The Effects of Sarbanes-Oxley on Reporting Profits

David Aron Smith*
Associate at Analysis Group, Chicago, USA

Abstract
This paper looks at the effects of Sarbanes-Oxley on the practice known as Earnings Management. Using earnings as reported in quarterly corporate filings, I employ a test of discontinuity used by DeGeorge, Patel and Zeckhauser to estimate the discontinuity in reported earnings per share (EPS) following a major accounting reform resulting from Sarbanes-Oxley. I find that the previous discontinuities found in the literature in reported EPS at 0¢ and 1¢ disappear after the accounting reform. Using a standard market model on security prices, I also find that short-term abnormal returns to reporting 1¢ EPS decrease following this reform, which could indicate positive short-term returns to Earnings Management.

Keywords: Sarbanes-Oxley; Earnings per share; Accounting reform; Earnings management

JEL Classification: D03, D22, G02, G18, K22

Introduction
In 1998, then-chairman of the Securities and Exchange Commission (SEC) Arthur Levitt Jr. drew attention to the common practice of Earnings Management (EM), whereby companies use flexibility in accounting rules to make short-term adjustments to profits [1]. Burgstahler and Dichev and DeGeorge [2], Patel and Zeckhauser [3] (hereafter DPZ) were among the first to argue that profits [1]. Burgstahler and Dichev and DeGeorge [2], Patel and Zeckhauser [3] (hereafter DPZ) were among the first to argue that

SOX was passed to “protect investors by improving the accuracy and reliability of corporate disclosures” [4]. In 2004, the Public Company Accounting Oversight Board (PCAOB), created by SOX, issued Auditing Standard No. 1 (AS1) in attempt to standardize accounting practices. This standard was effective beginning May 24, 2004. Previous studies such as Bartov and Cohen [5] and Li, Pincus, and Rego [6] investigate the effects of SOX on earnings expectations management and the stock price of firms that had previously managed earnings. Davis, Soo, and Trompeter [7] find evidence consistent with a decline in the use of discretionary accruals to meet or beat earnings forecasts post-SOX. Singer and You [8] find that the quality of reported earnings increased significantly due to a provision of Auditing Standard No. 2. However, to my knowledge no studies have examined how the EM practice of “avoiding red ink” has been affected by SOX. For simplicity, I hereafter refer mainly to SOX rather than AS1 except where the distinction is necessary.

Figure 1 shows the truncated distributions of reported EPS for roughly seven years before and after SOX. Before SOX, there is a noticeable dip at -1¢ and a jump at 0¢ and 1¢ EPS. After SOX, the dip at -1¢ is absent, the jump at 0¢ has diminished, and although there is still a peak at 1¢ EPS, it is much less pronounced. To go about statistically determining a discontinuity, I use a form of a t-test, which is used by DPZ. On the full pre-SOX sample, I find significant discontinuities at -1¢, 0¢, and 1¢ EPS, which indicate that fewer firms report -1¢ EPS than would be expected under a smooth EPS distribution, while more firms than expected report 0¢ and 1¢ EPS—consistent with the literature on earlier data. Post-SOX, I find no significant discontinuities, indicating that the current distribution better resembles a smooth distribution around 0¢.

Lobo and Zhou [9], the mainstream media, and others have argued that it became more costly for managers to engage in EM activities after SOX due to increased regulatory scrutiny and stricter regulatory violations. AS1 in particular standardized accounting guidelines, reducing some flexibility that existed previously, increasing the shadow costs to EM through stringent accounting guidelines and heightened vigilance of auditing firms and regulators. This paper presents evidence consistent with the hypothesis that fewer firms have engaged in “avoiding red ink” since AS1 took effect. If this hypothesis is valid, it provides support for the argument that SOX has improved the accuracy of reported EPS, particularly around 0¢—or at least that it has increased the standardization of such figures.

