Detecting Abnormal Operating Performance: Revisited

Erik Lie*

Firms that undertake corporate events often exhibit atypical financial characteristics, such that future performance might be expected to change even before the event is announced. I investigate five methods for generating control samples in various sampling situations. The results indicate that the methods sometimes produce severely biased test statistics. Overall, the method that generates control firms with similar prior levels and changes of performance and similar market-to-book ratios produces the most reliable test statistics. Furthermore, I find that it is more important to control for levels of performance than for changes in performance and market-to-book ratios.

Numerous studies in accounting and finance investigate operating performance following corporate events. The goal of these studies is to assess whether the events convey information about future operating performance, which entails identifying the unexpected portion of any performance changes. However, several studies (e.g., Easton and Harris, 1991; Ali and Zarowin, 1992; and Fama and French, 2000) indicate that operating performance consists of both permanent and transitory components.[1] Furthermore, corporate events are frequently preceded by either significant changes in performance (e.g., Healy and Palepu, 1988; Denis and Denis, 1995; Benartzi, Michaely, and Thaler, 1997; and Denis and Kruse, 2000) or abnormal levels of performance (e.g., Shah, 1994; Loughran and Ritter, 1997; Mikkelson, Partch, and Shah, 1997; Lie and McConnell, 1998; McLaughlin, Safieddine, and Vasudevan, 1998; and Friday, Howton, and Howton, 2000). If these significant changes or abnormal levels indicate that some portion of a firm's earlier operating performance is transitory, then we might expect that the future operating performance will change.

Barber and Lyon (1996) use simulations of random samples to evaluate various models of expected operating performance. Perhaps their most important finding is that failure to consider an expected reversion in operating performance could introduce bias when examining post-event operating performance: models that do not control for levels of pre-event performance yield misspecified test statistics if the sample firms exhibit inferior or superior pre-event performance. Therefore, Barber and Lyon suggest that researchers control for expected reversion by comparing the performance of the sample firms to that of control firms with similar pre-event performance.

Some recent studies use methods for developing control firms that vary somewhat from those examined by Barber and Lyon (1996). In their study of operating performance following dividend changes, Benartzi, Michaely, and Thaler (1997) control for reversion by comparing the performance of their sample firms to that of control firms with similar pre-event changes in performance, rather than firms with similar pre-event levels of performance as Barber and Lyon suggest. Alternatively, in their study of operating performance following self-tender offers, Nohel and

[1]Potential sources of the transitory component include extraordinary income or expenses, temporary shifts in demand or supply, or earnings management (Rangan, 1998; and Teoh, Welch, and Wong, 1998).

I thank Dave Denis, Heidi Lie, seminar participants at the College of William & Mary, and two anonymous referees for helpful comments.

Erik Lie is an Assistant Professor of Finance at the College of William and Mary in Williamsburg, Virginia.

Tarhan (1998) extend Barber and Lyon's suggested procedure. In particular, they generate control firms with similar pre-event level of performance and ratio of market-to-book value of assets.

In my study, I investigate which procedure for developing control firms is more appropriate in various sampling situations. I extend the study of Barber and Lyon (1996) in several directions. First, in addition to evaluating a model of expected operating performance based on past levels of performance, I evaluate models based on past changes in performance and market-to-book ratios. Second, in addition to assessing the various models when the sample firms have extreme pre-event performance or book value of assets, I examine sample firms that exhibit dramatic changes in past performance or extreme market-to-book ratios. Third, I separately report rejection levels of the null hypothesis of no abnormal performance in favor of the alternative hypotheses of negative abnormal performance and positive abnormal performance instead of total rejection levels.

In my preliminary analysis, I investigate the relation between future changes in performance and past changes in performance, past levels of performance, and market-to-book ratios. I find that future changes in performance are negatively related to past levels and negatively related to changes over the past year. The exception is when performance is high but declining. In those cases, future changes in performance are positively related to past changes (i.e., there is a momentum effect rather than a reversion effect). I find that future changes are positively related to market-to-book ratios, except when past level of performance is low and the market-to-book ratio is high. These results indicate that future changes in performance are partially predictable. Therefore, researchers should control for this predictable portion as they attempt to assess what portion of future changes is unexpected.

Next, I examine the impact of the choice of control firms on test statistics designed to detect abnormal (i.e., unexpected) operating performance. I assess five methods for constructing samples of control firms. My results indicate that, when the original sample firms are randomly drawn from the whole universe of firms, all methods produce well-specified test statistics. Further, industry clustering has little effect on the test statistics. When the samples are drawn from firms with pre-defined characteristics, the methods often produce severely biased test statistics. Overall, the method that generates control firms with similar past levels of performance, similar past changes in performance, and similar market-to-book ratios, seems superior to the other methods. Furthermore, it appears more important to control for levels of performance than changes in performance and market-to-book ratios.

The paper proceeds as follows. The next section describes the sample. Section II presents the results from regressing changes in performance against past levels of performance, past changes in performance, and market-to-book ratios. Section III describes the expected performance models. Section IV presents the empirical results for the different methods of generating control samples. Section V summarizes and concludes.

I. Data

The database I use in this study comprises all firms with data available on Compustat. The sample period starts in 1978 and ends in 1997.[2] I measure operating performance as operating income (Compustat data item 13) scaled by the average of the beginning- and ending-period book value of assets for the same period. I call this ratio return on assets (ROA). I measure the

[2]In comparison, Barber and Lyon (1996) use a database of only NYSE and Amex firms, and their data are taken from the period 1977-1992. To the extent that empirical studies include Nasdaq firms, my simulation results are, therefore, more relevant.

market-to-book ratio (M/B) as the book value of assets minus the book value of equity, plus the market value of equity at the end of the fiscal year, all scaled by the book value of assets at the end of the fiscal year.

To be included in the analyses, I require that ROAs be available for years t-1, t, and t+1. Doing so enables me to find the level of ROA in year t as well as the changes in ROA between years t-1 and t (prior change) and between years t and t+1 (future change). I also require that M/B be available for year t, and that the book value of assets in year t exceed one million dollars.

