1999) database provides tickers for those companies that have ticker identification numbers in COMPUSTAT. We exclude those cases of apparent audit failure which do not have tickers in the Palmrose (1999) database.

[4] COMPUSTAT data item 149 provides both auditor and audit opinion information. However, it does not provide reasons why the qualified audit opinions were issued. We include auditor opinions with a value 1 (unqualified opinion) and 4 (unqualified opinion with explanatory language).

value of stockholders' equity, and stock price data must be available from the COMPUSTAT and CRSP databases. Given these considerations, the resulting sample of apparent audit failures is 346 firms (616 firm/years). See Table 2 for sample size and industry distribution information. Given the matched-pairs design used in this study, each audit failure firm in the sample is matched with a control firm counterpart based on year of misstated financial statements, industry (SIC code), company size (total assets), and auditor type (Big 8/4 vs. non-Big 8/4).[5] In searching the matched-pairs, we select similar size firms in the same industry (2 to 4-digit SIC code). If there is more than one similar size firms available, we select the one with the closest SIC code. If there is no firm of similar size available in the same industry, we select the firm closest in total assets. If, following the above selecting procedure, there is no firm available to serve as a match in the same industry as the audit failure firm, we drop the audit failure observation.

Insert Table 2 here.

Testing of the Hypothesis

Prior accounting research frequently has used a model's R^2 statistic to measure the value relevance of accounting information (Lang et al. 2003; Sami and Zhou 2004; Francis and Schipper 1999; Nwaeze 1998; Collins et al. 1997; Amir and Lev 1996; Harris et al. 1994). These studies have measured value relevance as the R^2 resulting from the regressions of stock prices on per share values of earnings and book values of equity. These studies compare value relevance measured by R^2 either over different time periods or across different samples. In this study, R^2s are compared across audit failure and non-

[5] The matched-pairs control group also excludes firms with qualified audit opinions.

audit failure groups to investigate whether the market might assess audit quality accurately. Except in Harris et al. (1994), Lang et al. (2003), and Sami and Zhou (2002), value-relevance studies have not included a formal test for the difference of R^2s. For example, Dechow (1994) simply compares the adjusted R^2s of earnings/returns model across different aggregate accrual levels.

Harris et al. (1994), Lang et al. (2003) and Sami and Zhou (2004) use the procedure demonstrated in Cramer (1987) to obtain the mean and the variance of R^2s. They then conduct a z-test to compare the means of two R^2s. A major problem with applying the Cramer method in this context of a z-test is that it depends on the assumption that R^2 is normal, but R^2 is not normal, even asymptotically (Anderson 2003, 155). As an example, see Figure1 for a demonstration of an R^2 distribution. Figure 1 is the distribution of the sample R^2 when the true R^2=0.9, the sample size n=100 and the number of independent variables p=10. Visual inspection of this figure reveals that the distribution of R^2 is not normal. An additional complication is that calculation of the moments of R^2, which is required to apply Cramer's method, is prone to computational difficulties that may produce incorrect answers, even when the requisite formulae seem to be programmed correctly. See McCullough and Dang (2004) for a discussion of this point.

Insert Figure 1 here

Because of the difficulty and possible unreliability of using the Cramer procedure, although testing results applying the Cramer procedure are also included, this study primarily uses the bootstrap method to create tests for the difference in R^2s. Bootstrapping is a resampling method that requires fewer assumptions than traditional methods and generally provides more accurate results. For example, the bootstrap method does not require normality of the distribution of the R^2 and it can provide a faster convergence to the expected value of the parameter of interest.

Since the models cannot be nested to formally test H_0: $R^2_a = R^2_b$ against H_a: $R^2_a \neq R^2_b$, we employ the usual approach of comparing the confidence interval for R^2_a and the confidence interval for R^2_b. Barr (1969) illustrates that the length of the confidence intervals for the two-interval test must be constructed with the multiplier

$z' = \frac{\sqrt{n_1 + n_2}}{\sqrt{n_1} + \sqrt{n_2}} z_{0.975}$ if the significance level of 0.05 is desired. When the sample sizes of the two samples are the same, i.e., both have n observations, the multiplier becomes $z' = \frac{\sqrt{2n}}{2\sqrt{n}} z_{0..975} = \frac{\sqrt{2}}{2} z_{0.975}$. As a result, the percentage of the confidence intervals is 83.4%. Therefore, comparing two 95% intervals is incorrect if α=0.05 is required.[6]

Descriptive Statistics

The sample selection procedure yielded 616 firm/year observations of apparent audit failures, which were matched with a control group in order to perform the comparisons required to test the hypothesis. In the audit failure group, 502 (114) firm/years were audited by Big 8/4 auditors (non-Big 8/4 auditors). See Table 3 for the observations included in each comparison.[7]

Insert Table 3 here

---

[6] For example, if the sample size for two samples are both 300, the multiplier for the confidence interval is

$$z' = \frac{\sqrt{2 \times 300}}{2\sqrt{300}} \cdot z_{0.975} = \frac{\sqrt{2} \cdot \sqrt{300}}{2\sqrt{300}} \cdot z_{0.975} = \frac{\sqrt{2}}{2} \cdot z_{0.975} = 0.834$$

[7] Note that in comparison 4 in Table 2, 39 of the 502 audit failure cases were dropped due to inability to match these firm/years with those of a non-Big 8/4 control group.

