THE ACCOUNTING REVIEW
Vol 70, No. 2
April 1995
pp. 193–225

Detecting Earnings Management

Patricia M. Dechow
Richard G. Sloan
University of Pennsylvania
Amy P. Sweeney
Harvard University

ABSTRACT: This paper evaluates alternative accrual-based models for detecting earnings management. The evaluation compares the specification and power of commonly used test statistics across the measures of discretionary accruals generated by the models and provides the following major insights. First, all of the models appear well specified when applied to a random sample of firm-years. Second, the models all generate tests of low power for earnings management of economically plausible magnitudes (e.g., one to five percent of total assets). Third, all models reject the null hypothesis of no earnings management at rates exceeding the specified test-levels when applied to samples of firms with extreme financial performance. This result highlights the importance of controlling for financial performance when investigating earnings management stimuli that are correlated with financial performance. Finally, a modified version of the model developed by Jones (1991) exhibits the most power in detecting earnings management.

Key Words: *Earnings management, Discretionary accruals, Models selection, SEC.*

Data Availability: *Data used in this study are publicly available from the sources identified in the paper. A listing of the firms investigated by the SEC is available from the authors.*

We appreciate the input of workshop participants at the University of Arizona, Harvard University (1994 Financial Decision and Control Seminars), Michigan State University, New York University, the University of Pennsylvania, Pennsylvania State University, Purdue University, the University of Rochester, Rutgers University, Stanford University (1993 summer camp), Temple University, Texas Christian University and the 1994 AAA annual meetings. We are particularly grateful for the suggestions of Bob Holthausen and two referees.

Submitted February 1994
Accepted December 1994

I. INTRODUCTION

ANALYSIS of earnings management often focuses on management's use of discretionary accruals.[1] Such research requires a model that estimates the discretionary component(s) of reported income. Existing models range from simple models in which discretionary accruals are measured as total accruals, to more sophisticated models that attempt to separate total accruals into discretionary and nondiscretionary components. There is, however, no systematic evidence bearing on the relative performance of these alternative models at detecting earnings management.

We evaluate the relative performance of the competing models by comparing the specification and power of commonly used test statistics. The specification of the test statistics is evaluated by examining the frequency with which they generate type I errors. Type I errors arise when the null hypothesis that earnings are *not* systematically managed in response to the stimulus identified by the researcher *is rejected* when the null is true. We generate type I errors for both a random sample of firm-years and for samples of firm-years with extreme financial performance. We focus on samples with extreme financial performance because the stimuli investigated in previous research are frequently correlated with financial performance. Thus, our findings shed light on the specification of test statistics in cases where the stimulus identified by the researcher does *not* cause earnings to be managed, but is correlated with firm performance.

The power of the test statistics is evaluated by examining the frequency with which they generate type II errors. Type II errors arise when the null hypothesis that earnings are *not* systematically managed in response to the stimulus identified by the researcher is *not rejected* when it is false. We generate type II errors in two ways. First, we measure rejection frequencies for samples of firm-years in which we have artificially added a fixed and known amount of accruals to each firm-year. These simulations are similar to those performed by Brown and Warner (1980, 1985) in evaluating alternative models for detecting abnormal stock price performance. However, our simulations differ in several respects. In particular, we must make explicit assumptions concerning the components of accruals that are managed and the timing of the accrual reversals. To the extent that our assumptions are not representative of the circumstances of actual earnings management, our results lack external validity. To circumvent this problem, we generate type II errors for a second set of firms, for which we have strong priors that earnings have been managed.[2] This sample consists of firms that have been targeted by the Securities and Exchange Commission (SEC) for allegedly overstating annual earnings. The external validity of these results rests on the assumption that the SEC has correctly identified firm-years in which earnings have been managed. This assumption seems reasonable, since the SEC (1992) indicates that out of the large number of cases that are brought to its attention, it only pursues cases involving the most significant and blatant incidences of earnings manipulation.

The empirical analysis generates the following major insights. First, all of the models appear well specified when applied to a random sample of firm-years. Second, the models all generate tests of low power for earnings management of economically plausible magnitudes (e.g., one to five percent of total assets). Third, all models reject the null hypothesis of no earnings

[1] See, for example, Healy (1985), DeAngelo (1986) and Jones (1991). Other constructs that have been used to detect earnings management include accounting procedure changes (Healy 1985; Healy and Palepu 1990; Sweeney 1994), specific components of discretionary accruals (McNichols and Wilson 1988; DeAngelo et al. 1994) and components of discretionary cash flows (Dechow and Sloan 1991).

[2] Schipper (1989) defines earnings management as purposeful intervention in the external financial reporting process, with the intent of obtaining some private gain (as opposed to, say, merely facilitating the neutral operation of the process) In the spirit of Schipper's definition, our procedure assumes that reported earnings in the firm-years targeted by the SEC are higher than they would have been under the neutral application of GAAP.

management at rates exceeding the specified test-levels when applied to samples of firms with extreme financial performance.[3] Finally, a version of the model developed by Jones (1991) that is modified to detect revenue-based earnings management generates the fewest type II errors.

The paper is organized as follows. Section II outlines the statistical testing procedure used to detect earnings management and highlights the effects of model misspecification on statistical inference. Section III introduces the competing models for measuring discretionary accruals. The experimental design is described in section IV and the empirical results are analyzed in section V. Section VI concludes the paper and provides suggestions for future earnings management research.

II. STATISTICAL ISSUES

This section considers potential misspecifications in tests for earnings management and their impact on inferences concerning earnings management. The analysis builds on a related analysis in McNichols and Wilson (1988). Following McNichols and Wilson, accrual-based tests for earnings management can be cast in the following linear framework:

$$DA_t = \alpha + \beta\,PART_t + \sum_{k=1}^{K} \gamma_k X_{kt} + \varepsilon_t, \tag{1}$$

where

DA = discretionary accruals (typically deflated by lagged total assets);
PART = a dummy variable partitioning the data set into two groups for which earnings management predictions are specified by the researcher;
X_k = (for k=1, .., K) other relevant variables influencing discretionary accruals; and
ε = an error term that is independently and identically normally distributed.

In most research contexts, PART will be set equal to one in firm-years during which systematic earnings management is hypothesized in response to the stimulus identified by the researcher (the "event period") and zero during firm-years in which no systematic earnings management is hypothesized (the "estimation period"). The null hypothesis of no earnings management in response to the researcher's stimulus will be rejected if $\hat{\beta}$, the estimated coefficient on PART, has the hypothesized sign and is statistically significant at conventional levels.

Unfortunately, the researcher cannot readily identify the other relevant variables, (the X_ks), and so excludes them from the model. Similarly, the researcher does not observe DA, and is forced to use a proxy, (DAP), that measures DA with error, (υ):

$$DAP_t = DA_t + \upsilon_t.$$

Thus, the correctly specified model can be expressed in terms of the researcher's proxy for discretionary accruals as

$$DAP_t = \alpha + \beta\,PART_t + \sum_{k=1}^{K} \gamma_k X_{kt} + \upsilon_t + \varepsilon_t. \tag{1'}$$

This model can be summarized as:

$$DAP_t = \alpha + \beta\,PART_t + \mu_t + \varepsilon_t, \tag{1''}$$

[3] The excessive rejection rates in the samples with extreme financial performance have two potential causes. First, nondiscretionary accruals (that are not extracted by the models) may be correlated with firm performance. Second, other factors that are correlated with firm performance may cause earnings to be systematically managed. In the first case, the null hypothesis is falsely rejected because of correlated measurement error in the proxy for discretionary accruals. In the second case, the tests are correctly detecting earnings management, but the cause of earnings management is not known Thus, if a researcher selects a stimuli that does not cause earnings to be managed but is correlated with firm performance, the test will be misspecified. We expand on these issues in section II.

where μ_t captures the sum of the effects of the omitted relevant variables on discretionary accruals and the error in the researcher's proxy for discretionary accruals. Given the regular Gaussian assumptions,[4] the OLS estimate of β, ($\hat{\beta}$), from a multiple regression of DAP on PART and μ is the best unbiased estimator of β. Also, the ratio of ($\hat{\beta}-\beta$) to its standard error, SE($\hat{\beta}$), has a t-distribution, which can be used to test for earnings management. This framework therefore provides a benchmark for evaluating the case where μ is omitted from the regression.

The model of earnings management typically estimated by the researcher can be represented as

$$DAP_t = \hat{a} + \hat{b}\,PART_t + e_t. \tag{2}$$

The researcher's model is misspecified by the omission of the relevant variable μ. Recall that the μ can represent either measurement error in DAP or omitted relevant variables influencing DA. Estimating model (2) using OLS has two undesirable consequences:[5]

(i) $\hat{b}$ is a biased estimator of β, with the direction of the bias being of the same sign as the correlation between PART and μ; and

(ii) SE ($\hat{b}$) is a biased estimator of SE ($\hat{\beta}$). In particular, if PART and μ are uncorrelated, SE($\hat{b}$) will provide an upwardly biased estimate of SE ($\hat{\beta}$).

These consequences lead to the following three problems for statistical inference in tests for earnings management:

Problem 1: *Incorrectly attributing earnings management to PART*

If the earnings management that is hypothesized to be caused by PART does not take place (i.e., the true coefficient on PART is zero) and μ is correlated with PART, then the estimated coefficient on PART, will be biased away from zero, increasing the probability of a type I error.

This problem will arise when (i) the proxy for discretionary accruals contains measurement error that is correlated with PART and/or (ii) other variables that cause earnings management are correlated with PART and are omitted from the analysis. In this latter case, earnings management is correctly detected by the model, but causality is incorrectly attributed to PART.

Problem 2: *Unintentionally extracting earnings management caused by PART*

If the earnings management that is hypothesized to be caused by PART does take place and the correlation between μ and PART is opposite in sign to the true coefficient on PART, then the estimated coefficient on PART will be biased toward zero. This will increase the probability of a type II error.

This problem will arise when the model used to generate the discretionary accrual proxy unintentionally removes some or all of the discretionary accruals. Under such conditions, the measurement error in the proxy for discretionary accruals (i.e., μ) will be

---

[4] The required assumptions are (i) ε_t is distributed independent normal with zero mean and common variance, σ_ε^2; and (ii) $PART_t$ and μ_t are distributed independently of ε_τ for all t and τ. The assumption that the residuals are normally distributed is not one of the original Gaussian assumptions. It is, however, required (i) for the OLS estimate to be the best of all (linear and nonlinear) unbiased estimators; and (ii) to derive the distribution of the test-statistic. Throughout the remainder of the paper, references to the Gaussian assumptions will therefore include the normality assumption.

[5] The derivation of these properties is identical to the standard derivation for the properties of OLS estimators in the case of the exclusion of a relevant regressor (e.g., Johnston 1984, 260–261).

negatively correlated with the discretionary accrual proxy, causing the coefficient on PART to be biased toward zero.

Problem 3: *Low power test*

If μ is not correlated with PART, then the estimated coefficient on PART will not be biased. However, the exclusion of relevant (uncorrelated) variables leads to an inflated standard error for the estimated coefficient on PART. This will increase the probability of a type II error.

We will refer back to each of these problems in our discussion of the models for detecting earnings management.

III. MEASURING DISCRETIONARY ACCRUALS

The usual starting point for the measurement of discretionary accruals is total accruals. A particular model is then assumed for the process generating the nondiscretionary component of total accruals, enabling total accruals to be decomposed into a discretionary and a nondiscretionary component. Most of the models require at least one parameter to be estimated, and this is typically implemented through the use of an "estimation period," during which no systematic earnings management is predicted. This paper considers five models of the process generating nondiscretionary accruals. These models are general representations of those that have been used in the extant earnings management literature. We have cast all models in the same general framework to facilitate comparability, rather than trying to exactly replicate the models as they may have appeared in the literature.

