Financial Analysts Journal
Volume 69 • Number 2
©2013 CFA Institute

Earnings Manipulation and Expected Returns

Messod D. Beneish, Charles M.C. Lee, and D. Craig Nichols

*An accounting-based earnings manipulation detection model has strong out-of-sample power to predict cross-sectional returns. Companies with a higher probability of manipulation (*M*-score) earn lower returns on every decile portfolio sorted by size, book-to-market, momentum, accruals, and short interest. The predictive power of* M*-score stems from its ability to forecast changes in accruals and is most pronounced among low-accrual (ostensibly "high-earnings-quality") stocks. These findings support the investment value of careful fundamental and forensic analyses of public companies.*

In our study, we examined the investment value of a particular form of detailed financial analysis associated with the detection of earnings manipulation. Closely related to the "quality of earnings" analysis popularized by O'Glove (1987), Kellogg and Kellogg (1991), and Siegel (1991), adherents of this form of analysis pore over corporate financial statements looking for inconsistencies, irregularities, and other signs of trouble. Although these efforts have yielded individual success stories, the evidence to date has been largely anecdotal.[1]

The statistical model that we examined (Beneish 1999a) represents a systematic distillation of forensic accounting principles described in the practitioner literature. Using model coefficients that were estimated in a prior period, we found clear *out-of-sample* evidence that these techniques have had substantial investment value.[2] In particular, we found that this model correctly identified, in advance of public disclosure, a large majority (71%) of the most famous accounting fraud cases that surfaced after the model's estimation period.

The main focus of our study, however, was return prediction, not fraud detection. Specifically, we investigated a potential link between the probability of manipulation (*M*-score) generated by the Beneish model and subsequent returns. Although relatively few companies are indicted ("caught") for fraud, the incidence of earnings management is likely much higher. For example, in a survey of 169 chief financial officers of public companies, Dichev, Graham, and Rajgopal (2012) reported that their survey respondents estimated that approximately 20% of all companies manage earnings to misrepresent economic performance.

We posited that the *M*-score generated by the model will be informative of a company's expected returns. The reason is that the "profile of a typical earnings manipulator" as defined by Beneish (1999a) is a company that is (1) growing quickly (extremely high year-over-year sales growth), (2) experiencing deteriorating fundamentals (as evidenced by a decline in asset quality, eroding profit margins, and increasing leverage), and (3) adopting aggressive accounting practices (e.g., receivables growing much faster than sales, large income-inflating accruals, and decreasing depreciation expense).

Our main hypothesis was that companies that share traits with past earnings manipulators (i.e., those that "look like manipulators") represent a particularly vulnerable type of growth stock. Because of their strong recent growth trajectory, these companies are likely to be more richly priced. At the same time, they exhibit a number of potentially problematic characteristics, indicative of either lower *earnings quality* or a more challenging economic environment.[3] Although the accounting games such companies engage in might not be serious enough to warrant legal action, we posited that their earnings trajectory is more likely to disappoint investors (i.e., they have lower earnings quality). To the extent that the pricing implications of these accounting-based indicators are not fully transparent to investors, companies that "look like" past earnings manipulators will also earn lower future returns.

■ *Discussion of findings.* Our evidence shows that companies with a higher probability of manipulation (*M*-score) earn lower returns in every decile

Messod D. Beneish is the Sam Frumer Professor of Accounting at Indiana University, Bloomington. Charles M.C. Lee is the Joseph McDonald Professor of Accounting at Stanford University, California. D. Craig Nichols is assistant professor of accounting at Syracuse University, New York.

portfolio sorted by size, book-to-market, momentum, accruals, and short interest. These returns are economically significant (averaging just below 1% a month on a risk-adjusted basis) and survive the usual risk controls. Our evidence further shows that a large proportion of the abnormal return is earned in the short three-day windows centered on the next four quarterly earnings releases, suggesting that our results are due to a delayed reaction to earnings-related news rather than risk-based factors. The robustness of our results, even among highly liquid companies, implies that they are unlikely to be fully explained by transaction costs.

Having documented *M*-score's ability to predict risk-adjusted returns, we aimed to better understand the nature of the information it conveys. In particular, extensive evidence suggests that accruals, as well as components and variants of accruals, predict one-year-ahead returns (see, e.g., Sloan 1996; Fairfield, Whisenant, and Yohn 2003; Richardson, Sloan, Soliman, and Tuna 2005; Cooper, Gulen, and Schill 2008; Hirshleifer, Hou, and Teoh 2012). Given *M*-score's close ties to accruals, we wished to better understand (1) how *M*-score differs from accruals in terms of each variable's predictive ability and (2) the source and nature of *M*-score's incremental predictive power over accruals.

We performed three sets of analyses. First, we conducted detailed tests on the joint ability of accruals and *M*-score to predict returns. We found that the dominance of *M*-score over accruals is evident in both independent and nested sorts. When companies are sorted on these two variables independently, *M*-score is particularly effective in predicting returns among low-accrual companies (i.e., companies with *high earnings quality* according to their accrual ranking). For the lowest-accrual quintile, the spread in size-adjusted returns between high-*M*-score companies and low-*M*-score companies was –19.8% over the next 12 months. For companies in the second-lowest-accrual quintile, the spread was –10.5% a year. In general, after controlling for *M*-score, we found that accruals exhibit limited predictive power, which is concentrated primarily in the mid-*M*-score quintiles.

Second, we analyzed individual components of the model in specific subpopulations. Noting that *M*-score is particularly effective in separating future winners from losers among low-accrual companies, we designed a difference-in-difference test that sheds light on the elements of the model that contribute the most to this effect. Our analysis shows that the variables related to the *predisposition to commit fraud*, rather than the variables associated with the *level* of aggressive accounting, are the primary drivers of the model's incremental power. Specifically, the model's incremental predictive power in the low-accrual group is most directly related to its ability to identify fast-growing companies that have recently experienced some economic headwind. Elements of the model associated with aggressive accounting (i.e., those components most closely aligned with accruals) provide no incremental predictive power.

Third, we discovered that the Beneish model's efficacy is associated with its ability to predict the *directional change* in current-year accruals (i.e., whether the accruals component of current-year earnings will continue into the next year or disappear). Specifically, we found that high-*M*-score companies (those that look like manipulators) have income-increasing accruals that are more likely to disappear next year and income-decreasing accruals that are more likely to persist, or reappear, next year. We observed the exact opposite among low-*M*-score companies (those that look least like manipulators). In other words, *M*-score provides useful information about the future *persistence* of accruals.

Our study is related to a growing body of literature on the effective use of corporate financial information. Aside from fraud detection, researchers have used financial information in different contexts to better understand earnings quality (Lev and Thiagarajan 1993; Sloan 1996; Richardson et al. 2005), bankruptcy risk (Altman 1968; Ohlson 1980), the direction of future earnings (Ou and Penman 1989), and future returns (Holthausen and Larcker 1992; Piotroski 2000; Beneish, Lee, and Tarpley 2001; Mohanram 2005). Our study builds on and extends this line of research by documenting the usefulness of earnings manipulation detection techniques for earnings quality assessment and return prediction. By parsing the various elements of the model, we found new evidence on how and why such techniques work. These insights suggest new directions for earnings quality analysis and should enhance future efforts to identify potential over- and undervaluations. Later in the article, we discuss in more detail how our research is related to, but distinct from, these prior studies.

Our findings also extend the literature on market learning and the limits of arbitrage (Berk and Green 2004; Green, Hand, and Soliman 2011; Bebchuk, Cohen, and Wang, forthcoming). At the heart of the issue is the definition of "publicly available" information. Although all the individual components of the model are publicly available, our evidence implies that public availability alone does not ensure that these elements are fully integrated into price in a timely manner. Learning takes time. Indeed, our results suggest that even publication in a journal does not guarantee that a strategy will become fully transparent to the market immediately.[4]

Overall, our analyses provide substantial support for the use of forensic accounting in equity investing. Our evidence indicates that the efficacy of the model derives from its ability to separate companies whose accruals are more likely to persist from those whose accruals are more likely to reverse. Because Beneish developed this model by using forensic accounting principles and the model's parameters were estimated in a prior period, our findings offer out-of-sample validation for the approach advanced by forensic accountants.

Related Literature

Our study is related to two main bodies of work: (1) the large literature on accruals and future returns and (2) studies that use financial statement variables in other decisional contexts. Here, we relate the Beneish (1999a) model, in broad terms, to these prior studies. Later in the article (and also in Appendix A), we present a detailed description of the Beneish model, as well as the intuition behind each of the eight variables used in its estimation.

Accruals and Future Returns. In a seminal study, Sloan (1996) showed that companies with higher (income-inflating) accruals earn lower returns than companies with lower accruals. He offered evidence that the predictive power of accruals derives from the fact that the cash-based component of earnings is more persistent (i.e., is of "higher quality") than the accruals-based component. Sloan's original study spawned a large number of studies seeking to explain, confirm, refute, or recast his findings.[5] Although most of these studies have confirmed or extended Sloan's original findings, a few have raised questions as to his findings' robustness or interpretation.[6]

Although our study is related to the accruals literature, the model we tested was designed for a different purpose, and the individual components of the model reflect this broader mandate. In general terms, the profile of an earnings manipulator that emerges from the Beneish model is a company that is (1) growing extremely fast (sales growth index), (2) experiencing some economic headwind (asset quality index; gross margin index; sales, general, and administrative expenses [SGA] index; leverage index), and (3) practicing aggressive accounting (days in receivables, depreciation index, and the ratio of accruals to total assets).[7]

Of these three categories, only the last is aligned with the accrual measures. The explanatory variables from the first two groups are designed primarily to detect the *propensity to commit fraud* (i.e., exceptionally high year-over-year sales growth, eroding profit margins, a decline in asset quality, and an increase in leverage) rather than the effect of aggressive accounting. In other words, five of the eight variables in Beneish's fraud detection model are not associated with the various manifestations of accruals.

In fact, our test results show that the incremental power of the model over accruals stems mainly from variables in the first two categories and is not a function of aggressive accounting per se. Specifically, our findings show that the efficacy of *M*-score is associated with its ability to predict *changes* in current-year accruals. In other words, *M*-score provides additional information about the quality of earnings beyond the current-year *level* of reported accruals.

Other Financial Analysis Research. Our study is also related to studies that have examined the usefulness of financial statement information in other decisional contexts. Broadly speaking, these studies fall into three categories: (1) distress analysis—studies that aim to predict bankruptcy and financial distress (e.g., Beaver 1966; Altman 1968; Ohlson 1980; Beaver, McNichols, and Price 2007); (2) basic ratio analysis—studies that combine a large set of financial ratios in a less structured manner to predict either future earnings or stock returns (Ou and Penman 1989; Holthausen and Larcker 1992); and (3) contextual analysis—studies that apply financial analysis in such targeted settings as value stocks (Piotroski 2000), growth stocks (Mohanram 2005), or extreme performers (Beneish et al. 2001). Although all these studies used accounting information, each group is designed for a different purpose and features different financial variables.[8]

The primary characteristic of Beneish (1999a) that distinguishes it from the other studies is its close allegiance to the fraud detection literature espoused by financial practitioners. For example, the model is related to the analysis by Foster (1979) of Abraham Briloff, an accounting professor who successfully used public reports to identify accounting irregularities. Other sources that influenced the original model include O'Glove (1987), Kellogg and Kellogg (1991), Siegel (1991), and an earlier edition of Schilit (2010). Unlike Ou and Penman (1989) and Holthausen and Larcker (1992), the model in Beneish (1999a) does not use a large number of ratios; and unlike the contextual studies, the Beneish model was not designed to predict returns in particular subpopulations of companies. Rather, it was built by carefully observing the financial information most often discussed by experts in forensic accounting, who are concerned with detecting accounting irregularities, particularly among fast-growing companies.[9]

In sum, prior studies have examined the usefulness of financial information in a variety of contexts. Our study extends this literature by focusing on a model built on forensic accounting principles and by demonstrating its strong out-of-sample ability to both identify famous fraud cases and predict cross-sectional stock returns. Our study also provides new evidence on why forensic accounting techniques work and how they inform us about companies' earnings quality.

The Earnings Manipulation Detection Model

Beneish (1999a) profiled companies that had manipulated earnings (companies either charged with manipulation by the U.S. SEC or that admitted to manipulation in the public press) and developed a statistical model to distinguish manipulators from nonmanipulators. The model presented in Beneish (1999a) relies exclusively on financial statement data and is thus useful even in assessing the fraud potential of companies without security prices (e.g., in pricing an IPO). In the original paper, Beneish estimated the model by using data for 1982–1988, and its holdout sample performance was assessed for 1989–1992.

Since the publication of the original study, the model has attained some notoriety for flagging Enron Corporation well in advance of its eventual demise.[10] The model has been featured in financial statement analysis textbooks (Fridson 2002; Stickney, Brown, and Wahlen 2003) and in articles directed at auditors, certified fraud examiners, and investment professionals (Ciesielski 1998; Merrill Lynch 2000; Wells 2001; Dresdner Kleinwort Wasserstein, Inc. 2003; Harrington 2005). Evidence of its out-of-sample performance, however, is *ad hoc* and anecdotal.

Appendix A provides a detailed description of each variable, the loadings on each variable, and a tabulation of the sample distribution over time and across industries.

