# François Degeorge

*Hautes Études Commerciales and the Centre for Economic Policy Research*

# Jayendu Patel

*Boston University*

# Richard Zeckhauser

*Harvard University and National Bureau of Economic Research*

# Earnings Management to Exceed Thresholds*

## I. Introduction

Analysts, investors, senior executives, and boards of directors consider earnings the single most important item in the financial reports issued by publicly held firms. In the medium to long term (1–10-year intervals), returns to equities appear to be explained overwhelmingly by the firm's cumulative earnings during the period; other plausible explanations—such as dividends, cash flows,

Earnings provide important information for investment decisions. Thus executives—who are monitored by investors, directors, customers, and suppliers—acting in self-interest and at times for shareholders, have strong incentives to manage earnings. We introduce behavioral thresholds for earnings management. A model shows how thresholds induce specific types of earnings management. Empirical explorations identify earnings management to exceed each of three thresholds: report positive profits, sustain recent performance, and meet analysts' expectations. The positive profits threshold proves predominant. The future performance of firms suspect for boosting earnings just across a threshold is poorer than that of control group firms.

or capital investments—have marginal correlations close to zero (Easton, Harris, and Ohlson 1992; Kothari and Sloan 1992). Even for short-term equity returns, earnings are an important explanatory factor.[1]

The rewards to a firm's senior executives—both employment decisions and compensation benefits—depend both implicitly and explicitly on the earnings achieved on their watch (Healy 1985). But such executives have considerable discretion in determining the figure printed in the earnings report for any particular period. Within generally accepted accounting principles (GAAP), executives have considerable flexibility in the choice of inventory methods, allowance for bad debt, expensing of research and development, recognition of sales not yet shipped, estimation of pension liabilities, capitalization of leases and marketing expenses, delay in maintenance expenditures, and so on. Moreover, they can defer expenses or boost revenues, say, by cutting prices. Thus, executives have both the incentive and ability to manage earnings. It is hardly surprising that the popular press frequently describes companies as engaged in earnings management—sometimes referred to as manipulation.[2]

This article studies earnings management as a response to implicit and explicit rewards for attaining specific levels of earnings, such as positive earnings, an improvement over last year, or the market's consensus forecast. We label as "earnings management" (EM) the strategic exercise of managerial discretion in influencing the earnings figure reported to external audiences (see Schipper 1989). It is accomplished principally by timing reported or actual economic events to shift income between periods.

We sketch a model that predicts how executives strategically influence the earnings figures that their firms report to external audiences and then examine historical data to confirm such patterns. Our model incorporates behavioral propensities and a stylized description of the interactions among executives, investors, directors, and earnings analysts to identify EM patterns that generate specific discontinuities and distortions in the distribution of observed earnings.[3]

We do not determine which components of earnings or of supplementary disclosures are adjusted. Nor do we attempt to distinguish empirically between "direct" EM—the strategic timing of investment,

---

1. Ball and Brown (1968) is the classic early work; see Dechow (1994) and references there for subsequent research that details the relevance of earnings.
2. See, e.g., the multipage stories "Excuses Aplenty When Companies Tinker with Their Profits Reports," *New York Times* (June 23, 1996), and "On the Books, More Facts and Less Fiction," *New York Times* (February 16, 1997). A recent study—Bruns and Merchant (1996, p. 25)—concludes that "we have no doubt that short-term earnings are being manipulated in many, if not all, companies."
3. DeBondt and Thaler (1995, pp. 385–410) provide a discussion of behaviorally motivated financial decisions by firms.

sales, expenditures, or financing decisions—and "misreporting"—EM involving merely the discretionary accounting of decisions and outcomes already realized.[4]

We identify three thresholds that help drive EM: the first is to report profits—for example, 1 penny a share. This threshold arises from the psychologically important distinction between positive numbers and negative numbers (or zero). The second and third benchmarks rely on performance relative to widely reported firm-specific values. If the firm does as well or better than the benchmark, it is met; otherwise it is failed. The two benchmarks are performance relative to the prior comparable period and relative to analysts' earnings projections. Performance relative to each benchmark is assessed by examining the sprinkling of quarterly earnings reports in its neighborhood. A big jump in density at the benchmark demonstrates its importance.

Burgstahler and Dichev (1997) examine the management of earnings to meet our first two thresholds, though not in relation to analysts' estimates.[5] Their analysis delves more deeply into accounting issues and identifies the "misreporting" mechanisms—for example, the manipulation of cash flow from operations, or changes in working capital—that permit earnings to be moved from negative to positive ranges. We devote considerably more attention to the motivations for EM, consider direct EM (e.g., lowering prices to boost sales) in addition to misreporting, provide an optimizing model on how earnings are managed, and analyze the consequences of management for future earnings. In addition, we explore EM as the executive's (agent's) response to steep rewards—reaping a bonus or retaining a job—that depend on meeting a bright threshold.[6] Finally, we look at the hierarchy among our thresholds.

Earnings management arises from the game of information disclosure that executives and outsiders must play. Investors base their decisions on information received from analysts—usually indirectly, say, through a broker—and through published earnings announcements. To bolster investor interest, executives manage earnings, despite the real earnings sacrifice. Other parties, such as boards of directors, analysts, and accountants, participate in this game as well, but their choices are exogenous to our analysis. For example, the contingent remuneration actions of boards are known to executives. Presumably such pay packages are structured to take distorting possibilities into account and may

---

4. Foster (1986, p. 224) discusses mechanisms for misreporting transactions or events in financial statements.

5. Payne and Robb (1997) show that managers use discretionary accrual to align earnings with analysts' expectations.

6. Burgstahler (1997) adds a model in which earnings are manipulated because the marginal benefit of reporting higher earnings is greatest in some middle range.

have been adjusted somewhat to counter EM.[7] If so, finding evidence of management is more significant.

Executives may also distort earnings reports in a self-serving manner, imposing an agency loss that reduces the firm's value if their incentives are not fully aligned with those of shareholders. Full alignment is unlikely. First, while the value of the stock is the present value of dividends stretching to infinity, the executive's time horizon is relatively short. Since it is difficult for boards, shareholders, or the stock market to assess future prospects, executives have an incentive to pump up current earnings at the expense of the hard-to-perceive future beyond their reign. Accordingly, a major benefit of stock options is that they extend the time horizon for executives.

Second, an executive's compensation, including the probability of keeping his job, is likely linked to earnings, stock price performance, or both. (See Healy 1985; Gaver, Gaver, and Austin 1995.) If accepting lower earnings today might result in a termination or a lost bonus, substantially greater earnings tomorrow may not represent a desirable trade-off. When earnings are near the unacceptable range, executives' incentives to manage them upward will be significant. However, when bonuses are near maximum, further earnings increases will be rewarded little, generating an incentive to rein in today's earnings—that is, shift them forward—making future thresholds easier to meet.[8] Executives may also be reluctant to report large gains in earnings because they know their performance target will be ratcheted up in the future. Earnings so poor as to put thresholds and bonuses out of reach may also be shifted to the future; the executive saves for a better tomorrow.

Earnings can be managed by actually shifting income over time, which we label "direct management," or by misreporting. A typical misreport, failing to mark down "stale" inventory or incurring extraordinary charges beyond what prudence requires, simply relocates an amount from 1 year to another. Such misreports must pass through the hands of accountants, who are reliable professionals. Accountants' procedures prevent simple misreporting of earnings; indeed, only their oversight makes earnings reports meaningful. But accountants are neither omniscient nor disinterested. They can be misled, but only at a cost. The executive may need to co-opt the auditor, say, with an unneeded consulting contract. Alternatively, he may make his misreporting hard

7. Dechow, Hudson, and Sloan (1994) document that compensation committees often override the provisions of incentive plans to avoid providing incentives for executives to behave opportunistically.

8. Healy (1985, p. 106) reports that "managers are more likely to choose income-decreasing accruals when their bonus plan's upper and lower bounds are binding, and income-increasing accruals when these bounds are not binding." Holthausen, Larcker, and Sloan (1995) find that managers manipulate earnings downward when they are at the upper bounds of their bonus contracts. However, they find no manipulation downward below their contract's lower bounds.

to detect, but that requires weakening internal control mechanisms, which help the manager in allocating resources or detecting shirking or misappropriation at lower levels in the firm.

Direct management of earnings upward—through delaying desirable training or maintenance expenditures or cutting prices to boost sales—has real consequences and can impose costs beyond today's benefits plus imputed interest. Earnings delays, perhaps, by accelerating costs to this year to pave the way for a brighter future—common behavior when a new team takes over and blames poor initial results on past leadership—are also costly. Both misreporting and direct earnings management, whether pushing earnings forward or back, are costly activities. Their marginal cost increases with scale since cheap transfers are undertaken first.

Section II of this article reports briefly on relevant literature from psychology, develops a model of EM around threshold targets, and draws inferences for real world data. Section III reports on empirical explorations relating to thresholds, studying conditional and unconditional distributions of quarterly earnings over the period 1974–96. Section IV examines whether firms that are more likely to have managed earnings upward to attain a threshold in a particular year underperform in the subsequent year. Section V suggests future directions and concludes.

## II. A Threshold Model of Earnings Management

Executives manage earnings to influence the perceptions of outsiders—such as investors, banks, and suppliers—and to reap private payoffs.[9] In our stylized model, outsiders utilize thresholds as a standard for judging and rewarding executives. When executives respond to these thresholds, distributions of reported earnings get distorted: far too few earnings fall just below a threshold, too many just above it.

### A. *Why Thresholds?*

Executives focus on thresholds for earnings because the parties concerned with the firm's performance do. Executives may also manipulate earnings for their own reasons if, for example, they derive personal satisfaction from making a target; however, the biases of outsiders are our focus.

---

9. Even if EM is costly, it may be in the interest of shareholders ex ante if it increases the information available to important parties. In some settings, manipulated earnings may contain more, not less, information about the firm's true prospects. For example, if a firm's earnings barely meet some threshold, it is likely that the figure has been inflated. But this implies that executives are confident that the cost of manipulation—reduced profits next year—will not be so large as to reduce dramatically the prospect that the firm will meet the threshold next year. Thus, small manipulated profits may contain more information than small unmanipulated profits.

Beyond boards, investors, and analysts, earnings reports are important to those people concerned with the firm's viability and profitability because they make firm-specific investments, such as customers and suppliers, bankers, and workers. Many of these outsiders exhibit what we call a "threshold mentality," for both rational and perceptual reasons. In a range of circumstances, individuals perceive continuous data in discrete form; indeed "the tendency to divide the world into categories is a pervasive aspect of human thought" (Glass and Holyoak 1986, p. 149). For example, we perceive the continuous color spectrum discretely, recognizing seven primary colors. Similarly, if a diagram shades from dark to light and then remains light, humans perceive a bright line where the shading to light stops (Cornsweet 1974, pp. 276–77). Below we discuss three established demarcations for corporate earnings. Unlike our vision examples, earnings demarcations draw strongly on external cues.