The Healy Model

Healy (1985) tests for earnings management by comparing mean total accruals (scaled by lagged total assets) across the earnings management partitioning variable. Healy's study differs from most other earnings management studies in that he predicts that systematic earnings management occurs in every period. His partitioning variable divides the sample into three groups, with earnings predicted to be managed upwards in one of the groups and downward in the other two groups. Inferences are then made through pairwise comparisons of the mean total accruals in the group where earnings is predicted to be managed upwards to the mean total accruals for each of the groups where earnings is predicted to be managed downwards. This approach is equivalent to treating the set of observations for which earnings are predicted to be managed upwards as the estimation period and the set of observations for which earnings are predicted to be managed downwards as the event period. The mean total accruals from the estimation period then represent the measure of nondiscretionary accruals. This implies the following model for nondiscretionary accruals:

$$NDA_\tau = \frac{\sum_t TA_t}{T}, \tag{4}$$

where

NDA = estimated nondiscretionary accruals;
TA = total accruals scaled by lagged total assets;
t = 1, 2,...T is a year subscript for years included in the estimation period; and
τ = a year subscript indicating a year in the event period.

The DeAngelo Model

DeAngelo (1986) tests for earnings management by computing first differences in total accruals, and by assuming that the first differences have an expected value of zero under the null hypothesis of no earnings management. This model uses last period's total accruals (scaled by lagged total assets) as the measure of nondiscretionary accruals. Thus, the DeAngelo Model for nondiscretionary accruals is:

$$NDA_\tau = TA_{\tau-1}\ . \tag{5}$$

The DeAngelo Model can be viewed as a special case of the Healy Model, in which the estimation period for nondiscretionary accruals is restricted to the previous year's observation.

A common feature of the Healy and DeAngelo Models is that they both use total accruals from the estimation period to proxy for expected nondiscretionary accruals. If nondiscretionary accruals are constant over time and discretionary accruals have a mean of zero in the estimation period, then both the Healy and DeAngelo Models will measure nondiscretionary accruals without error. If, however, nondiscretionary accruals change from period to period, then both models will tend to measure nondiscretionary accruals with error. Which of the two models is more appropriate then depends on the nature of the time-series process generating nondiscretionary accruals. If nondiscretionary accruals follow a white noise process around a constant mean, then the Healy model is appropriate. If nondiscretionary accruals follow a random walk, then the DeAngelo model is appropriate. Empirical evidence suggests that total accruals are stationary in the levels and approximate a white noise process (e.g., Dechow 1994).

The assumption that nondiscretionary accruals are constant is unlikely to be empirically descriptive. Kaplan (1985) points out that the nature of the accrual accounting process dictates that the level of nondiscretionary accruals should change in response to changes in economic circumstances. Failure to model the impact of economic circumstances on nondiscretionary accruals will cause inflated standard errors due to the omission of relevant (uncorrelated) variables (problem 3). In addition, if the firms examined are systematically experiencing abnormal economic circumstances, then failure to model the impact of economic circumstances on nondiscretionary accruals will result in biased estimates of the coefficient on PART (problem 1).

The Jones Model

Jones (1991) proposes a model that relaxes the assumption that nondiscretionary accruals are constant. Her model attempts to control for the effect of changes in a firm's economic circumstances on nondiscretionary accruals. The Jones Model for nondiscretionary accruals in the event year is:

$$NDA_\tau = \alpha_1(1/A_{\tau-1}) + \alpha_2(\Delta REV_\tau) + \alpha_3(PPE_\tau)\ , \tag{6}$$

where

ΔREV_τ = revenues in year τ less revenues in year $\tau-1$ scaled by total assets at $\tau-1$;
PPE_τ = gross property plant and equipment in year τ scaled by total assets at $\tau-1$;
$A_{\tau-1}$ = total assets at $\tau-1$; and
$\alpha_1, \alpha_2, \alpha_3$ = firm-specific parameters.

Estimates of the firm-specific parameters, α_1, α_2 and α_3 are generated using the following model in the estimation period:

$$TA_t = a_1(1/A_{\tau-1}) + a_2(\Delta REV_t) + a_3(PPE_t) + \upsilon_t\ , \tag{7}$$

where

a_1, a_2 and a_3 denote the OLS estimates of α_1, α_2 and α_3 and TA is total accruals scaled by lagged total assets. The results in Jones (1991) indicate that the model is successful at explaining around one quarter of the variation in total accruals.

An assumption implicit in the Jones model is that revenues are nondiscretionary. If earnings are managed through discretionary revenues, then the Jones Model will remove part of the managed earnings from the discretionary accrual proxy (problem 2). For example, consider a situation where management uses its discretion to accrue revenues at year-end when the cash has not yet been received and it is highly questionable whether the revenues have been earned. The result of this managerial discretion will be an increase in revenues and total accruals (through an increase in receivables). The Jones model orthogonalizes total accruals with respect to revenues and will therefore extract this discretionary component of accruals, causing the estimate of earnings management to be biased toward zero. Jones recognizes this limitation of her model (see Jones 1991, footnote 31).

The Modified Jones Model

We consider a modified version of the Jones Model in the empirical analysis. The modification is designed to eliminate the conjectured tendency of the Jones Model to measure discretionary accruals with error when discretion is exercised over revenues. In the modified model, nondiscretionary accruals are estimated during the event period (i.e., during periods in which earnings management is hypothesized) as:

$$NDA_\tau = \alpha_1(1/A_{\tau-1}) + \alpha_2(\Delta REV_\tau - \Delta REC_\tau) + \alpha_3(PPE_\tau), \tag{8}$$

where

ΔREC_τ = net receivables in year τ less net receivables in year τ-1 scaled by total assets at τ-1.

The estimates of α_1, α_2, α_3 and nondiscretionary accruals during the estimation period (in which no systematic earnings management is hypothesized) are those obtained from the original Jones Model. The only adjustment relative to the original Jones Model is that the change in revenues is adjusted for the change in receivables in the event period. The original Jones Model implicitly assumes that discretion is not exercised over revenue in either the estimation period or the event period. The modified version of the Jones Model implicitly assumes that all changes in credit sales in the event period result from earnings management. This is based on the reasoning that it is easier to manage earnings by exercising discretion over the recognition of revenue on credit sales than it is to manage earnings by exercising discretion over the recognition of revenue on cash sales. If this modification is successful, then the estimate of earnings management should no longer be biased toward zero in samples where earnings management has taken place through the management of revenues.

The Industry Model

The final model considered is the Industry Model used by Dechow and Sloan (1991). Similar to the Jones Model, the Industry Model relaxes the assumption that nondiscretionary accruals are constant over time. However, instead of attempting to directly model the determinants of nondiscretionary accruals, the Industry Model assumes that variation in the determinants of nondiscretionary accruals are common across firms in the same industry. The Industry Model for nondiscretionary accruals is:

$$NDA_\tau = \gamma_1 + \gamma_2 \text{median}_I(TA_\tau), \tag{9}$$

where

$\text{median}_I(TA_\tau)$ = the median value of total accruals scaled by lagged assets for all non-sample firms in the same 2-digit SIC code.[6]

The firm specific parameters γ_1 and γ_2 are estimated using OLS on the observations in the estimation period.

The ability of the Industry Model to mitigate measurement error in discretionary accruals hinges critically on two factors. First, the Industry Model only removes variation in nondiscretionary accruals that is common across firms in the same industry. If changes in nondiscretionary accruals largely reflect responses to changes in firm-specific circumstances, then the Industry Model will not extract all nondiscretionary accruals from the discretionary accrual proxy. Second, the Industry Model removes variation in discretionary accruals that is correlated across firms in the same industry, potentially causing problem 2. The severity of this problem depends on the extent to which the earnings management stimulus is correlated across firms in the same industry.

IV. EXPERIMENTAL DESIGN

Sample Construction

The empirical analysis is conducted by testing for earnings management using four distinct samples of firm-years as event-years:

(i) a randomly selected sample of 1000 firm-years;
(ii) samples of 1000 firm-years that are randomly selected from pools of firm-years experiencing extreme financial performance;
(iii) samples of 1000 randomly selected firm-years in which a fixed and known amount of accrual manipulation has been artificially introduced; and
(iv) a sample of 32 firms that are subject to SEC enforcement actions for allegedly overstating annual earnings in 56 firm-years.

Sample (i) is designed to investigate the specification of the test statistics generated by the models when the measurement error in discretionary accruals (μ) is uncorrelated with the earnings management partitioning variable (PART). Because the earnings management partitioning variable is selected at random, it is expected to be uncorrelated with any omitted variables. Note that this is simply a test of whether the Gaussian assumptions underlying the regression are satisfied. The existence of uncorrelated omitted variables reduces the power of the test (problem 3), but will not systematically bias the type I errors.

The 1000 randomly selected firm-years are selected from the 168,771 firm-years on the COMPUSTAT industrial files with the necessary data between 1950 and 1991. The 1000 firm-years are selected in a sequential fashion and without replacement. A firm-year is not selected if its inclusion in the random sample leaves less than ten unselected observations for the estimation period. Selected firms have an average of 21.5 observations. The requirement of more than 10 observations is necessary to efficiently estimate the parameters of the nondiscretionary accrual models for each firm. This sequential selection procedure continues until the random sample consists of 1000 firm-years.

Sample (ii) is designed to test the specification of each model when the earnings management partitioning variable, PART, is correlated with firm performance. The earnings management stimulus investigated in many existing studies are correlated with firm performance. For

[6] The use of two-digit SIC levels represents a trade-off between defining industry groupings narrowly enough that the Industry Model captures the industry specific effects versus having enough firms in each industry grouping so that the model can effectively diversify firm-specific effects.

example, Healy (1995) hypothesizes that management reduce earnings when either earnings are below the lower bound or cash from operations are above the upper bound of top executive bonus plans. Researchers have also investigated whether management attempt to loosen debt covenant restrictions through their accrual choices (e.g., Defond and Jiambalvo 1994; DeAngelo et al. 1994). Firms close to debt covenant restrictions are often experiencing poor earnings and/or cash flow performance. A final example is studies investigating accrual manipulation around non-routine management changes (e.g., Pourciau 1993; DeAngelo 1988). DeAngelo (1988) points out that poor prior earnings performance is often cited as a reason for management change. Thus, sample (ii) is used to examine the impact of firm performance on model misspecification.

To investigate the estimates of discretionary accruals produced by the models when firm performance is unusual, firm-years are selected to have either extreme earnings performance or extreme cash from operations performance.[7] A "high" and a "low" sample is formed for each of the performance measures, resulting in a total of four samples. These samples are formed using the following procedure. Each of the performance measures is standardized by lagged total assets. All firm-years with available data on the COMPUSTAT industrial files are then separately ranked on each performance measure. For each measure, firm-years are assigned in equal numbers to decile portfolios based on their ordered ranks. Each portfolio contains approximately 17,000 firm-years. Samples of 1000 firm-years are randomly selected from the highest and lowest portfolios for each performance measure using the same procedure that was discussed for sample (i).

Sample (iii) is designed to evaluate the relative frequency with which the competing models of nondiscretionary accruals generate type II errors. Brown and Warner (1980, 1985) investigate the type II errors of alternative models for measuring security price performance by artificially introducing a fixed and known amount of abnormal stock price performance into a randomly selected sample of firm-years. Inducing abnormal accruals is more complex than inducing abnormal stock returns for two reasons. First, we have to make explicit assumptions concerning the component(s) of accruals that are managed. This assumption is critical for the Jones Model, because if we introduce earnings management by artificially inflating revenues, then both accruals and revenue increase. The increase in revenue will affect the estimate of nondiscretionary accruals generated by the Jones Model. Second, since accruals must sum to zero over the life of the firm, artificially inducing discretionary accruals requires additional assumptions about the timing of the accrual reversals. Thus, we artificially introduce earnings management into sample (iii), but recognize that the external validity of the results is contingent upon how representative our assumptions are of actual cases of earnings management.

We obtain sample (iii) by beginning with the 1000 randomly selected firm-years in sample (i) and then adding accrual manipulation ranging in magnitude from zero percent to 100 percent of lagged assets (in increments of ten percent). In all cases, we assume that the accruals fully reverse themselves in the next fiscal year. We make three different sets of assumptions regarding the components of accruals that are managed:

(1) *Expense Manipulation* - delayed recognition of expenses. This approach is implemented by adding the assumed amount of expense manipulation to total accruals in the earnings management year, and subtracting the same amount in the following year. Since none of the models use expenses to estimate nondiscretionary accruals, none of the other variables used in the study need to be adjusted.