Table 4 presents descriptive statistics, including the mean, median, and standard deviation for total assets (TA), stock prices (P), earnings per share (EPS), and book value of equity per share (BVPS) for both groups in each comparison. Table 4 also includes a comparison of the means for each of these variables. Since the distributions of these variables might not be normal, both a two-sample t test and a nonparametric Wilcoxon test are conducted. In general, the audit failure group and the matched control group are not significantly different in total assets, which reflects a successful control for company size. Not surprisingly, in comparison 4, because of the special matching requirements, the audit failure and non-audit failure groups exhibit different characteristics in firm size. As stated earlier, in cases where there are no companies available in the same industries with the similar firm sizes as audit failure observations, we select matched-pairs with the closest total assets. Consistent with the tendency of larger firms to have Big 8/4 auditors, comparison 4 shows that the average company size of the audit failure group is significantly larger than that of the matched control group. In comparison 5, the Wilcoxon test results in a significant difference in company size, while the two-sample t test does not indicate a significant difference. The control firms also appear to have better performance compared with audit failure firms. With the exception of comparison 3, the matched control groups have a higher EPS than the audit failure group.

Insert Table 4 here

Testing the Hypothesis Using Bootstrap Confidence Intervals

For the purpose of the current study, 1,000 bootstrap resamples were created by sampling with replacement from each of the original samples. Each bootstrap resample is the same size as the original sample.[8] For each bootstrap resample, there is one R^2 generated from the bootstrap regression, which is called R^{2*}. Therefore, the 1,000 bootstrap resamples generated 1,000 R^{2*}s.[9]

To test whether R^2s of the audit failure group and the matched control group differ, we can compare the 83.4% bootstrap percentile confidence intervals of R^2s. Figures 2 through 6 present the histograms which provide information about the distribution of R^2s based on bootstrap resamples. The histograms of R^{2*}s sometimes indicate approximate normality, but other times show an approximately normal central region with non-normal tails. More importantly, there are three cases where even the central region is decidedly non-normal. Of these three, two cases are the distributions of R^2s for the audit failure groups audited by non-Big 8/4 auditors in comparison 3 and comparison 5, which exhibit an asymmetric pattern because most R^2s from the bootstrap resamples are quite small and very close to zero. The third case appears in comparison 5, where R^2s for the matched control group audited by Big 8/4 auditors appear to have a bimodal distribution, a probability distribution characterized by two peaks or humps rather than the more common single peak that characterizes the normal distribution. One peak is near 0.15 and the other peak is near 0.46. Since we run bootstrap regressions by resampling the residuals from the original regression, the bimodal distribution of R^2 appears to be due to an outlier in the residuals. Examining the regression output, we find one residual with a very extreme value. The bimodality exists when a highly influential point is included in

---

[8] For example, for the original sample size of 616 firm/years, the bootstrap resamples also will contain 616 observations, but will not be identical to the set of observations in the original sample.

[9] See McCullough and Vinod (1993) for details on implementing the bootstrap procedure.

some bootstrap resamples but not in others. When we delete this outlier, the distribution of R^2 exhibits unimodality. See the histogram of R^2 after removing the outlier in Figure 6.

Insert Figures 2-6 here

As stated earlier, one advantage of bootstrap methods is that they do not require distributions to be normal. Hence, the above descriptions of non-normal behavior are not troublesome. To create 83.4% bootstrap percentile confidence intervals, the 1,000 R^{2*}s from bootstrap regressions were sorted in ascending order. The lower value of the confidence interval is the 83rd (0.083*1000) R^{2*} and the upper value is the 917th (0.917*1000) R^{2*}. The results of the bootstrap percentile confidence intervals are shown in Table 5. Comparisons of bootstrap percentile confidence intervals indicate significant differences in R^2s for the audit failure group and the matched control group.[10]

Insert Table 5 here

Testing the Hypothesis Using the Cramer Procedure

In order to be consistent with prior literature (Harris et al. 1994; Lang et al. 2003; Sami and Zhou 2004), we also perform the test using the Cramer procedure. Cramer (1987) provides a procedure to calculate the first moment and second moment of R^2. Once we have the mean and variance of R^2, we can calculate the z-statistic as in Harris et al. (1994) and Lang et al. (2003) and compare means of R^2s across two samples. The expression of a z-test can be modified as follows since the number of observations of R^2 is only one for each sample:

[10] In comparison 5, the result is not significant before the outlier is removed. The confidence interval, (0.0822, 0.5369), is wide when the outlier is included in the sample because of the bimodal distribution of R^2.

$$z = \frac{E\left(R_1^2\right) - E\left(R_2^2\right)}{\sqrt{V\left(R_1^2\right) + V\left(R_2^2\right)}}$$

The results using the Cramer procedure and z-test are presented in Table 6. These results are similar to our results employing the bootstrap method. Therefore, our results are robust to different testing methods[11]. However, as discussed in McCullough and Dang (2004), caution should be given when using the Cramer procedure to calculate the mean and standard deviation of the R^2. We did encounter computer operation overflows, resulting in incorrect outputs, before we modified our programming to resolve such problems.