Table I provides descriptive statistics for the sample of 79,100 firm-years. I provide medians and the tenth and ninetieth percentiles because they are used as cutoff values in later analyses.

II. Predicting Future Changes in ROA

Fama and French (2000) show evidence suggesting that earnings are mean reverting. They find that one-year changes in earnings are negatively related to past levels of earnings and the prior year's changes in earnings. They further find evidence of nonlinear relations in their regression specifications. Mean reversion is stronger when earnings are below the mean earnings and when earnings are further away from the mean in either direction.

I regress the change in ROA from year t to t+1 against the ROA in year t and the change in ROA from year t-1 to t. I also include the market-to-book ratio in year t as an independent variable because the market's expectations of future earnings changes should be reflected in the market value. The market value of the firm is, at least in theory, equal to the discounted value of future cash flow. Because earnings are one of the major determinants of cash flow, higher future earnings imply a higher value. Therefore, the market value scaled by the book value embeds information about the capital market's expectation of future scaled earnings.

Table II, Panel A, shows the results of the regressions for the entire sample. The results indicate that future changes are negatively related to past levels and past changes, and positively related to the market-to-book ratio. The R-squared is 6.7% when I include only the past levels independent variable. The R-squared increases to 8.1% when I include all three independent variables. Thus, the incremental explanatory power from past changes and the market-to-book ratio is only 1.4%.

However, as suggested by Fama and French (2000), there could be complex nonlinear relations between the future changes in ROA and the independent variables in the regressions. For example, Basu (1997) argues that accounting conservatism leads firms to report losses immediately, but spread out earnings gains over time. Basu shows that positive earnings changes are more persistent than negative earnings changes. Alternatively, the results in Burgstahler and Dichev (1997) suggest that managers are more likely to temporarily manipulate earnings upward than downward, suggesting that earnings reversion is more likely to take place following an increase in earnings than following a decrease. To uncover some of these complex relations, I partition my sample into eight groups, based on whether the ROA in year t and the change in ROA from year t-1 to year t are above or below the median, and whether the market-to-book ratio is above or below one. Table II, Panel B, reports the regression results for each of these groups.

I find that the relation between future changes and past levels is negative for all eight groups, and that although the strength of the relation varies somewhat across the groups, it is consistently statistically significant. In contrast, the sign of the relation between past and

Table I. Descriptive Statistics

This table presents descriptive statistics for the book value of total assets (in millions of dollars) in year t ($Size_t$), the ratio of the market value of assets to book value of assets in year t (M/B_t), ROA in year t (ROA_t), the change in ROA from year t-1 to t ($\Delta ROA_{t-1,t}$), and the change in ROA from year t to t+1 ($\Delta ROA_{t,t+1}$) for a sample of 79,100 firm-years.

	Mean	Median	10th Percentile	90th Percentile
$Size_t$	1,958.3	95.4	6.8	2,974.5
M/B_t	1.698	1.189	0.835	2.823
ROA_t	0.099	0.115	-0.050	0.251
$\Delta ROA_{t-1,t}$	-0.008	-0.001	-0.099	0.076
$\Delta ROA_{t,t+1}$	-0.006	-0.002	-0.100	0.079

future changes varies across the groups. Although the relation is negative for most of the groups, it is positive for the group with high levels of ROA and low (i.e., typically negative) prior changes in ROA. In other words, firms with high, but recently declining, earnings exhibit a momentum effect rather than a reversion effect: the greater the recent drop in earnings, the larger the expected future drop in earnings.

I also partition the sample into more groups to examine whether the mean reversion is weaker for firms with less extreme levels or changes in ROA. (These results are not tabulated.) Firms in the middle third based on the level of ROA show a weaker relation between levels of ROA and future changes than do other firms. Similarly, firms in the middle third based on past changes in ROA show a weaker relation between past and future changes in ROA.

I find that the relation between the market-to-book ratio and future changes in ROA is positive for most groups. Therefore, a high M/B indicates that earnings are likely to increase. However, for the groups with a low ROA but high M/B, the relation between M/B and future changes in ROA is negative. This is a surprising result, but further exploration is beyond the scope of this paper.

The complex relations between the variables also suggest that the incremental explanatory power associated with the various independent variables in Panel A could have been higher if the specification allowed for nonlinear relations. For example, the specification in Panel A does not allow for the relation between future changes in ROA and M/B to vary across different types of firms, but the results in Panel B suggest that it does. Since the relation is actually positive for certain firm types and negative for others, the results in Panel A are misleading.

My regression results suggest that a model of expected future earnings should incorporate past levels of ROA, past changes in ROA, and market-to-book ratios. The complexity of the relations further suggests that it is difficult to capture expected changes in earnings with a simple prediction model based on the universe of data. As a measure for expected future changes, researchers should use actual future changes of control firms with similar past performance patterns (i.e., past levels of and changes in ROA) and market-to-book ratios. This approach also implicitly controls for changes in future market-wide conditions.

III. Modeling Expected Performance

To assess whether operating performance is abnormal, I develop a measure of expected operating performance that relies heavily on the performance for samples of control firms.

Table II. Earnings Regressions

This table presents results from regressions of the change in ROA from year t to t+1 ($\Delta ROA_{t,t+1}$) against the ROA in year t (ROA_t), the change in ROA from year t-1 to t ($\Delta ROA_{t-1,t}$), and the ratio of the market value of assets to book value of assets in year t (M/B_t). In Panel B, the sample is partitioned into eight groups based on whether $\Delta ROA_{t-1,t}$ and ROA_t are above their respective medians and whether M/B_t is above one. (t-statistics appear in parentheses to the right of the corresponding regression coefficient.)