[7] We focus on the most extreme deciles of each performance measure to generate powerful tests for possible performance related biases. Our samples are therefore likely to have more extreme performance than that occurring in specific earnings management studies. Thus, we expect the performance related misspecification to be more severe in our extreme decile samples. In additional tests (not reported) we confirm that the performance induced misspecifications are not limited to the extreme deciles

(2) *Revenue Manipulation* - premature recognition of revenue (assuming all costs are fixed). This approach is implemented by adding the assumed amount of revenue manipulation to total accruals, revenue and accounts receivable. The same amount is subtracted from total accruals, revenue and accounts receivable in the following year; and

(3) *Margin Manipulation* - premature recognition of revenue (assuming all costs are variable). This approach is implemented by adding the assumed amount of margin manipulation to total accruals and by adding the following to revenue and accounts receivable:

(assumed amount of margin manipulation) / (net income ratio),

where the net income ratio is the ratio of the firm's net income to revenue, estimated using the median value of the ratio from observations in each firm's estimation period. For example, to artificially introduce earnings management of one percent of lagged assets in a firm with a net income ratio of ten percent, we add one percent of lagged assets to total accruals and ten percent of lagged assets to revenue and accounts receivable. The same amounts are subtracted from total accruals, revenue and accounts receivable in the following year.

The difference between assumptions (2) and (3) relate to the matching of expenses to manipulated revenues. Assumption (2) corresponds to 'pure' revenue manipulation, in which revenues are manipulated upwards, but expenses do not change. Assumption (3) corresponds to premature recognition of a sale in a setting where all costs are variable. Revenues are manipulated upwards, but expenses are matched to the manipulated revenues. The crucial difference between (2) and (3) is that (3) requires much greater revenue manipulation in order to achieve a given increase in earnings. Assumptions (2) and (3) are extremes on a continuum, and in practice, we would expect most revenue-based earnings management to lie between these two extremes.

Interpretation of the type II errors for sample (iii) is contingent on the explicit assumptions that are made concerning how earnings are managed. In order to reinforce the external validity of our conclusions concerning type II errors, we examine a sample of firm-years for which we have strong a priori reasons to expect earnings management of a known sign. Sample (iv) consists of firm-years that are subject to accounting-based enforcement actions by the SEC. The SEC takes enforcement actions against firms and individuals having allegedly violated the financial reporting requirements of the securities laws. Since April 1982, the SEC has published the details of its major enforcement actions in a series of Accounting and Auditing Enforcement Releases (AAERs).[8]

Enforcement actions in which the Commission alleges that a firm has overstated annual earnings in violation of Generally Accepted Accounting Principles (GAAP) are brought pursuant to Section 13(a).[9] A total of 134 firms are the subject of AAERs brought pursuant to Section 13(a). We further require that (i) each firm has at least ten years of the required financial statement data on the COMPUSTAT industrial files (excluding the years in which the alleged overstatements of earnings occurred); (ii) the AAER alleges that annual earnings have been overstated (many of the AAERs relate to overstatements of quarterly earnings that are reversed before the fiscal year end); and (iii) the AAER does not relate to a financial institution (since the current asset and current liability variables that we use to compute accruals are not available for these firms). These

[8] Feroz et al. (1991) provide descriptive evidence on the AAERs and their financial and market effects Pincus et al. (1988) describe the events leading to a formal SEC investigation and the publication of an AAER.

[9] Section 13(a) requires issuers whose securities are registered with the Commission to file reports (including the annual financial statements on form 10-K) as specified by Commission rules and regulations. The financial statements contained in these filings are required to comply with Regulation S-X, which in turn requires conformity with GAAP.

restrictions result in a sample of 32 firms that are alleged to have overstated earnings in a total of 56 firm-years. Fifteen of the sample firms are targeted for overstating revenue alone, three are targeted for overstating revenue in combination with understating expenses and the remaining 14 firms are alleged to have understated a variety of expenses.

Data Analysis

The empirical tests for earnings management follow from the regression framework developed in section II. The empirical tests are separately applied to each of the samples described above. The firm-years in each sample represent the event-years that are to be tested for earnings management. We therefore begin by matching each firm-year represented in a sample with the remaining non-event-years for that firm on COMPUSTAT to form the estimation period. The sample selection procedures ensure that all firms have at least ten observations in their estimation period.

Consistent with previous studies of earnings management (Healy 1985 and Jones 1991), total accruals (TA), are computed as:[10]

$$TA_t = (\Delta CA_t - \Delta CL_t - \Delta Cash_t + \Delta STD_t - Dep_t)/(A_{t-1}),$$

where

ΔCA	= change in current assets (COMPUSTAT item 4);
ΔCL	= change in current liabilities (COMPUSTAT item 5);
ΔCash	= change in cash and cash equivalents (COMPUSTAT item 1);
ΔSTD	= change in debt included in current liabilities (COMPUSTAT item 34);
Dep	= depreciation and amortization expense (COMPUSTAT item 14); and
A	= Total Assets (COMPUSTAT item 6).

Earnings is measured using net income before extraordinary items and discontinued operations (COMPUSTAT item 18) and is also standardized by lagged total assets. Cash from operations is computed as:

$$\text{Cash from operations} = \text{Earnings} - \text{TA}.$$

Using each of the competing models, discretionary accruals are then estimated by subtracting the predicted level of nondiscretionary accruals (NDAP) from total accruals (standardized by lagged total assets):

$$DAP_{it} = TA_{it} - NDAP_{it}. \tag{10}$$

To test for earnings management, the estimated discretionary accruals are regressed on the partitioning variable, PART. Recall that the regression pools across observations in the event period and the estimation period. PART is set equal to one if the observation is from the event period and zero if the observation is from the estimation period:

$$DAP_{it} = \hat{a}_i + \hat{b}_i\, PART_{it} + e_{it}. \tag{11}$$

The coefficient on PART, $\hat{b}_i$, provides a point estimate of the magnitude of the earnings management attributable to the stimulus represented by PART. The null hypothesis of no earnings management in response to this factor is tested by applying a t-test to the null hypothesis

[10] All data required to estimate the nondiscretionary accruals models and conduct the empirical analysis are initially obtained from the COMPUSTAT industrial files. Data for the 56 event-years in the SEC sample are manually checked to hard copies of the sample firms' annual reports In some of the cases where the SEC requires a firm to restate its earnings, we found that the COMPUSTAT files contained the restated numbers. In these cases, we substitute the original figures reported in the hard copies of the annual reports.

that $b_i = 0$.[11] The null hypothesis that the average t-statistic is zero for the N firms in the sample is also tested by aggregating the individual t-statistics to form a Z-statistic:

$$Z = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} \frac{t_j}{\sqrt{k_j/(k_j-2)}}, \tag{12}$$

where

t_j = t-statistic for firm j; and
k_j = degrees of freedom for t-statistic of firm j.

The Z-statistic is asymptotically distributed unit normal if the t_j's are cross-sectionally independent.

V. EMPIRICAL RESULTS

Random Sample of Firm-Years

Table 1 provides descriptive statistics on the parameter estimates and test statistics generated by each of the discretionary accrual models when applied to the sample of 1000 randomly selected firm-years. For each model, the row labeled "earnings management" represents the estimated coefficient on PART, ($\hat{b}_i$), the row labeled "standard error" represents the standard error of this coefficient estimate, and the row labeled "t-statistic" represents the t-statistic for testing the null hypothesis that this coefficient is equal to zero. The mean and median values of earnings management are close to zero for all models indicating, as expected, that there is no systematic evidence of earnings management in the randomly selected event-years relative to years in the estimation period. The standard errors tend to be highest for the DeAngelo Model and lowest for the Jones and Modified Jones models, suggesting that the latter models are more effective at modeling the time-series process generating nondiscretionary accruals and suffer less from misspecifications caused by omitted determinants of nondiscretionary accruals. Note, however, that from a researcher's perspective, the standard errors are high in all models. For example, the mean standard error exceeds 0.09 for all models. Earnings management would therefore have to exceed 18 percent of lagged assets before we would expect to generate a t-statistic greater than two for an individual firm. Alternatively, if a Z-statistic were computed for a sample of firms that had all managed earnings by one percent of total assets, over 300 firms would be required in the sample before the Z-statistic is expected to exceed two. Thus, none of the models are expected to produce powerful tests for earnings management of economically plausible magnitudes.

Table 2 reports the incidence of type I errors for the sample of 1000 randomly selected firm-years using the conventional test levels of five percent and one percent. Since the earnings management partitioning variable is selected at random in this sample, it is expected to be uncorrelated with any omitted variables. Thus, the type I errors should correspond to the test levels applied, so long as the Gaussian assumptions are satisfied. Type I errors are reported for both the null hypothesis that discretionary accruals are less than or equal to zero and the null hypothesis

[11] The computation of the standard error of $\hat{b}_i$ requires special attention because the measures of discretionary accruals in the event period (estimation period) are prediction errors (fitted residuals) from a first-pass estimation process. An adjustment must therefore be made to reflect the fact that the standard errors of the prediction errors are greater than the standard errors of the fitted residuals. Likewise, the degrees of freedom in the t-test must reflect the degrees of freedom used up in the first-pass estimation. This can be accomplished by either explicitly adjusting the standard error and degrees of freedom of the prediction errors (see Jones 1991) or by estimating a single stage regression that includes both PART and the determinants of nondiscretionary accruals (see Dechow and Sloan 1991). The two approaches are econometrically equivalent and we therefore use the latter approach for its computational ease (see Salkever 1976 for an extended discussion and proof on this issue).

TABLE 1
Results of Tests for Earnings Management Using Alternative Models to Measure Discretionary Accruals. The Results are Based on a Sample of 1000 Randomly Selected Firm-Years.

Model	*Mean*	*Standard Deviation*	*Lower Quartile*	*Median*	*Upper Quartile*
Healy Model:					
earnings management	0.002	1.241	–0.035	–0.001	0.040
standard error	0.195	4.573	0.039	0.065	0.104
t-statistic	0.012	1.174	–0.583	0.010	0.598
DeAngelo Model:					
earnings management	0.002	0.151	–0.048	0.001	0.052
standard error	0.281	6.799	0.054	0.090	0.143
t-statistic	0.002	1.135	–0.577	0.018	0.637
Jones Model:					
earnings management	0.001	0.118	–0.037	–0.001	0.036
standard error	0.092	0.438	0.036	0.060	0.095
t-statistic	0.013	1.155	–0.647	–0.022	0.644
Modified Jones Model:					
earnings management	0.002	0.119	–0.035	0.001	0.041
standard error	0.092	0.437	0.036	0.060	0 095
t-statistic	0.062	1.204	–0.613	0.027	0.745
Industry Model:					
earnings management	0.002	0.662	–0.032	0.000	0 039
standard error	0.211	5.363	0.038	0.063	0.101
t-statistic	0.028	1.165	–0.555	0.006	0.637

Notes:
Earnings management represents the estimated coefficient on PART, ($\hat{b}_i$), from firm-specific regressions of $DAP_{it} = \hat{a}_i + \hat{b}_i\ PART_{it} + e_{it}$; where DAP is the measure of discretionary accruals produced by each of the models and PART is an indicator variable equal to 1 in a year in which earnings management is hypothesized to occur in response to the stimulus identified by the researcher and 0 otherwise. Standard error is the standard error of the coefficient on PART for each of the regressions and t-statistic is the t-statistic testing the null hypothesis that the coefficient on PART is equal to zero.

that discretionary accruals are greater than or equal to zero. A binomial test is also conducted to assess whether the empirical rejection frequencies are significantly different from the specified test levels. The empirical rejection frequencies are close to the specified test levels for all models, and none of the differences are significant at conventional levels. Thus, all models appear well specified for a random sample of firm-years.

Samples of Firm-Years Experiencing Extreme Financial Performance

This section considers the four samples of firm-years experiencing extreme financial performance. The first two samples exhibit high and low earnings performance, respectively. Figure 1 contains plots in event time of earnings and its components for each of the two samples.

TABLE 2
Comparison of the Type I Errors for Tests of Earnings Management Based on Alternative Models to Measure Discretionary Accruals. Percentage of 1000 Randomly Selected Firm-Years for which the Null Hypothesis of No Earnings Management is Rejected (One-Tailed Tests).