Insert Table 6 here

Sensitivity Analysis

In order to ensure that our analyses are not affected by potentially confounding factors, we also perform several additional analyses. First, as shown in descriptive statistics, the audit failure group on average shows a lower EPS compared with the matched control group. According to prior literature (e.g., Hayn 1995), losses are less informative for prices than profits. To assure that our results are not driven by the relatively higher frequency of losses in the audit failure group, we also analyze the data after removing all these observations with negative EPS and their pairs in the matched control group. Second, in our original analysis, we include material misstatements that received both “unqualified opinion” and “unqualified opinion with explanatory language” as audit failure cases. Some of the “explanatory language” cases refer to an auditor’s

[11] If the results are different using these two methods, the bootstrap method should be more reliable since the z-test requires normality of the R^2 and the distribution of R^2 is not exactly normal.

substantial doubts about a firm's ability to continue as a going concern. Although the "going concern" opinion also indicates audit failure, insofar as the auditor's opinion on the fairness of the firm's financial statements is not qualified, it is possible that firms that receive going concern opinions are fundamentally different from other firms. To exclude the possibility that our results are driven by this factor, we perform additional analyses after removing all going concern cases and their pairs in the matched control group. Finally, although there are not many, we did find that a few audit failure firms changed their auditors during the period their financial statements were misstated. It is possible that these firms might be different from other firms, e.g., they might be priced differently. Therefore, we analyze the data without those firms that changed auditors during their misstatement years. The results of all additional analyses are consistent with our original analyses, which support our hypothesis.[12]

VI. CONCLUSIONS

In this study, we examine whether the market might assess audit quality accurately based on a value relevance perspective. The bootstrap method is used to create formal tests for differences in R^2s across samples. The results suggest that the value relevance of earnings and book values of equity is lower for firms experiencing apparent audit failures than for firms that have not experienced apparent audit failures. Although the market may exhibit less confidence in accounting information from firms audited by non-Big 8/4 auditors, the results provide evidence that the market generally exhibits less confidence in firms experiencing audit failure, regardless of their auditors' size. These results provide

[12] We deleted 195, 128, and 61 firm/years with negative EPS, unqualified opinion with explanatory language, and multiple auditors during the misstatement years respecitively. Results from additional analyses are not tabulated.

support for the hypothesis that the explanatory power of accounting information of firms experiencing apparent auditor failures is lower than that of firms that have not experienced apparent auditor failures.

The market appears to accurately differentiate *ex ante* between firms with audit quality differences even though an apparent audit failure may not be confirmed by future actions and events until well after its occurrence. Moreover, in comparison 4, accounting information provided by firms experiencing audit failures and audited by Big 8/4 auditors is less value-relevant compared with firms audited by non-Big 8/4 auditors that have not experienced audit failures. This result suggests that the market's ability to evaluate audit quality is not affected by auditor reputation. Thus, as illustrated in comparisons 1 through 5, the results indicate that the market appears to assess audit quality accurately.

In addition to providing evidence suggesting that auditor reputation does not impact the value relevance of accounting information in a dysfunctional manner, this study demonstrates an innovative application of bootstrapping to test for differences in R^2s in the context of value-relevance research. This method may allow researchers to make more precise conclusions when conducting future value-relevance studies.

REFERENCES

Amir, E. 1996. The effect of accounting aggregation on the value-relevance of financial disclosures: the case of SFAS No. 106. *The Accounting Review*. Vol. 71: 573-590.

______ and B. Lev. 1996. Value-relevance of nonfinancial information: the wireless communications industry. *Journal of Accounting and Economics* 22: 3-30.

Anderson, T. W. 2003. *An Introduction to Multivariate Analysis*. New York. John Wiley and Sons, Inc.: 149-157.

Balsam, S., J. Krishnan, and J. S. Young. 2000. Auditor industry specialization and the earnings response coefficient. Working Paper.

Barr, D. R. 1969. Using confidence intervals to test hypotheses. *Journal of Quality Technology*. Vol. 1: 256-258.

Bauwhede, H. V., M. Willekens, and A. Gaeremynck. 2000. Audit quality, public ownership and firms' discretionary accruals management. Working Paper.

Becker, C. L., M. L. DeFond, J. Jiambalvo, and K. R. Subramanyam. 1998. The effect of audit quality on earnings management. *Contemporary Accounting Research*. Vol. 15, No. 1: 1-21.

Clarkson, P. M. and D. A. Simunic. 1994. The association between audit quality, retained ownership, and firm-specific risk in U. S. vs. Canadian IPO markets. *Journal of Accounting and Economics*. Vol.17: 207-228.

Collins, D.W., E. L. Maydew, and I. S. Weiss. 1997. Changes in the value-relevance of earnings and book values over the past forty years. *Journal of Accounting and Economics* 24: 39-67.

Copley, P. A. 1991. The association between municipal disclosure practices and audit quality. *Journal of Accounting and Public Policy*. Vol. 10, No. 4: 245.

Cramer, J. S. 1987. Mean and variance of R^2 in small and moderate samples. *Journal of* Econometrics 35: 253-266.

Davidson, R. A. and D. Neu. 1993. A note on association between audit firm size and audit quality. *Contemporary Accounting Research*. Vol. 9, No. 2: 479-488.

DeAngelo, L. E. 1981. Auditor size and audit quality. *Journal of Accounting and Economics.* Vol. 3, No. 3: 183-199.