The results of a mean comparison of short-term abnormal return to firms reporting 0¢ and 1¢ EPS before and after SOX indicate that...
returns have decreased for the pool of firms that report 0¢ and 1¢ EPS. If we accept the results of the test of discontinuity indicating that fewer firms manage earnings around 0¢ EPS, then the decrease in short-term abnormal returns for all firms reporting 1¢ EPS could indicate that the firms that manage earnings see a higher return than firms that don’t. However, changes over time in the returns to EM could cloud this result.

Earnings Management in the Literature

At the root of “avoiding red ink” is prospect theory and loss aversion [10], which suggests that the largest increase in utility is made when moving from a loss to a gain relative to a particular reference point. This is due to a kink in the utility function at the reference point, with concavity in gains and convexity in losses. If loss averse investors focus on EPS with a reference point at 0¢ EPS, then the largest incentives to manage earnings are from a slightly negative EPS to 0¢ or 1¢ EPS. Traditionally, loss aversion and reference-dependent preferences have been used in finance to help explain the equity premium puzzle and disposition effects. In fact, DellaVigna [11] notes that these are “two of the most important applications of reference-dependent preferences”. See Baker and Wurgler [12] for a survey of recent publications, as well as for a broader review of the behavioral finance literature.

My analysis is most similar to Keung, Lin, and Shih [13], who find that reporting 1¢ EPS is a bad signal to investors. However, their regression on returns includes earnings surprises scaled by share price, and as a test for investor skepticism, they look at whether $\text{EPS}_t - \text{EPS}_t^+$ falls in the range [0, 1¢]. I assume investors cannot identify which firms are engaging in EM, which allows me to put aside issues of using earnings surprises and correlations of earnings over time. More importantly, their paper does not discuss the effects of SOX, which is the primary purpose of this paper. They do find that skepticism has increased over time, finding significant skepticism at the 1% level in 2002-2006. However, as Bartov and Cohen argue, earnings surprises became increasingly manipulated post-SOX, as firms increasingly managed analyst expectations, which could cause the earnings surprise variable to be endogenously determined. Davis, Soo, and Trompeter find that the use of discretionary accruals to meet or beat analyst expectations declines post-SOX. However, none of these papers, nor any other of which I am aware, shows positive abnormal returns for firms that engage in EM.

Durschi and Easton [14] contend that kinks in the EPS distribution are not ipso facto evidence of EM. This is technically true. However, it is difficult to imagine what data would provide ipso facto evidence. While EM is not the only explanation for the kink in the distribution, it is a plausible and straight-forward one. Those authors explain that kinks can exist in the distribution without EM. However, I do not believe those arguments are relevant for this study. If EM is not the underlying reason, another explanation is required that would explain the smoothing in the EPS distribution precisely after regulatory actions intended to reduce EM. To refuse inference to the best explanation in this case is to prefer no answer over a plausible one.

Data and Empirical Methodology

Compustat data

I use a Compustat dataset that consists of quarterly data on 27,584 firms providing partial or complete financial reports over the 1974-2011 period. While the total number of observations exceeds 1,211,710, the number of available observations is much smaller for my analyses. I focus on Compustat item EPSFXQ—diluted earnings per share excluding extraordinary items, as reported—to best capture the figures most likely scrutinized by investors. This is also the item commonly used in the literature. Any attempt at EM by firms using extraordinary items will not be captured by these data. However, according to I/B/E/S, analysts’ earnings forecasts do not include unusual or non-recurring charges. Therefore, it is unlikely that firms are using extraordinary items to engage in EM.

The practice of deflating EPS by price per share is criticized by Durschi and Easton and DPZ, as it can create a buildup in the density around zero, which is also our point of interest. By excluding firms with extreme prices, EPS deflation is no longer necessary. As such, I exclude firms with prices at the 10% extremes, consistent with DPZ. Following Burgstahler and Dichev, I remove regulated industries (SIC codes 4400-5000) and financial institutions (SIC codes 6000-6500). Summary statistics for the sample can be found in Table 1. Table 2 reports selected summary statistics for EPS in the sample. The summary statistics are consistent with the hypothesis that fewer firms are managing earnings around 0¢ EPS after SOX.