The salience of thresholds arises from at least three psychological effects. First, there is something fundamental about positive and nonpositive numbers in human thought processes.[10] Hence, this dividing line carries over for the threshold on absolute earnings. When looking at the benchmarks of quarterly earnings a year back and the analysts' consensus forecast, there is a salient dividing line between meeting and failing to meet the norm. Meeting the norm is critical, as opposed to beating it by 10% or falling short by 3%. Saliency makes the norm itself a focal point, which reinforces its psychological properties.[11]

Second, as prospect theory tells us, individuals choosing among risky alternatives behave as if they evaluate outcomes as changes from a reference point (Kahneman and Tversky 1979). The reference point is usually some aspect of the decision maker's current state (e.g., wealth), and it shifts over time, sometimes with how the decision is framed. The amount of shifting can dramatically affect choices for two reasons: there is a kink in the utility function at the reference point (zero change); and the overall curve is S-shaped (i.e., it is convex for losses and concave for gains). If the preferences of executives, the boards that review them, or the investors who trade the firm's stock are consistent with the predictions of prospect theory, then executives will have a

----

10. The symbol for zero came late and with difficulty to mathematicians, except in India. For example, China imported it from India in the eighth century, and "the mathematicians and astronomers of Sumer and Babylon labored for nearly 1500 years before they introduced the notion of a 'zero' symbol." Negative numbers were much harder still, not becoming "generally recognized as 'numbers' until the sixteenth century" (Barrow 1992, pp. 89–90). In contrast, positive numbers appear to be a more directly grasped concept for humans.

11. Any assessor of earnings will worry about the consistency of his judgment with that of others, which makes focal points critical. When comparing performance to a yardstick, just meeting the standard is a spotlighted property. For a seminal analysis of focal points see Schelling (1960). See Young (1996) for a discussion of conventions.

threshold-related reward schedule and are likely to manage reported earnings in response. The thresholds they will wish to reach are the reference points in the value functions of the participants; such points are likely to be perceptually salient.

Third, thresholds come to the fore because people depend on rules of thumb to reduce transactions costs. The discreteness of actions, whether by investment analysts recommending sell, hold, or buy, rating agencies giving letter grades, bankers making or refusing loans, or boards retaining or dismissing the CEO promotes the use of thresholds of acceptable performance.[12] Banks, for example, may grant loans only to firms that report positive earnings; that is, banks use a threshold of zero earnings as an initial screen since judiciously adjusting interest rates in response to differential performance may be too hard. Earnings management across thresholds can also simplify executives' relations with shareholders and boards of directors. A report to shareholders that earnings have been up 6 years in a row is cheaply communicated. A statement that they have been up 5 out of 6 years, and only fell by 1% in the off year, is less easily understood, so that struggling across the threshold of last year's earnings becomes worthwhile. When a firm falls short of analysts' earnings projections, the board may think that the executives did a poor job; bonuses and stock option awards may suffer. Such doubts are much less likely to arise if the analysts' earnings are just met.[13]

Threshold effects may be important even if few participants respond to them directly. Suppose that only the firm's bankers care directly whether the firm reaches a specific performance threshold but all parties know how the bankers feel. Since analysts and shareholders know that executives cannot lightly risk raising the bankers' ire, they will want to know whether the firm meets the banker's performance threshold. Thus, reaching the threshold caters both to the bankers and to other participants' rational perceptions through inference.

In contrast to a world in which all participants care about thresholds, threshold-regarding (TR) behavior by merely a minority may have a much more than proportional effect. For example, the level of EM in a world where 25% of boards of directors respond naturally to thresholds may be much more than 25% as great as in a world where all boards are threshold-driven. Consider an executive threatened with modestly negative results who does not know how his board will re-

12. Burgstahler (1997) shows empirically that the net probability of improvements in outside ratings of both debt and equity are greatest in the neighborhoods of zero earnings and zero changes in earnings.
13. President Clinton, recognizing the role of thresholds, announced that he was seeking to secure 50% of the 1996 presidential vote so as to claim a mandate. Not surprisingly, he struggled hard in the final days to get more than 50%. (In fact, he won 49.2% of the actual vote.)

spond. If it is TR, it will fire him with probability 0.4; otherwise, his job is safe. If he knew he faced a TR board, he would manage earnings to the positive safe zone. But even with only a 25% chance it is TR, he may do the same thing. A 10% probability of being fired may be sufficient stimulus. Signaling and lemons-type unraveling can also lead to spillovers, for instance, if higher-quality firms are more able and likely to manipulate to the safe range. If so, TR behavior by a modest proportion of boards spills over to affect potentially the behavior of large numbers of executives.

### B.   Three Thresholds

Reports in the financial press suggest that executives care about three thresholds when they report earnings:

1) to report positive profits, that is, report earnings that are above zero;
2) to sustain recent performance, that is, make at least last year's earnings; and
3) to meet analysts' expectations, particularly the analysts' consensus earnings forecast.

The analysts' consensus estimate, unlike our other thresholds, is endogenous. Although executives try to report earnings that exceed analysts' forecasts, analysts try to anticipate reported earnings.[14] A complicated game ensues, in which analysts predict an earnings number that will then be manipulated in response to their prediction. Anecdotal evidence suggests that executives, realizing the importance of meeting or exceeding the analysts' consensus, actively try to influence analysts' expectations downward, especially when the earnings announcement date draws near.[15]

### C.   A Two-Period Model with Last Period's Earnings as Threshold

Earnings management to reach thresholds affects the distribution of reported earnings. We study a simple 2-period model where the threshold to be met is last year's earnings. In each period $t = 1, 2$, the firm gets a random, independent, and identically distributed draw of "latent" or true earnings, $L_t$. Outsiders cannot observe these latent earnings. They only see reported earnings, $R_t$. In period 1, executives can manipulate

14. See Abarbanell and Bernard (1992) and references there on possible biases in analysts' forecasts.
15. See "Learn to Manage Your Earnings, and Wall Street Will Love You," *Fortune* (March 31, 1997). This article tells the story of a meeting of Microsoft's Bill Gates, his chief financial officer, and financial analysts, during which the Microsoft executives paint a particularly bleak picture of the company's future. At the end of the meeting, Gates and his chief financial officer congratulate each other when they realize that their goal of depressing analysts' expectations has been achieved.
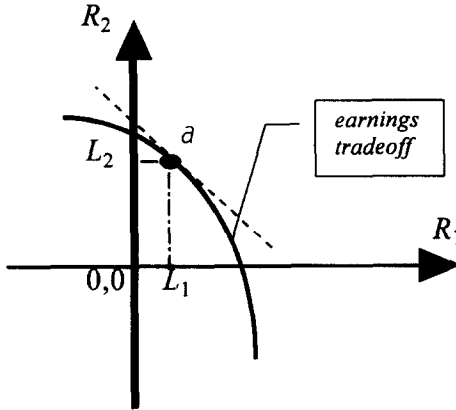
Fig. 1.—Reported earnings in period 2 as a function of those in period 1, holding $L_1$ and $L_2$ fixed.

reported earnings $R_1$ by choosing an amount $M_1$ (possibly negative) to add to earnings, such that $R_1 = L_1 + M_1$. The cost of manipulation is paid when there is full settling up in period 2:

$$R_2 = L_2 - k(M_1),$$

where $k(0) = 0$ and there are positive and increasing marginal costs for moving away from zero. For simplicity, assume a zero discount rate. Pumping up reported earnings today reduce earnings tomorrow by more than 1 dollar. If period 1 manipulation is negative (executives rein in earnings), another dollar reduction boosts next year's earnings by less than 1 dollar.

The executive exits after period 2, and we assume that all is revealed at that point. This produces the trade-off indicated in figure 1. Point $a$ corresponds to $M_1 = 0$, and thus $R_1 = L_1$. As shown, the slope of the trade-off curve at $a$ is $-1$. (More generally, the slope will be $-(1 + r)$, where $r$ is the 1-period interest rate corresponding to a nonzero time value of money.)

We assume that the executive's expected reward schedule falls sharply at one or more thresholds, such as negative earnings, or earnings below last year. Below such thresholds, he or she might risk termination or at least a substantial cut in bonus. For simplicity, we assume that at all earnings levels other than at the thresholds the incentives for better performance are positive and constant. (In practice, we suspect that they are steep near a threshold and more tempered at either extreme.)

The self-interested executive manages earnings to maximize his personal payoff. In each period, he receives a payoff $f(R_t, R_{t-1})$, where $R_t$ is the reported earnings performance in period $t$. If the manager

meets or surpasses the benchmark, he receives a bonus $v(R_t, R_{t-1})$.[16] Thus, we posit the following form for $f$:

$$f(R_t, R_{t-1}) = \beta R_t + v(R_t, R_{t-1}),$$

where

$$v(R_t, R_{t-1}) = \gamma \text{ if } R_t \geq R_{t-1}$$

$$= 0 \text{ otherwise.}$$

The executive's direct rewards for the current period performance (at a rate $\beta$) are captured by the first term. The previous period's reported earnings serve as a benchmark for a second effect. For period 1, the benchmark is $R_0$, which is normalized to zero for exposition, and the benchmark is $R_1$ for period 2. Thus the $v(R_t, R_{t-1})$ term induces a ratcheting effect.[17]

Executives are assumed to be risk neutral for convenience; this assumption could easily be relaxed. For our 2-period illustration, the executive selects $M_1$ to maximize the net present value of the expected payoffs in the two periods, that is, to maximize

$$f(R_1, R_0) + \delta E[f(R_2, R_1)],$$

where $E$ denotes expectation and $\delta$ is the discount factor.

Managing earnings is an imprecise science, relying on estimates of both latent earnings and the effects of any attempts to boost earnings. Latent earnings may well prove higher or lower than expected. We analyze two cases, depending on whether the executive knows $L_1$ precisely or imprecisely when he selects $M_1$.

CASE 1. The executive knows $L_1$ precisely when he selects $M_1$. In this setting, the primary element of the executive's strategy is intuitively clear. If $L_1 < R_0$, the executive should select $M_1$ to achieve the threshold and reap the bonus, unless the entailed loss on $L_2$ in expected value terms proves too costly.

We set $R_0$ equal to zero for convenience. If $L_1$ is slightly below zero, then it will be worthwhile to select a positive $M_1$—the executive should borrow future earnings to make the bonus. While manipulation will sacrifice a greater amount of second-period earnings and raise the hurdle for the second period, it will allow the executive to earn the bonus

16. If effort boosts earnings, incentives should be strongest where earnings outcomes, given optimal effort, are most likely. Strong incentives cannot be provided across all outcomes because executives cannot be paid negatively for poor outcomes. Thus, strong incentives will not be provided for very favorable outcomes either lest executives be overpaid on average. The optimal reward schedule will be steeply responsive near the benchmark since such earnings outcome are most likely.

17. Ratcheting of standards is well known in the contexts of worker productivity, procurement, and regulation and is primarily studied for its disincentive effects on first-period effort. See Milgrom and Roberts (1992, pp. 233–36) and Laffont and Tirole (1993, pp. 381–87).

for sure now, only sacrificing it with some chance in period 2. The borrowing will prove well worthwhile, except in the unlikely case when it turns out to sacrifice next period's bonus.

