Open Research Online

The Open University's repository of research publications
and other research outputs

An improved test for earnings management using kernel density estimation

Journal Item

For guidance on citations see FAQs.

Version: Accepted Manuscript

oro.open.ac.uk

An improved test for earnings management using kernel density estimation[i]

Henry Lahr[ii]

Centre for Business Research, Judge Business School
University of Cambridge
Trumpington Street
Cambridge, CB2 1AG
United Kingdom

Phone: +44 1223 765340
Fax: +44 1223 765338
h.lahr@cbr.cam.ac.uk

1

An improved test for earnings management using kernel density estimation

Abstract

This paper describes improvements on methods developed by Burgstahler and Dichev (1997) and Bollen and Pool (2009) to test for earnings management by identifying discontinuities in distributions of scaled earnings or earnings forecast errors. While existing methods use preselected bandwidths for kernel density estimation and histogram construction, the proposed test procedure addresses the key problem of bandwidth selection by using a bootstrap test to endogenize the selection step. The main advantage offered by the bootstrap procedure over prior methods is that it provides a reference distribution that cannot be globally distinguished from the empirical distribution rather than assuming a correct reference distribution. This procedure limits the researcher's degrees of freedom and offers a simple procedure to find and test a local discontinuity. I apply the bootstrap density estimation to earnings, earnings changes, and earnings forecast errors in U.S. firms over the period 1976–2010. Significance levels found in earlier studies are greatly reduced, often to insignificant values. Discontinuities cannot be detected in analysts' forecast errors, while such findings of discontinuities in earlier research can be explained by a simple rounding mechanism. Earnings data show a large drop in loss aversion after 2003 that cannot be detected in changes of earnings.

1. Introduction

Earnings benchmarks are widely used in the literature investigating earnings management. Burgstahler and Dichev (1997) (BD henceforth) and Degeorge et al. (1999) identify three benchmarks at which earnings management might be observed: Firms tend to avoid small losses, earnings decreases, and earnings that fall short of analysts' forecasts. These benchmarks are used in tests based on distributions of earnings, changes of earnings, and forecast errors, respectively, to derive conclusions about the existence of earnings management. For example, the deviation of reported earnings from analysts' forecasts is found to be skewed or discontinuous[3] at zero, which can indicate that earnings or forecasts are managed (Degeorge et al., 1999; Burgstahler and Eames, 2003; Burgstahler and Eames, 2006; Eames and Kim, 2012). If firms have incentives and discretion to achieve earnings above some threshold, the distribution of reported earnings will have fewer than the expected number of observations for earnings just below this threshold and more just above. The emerging consensus is that the distribution of earnings has a discontinuity around zero and is not symmetric (Cohen and Lys, 2003; Dechow et al., 2003; McNichols, 2000; Beaver et al., 2003; Burgstahler and Dichev, 1997).

In this paper, I evaluate the common distributional assumptions underlying most discontinuity tests and propose an improved test procedure for earnings management that corrects the tendency of earlier methods to employ overly optimistic distributional assumptions. In general, the distributional approach to testing for earnings management compares the observed number of observations in a region around the suspected discontinuity with the number expected in the absence of earnings management. If the difference between the observed number and the expected number derived from a continuous distribution is significant, the location at which this occurs can be considered a discontinuity and can be

---

[3] More precisely, a discontinuity in the sense of this paper is a point at which a density function is discontinuous, jumping from a region of lower density to one of higher density or vice versa.

interpreted as a benchmark for earnings management. The likelihood of detecting a discontinuity, however, will depend on the assumptions underlying this expected number. BD derive the expected number of observations by interpolating histogram bins. Their test statistic is the difference between the expected and actual number of observations in a histogram bin, divided by its standard deviation. Degeorge et al. (1999) propose a similar test statistic. Despite the large and steadily growing number of studies that use BD's method,[4] it has several shortcomings that reduce its validity under many circumstances. For example, as noted by Glaum et al. (2004), the choice of histogram interval width is a critical consideration that is often neglected. Holland (2004) demonstrates that if the peak of the distribution falls adjacent to a threshold, the BD method will not provide statistically reliable, robust results. The BD method does not locate the exact point of discontinuity and gives no hint of the structure of earnings management. Even if a plausible bin width can be determined, the researcher can arbitrarily shift the histogram's origin to the left or right.

The major difficulty in testing discontinuities is to derive a distribution of earnings under the null hypothesis of no earnings management or, more generally, under the null hypothesis of no discontinuity, against which the actual number of observations can then be compared. Choosing a reference (or "true") distribution that is biased can severely affect results. Bollen and Pool (2009) recognize the limitations of using histograms in their analysis of hedge fund returns. Instead of assuming local linearity in the distribution of fund returns, they fit a nonparametric reference distribution by kernel density estimation. Their method improves BD's test by eliminating some spurious results in histogram bins adjacent to the region around zero. However, the critical problem of choosing the appropriate kernel bandwidth (or bin width in BD's test) remains. Most research relies on Silverman's (1986) rule of thumb, which

---

[4] For a cross-section of studies discussing empirical evidence for earnings management and alternative explanations for discontinuities in standardized earnings, earnings per share, earnings increases, or analysts' forecast errors, see Burgstahler and Chuk, 2012; Beatty et al., 2002; Burgstahler and Eames, 2003; Beaver et al., 2003, 2007; Cohen and Lys, 2003; Dechow et al., 2003; Glaum et al., 2004; Brown and Caylor, 2005; Coulton et al., 2005; Durtschi and Easton, 2005; Burgstahler and Eames, 2006; Gore et al., 2007; Pinnuck and Lillis, 2007; Talha et al., 2008; Tung et al., 2008; Charoenwonga and Jiraporn, 2009.

may or may not suit the data at hand. Choosing a wrong bandwidth can have a severe impact on results. Gaussian kernels in particular, which are predominantly used for density estimation in studies employing Bollen and Pool's methodology (for example, Cassar and Gerakos, 2011; Brown et al., 2010), can result in highly inflated test statistics if bandwidth is not adjusted to suit the kernel.

This paper proposes a method to select a reference distribution free from these shortcomings when testing discontinuities in distributions. The aim is to construct the bandwidth needed for kernel density estimation from a bootstrap[5] test rather than use a preselected one that is assumed correct. Based on the null hypothesis that the reference distribution is continuous, I select the kernel bandwidth in such a way that the resulting continuous density estimate cannot be globally distinguished from the data by a bootstrap test. After fitting a reference distribution, a local test for a discontinuity can be carried out in the usual way by comparing the expected and actual number of observations. Throughout this paper, I refer to this method as bootstrap kernel density estimation (bootstrap KDE). This previously missing specification test ensures that the kernel estimate can serve as a reference distribution. The crucial step of bandwidth selection now depends on the fit of the density estimate and is not arbitrary as in previous studies. The density estimate obtained from the bootstrap test can be used as a nonparametric reference distribution in place of BD's assumptions of local linearity and correct bin width or Bollen and Pool's assumption of a correct bandwidth selector.

The full test procedure uses this bandwidth selection step as a key element in determining the significance of a discontinuity. First, the user supplies an a-priori bandwidth estimate as a starting value for iteration, such as one derived from Silverman's (1986) rule of thumb, and a kernel function and confidence level for the bootstrap step. In the bootstrap step that follows,

---

[5] I refer to bootstrapping as a procedure that draws repeated samples with replacement (resampling) from an empirical distribution to establish the uncertainty of a statistic of this distribution. In this paper, bootstrapping is used to estimate the uncertainty about the empirical cumulative distribution at the point of a suspected discontinuity.

the kernel bandwidth is iteratively adjusted to produce a suitable reference distribution. As a by-product, the procedure identifies a point at which a discontinuity might be located by finding the maximum difference between the empirical distribution and the integrated density estimate. Finally, a standard z-test or binomial test can be used to test the discontinuity's statistical significance in a region around the location of a suspected discontinuity.

Kernel density estimation can be used for a wide range of distributions and is not limited to a linear relationship between the numbers of observations in adjacent histogram bins. To test prior findings of discontinuities and demonstrate the effect of bandwidth selection, I apply the bootstrap KDE procedure to three classical earnings measures: earnings scaled by market capitalization, changes in scaled earnings, and errors in analysts' forecasts of earnings per share (EPS). Results for scaled earnings and changes of earnings[6] of U.S. firms in the period 1976–2010 show a discontinuity at zero. However, there is a large drop in significance after 2003, which might indicate a decrease in earnings management over time. Prior research suggests a declining importance of avoiding small losses or earnings decreases compared to meeting or beating analysts' forecasts (Herrmann et al., 2011; Burgstahler and Eames, 2006; Brown, 2001; Degeorge et al., 1999), which I find only for standardized earnings, but not for year-on-year changes in earnings.

The bootstrap KDE procedure provides evidence against a discontinuity in analysts' earnings forecast errors. Prior findings of discontinuities are most likely caused by inflated bandwidths and EPS forecasts that are rounded to two digits. When EPS observations that exactly meet the median forecast are kept in the sample, the bootstrap test fails to establish a suitable reference distribution due to the large point mass of zeroes at the origin. Hence, any attempt to use a continuous reference distribution including zeroes has a large potential to produce misleading results. If zeroes are excluded from the sample, the bootstrap approach does not detect a discontinuity in forecast errors based on stock split-adjusted I/B/E/S data. If

---

[6] "Earnings and changes of earnings" both refer to scaled, or standardized, earnings. To enhance readability, I omit "scaled" when referring to earnings scaled by market capitalization.

I/B/E/S data are manually adjusted for stock splits to increase the precision of forecast errors, an apparent discontinuity can be detected. This discontinuity disappears, however, with re-introduction of the previously omitted zeroes while accounting for the fact that EPS estimates are usually rounded to two digits.

The remainder of this paper is organized as follows: Section 2 outlines the idea behind a bootstrap specification test for the reference distribution and reviews BD's (1997) test and Bollen and Pool's (2009) method. Section 3 describes the proposed bootstrap KDE procedure. Sections 4 and 5 apply the bootstrap KDE procedure to earnings benchmarks, and section 6 concludes.

2. In search of a reference distribution

There are two closely related sides to the problem of measuring earnings management by identifying a discontinuity in a given empirical distribution: We need to define *discontinuity* and there must be a statistical test of significance. Although a discontinuity in density functions is well defined, empirical densities lack an unambiguous definition. To make inferences about the observed data, a "true" density function that would be obtained if earnings management was absent has to be estimated in a most general way. Once such a reference distribution is established, it can be compared with the observed data. A mismatch between the two can then indicate earnings management.

[Figure 1 about here]

The search for a reference distribution starts with a general assumption about its shape. BD assume a linear relationship between the number of observations in a histogram bin and its adjacent bins, which translates into a locally linear reference density function as shown in Figure 1 (right panel). A more flexible and nonparametric method to define a reference distribution is to construct it from the data by smoothing the observed data through kernel

density estimation, which is a commonly used method if a continuous density function is needed (Bollen and Pool, 2009). The only key assumption about the reference density is its continuity. The right panel in Figure 1 shows two such density estimates for particular choices of bandwidth. One is based on a bandwidth that coincides with the histogram's bin width, the other is chosen by an automatic bandwidth selector similar to the one used by Bollen and Pool (2009).

There is an infinite number of possible kernel density estimates, one for each combination of kernel bandwidth and kernel function. Which one is the correct density estimate? One way to assess whether a particular density estimate can be used as a reference distribution in a discontinuity test is to determine whether it agrees with the observed data. Since there is uncertainty about the underlying true distribution, some density estimates are statistically indistinguishable from this true distribution. For insight into this uncertainty, consider the cumulative distribution function for the observed data shown in Figure 2, which corresponds to a narrow region of the histogram in Figure 1. The uncertainty about the distribution's exact location is reflected in a confidence interval that I obtain from resampling (bootstrapping) the empirical distribution. All kernel density estimates that lie entirely within this confidence band are candidates for the reference distribution, because they cannot be statistically distinguished from the data.

[Figure 2 about here]

The crucial step in establishing a reference distribution is the choice of bin width or kernel bandwidth. Results contrasting two different choices for the kernel bandwidth are shown in Figure 2. Both panels show the empirical distribution of scaled earnings and an integrated kernel density estimate with bootstrap confidence bands. Note the marked change in slope at zero, which is often attributed to earnings management. In the left panel, bandwidth is

8

determined in the same way as in prior studies—by calculating Silverman's (1986) rule of thumb and applying a Gaussian kernel. As can be seen directly from the graph, the integrated density estimate (the reference distribution) lies far outside the bootstrap confidence bands and even outside the confidence bands of a Kolmogorov-Smirnov distribution test. Therefore, the null hypothesis that the observed data are generated by the reference distribution is strongly rejected. Any inference based on this unsuitable reference distribution is likely to produce misleading results. A contribution of this paper is to test whether specific choices of bandwidth in previous studies in fact generate suitable reference distributions.

A suitable and unique continuous reference distribution can be constructed by varying the kernel density estimate's smoothness. The right panel in Figure 2 shows a reference distribution based on a kernel density estimate whose bandwidth has been adjusted so that it exactly meets the bootstrap confidence bands.[7] In other words, this reference distribution now agrees with the data at a specified level of confidence and can be used for a discontinuity test. The remaining part of this section and section 3 describe the bandwidth selection procedure and discontinuity tests in detail.

Determining a correct bandwidth can be seen as a specification test analogous, for example, to testing normality of the data before performing a *t*-test. An invalid *t*-test still produces a result, although this result is sensitive to departures from normality. In particular, *t*-tests performed on skewed data reject the null hypothesis too frequently. In this sense, *t*-test statistics for nonnormal data can be described as inflated. Similarly, my results show that prior tests of discontinuities are indeed sensitive to bandwidth specification. If bandwidth is not specified correctly, test statistics can be highly inflated. Hence, whenever a given reference distribution lies outside the bootstrapped confidence region, a discontinuity test based on this reference distribution violates its assumptions.

---

[7] The empirical distribution and kernel density estimate shown in the right panel correspond to Figure 2.