Test of discontinuity

As a test of discontinuity, I use a form of a t-test for univariate distributions used by DPZ which I refer to as the r-test and describe in...
I estimate the following market model over the 255-day period ending measure abnormal stock returns around the earnings announcement, from the Center for Research in Security Prices (CRSP) database. To

\[ R_{jt} = \alpha_j + \beta_j R_{mt} + \epsilon_{jt}, \]

where \( R_{jt} \) is the stock return for firm \( j \) on day \( t \), and \( R_{mt} \) is the weighted market return on day \( t \). The results of this market model are used to predict daily returns

\[ AR_{jt} = R_{jt} - R_{mt}, \]

which is in turn used to estimate daily abnormal returns using the equation:

\[ AR_j = R_j - R_{mt}, \]

where \( AR_j = 0 \) indicates no abnormal return. Borrowing the notation of Keung, Lin, and Shih (2010), I define cumulative abnormal returns for firm \( j \) from \( b \) days before the earnings announcement to \( a \) days after as

\[ CAR_{[b,a]} = \left( \prod_{t=b}^{a} AR_j + 1 \right) - 1 \]

Here, \( CAR_{[b,a]} = 0 \) indicates no cumulative abnormal return. The market model is run on firm observations after January 1, 1997 with at least 50 trading days of data in the relevant time frame.

### Results

#### Test of discontinuity

The results of the \( \tau \)-test are summarized in Table 3. I find that the measure of discontinuity diminished significantly at both 0¢ and 1¢ EPS after SOX. Specifically, I find that \( r(0¢) \) falls to 0.32, from 7.38 in the 1997-2004 data and 7.68 in the 1974-2004 data. I also find that \( r(1¢) \) falls to 0.21, from 3.97 in the 1997-2004 data and 3.55 in the 1974-2004 data. Before SOX, there appear to be too few companies reporting -1¢ EPS, given the high values for \( r(-1¢) \), whereas post-SOX there are roughly the number of firms reporting -1¢ EPS as we would predict given a smooth distribution. This is consistent with the hypothesis that some firms with latent EPS of -1¢ were able to manage their earnings upward to an EPS of 1¢ using the accounting flexibility allowed at the time, and that fewer firms are doing so today.

#### Returns to reporting 1¢ EPS before and after SOX

Tables 4 and 5 show the results of a simple \( t \)-test on the equality of the mean abnormal returns for firms reporting 0¢ and 1¢ EPS before and after AS1 was in effect. The test allows for the two samples to have unequal variances. There is a small decrease in returns for the pool of firms that report 1¢ EPS post-SOX, and this decrease is significant at the 1% level. This is true over a very short window of three days

### Summary Statistics of EPSFXQ.

<table>
<thead>
<tr>
<th>Time Period</th>
<th>(-2)</th>
<th>(-1)</th>
<th>0¢</th>
<th>1¢</th>
<th>2¢</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of times reported</td>
<td>1,337</td>
<td>1,248</td>
<td>1,746</td>
<td>2,379</td>
<td>2,349</td>
</tr>
<tr>
<td>Number of reporting firms</td>
<td>1,907</td>
<td>1,788</td>
<td>2,410</td>
<td>3,140</td>
<td>3,044</td>
</tr>
<tr>
<td>Pre-SOX</td>
<td>1,172</td>
<td>1,042</td>
<td>1,285</td>
<td>1,481</td>
<td>1,508</td>
</tr>
<tr>
<td>Number of times reported</td>
<td>1,882</td>
<td>2,055</td>
<td>2,218</td>
<td>2,357</td>
<td>2,263</td>
</tr>
<tr>
<td>Number of reporting firms</td>
<td>984</td>
<td>1,052</td>
<td>1,170</td>
<td>1,146</td>
<td>1,082</td>
</tr>
<tr>
<td>Source: Compustat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Sample excludes Financial Institutions (SIC codes 6000-6500), Regulated Industries (SIC codes 4400-5000), and companies in the top and bottom 10% of prices. The "Pre-SOX" period indicates Jan 1997 thru May 23, 2004, and the "Post-SOX" period indicates May 24, 2004 thru Dec 2011.