If $L_1$ is significantly below zero, then borrowing to cross the threshold may be too costly. To determine whether it is, the executive compares two quantities. The first is his expected payoff if he manipulates just enough—that is, selects $M_1$ so that $R_1 = 0$—to secure the bonus. The second is his optimal strategy if he forgoes the bonus. For the second, he actually selects a negative value of $M_1$, lowering the next period's threshold and pushing earnings forward thus in two ways. We call this "saving for a better tomorrow." Reducing earnings when latent earnings are disappointing is referred to in the literature as "taking the Big Bath."

If $L_1$ is above $R_0$, then there is no reason to boost earnings. Indeed, for $L_1 > R_0$, some reining in is desirable since it increases the likelihood that the executive will earn the next period's bonus.

To illustrate, we choose $R_0 = 0$, $\beta = 1$, $\gamma = 10$, and $\delta = 1$; hence, $f(R_1) = R_1 + v(R_1, 0)$. Earnings have a normal distribution each period, with mean of zero and standard deviation of 10.[18] The second-period cost from manipulation is $k(M) = e^M - 1$, which is greater than $M$, implying that any manipulation is costly on net. (If $M < 0$, earnings are manipulated downward in the first period and boosted in the second period—but the second period boost is smaller than the first period hit.)

Figure 2 illustrates the executive's optimal strategy as a function of latent earnings $L_1$. The initial threshold is achieved where $L_1 + M_1 = R_0 = 0$. Our key finding is that, just below zero, the optimal strategy is to set $M_1 = -L_1$; future earnings are borrowed to meet today's earnings threshold. At point $Z$, the payoff from choosing $M_1 = -L_1$ (and therefore a positive $M_1$, indicating borrowing) just equals the payoff from saving for a better tomorrow (taking the optimal sacrifice in earnings). Left of $Z$, the optimal bath gives a higher payoff than striving. Right of $Z$, borrowing gives a higher 2-period payoff. Hence the discontinuity in the graph.

When $L_1$ is small and positive, it pays to rein in, so that reported earnings just sneak beyond the threshold (recall that in this initial version of the model, $R_1$ can be targeted perfectly, so there is no risk of missing zero earnings). As $L_1$ becomes larger, reining in becomes less

---

18. We assume stationarity in the latent earnings distribution. This might be considered unrepresentative if the real earnings process has a random walk characteristic. If latent earnings do follow a random walk and we keep the same ratcheting structure in the payoff function, the manipulation behavior will be identical close to the threshold ($M_1 = L_1$). Away from the threshold, firms will manipulate by a constant amount regardless of $L_1$: ratcheting combined with the random walk assumption ensures that the executive's decision problem is invariant to $L_1$.
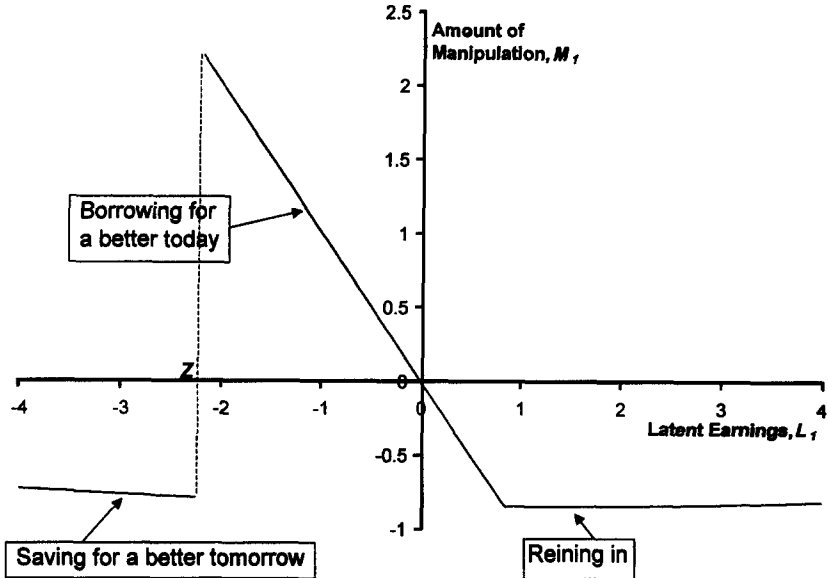
FIG. 2.—Optimal amount of period 1 manipulation, $M_1$, as a function of latent period 1 earnings $L_1$. Latent earnings $L_1$ are normally distributed with mean zero and standard deviation 10. If reported earnings $R_1 = L_1 + M_1$ reach at least $R_0 = 0$, the executive reaps a bonus of 10. The period 2 cost of manipulation is $k(M_1) = e^M - 1$. The executive knows $L_1$ exactly when choosing the manipulation level $M_1$.

attractive since ratcheting upward makes the benchmark less likely to be attained in period 2 and the $k$ function is convex. Indeed, for large values of $L_1$ (not shown), reining in is abandoned since next year's bonus is unlikely to be reaped in any circumstance.[19]

Figure 2 identifies three phenomena that arise if executives misreport earnings. First, for a range of values of $L_1$, a profit just sufficient to meet the threshold is recorded. Second, EM creates a gap in the earnings distribution just below the threshold (zero in this case). Third, the level of reported earnings will be a sharply discontinuous function of latent earnings.[20]

19. Healy (1985, p. 90), who focuses on misreporting (discretionary accruals) and does not consider ratcheting, provides intuition for a three-component linear schedule. Whereas our schedule has a jump at the threshold, with a shallow positive slope to either side, Healy assumes a schedule that has a slope over a middle range, with a zero slope to either side. Unlike Healy, we assume improved performance is rewarded everywhere and that there is a sharp reward at the threshold.

20. This will make reported earnings very difficult to predict. Thus, executives' manipulations could explain why analysts' forecasts are often wrong. Roughly 45% of analysts' estimates fall outside a band of 15% plus or minus the actual earnings (Dreman and Berry 1995a, p. 39).

CASE 2. The executive has an imprecise estimate of $L_1$ when he chooses $M_1$. The executive has a prior probability distribution on $L_1$ centered on its true value with variance $\sigma^2$. Now, when the executive sets an $M_1 > 0$ seeking to meet or exceed the threshold, he must choose a value higher than in case 1 to be sure the threshold is met. Also, when by chance $L_1$ ends up toward the bottom of its expected range, small negative earnings will be recorded.

Case 2 incorporates uncertainty, sets $\sigma^2 = 1$, and uses the same parameter values employed in our prior example. Figure 3 shows the distribution of reported earnings for 20,000 draws of latent earnings with a bin width of 1 unit. The density of reported earnings dips just below zero and piles up above the threshold. The extreme outcome of zero density just below zero, which occurs when the executive has perfect knowledge of $L_1$, does not appear in case 2. The maximum hump is shifted to the right of zero because executives hedge against uncertainty by undertaking some positive EM even when the mean of their prior distribution of $L_1$ is somewhat above the threshold. This simulated distribution of $R_1$ is the pattern to which we compare the empirical distributions below. The dark shaded areas below the horizontal show shortfalls in the density, if any, relative to the equidistant bin on the other side of the threshold (which, by construction, is at the peak of the latent symmetric distribution).

In results not shown, we explored the consequences of changing the parameter values of the model. We find that the changeover point $Z$ (where $L_1$ is negative and the payoff from borrowing to meet the threshold just equals the payoff from taking the optimal sacrifice) moves to the left as the discount factor increases since, the higher the executive's discount rate, the more valuable it is to get high earnings this period, and the costlier it is to take a bath.

The more uncertain are second-period earnings, the more the executive will manipulate to secure the bonus in the first period since big borrowings are less likely to sacrifice the second year's threshold. As the bonus for crossing the threshold (indicated by $\gamma$) falls in importance relative to the rewards per unit of reported earnings (indicated by $\beta$), EM becomes more costly and decreases. For any level of $L_1$, as $\beta$ increases, the optimal $M_1$ moves closer to zero, and the dip below and pile-up above the threshold both diminish. Where borrowing earnings had been most extreme, the executive saves instead for a better tomorrow.[21]

---

21. We have extended results to a three-period setting in results not shown. With more than two periods, there are factors that make saving earnings from the first period both more and less valuable. They would be more valuable because there would be no danger that they would be "wasted," i.e., more than enough to secure the second-period bonus. They would be less valuable because executives could always borrow in the second period to make that period's bonus.
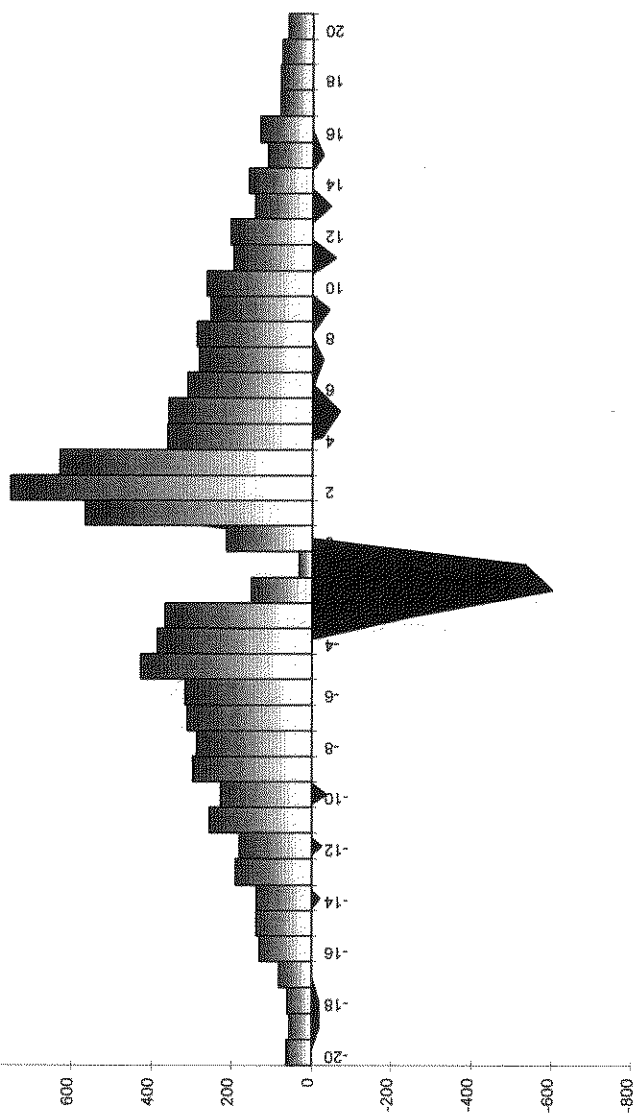
FIG. 3.—Simulated distribution of reported earnings $R_1$. Latent earnings $L_1$ are normally distributed with mean 0 and standard deviation 10. If reported earnings $R_1 = L_1 + M_1$ reach at least $R_0 = 0$, the executive reaps a bonus of 10. The period 2 cost of manipulation is $k(M_1) = e^M - 1$. The executive knows $L_1$, imprecisely when choosing the manipulation level $M_1$ (he has a probability distribution centered on $L_1$ with a variance of one). The dark shaded areas below the horizontal show shortfalls relative to the equidistant bin on the other side of the threshold of zero.