2.1. Burgstahler and Dichev's test

In order to estimate the effect of bandwidth selection on test results, I compare my bootstrap KDE results with BD's widely used method. They construct histograms to derive a test statistic based on the expected number of observations in each histogram bin. Let $X_1, ..., X_N$ be N independent random variables with distribution function F. Construct a histogram with equally spaced bin boundaries $-\infty = c_0 < c_1 < ... < c_m = \infty$, where $c_j - c_{j-1} = h$ for $j = 1,...,m$. The number of observations in bin i is then defined as

$$n_i \equiv \sum_{i=1}^{N} \mathbf{1}\left(X_k \in (c_{i-1}, c_i] \right), \; i = 1,...,m, \tag{1}$$

with $\mathbf{1}(\cdot)$ being the indicator function, which is equal to unity if the condition enclosed in parentheses holds and zero otherwise. The number of observations n_i follows a multinomial distribution with $p_i \equiv P(X \in (c_{i-1}, c_i])$. The test statistic is

$$BD = \frac{(n_{i-1} + n_{i+1})/2 - n_i}{\sqrt{Var\left((n_{i-1} + n_{i+1})/2 - n_i \right)}}, \tag{2}$$

where

$$Var\left(\frac{n_{i-1} + n_{i+1}}{2} - n_i \right) = Np_i(1 - p_i) + \frac{N}{4}(p_{i-1} + p_{i+1})(1 - p_{i-1} - p_{i+1}) + Np_i(p_{i-1} + p_{i+1}), \tag{3}$$

N is the total number of observations, n_i is the number of observations in bin i and $p_i = n_i / N$. Burgstahler and Dichev ignore the last term in Equation (3), as noted by Takeuchi (2004). More importantly, they assume that $E[(n_{i-1} + n_{i+1})/2 - n_i] = 0$ to derive Equation (3) (see Appendix). This linearity assumption can be overly restrictive when applied to multimodal or skewed densities, particularly in regions around the mode(s) of a distribution. Additionally, histogram bin width is chosen arbitrarily based on a priori considerations.

2.2. Kernel density estimation and bandwidth selection

The more general method used in this paper employs kernel density estimators as introduced by Rosenblatt (1956) and Parzen (1962) and applied in the context of managed hedge fund returns by Bollen and Pool (2009). This family of density estimators $\hat{f}_h$ is defined by

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right), \tag{4}$$

where h is the kernel bandwidth, n is the sample size, and K, the kernel, is some smoothing density function. The subscript h indicates that the density estimate depends on the parameter h which is our main focus here. I use three kernel functions to compute the results in this paper: The Epanechnikov kernel,

$$K(z) = \frac{3}{4}(1 - z^2)\mathbf{1}(|z| \leqslant 1), \tag{5}$$

the Gaussian kernel

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} z^2\right) \tag{6}$$

and the uniform kernel

$$K(z) = \frac{1}{2}\mathbf{1}(|z| \leqslant 1), \tag{7}$$

where $\mathbf{1}(\cdot)$ is the indicator function. Applying different kernel functions provides a robustness check and offers some insights into the importance of selecting an appropriate kernel relative to the importance of choosing the optimal bandwidth h.

The choice of the bandwidth h is a key aspect of a practical implementation of kernel density estimation. Choosing a small h leads to an estimator with a small bias and large variance, whereas a large h causes lower variance at the expense of concealing features that might be present in the data. Techniques for bandwidth selection include Silverman's (1986)

rules of thumb, oversmoothing, cross-validation, direct plug-in methods, the solve-the-equation method, and the smoothed bootstrap (Jones et al., 1996). Studies estimating the density of standardized earnings or related variables commonly use Silverman's (1986: 45, eq. 3.28) rule of thumb for bandwidth selection, which is theoretically optimal for normally distributed data and a Gaussian kernel:

$$h_{opt} = 1.06\sigma n^{-1/5} \tag{8}$$

where

$$\sigma = \min\left(\sqrt{\hat{\sigma}_X}, \frac{Q_X}{1.349} \right) \tag{9}$$

and $\hat{\sigma}_X$ is the sample standard deviation and Q is the interquartile range. This bandwidth selector is sensitive to heavily skewed data but not so much to kurtosis of the underlying empirical distribution. Silverman therefore proposes a variation using a factor of 0.9 instead of 1.06, which provides robust bandwidth estimation for a wide range of t-, log-normal, and normal mixture distributions (see Silverman, 1986: 48, eq. 3.31):

$$h_{opt} = 0.9\sigma n^{-1/5} \tag{10}$$

Standardized earnings and other data in financial applications are often moderately skewed or show excess kurtosis. Equation (10) should therefore provide a good approximation of the optimal bandwidth from a theoretical perspective. Studies estimating the density of earnings or other variables to find discontinuities usually use Equation (8) or a close variation thereof (Bollen and Pool, 2009; Cassar and Gerakos, 2011; Green, 2010; Brown et al., 2010).

If the distribution is not only skewed but also suspected to be discontinuous, the kernel density estimated with bandwidths obtained from Equations (8) or (10) can be highly inaccurate, as seen in Figure 2 (left panel). If a Gaussian kernel is used in combination with any of these bandwidths, results can be strongly biased in favor of rejecting the null

hypothesis of no discontinuity, as I will show in section 4.3 below. The general rule is, the larger the bandwidth, the more likely the null is rejected. In many cases in which a bandwidth chosen by the above rule seems optimal, even a simple Kolmogorov-Smirnov test can reject the assumption that the kernel density estimate can serve as the "true" reference distribution when testing for a discontinuity. Hence, the aim of this paper is to construct a density estimate that is not rejected by a bootstrap test, because it cannot be distinguished *globally* from the data. After a continuous reference distribution has been constructed, a *local* discontinuity test can then be applied.

Note that optimal bandwidth depends on sample size, which may lead to different estimates for the extent of earnings management, measured by the difference between the empirical cumulative distribution function (ECDF) and integrated density estimate at some point, for the same empirical distribution but different sample sizes. In fact, the difference between the estimated and empirical distributions can be made arbitrarily small for large samples, since the estimated density becomes increasingly fine-grained. As a result of this dependence on sample size, the number of managed earnings cannot be measured without making further assumptions about the distribution of unmanaged earnings. This feature is also inherent in the BD method in which optimal bin width must be determined. The number of companies that manage earnings as found, for example, by BD and Donelson et al. (forthcoming) depends on their assumptions of a locally linear density and a specific bin width and can be made smaller or larger by employing different assumptions or sample sizes.

3. Adjusting the bandwidth and testing the discontinuity

The underlying distribution of unmanaged earnings is unknowable from the sample data alone, because we observe earnings only after potential manipulations. We can, however, try to make as few assumptions as possible when estimating a reference distribution from the data. The only key assumption about the underlying distribution in my test procedure is its smoothness. More precisely, the reference distribution's density function is only assumed to

be continuous as opposed to BD's above assumption of local linearity (their definition of "smooth"). The main objective of this paper is to find a suitable bandwidth that can be used in kernel density estimation to construct this smooth density function.

The reference distribution must be chosen carefully, since its shape will determine the level of statistical significance. The idea is to estimate a kernel density that globally fits the empirical distribution and to locally test for discontinuities based on this density estimate. A significant discontinuity test can indicate that (1) the data have an abnormal discontinuity or (2) the kernel density estimate is simply oversmoothed and thus missing this discontinuity. If the kernel density estimate is not a good approximation of the data, tests of discrepancies between the two can produce spurious inferences. Therefore, the data should be a plausible realization of the estimated kernel density. Instead of assuming a specific shape for the distribution by selecting some a priori "optimal" bandwidth, I test directly whether a density estimate corresponds to the data in a statistically meaningful way.

A general method to detect a mismatch between the empirical and reference distributions is to construct confidence bands for the empirical distribution that the kernel density estimate has to meet. Such confidence bands can be obtained by resampling the original data (bootstrapping, see Scott, 1992: chap. 9.3.2), since the kernel density estimate should be a plausible distribution considering that the data are only a single realization of some unknown density. The distance between the integrated kernel density estimate and these confidence bands around the ECDF can be adjusted by varying the kernel bandwidth to yield a kernel density estimate that agrees with the data at a confidence level selected by the researcher. If, for example, Silverman's rule of thumb suggests a bandwidth that is rejected as too large by this bootstrap test, bandwidth will be reduced until we find the maximum bandwidth that still yields a continuous reference distribution that agrees with the bootstrapped confidence interval for the empirical distribution. Endogenizing bandwidth selection in this way replaces

an a priori selected bandwidth that may or may not be appropriate for the data at hand with a bandwidth that is correct at a precise confidence level.

After a reference distribution is established, there needs to be a measure of statistical significance where a discontinuity is suspected. For example, BD calculate the standardized difference between the expected and observed number of observations in histogram bins and perform a *z*-test. They test for a discontinuity at a location determined by a priori considerations. For this reason, their method is not a test of discontinuities *somewhere* in the distribution but at a specific location. The method proposed in this paper can be used in this way but can also take a different approach to finding and testing the discontinuity: A test can be conducted at the point of maximum difference between the empirical cumulative distribution and the integrated kernel density estimate, which is a likely candidate for a discontinuity. Irrespective of whether the discontinuity's location is chosen a priori or by finding the maximum difference, the test should then be based on the expected number of observations within a reasonable interval around this point of maximum difference.

The test procedure is as follows:

1. Select a first estimate for the kernel bandwidth *h* as a starting point for iteration. This bandwidth can be calculated from the data, for example, by Silverman's (1986) rule of thumb.

2. Choose a kernel function and construct a kernel density estimate. Calculate the maximum difference to the ECDF. Denote the point of maximum difference as d_{max}.

3. Draw a large number of samples from the original data with replacement (that is, bootstrap) and construct a confidence interval for the empirical distribution at d_{max}.

4. If the integrated kernel density at d_{max} is outside this confidence interval, reduce bandwidth *h* and proceed from step 2. If the ECDF is inside the confidence interval, increase *h*. Iterate from step 2 until the ECDF meets the confidence band.

5. Test the expected number of observations within the intervals $(d_{max} - h, d_{max}]$ and $(d_{max}, d_{max} + h]$ simultaneously against the observed number using a binomial test or z-test to determine the discontinuity's statistical significance.

3.1. Bootstrap confidence intervals for a kernel density estimate

Bootstrapping is used in step 3 to test whether the unknown earnings process may have generated the kernel density estimate representing the true distribution given that the data are only one realization of some underlying true distribution. Under the assumption of independently and identically distributed observations, the sampling error can be estimated by bootstrapping from the empirical distribution.[8] To reflect the error due to observing only a finite sample from the underlying distribution of "true" standardized earnings, a large number of samples with size equal to the original sample size are drawn with replacement from the data. From these draws, I construct confidence bands for the empirical distribution at a predetermined confidence level. If the integrated kernel density estimate for some bandwidth lies outside the confidence interval, it is unlikely that the density estimate derived from the original data describes the data reasonably well. Bandwidth must be reduced until the integrated kernel density estimate exactly meets the confidence interval (see Figure 2 for an illustration). Sometimes, after fitting the bandwidth to the empirical distribution at a specific point, the point of maximum difference between integrated kernel density and empirical distribution is different from the one at which the iteration started. In that case, the iteration procedure should be repeated at this point, and the final bandwidth should be the smallest bandwidth obtained from iterations at all points of maximum difference that were found.

This technique leads to reasonable bandwidths and reduces the researcher's degrees of freedom to selecting the kernel function. Its benefit is that it gives a precise answer to the

---

[8] Bollen and Pool (2009: 2270) use a smoothed bootstrap to test the robustness of their findings. Although it shares the name, their bootstrapping approach should not be confused with the one proposed here. While they employ bootstrapping as part of a discontinuity test after assuming a reference distribution (by preselecting a bandwidth), I use it to construct a plausible reference distribution.

question of how likely it is that the kernel density estimate generates the empirical data, as opposed to the assumption of having picked a bandwidth suitable for the data at hand. The proposed method can also be employed to test the validity of reference distributions used in prior studies. Kernel density estimates using the original data and bandwidths documented in prior studies can be compared with a bootstrapped confidence interval for the empirical distribution at zero. If the original data and bandwidth produce a kernel density estimate that lies outside this confidence interval, we can conclude that the implied reference distribution does not represent the data and thus cannot be used to conduct statistical tests to find a discontinuity.

3.2. Testing the discontinuity

After a plausible kernel density estimate has been constructed, the discontinuity itself can be tested using the KDE as a reference distribution. The discontinuity is likely to be located at d_{max}, the maximum difference between the integrated density estimate and the ECDF.[9] The method proposed in the literature and used here is to compare the expected number of observations under the null hypothesis of being generated by the density estimate with the actual number on both sides of the discontinuity. Since the number of observations that fall within a specific interval follows a binomial distribution, the test statistic is

$$z = \frac{p - \hat{p}}{\sqrt{\hat{p}(1 - \hat{p})}} \sqrt{N}, \tag{11}$$

if we use the normal approximation to the binomial distribution and assume independence of observations. The empirical probability p is the actual number of observations in this interval divided by sample size N. For small pN, the binomial distribution should be used instead of the normal approximation. The expected number of observations in terms of the integrated density estimate, $\hat{F}_h$, over the interval of interest is

---

[9] This can be shown analytically for some cases, such as a uniform distribution and uniform kernel. If a distribution contains several and possibly smaller discontinuities, a graphical representation of the difference between empirical distribution and integrated KDE may reveal more potential locations for a discontinuity test.

$$\hat{p} = \hat{F}_h(d_{max}) - \hat{F}_h(d_{max} - h) \tag{12}$$

for the interval to the left of the discontinuity and

$$\hat{p} = \hat{F}_h(d_{max} + h) - \hat{F}_h(d_{max}) \tag{13}$$

for the interval to the right. A natural choice for the interval width is the kernel bandwidth h, especially for kernels with a finite support, such as the uniform or Epanechnikov kernels. Since two hypotheses are tested simultaneously, the test's rejection region must be adjusted, for example, by a Bonferroni correction. If the true density shows a positive jump at the discontinuity, the test statistic is negative to the left and positive to the right.