Dechow, P. M. 1994. Accounting earnings and cash flows as measures of firm performance: the role of accounting accruals. *Journal of Accounting and* Economics 18: 3-42.

Francis. J. and K. Schipper. 1999. Have financial statements lost their relevance? *Journal of Accounting Research*. Vol. 37, No. 2: 319-352.

Harris, T. S., M. Lang, and H. P. Moller. 1994. The value relevance of German accounting measures: an empirical analysis. *Journal of Accounting Research*. Vol. 32, No. 2: 187-209.

Hayn, C. 1994. The information content of losses. *Journal of Accounting and Economics.* Vol. 20: 125-153.

Hogan, C. E. 1997. Costs and benefits of audit quality in the IPO market: a self-selection analysis. *The Accounting Review*. Vol. 72, No. 1: 67-86.

Kim, J., R. Chung, and M. Firth. 2003. Auditor conservatism, asymmetric monitoring and earnings management. *Contemporary Accounting Research*. Vol. 20, No. 2: 323-359.

Krishnan, G. V. 2003. Audit quality and the pricing of discretionary accruals. *Auditing: A Journal of Practice and Theory*. Vol. 22, No.1:109-126.

Lam, S. and. S. Chang. 1994. Auditor service quality and auditor size: evidence from initial public offerings in Singapore. *Journal of International Accounting Auditing and Taxation*. Vol. 3, No. 1: Pg. 103.

Lang, M., J. S. Raedy and M. H. Yetman. 2003. How representative are firms that are cross-listed in the United States? An analysis of accounting quality. *Journal of Accounting Research*. Vol. 41, No. 2: 363-386.

Lyon, J. D. and D. A. Schroeder. 1992. Firm growth and the valuation relevance of earnings levels, earnings innovations, and dividends. *Journal of Accounting, Auditing, and Finance*: 531-551

McCullough, B. D. and H. D. Vinod. 1993. Implementing the single bootstrap: some computational considerations. *Computational Economics* 6 (1): 1-15

________________ and L. Dang. 2004. Programming the moments of R^2. Working Paper.

Moreland, K. A. 1995. Criticisms of auditors and the association between earnings and returns of client firms. *Auditing: A Journal of Practice & Theory*. Vol. 14, No. 1: 94-104.

Nichols, D. R. and D. B. Smith. 1983. Auditor credibility and auditor changes. *Journal* *of Accounting Research*. Vol.21, No. 2: 534-544.

Nwaeze, E. T. 1998. Regulation and the valuation relevance of book value and earnings: evidence from the United States. *Contemporary Accounting Research*. Vol. 15, No. 4: 547-573.

Ohlson, J. A. 1995. Earnings, book values, and dividends in equity valuation. *Contemporary Accounting Research*. Vol. 11, No. 2: 661-687.

Palmrose, Z. 1988. An analysis of auditor litigation and audit service quality. *The Accounting Review*. Vol. 64, No.1: 55-73.

__________. 1999. *Empirical Research in Auditor Litigation: Considerations and Data*. Sarasota, Florida: American Accounting Association.

Petroni, K. and M. Beasley. 1996. Errors in accounting estimates and their relation to audit firm type. *Journal of Accounting Research*. Vol. 34, No. 1: 151-171.

Rajgopal, S., M. Venkatachalam, and S. Kotha. 2002. The Value-relevance of network advantages: the case of e-commerce firms. *Journal of Accounting Research*. Vol., 41. No. 1: 135-162.

Rees, W. P. 1999. Influences on the value-relevance of equity and net income in the UK. *Managerial Finance.* Vol. 25 No.12: 58-65.

Sami, H. and H. Zhou. Forthcoming. A comparison of value relevance of accounting information in different segments of the Chinese stock market. *International Journal of Accounting*. Winter Issue.

Stober, T. L. 1999. Empirical applications of the Ohlson [1995] and Feltham and Ohlson [1995, 1996] valuation models. *Managerial Finance*. Vol. 25 No. 12: 3-16.

Tate, S. 2001. Differences in financial statement and compliance audit assessment between size and non-size auditors. Working Paper.

Teoh, S. H. and T. J. Wong. 1993. Perceived auditor quality and the earnings response coefficient. *The Accounting Review*. Vol. 68, No. 2: 346-366.

Zhou, J. and R. Elder. 2001. Audit firm size, industry specialization and earnings management by initial public offering firms. Working Paper.

Table 1

Comparisons Used to Test the Hypothesis

Comparison	Failure Group (AF) Auditors	Non-Failure Group (NAF) Auditors	Results and Implications	
			If $R^2_{AF} < R^2_{NAF}$	If $R^2_{AF} >= R^2_{NAF}$
1	Both Big 8/4 and non-Big 8/4 auditors	Both Big 8/4 and non-Big 8/4 auditors	The market appears to assess audit quality accurately	The market does not appear to assess audit quality accurately
2	Big 8/4 auditors	Big 8/4 auditors	The market appears to assess audit quality accurately for Big 8/4 auditors	The market does not appear to assess audit quality accurately for Big 8/4 auditors
3	Non-Big 8/4 auditors	Non-Big 8/4 auditors	The market appears to assess quality accurately for non-Big 8/4 auditors	The market does not appear to assess audit quality accurately for non-Big 8/4 auditors
4	Big 8/4 auditors	Non-Big 8/4 auditors	The market appears to assess audit quality accurately regardless of auditor reputation	Market assessment appears to be influenced by auditor reputation in a dysfunctional manner
5	Non-Big 8/4 auditors	Big 8/4 auditors	The market appears to assess audit quality accurately regardless of auditor reputation	Market assessment appears to be influenced by auditor reputation in a dysfunctional manner