The next two sections relate the results of the model to empirical data on earnings so as to evaluate the evidence of EM. Section III examines the distributions (both unconditional and conditional) of reported earnings over 1974–96. Section IV reports on statistical tests of the hypothesis that future earnings are lower when current earnings are likely to have been manipulated upward to attain a threshold.

## III. Evidence of Earnings Management to Exceed Thresholds

Theory suggests that simple thresholds will significantly influence executives' management of earnings. It is impossible to monitor manipulation, $M$, and test the theory directly, so we evaluate the indirect evidence provided by the values of reported earnings, $R$.[22]

Our empirical analyses explore the extent to which managers manage earnings to attain our three thresholds: (1) to "report profits," that is, to achieve 1 cent or more in earnings per share; (2) to "sustain recent performance," that is, to meet or surpass the most recent level of comparable earnings (which, given seasonal variation, is the corresponding quarter from the previous year); and (3) to "meet analysts' expectations," that is, to meet or exceed the consensus forecast of analysts. We study the density function for earnings near each threshold. If managers do indeed manage earnings to meet a threshold, we expect to observe "too few" earnings reports directly below it and "too many" at or directly above it. We do not expect findings as stark as those our model generates because of numerous additional factors, including heterogeneity among firms in both earnings distributions and EM potential.

Subsection IIIA briefly discusses the sample and the construction of variables. Subsection IIIB presents three univariate histograms that provide evidence of EM across the three thresholds. For each histogram, we report the results of a statistical test that the discontinuity at the conjectured threshold is significant. Details of the test method are discussed in the appendix. Finally, in Subsection IIIC, we explore conditional distributions to rank the importance of the three thresholds.

### A.  Data and Construction of Variables

Our data set consists of quarterly data on 5,387 firms providing partial or complete data over the 1974–96 period. To conveniently align quarterly observations, we drop firms whose fiscal years do not end in March, June, September, or December. While the total number of observations exceeds 100,000, the number of available observations is much smaller for many of the analyses. For the 1974–84 period, the

---

22. Dechow et al. (1995) address the problems of estimating the level of discretionary accrual activity.

sample includes only the midcapitalization or larger firms for which Abel-Noser (more recently Q-Prime) provides data on analysts' forecasts of earnings. The reported earnings per share are from Compustat item no. 8, which excludes extraordinary items. Our post-1984 sample of more than 83,000 observations draws from the databases provided by I/B/E/S International Inc. The I/B/E/S databases contain analysts' forecasts of quarterly earnings as well as the reported earnings.

We represent analysts' expectations by the mean of the analysts' forecasts for the contemporaneous quarter. Such forecasts are usually available around the middle of the last month of the quarter. (The financial results for a quarter are announced by firms about 4 weeks into the next quarter—typically slightly later for the fiscal-year-ending quarter and somewhat earlier for the other three quarters.) According to I/B/E/S, analysts' earnings forecasts do not include unusual or nonrecurring charges, and so the reported earnings per share (EPS) variable we use excludes extraordinary items. Thus, any evidence of EM we uncover excludes earnings-shifting strategies employing extraordinary items.[23]

In testing our hypotheses, we pool data from firms that vary widely in size and share price. For example, the median firm size in our sample during the 1980s, as measured by its average market capitalization, is $128 million; the interquartile range of market capitalization is $353 million. The corresponding values based on price per share are $12.77 and $11.88, respectively. We need to address the potential heterogeneity that results from drawing quarterly results from such a wide range of firms.

The literature commonly normalizes EPS by deflators such as price per share or assets per share in an attempt to homogenize the distribution from which the different observations are drawn. However, because EPS is measured (and reported and forecast) rounded to the closest penny, spurious patterns can arise in the distribution of such normalized EPS. (This problem appears to have been overlooked previously.)[24] For example, exactly zero EPS (or change in EPS or forecast

23. Philbrick and Ricks (1991) argue that analysts fail to account for special items, especially asset sales, that affect reported earnings. They recommend that the reported earnings before extraordinary items also be purged of the after-tax effects of asset sales. See also Keane and Runkle (in press). There are some large outliers in the set of reported earnings recorded by I/B/E/S in the post-1984 sample that could be corrected by cross-checking with Compustat data. However, since our analysis focuses on observations in a region far from the tails of the distributions, this problem of possibly spurious outliers is not significant for us. In our analyses, we do not make adjustments to the EPS numbers coded by I/B/E/S for the post-1985 sample.

24. This problem is analogous to the "aliasing problem" in the literature on the spectral analysis of time-series data (e.g., see Koopmans 1974, ch. 3). The classic aliasing problem arises when the spectrum of interest is a continuous-time series but the available sample was sampled at discrete intervals. In this situation, either lack of prior knowledge of the specific bounds of the frequency interval in which the spectrum is concentrated, or an
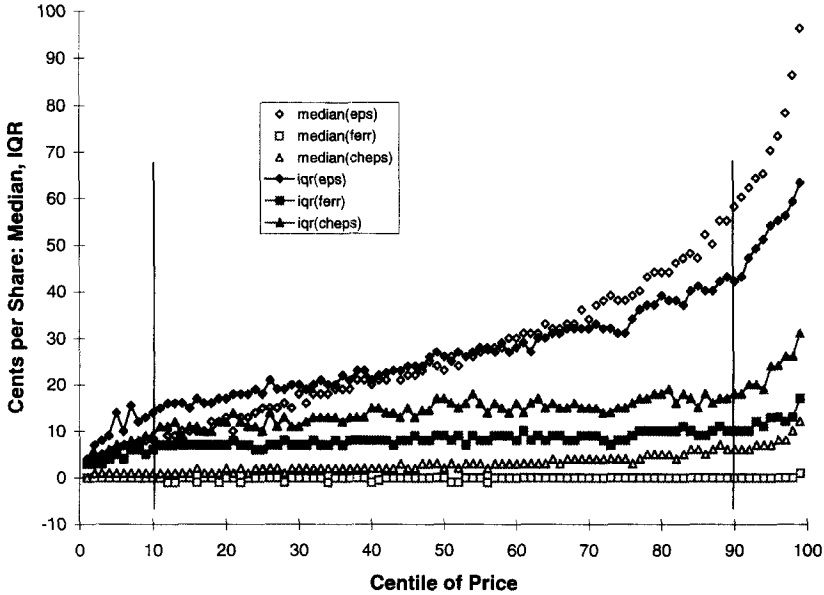
FIG. 4.—Medians and interquartile ranges for EPS, FERR, and ΔEPS as a function of centiles of price per share.

error) occur with nontrivial probability because of the rounding (as does any specific value like 1 penny). However, a zero remaps to zero after deflation compared to, for instance, a 1-penny EPS that can remap into a relatively large or small number depending on the deflator. Thus, deflation can lead to a spurious buildup in the density at zero, a critical area of interest for our study. In simulations not shown, this problem proves significant under conditions where EPS is rounded off to the nearest penny (as in practice).

Fortunately, if we exclude the extreme firms in terms of price, then deflation to correct for possible heterogeneity proves unnecessary for the important variables related to EPS that we study. Figure 4 shows the medians (represented by hollow symbols) and interquartile ranges (represented by corresponding solid symbols which are connected) of the important variables as a function of centiles of price per share.

The best situation for our study would arise if the measures of location (median) and dispersion (interquartile range) proved to be homogeneous across the different centiles. Consider for instance the analysts'

inability to sample often enough, results in accurate estimates of the *sampled* process spectrum providing poor or misleading estimates of the original spectrum. In our setting, the estimate of the probability density function risks distortion owing to the initial rounding off (discretization) of EPS and any subsequent renormalization.

forecast error (FERR), constructed as the reported EPS minus the mean of the analysts' forecasts. In figure 4, FERR's median and interquartile range are indicated by squares. These measures are reasonably independent of price per share if we focus on the middle 80% of the sample indicated as the region between the two vertical lines drawn at 10% and 90% in figure 4. Consider the case of the change in earnings per share, denoted $\Delta$EPS, which is simply EPS minus EPS of 4 quarters ago. The distribution of $\Delta$EPS, like FERR, appears stable in the middle 80% of the sample given in figure 4. In the analysis that follows, we restrict our sample to the middle 80% of the sample, which delivers reasonable homogeneity.

We further analyzed the sample for heterogeneity caused by variation across different time periods. For the culled sample of the middle 80%, time variation in the distribution proved not to be a major problem.

However, the situation for the basic EPS series itself is not resolved by restricting our sample to the middle 80%. Earnings-per-share medians as well as interquartile range increase steadily throughout the centiles of price per share, as is readily seen in figure 4. Therefore, in any analysis with EPS, we check whether results obtained for our entire sample hold for each of the quartiles of the middle 80% (i.e., 11%–30%, 31%–50%, 51%–70%, and 71%–90% from the preculled sample).

## B.  Historical Evidence of Earnings Management

The hypotheses about threshold-driven EM predict discontinuities in earnings distributions at specific values. As a first cut, we assess empirical histograms, focusing on the region where the discontinuity in density is predicted for our performance variables. Second, we compute a test statistic, $\tau$, that indicates whether or not to reject the null hypothesis that the distribution underlying the histogram is continuous and smooth at the threshold point. Since traditional statistical tests are not designed to test such hypotheses, we developed a test statistic, $\tau$, which extrapolates from neighborhood densities to compute expected density at the threshold assuming no unusual behavior there. The appendix discusses our testing method.

To construct empirical histograms requires a choice of bin width that balances the need for a precise density estimate with the need for fine resolution. Silverman (1986) and Scott (1992) recommend a bin width positively related to the variability of the data and negatively related to the number of observations; for example, one suggestion calls for a bin width of $2(\text{IQR})n^{-1/3}$, where IQR is the sample interquartile range of the variable and $n$ is the number of available observations. Given our sample sizes and dispersions of variables, such formulas imply a bin width of 1 penny (the minimum resolution for our data).
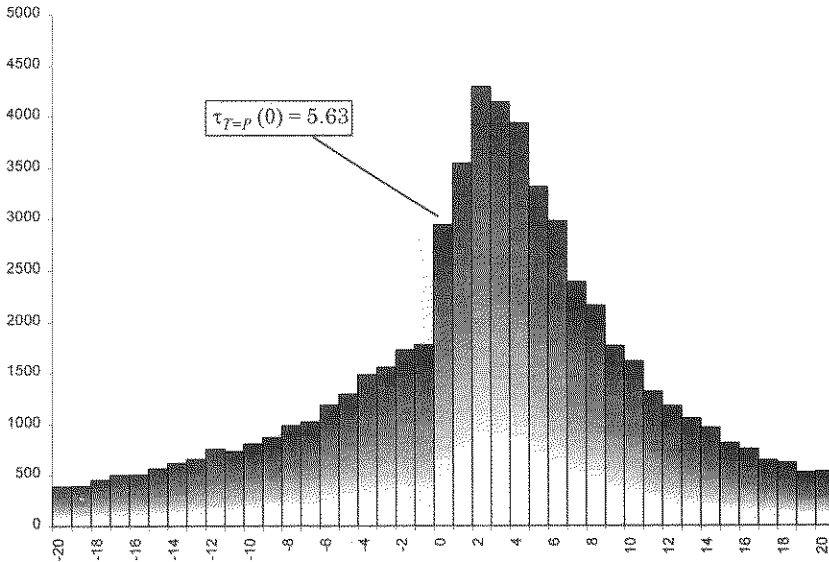
FIG. 5.—Histogram of change in EPS ($\Delta$EPS $=$ EPS$_t$ $-$ EPS$_{t-4}$): exploring the threshold of "sustain recent performance."