4. Earnings and changes in earnings

The purpose of this section is to demonstrate and evaluate the bootstrapping technique with data that have been used extensively in the literature and to extend the sample period to include recent years. I compare prior methods for identifying discontinuities and bootstrap KDE to various subsamples in order to highlight the importance of choosing an appropriate bandwidth (or bin width) and kernel function. The empirical findings challenge several results in prior studies on the existence and stability of discontinuities in earnings and earnings changes.

4.1.   Data

The data used in the empirical analyses are from the Compustat annual financial accounts database (fundamentals file) and include net income and market capitalization for U.S. firms in the period 1976 to 2010. I choose these variables and time period to produce results comparable to those of BD. The samples I use to investigate discontinuities in standardized earnings and earnings changes contain $N = 175\,349$ and $N = 165\,052$ firm-year observations, respectively. Standardized earnings are calculated as net income scaled by lagged market capitalization ($Earnings_t / MV_{t-1}$), which is obtained from multiplying ordinary shares

outstanding by financial year-end share price as taken from Compustat.[10] Changes in earnings are defined as the change in net income, divided by market capitalization two years earlier ($(Earnings_t - Earnings_{t-1}) / MV_{t-2}$).

Following BD's sample selection criteria, I exclude financial institutions with SIC codes in the ranges 4400–5000 and 6000–6500. Standardized earnings less than −1 or greater than 1 are excluded as well, as are observations of exactly zero, since these might be data errors.[11] However, the methods used in this paper are robust to outliers, which would exert an influence on outcomes only if kernel density estimation with a Gaussian kernel was used. The remaining earnings with outliers removed have a mean of $\mu = 0.0051$, a standard deviation of $\sigma = 0.2202$, skewness of $s = -1.1509$, and kurtosis of $k = 7.3465$. Their median of 0.0403 is more informative, since it is robust to outliers. The same statistics for earnings changes, excluding values outside the [−1, 1] interval, are $\mu = 0.0094$, $\sigma = 0.2107$, $s = 0.0724$, and $k = 8.8662$, and the median is 0.0077.

4.2. Method

For discontinuity tests on standardized earnings and earnings changes, I use the bootstrap procedure already described, a density estimate as used in the literature ("Rule-of-thumb KDE," see Bollen and Pool, 2009), and BD's method of interpolating histogram bins. Selecting a bandwidth and kernel function are the researcher's two degrees of freedom in density estimation. The kernel bandwidth h for the bootstrap KDE is determined by the procedure described above using 2000 bootstrap samples drawn from the original data and a one-sided confidence interval of 0.1 when fitting the bandwidth to the data. To compare my results with the literature, I compute a density estimate with Silverman's (1986: 48, eq. 3.31)

---

[10] Deflating by market capitalization is standard practice in the literature (for example, Burgstahler and Dichev, 1997; Dechow et al., 2003; Beaver et al., 2007). It is necessary to account for firm size, as size is related to the distribution of net income and to the extent of potential earnings management. Results are robust to scaling by book value of equity or total assets. For an analysis using undeflated net income, see section 4.3.

[11] Excluding zero earnings follows the literature, see Burgstahler and Dichev (1997), Bollen and Pool (2009), Brown et al. (2010).

rule of thumb using Equation (10). Since bandwidth can be chosen arbitrarily in BD's method and to directly compare results, I use the bootstrap KDE bandwidth for BD's test.

There is a great variety of available kernel functions, of which I test three: Epanechnikov kernel, Gaussian kernel, and uniform kernel. Gaussian kernels have been used extensively in the literature, especially in recent work on discontinuities (Bollen and Pool, 2009; Green, 2010; Brown et al., 2010). Epanechnikov kernels have theoretical appeal in a minimum variance sense and uniform kernels are the simplest possible. Both Epanechnikov and uniform kernels have bounded support, whereas Gaussian kernels take all sample observations into account. This property turns out to be important for the reliability of discontinuity tests.

4.3.   Results

Standardized earnings of U.S. firms show the pattern presented by BD. They are quite heavy-tailed, as can be seen in Figure 1, which shows a histogram of earnings as the starting point for BD's method. The kernel density estimate, which is the reference distribution for the bootstrap KDE procedure, mirrors the histogram's skewness and kurtosis. An anomaly around zero is clearly visible. The objective of the following analysis is to test whether the observations seemingly missing under the curve to the left of zero and the additional ones to the right constitute a significant discontinuity.

The gap visible in Figure 2 between the integrated kernel density estimate and the empirical cumulative distribution of standardized earnings suggests a discontinuity at zero. While the reference distribution is continuous at zero by design, the empirical distribution shows a marked change in density at this point. The cumulative distribution deviates negatively from the reference distribution in the interval $[-0.003, 0.003]$ as shown in Figure 3. This reproduces the result reported by BD, who find unusually low frequencies of small losses and unusually high frequencies of small positive income in cross-sectional distributions of earnings.

[Figure 3 about here]

The maximum difference between the integrated kernel density estimate and the empirical distribution is located almost exactly at zero. If we had no a priori reasons to believe that a possible discontinuity is located at zero, the algorithm would suggest a location at standardized earnings of 0.00607 percent, which is only 16 observations to the right of zero in my sample of 175 349 firm-year observations. For earnings changes, the maximum difference occurs at 0.00571 percent. These locations are independent of the kernel function chosen. This finding suggests that regardless of the method for testing the discontinuity's significance, its location can be established with high precision.

Discontinuities at zero standardized earnings and zero earnings changes are highly significant for all three kernels tested (see Tables 1 and 2). For standardized earnings, tests for all combinations of time period, estimation technique, and kernel function yield significant discontinuities. Negative test statistics for small losses and positive ones for small gains show the positive change in density around the discontinuity. Results for the 1976–1994 period investigated by BD are more strongly significant than those for the period 1995–2010 despite the smaller sample size, which indicates less loss aversion in more recent years.

Earnings management is only one of several explanations for a discontinuity in scaled earnings. Although testing alternative hypotheses is not the primary focus of this paper, the bootstrap test procedure can directly address an explanation put forward by Durtschi and Easton (2005, 2009). They argue that deflating net income by market capitalization pulls observations away from zero on the left of this threshold and draws them closer to zero on the right. This scaling supposedly creates a discontinuity in an otherwise smooth net income distribution.[12] They present results showing an absence of a discontinuity in intervals of width $100 000. When applying the bootstrap approach to undeflated net income, however, I find a

---

[12] For a counterargument of why scaling is necessary and a detailed discussion of Durschi and Easton's (2005, 2009) findings, see Burgstahler and Chuk (2012).

highly significant discontinuity. Contrary to claims by Durtschi and Easton (2005: 568), this discontinuity exists in undeflated net income regardless of whether I use the same firm-year observations as before (those 175 349 observations that have a beginning-of-year market price) or all Compustat net income observations in the sample period ($N = 244\,764$). This finding holds for all kernel functions used in this paper. Standardized differences for the interval to the right of zero and an Epanechnikov kernel are $z = 5.74$ ($p < 0.01$) for firm-year observation with existing market value and $z = 3.28$ ($p < 0.01$) for all net income observations in the sample period. Standardized differences on the left of zero are significant only for the sample with beginning-of-year market capitalization, but this does not affect the discontinuity's overall significance. This discontinuity in undeflated net income might have gone undetected because of the relatively narrow kernel bandwidth, which is $h = \$67\,719$ for an Epanechnikov kernel and $h = \$34\,516$ for a Gaussian kernel.

[Table 1 about here]

[Table 2 about here]

When comparing alternative discontinuity tests, the bootstrap KDE always shows the lowest level of significance relative to Rule-of-thumb KDE and BD's method. BD's test statistic seems to overstate the significance of discontinuities even if the relatively small bootstrap bandwidth is used. I compute BD test statistics of -10.24 to the left and 7.57 to the right of zero earnings in the case that is best comparable to BD's results ($h = 0.00267$, BD use $h = 0.0025$), where BD find -8.00 and 5.88, respectively. Bootstrap KDE test statistics are -7.10 and 6.26 for the same sample. In general, BD significance levels exceed the significance levels in bootstrap KDE tests by far, which is especially severe for bandwidths obtained from Epanechnikov and uniform kernels. By contrast, Rule-of-thumb KDE performs

worst when using a Gaussian kernel, which is the method preferred in the extant literature. A Gaussian kernel used in combination with Silverman's rule of thumb to determine bandwidth seems to be a particularly bad idea when testing discontinuities. Overstated test statistics are likely due to exceedingly large bandwidths, which are about three to four times as big for Rule-of-thumb KDE compared to bootstrap KDE using Gaussian kernels. Test statistics are inflated for Epanechnikov and uniform kernels as well, but to a lesser extent.

For changes in earnings, all tests are highly significant, albeit less so when compared to standardized earnings. The period 1976–1994 appears slightly more significant than 1995–2010, a trend that cannot be detected by the Rule-of-thumb KDE or BD methods. Again, test statistics obtained from the bootstrap KDE are always smaller than those from the Rule-of-thumb KDE and BD methods, with the exception of the Gaussian kernel case where BD's test statistic and the bootstrap KDE give about the same result for the 1995–2010 period. The overall picture of strongly inflated test statistics when smoothing with Gaussian kernels also holds for earnings changes, which can be attributed to the weight this kernel places on observations outside the interval $\pm h$ from the point of interest. Both the uniform kernel and the Epanechnikov kernel consider observations within this finite interval only, which might be superior to using a Gaussian kernel when looking for irregularities in the *neighborhood* of some point. Hence, the bootstrap bandwidth needs to be much smaller for a Gaussian kernel to produce a reference distribution that is not significantly different from the data.

The bootstrap approach can be used to test for discontinuities in yearly subsamples over the period 1976–2010 by calculating the kernel density estimate for each year, using an individually appropriate bandwidth and then applying a z-test or binomial test at zero. Table 3 shows test results for standardized earnings and earnings changes over time, using an Epanechnikov kernel for density estimation. If discontinuities in standardized earnings can be interpreted as earnings management, then the prevalence of this practice has been decreasing

since 1999. After 2001, no discontinuity can be found.[13] The highest significance occurs between 1991 and 1994 independently from the method used to derive the test statistic. This result is not driven by sample size, which is comparatively small in these years. However, which method is used does make a difference for the overall level of significance. BD's test almost always returns significant results, except for the years 2005, 2006, and 2008. Bootstrap KDE produces significant results slightly more often than Rule-of-thumb KDE,[14] but with gaps and not since 2003.

A striking result in yearly subsamples is the complete absence of loss aversion after 2003. Inspection of discontinuity graphs similar to Figure 3 for yearly subsamples confirms this result: After 2003, no dip in the ECDF around zero can be found. To illustrate the effects underlying the difference in significance in yearly subsamples (see Table 3), Figure 4 shows the difference between expected and observed earnings numbers in the years 2001 and 2009.[15]

Contrary to scaled earnings, there is no clear trend in the tendency to avoid small negative earnings *changes*. Insignificant tests appear seemingly randomly over the test period. Closer visual inspection of yearly distributions reveals that both the bootstrap KDE and Rule-of-thumb KDE seem to detect and reject discontinuities correctly most of the time, whereas the BD test statistic rejects the null hypothesis of no discontinuity much too often.

Yearly results for Gaussian and uniform kernels, which are shown in Figure 5, mirror the results obtained with Epanechnikov kernels. Discontinuities in earnings are harder to detect after 2003, while there is no such trend for earnings changes. The bootstrap approach yields

---

[13] A prior version of this paper included results for discontinuities in standardized earnings of German firms. In these data, significance peaked in 2000 and has been declining since, but is still detectable in 2008. If a Bonferroni correction for multiple hypothesis tests is applied, discontinuities during the periods 1996–1998 and 2006–2008 are insignificant. This pattern resembles stock market movements over the period 1991–2008, which is supported by evidence provided by Günther et al. (2009): They suggest that neither voluntary nor mandatory IFRS adoption around 2005 unambiguously leads to higher earnings quality and that capital market phases are an important determinant of earnings quality.

[14] If the bootstrap KDE approach detects more discontinuities than Rule-of-thumb KDE, then this could be due to bandwidths being larger, but that does not necessarily lead to higher significance in yearly subsamples.

[15] While the difference peaks at zero in 2001, no such anomaly can be found in 2009. Although time-specific events are hard to disentangle from other confounding time effects, it might be worthwhile to investigate whether the apparent reduction in earnings management after 2003 could be due to the Sarbanes-Oxley Act of 2002.

conservative estimates for all kernel functions used. Significance is highly inflated for the popular combination of rule-of-thumb bandwidth and Gaussian kernels, which is particularly striking for earnings changes in Figure 5c. BD's test statistic seems somewhat inflated as well, even when using the comparatively small bootstrap bandwidth.

[Tables 3, 4 and 5 about here]

All results taken together, loss aversion has likely declined since 2003, whereas I find no such tendency for earnings changes. The bootstrap KDE approach detects discontinuities far less often than BD's (1997) test and about as often as Bollen and Pool's (2009) test when using an Epanechnikov kernel. Compared with Bollen and Pool's (2009) specification that uses a Gaussian kernel as shown in Figure 5c, test statistics are reduced to half their size, rendering many tests insignificant. These results imply that prior positive findings of discontinuities often suffer from misspecified reference distributions.

Empirical studies suggest that the tendency to meet or beat analysts' forecasts has gained importance over avoiding small losses or earnings decreases in recent years (Herrmann et al., 2011; Burgstahler and Eames, 2006; Brown, 2001; Degeorge et al., 1999). Brown and Caylor (2005) show that since the mid-1990s, managers seek to avoid either small losses or earnings decreases less than negative earnings surprises. They argue that the economic rationale for this shift in hierarchy is that since the mid-1990s investors unambiguously rewarded firms for reporting earnings meeting analysts' estimates more than they did so for meeting the other two thresholds. While earnings management cannot be ruled out in many subperiods in my results, the next section presents evidence against earnings management to meet or beat earnings forecasts.