Panel A. Regressions Using the Overall Sample

	(a)		(b)		(c)		(d)	
Intercept	0.008	(17.9)	0.006	(12.9)	0.005	(9.2)	0.003	(4.9)
ROA_t	-0.163	(-75.6)	-0.145	(-65.8)	-0.163	(-76.0)	-0.146	(-66.2)
$\Delta ROA_{t-1,t}$			-0.105	(-30.1)			-0.105	(-30.3)
M/B_t					0.002	(9.2)	0.002	(9.7)
Adj. R^2	0.067		0.078		0.069		0.081	
N	79,100		79,100		79,100		79,100	

Panel B. Regressions for Subsamples

		$M/B_t < 1$				$M/B_t \geq 1$			
		ROA_t< Median		$ROA_t \geq$ Median		ROA_t< Median		$ROA_t \geq$ Median	
$\Delta ROA_{t-1,t}$ < Median	Intercept	-0.048	(-7.8)	-0.047	(-3.5)	0.011	(5.5)	-0.014	(-10.4)
	ROA_t	-0.349	(-27.2)	-0.170	(-5.8)	-0.160	(-19.3)	-0.172	(-20.6)
	$\Delta ROA_{t-1,t}$	-0.098	(-6.4)	0.248	(8.4)	-0.127	(-10.0)	0.111	(9.4)
	M/B_t	0.065	(9.1)	0.065	(4.7)	-0.005	(-8.6)	0.015	(24.9)
	Adj. R^2	0.144		0.058		0.071		0.045	
	N	9,093		2,740		13,345		14,372	
$\Delta ROA_{t-1,t} \geq$ Median	Intercept	-0.044	(-7.1)	-0.029	(-2.5)	0.009	(4.8)	0.011	(9.4)
	ROA_t	-0.176	(-10.3)	-0.317	(-13.6)	-0.187	(-20.7)	-0.160	(-25.6)
	$\Delta ROA_{t-1,t}$	-0.060	(-3.7)	-0.182	(-9.4)	-0.014	(-1.3)	-0.313	(-41.8)
	M/B_t	0.055	(7.7)	0.072	(5.8)	-0.002	(-4.4)	0.009	(18.1)
	Adj. R^2	0.023		0.134		0.043		0.152	
	N	6,341		3,588		10,772		18,849	

A. Measure of Expected Operating Performance

Following Barber and Lyon (1996), I define a firm's expected performance as its past performance, plus the change in the performance of a control firm:

$$E\left(P_{i,t+1}\right) = P_{i,t} + \left(PC_{i,t+1} - PC_{i,t}\right) \quad (1)$$

E(·) is an expectations operator, year t is the year preceding the event year, $P_{i,t}$ is the performance of firm i in year t, and $PC_{i,t}$ is the performance of control firm i in year t.

B. Control Samples

In their study, Barber and Lyon (1996) note that the choice of an appropriate control sample is critical to investigating the operating performance following corporate events. The choice of an inappropriate control sample can cause severely biased results and erroneous conclusions. I investigate five methods for developing control firms, and I use these control firms as a benchmark against which I assess the change in operating performance for the sample firms from year t to year $t+1$. To capture industry trends in performance, all five methods use control firms in similar industries as the original sample firms, insofar as such control firms are available.

I base my first method (M1) on industry classification and pre-event performance. For each original sample firm, I identify all firms with the same two-digit SIC code. Of these firms, I identify all firms with ROA within ±10% or within ±0.01 of the sample firm's ROA in year t.[3] As the control firm, I choose the firm with ROA closest to that of the original sample firm. If I cannot identify a potential control firm among firms that have the same two-digit SIC code, I repeat the process for firms with the same one-digit SIC code. If I still cannot identify a potential control firm, I repeat the process for all firms without regard to SIC code. Finally, if I still cannot identify a potential control firm, I choose the firm with ROA closest to the sample firm as the control firm, regardless of its SIC code and whether the ROA is within ±10%.

The second method (M2) generates control firms based on industry classification and one-year pre-event changes in performance. From all firms with the same two-digit SIC code, I identify those with a change in ROA within ±10% or within ±0.01 of the sample firm's change in ROA from year t-1 to year t. As the control firm, I choose the firm with the change in ROA closest to that of the original sample firm. If no firms with the same two-digit SIC code meet the matching criteria, I repeat the process first for firms with the same one-digit SIC code, and then for all firms without regard to SIC code. If still no firms meet the criteria, I choose the firm with the change in ROA closest to that of the sample firm, without regard to SIC code and ROA change filter.

The third method (M3) combines M1 and M2. I identify firms within the same two-digit SIC code that have an ROA within ±20% or within ±0.01 of the sample firm's ROA in year t, and a change in ROA within ±20% or within ±0.01 of the sample firm's change in ROA from year t-1 to year t. From these firms, I choose the firm with the lowest sum of absolute differences, defined as follows:

$$\left|\text{ROA}_{\text{t,Sample firm}} - \text{ROA}_{\text{t, Firm i}}\right| + \left|\Delta\text{ROA}_{\text{t-1 to t,Sample firm}} - \Delta\text{ROA}_{\text{t-1 to t, Firm i}}\right| \tag{2}$$

If no firms meet the criteria, then I repeat the process first for firms with the same one-digit SIC code, and then for all firms without regard to SIC code. If there are still no firms that meet the criteria, then I choose the firm that has the lowest sum of absolute differences regardless of the filters.

The fourth and fifth methods (M4 and M5) are extensions of M1 and M3, respectively. In M4, I identify firms with ROA within ±20% or within ±0.01 of the sample firm's ROA in year t, and an M/B within ±20% or within ±0.1 of the sample firm's M/B in year t. From these firms, I choose the firm with the level of ROA closest to that of the original sample firm. If no firms meet the criteria, I repeat the process first for firms with the same one-digit SIC code, and then for all firms without regard to SIC code. If still no firms meet the criteria, I choose the firm with ROA closest to the sample firm, regardless of the filters.