Table 2

Sample Determination and Industry Distribution

Panel A: Sample Determination for Apparent Audit Failure Cases

	number of firms	number of firm/years
Audit failure cases:	416	779
Less: observations without price data:	70	163
Sample for testing	**346**	**616**

Panel B: Industry Distribution for Apparent Audit Failure Cases

Industry	**SIC Code**	**Number of Observations**	**Percentage**
Agriculture, Forestry, Fishing	01-09	1	0.29%
Mining	10-14	15	4.34%
Construction	15-17	7	2.02%
Manufacturing	20-39	127	36.71%
Transportation & Public Utilities	40-49	31	8.96%
Wholesale Trade	50-51	12	3.47%
Retail Trade	52-59	38	10.98%
Finance, Insurance, Real Estate	60-67	28	8.09%
Services	70-89	82	23.70%
Public Administration	90-99	5	1.45%
Total		346	100.00%

Table 3

Number of Firm/Year Observations in Each Comparison

	Audit Failure Group	Matched Control Group
	Comparison 1	
Number of Observations	616	616
Big auditors	502	502
Non-Big auditors	114	114
	Comparison 2	
Number of Observations	502	502
Big auditors	502	502
Non-Big auditors	0	0
	Comparison 3	
Number of Observations	114	114
Big auditors	0	0
Non-Big auditors	114	114
	Comparison 4	
Number of Observations	463	463
Big auditors	463	0
Non-Big auditors	0	463
	Comparison 5	
Number of Observations	114	114
Big auditors	0	114
Non-Big auditors	114	0

Table 4
Descriptive Statistics for Observations Included in Hypothesis Testing

Comparison 1

	Auditor Failure Group			Matched Control Group			Compare Means[a]	
	Mean	Median	Stddev	Mean	Median	Stddev	t test	Wilcoxon Test
TA	2421.1010	135.5100	8309.7470	1860.2880	114.5100	6508.7950	0.1875	0.332
P	17.4288	12.2500	17.7024	18.3725	12.0630	22.4992	0.4135	0.9602
EPS	-0.0065	0.3450	7.0179	0.7458	0.5000	5.7049	**0.0392**[b]	**0.0022**
BVPS	7.6564	12.2500	9.7234	8.6550	5.7230	16.2675	0.1913	0.2421

Comparison 2

	Auditor Failure Group			Matched Control Group			Compare Means	
	Mean	Median	Stddev	Mean	Median	Stddev	t test	Wilcoxon Test
TA	2958.4940	224.6000	9121.2050	2243.6640	182.2900	7122.9120	0.1667	0.2256
P	19.4984	14.7500	17.9248	21.2554	15.2500	23.8012	0.1867	0.5052
EPS	0.01631	0.4900	7.7616	0.9306	0.7850	6.2850	**0.0405**	**0.0007**
BVPS	8.7918	6.2610	10.2356	9.9054	6.9360	17.6025	0.2208	0.3822

Comparison 3

	Auditor Failure Group			Matched Control Group			Compare Means	
	Mean	Median	Stddev	Mean	Median	Stddev	t test	Wilcoxon Test
TA	54.6836	16.7150	138.4634	172.0858	15.0150	1447.4700	0.3904	0.9080
P	8.3157	4.0000	13.3492	5.6772	3.4380	6.6272	0.0605	0.2113
EPS	-0.1072	-0.0100	0.9703	-0.0679	0.0200	1.0853	0.7734	0.8574
BVPS	3.1489	1.1916	4.3721	2.6570	1.7497	5.3825	0.4496	0.0890

Comparison 4

	Auditor Failure Group			Matched Control Group			Compare Means	
	Mean	Median	Stddev	Mean	Median	Stddev	t test	Wilcoxon Test
TA	3008.3940	211.5500	9391.1080	1393.2470	64.5000	9913.3340	**0.0111**	**0.0000**
P	19.6246	14.5000	18.3666	15.1947	8.8750	22.0622	**0.0009**	**0.0000**
EPS	-0.0472	0.4700	8.0709	0.5576	0.3700	1.6394	0.1147	0.9536
BVPS	8.5995	6.1680	10.4589	7.8357	4.9530	13.6130	0.3386	**0.0002**

Comparison 5

	Auditor Failure Group			Matched Control Group			Compare Means	
	Mean	Median	Stddev	Mean	Median	Stddev	t test	Wilcoxon Test
TA	54.6836	16.7150	138.4634	46.9001	23.3100	70.6154	0.5932	**0.0426**
P	8.3157	4.0000	13.3492	10.3906	6.7500	14.9525	0.2691	**0.0038**
EPS	-0.1072	-0.0100	0.9703	0.1656	0.0900	1.3441	0.0796	**0.0309**
BVPS	3.1489	1.1916	4.3721	4.7300	3.3794	5.4881	0.0018	**0.0000**

TA: Total asset at the end of the fiscal year t;
P: Stock price, 3 month after the end of the fiscal year t;
EPS: Earnings per share excluding extraordinary items for fiscal year t;
BVPS: Book value of equity per share at the end of the fiscal year t;

a. Both p-values of the two-sample t test and the nonparametric two-sample Wilcoxonon test are given.