5. Earnings forecast errors

The practical importance of earnings management to meet the consensus estimate and the tendency of BD's test statistic to inflate test statistics call for a reinvestigation of the distributional approach to detect discontinuities in forecast errors. In their influential studies on earnings management to beat benchmarks, Degeorge et al. (1999) and Burgstahler and Eames (2006) investigate earnings surprises in analysts' forecasts. Burgstahler and Eames (2006) apply the BD test to histograms of forecast errors, while Degeorge et al. (1999) employ a very similar test to find anomalies in histograms. Although they use data from different sources, they find similar evidence for an anomaly in the distribution of forecast errors. There seems to be a lack of small negative earnings surprises, which they attribute to earnings management or forecast management to beat the benchmark. Dechow et al. (2003) find an increasing tendency over time for firms to beat the forecast benchmark relative to other benchmarks.

Although a substantial number of studies have focused on positive abnormal accruals and earnings guidance to hit earnings targets (Payne and Robb, 2000; Bartov et al., 2002; Koh et al., 2008; Athanasakou et al., 2009, 2011), survey evidence on the importance of earnings benchmarks collected by Graham et al. (2005) is mixed. Their results show a higher percentage of CFOs who consider it important to meet the previous year's earnings than CFOs who agree that analysts' consensus estimates are an important benchmark. However, CFOs also believe that market reactions after missing the forecast benchmark are more severe than after underperforming prior earnings.

1.1. Data and method

To calculate a standardized measure of earnings surprises over the time period 1986–2010, I obtain analyst forecasts for EPS and actual values from I/B/E/S and year-end share price data from Compustat. Following BD and Burgstahler and Eames (2006), I exclude both financial and utility firms from my analyses and scale forecast errors by the firm's share price

at the end of the previous financial year. Subtracting the median EPS forecast from the actual value yields the correct forecast error if the number of shares is constant for both figures. If stock splits occur during the observation period, these must be taken into account to arrive at the same per-share basis for forecasts and actual earnings.

Working with I/E/B/S data can be challenging due to split adjustments and rounding errors. Payne and Thomas (2003) note that adjusting past EPS for stock splits and rounding to the nearest penny can cause severe problems, because adjusted EPS values decrease in magnitude and rounding errors become relatively more important. Historically, I/B/E/S provided only split-adjusted data, which suffered from this problem. Recognizing this difficulty for researchers, I/B/E/S made unadjusted data available, which contain unadjusted individual and consensus forecasts. However, these data can still be subject to errors due to the fact that I/B/E/S records splits only once per month, which often is not the exact true split date (Robinson and Glushkov, 2006). As a robustness test and to compare different approaches to accounting for rounding errors due to split adjustments, I compute forecast errors in three ways:

"Adjusted" method: I derive forecast errors from median estimates obtained from the unadjusted I/B/E/S summary file and *adjusted* actuals from the adjusted details file, which are converted to unadjusted values by matching the I/B/E/S adjustment factor at the date the forecast is made. Forecast errors are defined as

$(Actual_t \times AdjFac_{t^*} - MedEst_{t^*}) \times Shares_{t^*}/MktCap_{t-1}$ where t^* is the month in which the forecast for period t is recorded (the I/B/E/S statistical period) and $Shares_{t^*}$ is the number of shares outstanding at the forecast date. An advantage of this method is that adjustment factors for both the forecast and actual value are precisely known at the forecast date.

Unfortunately, rounding errors can still play a role. I/B/E/S provides data both on an unadjusted basis and adjusted for stock splits. Forecast values are rounded to the nearest

penny (two decimal places), while actual values are reported with four-digit accuracy in the adjusted files and with two digits in the unadjusted files. Although I use the higher precision number, forecast errors might not be zero even in the case of a perfect forecast. For example, consider an actual EPS value of 1.16 that is shrunk to 0.0967 due to a 12:1 stock split. The corresponding induced forecast error is $0.0967 \times 12 - 1.16 = 0.0004$ per share.

"Unadjusted" method: One solution that does not suffer from these rounding errors is to use *unadjusted* actuals and forecasts. This requires correcting for stock splits that occur between the forecast date and the earnings announcement date. I obtain actual EPS values from the unadjusted details file and match adjustment factors in the reporting month. Forecast errors are defined as $(AdjFac_{t^*} / AdjFac_t \times Actual_t - MedEst_{t^*}) \times Shares_{t^*} / MktCap_{t-1}$ where t^* again is the I/B/E/S statistical period. A small disadvantage of this method is that adjustment factors are incorrect in some cases, since the I/B/E/S effective split date (once per statistical period = once per month) is not necessarily the exact true date of a stock split.

"Unadjusted" actuals plus random error: A third methodology is inspired by the observation that reported EPS and EPS forecasts are usually rounded to the nearest penny by firms and data providers—in addition to rounding errors caused by stock splits. Although the underlying (latent) distribution of earnings and forecasts is likely to be continuous, analysts usually report their estimates as a rounded dollar figure. Similarly, databases record estimates and actual EPS on a rounded basis with two or four digits. Rounding can be seen as a source of measurement error, since the true value lies somewhere within the range [EPS−0.005, EPS+0.005), if EPS values are rounded to the nearest penny. EPS estimates are rounded in a similar way, aggravating the uncertainty in earnings forecast errors.

We can account for this uncertainty due to discretization by re-introducing just enough additional variance to cover the whole real line and not just discrete 0.01-spaced values. I therefore use original "unadjusted" forecast errors $(AdjFac_{t^*} / AdjFac_t \times Actual_t - MedEst_{t^*})$

and add an error term distributed uniformly over the interval (−0.005, 0.005). This conservative approach approximates a data-generating process in which only final forecast errors are rounded, without individually rounding EPS estimates and actual values. Since the distribution of raw forecast errors can be recovered by rounding to the nearest penny, results from this method using an uncertainty-augmented distribution can be directly compared with the "unadjusted" method.

Conforming to earlier studies, I measure forecast errors at five forecast horizons: in the reporting month and one, three, six, and nine months before the earnings announcement date. I perform separate analyses for the time period 1986–2000 (chosen by Burgstahler and Eames, 2006) and the whole sample period 1986–2010. The yearly subsamples' sizes vary between 10 605 and 46 175, which are only slightly smaller on a yearly basis than those reported by Burgstahler and Eames (2006).

5.1.   Results

5.1.1.   "Adjusted" vs. "unadjusted" EPS and the role of zeroes

At first glance, histograms of forecast errors computed from I/B/E/S data look similar to those presented by Burgstahler and Eames (2006). A distinct feature of the data is the presence of many forecast errors that are exactly zero. Their number decreases in subsamples for longer forecast horizons (see the shaded areas in Figure 6), which can indicate better forecasts by analysts or, as Burgstahler and Eames argue, forecast management by firms when the earnings announcement date draws closer. Notice the drop in frequency on both sides of the origin if zero earnings errors are excluded. This apparent lack of small negative and positive observations can be evidence against earnings management to *beat* the threshold. It can be interpreted as evidence for management to *meet* the benchmark or as a result of rounding that is unrelated to earnings management, as I will argue in this section. In all cases, the treatment and interpretation of zeroes is crucial to the analysis. I will discuss possible causes for the large number of zeroes and results for alternative ways to deal with them.

[Figure 6 about here]

If we keep observations of exactly zero in the sample as in previous research, results for both BD's and Bollen and Pool's (2009) discontinuity test are highly significant.[16] The bootstrap KDE approach, however, fails to find a reference distribution that can be used for a local discontinuity test. Any reference distribution based on an arbitrarily small bandwidth lies outside a 10 percent confidence interval around the empirical distribution, rendering the distributional approach underlying BD's (1997) and Bollen and Pool's (2009) methodology inadequate to study EPS forecasts due to an inability to establish a reference distribution. Significant discontinuity tests reported by Burgstahler and Eames (2006) that are based on a sample including zero EPS forecast errors are therefore likely to rely on an oversmoothed reference distribution. In other words, the bootstrap test finds that the empirical distribution is discontinuous because of the large point mass at zero, not because earnings have been shifted from the left of zero to the right. This pattern is consistent, however, with firms managing earnings or forecasts to exactly meet the benchmark by eliminating small positive and negative earnings surprises.

To explore this potential effect of management to meet the benchmark in the region around zero, I exclude all zeroes from my analyses in the next step. There are far fewer observations on both sides of zero than expected, which is reflected in the empirical distribution's flat parts around zero and the resulting gaps between the empirical and reference distributions in Figure 7a on both sides of zero. Missing observations on both sides of zero can indicate earnings management or forecast management to exactly hit the benchmark or, as I will argue, the effect of rounding small errors to zero. However, the large

---

[16] Results using Burgstahler and Dichev's (1997) and Bollen and Pool's (2009) tests are similar to those reported in the respective studies. Exact test statistics are available from the author.

spike in the density of forecast errors in a tiny range around zero—corresponding to a steep empirical distribution function at zero—appears in adjusted forecast errors even if errors at zero are excluded. This result is most likely caused by small rounding errors in I/B/E/S-adjusted actual EPS values (see Figure 7a) as described in Section 1.1.

The density spike at zero in adjusted forecast errors disappears if unadjusted actual EPS values are used ("unadjusted" method in Figure 7b, again excluding zeroes). Compared to the "adjusted" method, the region of relatively low density between −0.0002 and 0.0002 persists. If there was earnings management to beat the consensus estimate, it would be indicated by a lower than expected density of earnings to the left of zero and an *increased* density at zero or to the right of zero, with the largest difference between the expected and empirical cumulative distributions being at zero. As can be seen from Figure 7, there is no such difference. The observed cumulative number of observations almost exactly agrees with the expected number at zero.

[Figure 7 about here]

5.1.2. *Comparing alternative discontinuity tests*

Test results for the bootstrap approach differ substantially from prior discontinuity tests. The bootstrap KDE procedure does not detect a discontinuity for all but one of the forecast horizons tested in Table 4 when using adjusted actual EPS. It picks up a significant discontinuity ($z = 2.05$) only for a forecast horizon of nine months in the period 1986–2000, which might well be a chance finding, since it contradicts the BD test ($z = −2.09$) for the same sample. Contrary to the insignificant bootstrap KDE results, the BD test statistics show a consistent pattern with negative and significant tests to the right of zero.

A negative BD test statistic to the right of zero can be interpreted as downward management of earnings to meet the analysts' forecast, which, however, is not supported by the bootstrap KDE approach. The Rule-of-thumb KDE method seemingly detects a

discontinuity at zero in many subsamples, even with the "correct" minus and plus signs to the left and right of zero. This result is due to an inappropriate reference distribution caused by a large bandwidth that is rejected by the bootstrap test. It underlines the Rule-of-thumb KDE test's tendency to produce spurious significance in light of conflicting results produced by the bootstrap KDE approach, which tests the reference distribution instead of assuming its appropriateness.

[Table 4 about here]

Applying all three methodologies to unadjusted forecast errors (middle columns in Table 4) produces significant test statistics of the same magnitude as for adjusted EPS, but provides evidence against earnings management to beat the forecast benchmark. Test statistics for the bootstrap KDE method are negative on both sides of zero, which contradicts the hypothesis that additional observations to the right of zero are generated by firms that clear the zero benchmark. Rule-of-thumb KDE test results, on the other hand, are significant with the expected sign under earnings management to beat the benchmark for all forecast horizons. Again, these results are rejected by the bootstrap test as using an inappropriate reference distribution with an unsuitably large kernel bandwidth. A conclusion could be that firms try to avoid earnings surprises regardless of sign and aim to *exactly* meet the forecast. This type of earnings management can account for the large number of zeroes we observe.

5.1.3. *A discontinuity caused by rounding of published EPS figures*

Despite the plausibility of earnings management to exactly meet the benchmark, an investigation of the distribution of EPS forecast errors cannot distinguish this explanation from a more parsimonious one based on rounding, which can also explain the large number of zeroes. If analysts' net income forecasts are very close to the actual value reported by the

company, this small difference does not show up in EPS figures rounded to the nearest penny, effectively suggesting a zero forecast error. The corresponding number of forecast errors that are exactly zero will be highly inflated, which is supported by the large number of zeroes in Figure 6. If EPS was measured on a continuous scale that is not limited to multiples of 0.01, the probability of an exactly zero error would be zero, since there would be no need for rounding.

To see the impact of rounding on the underlying distributions of forecasts and reported earnings, consider the following data generation process. Suppose that reported *total* earnings and estimates can be measured with arbitrary (subpenny) precision. A data provider (e. g., I/B/E/S) or an analyst collects forecasts and actual values for *total* net income, computes the difference between the two per share, and *then* rounds it to two decimal places. We only observe this rounded EPS figure. For example, a firm might report a total net income of USD 90,000.00 against a forecast of USD 91,000.00. On the basis of one million common shares, this yields actual EPS of USD 0.09 and an EPS forecast error of USD −0.001. Since some forecast errors—like this one—are rounded to zero before we observe them, scaled EPS are zero as well, and we can choose to analyze the distribution of scaled EPS either including or excluding these zeroes. If we choose to keep the zeroes in our analysis, the large number of them—the positive point mass at zero—prevents us from establishing a reference distribution, and statistical tests based on this distribution cannot be applied. On the other hand, if we exclude all zeroes, tests for discontinuities will often, and unsurprisingly, find this lack of zeroes that were small forecast errors prior to rounding. Both the failure to establish a reference distribution and a significant discontinuity at zero are caused entirely by rounding.

A simple test can be devised to assess the sensitivity of prior findings of discontinuities to this type of rounding. We can use the difference between EPS estimates and actual values and add a small error to each observation representing our uncertainty about the underlying precise value. An error distributed uniformly over the interval (USD −0.005, USD 0.005) is a

natural choice, because it resembles the rounding mechanism and does not change the order of nonidentical observations. When added to each forecast error before scaling by share price, this choice returns the original distribution when rounding again to two digits. Note that previously excluded zeroes remain in the sample, as they become nonzero due to the random error component.