In M5, I identify firms with ROA within ±20% or within ±0.01 of the sample firm's ROA in year t, a change in ROA within ±20% or within ±0.01 of the sample firm's change in ROA from

[3]In comparison, Barber and Lyon (1996) require that all performance-based control firms have ROA within ±10% of those of the original sample firms. However, this requirement makes it hard to generate control firms if the ROA of the original sample firm is close to zero.

year t-1 to t, and a M/B within ±20% or within ±0.1 of the sample firm's M/B in year t. From these firms, I choose the firm with the lowest sum of absolute differences, defined exactly as in M3. If no firms meet the criteria, I repeat the process first for firms with the same one-digit SIC code, and then for all firms regardless of SIC code. If still no firms meet the criteria, I choose the firm that has the lowest sum of absolute differences, regardless of the filters.

Because past levels of ROA, past changes in ROA and M/B all predict future changes in ROA, a method that controls for all three of these variables should be superior. However, there is a disadvantage of controlling for too many dimensions: each dimension suffers because it is difficult to find control firms that are close on all dimensions. Only an empirical analysis can show which method is most appropriate for different sampling situations.

IV. Empirical Results from Simulations

I rely on simulations to assess the performance of the five methods for generating control firms. I draw 10,000 random samples of 50 firm-years without replacement. Each sample is drawn either from the total population of firm-years, or from a subpopulation with pre-defined characteristics.

A. Simulations Using Random Samples

Table III, Panel A, presents the percentages of firms that meet the various criteria of the five procedures used to generate control firms. As expected, M1 and M2 generate control firms that are closer to the industry of the original firms than do other methods. For example, more than 90% of the control firms generated by M1 and M2 have the same two-digit SIC codes as the original firms, while comparable figures for M3, M4, and M5 are 64%, 81%, and 47%, respectively. These percentages show that there is a potential disadvantage of controlling for too many dimensions: control firms often cannot be found in the same industry as the original firm. To further illustrate the similarities between the control firms and the sample firms, Panel B presents median absolute differences in matching characteristics (i.e., ROA, ROA change, and market-to-book ratio).

Table IV, Panel A, presents the empirical rejection rates for the five methods. I define the rejection rates as the percentage of 10,000 random samples of 50 firm-years that rejects the null hypothesis of no abnormal operating performance. I estimate the rates using nonparametric Wilcoxon tests at a theoretical significance level of 5%.[4] I separately report rejection levels of the null hypothesis of no abnormal performance in favor of the alternative hypotheses of negative abnormal performance and positive abnormal performance.[5] The different methods for generating control samples seem to produce equally well-specified test statistics, although they are all slightly conservative (i.e., they reject the null hypothesis too infrequently). Hence, in random samples, the method used to generate control firms has little impact on the specification of the test statistics.

To estimate power functions for the various methods, I induce abnormal performance by adding a constant with a known mean to the observed performance (see Figure I). The power functions show that the methods for generating control firms are about equal in power, although

[4]I also estimate the rejection rates at theoretical significance levels of 1% and 10%. These rates exhibit similar trends as the rates using a level of 5%, and thus are not reported.

[5]Barber and Lyon (1996) report the rejection levels of the null hypothesis of no abnormal performance in favor of the alternative hypothesis of abnormal performance. However, this might partially disguise poorly specified test-statistics. For example, if the test rejects the null hypothesis in favor of negative abnormal performance twice as often as it theoretically should, but never rejects in favor of positive abnormal performance, then the total rejection level as reported by Barber and Lyon would still be theoretically correct. An additional advantage of reporting rejection levels for both tails is that they give a better sense of the direction of potential bias.

Table III. Characteristics of Control Firms

This table presents characteristics of control firms relative to sample firms. The figures are based on 10,000 random samples of 50 firms. The first method for generating control firms is based on industry classification and the pre-event ROA. The second method is based on industry classification and the one-year pre-event change in ROA. The third method is based on industry classification, the pre-event ROA, and the one-year pre-event change in ROA. The fourth method is based on industry classification, the pre-event ROA, and the pre-event market-to-book ratio. The fifth method is based on industry classification, the pre-event ROA, the one-year pre-event change in ROA, and the pre-event market-to-book ratio.

Panel A. Percentage of Firms with Available Control Firms in Different Control Categories

	2-Digit SIC & Filter	1-Digit SIC & Filter	No SIC & Filter	No SIC & No Filter
1. Matched on ROA	91.5	6.4	1.7	0.3
2. Matched on ROA Change	92.2	6.0	1.5	0.3
3. Matched on ROA and ROA Change	63.5	19.4	11.2	6.0
4. Matched on ROA and M/B	80.7	10.8	6.2	2.2
5. Matched on ROA, ROA Change, and M/B	47.1	21.7	16.1	15.2

Panel B. Median Absolute Differences in Matching Characteristics Between Sample Firms and Control Firms

	ROA	ROA Change	M/B
1. Matched on ROA	0.001	0.045	0.31
2. Matched on ROA Change	0.076	0.001	0.38
3. Matched on ROA and ROA Change	0.004	0.003	0.31
4. Matched on ROA and M/B	0.002	0.042	0.11
5. Matched on ROA, ROA Change, and M/B	0.005	0.004	0.13

M5 is marginally more powerful and M2 marginally less powerful than the others. The power functions based on parametric test statistics are much flatter, suggesting less power, than those based on nonparametric test statistics. Because the parametric test statistics are considerably less powerful than non-parametric Wilcoxon test statistics, I only report results based on non-parametric test statistics.

Another concern is the effect of industry clustering in the sample. If performance changes are more correlated across firms in the same industry than across firms in different industries, then industry effects are relevant. Indeed, this logic forms my basic premise for identifying control firms in the same industry as the original firm. However, as Table III shows, it is often difficult to identify firms in the same industry that have the desired characteristics. To test the effect of industry clustering, I randomly draw samples of firm-years where all the firms in a sample come from one of the 30 two-digit industries with the largest number of firm-years out of 74 possible two-digit industries.[6] Table IV, Panel B, presents the empirical rejection rates. The rejection rates tend to be slightly higher than those in Panel A, but again, all procedures appear to generate well-specified test statistics. Thus, industry clustering does not appear to be of great concern.