To compute a probability of manipulation (*M*-score), Beneish (1999a) estimated a probit regression by using a portion of his sample and validated the model's efficacy through a holdout sample.[11] As in all statistical discriminant analyses of this genre, a Type I classification error (the probability of "missing a culprit") needs to be traded off against a Type II classification error (the probability of "nabbing an innocent company"). Beneish (1999a) contains an extensive discussion of the relative implied cost of the two types of error for alternative *M*-score threshold (or "cutoff") values. His results show that the model performs best relative to a naive model when the relative cost of a Type I error to a Type II error is between 20:1 and 30:1 (Beneish 1999a, Table 5). This relative cost function corresponds to an *M*-score cutoff value of –1.78 (i.e., companies with an *M*-score that exceeds –1.78 would be flagged as potential manipulators).[12] Using this cutoff value, Beneish demonstrated that the model flagged approximately 13% of the holdout sample companies as manipulators. Strikingly, the 13% of sample companies that were flagged included approximately half of all actual manipulators.

In our study, we closely replicated the model as published in Beneish (1999a). We used the exact coefficients as estimated for the model (Beneish 1999a, Table 3). We also used the same cutoff value suggested in the original study to classify companies as manipulators (i.e., *M*-scores exceeding –1.78). For seven of the eight explanatory variables, we followed exactly the same construction as in the original paper. For the accruals variable, the original paper used a balance sheet estimation method (see Sloan 1996). For our study, consistent with the evolution in the accruals literature since 1999, we derived the same conceptual measure as in Beneish (1999a) but used information from the statement of cash flows rather than the noisier balance sheet estimate.[13]

We examined the model's performance with respect to well-known fraud cases over 1998–2002 (as reported by Audit Integrity), thus demonstrating the continued relevance of the model in detecting fraud (see Appendix B). This period was marked by an unusual number of high-profile fraud cases that helped bring forensic accounting to prominence. In the aftermath, Audit Integrity listed 20 companies that it deemed the most egregious examples of earnings manipulators.[14] Two of these companies are financial firms (to which the model does not apply), and one is not an actual fraud case (Motorola did not manipulate its own earnings; it only abetted Adelphia Communications Corporation). Because the holdout sample for the original Beneish model ends in February 1993, the remaining 17 companies identified by Audit Integrity represent an entirely out-of-sample test for the model.

As Appendix B shows, the model predicted the fraud for 12 of the 17 companies, including Cendant Corporation, Enron, Global Crossing, Qwest Communications International, and several other famous cases. On average, the model detected the fraud by using financial information that was available a year and a half *before* the public revelation of problems. Of particular note, the model received attention subsequent to the Enron scandal as the investing public learned that the model had flagged Enron prior to the debacle.[15]

In sum, the model appears to have performed quite well in identifying the most famous accounting fraud cases that surfaced after its publication.

Does *M*-Score Predict Future Returns?

We examined the model's ability to predict out-of-sample, one-year-ahead stock returns for a broad cross section of companies over 1993–2010. We report our test period, abnormal return computations, and the results of various tests that we ran to assess whether *M*-score has incremental predictive power over previously documented return predictors.

The Sample. We selected the initial sample from the Compustat Industrial, Research, and Full Coverage files for 1993–2010. We eliminated (1) financial services companies (SIC codes 6000–6899), (2) companies with less than $100,000 in sales (Compustat item #12) or in total assets (Compustat item #6), (3) companies with a market capitalization of less than $50 million at the end of the fiscal period preceding portfolio formation, and (4) companies without sufficient data to compute the probability of manipulation. Following Beneish (1999a), we winsorized the predictive variables in the model at the 1% and 99% levels for each year in our sample period to address problems caused by small denominators and to control for the effect of potential outliers.

Following a slightly modified version of the procedure outlined in Lyon, Barber, and Tsai (1999), we computed size-adjusted returns.[16] To form reference portfolios, we first identified decile portfolio breakpoints on the basis of all NYSE companies. We then assigned all NYSE, Amex, and NASDAQ companies to portfolios on the basis of those breakpoints. Because the smallest portfolio had a disproportionately large number of stocks, we further sorted those stocks into five portfolios on the basis of market capitalization. The end result was 14 size-based portfolios. We then accumulated returns for 12 months starting with the first day of the next month following portfolio assignment. If a company was delisted, we included returns up to the delistment date as well as any delistment return reported by CRSP. If a delistment return was missing, we estimated it by using the procedures outlined in Beaver et al. (2007). As in Lyon et al. (1999), from the month following delistment to the end of the holding period, we assumed that the proceeds from delisting, if any, were invested in the CRSP size-based portfolio to which the company belonged.

To compute size-adjusted returns, we accumulated returns for 12 months, starting with the fifth month after the fiscal year-end, by using the same delisting procedures described earlier, if necessary. We used the stock's market cap at the end of the fourth month following the fiscal year-end to identify its reference portfolio. We then subtracted the return for the reference portfolio from the return for the company.

To ensure that the trading strategies we examined were implementable, we required all companies in our rankings to have stock return data available in the CRSP files at the time the rankings were made, and we used prior-year decile cutoffs to assign companies to decile portfolios of the ranking variable (the probability of manipulation, accruals, momentum, etc.) in the current year. We based our trading strategy return computations on taking positions four months after the end of the fiscal year. The final sample consisted of 43,544 company-year observations from 1993 to 2010.[17]

***M*-Score and Future Returns.** Although Appendix B and prior research (e.g., Beneish 1997) demonstrate the ability of the Beneish model to identify companies that commit fraud, very few fraud cases lead to indictments.[18] The Beneish model, however, flagged 17.4% of company-year observations as potential frauds. As discussed earlier, the much higher number of potential frauds flagged by the model is a function of a trade-off between the costs of Type I and Type II errors. On the basis of results reported in the original study (Beneish 1999a)—as well as an intuitive assessment of the cost to an asset manager of missing a manipulating company—we adopted a threshold (cutoff) value for *M*-score of –1.78.[19]

Table 1 compares returns for companies flagged as probable manipulators with the returns of not-flagged companies by using this cutoff value. For the full sample period, flagged companies generated one-year-ahead size-adjusted returns of –7.5%, whereas not-flagged companies had positive returns of 2.4%. Both of these average returns are statistically significant. Not-flagged companies outperformed flagged companies by 9.9 percentage points (pps), on average—also statistically significant.

Table 1 also compares flagged and not-flagged companies by year and in subperiods. The spread in returns across not-flagged and flagged companies is negative in only four years (1994, 2002, 2004, and 2008) but is not significantly negative in any year. Not-flagged companies significantly outperformed flagged companies in 13 years. Splitting the sample into two subperiods (Panel B), we can see that the effect was weaker after the tech

Table 1. Year-by-Year Size-Adjusted Returns to Flagged Companies, 1993–2010

		Not Flagged		Flagged		Spread
	N	Portion	BHSAR	Portion	BHSAR	(pps)
A. Yearly results						
1993	2,304	82.4%	3.5%***	17.6%	–5.1%**	8.6***
1994	2,474	79.0	1.4	21.0	2.7	–1.3
1995	2,786	79.8	1.2	20.2	–15.4***	16.5***
1996	2,972	77.1	–0.7	22.9	–8.7***	8.0***
1997	3,139	76.6	0.4	23.4	–9.4***	9.9***
1998	2,798	78.1	9.1***	21.9	5.9	3.2
1999	2,789	77.2	6.6***	22.8	–21.6***	28.2***
2000	2,616	72.8	4.1***	27.2	–28.1***	32.2***
2001	2,480	86.6	–1.5*	13.4	–17.5***	16.0***
2002	2,278	89.6	5.7***	10.4	11.0**	–5.3
2003	2,550	88.1	0.8	11.9	–5.5**	6.3**
2004	2,571	84.7	2.2**	15.3	5.2	–3.0
2005	2,534	87.3	1.1	12.7	–3.3	4.4**
2006	2,480	86.3	5.3***	13.7	1.5	3.8*
2007	2,348	86.8	–0.7	13.2	–3.5	2.9
2008	1,997	88.6	3.1**	11.4	6.5	–3.3
2009	1,988	89.7	6.2***	10.3	–6.0*	12.2*
2010	430	91.6	0.1	8.4	–12.2**	12.2**
Full sample	43,534	82.6	2.4***	17.4	–7.5***	9.9***
B. Subperiod results						
1993–2001	24,358	78.68%	3.40%***	21.32%	–11.17%***	–14.57***
2002–2010	19,176	87.61	3.17***	12.39	0.50	–2.67**

Notes: This table reports the year-by-year size-adjusted returns for companies flagged by the Beneish (1999a) model and those that were not. BHSAR denotes annual buy-and-hold returns to an equal-weighted portfolio formed at the start of the first day of the fifth month following the end of the fiscal year less the returns to a portfolio of companies from the same NYSE/Amex/NASDAQ size decile (size decile membership determined at the beginning of return window). For delisted companies, any proceeds upon delisting are reinvested in the size portfolio to which the company belongs. "Flagged" denotes companies that fit the profile of an earnings manipulator on the basis of the *M*-score model in Beneish (1999a) and a cutoff of –1.78.

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

bubble burst. The percentage of companies flagged as likely manipulators has also declined in recent years, indicating a drop in companies fitting the manipulator profile in the post-bubble period (following the Sarbanes–Oxley Act of 2002). Overall, Table 1 suggests that flagged companies (those that merely look like a manipulator) are associated with lower expected returns.

Distinguishing *M*-Score from Other Predictors of Future Returns. Prior research has shown that a number of characteristics are correlated with subsequent returns: (1) the difference between earnings and cash flows from operations (accruals), following Sloan (1996); (2) price momentum (momentum), following evidence in Jegadeesh and Titman (1993) that past 3- to 12-month returns tend to continue in the subsequent year; (3) company size (market value of equity, or MVE), following evidence in, among others, Fama and French (1992); (4) the book-to-market ratio (BTM), following evidence in Davis (1994), Haugen and Baker (1996), and Fama and French (1992); and (5) the short-interest ratio (SIR), following evidence in Drake, Rees, and Swanson (2011) that companies with high SIRs subsequently earn lower returns.

Table 2 reports the correlation matrix for these characteristics (Pearson correlations are above the diagonal and Spearman correlations are below the diagonal). Correlations of *M*-score with three variables are noteworthy. First, *M*-score and accruals are highly correlated (Pearson correlation = 0.452, $p < 0.001$). Many observers have speculated that earnings management is an important reason why the persistence of accounting accruals differs from that of cash flows, suggesting that earnings management misleads investors. Therefore, it is possible that both *M*-score and accruals measure earnings

Table 2. Correlation Matrix

	M-Score	Accruals	Momentum	ln(MVE)	BTM	SIR	BHSAR
M-score		0.452***	–0.039***	–0.042***	–0.020***	0.034***	–0.065***
Accruals	0.645***		–0.036***	–0.016***	–0.019***	–0.014***	–0.036***
Momentum	–0.064***	–0.039***		0.033***	–0.120***	–0.031***	0.025***
ln(MVE)	–0.048***	–0.048***	0.114***		–0.160***	0.126***	–0.011**
BTM	–0.113***	0.000	–0.278***	–0.280***		–0.037***	0.027***
SIR	0.013***	–0.046***	–0.048***	0.379***	–0.155***		–0.024***
BHSAR	–0.093***	–0.034***	0.062***	0.065***	0.053***	–0.028***	

Note: This table reports Pearson (above diagonal) and Spearman (below diagonal) correlations for sample variables; N = 43,534.

**Significant at the 5% level.
***Significant at the 1% level.

manipulation and that little incremental value exists in studying *M*-score. Second, the negative correlations between *M*-score and both momentum (–0.039, p-value < 0.001) and BTM (–0.020, p-value < 0.001) suggest that companies with a high probability of earnings overstatement have "momentum" and "glamour" (a low BTM) characteristics. Third, the Pearson correlation between *M*-score and SIR is positive and significant (0.034, p-value < 0.001), consistent with high-*M*-score companies attracting the attention of short sellers.

To examine whether the returns to a strategy based on *M*-score are subsumed by other potential predictors of future returns, we estimated Fama–MacBeth (1973) cross-sectional regressions of one-year-ahead, buy-and-hold, size-adjusted returns (BHSAR_{t+1}) on the scaled decile ranks of several predictors:

$$\begin{aligned}\text{BHSAR}_{t+1} = a_0 + a_1 M\text{-score}_t + a_2\text{Accruals}_t \\ + a_3\text{Momentum}_t + a_4\text{MVE}_t \\ + a_5\text{BTM}_t + a_6\text{SIR}_t + e_{t+1}.\end{aligned} \tag{1}$$

Table 3 reports the coefficients from 18 annual cross-sectional regressions based on Equation 1, as well as their time-series averages. The dependent variable is one-year-ahead, size-adjusted returns computed by using reference portfolio returns calculated as outlined in Lyon et al. (1999). Because the independent variables are scaled decile ranks, the values range from 0 to 1.