Results for this third approach using "unadjusted" actuals plus a random rounding error suggest that the significance of discontinuity tests heavily depends on whether rounding is applied to forecasts before they are made public. To get an intuition for the effect of adding random noise to forecast errors, consider Figure 7, which shows a fitted kernel density estimate for the period 1986–2010 and a three-month forecast horizon in panel (c). While results based on forecast errors without error component are highly significant for the three-month horizon (panel b), forecast errors are very smooth around zero without a trace of a discontinuity in panel (c) if we account for rounding. Additional observations introduced by adding back the "noisy" zeroes seem to fill the gap around zero in such a way that the discontinuity disappears. Test statistics using the bootstrap KDE test on forecast errors with random component shown in Table 4 (rightmost columns) are insignificant now for all forecast horizons that previously showed discontinuities. Both the Rule-of-thumb KDE test and the BD test still find discontinuities for all but one forecast horizon.

In summary, there is evidence against a discontinuity in forecast errors within a three-month period prior to earnings announcements. For six- and nine-month horizons, the bootstrap approach fails to find a reduced number of firms to the left of zero but suggests a slightly increased number of small positive forecast errors, contrary to Burgstahler and Eames' (2006) findings. An implication is that firms might engage in tactics to beat analysts' forecasts long before the actual earnings announcement, perhaps through real earnings management. Closer inspection of the distribution of forecast errors at long horizons,

however, reveals that there is no marked change in the density around zero, which casts doubt on earnings management as an explanation for the positive test result.

Not being able to detect a discontinuity in the distribution of forecast errors with added rounding noise does not, of course, disprove the existence of earnings management to beat or meet analysts' forecasts within the half penny range around zero. Firms that would otherwise show small negative or positive surprises might still manage earnings to exactly meet the forecast benchmark within this half penny range that cannot be observed in rounded EPS data. In view of the absence of firms in the flat regions of the empirical distribution around zero, it seems unlikely, however, that *all* firms engage in such earnings management. Finally, these findings underline how easily a discontinuity-free distribution can be generated that shows a discontinuity that looks like earnings management when rounding is applied to the underlying forecast errors.

6. Conclusion

In this paper, I evaluate the distributional assumptions underlying discontinuity tests for earnings management and propose a new test procedure to detect discontinuities in empirical distributions. This procedure is based on a continuous reference distribution constructed by kernel density estimation and a bootstrap test that fits the kernel density estimate to the data and thereby ensures the appropriateness of this reference distribution. Prior studies did not test whether their assumptions regarding bandwidth selection and reference distributions were supported by the data. The bootstrap approach uses fewer assumptions, performs better and overcomes several shortcomings of BD's and Bollen and Pool's (2009) methods. A wide range of empirical data can be tested, such as standardized earnings, earnings changes, or forecast errors relative to analysts' predictions.

Comparisons between several bandwidth selection criteria and kernel functions reveal a substantial tendency towards significant results in prior research. This bias is particularly severe for the Gaussian kernel function that is predominantly used in the literature.

Bandwidths selected by the bootstrap method are often much smaller than those in prior studies, indicating that the reference distributions used in these studies do not represent the data. The bootstrap kernel density estimate generates more conservative and more stable results than the procedures employed by BD and Bollen and Pool (2009). Since a source of errors in significance tests appears to be the choice of kernel function, working with finite kernels such as Epanechnikov or uniform kernels seems superior to kernels that take *all* available data into account when testing for a *local* discontinuity. Overall, results indicate that selecting the correct bandwidth for any given kernel function is more important than the researcher's choice of a kernel function.

Earnings data from U.S. firms in the time period 1976–2010 show a discontinuity at zero for standardized earnings and earnings changes. In contrast to findings of discontinuities in the whole sample of earnings and earnings changes, discontinuities do not exist in many yearly subsamples. Results show a large drop in loss aversion after 2003. Although other studies suggest an increasing importance to meet or beat analysts' forecasts compared to avoiding small losses or earnings decreases, I find a trend only for standardized earnings but not for changes in earnings.

When applied to errors in analysts' forecasts of standardized EPS, the bootstrap KDE method provides evidence against earnings management to beat analysts' forecasts. Results suggest that prior findings of discontinuities in forecast errors are due to misspecifying the reference distribution that is used to calculate test statistics. The bootstrap test rejects all reference distributions with positive bandwidths if zeroes are not excluded from the sample. If we exclude zeroes, the bootstrap KDE method does not detect a discontinuity in forecast errors based on split-adjusted actual EPS. A gap on both sides of zero in forecast errors based on "unadjusted" actual EPS, however, can be interpreted as earnings management to meet the benchmark.

Discontinuities found in forecast errors can be explained more parsimoniously by a simple rounding mechanism in reported EPS. Forecast errors are often rounded off to zero, because EPS figures used in the numerator are usually provided with only two-digit accuracy. This rounding simultaneously generates a lack of observations on both sides of zero and a large point mass at zero, which can be mistaken for earnings management. Discontinuities disappear if we add back a small error to each observation that represents the actual rounding in two-digit EPS forecasts. This suggests that earnings management detected in EPS forecasts takes place within a half-penny range around the analyst forecast or is an artifact created by rounding. A measure of forecast errors with higher accuracy, possibly based on total net income, may be used in future research to resolve this question.

References

Athanasakou, V., Strong, N. C. and Walker, M. (2011) The market reward for achieving analyst earnings expectations: Does managing expectations or earnings matter? *Journal of Business Finance & Accounting*, 38(1-2), pp. 58–94.

Athanasakou, V. E., Strong, N. C. and Walker, M. (2009) Earnings management or forecast guidance to meet analyst expectations? *Accounting & Business Research*, 39(1), pp. 3–35.

Bartov, E., Givoly, D. and Hayn, C. (2002) The rewards to meeting or beating earnings expectations, *Journal of Accounting and Economics*, 33(2), pp. 173–204.

Beatty, A. L., Ke, B. and Petroni, K. R. (2002) Earnings management to avoid earnings declines across publicly and privately held banks, *The Accounting Review*, 77(3), pp. 547–570.

Beaver, W. H., McNichols, M. F. and Nelson, K. K. (2003) Management of the loss reserve accrual and the distribution of earnings in the property-casualty insurance industry, *Journal of Accounting and Economics*, 35(3), pp. 347–376.

Beaver, W. H., McNichols, M. F. and Nelson, K. K. (2007) An alternative interpretation of the discontinuity in earnings distributions, *Review of Accounting Studies* 12(4), pp. 525–56.

Bollen, N. P. B. and Pool, V. K. (2009) Do hedge fund managers misreport returns? Evidence from the pooled distribution, *Journal of Finance*, 64(5), pp. 2257–2288.

Brown, L. D. (2001) A temporal analysis of earnings surprises: Profits versus losses, *Journal of Accounting Research*, 39(2), pp. 197–219.

Brown, L. D. and Caylor, M. L. (2005) A temporal analysis of quarterly earnings thresholds: Propensities and valuation consequences, *The Accounting Review*, 80(2), pp. 423–440.

Brown, S., Kang, M., In, F. and Lee, G. (2010) Resisting the manipulation of performance metrics: An empirical analysis of the manipulation-proof performance measure, Working Paper, Available at http://ssrn.com/abstract=1536323.

Burgstahler, D. and Dichev, I. (1997) Earnings management to avoid earnings decreases and losses, *Journal of Accounting and Economics*, 24(1), pp. 99–126.

Burgstahler, D. and Chuk, E. (2012) What have we learned about earnings management? Correcting disinformation about discontinuities. Working paper, University of Washington, Available at www.ssrn.com/abstract=1866008.

Burgstahler, D. and Eames, M. (2006) Management of earnings and analysts' forecasts to achieve zero and small positive earnings surprises, *Journal of Business Finance & Accounting*, 33(5/6), pp. 633–652.

Burgstahler, D. C. and Eames, M. J. (2003) Earnings management to avoid losses and earnings decrease: Are analysts fooled? *Contemporary Accounting Research*, 20(2), pp. 253–294.

Cassar, G. and Gerakos, J. (2011) Hedge funds: Pricing controls and the smoothing of self-reported returns. *Review of Financial Studies* 24(5), pp. 1698–1734.

Charoenwonga, C. and Jiraporn, P. (2009) Earnings management to exceed thresholds: Evidence from Singapore and Thailand, *Journal of Multinational Financial Management*, 19(3), pp. 221–236.

Cohen, D. A. and Lys, T. Z. (2003) A note on analysts' earnings forecast errors distribution, *Journal of Accounting and Economics*, 36(1-3), pp. 147–164.

Coulton, J., Taylor, S. and Taylor, S. (2005) Is 'benchmark beating' by Australian firms evidence of earnings management? *Accounting and Finance*, 45(4), pp. 553–576.

Dechow, P. M., Richardson, S. A. and Tuna, I. (2003) Why are earnings kinky? An examination of the earnings management explanation, *Review of Accounting Studies*, 8(2-3), pp. 355–384.

Dechow, P. M., Sloan, R. G. and Sweeney, A. P. (1995) Detecting earnings management, *The Accounting Review*, 70(2), pp. 193–225.

Degeorge, F., Patel, J. and Zeckhauser, R. (1999) Earnings management to exceed thresholds, *Journal of Business*, 72(1), pp. 1–33.

Donelson, D. C., McInnis, J. M. and Mergenthaler, R. D. (forthcoming) Discontinuities and earnings management: Evidence from restatements related to securities litigation. *Contemporary Accounting Research*.

Durtschi, C. and Easton, P. (2005) Earnings management? The shapes of the frequency distributions of earnings metrics are not evidence ipso facto, *Journal of Accounting* Research, 43(4), pp. 557–592.

Durtschi, C. and Easton, P. (2009) Earnings management? Erroneous inferences based on earnings frequency distributions, *Journal of Accounting Research* 47(5), pp. 1249–1281.

Eames, M. J. and Kim, Y. (2012) Analyst vs. Market Forecasts of Earnings Management to Avoid Small Losses, *Journal of Business Finance and Accounting*, 39(5-6), pp. 649–674.

Glaum, M., Lichtblau, K. and Lindemann, J. (2004) The extent of earnings management in the U.S. and Germany, *Journal of International Accounting Research*, 3(2), pp. 45–77.

Gore, P., Pope, P. F. and Singh, A. K. (2007) Earnings management and the distribution of earnings relative to targets: UK evidence, *Accounting and Business Research*, 37(2), pp. 123–150.

Graham, J. R., Harvey, C. R. and Rajgopal, S. (2005) The economic implications of corporate financial reporting, *Journal of Accounting and Economics*, 40(1-3), pp. 3–73.

Green, J. R. (2010) Discretion, managerial incentives, and market conditions: The misreporting of hedge fund returns, Ph.D. dissertation, University of North Carolina at Chapel Hill.

Günther, N., Gegenfurtner, B., Kaserer, C. and Achleitner, A.-K. (2009) International financial reporting standards and earnings quality: The myth of voluntary vs. mandatory adoption, CEFS Working Paper, Technical University Munich, Available at http://ssrn.com/abstract=1413145.

Herrmann, D., Hope, O-K., Payne, J. L. and Thomas, W. B. (2011) The market's reaction to unexpected earnings thresholds, *Journal of Business Finance and Accounting*, 38(1-2), 34–57.

Holland, D. (2004) Earnings management: A methodological review of the distribution of reported earnings approach, Working Paper, Department of Accounting and Finance, Monash University.

Jones, M. C., Marron, J. S. and Sheather, S. J. (1996) A brief survey of bandwidth selection for density estimation, *Journal of the American Statistical Association*, 91(433), pp. 401–407.

Koh, K., Matsumoto, D. A. and Rajgopal, S. (2008) Meeting or beating analyst expectations in the post-scandals world: Changes in stock market rewards and managerial actions, *Contemporary Accounting Research*, 25(4), pp. 1067–1098.

McNichols, M. F. (2000) Research design issues in earnings management studies, *Journal of Accounting and Public Policy*, 19(4-5), pp. 313–345.

Parzen, E. (1962) On estimation of a probability density function and mode, *The Annals of Mathematical Statistics*, 33(3), pp. 1065–1076.

Payne, J. L. and Robb, S. W. (2000) Earnings management: The effect of ex ante earnings expectations, *Journal of Accounting, Auditing & Finance*, 15(4), pp. 371–392.

Payne, J. L. and Thomas, W. B. (2003) The implications of using stock-split adjusted I/B/E/S data in empirical research, *The Accounting Review*, 78(4), pp. 1049–1067.

Pinnuck, M. and Lillis, A. M. (2007) Profits versus losses: Does reporting an accounting loss act as a heuristic trigger to exercise the abandonment option and divest employees? *The Accounting Review*, 82(4), pp. 1031–1053.

Robinson, D. and Glushkov, D. (2006) A note on IBES unadjusted data, Technical Note, Wharton Research Data Services.

Rosenblatt, M. (1956) Remarks on some nonparametric estimates of a density function, *The Annals of Mathematical Statistics*, 27(3), pp. 832–837.

Scott, D. W. (1992) *Multivariate Density Estimation*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons, Inc.

Silverman, B. W. (1986) *Density estimation for statistics and data analysis*. London: Chapman and Hall.

Takeuchi, Y. (2004) On a statistical method to detect discontinuity in the distribution function of reported earnings, *Mathematics and Computers in Simulation*, 64, pp. 103–111.

Talha, M., Mohamad, M. and Sallehhuddin, A. (2008) Avoiding losses and earnings decrease by malaysian companies before and after major corporate governance initiatives, *International Journal of Managerial and Financial Accounting*, 1(1), pp. 45–63.

Tung, S., Lan-Fen, W., Chen-Chang, L., Ching-Hui, L. and Hsu, A. (2008) The use of extended credit (channel stuffing) to avoid reporting losses, *Asia-Pacific Journal of Accounting & Economics*, 15(2), pp. 161–182.