B. Simulations Using Random Samples with Pre-Specified Characteristics

Researchers often assess firms' operating performance following certain corporate events. Thus, the sample firms are not drawn randomly from the whole universe of firms, but are only selected if they undergo the event in question. These carefully chosen sample firms often exhibit systematic performance patterns preceding the event. For example, Healy and Palepu (1988) and Benartzi, Michaely, and Thaler (1997) document increases (decreases) in operating performance before dividend increases (decreases), while Shah (1994) finds abnormally good (poor) performance before leverage-increasing (leverage-decreasing) exchange offers. Other studies require that sample firms exhibit certain performance characteristics. For example, Denis and Kruse (2000) examine corporate restructuring activities and performance for a sample of firms that

[6]About 83% of the firm-years are in one of these 30 two-digit industries.

Table IV. Rejection Rates for Random Samples

This table presents the percentage of 10,000 random samples of 50 firms that reject the null hypothesis of no abnormal operating performance at a theoretical significance level of 5%. The rejection rates are separated for the alternative hypotheses of negative performance (the estimated test statistic is less than the theoretical test statistic with a cumulative probability of 0.025) and positive performance (the estimated test statistic is larger than the theoretical test statistic with a cumulative probability of 0.975). The first method for generating control firms is based on industry classification and the pre-event ROA. The second method is based on industry classification and the one-year pre-event change in ROA. The third method is based on industry classification, the pre-event ROA, and the one-year pre-event change in ROA. The fourth method is based on industry classification, the pre-event ROA, and the pre-event market-to-book ratio. The fifth method is based on industry classification, the pre-event ROA, the one-year pre-event change in ROA, and the pre-event market-to-book ratio. All tests are based on non-parametric (Wilcoxon) test statistics. The rejection levels are significantly different from the theoretical level of 2.5% if they fall outside the range 2.1%-2.9% (α=0.01), in which case, the figures are marked with *. (The range for α=0.05 is 2.2%-2.8%.)

	Cumulative Probability	
	0.025	0.975
Panel A. Samples without Industry Clustering		
1. Matched on ROA	2.1	2.4
2. Matched on ROA Change	2.1	2.3
3. Matched on ROA and ROA Change	2.2	2.2
4. Matched on ROA and M/B	2.4	2.2
5. Matched on ROA, ROA Change, and M/B	2.1	2.1
Panel B. Samples with Industry Clustering		
1. Matched on ROA	2.2	2.8
2. Matched on ROA Change	2.4	2.2
3. Matched on ROA and ROA Change	2.3	2.5
4. Matched on ROA and M/B	2.1	2.7
5. Matched on ROA, ROA Change, and M/B	2.3	2.5

have experienced a substantial drop in performance. Yet other studies partition samples into groups based on financial characteristics, and separately examine the operating performance of these groups of firms. For example, Nohel and Tarhan (1998) compare operating performance following self-tender offers for firms that have high market-to-book ratios with that of firms with low market-to-book ratios.

1. Samples with Extreme Pre-Event ROA

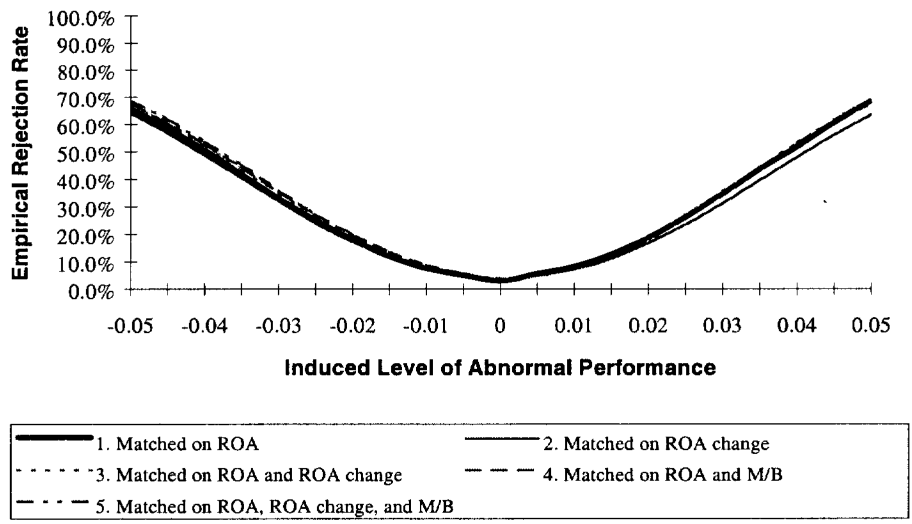
Table V, Panel A, presents the results for sample firms from either the lowest or highest deciles based on the pre-event ROA. Not surprisingly, when I choose control firms based on the past level of ROA (M1), the test statistics seem fairly well specified. However, there is a mild positive bias (i.e., a tendency for the test statistics to reject the null hypothesis in favor of positive performance too frequently, and negative performance too infrequently) for samples with low ROA. Test statistics based on M1 reject the null hypothesis in favor of negative abnormal performance in 1.2% of the 10,000 samples (compared to the theoretical level of 2.5%) and in favor of positive abnormal performance in about 3.5% of the samples. Both of these percentages are statistically different from the theoretical levels. This bias might be due to imperfect matching. For firms with very low ROA, there are more potential control firms that have slightly larger ROA than there are with slightly lower ROA (because the probability density function is increasing at the lower tail of the distribution). Thus, the chosen control firms tend to have marginally higher ROA than the original sample firms. This tendency to choose control firms with slightly higher ROA could create the positive bias, because future changes in ROA are negatively related to past levels of ROA.

M3, M4, and M5 yield as well-specified test statistics as M1, even though they control for

Figure I. Power Functions

This figure shows empirical rejection rates of parametric and nonparametric test statistics at a theoretical significance level of 5%. I induce abnormal performance by adding a constant to the observed performance for each of 50 randomly selected firm-years in 10,000 random samples. The first method for generating control firms is based on industry classification and the pre-event ROA. The second method is based on industry classification and the one-year pre-event change in ROA. The third method is based on industry classification, the pre-event ROA, and the one-year pre-event change in ROA. The fourth method is based on industry classification, the pre-event ROA, and the pre-event market-to-book ratio. The fifth method is based on industry classification, the pre-event ROA, the one-year pre-event change in ROA, and the pre-event market-to-book ratio.