Null Hypothesis	*Earnings management ≤ 0*		*Earnings management ≥ 0*	
Test Level:	*5%*	*1%*	*5%*	*1%*
Healy Model:				
t-test	5.0%	1.3%	5.1%	1.4%
DeAngelo Model:				
t-test	4.8	1.0	5.2	1.1
Jones Model:				
t-test	4.9	1.4	5.9	1.5
Modified Jones Model:				
t-test	4.9	1.3	5.9	1.3
Industry Model:				
t-test	4.2	1 4	5.5	1.2

* Significantly different from the specified test level at the 5 percent level using a two-tailed binomial test.
** Significantly different from the specified test level at the 1 percent level using a two-tailed binomial test

Year 0 represents the year in which the firm-years are selected based on their extreme earnings performance. There are separate plots for total accruals, cash from operations and earnings. Each of the variables is scaled by lagged total assets and the median values for each of the two samples are shown in the plots. The bottom plot is of earnings performance. As expected, the high earnings performance sample gradually increases to a peak in year 0 and then declines thereafter. Similarly, the low earnings performance sample gradually declines to a trough in year 0 and then increases thereafter. The total accruals and cash from operations plots mirror the earnings plots, though the peaks and troughs are less extreme. This reflects the fact that earnings is the sum of cash from operations and accruals. Firms with high earnings tend to have high cash flows and high accruals. Similarly, firms with low earnings tend to have low cash flows and low accruals.

Table 3 reports the rejection frequencies for tests of earnings management in response to the stimulus represented by PART. Since PART is measured by randomly selecting firms with extreme earnings performance, PART is constructed so that it is not itself *a causal determinant* of earnings management (although it may be imperfectly correlated with causal determinants). Thus, we have constructed a scenario which is analogous to the case where a researcher has selected a stimulus that is correlated with firm performance, but *where the stimulus is not itself a causal determinant* of earnings management. As such, any rejections of the null hypothesis of no earnings management represent type I errors. However, these results do not permit a direct assessment of the extent of misspecification in existing studies. Such an assessment requires a detailed reexamination of the stimulus in question [e.g., Holthausen et al. 1995].

FIGURE 1

Time Series of Median Annual Total Accruals, Cash from Operations and Earnings all Standardized by Lagged Total Assets. Year 0 is the Year in which Firm-Years are Selected from the Lowest and Highest Decile of Earnings Performance. Sample Consists of 1000 Firm-Years Randomly Selected from Firm-Years in the Lowest and Highest Decile of Earnings Performance.

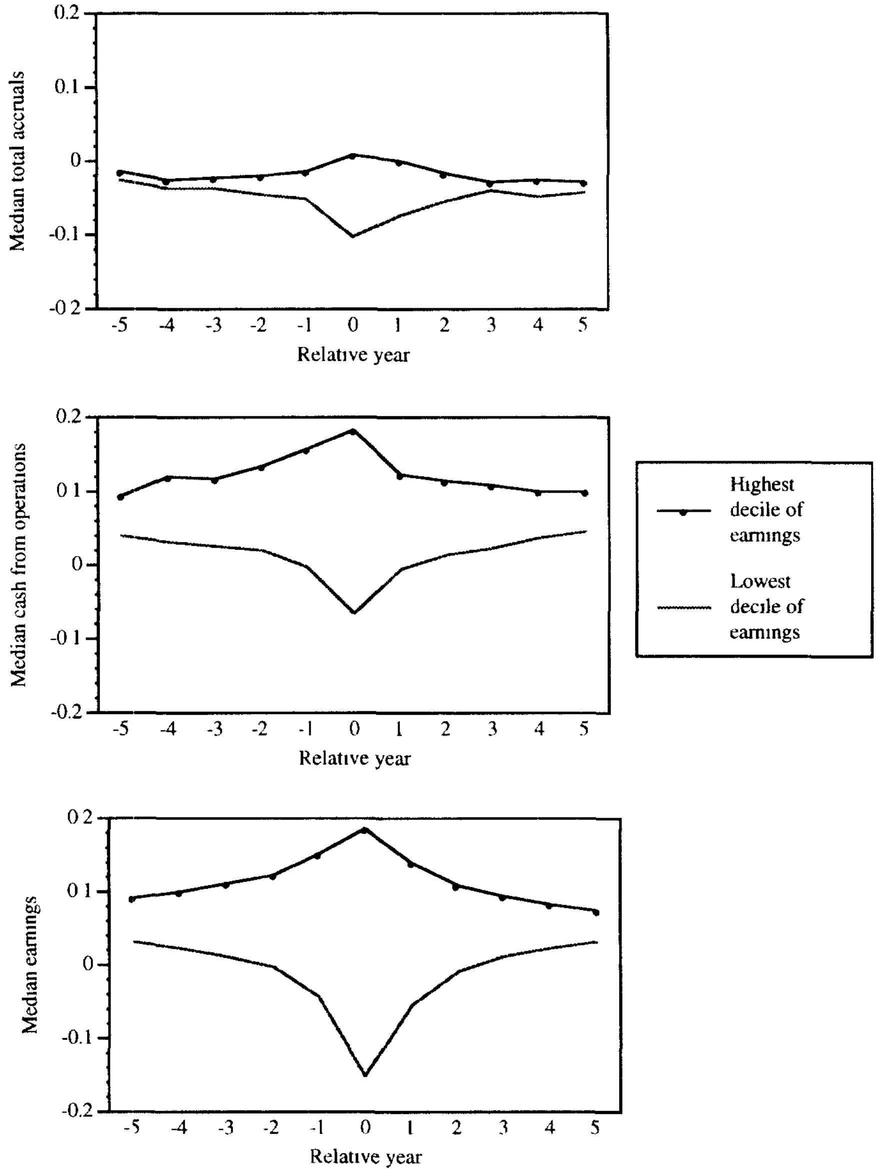


TABLE 3

Comparison of the Type I Errors for Tests of Earnings Management Based on Alternative Models to Measure Discretionary Accruals. Percentage of 1000 Firm-Years Randomly Selected from Firm-Years in the Lowest Decile and Highest Decile of Earnings Performance for which the Null Hypothesis of No Earnings Management is Rejected (One-Tailed Tests).

Null Hypothesis	*Earnings management ≤ 0*		*Earnings management ≥ 0*	
Test Level:	*5%*	*1%*	*5%*	*1%*
Panel A: Lowest decile of earnings performance				
Healy Model:				
t-test	1.7%**	0 4%*	25.9%**	9 9%**
DeAngelo Model:				
t-test	2.8**	0 4*	13.5**	3.3**
Jones Model:				
t-test	2.7**	0 7	16.6**	5 4**
Modified Jones Model:				
t-test	2.7**	0.6	17.6**	6.5**
Industry Model:				
t-test	1.7**	0 4*	22.6**	8.4**
Panel B: Highest decile of earnings performance				
Healy Model:				
t-test	12.8%**	4 2%**	4.4%	1.4%
DeAngelo Model:				
t-test	9.5**	1.6*	4.2	0.9
Jones Model:				
t-test	6.5*	1.3	6.3	1.4
Modified Jones Model:				
t-test	7.6**	1.8*	5.5	1.4
Industry Model:				
t-test	10.3**	3 2**	4.4	1.5

* Significantly different from the specified test level at the 5 percent level using a two-tailed binomial test.
** Significantly different from the specified test level at the 1 percent level using a two-tailed binomial test

Panel A of table 3 reports the results for the low earnings performance sample. The proportion of type I errors for tests of the null hypothesis that earnings management ≤ 0 are all less than the specified test levels and many of the differences are statistically significant. Conversely, the proportion of type I errors for tests of the null hypothesis that earnings management ≥ 0 are appreciably greater than the corresponding test levels and the differences are statistically significant in all cases. For example, using a test level of five percent results in rejection rates ranging from 13.5% for the DeAngelo Model to 25.9% for the Healy Model. The high rejection rates arise because firm-years with low earnings also tend to have low total accruals and all the models attribute part of the lower accruals to negative discretionary accruals. Thus, the null hypothesis that earnings are not managed in response to the stimulus represented by PART is rejected in favor of the alternative that earnings are managed downwards.

Panel B of table 3 reports rejection frequencies for the sample of firm-years selected on the basis of high earnings performance. In this case, the results are opposite to those for the low earnings performance sample. The null hypothesis that earnings management ≥ 0 is rejected at rates similar to those reported for the random sample in table 2. However, the null hypothesis that earnings management ≤ 0 is rejected at rates that are appreciably greater than the specified test levels and the differences are statistically significant in nearly all cases. For example, the test level of five percent yields rejection rates ranging from 6.5% for the Jones Model to 12.8% for the Healy Model. This reflects the fact that firm-years with high earnings tend to have high accruals and the models of nondiscretionary accruals do not completely extract the higher accruals. In both panels A and B, the misspecifications are less severe for the Jones and Modified Jones models than for the Healy Model. This is consistent with part of the systematic behavior in accruals being extracted by these more sophisticated models.

The results reported in panels A and B of table 3 are open to two interpretations (see the discussion of problem 1 in section II): (i) Earnings performance is correlated with the error in measuring discretionary accruals (i.e., earnings performance is correlated with nondiscretionary accruals that are not completely extracted by any of the models); and/or (ii) earnings performance is correlated with *other variables* that cause earnings to be managed. If a researcher selects a stimulus that *does not* cause earnings to be managed but is correlated with earnings performance, then the tests for earnings management will generate excessive type I errors. That is, using the models evaluated here, the researcher will detect low discretionary accruals when earnings are low and high discretionary accruals when earnings are high, even if the cause of the earnings management is not the stimulus investigated by the researcher

The evidence in table 3 suggests that before attributing causation to the investigated stimulus, the researcher should ensure that the results are not induced by omitted variables correlated with earnings performance. Holthausen et al (1995) illustrate this point in their extension of Healy's (1985) paper on executive bonus plans. They conclude that Healy's lower bound results are induced by the correlation between his partitioning variable and earnings performance and that Healy prematurely attributes the earnings management to bonus plans. We provide further discussion of this problem in section VI.

The second two samples of firm-years are selected on the basis of high cash from operations and low cash from operations performance, respectively. Event time plots for these two samples of firms are provided in figure 2. The middle plot is of cash from operations. As expected, the high cash from operations sample climbs to a peak in year 0 and declines thereafter. The low cash from operations sample exhibits the opposite behavior, falling to a trough in year 0 and improving thereafter. The bottom plot is of earnings, which follow a similar, though less pronounced pattern to cash from operations. The top plot is of total accruals and is markedly different from the other two plots. In every year except for the event-year, total accruals are very similar for the two samples. In the event-year, the *low* cash from operations firms experience a sharp *increase* in total accruals, while the *high* cash from operations firms experience a sharp *decrease* in total accruals. The event-year accrual changes are opposite in sign, but about half as large as the corresponding changes in cash from operations. These results are consistent with the findings of Dechow (1994), who hypothesizes that this negative correlation results from the application of the matching principle under accrual accounting. Dechow's evidence suggests that the event-year accrual changes represent nondiscretionary accruals that are made with the objective of eliminating temporary mismatching problems in cash from operations. If matching is the cause of the negative correlation, then a well specified model of nondiscretionary accruals should control for this effect. However, the results in table 4 indicate that existing models do not completely control for this negative correlation.

FIGURE 2

Time Series of Median Annual Total Accruals, Cash from Operations and Earnings all Standardized by Lagged Total Assets. Year 0 is the Year in which Firm-Years are Selected from the Lowest and Highest Decile of Cash from Operations. Sample Consists of 1000 Firm-Years Randomly Selected from Firm-Years in the Lowest and Highest Decile of Cash from Operations.