A Appendix

6.1.1. A.1 Proof of equation (3)

Assume $E\left[(n_{i-1}+n_{i+1})/2 - n_i\right] = 0$, where E is the expectation operator with respect to the

sampling population. Then

$$
\begin{aligned}
Var\left(\frac{n_{i-1}+n_{i+1}}{2} - n_i\right) &= E\left[\left(\frac{n_{i-1}+n_{i+1}}{2} - n_i - E\left(\frac{n_{i-1}+n_{i+1}}{2} - n_i\right)\right)^2\right] \\
&= E\left[\left(\frac{n_{i-1}+n_{i+1}}{2}\right)^2 - 2n_i\frac{n_{i-1}+n_{i+1}}{2} + n_i^2\right] \\
&= \frac{1}{4}E\left[n_{i-1}^2 + 2n_{i-1}n_{i+1} + n_{i+1}^2\right] - E\left[n_in_{i-1} + n_in_{i+1}\right] + E\left[n_i^2\right] \quad (14) \\
&= \frac{1}{4}\left(Np_{i-1}(1-p_{i-1}) - 2Np_{i-1}p_{i+1} + Np_{i+1}(1-p_{i+1})\right) \\
&\quad + Np_ip_{i-1} + Np_ip_{i+1} + Np_i(1-p_i) \\
&= Np_i(1-p_i) + \frac{N}{4}(p_{i-1}+p_{i+1})(1-p_{i-1}-p_{i+1}) + Np_i(p_{i-1}+p_{i+1}).
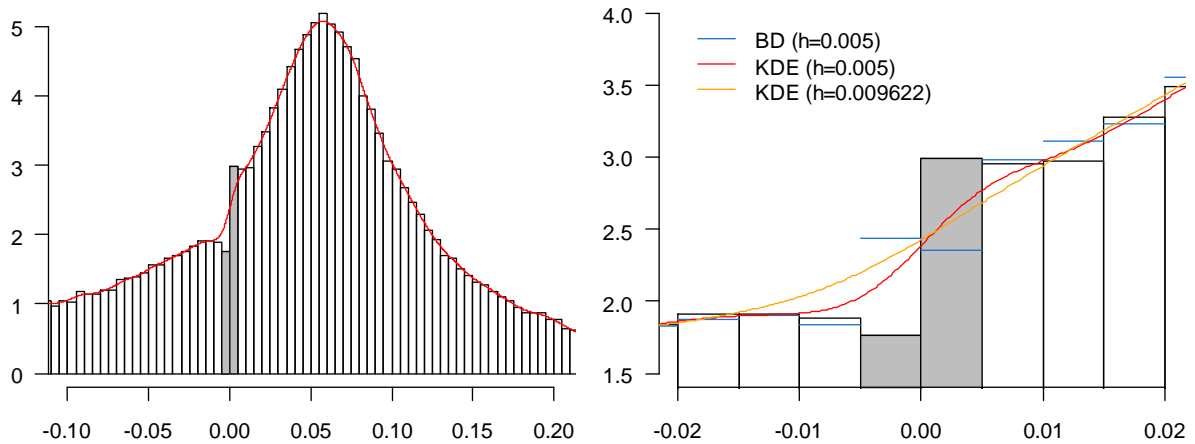\end{aligned}
$$

Figure 1. Relaxing the linearity assumption with kernel density estimation.
A smooth density function produced by kernel density estimation can relax the assumption of a locally linear density. Both graphs show the density of standardized earnings (net income divided by market capitalization as described in section 4) at bin width 0.005; the graph on the right is magnified to highlight the region of interest. Blue horizontal lines represent the density obtained from linearly interpolating histogram bins in Burgstahler and Dichev's (1997) method. The red and orange lines are two Gaussian kernel density estimates with bandwidths 0.005 (red) and 0.009622 (orange). Tails of the distribution are not displayed. A discontinuity test is performed on the difference between the number of firms in each bin (highlighted in grey) and the area under the blue lines in Burgstahler and Dichev's test or the area under the kernel estimate in Bollen and Pool's (2009) method and the bootstrap KDE. Note that this difference is smaller for both density estimates compared to BD's method.
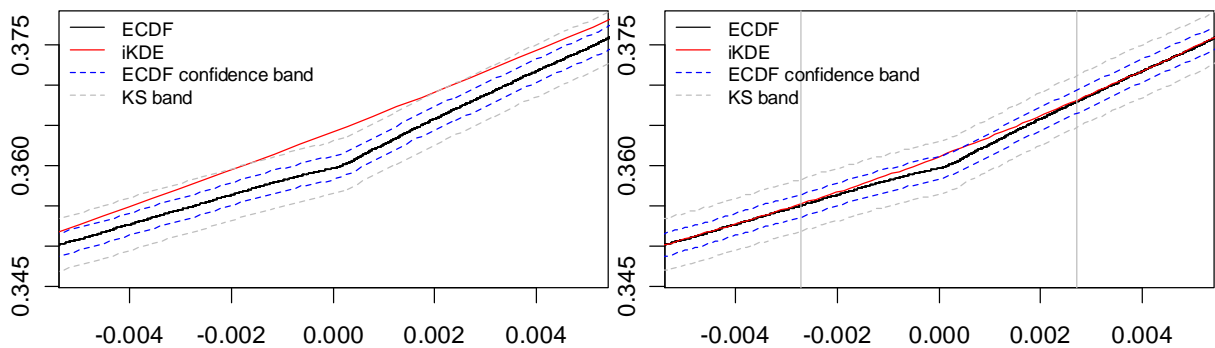
Figure 2. Choosing an appropriate bandwidth for kernel density estimation.
Bandwidth is chosen such that the kernel density estimate exactly meets a confidence band for the empirical distribution of scaled earnings (data description in section 4). The solid black line in both panels is the empirical cumulative distribution (ECDF) of net income scaled by market capitalization, corresponding to the histogram in Figure 1. The red line shows an integrated kernel density estimate (iKDE) of the underlying distribution with bandwidth $h=0.009594$ in the left panel and $h=0.002713$ in the right panel. The left panel shows an oversmoothed density estimate (iKDE), whereas in the right panel bandwidth has been adjusted to produce a density estimate that fits the data. Bandwidth h is determined by Silverman's (1986) rule of thumb on the left and the bootstrap method on the right, using a Gaussian kernel in both cases. Blue dashed lines are 0.1/0.9-confidence bands for the empirical distribution. The grey dashed lines are the Kolmogorov-Smirnov confidence bands at the 5 percent level. Vertical lines are at $0\pm h$, which is outside the region shown in the left panel.
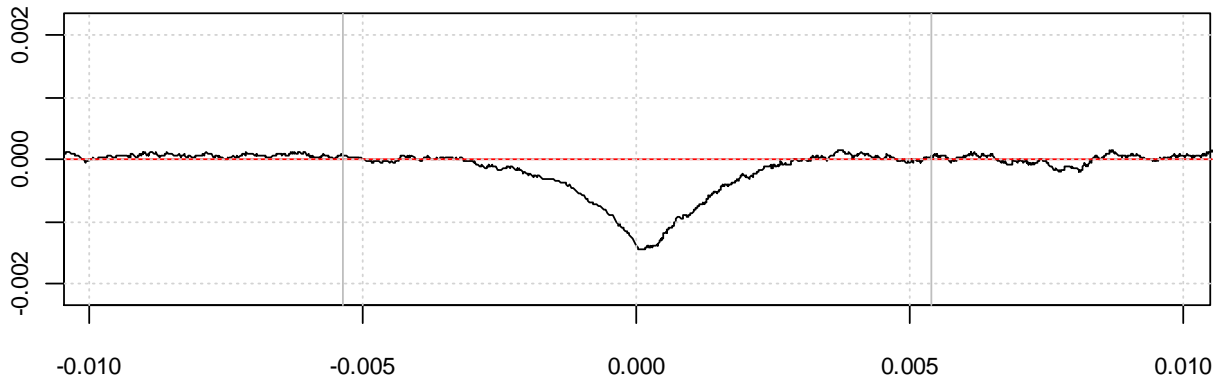
Figure 3. Difference between empirical cumulative distribution and reference distribution $(|S_N(X) - F(X)|)$ for standardized earnings in the period 1976–2010.

Grey solid lines show left and right boundaries for an Epanechnikov kernel centered at zero.
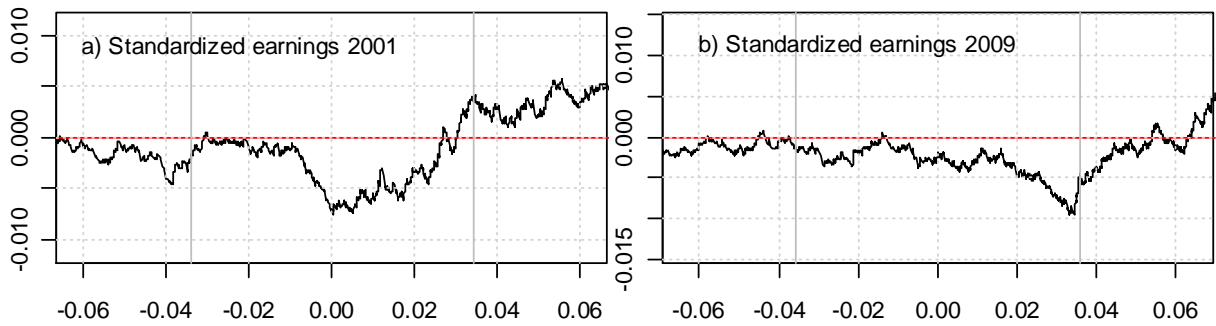
Figure 4. Difference between empirical cumulative distribution and reference distribution for standardized earnings in 2001 (left) and 2009 (right).