### Results of t-test.

<table>
<thead>
<tr>
<th>Time Period</th>
<th>(-2)</th>
<th>(-1)</th>
<th>0¢</th>
<th>1¢</th>
<th>2¢</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974 - May 2004</td>
<td>1.86</td>
<td>2.49</td>
<td>7.38</td>
<td>3.55</td>
<td>0.17</td>
</tr>
<tr>
<td>Pre-SOX</td>
<td>2.97</td>
<td>4.25</td>
<td>7.68</td>
<td>3.97</td>
<td>0.17</td>
</tr>
<tr>
<td>Post-SOX</td>
<td>2.16</td>
<td>0.51</td>
<td>3.22</td>
<td>0.21</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Notes: Only firms with reported earnings per share of 0¢ are included. Negative differences indicate higher returns before SOX.

I find different values of \( \tau \) than DPZ for 1974-1996. I cannot fully account for the differences in my results, other than to note that my sample contains more firms. It is possible that the DPZ eliminated some firms in order to meet the criteria for their other tests that were unrelated to "avoiding red ink," although they do not explicitly note that in their paper. Additionally, as my distribution peak is at 1¢ EPS, I perform a slightly different version of the t-test (as described in Appendix A), which could account for some of the difference.
surrounding the earnings announcement, as well as for nine days surrounding the earnings announcement. While there is a decrease in returns for the pool of firms that report 0¢ EPS post-SOX as well, it is not significant at the 5% level.

This evidence is consistent with the hypothesis that firms that manage earnings earn a higher return on average than firms that do not. The intuition is as follows: with semi-strong market efficiency, investors will accurately forecast the latent earnings of firms on average. However, if a firm manages earnings upward to 0¢ or 1¢ EPS from negative earnings, investors will be pleasantly surprised, on average. Although both populations theoretically contain a pool of firms that have managed earnings and those that have not, the results of the discontinuity test above are consistent with the hypothesis that there are fewer firms managing their earnings in the post-SOX pool.

### Conclusion

The results above show that SOX has resulted in a smoothing of the EPS distribution around 0¢ EPS, indicating that fewer firms are reporting 0¢ or 1¢ EPS relative to the rest of the distribution. However, it is important to note that the distribution does not itself indicate earnings management, as some authors have pointed out. Still, a reduction in EM around 0 is an intuitive explanation for the distribution smoothing evident after SOX. The change in abnormal returns after SOX combined with the smoothing of the EPS distribution around 0¢ could indicate positive short-term returns to EM, which has yet to be shown in the literature. Whether any abnormal returns are reversed or maintained in the long-run is left for future study.

### Appendices

#### Testing for discontinuity in a univariate distribution

Although not controversial in this particular setting, the decision of bin width for a test of discontinuity is an important one, particularly because discontinuity tests can be sensitive to width choice. I am not aware of any variation in the literature from the intuitive choice of 1¢. This is consistent with Silverman [29] and Scott [30], who recommend the Freedman-Diaconis rule, for a bid width of \(2 \times IQR \times n^{-1/3}\), where IQR is the sample interquartile range of the variable in question (in this case EPS). Given the high degree of outliers in the data, this is generally preferred to Scott's normal reference rule. Although 1¢ is a larger width than the formula implies, it is the minimum resolution for the data.

As this test is taken from DPZ, the description and notation is largely borrowed from their original paper. Let \(x\) be the variable of interest, such as earnings per share. The null hypothesis is that the probability density function of \(x\), call it \(f(x)\), is smooth at \(T\), a point of interest. The alternative hypothesis is that point \(T\) is a threshold, a point of discontinuity in \(f(x)\). Given a random sample of \(x\) of size \(N\), I estimate the density for discrete ordered points \(x_0, x_1, \ldots, x_n\), and so forth. I suppose these points are equispaced (as they are in our sample) and normalize the distance between the points to be of length one. Compute the proportion of observations that lie in bins covering \((x_0, x_1), (x_1, x_2), \ldots, (x_n-1, x_n)\), and so forth. These proportions, \(p(x)_i\), provide empirical estimates of \(f(x)\) at \(x_0, x_1, \ldots, x_n\).