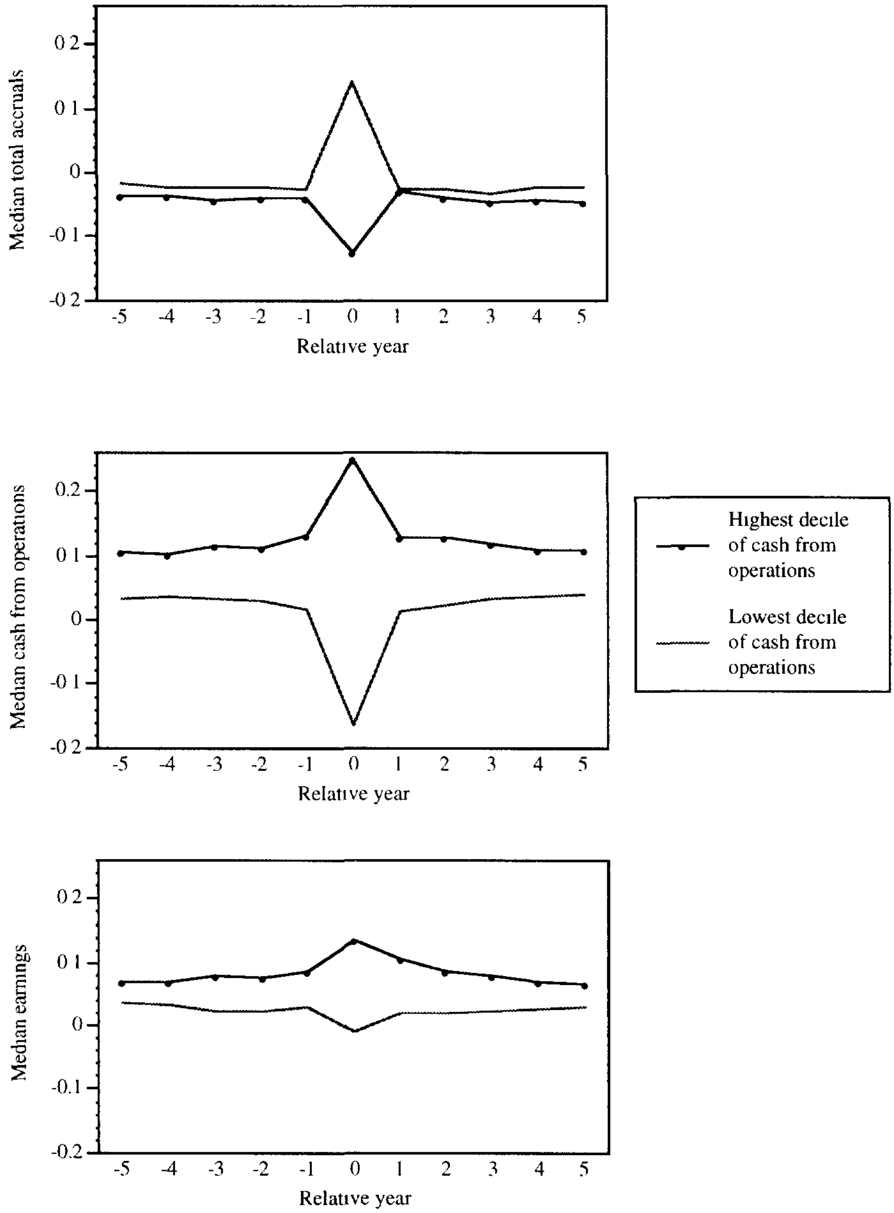


Table 4 reports the proportion of type I errors for the high and low cash from operations samples. Panel A indicates that the low cash from operations sample generates type I errors that are all significantly greater than the specified test levels for the null hypothesis that earnings management ≤ 0. For example, at the five percent test level the rejection frequencies range from a low of 32.9% for the DeAngelo Model to a high of 46.7% for the Healy Model. This stems from the regularity documented in figure 2 that firms with low cash from operations tend to have high total accruals. The opposite problem is observed when testing the null hypothesis that earnings management ≥ 0. Because total accruals tend to be high, discretionary accruals generated by the various models tend to be high, and the frequency of type I errors tend to be lower than the specified test levels.

Panel B of table 4 reports results for the high cash from operations sample. Recall that the high cash from operations sample has low total accruals in event year 0. The results for this sample indicate that the null hypothesis that earnings management ≤ 0 tends to be under-rejected relative to the specified test levels, while the null hypothesis that earnings management ≥ 0 tends to be over-rejected. The over-rejections are most serious for the Healy Model, 50.0%. These results illustrate the problem faced by Healy (1985) in his upper bound tests. Healy hypothesizes that the executives of firms in which cash from operations exceeds the upper bounds specified in their top executive bonus plans manage earnings downwards. However, panel B illustrates that estimated discretionary accruals generally tend to be low for firms with high cash flows. The upper bound results reported in Healy's table 2 are therefore likely to overstate the amount of earnings management that takes place at the upper bound. Healy recognizes this potential problem and controls for it through the use of a control sample in his table 4 results.

More generally, any earnings management study in which the stimulus under investigation is correlated with cash flow performance is likely to produce misspecified tests. For example, Gaver et al. (1995) replicate Healy's lower bound results using nondiscretionary earnings to classify firms relative to the lower bounds specified in their executive bonus plans. Gaver et al. measure nondiscretionary earnings as the sum of cash from operations plus nondiscretionary accruals, as generated by the Jones model. The resulting measure of nondiscretionary earnings is highly positively correlated with cash from operations (the mean Pearson correlation exceeds 0.8). Thus, their tests are likely to suffer from the misspecification demonstrated in panel A of table 4.[12] In particular, the lower bound sample is biased toward rejecting the null hypothesis that discretionary accruals are less than or equal to zero in favor of the alternative hypothesis that accruals are managed upwards. This result is documented by Gaver et al. and attributed to managerial "smoothing" of earnings.

Samples of Firm-Years with Artificially Induced Earnings Management

The results of the simulations using artificially induced earnings management are summarized in figures 3 and 4. Figure 3 provides information concerning bias in the estimates of earnings management produced by the competing models. For the sake of parsimony, we provide plots for only three models: the Healy Model; the Jones Model; and the Modified Jones Model. The results for the DeAngelo and Industry models are indistinguishable to those documented for the Healy and Modified Jones models. For each model and for each assumed source of earnings manipu-

[12] In additional tests (not reported) we reestimated the table 4 results using the Gaver et al. (1995) measure of nondiscretionary earnings in place of cash from operations. The results confirm that the low nondiscretionary earnings sample over-rejects the null hypothesis that discretionary accruals are less than or equal to zero in favor of the alternative hypothesis that they are greater than zero. For example, the Jones model (which is used by Gaver et al.) rejects the null hypothesis that earnings management is less than or equal to zero 37 1% (17 4%) of the time using a five percent (one percent) test level.

TABLE 4

Comparison of the Type I Errors for Tests of Earnings Management Based on Alternative Models to Measure Discretionary Accruals. Percentage of 1000 Firm-Years Randomly Selected from Firm-Years in the Lowest and Highest Decile of Cash From Operations Performance for which the Null Hypothesis of No Earnings Management is Rejected (One-Tailed Tests).

Null Hypothesis	*Earnings management ≤ 0*		*Earnings management ≥ 0*	
Test Level:	*5%*	*1%*	*5%*	*1%*
Panel A: Lowest decile of cash from operations performance				
Healy Model:				
t-test	46.7%**	24.1%**	1.2%**	0.3%**
DeAngelo Model:				
t-test	32.9**	12.4**	1.0**	0.2**
Jones Model:				
t-test	42.9**	19.2**	1.2**	0.5
Modified Jones Model:				
t-test	44.5**	21.7**	1 1**	0.5
Industry Model:				
t-test	45.0**	22.4**	1.2**	0.2**
Panel B: Highest decile of cash from operations performance				
Healy Model:				
t-test	0.0%**	0.0%**	50.0%**	23.9%**
DeAngelo Model:				
t-test	0.5**	0.1**	32.6**	12.4**
Jones Model:				
t-test	0.3**	0.1**	46.2**	19.9**
Modified Jones Model:				
t-test	0.3**	0.1**	46.4**	20.3**
Industry Model:				
t-test	0.2**	0.0**	46.7**	21.9**

* Significantly different from the specified test level at the 5 percent level using a two-tailed binomial test.
** Significantly different from the specified test level at the 1 percent level using a two-tailed binomial test.

lation, we provide a plot of detected earnings management (vertical axis) against induced earnings management (horizontal axis). Since our simulations are based on a large number of independent observations, an unbiased estimator is expected to result in a 45 degree line (i.e., detected earnings management is expected to equal induced earnings management). In each graph, the thin line represents the 45 degree line that would be generated by an unbiased estimator, and the thick line represents the results of our simulations.

The first column of graphs provides results for artificially induced expense manipulation. The thick line lies atop the thin line in all cases, indicating that all models provide unbiased tests of expensed-based earnings management. The second column of graphs provides results for

FIGURE 3

Simulation Results from Tests for Bias in Estimates of Earnings Management Based on Alternative Models for Measuring Discretionary Accruals. Simulations are Conducted for Artificially Induced Amounts of Earnings Management Ranging from Zero Percent to 100 Percent of Total Assets. Each Simulation Uses 1000 Firm-Years. The Thin Line Represents the Hypothetical Results for an Unbiased Estimator. The Thick Line Represents the Simulation Result.

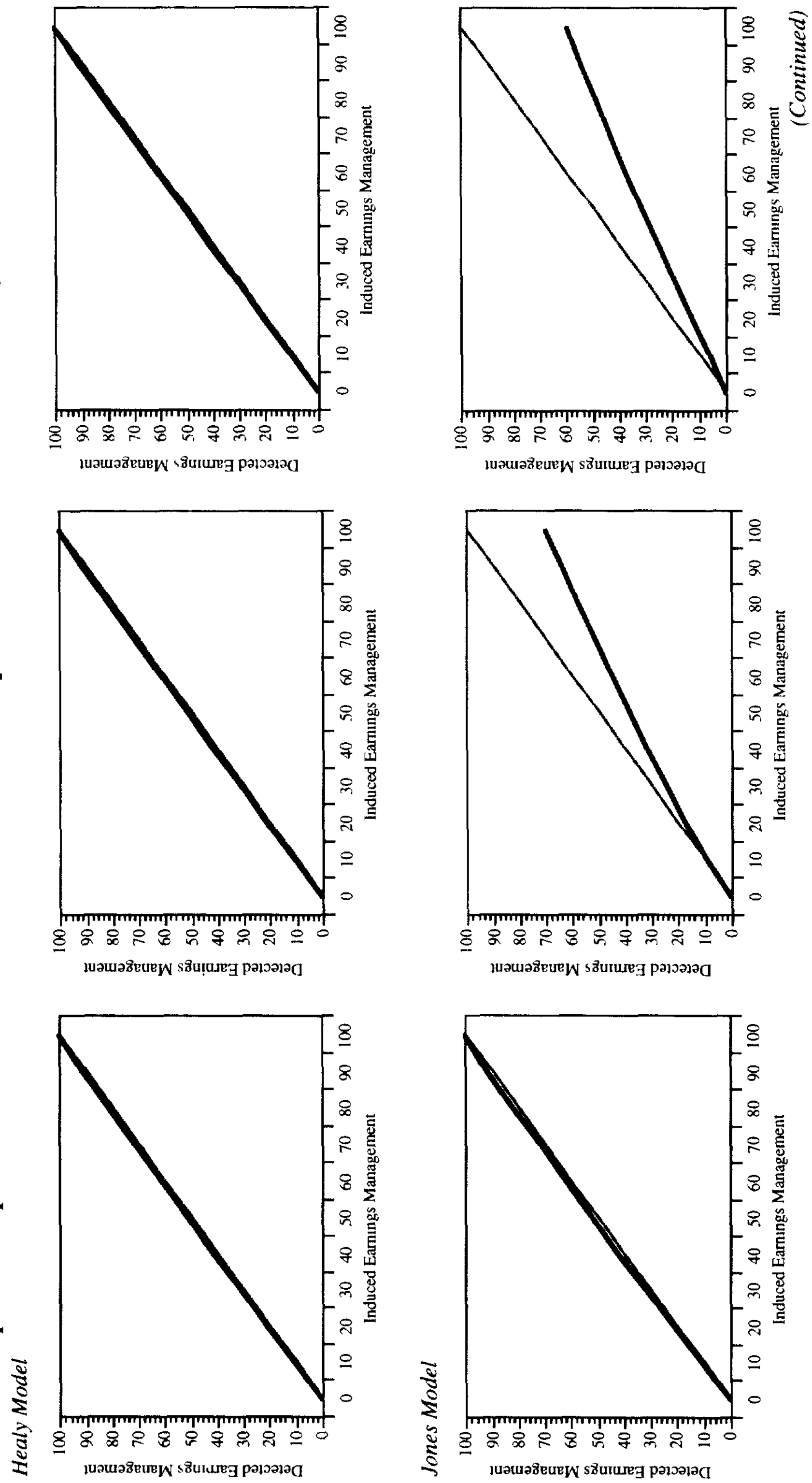
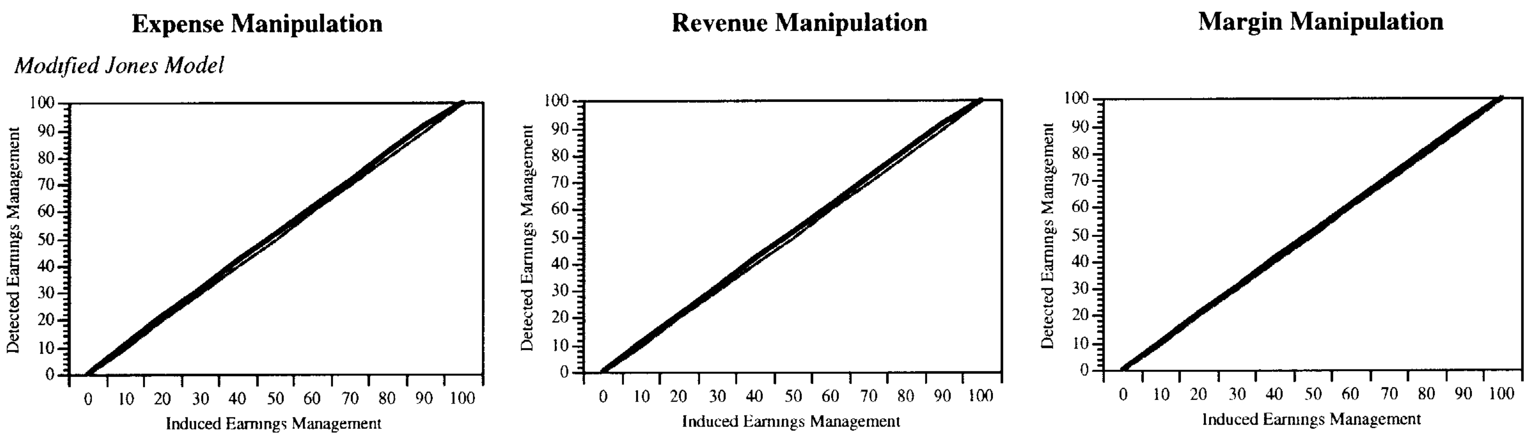