Panel A. Parametric Test Statistics



Panel B. Non-Parametric Wilcoxon Test Statistics

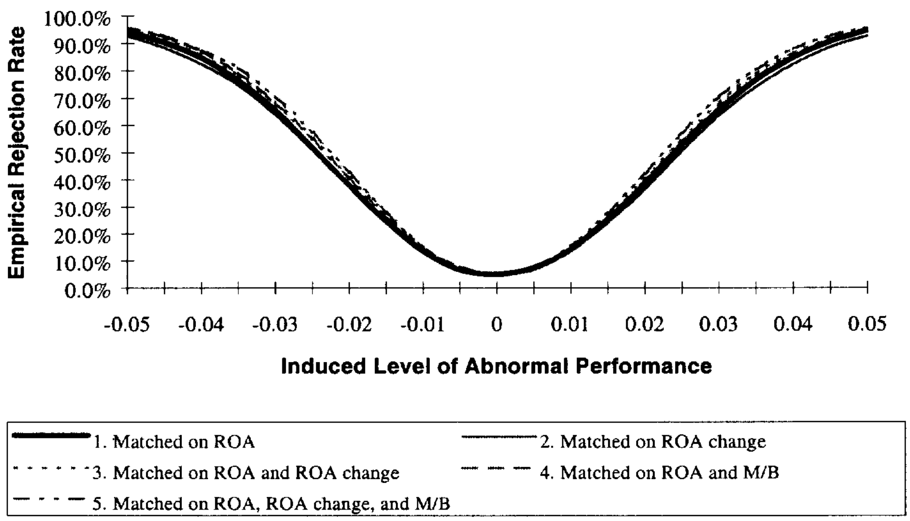


Table V. Rejection Rates for Samples with Pre-Specified Characteristics

This table presents the percentage of 10,000 random samples of 50 firms that reject the null hypothesis of no abnormal operating performance at a theoretical significance level of 5%. The rejection rates are separated for the alternative hypotheses of negative performance (the estimated test statistic is less than the theoretical test statistic with a cumulative probability of 0.025) and positive performance (the estimated test statistic is larger than the theoretical test statistic with a cumulative probability of 0.975). I draw random samples from either the lowest or highest deciles based on level of ROA (Panel A), past year's change in ROA (Panel B), the ratio of market value of assets to book value of assets (Panel C), or book value of total assets (Panel D). The first method for generating control firms is based on industry classification and the pre-event ROA. The second method is based on industry classification and the one-year pre-event change in ROA. The third method is based on industry classification, the pre-event ROA, and the one-year pre-event change in ROA. The fourth method is based on industry classification, the pre-event ROA, and the pre-event market-to-book ratio. The fifth method is based on industry classification, the pre-event ROA, the one-year pre-event change in ROA, and the pre-event market-to-book ratio. All tests are based on non-parametric (Wilcoxon) test statistics. The rejection levels are significantly different from the theoretical level of 2.5% if they fall outside the range 2.1%-2.9% (α=0.01), in which case the figures are marked with *. (The range for α=0.05 is 2.2%-2.8%.)

	Samples from Lowest Deciles		Samples from Highest Deciles	
	0.025	0.975	0.025	0.975
Panel A. Samples Selected on the Basis of ROA				
1. Matched on ROA	1.2 *	3.5 *	2.1	2.9
2. Matched on ROA Change	0.1 *	21.9 *	46.9 *	0.0 *
3. Matched on ROA and ROA Change	1.6 *	3.3 *	2.1	2.1
4. Matched on ROA and M/B	2.1	2.2	2.6	1.9 *
5. Matched on ROA, ROA Change, and M/B	1.4 *	3.4 *	1.9 *	1.8 *
Panel B. Samples Selected on the Basis of ROA Change				
1. Matched on ROA	1.9 *	4.0 *	6.6 *	0.3 *
2. Matched on ROA Change	3.2 *	2.2	1.4 *	1.7 *
3. Matched on ROA and ROA Change	2.3	1.8 *	1.7 *	1.7 *
4. Matched on ROA and M/B	2.5	2.5	7.0 *	0.2 *
5. Matched on ROA, ROA Change, and M/B	2.5	2.2	1.6 *	1.7 *
Panel C. Samples Selected on the Basis of M/B				
1. Matched on ROA	8.7 *	0.8 *	1.5 *	2.3
2. Matched on ROA Change	0.5 *	4.8 *	4.7 *	0.6 *
3. Matched on ROA and ROA Change	7.0 *	0.3 *	2.4	2.3
4. Matched on ROA and M/B	3.2 *	1.7 *	3.2 *	1.0 *
5. Matched on ROA, ROA Change, and M/B	3.5 *	0.9 *	2.5	1.5 *
Panel D. Samples Selected on the Basis of Size				
1. Matched on ROA	6.4 *	0.5 *	1.9 *	2.6
2. Matched on ROA Change	0.8 *	3.9 *	0.9 *	2.8
3. Matched on ROA and ROA Change	3.3 *	1.2 *	1.4 *	4.2 *
4. Matched on ROA and M/B	5.4 *	1.0 *	1.9 *	2.6
5. Matched on ROA, ROA Change, and M/B	3.0 *	1.6 *	1.6 *	2.9

characteristics beyond past ROA. In fact, M4 appears to produce slightly better statistics for samples with low ROA, because the test statistics associated with this method show almost no sign of bias. In contrast, M2 produces grossly misspecified and biased test statistics. For samples with low ROA, test statistics based on this method rarely reject the null hypothesis in favor of negative abnormal performance, but do reject in favor of positive abnormal performance in 22% of the samples. In contrast, for samples with high ROA, the test statistics rarely reject in favor of positive abnormal performance, but do reject in favor of negative abnormal performance in 47% of the samples.