The results indicate that scaled *M*-score ranks are negatively correlated with one-year-ahead abnormal returns (–0.083, t-statistic = –2.58) and that momentum is positively correlated with one-year-ahead abnormal returns (0.089, t-statistic = 2.74). BTM is also positively correlated with future abnormal returns (0.060, t-statistic = 1.67). The remaining variables, including accruals, MVE, and SIR, do not attain significance. This finding suggests that after controlling for accruals and other variables associated with future returns, a long–short portfolio strategy based on extreme *M*-score deciles earns an 8.3% one-year-ahead abnormal return.[20]

To better understand the source of the abnormal return on an *M*-score strategy, we also estimated time-series regressions of monthly excess returns on the Fama–French (1993) factors and the momentum factor of Carhart (1997):

$$\begin{aligned}\text{ExRet}_t = a_0 + a_1\text{MKT}_t + a_2\text{SMB}_t \\ + a_3\text{HML}_t + a_4\text{WML}_t + e_{t+1},\end{aligned} \tag{2}$$

where

ExRet = the monthly value-weighted *M*-score portfolio return in excess of the return on the one-month T-bill

MKT = the value-weighted market index return in excess of the one-month T-bill

SMB = returns on a factor-mimicking portfolio for the size factor (small minus big)

HML = the book-to-market factor (high minus low)

WML = the momentum factor (winners minus losers)

To calculate ExRet, we sorted companies into portfolios each month on the basis of each company's most recent *M*-score and the prior-year *M*-score cutoff. We calculated value-weighted returns for each portfolio each month by using the market value of equity at the beginning of the month. We then subtracted the return on the one-month T-bill to calculate the excess return (ExRet). Each portfolio consisted of 222 monthly observations over July 1993–December 2011.

Table 4 shows that, on average, an *M*-score strategy involves a negative bet on market beta (the average coefficient on MKT is –0.191) as well as a positive exposure to differential expected returns based on company size (the average coefficient on SMB is 0.730). On average, an *M*-score strategy has marginally significant exposure to the value factor (the average coefficient on HML is –0.176) and

Table 3. Multivariate Cross-Sectional Regressions, 1993–2010

	Intercept	*M*-Score	Accruals	MVE	BTM	Momentum	SIR	Adjusted R^2
1993	0.074**	–0.127***	–0.002	–0.028	0.033	0.071***	–0.034	1.5%
1994	0.127**	–0.171***	0.020	–0.136***	–0.213***	0.238***	0.063	3.0
1995	–0.002	–0.145***	0.039	0.038	0.141***	0.046*	–0.135***	2.7
1996	–0.017	–0.015	–0.077**	0.011	0.096***	–0.006	–0.040	0.5
1997	–0.008	–0.162***	–0.015	–0.010	0.017	0.166***	–0.009	1.9
1998	0.256**	0.016	–0.319***	–0.311***	–0.283***	0.302***	0.373***	3.4
1999	–0.060	–0.348***	0.151***	0.117**	0.277***	0.081*	–0.105**	3.2
2000	–0.138***	–0.370***	0.120***	0.046	0.219***	0.340***	–0.158***	12.7
2001	–0.045	–0.158***	0.106***	0.109***	0.005	0.116***	–0.176***	4.4
2002	0.099	0.143**	–0.256***	–0.121**	0.183***	–0.053	–0.008	2.0
2003	–0.235***	0.107***	–0.052	0.086***	0.246***	0.179***	–0.055**	4.5
2004	–0.099**	0.063	–0.012	–0.045	0.057*	0.106***	0.069**	0.4
2005	–0.059**	0.014	–0.004	0.079***	0.153***	–0.022	–0.068***	2.2
2006	0.013	–0.060*	0.081***	0.013	–0.003	0.080***	–0.031	0.6
2007	0.007	–0.032	0.070***	–0.018	–0.034*	0.007	–0.022	0.3
2008	0.078	–0.074	–0.086	0.010	0.238***	–0.264***	–0.005	3.8
2009	0.152	–0.092**	–0.084*	–0.019**	–0.007	0.074**	–0.087	1.8
2010	0.028	–0.076**	–0.028	–0.040	–0.046**	0.138**	–0.072	1.4
Average	0.009	–0.083	–0.019	–0.012	0.060	0.089	–0.028	2.8%
t-Statistic	0.35	–2.58	–0.68	–0.51	1.67	2.74	–0.98	

Note: This table reports the time-series means of 18 annual cross-sectional (Fama–MacBeth) regressions; *t*-statistics are based on the time-series distribution of the parameter estimates.

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Table 4. Time-Series Asset-Pricing Regressions, July 1993–December 2011

	Intercept	MKT	SMB	HML	WML	Adjusted R^2
1 (low)	0.314%	1.087***	–0.076	0.224***	–0.052	69.9%
2	0.215	1.022***	0.075	0.041	0.026	75.3
3	0.346***	0.861***	0.189***	0.015	–0.001	82.2
4	0.222	0.959***	0.008	–0.036	0.008	81.7
5	–0.022	0.893***	–0.116**	–0.055	–0.014	80.8
6	0.221	0.858***	–0.107**	–0.075*	0.024	80.5
7	0.031	0.933***	–0.195***	–0.086*	0.019	78.6
8	–0.283	1.056***	–0.130**	0.099*	0.002	78.1
9	–0.340*	1.251***	–0.418***	0.307***	–0.119***	87.3
10 (high)	–0.658***	1.278***	–0.806***	0.400***	–0.062	84.9
Hedge	0.972***	–0.191**	0.730***	–0.176*	0.010	28.7

Note: N = 22 months for each portfolio.

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

no significant exposure to the price momentum (WML) factor.

More importantly, after controlling for monthly correlations with all four factor-mimicking portfolios, we found that a hedged portfolio of extreme *M*-score decile companies (decile 1 – decile 10) has a positive intercept (i.e., monthly alpha) of around 0.972% (11.7% a year), which is highly statistically significant. This effect is most pronounced in the highest-*M*-score decile (i.e., most of the return derives from a negative bet against growth companies that subsequently perform much worse than expected). As a group, these companies underperformed their size peers by around 66 bps a month.

To further evaluate *M*-score's ability to predict returns in various subpopulations of stocks, we separated the companies in each decile of MVE, BTM, momentum, SIR, and accruals into high- and low-*M*-score groups; the high-*M*-score group included the 7,569 observations in our sample that were flagged as potential manipulators. The results, reported in **Table 5** and depicted in Figures 1–5, are striking. The performance of the flagged subsample is worse than that of its not-flagged counterpart in all 50 deciles. Furthermore, the average size-adjusted return of the flagged companies is negative in 47 of the 50 deciles.

Panel A of Table 5 (**Figure 1**) shows that a size-based trading strategy that buys small companies (decile 1) and shorts large companies (decile 10) yields 2.5% a year. Combining *M*-score with size (e.g., buying small not-flagged companies and selling short large flagged companies), the strategy yields 15.2% a year. Similarly, Panel B of Table 5 (**Figure 2**) shows that the improvement to a BTM strategy is also quite substantial. A strategy of buying value (decile 10) and shorting glamour (decile 1) yields 8.0%. By combining BTM with *M*-score (e.g., buying not-flagged value companies and selling flagged glamour companies), the strategy's yield improves to 13.7% a year.

Panel C of Table 5 (**Figure 3**) reports that a momentum-based trading strategy that buys high (decile 10) and shorts low (decile 1) yields 10.2%. Combining momentum with *M*-score (e.g., buying not-flagged high-momentum companies and selling short flagged low-momentum companies), the strategy's yield improves to 25.3%. Similarly, in Panel D of Table 5 (**Figure 4**), trading on extreme SIR deciles yields an abnormal return of 5.8% a year, which improves to 16.3% by superimposing *M*-score.

Finally, in Panel E of Table 5 (**Figure 5**), accruals alone return 8.3% a year, which can be improved to 14.4% by selling short flagged high-accrual companies and buying not-flagged low-accrual companies. Note that *M*-score is especially helpful when accruals provide a positive signal of earnings quality but *M*-score offers a conflicting signal. When low-accrual companies (generally regarded as "high quality") are flagged, they have an average abnormal return of –19.6%. As a group, they underperform not-flagged low-accrual companies by 26.0%.

Given *M*-score's high correlation with accruals (Pearson correlation of 0.452 in Table 2), we conducted more-detailed tests to assess the joint ability of accruals and *M*-score to predict returns. Panel A of **Table 6** reports average size-adjusted returns when companies are sorted independently into quintiles by both accruals and *M*-score. Panels B and C report the results of nested (i.e., sequential) sorts. The strong correlation observed in Table 2 is apparent in Panel A: Approximately 25% of the sample observations reside in 2 of the 25 portfolios (upper left and lower right).

Again, we can see that *M*-score is particularly effective in predicting returns among low-accrual companies. The positive returns for low-accrual companies are concentrated among the low-*M*-score companies. Low-accrual companies with low *M*-scores generate returns of 6.6%. In contrast, low-accrual companies with high *M*-scores have very negative returns (–13.2%). In the lowest-accrual quintile, companies with low *M*-scores outperform high-*M*-score companies by 19.8 pps. In quintile 2, low-*M*-score companies outperform by 10.5 pps. But accruals do not distinguish companies in any of the *M*-score quintiles. The only exception is the high-*M*-score quintile, in which the high-accrual companies *outperform* the low-accrual companies.

Panel B of Table 6 further isolates the effect of accruals and *M*-score by sorting companies on *M*-score within each accrual quintile. This approach allows us to "spread out" the variation in *M*-score across companies that have relatively similar accrual rankings. Among low-accrual (income-decreasing) companies, low-*M*-score companies outperform high-*M*-score companies by 9.6 pps. Among high-accrual companies, the spread is smaller: Low-*M*-score companies outperform high-*M*-score companies by 8.0 pps. Returns are also significant for the second and fourth accrual quintiles. Interestingly, companies with large differences in accruals have no differences in returns when *M*-score is extremely high or low, but there are differences in returns for the three intermediate quintiles.

Panel C of Table 6 first sorts on *M*-score and then on accruals. The results are consistent across all *M*-score portfolios: Extreme differences in accruals do not result in differences in returns once companies are sorted on *M*-score. The only exception is the fourth quintile. In contrast, low-*M*-score companies outperform high-*M*-score companies across all accrual sorts, and the spread in returns is strikingly consistent. Overall, Table 6 confirms the findings in Panel E of Table 5 (as well as the evidence in Table 3) and demonstrates that *M*-score dominates accruals as a predictor of future returns.

The Incremental Usefulness of Individual *M*-Score Components. The evidence presented thus far indicates that *M*-score has significant ability to predict one-year-ahead cross-sectional returns. Our results show that this predictive power does not come from its correlation with value, momentum, size, accruals, or short interest.

Table 5. Size-Adjusted Returns to Decile Portfolios Conditional on *M*-Score

	Full Sample		Not Flagged		Flagged		Not-Flagged Stocks less Flagged Stocks
Portfolio	*N*	BHSAR	*N*	BHSAR	*N*	BHSAR	(pps)
A. MVE portfolios							
1 (low)	4,176	2.6%**	3,382	4.6%***	794	–5.8%*	10.4***
2	4,305	2.7**	3,405	6.4***	900	–11.5***	17.9***
3	4,297	3.5***	3,370	5.3***	927	–2.9	8.2***
4	4,385	0.9	3,444	3.6***	941	–8.9***	12.5***
5	4,302	1.6	3,455	3.9***	847	–7.5***	11.3***
6	4,403	0.7	3,642	3.1***	761	–10.6***	13.7***
7	4,402	1.0	3,682	3.0***	720	–9.0***	12.0***
8	4,432	0.5	3,830	1.6**	602	–6.5**	8.1***
9	4,274	0.6	3,745	1.3*	529	–4.2	5.5**
10 (high)	4,558	0.1	4,178	0.9	380	–9.6***	10.5***
Low – high (pps)		2.5*		3.7**		3.8	
B. Book-to-market portfolios							
1 (low)	3,105	–4.0%***	2,231	–2.0%	874	–9.4%***	7.4**
2	4,526	–1.8*	3,502	1.0	1,024	–11.4***	12.3***
3	4,429	0.6	3,553	3.0***	876	–9.3***	12.4***
4	4,493	0.9	3,686	2.4***	807	–6.3**	8.7***
5	4,562	2.0**	3,823	2.9***	739	–2.9	5.8**
6	4,337	2.8***	3,675	4.3***	662	–5.9**	10.2***
7	4,359	1.1	3,746	2.2***	613	–5.3**	7.5***
8	4,336	3.8***	3,780	4.8***	556	–3.0	7.9***
9	4,474	3.1***	3,888	4.9***	586	–8.5***	13.4***
10 (high)	4,913	4.0***	4,249	6.4***	664	–11.7***	18.1***
High – low (pps)		8.0***		8.4***		–2.3	
C. Momentum portfolios							
1 (low)	4,502	–3.0%**	3,141	3.0%*	1,361	–16.8%***	19.8***
2	4,386	–2.2**	3,519	0.1	867	–11.9***	12.0***
3	4,423	–2.0**	3,730	–0.2	693	–11.6***	11.4***
4	4,324	–0.4	3,736	0.5	588	–6.2***	6.7***
5	4,340	0.9	3,800	1.9**	540	–6.1**	7.9***
6	4,270	2.5***	3,770	3.5***	500	–4.4*	7.9***
7	4,365	2.4***	3,814	3.8***	551	–6.7**	10.4***
8	4,330	3.7***	3,695	5.4***	635	–6.1***	11.5***
9	4,134	5.3***	3,489	6.8***	645	–2.4	9.2***
10 (high)	4,460	7.2***	3,439	8.5***	1,021	2.7	5.8
High – low (pps)		10.2***		5.5**		19.5***	
D. SIR portfolios							
1 (low)	4,264	2.5%**	3,652	4.2%***	612	–7.7%***	11.9***
2	4,046	4.5***	3,407	5.0***	639	1.8	13.2
3	3,987	1.9**	3,419	3.3***	568	–6.3**	9.6***
4	4,065	3.6***	3,470	5.8***	595	–9.3***	15.2***
5	4,152	3.0***	3,576	4.0***	576	–2.9	6.9**
6	4,320	1.6*	3,665	3.5***	655	–8.7***	12.2***
7	4,338	2.4**	3,666	4.4***	672	–8.7***	13.1***
8	4,822	–0.1	4,012	1.3	810	–7.0**	8.3***
9	4,809	–0.7	3,805	1.7	1,004	–9.6***	11.3***
10 (high)	4,731	–3.3***	3,461	0.0	1,270	–12.1***	12.1***