FIGURE 3 *(Continued)*

Simulation Results from Tests for Bias in Estimates of Earnings Management Based on Alternative Models for Measuring Discretionary Accruals. Simulations are Conducted for Artificially Induced Amounts of Earnings Management Ranging from Zero Percent to 100 Percent of Total Assets. Each Simulation Uses 1000 Firm-Years. The Thin Line Represents the Hypothetical Results for an Unbiased Estimator. The Thick Line Represents the Simulation Result.

Modified Jones Model



artificially induced revenue manipulation. It is evident that the estimates of earnings management provided by the Jones Model are biased downward. The change in revenue is used as an independent variable to extract nondiscretionary accruals in the Jones Model, thereby extracting part of the revenue-based earnings management. The magnitude of the bias indicates that approximately one-quarter of the induced earnings management is not detected. The Modified Jones Model does not suffer from this bias. The third and final column presents the results for artificially induced margin manipulation. Again, only the Jones Model produces biased estimates of discretionary accruals. The downward bias is approximately one-third of the induced earnings management and is more serious than for the case of revenue manipulation because margin manipulation requires a larger amount of revenue management for a given amount of earnings management.

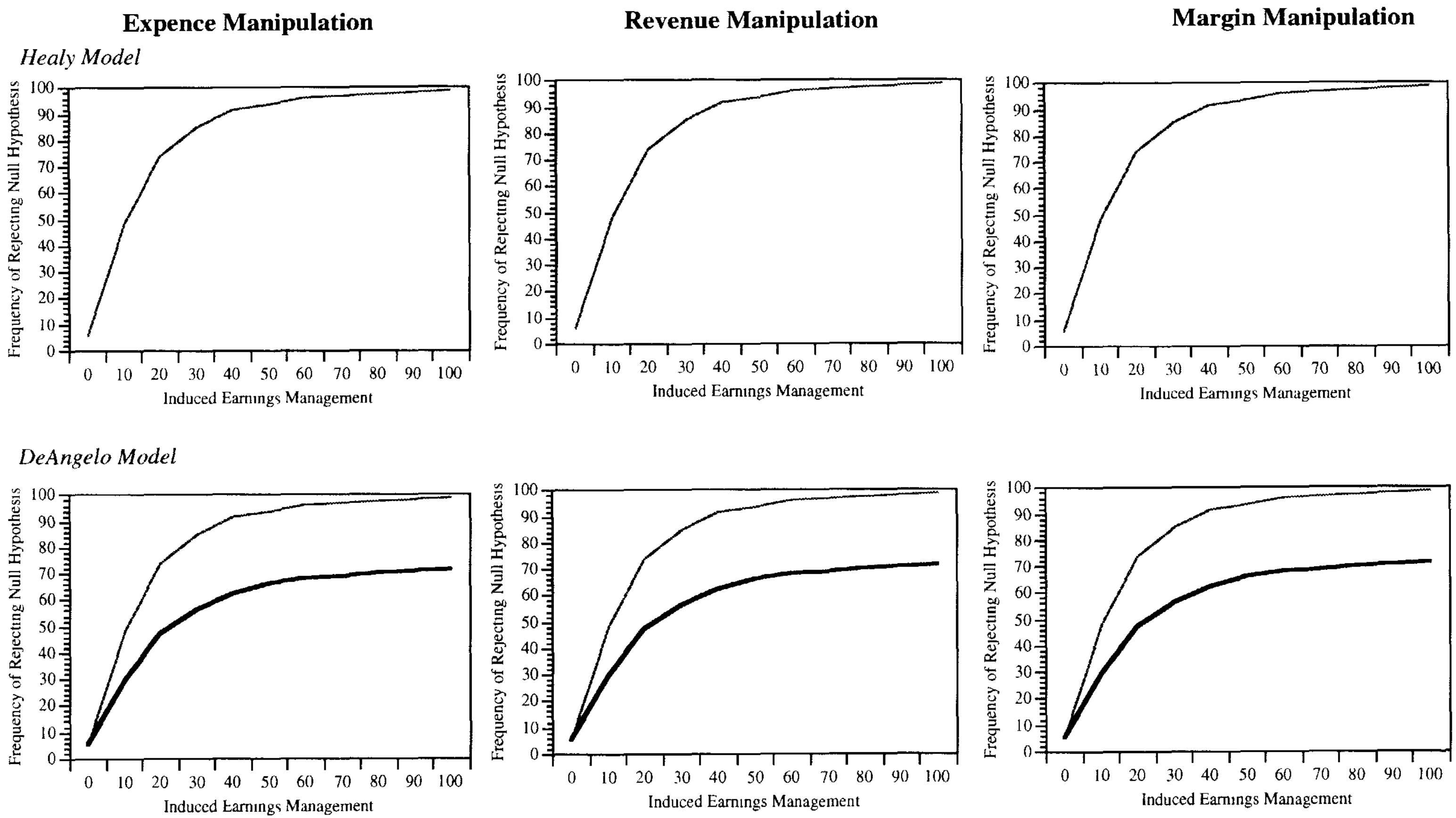
Figure 4 provides information concerning the relative power of the alternative models for detecting earnings management. These graphs plot the frequency with which the null hypothesis of no earnings management is rejected (vertical axis) against the magnitude of the induced earnings management (horizontal axis). A separate graph is provided for each model and for each assumed source of earnings manipulation. All rejection rates are computed at the five percent level using a one-tailed test.[13] The first graph reports the power function for the Healy Model (thin line). Healy's power function is also provided in the graphs of the remaining models to provide a benchmark for evaluating their relative power. The power functions for the remaining models are presented using the thicker lines.

The first column of graphs provide the power functions for expense manipulation. The DeAngelo Model lies substantially below the Healy Model because the standard errors of the estimate of earnings management (table 1) tend to be significantly higher for the DeAngelo Model. The Jones, Modified Jones and Industry models are all slightly more powerful than the Healy Model. Again, this arises because they have slightly lower standard errors. Though it is not readily apparent from the graphs, the Jones and Modified Jones models are more powerful than the Healy and Industry models for all levels of induced earnings management. The second column of graphs provides results for revenue-based earnings management. The only significant change from the preceding column is that the power function for the Jones Model now lies below that of the Healy Model, due to the bias results in figure 3. The Jones Model unintentionally extracts some of the revenue-based earnings management leading to a downwardly biased estimate of earnings management and correspondingly reducing the power of the test. The Modified Jones Model continues to dominate the other models. The third and final column provides the results for margin-based earnings management. The only significant change in this column is that the power of the Jones Model drops even further due to the downwardly biased estimate of earnings management. The Modified Jones Model still dominates all the other models, although it only dominates the Industry Model by a small margin. It should, however, be noted that the odds are stacked in favor of the Industry Model. We have implicitly assumed that earnings management is not clustered by industry (i.e., when we induce earnings management in a firm-year, we do not induce earnings management in the industry matched firm-years). To the extent that this assumption is violated, the power of the tests based on the Industry Model are overstated in our simulations.

[13] We replicated the results using a one percent test level. The relative rankings of the models are identical. We also performed identical tests assuming accruals are downwardly managed. The tenor of the bias and power results is unchanged.

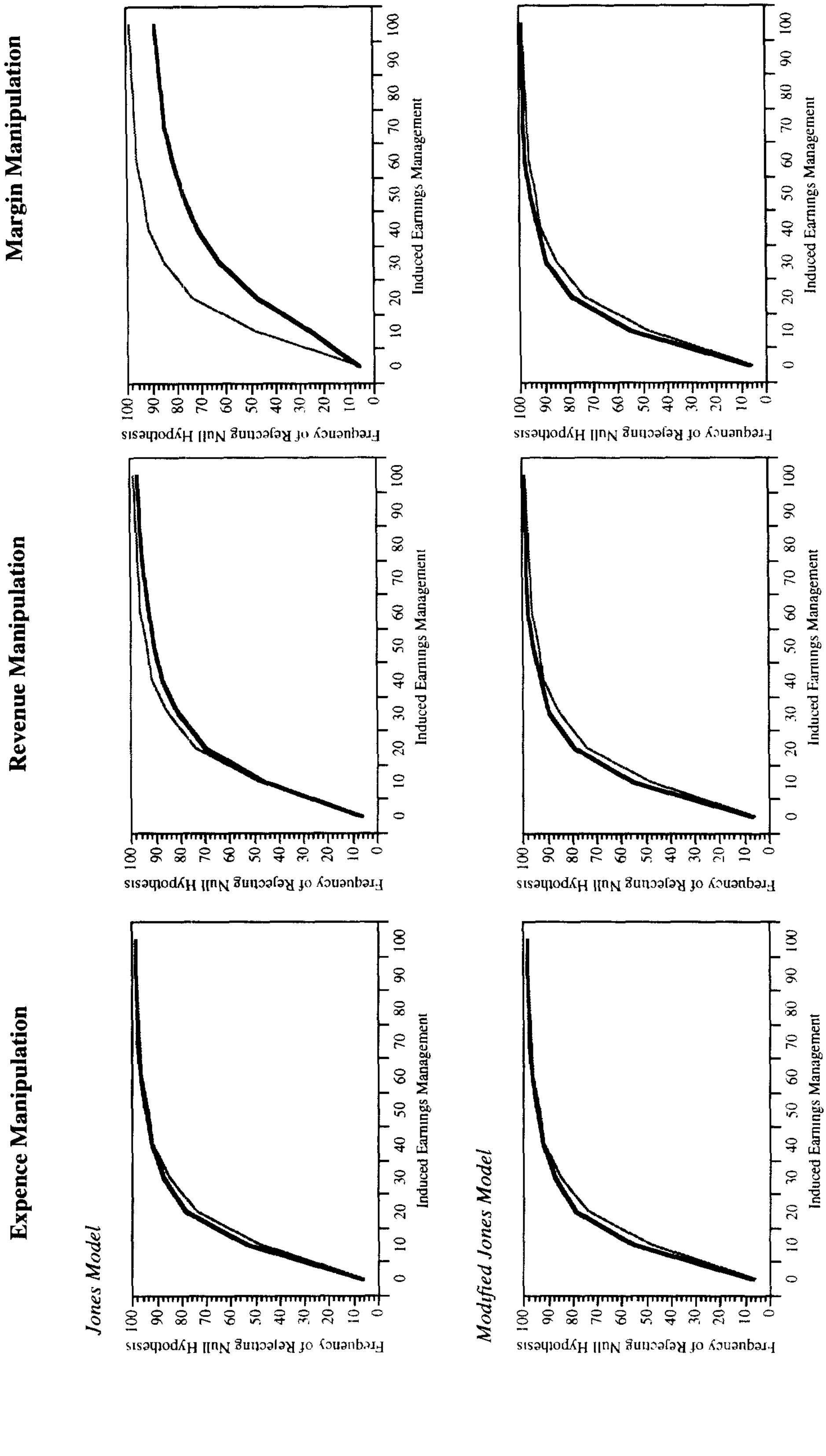
FIGURE 4

Simulation Results of Power Functions for Tests of Earnings Management Based on Alternative Models for Measuring Discretionary Accruals. Simulations are Conducted for Artificially Induced Amounts of Earnings Management Ranging from Zero Percent To 100 Percent and use a One-Sided Test Level of Five Percent. Each Simulation uses 1000 Firm-Years. The Thin Line Represents the Power Function for the Healy Model (Benchmark). The Thick Line Represents the Simulation Result for Each of the Other Models.