When I assess the future performance of firms that have experienced abnormal levels of past performance, the results suggest that it is critical to compare the projected future performance to that of benchmark firms that show similar levels of past performance. M1, M3, M4, and M5 all yield acceptable test statistics. Comparing future performance to benchmark firms with similar changes in performance, but not similar levels, results in severely biased and misspecified test statistics.[7]

2. Samples with Extreme Pre-Event Changes in ROA

Table V, Panel B, presents the results for sample firms from either the lowest or highest deciles based on the pre-event change in ROA. M2, M3, and M5, which all control for past ROA change, produce well-specified test statistics with little sign of bias. M1 and M4, which do not control for past ROA change, produce somewhat worse statistics. For M1, the test statistics show a modest bias for samples with extreme ROA decreases. For both M1 and M4, the test statistics are negatively biased for samples with extreme ROA increases. Thus, when the sample firms experience a dramatic change in ROA over the last year, ideally the control firms will experience a similar change.

It might be problematic to examine samples with extreme ROA changes without considering the level of ROA, because a change (of, e.g., 0.05) could have different implications for a firm with an ROA of 0.05 than for a firm with an ROA of 0.15. Thus, I further partition Panel B into firms with ROA above or below the median. The results are reported in Table VI. The test statistics are more biased for samples with ROA above the median (Panel B) than for samples with ROA below the median (Panel A). Further, M3 and M5 yield the least biased statistics. In fact, these two methods are the only ones that generate acceptable test statistics in Panel B. Thus, when the original sample firms exhibit a dramatic change in past ROA and either low or high pre-event ROA, researchers should generate control samples with similar past changes in, and levels of, ROA.[8]

3. Samples with Extreme Pre-Event M/B

Table V, Panel C, presents the results for samples with abnormally low (high) pre-event M/B. As expected, M4 and M5, which partially control for M/B, yield the least biased test statistics, although they are somewhat negatively biased.

4. Samples with Extreme Size

Table V, Panel D, shows that when the sample firms are either very small or very large, no method dominates. All methods produce somewhat biased test statistics, although the sign of the bias varies. I also note that the tests appear to be better specified for large, rather than for small firms.

C. Sensitivity Analysis

Researchers might be interested in assessing the operating performance for samples that do not quite conform to my simulated samples. For example, the sample might have more than 50 observations and/or exclude Nasdaq firms. Further, the sample in question might exhibit abnormal

[7]Since operating performance differs across industries, it might be more relevant to examine sample firms with extreme ROA relative to industry peers. An examination of firms with extreme industry-adjusted performance (not reported here) shows similar results to those in Panel A.

[8]I also examine samples from firms that are randomly drawn from either firm-years in the lowest decile based on level of ROA and the lowest decile based on the past change in ROA, or firm-years in the highest decile based on level of ROA and the highest decile based on the past change in ROA. Not surprisingly, M3 and M5 are the only ones that produce acceptable test statistics for these samples.

performance over several of the past years, and/or researchers might be interested in assessing future operating performance over several years.

1. Two-Year Changes in Past ROA

Following Fama and French (2000), I measure past change in ROA over the past year. There is no theoretical foundation for using one-year changes in past ROA rather than changes over several years, but one practical advantage of using one-year changes is that only one year of past data is required. Thus, using only one year of past data is less likely to impose sample-selection bias by excluding firms with insufficient past data.

Nevertheless, I also perform my experiments using two-year changes in past ROA. The results show that if the sample firms exhibit extreme two-year changes in past ROA, all the methods yield more biased test statistics than they would if the sample firms exhibit extreme one-year changes in past ROA. Further, in these sampling situations, M1, M2, M3, and M5 yield less biased test statistics if they control for changes in ROA over the past two years rather than over only the past year. These results suggest that the period over which the sample firms experience abnormal changes in ROA could be important. In any event, M3 and M5 still produce the least biased test statistics.

2. Two-Year Changes in Future ROA

Researchers typically investigate the operating performance of firms for several years after corporate events. Hence, I extend the earlier analysis using two-year future changes instead of one-year future changes in ROA. The results (not reported here) show the same tendencies as reported earlier. The main difference is that if a method generates poorly specified test statistics for a one-year future change, then the same method tends to generate even worse test statistics for a two-year window. In other words, any misspecification appears to intensify when future changes in ROA are measured over a longer window.

3. Sample Size

Although the results reported in this study are based on samples of 50 observations, most empirical studies of operating performance use larger samples. Obviously, larger samples result in greater statistical power, and any misspecification could be more pronounced. Simulations that use 100 and 200 observations rather than 50 confirm this reasoning. In addition, the rejection levels are higher for samples with industry clustering, perhaps because it is harder to generate a control sample with similar characteristics (including SIC code) when a larger portion of the firms in the industry are already in the original sample.

4. NYSE/Amex firms

As noted earlier, Barber and Lyon (1996) only include NYSE and Amex firms in their simulations. When I only include these firms in my simulations, the results are qualitatively similar to those reported here. The only exceptions are that my methods generate slightly more correctly specified test statistics for the largest size decile and the bias for the lowest size decile becomes slightly positive.

V. Conclusion

Many studies in accounting and finance analyze operating performance following corporate events. The problem is that future changes in operating performance are partially predicted by individual firm characteristics. Hence, any observed changes in performance after a corporate event could be attributable either to the event itself or to predictable patterns in performance. To control for the second possibility, performance studies must use control samples with similar

Table VI. Rejection Rates for Samples with Pre-Specified Characteristics

This table presents the percentage of 10,000 random samples of 50 firms that reject the null hypothesis of no abnormal operating performance at a theoretical significance level of 5%. The rejection rates are separated for the alternative hypotheses of negative performance (the estimated test statistic is less than the theoretical test statistic with a cumulative probability of 0.025) and positive performance (the estimated test statistic is larger than the theoretical test statistic with a cumulative probability of 0.975). I draw random samples from either the lowest or highest deciles of past year's change in ROA. I further partition the samples into firms with ROA below the median (Panel A) and firms with ROA above the median (Panel B). The first method for generating control firms is based on industry classification and the pre-event ROA. The second method is based on industry classification and the one-year pre-event change in ROA. The third method is based on industry classification, the pre-event ROA, and the one-year pre-event change in ROA. The fourth method is based on industry classification, the pre-event ROA, and the pre-event market-to-book ratio. The fifth method is based on industry classification, the pre-event ROA, the one-year pre-event change in ROA, and the pre-event market-to-book ratio. All tests are based on non-parametric (Wilcoxon) test statistics. The rejection levels are significantly different from the theoretical level of 2.5% if they fall outside the range 2.1%-2.9% (α=0.01), in which case the figures are marked with *. (The range for α=0.05 is 2.2%-2.8%.)