1. *"Sustain Recent Performance."* Press reports on corporate earnings typically compare current results with those from a year ago. Consistent with this practice, we provide evidence that earnings from 1 year ago constitute an important threshold for earnings reports, as we posited in our model. The distribution of the change in earnings, denoted $\Delta$EPS, is simply EPS minus EPS of 4 quarters ago. (The appropriate recent available benchmark proves to be the corresponding quarter from a year ago since earnings exhibit strong annual seasonal variation.) The distribution of $\Delta$EPS is plotted in figure 5.

Since corporate earnings tend to grow (surely in nominal terms), we do not expect the central tendency of the distribution to be close to zero. Indeed, the median and the mode of the distribution of the overall sample are 3 cents, while the mean is 0.81 cents. It is all the more remarkable, then, that we observe a large jump in the distribution at zero. In the region of small negative changes, the distribution appears to have been "shaved," with some density mass transferred to zero or slightly above. This pattern of $\Delta$EPS is consistent with executives' managing earnings to come in at or above the comparable figure for 4 quarters ago.[25]

25. A qualitatively similar pattern is reported in Burgstahler and Dichev (1997, fig. 1), although, since they deflate earnings, the extreme dip in density just below zero in their distribution of scaled earnings is most likely spurious (as discussed in IIIA above).
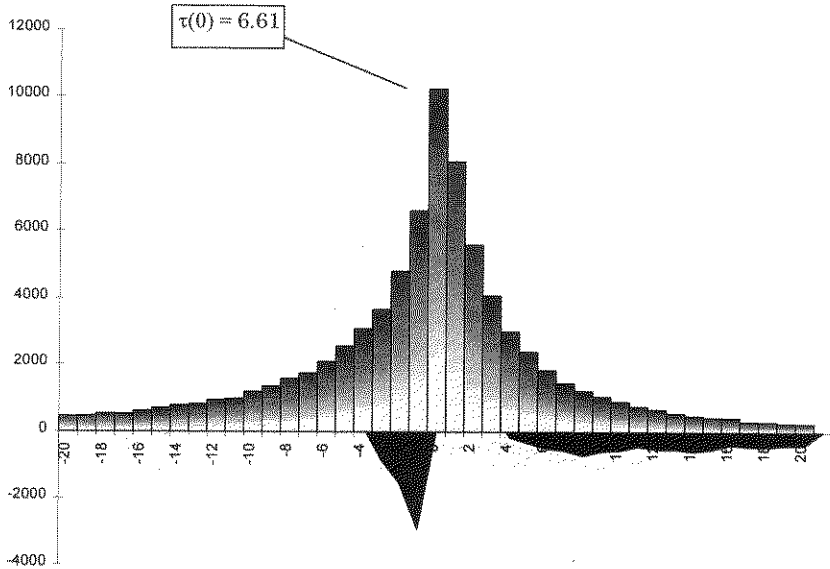
FIG. 6.—Histogram of forecast error for earnings per share: exploring the threshold of meeting analysts' expectations.

The easily discernible pileup of observations at the zero threshold for ΔEPS is confirmed by the τ-statistic of 6.61.[26] The value of 6.61 is the largest for all points in the neighborhood as well as being very significant. These findings are also confirmed with the subsamples of I/B/E/S and Q-Prime (unreported).

2. *"Meet Analysts' Expectations."* Figure 6 plots the empirical distribution of the forecast error, FERR (equal to EPS minus the analysts' consensus EPS forecast) in 1-penny bins in a range around zero, using quarterly observations over the 1974–96 period.

Consistent with the notion that ''making the forecast'' is an important threshold for managers, the distribution of FERR drops sharply below zero: we observe a smaller mass to the left of zero compared to the right. (Note that in the histogram, the bin starting with zero represents observations that are exactly zero.)

There is an extra pileup of observations at zero, although this is hard to see for a distribution like FERR that is centered on zero itself. The pileup is confirmed by the τ-statistic of 5.63, which is very significant.[27]

---

26. In this case the likely threshold is not at the peak of the distribution although its neighborhood includes the peak; see elaboration A1 discussed in the appendix.

27. In this case, we compute τ for the case in which the likely threshold is at the peak of the distribution (see elaboration A2 in the appendix).

This value exceeds the values of $\tau$ for all the neighboring points, none of which exceed 2.0 in absolute value (unreported).

Parallel to figure 3 (which, like FERR, had the latent distribution centered at the threshold), we show the shortfall in density below the histogram in figure 6 for the different outcomes—see the dark shaded areas below the horizontal axis. As predicted by consideration of earnings management to exceed the threshold of analysts' forecasts, we find that (1) the region just below zero exhibits a significant shortfall owing to "borrowing for a better today," and (2) the region of large positive forecast errors shows some shortfall owing to a combination of reduced density of "reining in" and excess density at the mirror bins from "saving for a better tomorrow."[28]

Previous studies on analysts' forecasts have reported an "optimistic bias": analysts' forecasts exceed reported earnings on average. Optimistic bias in the mean of forecasts works against our contention that executives will manage reported earnings to meet or exceed analysts' forecasts. This in turn suggests that a supportive finding will be more meaningful.

Fortunately, the two forces that may explain the data in figure 6— EM to attain or exceed the forecast, and a mean optimistic bias in the forecast—can be reconciled. It is sufficient that most of the time executives meet or slightly exceed analysts' forecasts but that they sometimes fall dramatically short. Given those forces, the forecast error distribution will be skewed, with a long left tail. This pattern appears in our sample: the mean of FERR is $-5.43$ while the median is zero; the skewness measure computes to $-43$ (whose $p$-value is near zero under the null hypothesis of a symmetric distribution). This confirms a statistically significant left-skewed distribution of earnings relative to forecast.

3. *"Report Positive Profits."* Our third possible important threshold is probably the most natural: positive earnings. To know whether this threshold has been reached, investors need no information on the company's performance history or the market's consensus forecast. This threshold also addresses the most important question for shareholders: is this firm profitable? The complication for studying a distribution of EPS, as discussed previously, is that the distribution is not homogeneous with respect to price per share. Thus, while we discuss the results for the overall sample, we confirm that similar findings emerge as well as for subsamples based on quartiles of price per share. In figure 7, we show the distribution of EPS in a window around zero.

Two patterns emerge. First, similar to $\Delta$EPS, the EPS distribution appears to be shaved in the negative region, consistent with the hypoth-

---

28. In results not reported, these findings are confirmed with the subsamples of I/B/E/S and Q-Prime.
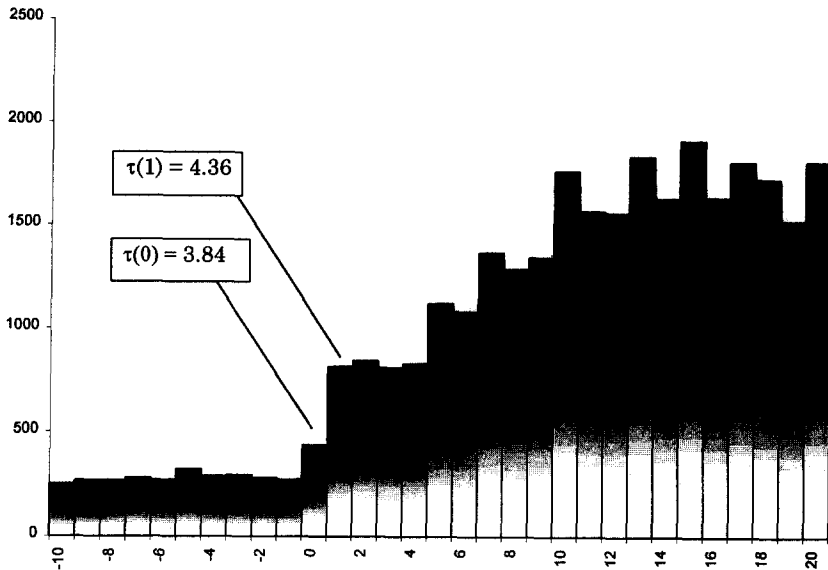
FIG. 7.—Histogram of EPS: exploring the threshold of "positive/zero profits"

esis of loss aversion. Second, the EPS distribution shows a considerable jump between zero and one (especially the latter); thus it appears that managers strongly desire to be able to report strictly positive earnings—as opposed to just breaking even.[29] The value of the $\tau$-statistic (based on the basic test in the appendix—the case in which the likely threshold is far from the peak of the distribution) confirms the observable pattern: at 1 cent per share, we obtain a $\tau$ value of 4.36, confirming a discontinuity in the EPS distribution there.[30]

Finally, we can also detect an upward kink in the EPS distribution from $-1$ cent/share to 0 cents/share, indicating a secondary threshold at zero, to "avoid red ink." The $\tau$-value for the secondary kink, 3.84, also proves significant, although our visual impression is that the threshold at zero is likely smaller than that at 1 cent/share.

In sum, we have established clear thresholds effects in the reporting of earnings, both visually and through statistical test results. The three thresholds affecting the reporting of earnings are to "sustain recent performance," to "meet analysts' forecasts," and to "report profits."

---

29. Note that this figure reinforces the impression of figure 1 in Hayn (1995), who, however, scales EPS by price per share and thus obtains a confounding density estimate at zero (as discussed above in Subsection IIIA).

30. Since we have two suspected thresholds adjacent to one another, the neighborhood used in the $\tau$-tests here always excludes the observation corresponding to the other threshold observation.

## C. *Conditional Distributions: Interaction among Thresholds*

If executives pay attention to more than one threshold, as seems likely, is one threshold more important than another? Is there a discernible hierarchy among them?

To investigate interactions between thresholds, an issue that appears to have been ignored by the literature, we analyze the conditional distributions of EPS, ΔEPS, and FERR, when each of the other thresholds is met or failed. For example, suppose that we find a significant threshold in the EPS distribution when we condition it on making the analysts' forecast, as well as when we condition it on failing to reach the analysts' forecast; and suppose that we also find that one of the "parallel" distributions of FERR (i.e., FERR conditional on EPS $> 0$, and FERR conditional on EPS $< 0$) exhibits no threshold effect: in this hypothetical example, we would conclude that "to report profits" is a more important threshold than "make the analysts' forecast."