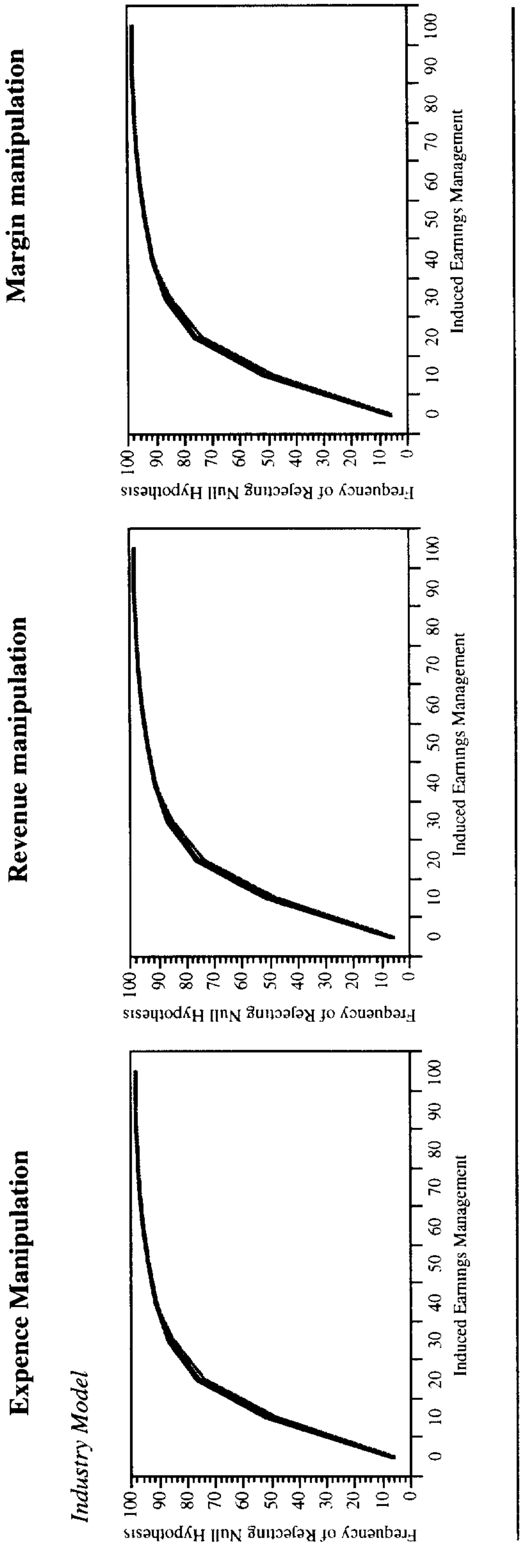
(Continued)

FIGURE 4 (*Continued*)



(*Continued*)

FIGURE 4 (*Continued*)



Sample of Firm-Years in which the SEC Alleges Earnings are Overstated

Figure 5 provides event time plots of total accruals, cash from operations and earnings for the sample of 32 firms alleged by the SEC to have overstated earnings. Year 0 represents the year in which the SEC alleges that earnings are overstated.[14] To provide a benchmark for comparison, plots are also provided for the sample of 1000 randomly selected event-years. The plot of median total accruals indicates that accruals are abnormally high in the years leading up to and including year 0 and are abnormally low thereafter. The fact that total accruals are higher for the SEC sample relative to the random sample in event-year 0 is consistent with the joint hypothesis that total accruals measure discretionary accruals and that discretionary accruals are positive. The plot also reveals a sharp decline in accruals in event year one, which is consistent with the managed accruals reversing.

The cash from operations plot indicates that cash flows tend to be slightly lower than normal for the SEC sample. The earnings plot indicates that earnings are close to the random sample in the years up to and including event-year 0, and substantially lower thereafter. Thus, the abnormally high accruals in years –5 through 0 have the effect of masking the lower cash flows and inflating reported earnings. This is consistent with management attempting to delay a decline in reported earnings through accrual management.

Table 5 summarizes the results from tests of earnings management using the alternative models to generate discretionary accruals. For each model of discretionary accruals, the table reports descriptive statistics on the estimates of earnings management, their standard errors and t-statistics, along with the aggregate Z-statistic. The Z-statistic is positive and highly statistically significant at conventional levels for all five models, supporting the hypothesis that earnings have been managed upwards. The statistic is the largest for the Modified Jones Model (5.76) followed by the Industry Model (5.00), the Healy Model (3.90), the Jones Model (3.69) and the DeAngelo Model (2.88). A comparison of the point estimates of earnings management and their associated standard errors permits the source of the differences in the Z-statistics to be examined. The Jones and Modified Jones Models have standard errors that are markedly lower than the other models. This reinforces our previous findings from table 1 that the Jones and Modified Jones Models are more successful at explaining variation in accruals. The lower standard errors explain the source of their power. The low power of the Jones Model relative to the Modified Jones Model stems from its smaller estimates of earnings management. These smaller estimates are consistent with the SEC sample including firms that overstate revenues and these overstatements not being detected by the Jones Model. This reason is investigated in more detail in table 6. Finally, the relatively high Z-statistic for the Industry Model stems from a combination of a high point estimate of earnings management relative to the Jones Model and a low standard error relative to the Healy and DeAngelo Models.[15]

[14] Some firms are alleged to have overstated earnings for two or more consecutive years. In figure 5, event year 0 pools across all observations for which overstatement is alleged, event year –1 is the year prior to the first year in which overstatement is alleged, and event year +1 is the year following the last year in which overstatement is alleged. Note that in the regression analysis, PART is coded as one in years when earnings management is alleged and zero otherwise.

[15] Firms subsequently restate earnings in 39 of the 56 firm-years in which earnings overstatement is alleged by the SEC. These 39 observations provide us with an opportunity to investigate the extent of earnings management detected by the models compared to that identified by the SEC. The mean (median) restatement is 4.6 (2.3)% of assets. The mean (median) detected earnings management as a percent of assets for the Healy Model is 14.7 (5.6); the DeAngelo Model is 14.6 (2.3); the Jones Model is 10.5 (5.3); the Modified Jones Model is 15.9 (7.1); and the Industry Model is 15.4 (8.1). These results are consistent with either (i) the SEC identifying or requiring only a subset of the total earnings management to be restated by the firms; or (ii) the models systematically overstating the magnitude of earnings management in this sample.

FIGURE 5

Time Series of Median Annual Total Accruals, Cash From Operations and Earnings all Standardized by Lagged Total Assets. Year 0 is the Year in which the SEC Alleges that the Firm has Overstated Earnings. The SEC Sample Consists of 32 Firms Identified by the SEC for Overstating Annual Earnings. The Random Sample Consists of 1000 Randomly Selected Firm Years.

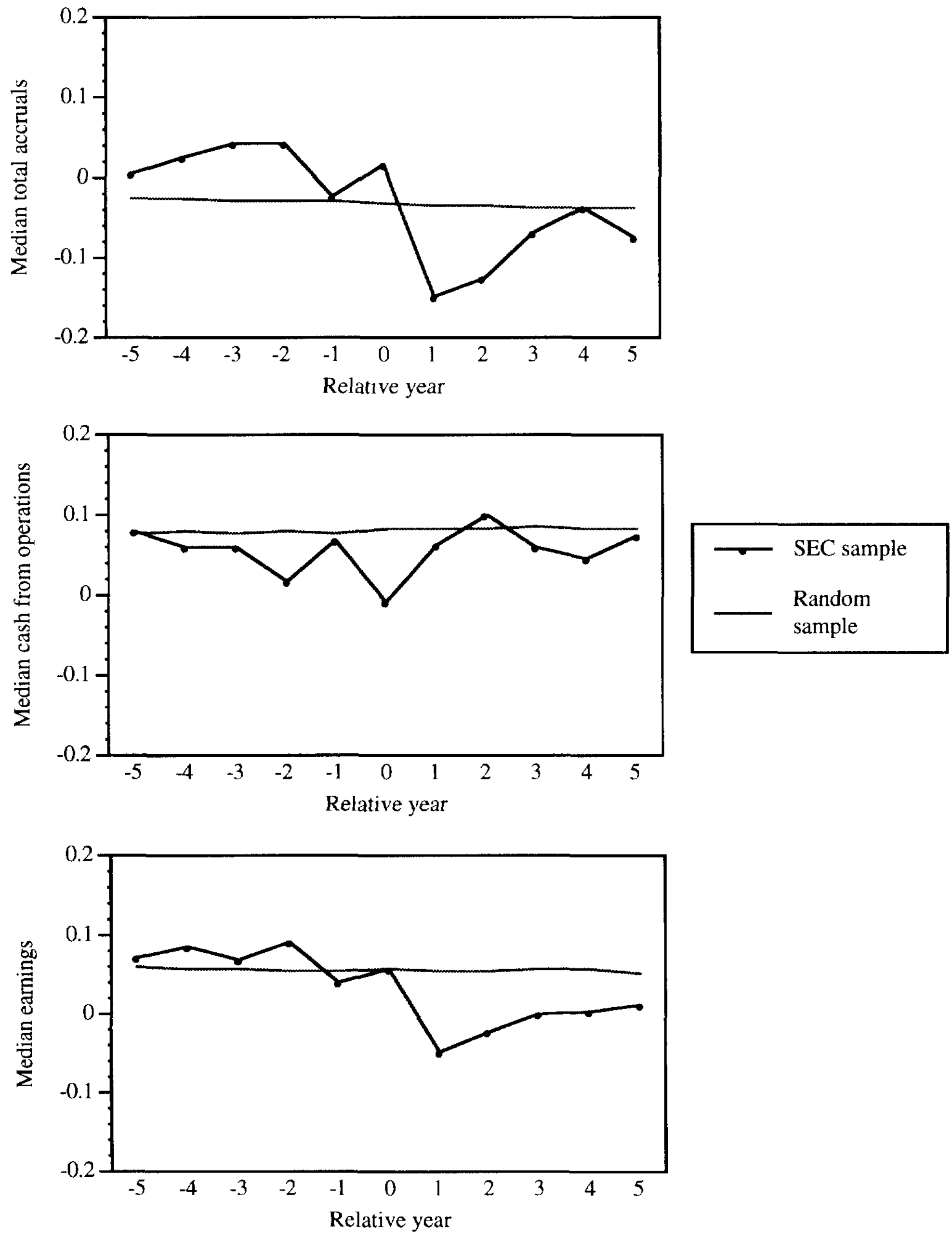


TABLE 5
Results of Tests for Earnings Management Using Alternative Models to Measure Discretionary Accruals. Sample of 32 Firms Targeted by the SEC in Accounting and Auditing Enforcement Releases (AAERs) between 1982 and 1992 for Allegedly Overstating Earnings.