	Samples from Lowest Decile of Past ROA Change		Samples from Highest Decile of Past ROA Change	
	0.025	0.975	0.025	0.975
Panel A. Samples with ROA Below the Median				
1. Matched on ROA	0.7*	6.3*	2.9	0.3*
2. Matched on ROA Change	1.1*	7.6*	0.1*	16.5*
3. Matched on ROA and ROA Change	1.8*	1.9*	0.8*	3.3*
4. Matched on ROA and M/B	1.3*	4.0*	4.2*	0.6*
5. Matched on ROA, ROA Change, and M/B	2.6	1.8*	1.2*	3.2*
Panel B. Samples with ROA Above the Median				
1. Matched on ROA	13.1*	0.1*	11.4*	0.1*
2. Matched on ROA Change	58.6*	0.0*	15.5*	0.1*
3. Matched on ROA and ROA Change	1.5*	0.9*	2.5	0.9*
4. Matched on ROA and M/B	13.3*	0.1*	12.4*	0.1*
5. Matched on ROA, ROA Change, and M/B	1.3*	1.6*	2.7	1.5*

characteristics. This study investigates five methods for generating such control samples. The results indicate that the methods of generating control samples produce very different test statistics in various sampling situations. In fact, the test statistics are often severely biased. The best overall method generates control firms with similar prior levels of performance, similar prior changes in performance, and similar market-to-book ratios. Of course, there might be situations in which it is impractical or impossible to find control firms that are similar along all dimensions. If so, the results presented here suggest that it is generally more important to control for prior changes in performance than for the other criteria.

The results have critical implications for researchers who study operating performance following corporate events, because failing to choose a proper method for generating control firms could lead to faulty conclusions about post-event abnormal operating performance. Researchers should first examine the pre-event performance patterns and market-to-book ratios of the sample firms. If the sample firms exhibit extreme characteristics, then researchers should choose a method for generating control firms that ensures that the control firms share these extreme characteristics. A further implication is that even though the overall sample does not exhibit extreme characteristics, researchers should not separately examine subsamples with extreme pre-event performance patterns or market-to-book ratios unless they match the control sample on these characteristics.■

References

Ali, A. and P. Zarowin, 1992, "The Role of Earnings Levels in Annual Earnings-Return Studies," *Journal of Accounting Research* 30, 286-296.

Barber, B. M. and J. D. Lyon, 1996, "Detecting Abnormal Operating Performance: The Empirical Power and Specification of Test Statistics," *Journal of Financial Economics* 41, 359-399.

Basu, S., 1997, "The Conservatism Principle and the Asymmetric Timeliness of Earnings," *Journal of Accounting and Economics* 24, 3-37.

Benartzi, S., R. Michaely, and R. Thaler, 1997, "Do Changes in Dividends Signal the Future or the Past?" *Journal of Finance* 52, 1007-1034.

Burgstahler, D. and I. Dichev, 1997, "Earnings Management to Avoid Earnings Decreases and Losses," *Journal of Accounting and Economics* 24, 99-126.

Denis, D. J. and D. K. Denis, 1995, "Performance Changes Following Top Management Dismissals," *Journal of Finance* 50, 1029-1057.

Denis, D. J. and T. A. Kruse, 2000, "Managerial Discipline and Corporate Restructuring Following Performance Declines," *Journal of Financial Economics* 55, 391-424.

Easton, P. D. and T. S. Harris, 1991, "Earnings as an Explanatory Variable for Returns," *Journal of Accounting Research* 29, 19-36.

Fama, E. F. and K. R. French, 2000, "Forecasting Profitability and Earnings," *Journal of Business* 73, 161-175.

Friday, H. S., S. D. Howton, and S. W. Howton, 2000, "Anomalous Evidence on Operating Performance Following Seasoned Equity Offerings: The Case of REITs," *Financial Management* 29, 76-87.

Healy, P. and K. Palepu, 1988, "Earnings Information Conveyed by Dividend Initiations," *Journal of Financial Economics* 21, 149-175.

Lie, E. and J. J. McConnell, 1998, "Earnings Signals in Fixed Price and Dutch Auction Self-Tender Offers," *Journal of Financial Economics* 49, 161-186.

Loughran, T. and J. R. Ritter, 1997, "The Operating Performance of Firms Conducting Seasoned Equity Offerings," *Journal of Finance* 52, 1823-1850.

McLaughlin, R., A. Safieddine, and G. K. Vasudevan, 1998, "The Information Content of Corporate Offerings of Seasoned Securities: An Empirical Analysis," *Financial Management* 27, 31-45.

Mikkelson, W. H., M. M. Partch, and K. Shah, 1997, "Ownership and Operating Performance of Companies that Go Public," *Journal of Financial Economics* 44, 281-307.

Nohel, T. and V. Tarhan, 1998, "Share Repurchases and Firm Performance: New Evidence on the Agency Costs of Free Cash Flow," *Journal of Financial Economics* 49, 187-222.

Rangan, S., 1998, "Earnings Management and the Performance of Seasoned Equity Offerings," *Journal of Financial Economics* 50, 101-122.

Shah, K., 1994, "The Nature of Information Conveyed by Pure Capital Structure Changes," *Journal of Financial Economics* 36, 89-126.

Teoh, S. H., I. Welch, and T. J. Wong, 1998, "Earnings Management and the Underperformance of Seasoned Equity Offerings," *Journal of Financial Economics* 50, 63-99.