Note the discontinuity at zero in 2001 and its absence in 2009. Grey solid lines show left and right boundaries for an Epanechnikov kernel centered at zero.
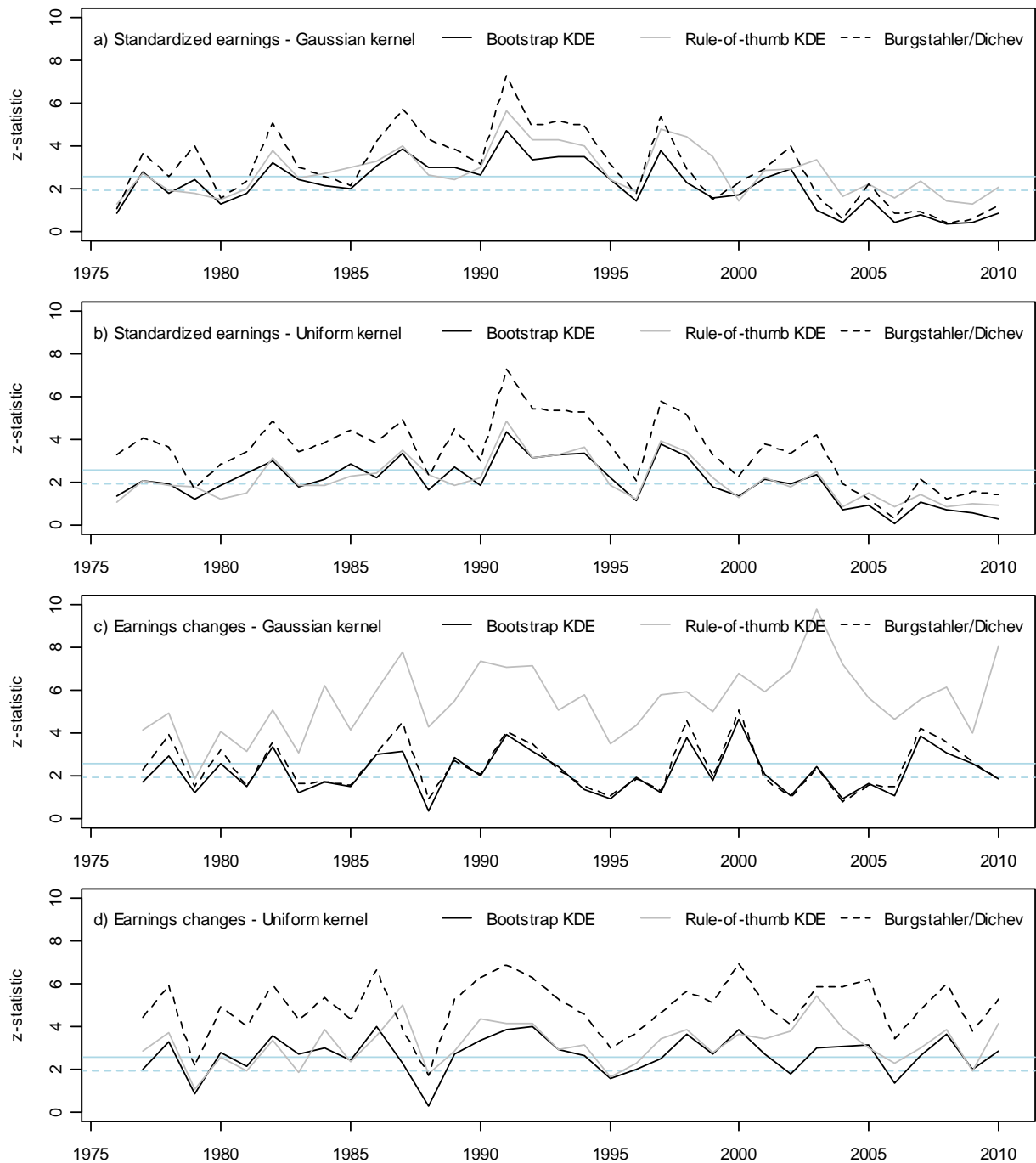
Figure 5. Standardized earnings and earnings changes for Gaussian and uniform kernel.
The upper two panels show results for standardized earnings calculated as those in Table 3 but using a Gaussian kernel (upper panel) and a uniform kernel (lower panel) to estimate the reference distribution. The lower two panels show results for the same estimation procedures applied to earnings changes. The blue lines are at the 2 percent (solid line) and 5 percent (dashed line) confidence levels.
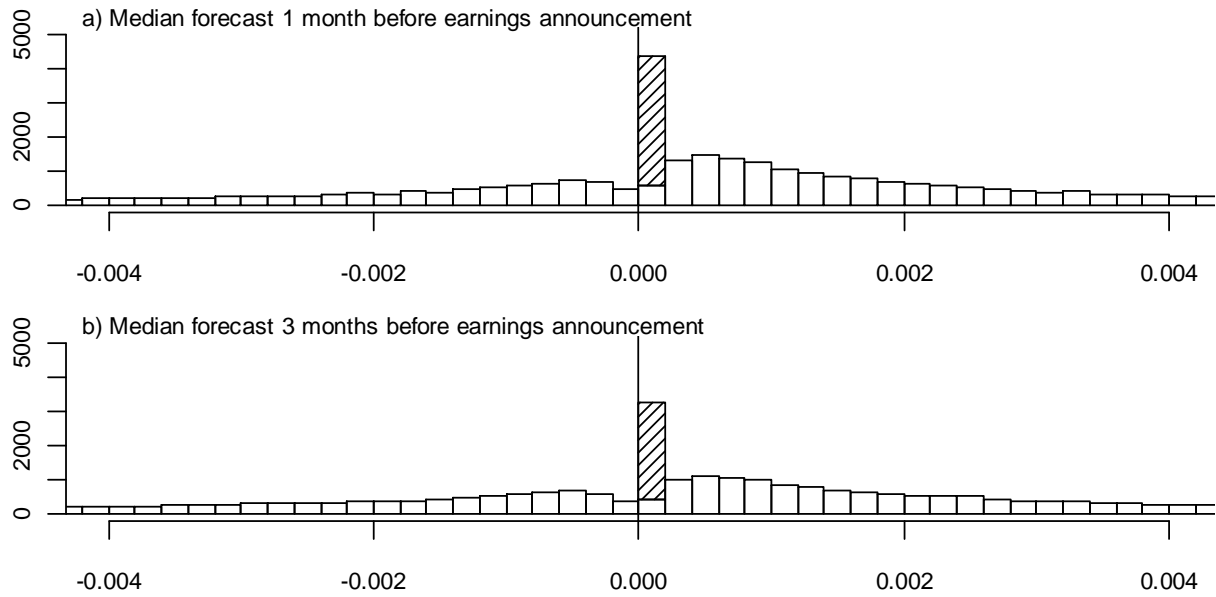
Figure 6. Earnings forecast errors from adjusted actuals.
Forecast errors are calculated using unadjusted I/B/E/S median estimates one month (panel a) and three months (panel b) before the respective earnings announcement over the time period 1986–2010. Actual values are taken from the I/B/E/S adjusted summary file and corrected for stock splits by matching the I/B/E/S adjustment factor at the forecast date. The shaded area shows errors that are exactly zero. Histograms for "unadjusted" errors are very similar to the ones shown here.
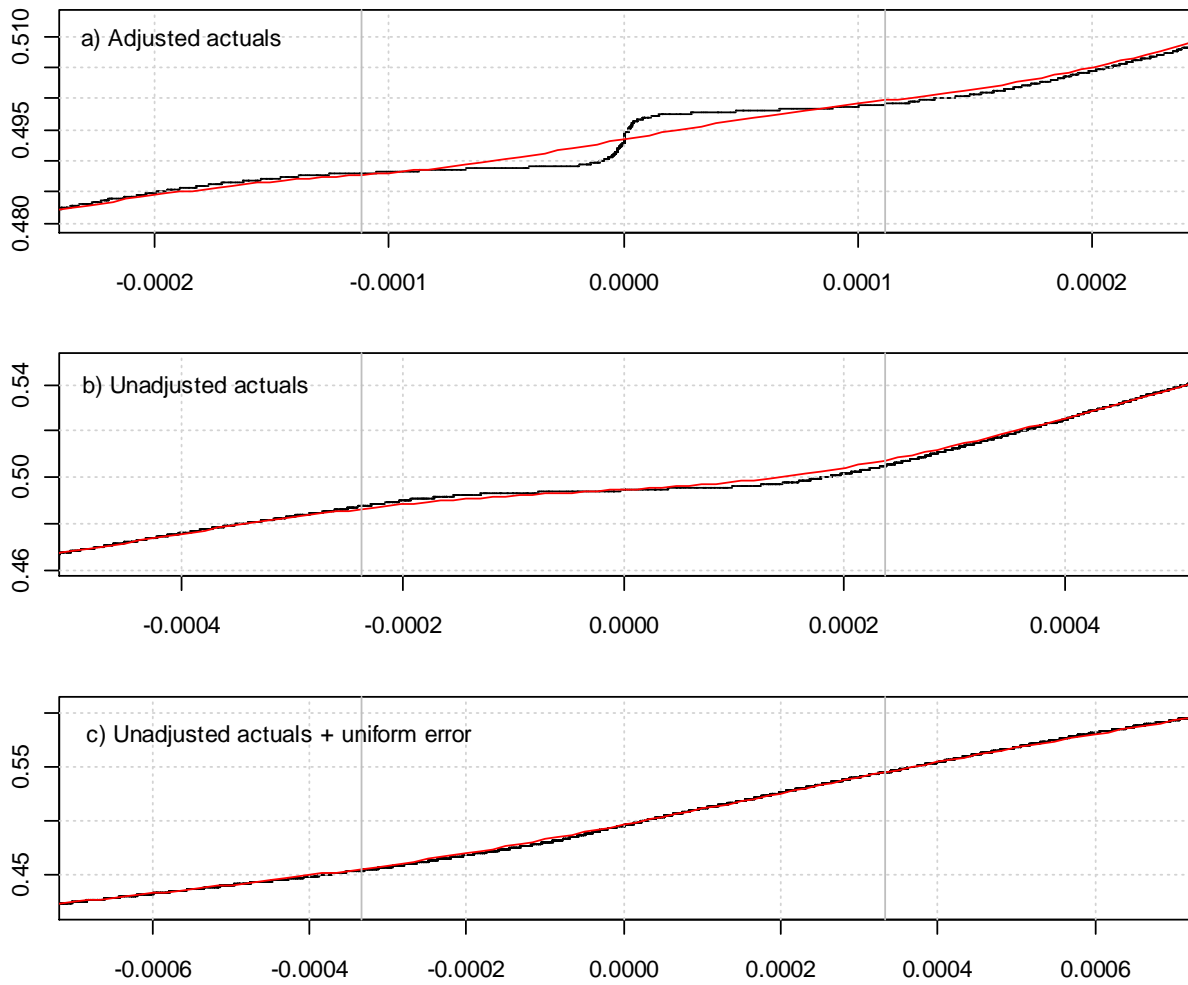
Figure 7. Observed vs. expected distribution of forecast errors.

The empirical distribution of forecast errors three months before the earnings announcement is shown as the black line relative to an integrated kernel density estimate in red over the time period 1986–2010. Forecast errors in panel (a) are calculated using unadjusted I/B/E/S median estimates and actual values taken from the I/B/E/S adjusted summary file and corrected for stock splits by matching the I/B/E/S adjustment factor at the forecast date. Panel (a) corresponds to panel (b) in Figure 6. Forecast errors in panel (b) are calculated from unadjusted median estimates and actuals, while both are put on the same per-share basis by matching I/B/E/S adjustment factors at the forecast and report dates. Panel (c) shows results similar to panel (b) with a uniform (-0.005, 0.005) error added to earnings surprises before scaling by share price. Vertical lines show the boundaries of an Epanechnikov kernel. All calculations exclude forecast errors that are exactly zero, since no reference distribution can be established when keeping zeroes in the sample. Adding a random error, however, adds back zeroes in panel (c) that are excluded in (a) and (b).

Table 1. Standardized earnings

This table shows results for discontinuity tests on standardized earnings ($Earnings_t$ / MV_{t-1}). The main difference between panels (and cells) is the method of obtaining the kernel bandwidth or bin width, h. *Bootstrap KDE* fits a kernel density estimator to serve as a reference distribution by using the bandwidth obtained from the procedure described in section 3 with a significance level of 0.1 at the bootstrap step. *Rule-of-thumb KDE* uses Silverman's (1986: 48, eq. 3.31) rule of thumb and *Burgstahler/Dichev* employs the test statistic in Equation (2) based on histogram bins. Bin width for the Burgstahler/Dichev method is chosen to be equal to the bootstrap KDE. Significance levels (*P*) are calculated by performing a *z*-test on Equation (12) and a binomial test on the number of observations. Results for the binomial test are shown for the bootstrap KDE test only, since they are very similar to the *z*-test. *P*-values are not corrected for testing the joint hypothesis of no effect to the left and right of zero and therefore represent lower bounds on the probability that the null is confirmed. For example, "4.593" in the 1976–1994 block in the upper panel shows the test statistic for an interval of width h=0.005739 to the right of zero, which is obtained by the bootstrap KDE method.

Period/N	Interval	Bootstrap KDE			Rule-of-thumb KDE		Burgstahler/Dichev	
		Std. Diff.	P(normal)	P(bin.)	Std. Diff.	P(normal)	Std. Diff.	P(normal)
Epanechnikov kernel								
		h=0.005739			h=0.010622		h=0.005739	
1976–1994	(-h, 0]	-6.542	0.000	0.000	-8.311	0.000	-14.292	0.000
N=78923	(0, h]	4.593	0.000	0.000	6.277	0.000	10.023	0.000
		h=0.011515			h=0.011474		h=0.011515	
1995–2010	(-h, 0]	-3.179	0.001	0.001	-3.215	0.001	-6.449	0.000
N=96426	(0, h]	2.250	0.024	0.025	2.200	0.028	3.090	0.002
		h=0.005385			h=0.009622		h=0.005385	
1976–2010	(-h, 0]	-5.942	0.000	0.000	-7.221	0.000	-12.415	0.000
N=175349	(0, h]	5.038	0.000	0.000	5.897	0.000	10.006	0.000
Gaussian kernel								
		h=0.002670			h=0.010622		h=0.002670	
1976–1994	(-h, 0]	-7.098	0.000	0.000	-12.345	0.000	-10.240	0.000
N=78923	(0, h]	6.261	0.000	0.000	8.597	0.000	7.571	0.000
		h=0.004986			h=0.011474		h=0.004986	
1995–2010	(-h, 0]	-3.641	0.000	0.000	-5.470	0.000	-4.715	0.000
N=96426	(0, h]	3.165	0.002	0.002	1.692	0.091	3.891	0.000
		h=0.002474			h=0.009622		h=0.002474	
1976–2010	(-h, 0]	-6.693	0.000	0.000	-11.243	0.000	-9.433	0.000
N=175349	(0, h]	6.276	0.000	0.000	7.547	0.000	7.981	0.000
Uniform kernel								
		h=0.004301			h=0.010622		h=0.004301	
1976–1994	(-h, 0]	-7.405	0.000	0.000	-10.512	0.000	-12.923	0.000
N=78923	(0, h]	6.081	0.000	0.000	8.114	0.000	9.009	0.000
		h=0.009190			h=0.011474		h=0.009190	
1995–2010	(-h, 0]	-3.844	0.000	0.000	-4.094	0.000	-6.075	0.000
N=96426	(0, h]	3.018	0.003	0.003	2.560	0.010	4.020	0.000
		h=0.004080			h=0.009622		h=0.004080	
1976–2010	(-h, 0]	-6.706	0.000	0.000	-9.198	0.000	-11.169	0.000
N=175349	(0, h]	6.312	0.000	0.000	7.494	0.000	9.490	0.000

Table 2. Earnings changes

This table shows results for discontinuity tests on earnings changes ($(Earnings_t - Earnings_{t-1}) / MV_{t-2}$). The main difference between panels (and cells) is the method of obtaining the kernel bandwidth or bin width, h. *Bootstrap KDE* fits a kernel density estimator to serve as a reference distribution by using the bandwidth obtained from the procedure described in section 3 with a significance level of 0.1 at the bootstrap step. *Rule-of-thumb KDE* uses Silverman's (1986: 48, eq. 3.31) rule of thumb and *Burgstahler/Dichev* employs the test statistic in Equation (2) based on histogram bins. Bin width for the Burgstahler/Dichev method is chosen to be equal to the bootstrap KDE. Significance levels (*P*) are calculated by performing a *z*-test on Equation (12) and a binomial test on the number of observations. Results for the binomial test are shown for the bootstrap KDE test only, since they are very similar to the *z*-test. *P*-values are not corrected for testing the joint hypothesis of no effect to the left and right of zero and therefore represent lower bounds on the probability that the null is confirmed. Intervals are denoted by their upper boundary. For example, "3.410" in the 1977–1994 block in the upper panel shows the test statistic for an interval of width $h=0.004916$ to the right of zero, which is obtained by the bootstrap KDE method.