Model	*Mean*	*Standard Deviation*	*Lower Quartile*	*Median*	*Upper Quartile*
Healy Model:					
earnings management	0.236	0.475	−0.022	0.058	0.258
standard error	0.203	0.255	0.084	0.126	0.201
t-statistic	0.760	1.310	−0.258	0.670	1.606
Z-statistic = 3.90**					
DeAngelo Model:					
earnings management	0.278	0.581	−0.011	0.089	0.310
standard error	0.269	0.277	0.118	0.168	0.279
t-statistic	0.564	0.907	0.088	0.467	1.224
Z-statistic = 2.88**					
Jones Model:					
earnings management	0.138	0.374	−0.023	0.061	0.172
standard error	0.158	0.183	0.075	0.105	0.158
t-statistic	0.754	1.414	−0.165	0.675	1.744
Z-statistic = 3.69**					
Modified Jones Model:					
earnings management	0.171	0.333	0.002	0.083	0.284
standard error	0.136	0.103	0.070	0.106	0.156
t-statistic	1.193	1.991	0.086	0.895	2.020
Z-statistic = 5.76**					
Industry Model:					
earnings management	0.218	0.418	−0.015	0.090	0.280
standard error	0.198	0.257	0.073	0.123	0.227
t-statistic	0.972	1.498	−0.123	1.038	1.488
Z-statistic = 5.00**					

Notes:
Earnings management represents the estimated coefficient on PART, ($\hat{b}_1$), from firm-specific regressions of $DAP_{it} = \hat{a}_1 + \hat{b}_1\ PART_{it} + e_{it}$; where DAP is the measure of discretionary accruals produced by each of the models and PART is an indicator variable equal to 1 in a year in which earnings management is hypothesized to occur in response to the stimulus identified by the researcher and 0 otherwise. Standard error is the standard error of the coefficient on PART for each of the regressions and t-statistic is the t-statistic testing the null hypothesis that the coefficient on PART is equal to zero.
** Significantly different from zero at the 1 percent level using a two-tailed test.

TABLE 6

Results of Tests for Earnings Management Using Alternative Models to Measure Discretionary Accruals. Comparison of the Jones and Modified Jones Models on the SEC Sample Stratified by the Source of the Alleged Earnings Overstatement. Sample of 32 Firms Targeted by the SEC in Accounting and Auditing Enforcement Releases (AAERs) between 1982 and 1992.

Model	*Mean*	*Standard Deviation*	*Lower Quartile*	*Median*	*Upper Quartile*
Panel A: Sample consists of 18 firms managing revenues					
Jones Model:					
earnings management	0.005	0.185	–0.030	0.038	0.095
Z-statistic = 1.56					
Modified Jones Model:					
earnings management	0.091	0.288	0.009	0.074	0.183
Z-statistic = 3.88**					
Panel B: Sample consists of 14 firms not managing revenues					
Jones Model:					
earnings management	0.310	0.482	–0.017	0.122	0.513
Z-statistic = 3.80**					
Modified Jones Model:					
earnings management	0.274	0.368	–0.005	0.118	0.515
Z-statistic = 4.31**					

Notes:
Earnings management represents the estimated coefficient on PART, ($\hat{b}_1$), from firm-specific regressions of $DAP_{it} = \hat{a}_i + \hat{b}_i\ PART_{it} + e_{it}$; where DAP is the measure of discretionary accruals produced by each of the models and PART is an indicator variable equal to 1 in a year in which earnings management is hypothesized to occur in response to the stimulus identified by the researcher and 0 otherwise
**Significantly different from zero at the 1 percent level using a two-tailed test

Table 6 provides an analysis of the impact of revenue-based earnings management on the performance of the Jones Model. The sample is stratified by the source of the earnings overstatement that is alleged by the SEC. Fifteen of the sample firms are accused of overstating revenues alone. A further three firms are accused of overstating revenues in combination with understating expenses. The remaining 14 firms are accused of understating expenses. We form two samples consisting of the 18 firms that are alleged to have overstated revenues and the 14 firms for which no overstatement of revenues is alleged. Table 6 reports the results of tests for earnings management applied to each of these two samples using the Jones and Modified Jones Models.

Panel A of table 6 reports the results for the sample for which revenue overstatements are alleged. The Z-statistic of 1.56 for the Jones Model is insignificantly different from zero at conventional levels, while the Z-statistic of 3.88 for the Modified Jones Model is highly significant. Inspection of the earnings management estimates for these two models indicates that the higher Z-statistic for the Modified Jones Model results from substantially larger estimates of earnings management. The mean (median) estimate of earnings management is 0.5% (3.8%) of lagged assets for the Jones Model and 9.1% (7.4%) of lagged assets for the Modified Jones Model. Panel B of table 6 reports results for the sample for which no revenue-based overstatements of earnings are alleged. The Z-statistics of 3.80 for the Jones Model and 4.31 for the Modified Jones Model are similar and statistically significant. Further inspection reveals that the earnings management estimates are also very similar. Thus, consistent with the results from our artificially managed samples, the two models appear to perform similarly in detecting expense-based earnings management. Overall, the results in table 6 provide confirmatory evidence that the Modified Jones Model is more powerful than the Jones Model in the presence of revenue-based earnings management.

The results in tables 5 and 6 provide descriptive evidence on the relative performance of the alternative models for measuring discretionary accruals. The results in table 7 directly investigate the frequency of type II errors for the competing models. Table 7 reports the proportion of the firms in the SEC sample for which the null hypothesis that discretionary earnings is less than or equal to zero is rejected. If it is assumed that all models are well specified and that the SEC has correctly identified firms that managed earnings, then the proportions of rejections in table 7 provide estimates of the relative power of the tests. The results indicate that the Modified Jones Model rejects the null hypothesis most frequently, followed by the Industry Model, the Jones Model, the Healy Model and the DeAngelo Model. These rankings correspond closely to the rankings of the power functions obtained in the simulation tests and reinforce the documented superiority of the Modified Jones Model.

VI. CONCLUSIONS AND IMPLICATIONS

This paper evaluates the ability of alternative models to detect earnings management. The results suggest that all the models considered appear to produce reasonably well specified tests for a random sample of event-years. However, the power of the tests is low for earnings management of economically plausible magnitudes. When the models are applied to samples of firm-years experiencing extreme financial performance, all models lead to misspecified tests. In this respect, our results highlight the conditions under which misspecified tests are likely to arise. However, we hasten to add that establishing the extent to which the results of an existing study are misspecified requires a detailed reexamination of that study (e.g., Holthausen et al.'s 1995 reexamination of Healy 1985). Finally, we find that a modified version of the model developed by Jones (1991) provides the most powerful tests of earnings management.

The findings in this study provide three major implications for research on earnings management. First, regardless of the model used to detect earnings management, the power of the tests is relatively low for earnings management of economically plausible magnitudes. Subtle cases of earnings management in the order of, say, one percent of total assets require sample sizes of several hundred firms to provide a reasonable chance of detection. Our analysis has focused primarily on documenting the properties of existing models. Further research to develop models that generate better specified and more powerful tests will further enhance our ability to detect earnings management.[16]

[16]Preliminary work in this direction is conducted by Beneish (1994).

TABLE 7

Comparison of Tests for Earnings Management Based on Alternative Models to Measure Discretionary Accruals. Percentage of Firms that are Alleged by the SEC to have Overstated Earnings for which the Null Hypothesis of No Earnings Management is Rejected (One-Tailed Tests). Sample of 32 Firms that are Targeted by the SEC in Accounting and Auditing Enforcement Releases (AAERs) between 1982 and 1992.

Model	*Test level of 5%*	*Test level of 1%*
Healy Model:		
t-test	12.5%*	6.3%**
DeAngelo Model:		
t-test	9.4	0.0
Jones Model:		
t-test	18.8**	6.3**
Modified Jones Model.		
t-test	28.1**	12.5**
Industry Model:		
t-test	18.8**	9.4**

* Significantly different from the specified test level at the 5 percent level using a two-tailed binomial test.
** Significantly different from the specified test level at the 1 percent level using a two-tailed binomial test.

Second, if the earnings management partitioning variable is correlated with firm performance, then tests for earnings management are potentially misspecified for all of the models considered. Pertinent measures of firm performance include earnings performance and cash from operations performance. Two recommendations can be made when facing this problem. First, the researcher can evaluate the nature of the misspecification and conduct a qualitative assessment of how it affects statistical inferences. For example, the nature of the performance-related bias may be such that the coefficient on the earnings management partitioning variable is negatively biased, while the researcher's hypothesis predicts a positive coefficient. Thus, if the researcher finds a significant positive coefficient, it would be reasonable to conclude that the hypothesis is supported, since the misspecification works against finding the result. Second, the researcher can attempt to directly control for the performance related misspecification. Possible approaches include the use of a control sample (e.g., Healy 1985), inclusion of firm performance in the earnings management regression (e.g., DeAngelo et al. 1994) or some other form of analysis of variance that controls for firm performance (e.g., Holthausen et al. 1995).

Finally, it is important to consider the relation between the context in which earnings management is hypothesized and the model of nondiscretionary accruals that is employed, because the model of nondiscretionary accruals may unintentionally extract the discretionary component of accruals. For example, if the Jones Model is used in a research context where discretion is exercised over revenues, then it is likely to extract the discretionary component of total accruals. Similarly, if the Industry Model is used in a research context where intra-industry correlation in discretionary accruals is expected, then it is likely to extract the discretionary component of total accruals. Consideration of the sample details should help avoid the use of a model of nondiscretionary accruals that unintentionally extracts discretionary accruals.

REFERENCES

Beneish, M. D. 1994. The detection of earnings manipulation. Working paper, Duke University, Durham, NC.

Brown, S., and J. Warner. 1980. Measuring security price performance. *Journal of Financial Economics* 14: 3–31.

———, and ———. 1985. Using daily stock returns: The case of event studies. *Journal of Financial Economics* 8: 205–258.

DeAngelo, L. 1986. Accounting numbers as market valuation substitutes: A study of management buyouts of public stockholders. *The Accounting Review* 61: 400–420.

———. 1988. Managerial competition, information costs, and corporate governance: The use of accounting performance measures in proxy contests. *Journal of Accounting and Economics* 10: 3–36.

DeAngelo, H., L. DeAngelo, and D. J. Skinner. 1994 Accounting choice in troubled companies. *Journal of Accounting and Economics* 17: 113–144.

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

———, and R. G. Sloan. 1991. Executive incentives and the horizon problem: An empirical investigation. *Journal of Accounting and Economics* 14: 51–89.

Defond, M. L., and J. Jiambalvo. 1994. Debt covenant violations and manipulation of accruals. *Journal of Accounting and Economics* 17: 145–176.

Feroz, E. H., K. Park, and V. Pastena. 1991 The financial and market effects of the SEC's accounting and auditing enforcement releases *Journal of Accounting Research* 29 (Supplement): 107–142.

Gaver, J. J., K. M. Gaver, and J. R. Austin. 1995. Additional evidence on the association between income management and earnings-based bonus plans. *Journal of Accounting and Economics* (forthcoming).

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

———, and K. G. Palepu. 1990. Effectiveness of accounting-based dividend covenants. *Journal of Accounting and Economics* 12: 97–123.

Holthausen, R., D. F Larcker, and R. G. Sloan. 1995. Annual bonus schemes and the manipulation of earnings. *Journal of Accounting and Economics* (forthcoming)

Johnston, J. 1984. *Econometric methods*. 3d ed. New York, NY: McGraw-Hill Book Company.

Jones, J. 1991. Earnings management during import relief investigations. *Journal of Accounting Research* 29: 193–228.

Kaplan, R. S. 1985. Comments on Paul Healy: Evidence on the effect of bonus schemes on accounting procedures and accrual decisions. *Journal of Accounting and Economics* 7: 109–113.

McNichols, M., and G. P. Wilson. 1988. Evidence of earnings management from the provisions for bad debts. *Journal of Accounting Research* 26 (Supplement): 1–31.

Pincus, K., W. H. Holder, and T. J. Mock. 1988. *Reducing the Incidence of Fraudulent Financial Reporting: The Role of the Securities and Exchange Commission.* Los Angeles, CA: SEC and Financial Reporting Institute of the University of Southern California.

Pourciau, S. 1993. Earnings management and nonroutine executive changes. *Journal of Accounting and Economics* 16: 150–180.

Salkever, D. 1976. The use of dummy variables to compute predictions, prediction errors, and confidence intervals. *Journal of Econometrics* 4: 393–397.

Securities and Exchange Commission. 1992. Annual Report. Washington, D.C.: GPO.

Schipper, K. 1989. Commentary on earnings management. *Accounting Horizons* 3: 91–102.

Sweeney, A. P. 1994. Debt covenant violations and managers' accounting responses. *Journal of Accounting and Economics* 17: 281–308.