Note: all p-values less than or equal to 0.05 indicated in boldface.

Table 5
Comparisons of Bootstrap Percentile Confidence Intervals for R^2s
Regression Model: $P_{it} = \alpha_1 EPS_{it} + \alpha_2 BVPS_{it}$

Comparison		Failure Group	Non-Failure Group
1	Auditor	Both Big and non-Big auditor	Both Big and non-Big auditor
	83% Bootstrap Percentile CI	**(0.1344, 0.2234)**	**(0.3555, 0.4825)**
2	Auditor	Big auditor	Big auditor
	83% Bootstrap Percentile CI	**(0.1175, 0.2072)**	**(0.3520, 0.4877)**
3	Auditor	Non-Big auditor	Non-Big auditor
	83% Bootstrap Percentile CI	**(0.0068, 0.1161)**	**(0.2640, 0.5030)**
4	Auditor	Big auditor	Non-Big auditor
	83% Bootstrap Percentile CI	**(0.1229, 0.2207)**	**(0.3965, 0.6054)**
5	Auditor	Non-Big auditor	Big auditor
	83% Bootstrap Percentile CI	**(0.0068, 0.1161)**	**(0.2045, 0.3894)**

Note: all p-values less than or equal to 0.05 indicated in boldface.

Table 6

Comparison of R^2s Using a z-test Based on the Cramer Procedure

Regression Model: $P_{it} = \alpha_1 EPS_{it} + \alpha_2 BVPS_{it}$

Comparison		Failure Group	Non-Failure Group
1	Auditor	Both Big non-Big auditor	Both Big non-Big auditor
	n	616	616
	R^2	0.1714	0.4179
	Mean(R^2)	0.1735	0.4190
	Stddev(R^2)	0.0264	0.0270
	z-statistic	**6.4977**	
2[a]	Auditor	Big auditor	Big auditor
	n	502	502
	R^2	0.1555	0.4129
	Mean(R^2)	0.1582	0.4142
	Stddev(R^2)	0.0286	0.0210
	z-statistic	**6.1826**	
3[b]	Auditor	Non-Big auditor	Non-Big auditor
	n	114	114
	R^2	0.0266	0.3555
	Mean(R^2)	0.0353	0.3624
	Stddev(R^2)	0.0369	0.0651
	z-statistic	**4.3668**	
4[a]	Auditor	Big auditor	Non-Big auditor
	n	463	463
	R^2	0.1628	0.4827
	Mean(R^2)	0.1656	0.4839
	Stddev(R^2))	0.0301	0.0291
	z-statistic	**7.6015**	
5[b]	Auditor	Non-Big auditor	Big auditor
	n	113[c]	113[c]
	R^2	0.0284	0.2862
	Mean(R^2)	0.0381	0.2946
	Stddev(R^2)	0.0369	0.0663
	z-statistic	**3.3805**	

a. In comparison 2 and 4, audit failure group is the same group except the number of observations differs. Failure group in both comparisons represents firms audited by Big auditors. In comparison 4, we dropped some failure group firms since we cannot find the matched pairs audited by non-Big auditors.

b. In comparison 3 and 5, audit failure group is the same group since it represents audit failure firms audited by non-Big auditors.

c. The number of observations is due to deleting the outlier existing in the matched control group and its pair in the audit failure group. Similar to the bootstrap testing results, z-statistic is not significant before deleting the outlier because the large magnitude of the variance of the R^2 for the matched control group, 0.1939.

Note: all p-values less than or equal to 0.05 indicated in boldface.