Twelve conditional distributions are of interest (three individual distributions $\times$ conditioning on the other two $\times$ two levels). We attend to the problem that there may be little or no room to meet or fail a threshold when another threshold is met or failed. For example, if the ΔEPS threshold is met, and analysts have predicted earnings below last year, there is no possibility of missing the FERR threshold. We thus focus only on samples where there is at least a 5-cent range within which the threshold could be failed or met. While this conditioning reduces the available sample size, we are assured that inferences in the neighborhood of the threshold are valid.

For our positive earnings threshold, we focus on the case of 1 cent in EPS (though the 0-cent threshold also appears important). Our other two thresholds imply a 0-cent threshold in the distribution of ΔEPS and FERR. Our results appear in the 12 panels of conditional distributions in figure 8, where the vertical lines indicate the thresholds.

To illustrate our requirements for inclusion, consider the second panel in the second row of figure 8 where we examine the ΔEPS threshold conditional on the EPS being positive. Since we want to have a 5-cent range where the firm could fall short of the ΔEPS threshold in this subsample, we consider only cases where lag(EPS) $> 5$. This provides a range of 1–5 cents/share where the current EPS can be positive and still fail to attain the ΔEPS threshold. Each of our 12 cases looks at performance relative to one threshold conditional on another threshold having been failed or met. When the conditional threshold is failed (met), we preserve a 5-penny range on the up (down) side.[31]

---

31. The restrictions we impose on included conditions influences the shape of our histograms. Considering the upper-right-hand diagram, for example. Apart from thresholds, it is unlikely that earnings will be nine cents when the analysts' forecast (AF) $< -4$, whereas $-9$ is not so unlikely even though FERR $\geq 0$. Our statistical test, which only looks at earnings in a range of 10 cents, mitigates this problem.

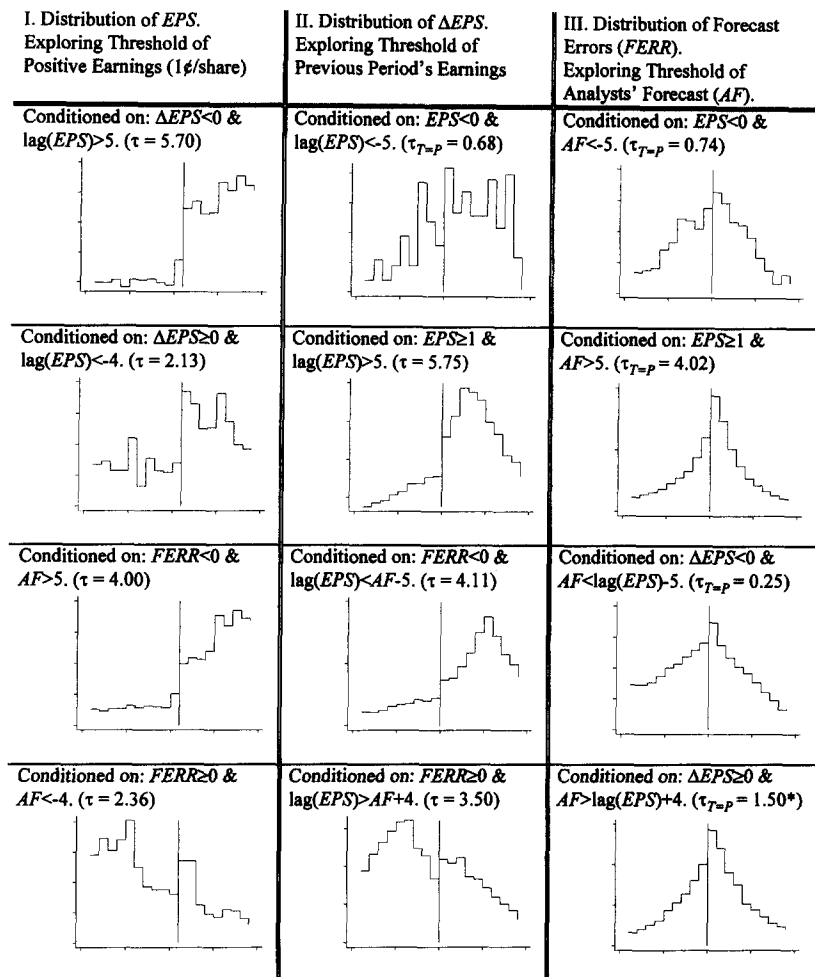| I. Distribution of *EPS*. Exploring Threshold of Positive Earnings (1¢/share) | II. Distribution of Δ*EPS*. Exploring Threshold of Previous Period's Earnings | III. Distribution of Forecast Errors (*FERR*). Exploring Threshold of Analysts' Forecast (*AF*). |
|---|---|---|
| Conditioned on: Δ*EPS*<0 & lag(*EPS*)>5. ($\tau = 5.70$) | Conditioned on: *EPS*<0 & lag(*EPS*)<-5. ($\tau_{T=P} = 0.68$) | Conditioned on: *EPS*<0 & *AF*<-5. ($\tau_{T=P} = 0.74$) |
| Conditioned on: Δ*EPS*≥0 & lag(*EPS*)<-4. ($\tau = 2.13$) | Conditioned on: *EPS*≥1 & lag(*EPS*)>5. ($\tau = 5.75$) | Conditioned on: *EPS*≥1 & *AF*>5. ($\tau_{T=P} = 4.02$) |
| Conditioned on: *FERR*<0 & *AF*>5. ($\tau = 4.00$) | Conditioned on: *FERR*<0 & lag(*EPS*)<*AF*-5. ($\tau = 4.11$) | Conditioned on: Δ*EPS*<0 & *AF*<lag(*EPS*)-5. ($\tau_{T=P} = 0.25$) |
| Conditioned on: *FERR*≥0 & *AF*<-4. ($\tau = 2.36$) | Conditioned on: *FERR*≥0 & lag(*EPS*)>*AF*+4. ($\tau = 3.50$) | Conditioned on: Δ*EPS*≥0 & *AF*>lag(*EPS*)+4. ($\tau_{T=P} = 1.50*$) |

FIG. 8.—Twelve conditional distributions: exploring the relative importance of thresholds. Note: the appendix provides motivation and construction details for the $\tau$ and $\tau_{T=P}$ statistics. Values in excess of 2.0 reliably reject the null hypothesis of no threshold effect. Whenever the values of the statistics are found to be in excess of 2.0 at the suspected thresholds for the above cases, they also prove to the largest relative to similar computations for all other points; the last panel reporting a $\tau_{T=P} = 1.50$ also obtains the largest value in its neighborhood.

Possible discontinuity in the histograms is evaluated for each of our 12 conditions employing the $\tau$-test used before for the univariate distributions, which is discussed in the appendix. The $\tau$-statistics confirm concretely what we see in figure 8. The EPS threshold is highly robust. Whether either of our other two thresholds fails or is met, there is a leap upward in density at our EPS threshold. Our other two thresholds

are less robust to the conditioning of whether another threshold is met or failed; that is, under some conditions, they are not significant. Specifically, in row 1 of column 2 in figure 8, we observe no clear discontinuity when the EPS earnings are negative. However, a discontinuity in $\Delta$EPS appears in the other three rows of column 2. We conclude that the threshold of the previous period's earnings is "robust" with respect to the threshold of analysts' forecast but not to that of positive earnings.

Similarly, in rows 1 and 3 of column 3 in figure 8, we observe no clear discontinuities in the distribution of FERR when EPS $<$ 0 and $\Delta$EPS $<$ 0, respectively. The effect of a FERR threshold reasserts itself when EPS $>$ 1 or $\Delta$EPS $>$ 0, as seen in rows 2 and 4 of column 3. We conclude that the threshold of analysts' forecast matters only if the other thresholds are met.

Overall, a threshold hierarchy emerges. The positive EPS threshold is the most important; it prevails regardless of whether or not the other two thresholds are met. The threshold of previous period's earnings is second in importance; it asserts itself only if the positive EPS threshold is met but is visibly present regardless of whether earnings make the analysts' forecast. The threshold of analysts' forecast is the weakest; it only matters if both the other thresholds are met.

## IV. The Effect on Future Earnings from Earnings Management

The previous section showed that executives manage earnings to meet three thresholds. When they do so, as discussed in Section II, current earnings are raised by "borrowing" from future earnings. From an empirical perspective, firms that barely attain a threshold are suspect of having engaged in upward earnings management. This section examines whether the future performance of firms barely meeting thresholds is inferior compared to control groups.

### A. The Implication of Earnings Management for Future Earnings

Earnings management to meet thresholds in this period will affect next period's earnings. Thus, we investigate whether following a period with likely EM there is any predictable effect on earnings. Our analysis in this section focuses first on the $\Delta$EPS threshold and then on the positive earnings threshold.[32]

We examine the performance of the suspect firms that just meet the threshold relative to the performance of firms that just miss the thresh-

---

32. We exclude the analysts' forecast threshold from this analysis. Even if a firm strives to meet the analysts' forecast in a given period, it is unlikely that it will find it harder to meet it in the following period, simply because the analysts' forecast is an endogenous target that will itself move according to firm performance or executives' announcements.

old or easily surpass it. Accordingly, we divide firms into three groups, depending on their earnings. Group A fails to meet the threshold, B just meets or exceeds it, C beats it easily. Each group has a 5-penny range. Group B is likely to include a number of firms that managed their earnings upward to meet the threshold. Group C is less likely to have boosted earnings and may have reined them in. Denote the average performance of a group by the corresponding lower-case letter, and indicate the period of the performance by a subscript (1 or 2). By assumption, we have $c_1 > b_1 > a_1$. Normally we would expect some persistence in both earnings level and in the change in earnings. Thus, absent any earnings management, we would expect $c_2 > b_2 > a_2$. How might EM affect these inequalities? Earnings recorded by group B are suspect of upward manipulation. Hence, $b_2$ would move down relative to both $c_2$ and $a_2$, giving $c_2 - b_2 > b_2 - a_2$. If EM is substantial, we might even have $a_2 > b_2$; that is, the lower performer in period 1 (those that just fall short of the threshold) would do better in period 2.

To facilitate comparisons, we add a fourth group D that strongly surpasses the threshold. Presumably, manipulation is less prominent for groups C and D, so their difference in performance in some sense provides a benchmark for what we should expect between adjacent groups.

### B.  Evidence on the "Borrowing" of Future Earnings

Consider first the threshold "sustain recent performance" (i.e., $\Delta$EPS = 0). We study the fiscal-year performance of firms since we expect fiscal-year effects to be the most powerful ones. We restrict ourselves to the subset of firms in which the fiscal year ends in December to avoid observations overlapping in time. We also restrict consideration to firms that show an uptick in the last quarter's performance since firms that manage earnings for a given year are likely to show an uptick in the last quarter's performance. (Concerns regarding spurious inferences induced by this artifact of the sample choice are addressed below.)