(continued)

Table 5. Size-Adjusted Returns to Decile Portfolios Conditional on *M*-Score (continued)

	Full Sample		Not Flagged		Flagged		Not-Flagged Stocks less Flagged Stocks
Portfolio	*N*	BHSAR	*N*	BHSAR	*N*	BHSAR	(pps)
Low – high (pps)		5.8***		4.2***		4.4	
E. Accrual portfolios							
1 (low)	4,339	3.5%***	3,868	6.3%***	471	–19.6%***	26.0***
2	4,372	3.9***	3,979	5.2***	393	–9.8**	15.0***
3	4,404	3.0***	4,052	3.8***	352	–5.6	9.4***
4	4,391	2.5***	4,040	3.4***	351	–7.3**	10.7***
5	4,228	3.4***	3,834	4.2***	394	–4.6	8.8***
6	4,218	1.0	3,819	1.4*	399	–2.8	4.2
7	4,173	1.0	3,684	2.2**	489	–8.0***	10.2***
8	4,369	1.2	3,705	2.1**	664	–3.8	5.9**
9	4,466	–0.3	3,389	1.7	1,077	–6.4***	8.1***
10 (high)	4,574	–4.8***	1,763	0.6	2,811	–8.1***	8.7***
Low – high (pps)		8.3***		5.7**		–11.5***	

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Figure 1. MVE Portfolios

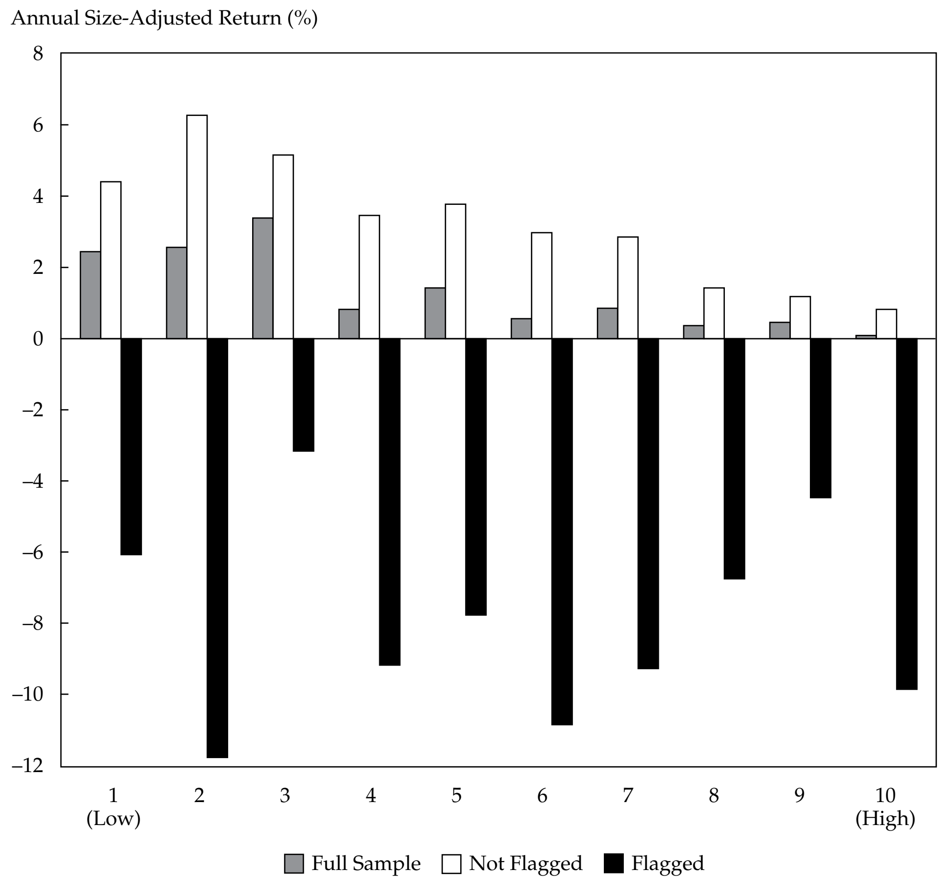
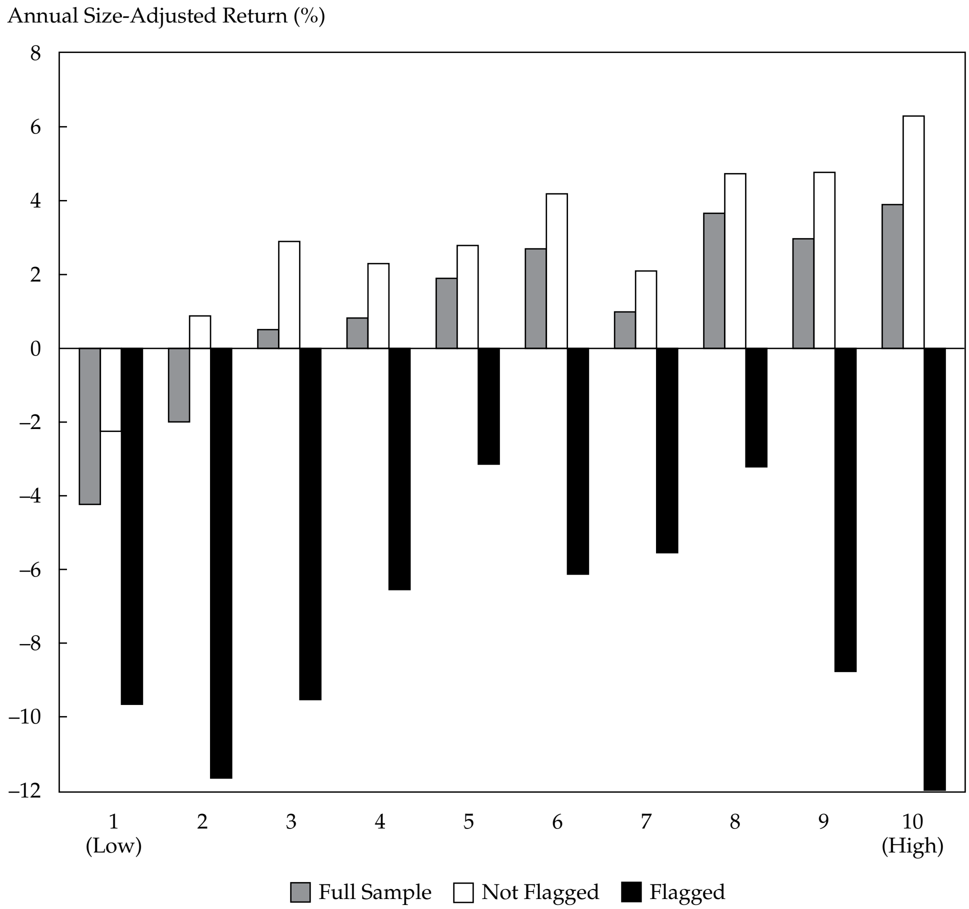


Figure 2. Book-to-Market Portfolios



Here, we present more-direct evidence on the nature of the information contained in *M*-score.

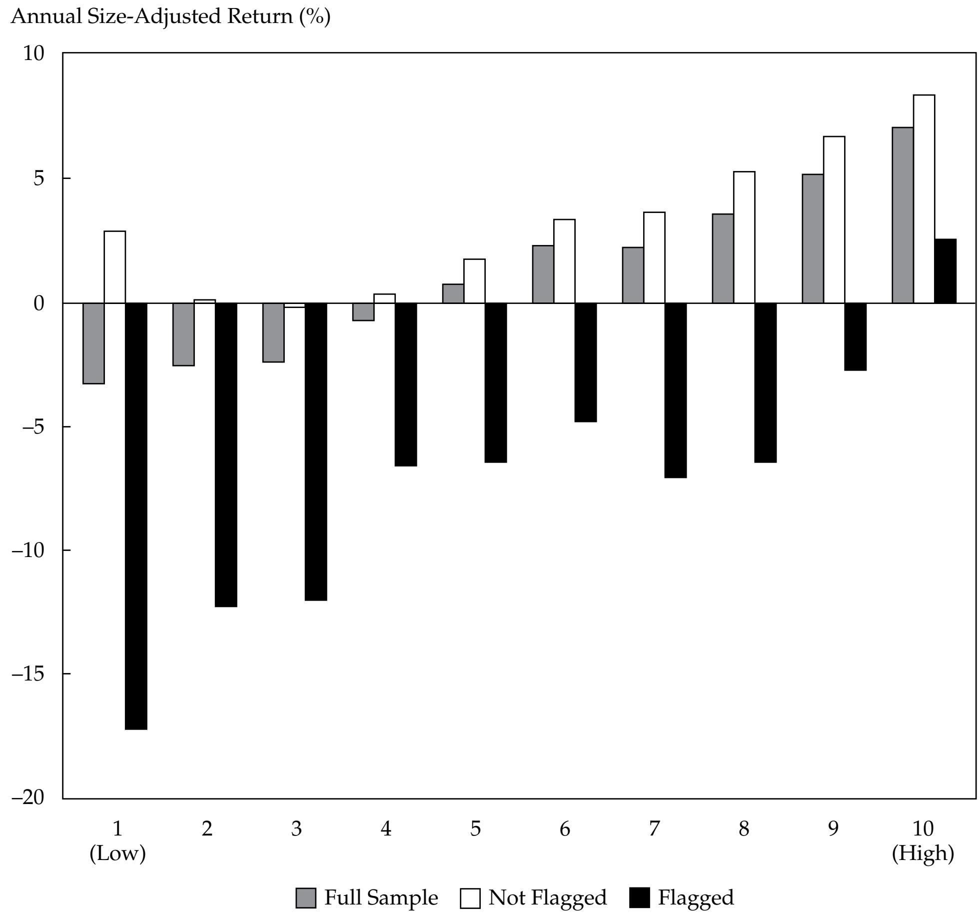
Table 7 compares the means for each of the seven components of the Beneish model besides accruals in various subpopulations of our sample. To construct this table, we independently sorted companies into quintiles according to accruals and *M*-score (as in Panel A of Table 6). We have seen from Table 6 that the Beneish model is particularly effective in separating winners from losers among low-accrual companies. We can study Table 7, which focuses sharply on companies in the upper right and upper left corners of Panel A of Table 6, to better understand which component of the model is contributing to the efficacy of *M*-score for return prediction among low-accrual companies.

The first two rows of Table 7 report the means for each variable for companies in the lowest-accrual quintile that are also in either the highest- or the lowest-*M*-score quintile. We then subtracted the mean of the high-*M*-score companies from the mean of the low-*M*-score companies, and we report the results in row 3. As expected, we found that high-*M*-score companies have higher means for the variables entering the model with a positive coefficient (DSR, AQI, GMI, SGI, DEPI) and lower means for the variables entering the model with a negative coefficient (SGAI, LEVI). On their own, these findings are not particularly informative, because we formed the two portfolios on the basis of *M*-score.

For comparison, we used a difference-in-difference test design. Specifically, Table 6 shows that *M*-score has little incremental predictive power for companies in the three high-accrual quintiles (3, 4, and 5). Exploiting this fact, we first grouped the companies in these three accrual quintiles and then further separated them according to their *M*-scores. Once again, we computed the mean for the same seven variables and report the results in rows 4–6 of Table 7. Finally, in the bottom row, we report descriptive statistics and the result of a statistical significance test for the difference-in-difference test across the two high – low *M*-score portfolios.

These results show that, compared with their counterparts in rows 4–6 of the table, the high-*M*-score companies in rows 1–3 had significantly higher sales growth (SGI), change in asset quality (AQI), and increase in leverage (LEVI). In other

Figure 3. Momentum Portfolios



words, the incremental predictive power of M-score among low-accrual companies is driven largely by these three factors.[21] Interestingly, all three factors are primarily indicators of a *predisposition to misstate earnings* rather than the result of the misstatement itself.