#### Basic test

Define \(\Delta p(x_i) = \log[p(x_i)] - \log[p(x_{i-1})]\). The expectation of \(\Delta p(x_i)\) is \(f'(x)\), and its variance depends on the higher derivatives at \(x_i\), as well as the available sample size \(N\). Consider a small region \(R_i\) around \(n\) of \(2r+1\) points (i.e., \(R_i = [x_i, i \in (n-r, n+r)]\)). Given our null hypothesis assumption regarding the smoothness of \(f(x)\), the distribution of \(\Delta p(x_i)\) will be approximately homogeneous.5

Using observations from \(R_i\), excluding \(\Delta p(x_i)\), compute a \(t\)-like test statistic,

\[
\tau(n) = \frac{\Delta p(x_i) - \text{mean}_{i} \{\Delta p(x_i)\}}{\text{s.d.}_{i} \{\Delta p(x_i)\}}
\]

where mean and s.d. denote the sample mean and standard deviation of the bracketed terms. I exclude the observations \(i = n\) to the sample means and standard deviations to increase the power in identifying a discontinuity at \(x_i\).

The alternative hypothesis conjectures a discontinuity at a preidentified threshold \(T\). (For our tests, \(T\) takes on both zero and 1¢.) The distribution of \(\tau(T)\) is likely to be well approximated by the Student’s \(t\)-distribution under the null hypothesis if the distribution of \(\Delta p(x_i)\) in \(R_i\) is approximately Gaussian. DPZ does not identify a cut-off value for \(\tau\) with regards to rejecting the null hypothesis, but rather compare it to adjacent measures of \(\tau\). They claim that their results are unambiguous (i.e., that their values of interest for \(\tau\) are so large). As this is a form of a \(t\)-test, values close to 2 may be thought of as ambiguous.

#### Special cases

The test above works well when \(T\) falls significantly on one side of the peak of the pdf\((P)\). However, when the symmetric neighborhood around \(T\) includes \(P\), small alterations to the test are required.

If \(T < P\) but \(T \neq P\), construct a neighborhood \(R_i\) that is the most symmetric region possible around \(T\) of \(2r+1\) points such that all points lie at or below \(P\).

If \(T > P\) but \(T \neq P\), construct a neighborhood \(R_i\) that is the most symmetric region possible around \(T\) of \(2r+1\) points such that all points above (but not inclusive of) \(P\).

If \(T = P\), the distribution of reported earnings is likely centered at this \(T\). Now, I identify an EM effect by testing whether the slope of the density function immediately to the left of \(T\) is significantly different from the corresponding slope to the immediate right of \(T\) after allowing

---

5For this analysis, \(n=7\) has been selected, creating 15 cent intervals. DPZ reported figures for \(n=5\), claiming little sensitivity at \(n=7\) and \(n=10\). However, as most of the tests in this paper are performed at or near the peak, and as adjacent points were often subject to their own discontinuity tests, a slightly larger \(r\) was chosen to provide a better representation of the neighborhood around \(n\) that is not subject to threshold tests.
for any general local skew in the distribution. Define $\nabla p_j = \Delta \log [p(x_{j}, \cdot)] - \Delta \log [p(x_{\cdot}, \cdot)]$. The test for this case amounts to examining whether $\nabla p_j$ is unusual. I use the observation $\nabla p_j$ from a small neighborhood $K(j > 1)$ to compute an estimate of the mean of $\nabla p_j$ as well as its standard deviation. As before, I compute a $t$-like test statistic, say, $t_{x_{j}}$, to assess the “unusualness” of $\nabla p_j$.

References