A widely accepted stylized fact is that a significant component of earnings changes is permanent.[33] But if there is significant mean reversion in $\Delta$EPS, then we may have difficulty discriminating earnings manipulation from overall behavior. Consider the regression of $\Delta\text{EPS}_{t+1} = \alpha + \beta\Delta\text{EPS}_t$. We estimated this relationship using fiscal-year observations but conditioning on $\Delta\text{EPS}_t > 5$ to exclude contamination of threshold effects. The estimated $\beta$ proved to be close to zero $(-0.05)$ in our sample; thus, a reasonable null hypothesis is that the ranking of firms by $\Delta$EPS over the next year will be unrelated to this year's

---

33. The earnings literature, e.g., Hayn (1995), notes that loss-reporting firms have different time-series properties of EPS. However, all our inferences in this section prove robust to conditioning on the sign of EPS.

TABLE 1     **Next Year's Relative Performance by Groups Formed around the No-Change Threshold of the Formation Quarter's ΔEPS; Subsample of Firms with ΔEPS ≥ 0 for Formation Quarter**

| Annual ΔEPS (Cents/Share) in Formation Quarter | Groups by Performance in Formation Year | | | |
|---|---|---|---|---|
| | A. Miss Threshold −5 to −1 | B. Meet Threshold 0 to +4 | C. Surpass Threshold +5 to +9 | D. Strongly Surpass Threshold +10 to +14 |
| No. of observations | 1,143 | 2,220 | 3,688 | 4,049 |
| Performance in following year: | | | | |
|   1. Mean ΔEPS | 9.48 | 7.10 | 6.53 | 9.44 |
|   2. Median ΔEPS | 13 | 8 | 9 | 12 |
|   3. Wilcoxon test* | 4.02 | −.77 | −7.54 | N.A. |
| *p*-values values (reporting column relative to next column) | .0001 | .4389 | .0000 | |

NOTE.—N.A. = not applicable.

\* The Wilcoxon test compares a group's performance in the postformation year with that of the next group. Under the null hypothesis that the distributions of performance of the two groups being compared are the same, the Wilcoxon test is distributed standard normal ($N(0,1)$).

performance. Moreover, in both the cases below, we find that the "strongly surpass threshold" never underperforms the "surpass threshold" group, reflecting some persistence in earnings growth in the neighborhood of interest.

If there is significant EM, we expect that the "meet threshold" group—many of whose members having presumably borrowed earnings—will underperform the groups immediately above ("surpass threshold") *and* below ("miss threshold"). This conjecture assumes that the EM effect exceeds any normal persistence in performance. This is a sharp and quite unusual prediction.

Table 1 column heads show the definitions of our four groups. Then it reports the mean and median of relative performance by group for the year following the formation of the groups. First note that, in our benchmark comparison, $d_2 > c_2$ as expected if heterogeneity in earnings growth among firms outweighs regression toward the mean effects. The salient comparisons for our purpose are between B and its neighbors. For both the mean and median, the performance of group B is worse than either of its neighbors, confirming our conjecture $c_2 - b_2 > b_2 - a_2$. In addition, $c_2 > b_2$, which is not surprising since C firms did better in period 1. The salient finding is that $b_2 < a_2$, presumably because of strong EM in the B group. All these differences prove statistically significant under the Wilcoxon test.[34]

34. The Wilcoxon test (also known as the Mann-Whitney two-sample statistic) is distributed standard normal under the null hypothesis that the performances of the two groups being compared have the same distribution. The test assumes independence across observa-

Firms that barely attain the threshold of sustain recent performance appear to borrow earnings from the next year.[35] The U-shaped pattern where B firms are outperformed by both A and C firms will be reinforced if groups just missing the threshold "save for a better tomorrow" and if those that surpass it "rein in."[36]

We repeat the analysis for the "report profits" threshold. Now groups are formed based on 4-quarter EPS performance. We divide firms as before using 5-cent bins. Since there are relatively few observations in the region of zero EPS, we do not apply further filters relating to fiscal-year end.

Under the null hypothesis discussed for table 1, there is no strongly expected order across groups for relative performance (i.e., annual $\Delta$EPS) in the postformation year, except perhaps because of heterogeneity in earnings growth potential among firms, which would predict $d_2 > c_2 > b_2 > a_2$. Summary results are reported in table 2.[37] The comparisons in table 2 yield one significant result. The meet-threshold group significantly underperforms the miss-threshold group in the sample A as well as B, a result indicative of strong EM.

However, the lack of significant difference between the meet and surpass thresholds suggests that evidence on earnings borrowings by the meet-threshold group is less conclusive than for the $\Delta$EPS threshold. While the meet-threshold group significantly underperforms the miss-threshold group, the meet-threshold group is not reliably outperformed by the surpass-threshold groups. In this case of the annual EPS threshold, evidence in earnings borrowings by the suspect meet-threshold group is less conclusive than it is for $\Delta$EPS.

---

tions, which is surely violated in our samples. This implies that the rejection rates using the traditional $p$-values based on the nominal size of our samples will be too high. However, the observed Wilcoxon test statistics for our sample are sufficiently large that the indicated U-pattern of performance is very unlikely to have arisen by chance.

35. The patterns are similar if we use a 10-penny range to define groups A, B, C and D, as well as if we use only quarters ending with the fiscal year.

36. Might the results in table 1 be spuriously induced because we select firms that have $\Delta$EPS $> 0$ in the most recent quarter? For instance, consider the miss-threshold group: it missed the annual threshold despite reporting relatively decent earnings in the latest quarter. This firm might be experiencing a rapid upward performance trend (compared to the meet-threshold group). If so, and given general persistence in earnings changes, the miss-threshold group would outperform the meet-threshold group in the next year absent earnings management. We check for this effect by using selection criterion of $\Delta$EPS $> 10$ and $\Delta$EPS $> 20$ for the most recent quarter. Given the construction of our groups, if the observed results in table 1 arise purely owing to the $\Delta$EPS $> 0$ filter, then we expect the meet-threshold group to outperform the surpass-threshold group with the $\Delta$EPS $> 10$ filter and the surpass-threshold group to outperform the strongly surpass-threshold group with the $\Delta$EPS $> 20$ filter. Neither turns out to be the case in our sample. The only performance reversal is observed between the miss-threshold group and the meet-threshold group.

37. Given the problems of heterogeneity for EPS identified in figure 4, we also studied the subsample of firms that were in the smallest quartile of price per share. Results (not reported) are qualitatively similar.

**TABLE 2    Next Year's Relative Performance by Groups Formed around the Zero Threshold of the Formation Year's EPS; Observations Are Restricted to ΔEPS > 0 for Formation Quarter**

| Annual EPS (Cents/Share) in Formation Quarter | Groups by Performance in Formation Year | | | |
|---|---|---|---|---|
| | A. Miss Threshold −5 to −1 | B. Meet Threshold 0 to +4 | C. Surpass Threshold +5 to +9 | D. Strongly Surpass Threshold +10 to +14 |
| No. of observations | 157 | 231 | 253 | 277 |
| Performance over following year: | | | | |
|   1. Mean ΔEPS | 37.29 | 10.89 | 8.36 | 21.50 |
|   2. Median ΔEPS | 35 | 10 | 18 | 18 |
|   3. Wilcoxon test* | 3.44 | −.51 | −.784 | N.A. |
| *p*-values (reporting column relative to next column) | .0006 | .6074 | .4329 | |

NOTE.—N.A. = not applicable.

* The Wilcoxon test compares a group's performance in the postformation year with that of the next group. Under the null hypothesis that the distributions of performance of the two groups being compared are the same, the Wilcoxon test is distributed standard normal ($N(0,1)$).

## V.    Concluding Discussion

Analysts, investors, and boards are keenly interested in financial reports of earnings because earnings provide critical information for investment decisions. Boards of directors charged with monitoring executives' performance recognize the importance of earnings to the firm's claimants and link managerial rewards to earnings outcomes. That this nexus of relations generates strong incentives for executives to manage earnings is hardly surprising. This analysis assesses the importance of thresholds for performance in this arena and the consequences thresholds have for patterns of reported earnings.

In work-in-progress that takes the threshold-based EM documented in this study as a given, we study how the equity market accounts for expected EM in resetting prices on announcement of earnings. Other ongoing investigation includes whether analysts efficiently account for EM in setting and revising their earnings forecasts, the salience of fiscal year thresholds, and whether different types of firms—for example, growth and value stocks—respond to the different incentives to manage earnings that are created because they suffer different penalties from falling short of thresholds.[38]

38. Dreman and Berry (1995*b*, pp. 23–24) find that low price-to-earnings ratio (P/E) (bottom-quintile) stocks fared better after a negative earnings surprise—actual earnings below the consensus forecast—than did high P/E (top-quintile) stocks. For a 1-year holding period, average annual market adjusted returns were +5.22% for the low P/E stocks but −4.57% for the high P/E stocks. The annualized differential for the quarter in which the surprise occurred was somewhat greater, +7.05% versus −5.69%.

Our model shows how efforts to exceed thresholds induce particular patterns of EM. Earnings falling just short of thresholds will be managed upward. Earnings far from thresholds, whether below or above, will be reined in, making thresholds more attainable in the future. Our empirical explorations find clear support for EM driven by three thresholds: report positive profits, sustain recent performance, and meet analysts' expectations. We observe discontinuities in the earnings distributions that indicate threshold-based EM. From explorations with conditional distributions, we infer that the thresholds are hierarchically ordered; it is most important first to make positive profits, second to report quarterly profits at least equal to profits of 4 quarters ago, and third to meet analysts' expectations. We also find evidence that the future performances of firms just meeting thresholds appear worse than those of control groups that are less suspect.[39]

Although earnings are a continuous variable, outsiders and insiders use psychological bright lines such as zero earnings, past earnings, and analysts' projected earnings as meaningful thresholds for assessing firms' performance. Theory suggests, and data document, that executives manage earnings in predictable ways to exceed thresholds.

## Appendix

### Testing for a Discontinuity in a Univariate Distribution

Let $x$ be the variable of interest, such as the change in earnings per share. The null hypothesis, $H_0$, conjectures that the probability density function of $x$, call it $f(x)$, is smooth at $T$, a point of interest because it may be a threshold under the alternative hypothesis, $H$. Given a random sample of $x$ of size $N$, we estimate the density for discrete ordered points $x_0, x_1, \ldots, x_n$, and so on.[40] Suppose the points are equispaced, and without loss of generality set the distance between the points to be of length one. Compute the proportion of the observations that lie in bins covering $[x_0, x_1), [x_1, x_2), \ldots, [x_n, x_{n+1})$, and so on. These proportions, denoted $p(x)$, provide estimates of $f(x)$ at $x_0, x_1, \ldots, x_n$, etc.[41]

## I.  Basic Test

The expectation of $\Delta p(x_n)$ [$\equiv p(x_n) - p(x_{n-1})$] is $f'(x_n)$, and its variance depends on the higher derivatives of $f(x)$ at $x_n$ as well as the available sample size $N$.