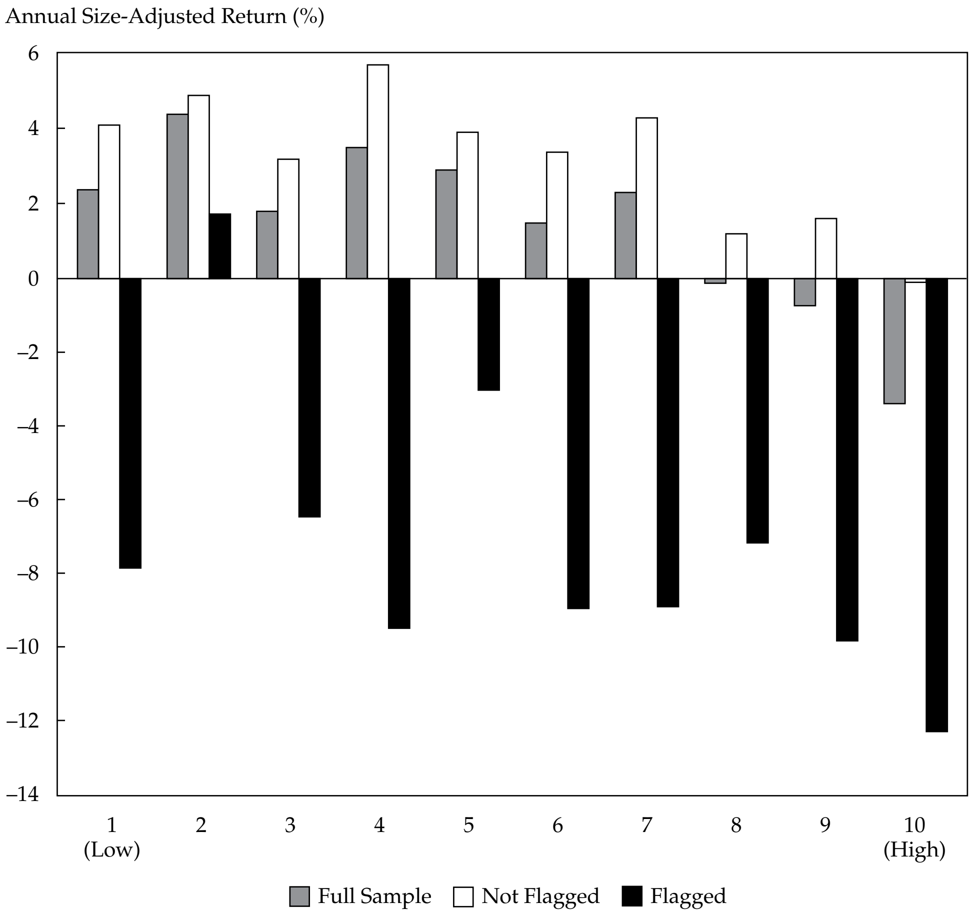
***M*-Score and the Persistence of Earnings Components.** Many researchers and practitioners have speculated that the accrual variable predicts returns because it provides information about earnings quality that market participants fail to use fully. We pursued this line of reasoning and examined the extent to which M-score is an indicator of a company's earnings quality. Specifically, we examined whether and how M-score might help us forecast companies' future earnings.

The notion that M-score might be useful in predicting future earnings is intuitive. As we have seen, high-M-score companies are fast-growing in terms of top-line sales. Moreover, they face operating environments that are becoming more challenging (e.g., declining margins and decreasing asset quality). Finally, these companies have adopted more-aggressive accounting practices in the most recent reporting period (higher receivables-to-sales, more income-inflating accruals, lower depreciation expense). Future earnings for these companies will be lower (i.e., current earnings will be less persistent) if either of two conditions occurs: (1) Current-period accounting distortions are corrected or reversed, or (2) the company succumbs to the difficult economic circumstances that are beginning to appear in its financial results.

We focused on the incremental ability of M-score to predict the persistence of companies' earnings. Prior studies have demonstrated that the cash flow component of earnings is more persistent than the accruals component (see, e.g., Sloan 1996; Richardson et al. 2005). We extended this analysis by exploring the differential persistence of one-year-ahead accruals for high- and low-M-score companies. If the forensic accounting principles that underpin this model are useful in separating companies with relatively high- or low-quality earnings, that fact should be evidenced in a difference in the persistence of future accounting accruals.

Specifically, we predicted that *income-increasing* accruals for companies with high (low) M-score

Figure 4. SIR Portfolios



should be less (more) persistent and *income-decreasing* accruals for companies with high (low) M-score should be more (less) persistent. In other words, for companies whose accruals component *increases* current-year income, we would expect higher M-score to be associated with decreased accrual persistence (leading to lower income next year). Conversely, for companies whose accruals component *decreases* current-year income, we would expect higher M-score to increase accrual persistence (leading to lower income next year).

To examine whether M-score contains such incremental information, we estimated the following relation between future-earnings and current-earnings components:

$$\begin{aligned}\text{EARN}_{t+1} = a_0 &+ a_1\text{CFO}_t \\ &+ a_2\text{ACCPOS}_t + a_3\text{ACCNEG}_t \\ &+ a_4\left(\text{ACCPOS}_t\right)\left(\text{SPM}_t\right) \\ &+ a_5\left(\text{ACCNEG}_t\right)\left(\text{SPM}_t\right) \\ &+ a_6\text{SPM}_t + e_{t+1},\end{aligned} \tag{3}$$

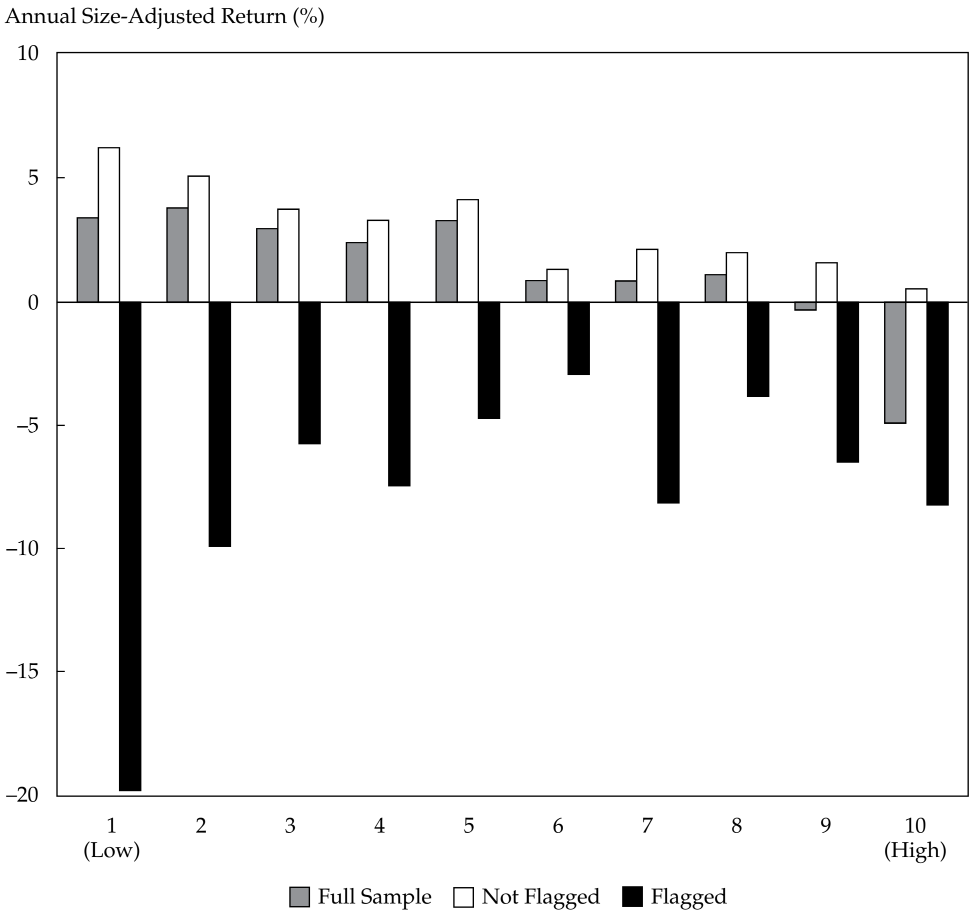
where

EARN = operating earnings before depreciation
CFO = cash flows from operations
ACCPOS = working-capital accruals when positive and zero otherwise
ACCNEG = working-capital accruals when negative and zero otherwise
SPM = M-score ranked into deciles and scaled to range from 0 (lowest M-score) to 1 (highest M-score)

All earnings and earnings components are deflated by average assets in year t.[22]

Table 8 provides a frame of reference. The first two columns report that the current year's earnings have a persistence coefficient of regression of 0.894 and that cash flows are more persistent than accruals (1.042 and 0.573). The result for cash flows is similar to what Sloan (1996) documented (0.860), but the persistence of accruals is smaller than his (0.765), largely owing to the fact that we used working-capital accruals and thus excluded

Figure 5. Accrual Portfolios



depreciation. The third column partitions accruals into positive and negative samples. Our results show that, consistent with Beneish and Vargus (2002), the persistence of both positive and negative accruals is significantly lower than that of cash flows.

The last column reports the results of tests examining the persistence of accruals conditional on *M*-score. We again found strong persistence for CFO (coefficient = 1.074). The coefficients on ACCPOS and ACCNEG reflect the persistence of positive and negative accruals of companies in the lowest-*M*-score decile (SPM = 0). For example, the coefficient of 0.982 on ACCPOS means that positive accruals for low-*M*-score companies have exceptionally high persistence, suggesting that they have a much higher likelihood of repeating. Conversely, the coefficient on ACCNEG is 0.282, indicating that for low-*M*-score companies, negative (or income-deflating) accruals are much less likely to repeat.

The coefficients on (ACCPOS)(SPM) and (ACCNEG)(SPM) capture the incremental effect of *M*-score on accrual persistence as we move from the lowest decile (SPM = 0) to the highest decile (SPM = 1). (ACCPOS)(SPM) is negative and significant (coefficient = –0.259), suggesting that for companies in the highest-*M*-score decile (SPM = 1), *income-increasing* accruals are much less likely to repeat (estimated coefficient = 0.982 – 0.259 = 0.723). For comparison, recall that the income-increasing accruals for companies in the lowest-*M*-score decile are highly persistent (0.996).

Similarly, (ACCNEG)(SPM) is positive and significant (coefficient = 0.699), indicating that for companies in the highest-*M*-score decile, *income-decreasing* accruals are much more likely to persist (estimated coefficient = 0.282 + 0.699 = 0.981). In other words, for high-*M*-score companies, any income-decreasing accruals this year will have a higher probability of repeating next year, leading to lower future earnings. Finally, SPM itself is positive and significant (coefficient = 0.031), implying that high-*M*-score companies (i.e., high-growth companies) have higher future earnings after controlling for differences in accrual persistence. This finding may reflect the fact that high-*M*-score companies, as a group, are continuing to grow more rapidly than the control companies; alternatively, it is possible

Table 6. Size-Adjusted Returns to Accrual and *M*-Score Quintile Portfolios

	1 (Low *M*-Score)		2		3		4		5 (High *M*-Score)		Low – High
Portfolio	*N*	BHSAR	*N*	BHSAR	*N*	BHSAR	*N*	BHSAR	*N*	BHSAR	(pps)
A. Independent sorts											
1 (low accruals)	5,521	6.6%***	1,183	4.0%**	553	8.8%**	490	–2.8%	964	–13.2%***	19.8***
2	2,353	4.9***	3,388	4.3***	1,365	1.5	837	1.2	852	–5.6***	10.5***
3	687	–0.1	2,741	3.6***	2,666	2.3**	1,431	2.1	921	–0.5	0.3
4	242	–1.2	1,019	4.2**	3,019	2.9***	2,838	0.9	1,424	–4.2**	3.0
5 (high accruals)	75	–2.9	207	6.2	786	2.6	3,148	–1.0	4,824	–4.7***	1.9
Low – high (pps)		9.5		–2.2		6.2		–1.8		–8.4***	
B. M-*score sorted within accrual portfolios*											
1 (low accruals)	1,735	5.9%***	1,747	8.6%***	1,743	5.9%***	1,747	1.9%	1,739	–3.7%**	9.6***
2	1,752	4.2***	1,762	6.3***	1,763	2.3**	1,762	3.2**	1,756	–2.2	6.4***
3	1,681	3.7***	1,694	2.1**	1,690	2.2**	1,694	2.1	1,687	0.9	2.8
4	1,703	3.3**	1,712	1.9*	1,709	3.1***	1,712	1.4	1,706	–4.1**	7.4***
5 (high accruals)	1,800	1.7	1,814	0.7	1,808	–3.3**	1,814	–5.5***	1,804	–6.3***	8.0***
Low – high (pps)		4.2*		7.9***		9.2***		7.4***		2.6	
C. Accruals sorted within M-*score portfolios*											
	1 (Low Accruals)		2		3		4		5 (High Accruals)		
1 (low *M*-score)	1,767	5.7%**	1,778	9.0%***	1,780	4.8%***	1,778	5.2%***	1,774	2.2%*	3.5
2	1,701	4.9***	1,710	4.5***	1,713	3.5***	1,710	3.9***	1,702	3.6***	1.3
3	1,671	4.2**	1,680	2.9***	1,681	2.0	1,680	2.9**	1,675	2.1*	2.1
4	1,741	0.4	1,751	0.5	1,755	1.1	1,751	2.9*	1,743	–3.8***	4.2**
5 (high *M*-score)	1,790	–6.3***	1,800	–5.0***	1,800	–2.7	1,800	–6.2***	1,794	–5.7***	0.6
Low – high (pps)		12.0***		14.0***		7.5***		11.4***		7.9***	

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

Table 7. Comparing *M*-Score Components for High- and Low-Accrual Companies

Accrual Quintile	*M*-Score Quintile	*N*	DSR	GMI	AQI	SGI	SGAI	DEPI	LEVI
Lowest	Lowest	5,521	0.873	1.031	0.948	1.073	1.080	0.966	1.159
	Highest	964	1.428	1.169	5.211	2.355	0.936	1.071	1.062
	Highest – lowest		0.555	0.138	4.263	1.282	–0.143	0.105	–0.097
3, 4, 5	Lowest	1,004	0.756	0.906	0.856	0.975	1.111	1.010	1.211
	Highest	7,169	1.341	1.072	2.299	1.554	0.952	1.104	0.989
	Highest – lowest		0.586	0.166	1.443	0.578	–0.159	0.094	–0.222
Low accruals – high accruals			–0.030	–0.028	2.820***	0.704***	0.016	0.010	0.124***

***Significant at the 1% level.

Table 8. Regression of Future Earnings on Current-Period Earnings Components

	Model 1	Model 2	Model 3	Model 4
Intercept	0.017**	0.003	–0.012**	–0.024***
EARN	0.894***			
CFO		1.042***	1.076***	1.074***
ACC		0.573***		
ACCPOS			0.845***	0.982***
ACCNEG			0.424***	0.282***
(ACCPOS)(SPM)				–0.259*
(ACCNEG)(SPM)				0.699***
SPM				0.031***
Adjusted R^2	39.5%	42.9%	43.3%	43.9%
N = 41,053				

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

that their inflated earnings in the current year do not fully reverse one year ahead.