Figure 1

An Example of an R^2 Distribution

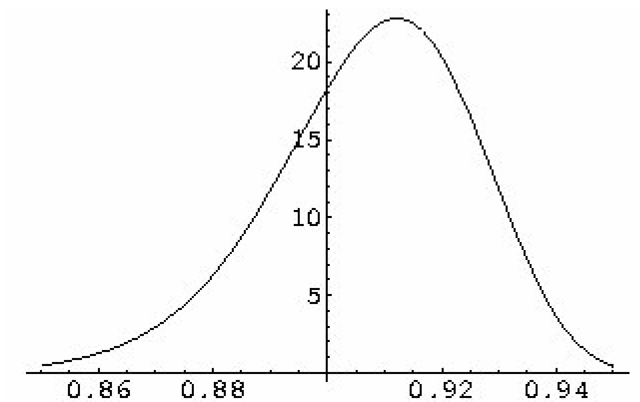
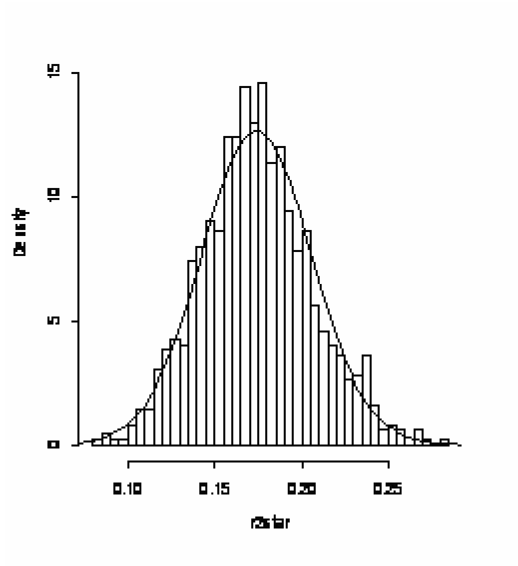
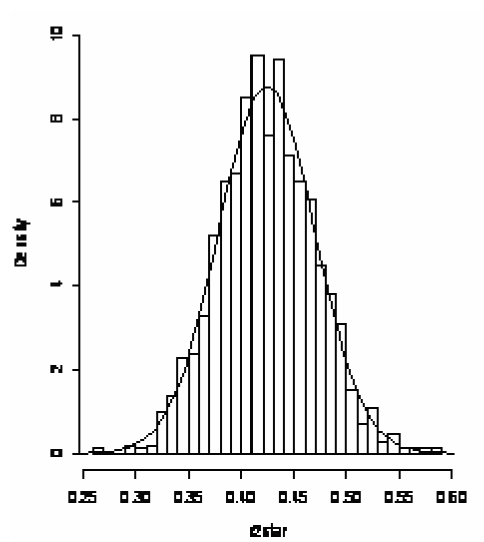


Figure 2

Histograms for Comparison 1



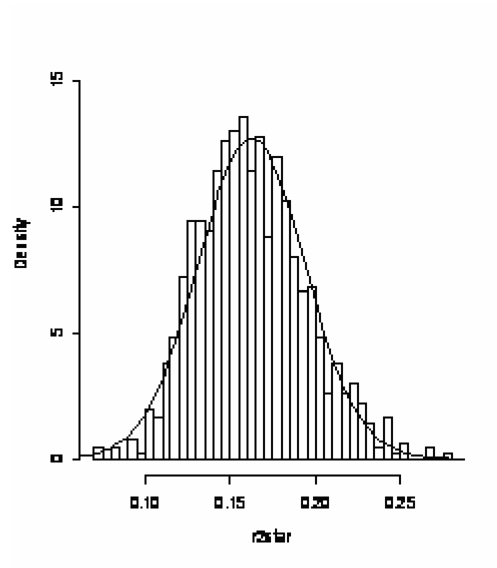
R^{2*}: Audit Failure Group---Both Big Auditors & Non-Big Auditors



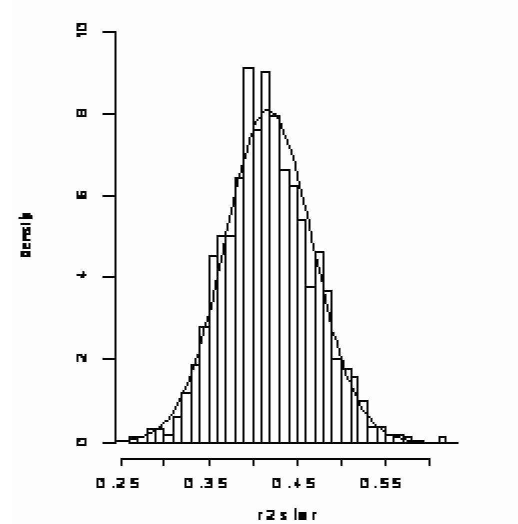
R^{2*}: Matched Control Group---Both Big Auditors and Non-Big Auditors

Figure 3

Histograms Comparison 2



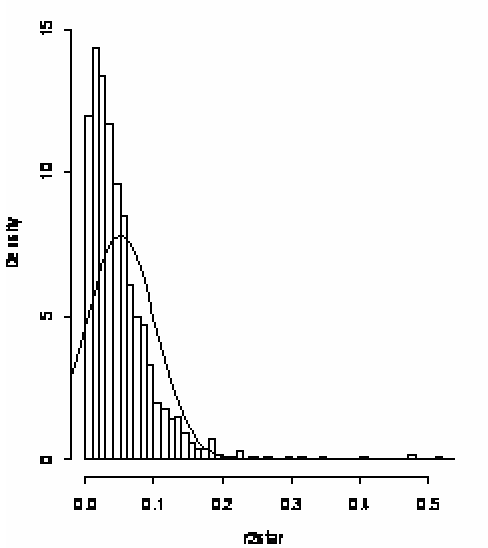
R^{2*}: Audit Failure Group---Big Auditors



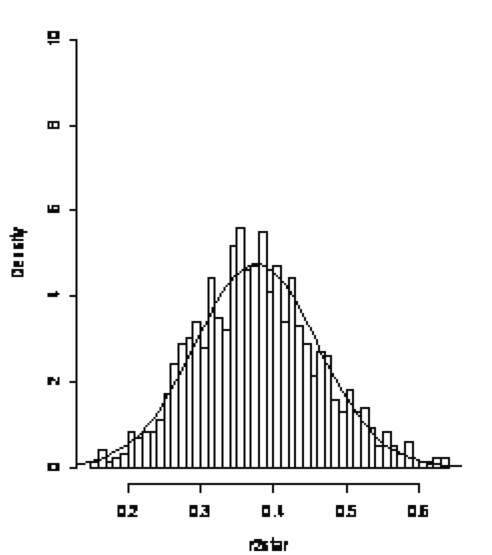
R^{2*}: Matched Control Failure Group-Big Auditors