---

39. In related work not reported in this article, we have explored whether the special saliency of annual reports creates additional incentives to manipulate earnings. We find that the pressure to sustain recent performance at the fiscal-year horizon induces extra noisiness in fourth-quarter earnings that varies predictably with the temptation to "generate" earnings to meet the threshold.

40. For our analyses the $x$'s are integers, though nothing in our test approach requires this.

41. Under $H_0$, improved estimates of $f(x)$ are possible using neighborhood bins. However, the power of tests to reject $H_0$ (especially given our alternate hypotheses discussed below) may be compromised by such an approach. Fortunately in our case, unambiguous results obtain with this most simple estimation strategy.

Consider a small symmetric region $R_n$ around $n$ of $2r + 1$ points (i.e., $R_n = \{x_i;$ $i \in (n - r, n + r)\}$); given the smoothness assumption for $f(x)$ under $H_0$, the distribution of $\Delta p(x_i)$ will be approximately homogeneous.[42]

Use the observations $\Delta p(x_i)$ from $R_n$, excluding $\Delta p(x_n)$, to compute a $t$-like test statistic, $\tau$. Specifically, compute

$$\tau_n = \frac{\Delta p(x_n) - \underset{i \in R, i \neq n}{\text{mean}} \{\Delta p(x_i)\}}{\underset{i \in R, i \neq n}{\text{s.d.}} \{\Delta p(x_i)\}},$$

where mean and s.d. denote the sample mean and standard deviation of $\{\cdot\}$. We exclude observations corresponding to $i = n$ in the computation of the mean and standard deviation to increase power in identifying a discontinuity in $f(x)$ at $x_n$.

Our alternative hypothesis, $H_1$, conjectures a discontinuity in $f(x)$ at a preidentified threshold $T$ (i.e., zeros in the distributions of $\Delta$EPS or forecast errors of earnings, or 1-penny in the distribution of EPS). The distribution of $\tau_T$ is likely to be well approximated by the Student's $t$-distribution under $H_0$ if the distribution of $\Delta p(x_i)$ in $R_T$ is approximately Gaussian. In unreported simulations, working with the log transformation of the estimated density improved the homogeneity of variance (of $\Delta \log\{p(x)\}$) across typical neighborhoods $R$—thus, all the tests that we report in the text are based on $\Delta \log\{p(x)\}$ rather than on $\Delta p(x)$, though the inferences appear similar. In any case, we do not solely rely on normality. Instead we compare the $\tau_T$ to other $\tau$-values computed for nearby points.

We examine the rank of $\tau_T$ relative to the other $\tau$'s as well as its relative magnitudes to assess whether a discontinuity at $T$ can be established. Fortunately, clear, unambiguous results obtain: using the full sample, the $\tau_T$ values always prove to be the largest when compared to the other $\tau$ values.[43]

## II. Elaborations

The basic test sketched above is satisfactory as long as the point at which the density being examined for a discontinuity ($T$) falls significantly on one side of the peak of the probability density distribution. Denote the peak by $P$. Now consider the case when the symmetric construction of $R_T$ sketched above would include $P$. Since points on different sides of $P$ are likely to have slopes of the density function of opposite signs, the symmetric $R_T$ will no longer be composed of similar points in the sense of similar slopes.

CASE A1. Symmetric neighborhood around $T$ would include peak, $P$, though $T \neq P$. For this case, we construct an asymmetric neighborhood $R_T$ around $T$. When $T < P$ ($T > P$), construct $R_T$ to be the most symmetric region possible around $T$ of $2r + 1$ points such that all the points lie at or below (above) $P$. The intuition for this construction is that by selecting points on the same side of $P$

---

42. For our analysis, we selected $r = 5$, which creates 11-penny intervals. Briefly, we explored $r = 7$ and $r = 10$ for $\Delta$EPS, and the qualitative findings remain unchanged.

43. Given the 10 neighborhood values to which we compare $t_T$ (see n. 3 above), the likelihood of obtaining $t_T$ as the largest value by chance is slightly less than 10%. Looking at the magnitudes themselves, the neighborhood $t$ values interestingly always compute to less than $|2|$ while the $t_T$ values always exceed 2.

we obtain a neighborhood with points that have similar slopes of the (log) density function. Given such an $R_T$, we compute $\tau_T$ as in the basic test approach above.

CASE A2.   Suspected threshold coincides with the peak, that is, $T = P$. Consider the case of the analysts' forecast as the threshold, $T$. In this case, the distribution of reported earnings is likely centered at this $T$ if analysts forecast the mode or if the latent distribution of earnings is nearly symmetric and forecasters minimize the mean squared forecast error or the mean absolute error. Now, we identify an earnings management effect by testing whether the slope of the density function immediately to the left of $T$ ($=P$) is significantly different from the corresponding slope (adjusted for sign) to the immediate right of $T$ after allowing for any general local skew in the distribution.

Define $\nabla p_j \equiv \Delta\log\{p(x_{T+j})\} - (-1 \times \Delta\log\{p(x_{T-j})\})$. As remarked before, log transformations of the density appear to stabilize variance across nearby $j$'s in simulations as well for our samples (not reported). The test for case B2 amounts to examining whether $\nabla p_1$ is unusual. We use the observations $\nabla p_j$ from a small neighborhood $R$ ($j > 1$,) to compute an estimate of the mean of $\nabla p_1$ as well as its standard deviation.[44] As before, we compute a $t$-like test statistic, say $\tau_{T=P}$, to assess the "unusualness" of $\nabla p_1$.

In simulations that mimic the statistical structure of our sample while assuming a Gaussian distribution for the latent earnings, the statistic $\tau_{T=P}$ proves to be greater than 2.0 less than 5% of the time. Nonetheless, since the real distribution is unlikely to be as well behaved as Gaussian in the absence of any discontinuity at $T = P$, the comparison of $\tau_{T=P}$ with the real samples to the reference level of 2.0 is only taken as suggestive of a discontinuity. Thus, we also examine the rank of $\nabla p_1$ to the corresponding values at nearby $j$'s: in our samples, when $\tau_{T=P}$ proves to be larger than 2.0, $\nabla p_1$ is always the largest in the neighborhood.

## References

Abarbanell, J., and Bernard, V. 1992. Tests of analysts' overreaction/underreaction to earnings information as an explanation for anomalous stock price behavior. *Journal of Finance* 47:1181–1207.
Ball, R., and Brown, P. 1968. An empirical evaluation of accounting income numbers. *Journal of Accounting Research* 6:231–58.
Barrow, J. D. 1992. *Pi in the Sky*. Boston: Little, Brown.
Bruns, W., and Merchant, K. 1996. The dangerous morality of managing earnings. *Management Accounting* 72 (August): 22–25.
Burgstahler, D. 1997. Incentives to manage earnings to avoid earnings decreases and losses: Evidence from quarterly earnings. Working paper. Seattle: University of Washington.
Burgstahler, D., and Dichev, I. 1997. Earnings management to avoid earnings decreases and losses. *Journal of Accounting and Economics* 24 (December): 99–126.
Cornsweet, T. C. 1974. *Visual Perception*. New York: Academic Press.
DeBondt, W., and Thaler, R. 1995. Financial decision-making in markets and firms: A behavioral perspective. In R. Jarrow, V. Maksimovic, and W. Ziemba (eds.), *Finance*. Vol. 9 of *Handbook in Research and Management Science*. Amsterdam: Elsevier/North Holland.
Dechow, P. 1994. Accounting earnings and cash flows as measures of firm performance:

44. In the tests reported in the main text of this article the $R$ for computing $\tau_{T=P}$ spans 10 nearby values, i.e., $j = 2, 3, \ldots, 11$. Similar results, not shown, obtain with fewer nearby values.

The role of accounting accruals. *Journal of Accounting and Economics* 18 (July): 3–42.

Dechow, P.; Huson, M.; and Sloan, R. 1994. The effect of restructuring charges on executives' cash compensation. *Accounting Review* 69 (January): 138–56.

Dechow, P.; Sloan, R.; and Sweeney, A. 1995. Detecting earnings management. *Accounting Review* 70 (April): 193–225.

Dreman, D. N., and Berry, M. A. 1995a. Analyst forecasting errors and their implications for security analysis. *Financial Analysts Journal* 51 (May/June): 30–41.

Dreman, D. N., and Berry, M. A. 1995b. Overreaction, underreaction, and the low-P/E effect. *Financial Analysts Journal* 51 (July/August): 21–30.

Easton, P.; Harris, T.; and Ohlson, J. 1992. Aggregate accounting earnings can explain most of security returns: The case of long return intervals. *Journal of Accounting and Economics* 15:119–42.

Foster, G. 1986. *Financial Statement Analysis.* Englewood Cliffs, N.J.: Prentice-Hall.

Gaver, J. J.; Gaver, K. M.; and Austin, J. R. 1995. Additional evidence on bonus plans and income management. *Journal of Accounting and Economics* 19:3–28.

Glass, A. L., and Holyoak, K. J. 1986. *Cognition.* New York: Random House.

Hayn, C. 1995. The information content of losses. *Journal of Accounting and Economics* 20:125–53.

Healy, P. 1985. The effect of bonus schemes on accounting decisions. *Journal of Accounting and Economics* 7:85–107.

Holthausen, R. W.; Larcker, D. F.; and Sloan, R. G. 1995. Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics* 19 (February): 29–74.

Kahneman, D., and Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47:263–91.

Keane, M., and Runkle, D. 1998. Are financial analysts' forecasts of corporate profits rational? *Journal of Political Economy* 106, no. 4 (August): 768–805.

Koopmans, L. H. 1974. *The Spectral Analysis of Time Series.* New York: Academic Press.

Kothari, S. P., and Sloan, R. G. 1992. Information in prices about future earnings: Implications for earnings response coefficients. *Journal of Accounting and Economics* 15 (June): 143–71.

Laffont, J.-J., and Tirole, J. 1993. *The Theory of Incentives in Procurement and Regulation.* Cambridge, Mass.: MIT Press.

Milgrom, P., and Roberts, J. 1992. *Economics, Organization, and Management.* Englewood Cliffs, N.J.: Prentice-Hall.

Payne, J. L., and Robb, S. W. 1997. Earnings management: The effect of ex ante earnings expectations. Working paper. Oxford: University of Mississippi.

Philbrick, D. R., and Ricks, W. E. 1991. Using Value Line and I/B/E/S analyst forecasts in accounting research. *Journal of Accounting Research* 29 (Autumn): 397–417.

Schelling, T. 1960. *The Strategy of Conflict.* Cambridge, Mass.: Harvard University Press.

Schipper, K. 1989. Earnings management. *Accounting Horizons* 3:91–102.

Scott, D. W. 1992. *Multivariate Density Estimation: Theory, Practice, and Visualization.* New York: Wiley.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis.* London: Chapman & Hall.

Young, P. 1996. The economics of convention. *Journal of Economic Perspectives* 10 (Spring): 105–22.