In sum, we found that *M*-score is incrementally informative about the persistence of the accruals component of earnings. First, the one-year-ahead persistence of *income-increasing* accruals is significantly lower for high-*M*-score companies. Second, any *income-decreasing* accruals reported in the current year are much more likely to persist for high-*M*-score companies than for low-*M*-score companies. Both results show that *M*-score is useful in predicting future earnings. This finding is consistent with *M*-score's possessing the ability to either (1) anticipate the reversal of transitory distortions in current-year reported accruals or (2) predict worsening economic conditions that will manifest themselves in subsequent accruals.

Predicting Short-Window Earnings Announcement Returns. Finally, to better understand the information content of *M*-score, we examined abnormal returns in three-day windows centered on subsequent earnings announcements. The main purpose of this test was to further differentiate between the market-inefficiency and risk-based explanations for the predictive power of *M*-score. If *M*-score's predictive power is due to a delayed market response to current information about future earnings realizations—*and* this misperception is corrected when further information about future earnings is released—then subsequent abnormal returns should cluster around the future announcements of earnings news. Conversely, if *M*-score's predictive power is due to omitted risk variables, then we should not observe higher abnormal returns concentrated around the release of subsequent earnings news.[23]

Table 9 reports the cumulative abnormal returns (CARs) around the next four earnings announcements after sorting companies into *M*-score-based portfolios. We added companies to portfolios on the first day of the fifth month following the end of the fiscal year. CAR denotes the cumulative raw return over Days –1, 0, and 1 relative to the earnings announcement date minus the cumulative return to the benchmark size portfolio to which the company belongs. To construct Panel A, we sorted companies into flagged and not-flagged categories on the basis of *M*-score. We report the short-window announcement period CAR for each of the next four quarters, as well as the total CAR for all four quarterly announcements.

Panel A shows that not-flagged companies earn a positive CAR over each of the next four quarters, whereas flagged companies experience, on average, negative returns. The quarter-by-quarter results show that short-window returns are reliably more positive for not-flagged (low-*M*-score) companies every quarter. The last column shows that the difference in total CAR over the next four quarters is 2.51 pps. In other words, approximately one-quarter (25%) of the annual hedge return to this simple *M*-score strategy is earned in the 12 trading days (5% of all trading days) around the next four earnings announcements. This evidence is consistent with a delayed market reaction to earnings-related news contained in *M*-score and inconsistent with risk-based explanations.

Panels B and C report average earnings announcement returns to alternative hedge strategies based on *M*-score deciles. To construct this panel, we first sorted companies into deciles on the basis of their current *M*-scores and the decile cutoffs from the prior year's *M*-score distribution. We then calculated the CAR for each decile portfolio (Panel B), as well as for alternative hedge portfolios (Panel C). In Panel C, we started with the most extreme portfolios (deciles 1 and 10) and progressively added less extreme *M*-score companies to the long and short positions.

Panel B shows that in each of the next four quarters, earnings announcement returns for low-*M*-score companies (deciles 1–5) are consistently more positive than those for high-*M*-score companies (deciles 6–10). Over our sample period, companies generally received positive announcement period returns. However, companies in the top two *M*-score deciles experienced consistently negative earnings announcement returns over the next two quarters.

Table 9. Earnings Announcement Returns by Quarter

	Quarter $t+1$		Quarter $t+2$		Quarter $t+3$		Quarter $t+4$		Total CAR for 4 Quarters
	N	CAR	*N*	CAR	*N*	CAR	*N*	CAR	
A. Earnings announcement returns by quarter for flagged and not-flagged companies									
Not flagged	35,423	0.33%***	34,679	0.12%**	33,723	0.17%***	29,968	0.53%***	1.03%***
Flagged	7,461	–0.45***	7,296	–0.39***	7,041	–0.74***	6,012	0.01	–1.48
Difference (pps)		0.78***		0.50***		0.91***		0.51***	2.51***
B. Earnings announcement returns by quarter and M*-score decile*									
Decile									
1	4,452	0.20%	4,354	0.29%*	4,213	0.03%	3,674	0.61%***	1.01%***
2	4,269	0.20	4,179	0.20	4,076	0.13	3,641	0.66***	1.09***
3	4,175	0.45***	4,079	0.23*	3,976	0.24*	3,590	0.53***	1.37***
4	4,238	0.28*	4,143	0.22*	4,048	0.44***	3,636	0.70***	1.52***
5	3,972	0.44***	3,881	0.17	3,772	0.43***	3,367	0.50***	1.43***
6	4,287	0.36***	4,206	0.01	4,097	0.27*	3,673	0.72***	1.24***
7	4,233	0.59***	4,139	0.11	4,021	0.15	3,574	0.40***	1.17***
8	4,390	0.16	4,316	–0.14	4,193	–0.15	3,668	0.28*	0.11
9	4,370	–0.13	4,269	–0.18	4,118	–0.52***	3,563	0.04	–0.75*
10	4,498	–0.53***	4,409	–0.57***	4,250	–0.83***	3,594	–0.04	–1.90***

C. Earnings announcement returns to alternative M*-score-based hedge portfolios*

Long	Short	Quarter $t+1$ CAR	Quarter $t+2$ CAR	Quarter $t+3$ CAR	Quarter $t+4$ CAR	
Decile 1	Decile 10	0.73%***	0.86%***	0.85%***	0.65%**	2.91%***
1–2	9–10	0.53***	0.62***	0.75***	0.64***	2.34***
1–3	8–10	0.27**	0.44***	0.47***	0.43***	2.01***
1–4	7–10	0.20**	0.40***	0.43***	0.35***	1.61***
1–5	6–10	0.23**	0.38***	0.47***	0.32***	1.33***

*Significant at the 10% level.
**Significant at the 5% level.
***Significant at the 1% level.

The first row of Panel C shows that when we have a long position in decile 1 (low-*M*-score) companies and short decile 10 (high-*M*-score) companies, we generate a CAR of 0.65%–0.86% over each of the next four quarters. The total CAR over the next four quarters is 2.91%, which is slightly higher than that for the strategy reported in Panel A. The following four rows show that the hedge returns are not sensitive to the decile cutoffs. As expected, the average hedge return decreases as we expand the number of companies included in the strategy. However, the CARs for these hedge strategies are all reliably positive. Moreover, the earnings announcement returns are, once again, disproportionately large relative to total annual returns for every hedge portfolio.

Conclusion

Fraudulent financial reporting imposes huge costs on financial markets. For example, shareholders of the companies listed in Appendix B collectively lost more than $180 billion when accounting "irregularities" were announced.[24] Perhaps even more important than the losses of investor wealth are the large welfare costs imposed by fraudulent financial reporting when resources are misdirected from their most productive use. These accounting misrepresentations increase transaction costs by eroding investor confidence in the integrity of the capital markets. In recent years, we have seen how accounting misrepresentations have triggered action by regulators, who impose (often costly) regulation on companies and markets. In short, when it comes to reporting frauds, many must pay for the transgressions of a relative few.

Efforts to combat accounting fraud involve both public and private initiatives. On the one hand, accounting and security market regulators can help curb the practice through legislation and enforcement actions. On the other hand, private parties, such as more sophisticated investors, play a role by identifying companies that are likely to have manipulated earnings and holding those companies accountable through market-based disciplining mechanisms.

In our study, we explored the implications of an earnings manipulation detection model for equity investors. Using the Beneish (1999a) model, which was estimated with data for 1982–1988 and whose holdout sample performance was assessed for 1989–1992, we showed that forensic accounting has significant out-of-sample ability not only to detect fraud but also to predict stock returns. Moreover, we found evidence that the efficacy of the model derives substantially from its ability to predict the likely persistence (or reversal) of the accruals component of current-year earnings.

A key feature of the Beneish model is its focus on both the results of aggressive accounting and the managerial predisposition to undertake such action in the first place. As Schrand and Zechman (2012) observed in their detailed analysis of 49 fraud cases, sometimes a company's initial misstatement reflects not so much a deliberate intent to deceive as simply an optimistic managerial bias. Our evidence is consistent with this observation in that much of the incremental predictive power of the model derives from variables that indicate deteriorating fundamentals in fast-growing companies rather than aggressive accounting per se. These findings point to new directions for future research on earnings quality. We hope that these results will spur further work in the area of forensic accounting.

We end on a cautionary note. Looking ahead, it appears likely that as the methods for earnings manipulation detection become more sophisticated, so too will the techniques of the perpetrators. For example, the evidence in Kama and Melumad (2011) shows that in recent years (since the Sarbanes–Oxley Act), U.S. companies seem to have adopted new methods—such as the strategic factoring of receivables—to mask the effect of their earnings manipulation. In the case of the Beneish model, the factoring of receivables will directly affect the usefulness of the DSR variable. Over time, one should reasonably expect evolving adaptations of this nature to diminish the overall efficacy of the model with respect to return prediction. Our weaker results in the second subperiod suggest that this process has already begun.

An earlier version of this article was presented at the June 2007 Corporate Ethics and Investing Conference of the Society of Quantitative Analysts, the May 2008 LSV-Penn State Conference, and workshops at Cornell University, Indiana University, the University of Maryland, the University of Notre Dame, the University of Arizona, and Stanford University. For helpful discussions and comments, we thank the participants at those workshops as well as C. Harvey, P. Hribar, J. Salamon, C. Trzincka, S. Bhojraj, D. Givoly, J. Lakonishok, M. Lang, A. Leone, R. Morton, B. Swaminathan, P. von Hippel, N. Yehuda, and X. Zhang. We also thank M. Drake, L. Rees, and E. Swanson for kindly sharing their short-selling data with us.

This article qualifies for 1 CE credit.

Appendix A. The Probability of Manipulation

In this appendix, we present a detailed description of the Beneish (1999a) model, including details of the sample, and the intuition behind each of the eight variables used in its estimation.

Estimating *M*-Score

We estimated the probability of manipulation (*M*-score) by using the following (unweighted) probit model, as in Beneish (1999a):

$$\begin{aligned} M\text{-score} = {} & -4.84 + 0.920(\text{DSR}) + 0.528(\text{GMI}) \\ & + 0.404(\text{AQI}) + 0.892(\text{SGI}) \\ & + 0.115(\text{DEPI}) - 0.172(\text{SGAI}) \\ & + 4.679(\text{Accruals}) - 0.327(\text{LEVI}). \end{aligned}$$

The model features eight accounting-based variables, each of which is so constructed that higher values are associated with a greater probability of earnings manipulation. A description of the variables and the rationale for their inclusion are provided in **Exhibit A1**.

Intuition behind the Eight Variables

Broadly speaking, the profile of a "typical earnings manipulator," as defined by Beneish (1999a), is a company that is (1) growing extremely quickly, (2) experiencing deteriorating fundamentals (as evidenced by a decline in asset quality, eroding profit margins, and increasing leverage), and (3) adopting aggressive accounting practices (e.g., receivables growing much faster than sales, large income-inflating accruals, decreasing depreciation expense).

More specifically, the model consists of eight financial ratios that capture either financial statement distortions that result from earnings manipulation (DSR, AQI, DEPI, Accruals) or a predisposition to engage in earnings manipulation owing to economic conditions (GMI, SGI, SGAI, LEVI).[25] Descriptive statistics for these ratios appear in **Table A1**. Each ratio is so constructed that a higher number increases the likelihood of manipulation.

The following is a summary of these variables. Not all eight are individually important, but collectively they create a "composite sketch," or profile, of a potential earnings manipulator.

Exhibit A1. Description of Variables and Rationale for Inclusion

Variable	Description (numbers in brackets are Compustat codes)	Rationale
DSR	(Receivables$_t$ [2]/Sales$_t$ [12])/(Receivables$_{t-1}$/Sales$_{t-1}$)	Captures distortions in receivables that can result from revenue inflation
GMI	Gross margin$_{t-1}$/Gross margin$_t$, where Gross margin = 1 – Costs of goods sold [8]/Sales	Deteriorating margins predispose companies to manipulate earnings
AQI	[1 – (PPE$_t$ + CA$_t$)/TA$_t$]/[1 – (PPE$_{t-1}$ + CA$_{t-1}$)/TA$_{t-1}$], where PPE is net [8], CA is current assets [4], and TA is total assets [6]	Captures distortions in other assets that can result from excessive expenditure capitalization
SGI	Sales$_t$ [12]/Sales$_{t-1}$	Managing the perception of continuing growth and capital needs predisposes growth companies to manipulate sales and earnings
DEPI	Depreciation rate$_{t-1}$/Depreciation rate$_t$, where Depreciation rate equals Depreciation [14–65]/(Depreciation + PPE [8])	Captures declining depreciation rates as a form of earnings manipulation
SGAI	(SGA$_t$ [189]/Sales$_t$ [12])/(SGA$_{t-1}$/Sales$_{t-1}$)	Decreasing administrative and marketing efficiency (larger fixed SGA expenses) predisposes companies to manipulate earnings
Accruals[a]	(Income before extraordinary items [18] – Cash from operations [308])/Total assets$_t$ [6]	Captures where accounting profits are not supported by cash profits
LEVI	Leverage$_t$/Leverage$_{t-1}$, where Leverage is calculated as debt to assets: (5 + 9)/6	Increasing leverage tightens debt constraints and predisposes companies to manipulate earnings

Notes: To gain insight into the relative importance of each individual variable, Beneish (1999a) re-estimated this model 100 times by using 100 random estimation samples. At the 5% level, DSR and SGI were significant in all 100 estimations, Accruals was significant in 95 of the 100 estimations, and GMI and AQI were significant in 84 of the 100 estimations. DEPI, SGAI, and LEVI were significant in only 18, 12, and 2 estimations, respectively (Beneish 1999a, Table 4).