Figure 4

Histograms Comparison 3



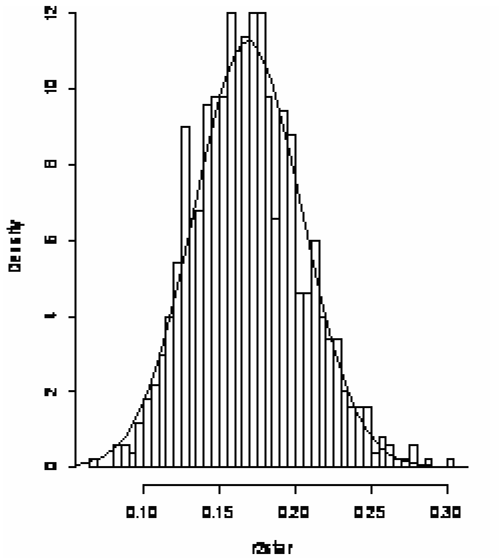
R^{2*}: Audit Failure Group---Non-Big Auditors



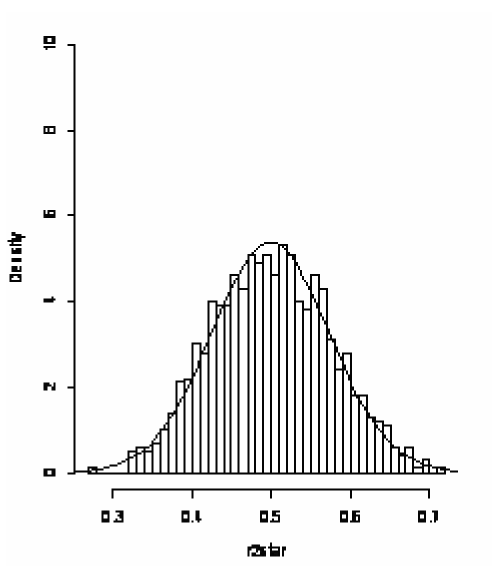
R^{2*}: Matched Control Group---Non-Big Auditors

Figure 5

Histograms Comparison 4



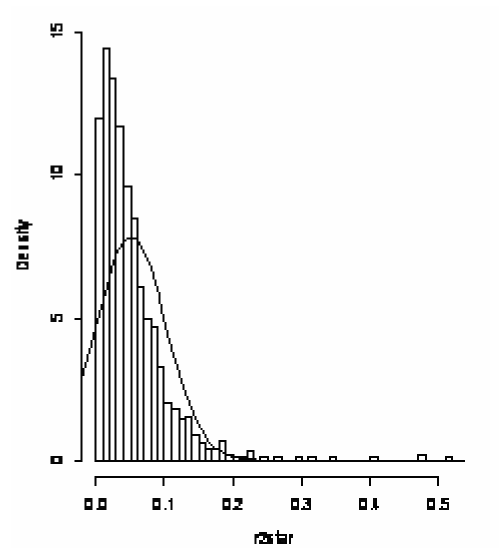
R^{2*}: Audit Failure Group---Big Auditors



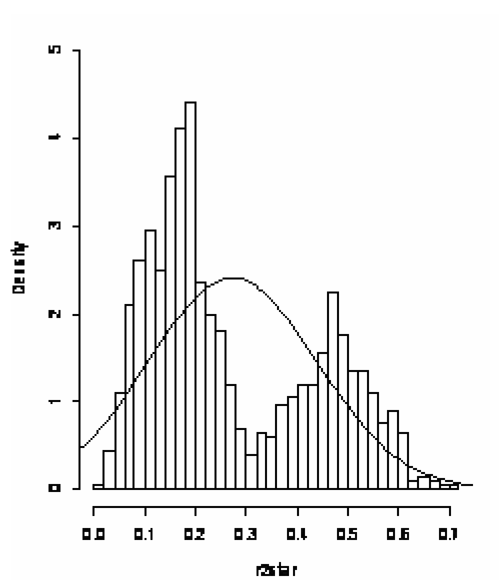
R^{2*}: Matched Control Group---Non-Big Auditors

Figure 6

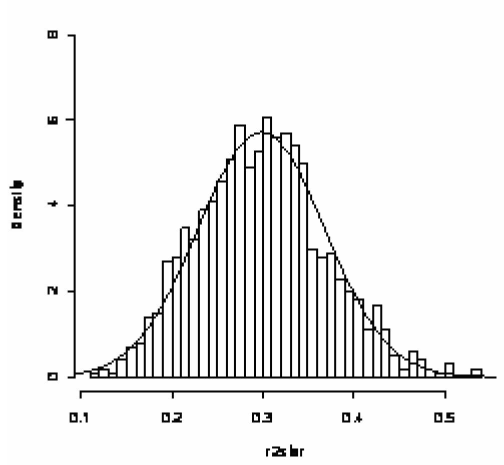
Histograms Comparison 5



R^{2*}: Audit Failure Group---Non-Big Auditors



R^{2*}: Matched Control Group---Big Auditors



R^{2*}: Matched Control Group---Big Auditors
After Removing the Outlier