		Bootstrap KDE			Rule-of-thumb KDE		Burgstahler/Dichev	
Period/N	Interval	Std. Diff.	P(normal)	P(bin.)	Period/N	Interval	Std. Diff.	P(normal)
Epanechnikov kernel								
		h=0.004916			h=0.008150		h=0.004826	
1977–1994	(-h, 0]	-3.283	0.001	0.001	-3.874	0.000	-6.691	0.000
N=71481	(0, h]	3.410	0.001	0.001	3.424	0.001	5.521	0.000
		h=0.006755			h=0.007506		h=0.006248	
1995–2010	(-h, 0]	-1.797	0.072	0.073	-1.257	0.209	-2.961	0.003
N=93571	(0, h]	3.092	0.002	0.002	3.845	0.000	8.826	0.000
		h=0.003951			h=0.006790		h=0.003716	
1977–2010	(-h, 0]	-2.868	0.004	0.004	-3.634	0.000	-5.508	0.000
N=165052	(0, h]	2.883	0.004	0.004	4.511	0.000	5.396	0.000
Gaussian kernel								
		h=0.002345			h=0.008150		h=0.002139	
1977–1994	(-h, 0]	-5.899	0.000	0.000	-5.706	0.000	-8.423	0.000
N=71481	(0, h]	4.339	0.000	0.000	8.133	0.000	5.888	0.000
		h=0.002990			h=0.007421		h=0.002998	
1995–2010	(-h, 0]	-1.035	0.301	0.309	-0.802	0.422	-1.154	0.248
N=93571	(0, h]	2.410	0.016	0.017	10.368	0.000	2.410	0.016
		h=0.001820			h=0.006746		h=0.001912	
1977–2010	(-h, 0]	-4.723	0.000	0.000	-4.606	0.000	-6.304	0.000
N=165052	(0, h]	3.835	0.000	0.000	10.658	0.000	5.099	0.000
Uniform kernel								
		h=0.003692			h=0.008150		h=0.003263	
1977–1994	(-h, 0]	-4.723	0.000	0.000	-4.800	0.000	-7.375	0.000
N=71481	(0, h]	3.577	0.000	0.000	4.685	0.000	5.325	0.000
		h=0.005289			h=0.007421		h=0.005262	
1995–2010	(-h, 0]	-2.175	0.030	0.029	-1.619	0.105	-2.828	0.005
N=93571	(0, h]	2.396	0.017	0.017	5.580	0.000	4.988	0.000
		h=0.002856			h=0.006746		h=0.002829	
1977–2010	(-h, 0]	-3.696	0.000	0.000	-4.444	0.000	-5.383	0.000
N=165052	(0, h]	3.647	0.000	0.000	6.168	0.000	4.929	0.000

Table 3. Standardized earnings and earnings changes over time

This table shows results for discontinuity tests on standardized earnings ($Earnings_t / MV_{t-1}$) and changes in earnings (($Earnings_t - Earnings_{t-1}) / MV_{t-2}$), which are calculated as those in tables 1 and 2 using an Epanechnikov kernel. Each row tests for a discontinuity to the left and right of zero simultaneously. *Std. Diff.* is the maximum absolute value of the two test statistics obtained from comparing the reference distribution (constructed by the bootstrap KDE, Rule-of-thumb KDE, or Burgstahler/Dichev method) to the empirical distribution within an interval of $\pm h$ from zero. Significance levels are adjusted for testing two simultaneous hypotheses using a Bonferroni correction of z-tests, *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

| | | Earnings | | | | | | Changes in earnings | | | | |
| | | Bootstrap KDE | | Rule-of-thumb KDE | | Burgstahler/ Dichev | | Bootstrap KDE | | Rule-of-thumb KDE | | Burgstahler/ Dichev |
Year	N	Std.diff.	h	Std.diff.	h	Std.diff.	N	Std.diff.	h	Std.diff.	h	Std.diff.
1976	3032	1.88	0.0378	0.98	0.0226	5.75 ***						
1977	2987	2.88 ***	0.0357	1.69	0.0177	6.05 ***	2977	2.05 *	0.01446	2.03 *	0.01422	5.54 ***
1978	2927	2.12 *	0.0366	1.60	0.0179	4.55 ***	2887	2.83 ***	0.01231	2.81 ***	0.01229	7.29 ***
1979	3105	0.96	0.0399	1.36	0.0189	2.47 **	2828	1.39	0.02218	0.81	0.01431	4.29 ***
1980	3402	1.39	0.0360	1.05	0.0178	3.26 ***	3017	3.44 ***	0.01980	2.10 *	0.01530	6.88 ***
1981	3569	1.62	0.0367	1.10	0.0189	3.67 ***	3283	3.09 ***	0.02573	1.59	0.01497	7.64 ***
1982	4042	2.02 *	0.0297	2.53 **	0.0201	4.44 ***	3438	2.93 ***	0.02303	2.53 **	0.01439	7.03 ***
1983	4235	1.43	0.0377	1.46	0.0195	3.97 ***	3940	2.20 *	0.02431	1.26	0.01567	5.86 ***
1984	4508	2.00 *	0.0221	1.41	0.0162	4.63 ***	4050	3.20 ***	0.01728	2.68 **	0.01439	8.10 ***
1985	4480	2.40 **	0.0346	1.94	0.0204	5.21 ***	4252	2.87 ***	0.01553	1.73	0.01146	6.75 ***
1986	4516	2.24 **	0.0250	1.99 *	0.0190	4.78 ***	4309	3.54 ***	0.01955	2.56 **	0.01545	8.50 ***
1987	4773	2.18 *	0.0226	2.93 ***	0.0171	3.94 ***	4364	2.97 ***	0.01295	3.47 ***	0.01439	7.08 ***
1988	4865	1.33	0.0264	1.91	0.0196	2.56 **	4517	1.44	0.01234	1.02	0.01196	5.22 ***
1989	4657	2.13 *	0.0351	1.52	0.0206	4.48 ***	4632	2.54 **	0.01737	1.75	0.01415	7.78 ***
1990	4541	2.28 **	0.0229	1.88	0.0198	5.22 ***	4529	3.72 ***	0.01746	3.01 ***	0.01516	9.19 ***
1991	4407	4.23 ***	0.0206	4.07 ***	0.0229	8.15 ***	4492	3.57 ***	0.01772	2.93 ***	0.01479	8.88 ***
1992	4594	2.98 ***	0.0200	2.36 **	0.0175	6.81 ***	4515	2.62 **	0.01720	2.76 **	0.01911	6.93 ***
1993	4883	2.26 **	0.0213	2.68 **	0.0166	5.33 ***	4635	2.92 ***	0.01888	1.85	0.01473	7.16 ***
1994	5400	2.89 ***	0.0149	3.04 ***	0.0135	4.78 ***	4816	2.13 *	0.01337	1.91	0.01282	6.69 ***
1995	5735	1.76	0.0215	1.68	0.0155	3.23 ***	5161	1.86	0.01240	1.07	0.01046	5.41 ***
1996	6312	0.77	0.0249	1.00	0.0150	3.11 ***	5528	2.31 **	0.01591	1.76	0.01215	6.89 ***
1997	6844	2.83 ***	0.0195	3.18 ***	0.0159	6.08 ***	6080	2.35 **	0.01337	2.39 **	0.01147	7.01 ***
1998	6815	2.65 **	0.0216	2.82 ***	0.0170	5.82 ***	6437	2.54 **	0.01627	2.90 ***	0.01176	7.49 ***
1999	6643	1.70	0.0251	1.75	0.0209	3.94 ***	6384	2.50 **	0.01537	1.83	0.01145	7.43 ***
2000	6703	1.39	0.0416	0.99	0.0213	3.71 ***	6280	3.00 ***	0.01858	2.33 **	0.01681	8.15 ***
2001	6402	2.74 **	0.0343	1.78	0.0254	5.63 ***	6363	2.46 **	0.01616	2.32 **	0.01550	6.47 ***
2002	6184	2.03 *	0.0307	1.44	0.0249	5.08 ***	6423	1.76	0.01582	2.29 **	0.01758	6.65 ***
2003	5924	2.07 *	0.0251	1.68	0.0263	4.93 ***	6241	3.22 ***	0.01446	3.40 ***	0.01606	9.50 ***
2004	5988	0.86	0.0236	0.72	0.0179	2.43 **	6069	2.50 **	0.01656	2.43 **	0.01726	7.39 ***
2005	5983	0.96	0.0153	1.23	0.0169	1.90	5831	2.75 **	0.01609	1.92	0.01132	8.23 ***
2006	5813	0.67	0.0177	0.61	0.0174	1.74	5696	1.73	0.01078	1.51	0.00996	5.12 ***
2007	5805	1.09	0.0164	0.92	0.0182	3.13 ***	5509	3.38 ***	0.01450	2.03 *	0.01046	9.10 ***
2008	5643	0.46	0.0202	0.71	0.0224	1.14	5469	2.82 ***	0.01505	2.73 **	0.01282	6.80 ***
2009	5033	0.61	0.0357	0.71	0.0272	1.42	5419	1.59	0.01428	1.20	0.01293	4.74 ***
2010	4599	0.73	0.0214	0.62	0.0193	2.90 ***	4681	2.28 **	0.01778	2.65 **	0.02215	6.57 ***

Table 4. Earnings forecast errors

This table shows results for discontinuity tests on earnings forecast errors. Forecast errors are calculated using unadjusted I/B/E/S median estimates and actual values that are taken from the I/B/E/S adjusted and unadjusted summary file (corresponding to main columns in this table) and corrected for stock splits by matching the appropriate I/B/E/S adjustment factor(s). The third main column adds a random uniform error in the interval (USD –0.005, USD 0.005) to forecast errors before scaling by market price in order to incorporate uncertainty about the true underlying per-share value for earnings estimates and reported earnings. Bandwidths used for kernel density estimation are obtained from the bootstrap KDE method and Silverman's (1986) rule of thumb, while $h=0.0002$ for Burgstahler and Dichev's (1997) test statistic as in Burgstahler and Eames (2006). Row entries show test statistics for the intervals to the left and right of zero. Bonferroni-adjusted significance levels: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Period	Interval	N	Adjusted actuals Bootstrap Std. Diff.	Rule of thumb Std. Diff.	BD Std. Diff.	N	Unadjusted actuals Bootstrap Std. Diff.	Rule of thumb Std. Diff.	BD Std. Diff.	N	Unadjusted actuals, random error Bootstrap Std. Diff.	Rule of thumb Std. Diff.	BD Std. Diff.
Median forecast in the reporting month													
1986–2000		10845	h=0.000121	h=0.000791	h=0.0002	10605	h=0.000333	h=0.000823	h=0.0002	11681	h=0.000384	h=0.000634	h=0.0002
	(-h,0]		-0.02	-3.31 ***	-0.59		-1.85	-3.56 ***	-6.50 ***		0.92	-0.38	3.24 ***
	(0,h]		-1.57	5.42 ***	-2.78 **		-1.17	6.17 ***	-5.63 ***		0.53	2.08 *	1.07
1986–2010		19998	h=0.000047	h=0.000596	h=0.0002	19672	h=0.000231	h=0.000614	h=0.0002	21837	h=0.000261	h=0.000471	h=0.0002
	(-h,0]		1.25	-3.37 ***	-4.43 ***		-3.46 ***	-4.20 ***	-11.31 ***		1.35	0.41	3.25 ***
	(0,h]		-0.75	4.52 ***	-5.94 ***		-2.79 **	4.89 ***	-8.78 ***		0.10	2.37 **	2.49 **
Median forecast 1 month before earnings announcement													
1986–2000		23260	h=0.000040	h=0.000744	h=0.0002	22774	h=0.000292	h=0.000780	h=0.0002	24966	h=0.000339	h=0.000626	h=0.0002
	(-h,0]		0.61	-2.33 **	-0.43		-2.08 *	-2.30 **	-8.69 ***		0.56	-0.62	1.94
	(0,h]		0.02	3.66 ***	-3.95 ***		-2.49 **	4.56 ***	-8.81 ***		1.13	3.99 ***	3.82 ***
1986–2010		42540	h=0.000041	h=0.000554	h=0.0002	41933	h=0.000206	h=0.000573	h=0.0002	46175	h=0.000278	h=0.000452	h=0.0002
	(-h,0]		0.71	-2.50 **	-5.29 ***		-3.87 ***	-2.91 ***	-14.06 ***		0.99	0.07	2.97 ***
	(0,h]		-0.05	2.53 **	-9.10 ***		-4.41 ***	3.47 ***	-13.84 ***		1.55	3.12 ***	4.48 ***
Median forecast 3 months before earnings announcement													
1986--2000		23780	h=0.000106	h=0.001068	h=0.0002	23491	h=0.000360	h=0.001106	h=0.0002	25048	h=0.000404	h=0.000958	h=0.0002
	(-h,0]		0.76	-1.10	-0.63		-2.84 ***	-1.14	-7.68 ***		0.12	-0.61	1.32
	(0,h]		-1.37	5.24 ***	-4.11 ***		-1.78	5.61 ***	-6.98 ***		0.73	4.84 ***	2.43 **
1986–2010		43355	h=0.000112	h=0.000761	h=0.0002	42991	h=0.000237	h=0.000781	h=0.0002	46137	h=0.000333	h=0.000666	h=0.0002
	(-h,0]		-0.65	-2.44 **	-5.45 ***		-3.26 ***	-2.44 **	-12.42 ***		1.03	-1.22	2.29 **
	(0,h]		-1.82	4.29 ***	-8.46 ***		-3.76 ***	4.74 ***	-11.94 ***		0.94	4.77 ***	3.15 ***
Median forecast 6 months before earnings announcement													
1986–2000		24297	h=0.000838	h=0.001949	h=0.0002	24156	h=0.000574	h=0.001986	h=0.0002	24926	h=0.001038	h=0.001863	h=0.0002
	(-h,0]		-0.20	-0.70	-0.42		-0.90	-0.96	-5.56 ***		-0.62	-0.93	-0.43
	(0,h]		1.69	3.75 ***	-4.27 ***		-0.18	4.19 ***	-5.67 ***		3.33 ***	3.90 ***	1.59
1986–2010		44421	h=0.000461	h=0.001422	h=0.0002	44243	h=0.000362	h=0.001439	h=0.0002	45775	h=0.000789	h=0.001345	h=0.0002
	(-h,0]		-0.54	-0.93	-4.35 ***		-2.06 *	-1.11	-9.45 ***		-0.71	-0.57	-0.08
	(0,h]		-0.20	4.47 ***	-7.72 ***		-0.83	4.70 ***	-9.41 ***		2.43 **	4.13 ***	1.43
Median forecast 9 months before earnings announcement													
1986–2000		24164	h=0.001364	h=0.002794	h=0.0002	24066	h=0.000894	h=0.002822	h=0.0002	24585	h=0.001567	h=0.002717	h=0.0002
	(-h,0]		-0.62	-0.96	-1.33		-0.66	-0.91	-4.41 ***		-0.22	-1.10	0.13
	(0,h]		2.05 *	4.60 ***	-2.09 *		1.40	4.54 ***	-3.73 ***		2.47 **	4.12 ***	2.17 *
1986–2010		44277	h=0.000729	h=0.002104	h=0.0002	44148	h=0.000567	h=0.002120	h=0.0002	45115	h=0.001174	h=0.002030	h=0.0002
	(-h,0]		-1.15	-0.93	-3.32 ***		-1.18	-0.58	-6.32 ***		-0.33	-0.75	0.21
	(0,h]		1.24	4.48 ***	-5.65 ***		-0.03	4.64 ***	-7.69 ***		2.36 **	4.26 ***	2.26 **