[a]Beneish (1999a) used a total accruals variable that is computed slightly differently but yields similar results; before the current presentation of the statement of cash flows became effective (in 1987), few companies reported cash flow from operations. Our current implementation follows the evolution of this variable in the accruals literature.

- *Rapid sales growth.* A striking characteristic of the manipulator population is its rapid revenue growth. Table A1 shows that the mean sales growth rate for the sample of manipulators (SGI) is 1.581, indicating that these companies, on average, increased year-over-year sales by 58%, compared with 13.3% for an industry control group. Sales growth per se is not a negative, but in this context, a company's high-past-growth trajectory may increase its predisposition to engage in manipulative behavior.
- *Deteriorating fundamentals.* Several variables are designed to capture deteriorating economic conditions. Specifically, manipulators are hypothesized to have deteriorating gross margins (GMI) and increasing SG&A expenses (SGAI). Their debt-to-assets ratio is increasing (LEVI), and a greater proportion of their total assets reflects noncurrent and non-PP&E investments (i.e., they have increased their proportion of "soft assets," or AQI).
- *Aggressive accounting.* The year-over-year ratio of the manipulators' receivables-to-sales (DSR) shows that their receivables are growing rapidly as a percentage of sales (1.412 for manipulators versus 1.030 for the control group). This number indicates an unusual buildup in receivables despite the rapidly increasing sales (a sign of potential revenue inflation). DEPI indicates that manipulators tend to slow down their depreciation expense as a percentage of their gross PP&E. Finally, Accruals indicates that the reported accounting profits of the manipulators are less supported by cash profits than those of nonmanipulators (Accruals was income inflating to the tune of 4.9% of total assets, compared with 1.5% for the control companies).

In short, the manipulators, as a group, appear to be growth companies that are running into some potential problems.

Incidence of Manipulation

The distribution of manipulators over the sample period suggests an increasing frequency of SEC accounting and auditing enforcement actions over time (see **Table A2**).

The distribution of manipulators by two-digit SIC code suggests that the highest concentration is in business services (10 companies, 13.5%), followed by industrial products (7 companies, 9.5%) and both electronic manufacturing and wholesale trade–durable goods (5 companies, 6.8%). **Table A3** reports information on the incidence of manipulation by industry.

Appendix B. Performance of Model for High-Profile Fraud Cases over 1998–2002

Table B1 reports the 20 companies identified by Audit Integrity as the "highest-profile" fraud cases uncovered over 1998–2002.[26] We examined the probability of manipulation (*M*-score) for each company on the basis of financial statement information reported by the company during the period of alleged manipulation but prior to public discovery. We flagged companies as manipulators if their *M*-score exceeded –1.78 at any time during the period of alleged (or admitted) violation.

We computed *M*-score as follows: *M*-score = –4.84 + 0.920(DSR) + 0.528(GMI) + 0.404(AQI) + 0.892(SGI) + 0.115(DEPI) – 0.172(SGAI) + 4.679(Accruals) – 0.327(LEVI). DSR denotes the

Table A1. Potential Predictive Variables: Descriptive Statistics for the Sample of 74 Manipulators and 2,332 Industry-Matched Nonmanipulators, 1982–1992

	Manipulators		Controls		Wilcoxon-Z	Median
Characteristic	Mean	Median	Mean	Median	*p*-Value[a]	*p*-Value[a]
Days in receivables	1.412	1.219	1.030	0.995	0.001	0.001
Gross margin index	1.159	1.028	1.017	1.001	0.019	0.078
Asset quality index	1.228	1.000	1.031	1.000	0.035	0.824
Sales growth index	1.581	1.341	1.133	1.095	0.001	0.001
Depreciation index	1.072	0.977	1.007	0.972	0.346	0.638
SGA index	1.107	1.028	1.085	0.990	0.714	0.098
Leverage index	1.124	1.035	1.033	1.000	0.107	0.039
Accruals to total assets	0.049	0.026	0.015	0.012	0.001	0.018

[a]The Wilcoxon rank-sum test and the median test compare the distribution of sample companies' characteristics with the corresponding distribution for nonmanipulators. The reported *p*-values are two-tailed and indicate the smallest probability of incorrectly rejecting the null hypothesis of no difference.

Table A2. Distribution of Manipulators

	1981–1985	1986–1989	1990–1993	Total
No. of companies	8	35	31	74

Table A3. Manipulators by Two-Digit Industry

SIC	Industry Description	*N*	Share
01	Agricultural production	1	1.4%
10	Metal mining	1	1.4
13	Oil and gas extraction	1	1.4
15	General building contractors	1	1.4
20	Food and kindred products	1	1.4
22	Textile mill products	3	4.1
23	Apparel and other textile products	1	1.4
24	Lumber and wood products	1	1.4
27	Printing and publishing	2	2.7
28	Chemicals and allied products	4	5.4
30	Rubber and miscellaneous plastics products	1	1.4
34	Fabricated metal products	1	1.4
35	Industrial and related products	7	9.5
36	Electronic and other electric equipment	5	6.8
37	Transportation equipment	2	2.7
38	Instruments and related products	2	2.7
45	Transportation by air	1	1.4
47	Transportation services	1	1.4
48	Communications	1	1.4
49	Electric, gas, and sanitary services	3	4.1
50	Wholesale trade–durable goods	5	6.8
51	Wholesale trade–nondurable goods	1	1.4
52	Building materials and garden supplies	1	1.4
54	Food stores	1	1.4
56	Apparel and accessory stores	1	1.4
57	Furniture and home furnishings stores	3	4.1
58	Eating and drinking places	1	1.4
59	Miscellaneous retail	2	2.7
70	Hotels and other lodging places	1	1.4
73	Business services	10	13.5
75	Auto repair, services, and parking	2	2.7
78	Motion pictures	3	4.1
80	Health services	1	1.4
82	Educational services	2	2.7
	Total	74	100.0%

ratio of receivables to sales in year t divided by the same ratio in year $t - 1$. GMI denotes the ratio of gross margin to sales in period $t - 1$ divided by the same ratio in period t. SGA denotes the ratio of sales, general, and administrative expenses to sales in period t divided by the same ratio in period $t - 1$. SGI equals sales in year t divided by sales in year $t - 1$. DEPI denotes the ratio of depreciation to depreciable base in year $t - 1$ divided by the same ratio in year t. AQI equals all noncurrent assets other than PP&E as a percentage of total assets in year t divided by the same ratio in year $t - 1$. Accruals equals the ratio of the difference between income before extraordinary items and operating cash flows to average total assets. LEVI equals the ratio of long-term debt plus current liabilities to total assets in year t divided by the same ratio in year $t - 1$. Year flagged refers to the first year the company was flagged as a manipulator by the *M*-score model. Year discovered refers to the year during which the fraud was first publicly revealed in the business press. Market cap

Table B1. Highest-Profile Fraud Cases, 1998–2002

Company	Flagged as Manipulator	Year Flagged	Year Discovered	Market Cap Lost ($ billions)	Market Cap Lost (%)
Adelphia Communications	Yes	1999	2002	4.82	96.8
American International Group[a]	na				
AOL Time Warner	Yes	2001	2002	25.77	32.2
Cendant Corporation	Yes	1996	1998	11.32	38.1
Citigroup[a]	na				
Computer Associates International	Yes	2000	2002	7.23	36.4
Enron	Yes	1998	2001	26.04	99.3
Global Crossing[b]	Yes	1999	2002		
HealthSouth Corporation	No		2002	2.31	57.3
JDS Uniphase Corporation	Yes	1999	2001	32.49	61.0
Lucent Technologies	Yes	1999	2001	11.15	24.7
Motorola[c]	na				
Qwest Communications International	Yes	2000	2002	9.84	41.8
Rite Aid Corporation	Yes	1997	1999	2.83	59.1
Sunbeam Corporation	Yes	1997	1998	1.28	58.8
Tyco International	No		2002	37.55	58.2
Vivendi Universal	No		2002	1.28	27.9
Waste Management	Yes	1998	1999	20.82	63.6
WorldCom/MCI Group	No		2002	1.03	69.8
Xerox Corporation	No		2000	7.73	43.8
Mean				10.89	51.94
Median				8.79	57.75

na = not applicable.
[a]Financial company.
[b]Delisted owing to bankruptcy.
[c]Did not manipulate its earnings—only abetted Adelphia.

lost represents the change in market capitalization during the three months surrounding the month the fraud was announced (i.e., Months –1, 0, 1). Market cap lost (%) represents the market capitalization lost in the three months surrounding the fraud announcement month as a percentage of market capitalization at the beginning of Month –1.

Notes

1. This type of analysis is sometimes called *forensic accounting*. See Schilit (2010) for a list of case studies and Foster (1979) for similar evidence from an earlier era.
2. By using the originally published coefficients, we avoided data-snooping and peek-ahead issues. Although such weights are unlikely to be optimal for either fraud detection or return prediction, our goal was not so much to optimize these outcomes as to demonstrate the out-of-sample efficacy of forensic accounting techniques.
3. The term *earnings quality* has been used in academic literature to describe such seemingly divergent ideas as the ability of earnings to represent past performance faithfully and the ability of earnings to predict future performance more accurately. In our study, we used the term to describe its "persistence"—that is, "higher earnings quality" refers to earnings that are less transitory and more likely to persist (or endure) in the future. Consistent with common usage among analysts, higher earnings quality, by our definition, deserves a higher price-to-earnings multiple.
4. A similar pattern was observed with the original accrual anomaly documented by Sloan (1996), which had strong out-of-sample return predictive power for a number of years after publication.
5. See, for example, Hribar and Collins (2002); Fairfield et al. (2003); Hirshleifer, Hou, Teoh, and Zhang (2004); Richardson et al. (2005); Chan, Chan, Jegadeesh, and Lakonishok (2006); Cooper et al. (2008); Hirshleifer et al. (2012). Other related studies have examined whether different market constituents understand the implications of accruals (Bradshaw, Richardson, and Sloan 2001; Beneish and Vargus 2002; Collins, Gong, and Hribar 2003; Barth and Hutton 2004). See Zachs (2011, ch. 2) for a good summary of accruals-related research.
6. See, for example, Desai, Rajgopal, and Venkatachalam (2004) and Khan (2005). More recently, the focus has shifted to the limits of arbitrage and market-learning-related explanations (Mashruwala, Rajgopal, and Shevlin 2006; Lev and Nissim 2006; Li and Sullivan 2011; Green et al. 2011).

7. In our study, we used three broad, and not necessarily mutually exclusive, categories to help explain the main motivations behind the eight variables. In fact, some variables (e.g., days in receivables and asset quality) could be indicators of either economic challenges or aggressive accounting and probably contain elements of both.
8. The bankruptcy prediction models, for example, are concerned with a company's "distance from default" and feature variables that measure profitability, leverage, and overall health. The "contextual analysis" studies are concerned with indicators of improving fundamentals for different types of stocks.
9. In taking seriously the work of financial practitioners, Beneish (1999a) is similar in spirit to Lev and Thiagarajan (1993) and Abarbanell and Bushee (1997). Unlike those studies, however, the Beneish model was built specifically to detect both the propensity to commit accounting fraud and the effect of such activities.
10. The model gained widespread recognition when a group of MBA students at Cornell University posted the earliest warning about Enron's accounting manipulation score, using the Beneish (1999a) model a full year before the first professional analyst reports (Morris 2009). This episode in U.S. financial history is preserved in the Enron exhibit at the Museum of American Finance in New York City (www.moaf.org) and is also recounted in Gladwell (2009).
11. Beneish (1999a) estimated both an unweighted and a weighted probit regression. Our discussion refers to the unweighted probit model.
12. The issue of classification errors was less pertinent to our study because, for purposes of return prediction, *M*-score often appears as a continuous variable. Nevertheless, it seems to us that a cost ratio in the neighborhood of 20:1 (or 30:1) is reasonable because, to most active asset managers, the cost of erroneously including an earnings manipulator in a portfolio is much higher than the cost of erroneously omitting an innocent company.
13. This option was unavailable to Beneish (1999a) because the statement of cash flows became a required disclosure in the United States only in 1987; much of his sample pre-dates that requirement. As a practical matter, in the absence of major business acquisitions, the two methods yield very similar estimates.
14. Audit Integrity was an independent research company that was later acquired by the RiskMetrics Group and is now called GMI Analyst. We are not affiliated with the company and are unfamiliar with the details of its selection method.
15. In January 2002, the *Wall Street Journal* reported that in seizing e-mails at Arthur Andersen, the U.S. Congress found evidence that the firm's Chicago office had issued two "alerts" to the Houston office in the spring of 2001 with respect to earnings manipulation at Enron. The alerts came from a tailored version of the model that Beneish had estimated under a consulting contract with Arthur Andersen. See "Andersen Knew of 'Fraud' Risk at Enron—October E-Mail Shows Firm Anticipated Problems before Company's Fall," *Wall Street Journal* (25 January 2002):A3.
16. Although Lyon, Barber, and Tsai (1999) formed their reference portfolios yearly, we performed our sorts and formed our reference portfolios monthly because the return windows for our stocks were not aligned by calendar date (i.e., they began in the fifth month after the end of the fiscal year for each stock).
17. Returns from CRSP are unavailable after December 2011. A one-year return window, together with a four-month period between the fiscal year-end and the start of the holding period, results in August 2010 as the latest date for financial statement data. Thus, 2010 has only 430 observations.
18. Beneish (1999a) examined 74 cases of fraud over 1982–1993. Dechow, Ge, Larson, and Sloan (2011), who investigated SEC accounting and audit enforcement actions, reported that less than 0.5% of the company-years in their sample were associated with fraud.
19. We computed results for a wide range of cutoff values around –1.78, and the results are similar. In later tests, we used *M*-score as a continuous variable and found consistent results.
20. As with all our analyses, we did not explicitly incorporate transaction costs and short-selling constraints. Because these costs vary by market participant, we leave any consideration of their impact to the reader.
21. Of the three, perhaps the LEVI result is least intuitive and thus merits an explanation. When we compute coefficients for high-*M*-score minus low-*M*-score companies, SGAI and LEVI should be negative because they have negative coefficients in the model. Our difference-in-difference test asks whether LEVI is *more* negative when we are better able to predict returns. The answer is no. In fact, we have greater predictive power in the sample in which LEVI is relatively more positive. In other words, the model is more effective in predicting returns when LEVI is relatively more positive for the companies we have a long position in (as compared with the companies we short). Therefore, controlling for other factors, a sharper increase in a company's leverage is bad news.
22. Key results are qualitatively identical when the dependent variable is deflated by year $t + 1$ average total assets.
23. Short-window returns are far less susceptible to risk-model misspecifications.
24. Beneish (1999b) and Karpoff, Lee, and Martin (2008) found evidence of large market value losses from public revelations of accounting manipulation.
25. Some variables, such as DSR and AQI, could be indicative of either deteriorating economic conditions or aggressive accounting (i.e., they may play a dual role).
26. This five-year period was marked by a large number of corporate accounting scandals. It also represents an out-of-sample test for the Beneish (1999a) model, which was estimated with data for 1982–1988 and tested on a holdout sample for 1989–1992. We have no affiliation with Audit Integrity (acquired by the RiskMetrics Group and now called GMI Analyst).

References

Abarbanell, J., and B. Bushee. 1997. "Fundamental Analysis, Future Earnings, and Stock Prices." *Journal of Accounting Research*, vol. 35, no. 1 (Spring):1–24.

Altman, E.I. 1968. "Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy." *Journal of Finance*, vol. 23, no. 4 (September):589–609.

Barth, M.E., and A.P. Hutton. 2004. "Analyst Earnings Forecast Revisions and the Pricing of Accruals." *Review of Accounting Studies*, vol. 9, no. 1 (March):59–96.

Beaver, W.H. 1966. "Financial Ratios as Predictors of Failure." *Journal of Accounting Research*, vol. 4:71–111.

Beaver, W., M. McNichols, and R. Price. 2007. "Delisting Returns and Their Effect on Accounting-Based Market Anomalies." *Journal of Accounting and Economics*, vol. 43, no. 2–3 (July):341–368.

Bebchuk, L., A. Cohen, and C. Wang. Forthcoming. "Learning and the Disappearing Association between Governance and Returns." *Journal of Financial Economics*.

Beneish, M.D. 1997. "Detecting GAAP Violation: Implications for Assessing Earnings Management among Firms with Extreme Financial Performance." *Journal of Accounting and Public Policy*, vol. 16, no. 3 (Autumn):271–309.

———. 1999a. "The Detection of Earnings Manipulation." *Financial Analysts Journal*, vol. 55, no. 5 (September/October):24–36.

———. 1999b. "Incentives and Penalties Related to Earnings Overstatements That Violate GAAP." *Accounting Review*, vol. 74, no. 4 (October):425–457.

Beneish, M.D., and M.E. Vargus. 2002. "Insider Trading, Earnings Quality, and Accrual Mispricing." *Accounting Review*, vol. 77, no. 4 (October):755–791.

Beneish, M.D., C.M.C. Lee, and R. Tarpley. 2001. "Contextual Fundamental Analysis through the Prediction of Extreme Returns." *Review of Accounting Studies*, vol. 6, no. 2–3 (June):165–189.

Berk, J.B., and R.C. Green. 2004. "Mutual Fund Flows and Performance in Rational Markets." *Journal of Political Economy*, vol. 112, no. 6 (December):1269–1295.

Bradshaw, M.T., S.A. Richardson, and R.G. Sloan. 2001. "Do Analysts and Auditors Use Information in Accruals?" *Journal of Accounting Research*, vol. 39, no. 1 (June):45–74.

Carhart, M.M. 1997. "On Persistence in Mutual Fund Performance." *Journal of Finance*, vol. 52, no. 1 (March):57–82.

Chan, K., L.K.C. Chan, N. Jegadeesh, and J. Lakonishok. 2006. "Earnings Quality and Stock Returns." *Journal of Business*, vol. 79, no. 3 (May):1041–1082.

Ciesielski, J. 1998. "What's Happening to the 'Quality of Assets'?" *Analyst's Accounting Observer*, vol. 7, no. 3 (February):19.

Collins, D.W., G. Gong, and S.P. Hribar. 2003. "Investor Sophistication and the Mispricing of Accruals." *Review of Accounting Studies*, vol. 8, no. 2–3 (June):251–276.

Cooper, M.J., H. Gulen, and M.J. Schill. 2008. "Asset Growth and the Cross-Section of Stock Returns." *Journal of Finance*, vol. 63, no. 4 (August):1609–1651.

Davis, J.L. 1994. "The Cross-Section of Realized Stock Returns: The Pre-Compustat Evidence." *Journal of Finance*, vol. 49, no. 5 (December):1579–1593.

Dechow, P.M., W. Ge, C.R. Larson, and R.G. Sloan. 2011. "Predicting Material Accounting Misstatements." *Contemporary Accounting Research*, vol. 28, no. 1 (March):17–82.

Desai, H., S. Rajgopal, and M. Venkatachalam. 2004. "Value-Glamour and Accruals Mispricing: One Anomaly or Two?" *Accounting Review*, vol. 79, no. 2 (April):355–385.

Dichev, I., J. Graham, and M. Rajgopal. 2012. "Earnings Quality: Evidence from the Field." Working paper, Emory University (1 July).

Drake, M.S., L. Rees, and E.P. Swanson. 2011. "Should Investors Follow the Prophets or the Bears? Evidence on the Use of Public Information by Analysts and Short Sellers." *Accounting Review*, vol. 86, no. 1 (January):101–130.

Dresdner Kleinwort Wasserstein, Inc. 2003. *Earnings Junkies.* London: Global Equity Research (29 October).

Fairfield, P.M., J.S. Whisenant, and T.L. Yohn. 2003. "Accrued Earnings and Growth: Implications for Future Profitability and Market Mispricing." *Accounting Review*, vol. 78, no. 1 (January):353–371.

Fama, E.F., and K.R. French. 1992. "The Cross-Section of Expected Stock Returns." *Journal of Finance*, vol. 47, no. 2 (June):427–465.

———. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics*, vol. 33, no. 1 (February):3–56.

Fama, E.F., and J.D. MacBeth. 1973. "Risk, Return, and Equilibrium: Empirical Tests." *Journal of Political Economy*, vol. 81, no. 3 (May–June):607–636.

Foster, G. 1979. "Briloff and the Capital Market." *Journal of Accounting Research*, vol. 17, no. 1 (Spring):262–274.

Fridson, M.S. 2002. *Financial Statement Analysis: A Practitioner's Guide*. New York: John Wiley & Sons.

Gladwell, M. 2009. *What the Dog Saw and Other Adventures.* New York: Little, Brown and Company.

Green, J., J.R.M. Hand, and M.T. Soliman. 2011. "Going, Going, Gone? The Apparent Demise of the Accruals Anomaly." *Management Science*, vol. 57, no. 5 (May):797–816.

Harrington, C. 2005. "Analysis of Ratios for Detecting Financial Statement Fraud." *Fraud Magazine* (March/April):24–27.

Haugen, R.A., and N.L. Baker. 1996. "Commonality in the Determinants of Expected Stock Returns." *Journal of Financial Economics*, vol. 41, no. 3 (July):401–439.

Hirshleifer, D., K. Hou, and S. Teoh. 2012. "The Accrual Anomaly: Risk or Mispricing?" *Management Science*, vol. 58, no. 2 (February):320–335.

Hirshleifer, D., K. Hou, S. Teoh, and Y. Zhang. 2004. "Do Investors Overvalue Firms with Bloated Balance Sheets?" *Journal of Accounting and Economics*, vol. 38 (December):297–331.

Holthausen, R.W., and D.F. Larcker. 1992. "The Prediction of Stock Returns Using Financial Statement Information." *Journal of Accounting and Economics*, vol. 15, no. 2–3 (June–September):373–411.

Hribar, D.W., and S.P. Collins. 2002. "Errors in Estimating Accruals: Implications for Empirical Research." *Journal of Accounting Research*, vol. 40, no. 1 (March):105–134.

Jegadeesh, N., and S. Titman. 1993. "Returns to Buying Winners and Selling Losers." *Journal of Finance*, vol. 48, no. 1 (March):65–91.

Kama, I., and N. Melumad. 2011. "Camouflaged Earnings Management." Working paper, Tel Aviv University and Columbia University (April).

Karpoff, J.M., D.S. Lee, and G.S. Martin. 2008. "The Cost to Firms of Cooking the Books." *Journal of Financial and Quantitative Analysis*, vol. 43, no. 3 (September):581–612.

Kellogg, I., and L.B. Kellogg. 1991. *Fraud, Window Dressing, and Negligence in Financial Statements.* New York: McGraw-Hill.

Khan, M. 2005. "Are Accruals Really Mispriced? Evidence from Tests of an Intertemporal Capital Asset Pricing Model." Working paper, MIT (11 January).

Lev, B., and D. Nissim. 2006. "The Persistence of the Accruals Anomaly." *Contemporary Accounting Research*, vol. 23, no. 1 (Spring):193–226.

Lev, B., and R. Thiagarajan. 1993. "Fundamental Information Analysis." *Journal of Accounting Research*, vol. 31, no. 2 (Autumn):190–215.

Li, X., and R.N. Sullivan. 2011. "The Limits to Arbitrage Revisited: The Accrual and Asset Growth Anomalies." *Financial Analysts Journal*, vol. 67, no. 4 (July/August):50–66.

Lyon, J., B. Barber, and C. Tsai. 1999. "Improved Methods for Tests of Long-Run Abnormal Stock Returns." *Journal of Finance*, vol. 54, no. 1 (February):165–201.

Mashruwala, C., S. Rajgopal, and T. Shevlin. 2006. "Why Is the Accrual Anomaly Not Arbitraged Away?" *Journal of Accounting and Economics*, vol. 42, no. 1–2 (October):3–33.

Merrill Lynch. 2000. "Financial Reporting Shocks" (31 March).

Mohanram, P. 2005. "Separating Winners from Losers among Low Book-to-Market Stocks using Financial Statement Analysis." *Review of Accounting Studies*, vol. 10, no. 2–3 (September):133–170.

Morris, G.D.L. 2009. "Enron 101: How a Group of Business Students Sold Enron a Year before the Collapse." *Financial History*, no. 94 (Spring/Summer):12–15.

O'Glove, T.L. 1987. *Quality of Earnings*. New York: Free Press.

Ohlson, J.A. 1980. "Financial Ratios and the Probabilistic Prediction of Bankruptcy." *Journal of Accounting Research*, vol. 18, no. 1 (Spring):109–131.

Ou, J.A., and S.H. Penman. 1989. "Financial Statement Analysis and the Prediction of Stock Returns." *Journal of Accounting and Economics*, vol. 11, no. 4 (November):295–329.

Piotroski, J. 2000. "Value Investing: The Use of Historical Financial Statement Information to Separate Winners from Losers." *Journal of Accounting Research*, vol. 38 (Supplement):1–41.

Richardson, S.A., R.G. Sloan, M.T. Soliman, and I. Tuna. 2005. "Accrual Reliability, Earnings Persistence and Stock Prices." *Journal of Accounting and Economics*, vol. 39, no. 3 (September):437–485.

Schilit, H. 2010. *Financial Shenanigans: How to Detect Accounting Gimmicks and Fraud in Financial Reports*. 3rd ed. New York: McGraw-Hill.

Schrand, C.M., and S.L.C. Zechman. 2012. "Executive Overconfidence and the Slippery Slope to Financial Misreporting." *Journal of Accounting and Economics*, vol. 53, no. 1–2 (February–April):311–329.

Siegel, J.G. 1991. *How to Analyze Businesses, Financial Statements, and the Quality of Earnings*. 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Sloan, R.G. 1996. "Do Stock Prices Fully Reflect Information in Accruals and Cash Flows about Future Earnings?" *Accounting Review*, vol. 71, no. 3 (July):289–315.

Stickney, C., P. Brown, and J. Wahlen. 2003. *Financial Reporting, Financial Statement Analysis, and Valuation*. 5th ed. Mason, OH: Thomson-Southwestern.

Wells, J.T. 2001. "Irrational Ratios." *Journal of Accountancy*, vol. 192, no. 2 (August):80–83.

Zachs, L., ed. 2011. *The Handbook of Equity Market Anomalies: Translating Market Inefficiencies into Effective Investment Strategies*. New York: Wiley.

