Journal of Accounting Research
Vol. 22 Supplement 1984
Printed in U.S.A.

A Comparison of Event Study Methodologies Using Daily Stock Returns: A Simulation Approach

THOMAS DYCKMAN,* DONNA PHILBRICK,† AND JENS STEPHAN††

1. *Introduction*

Accounting research studies commonly use measures of abnormal security returns to test hypotheses about accounting information and policies. Alternative methods are used in these studies to detect abnormal performance and thereby examine the information content of identified events. The ability of these methods to detect abnormal returns using simulation techniques is examined by Brown and Warner [1980; 1984]. Morse [1984] addresses this issue analytically. The analytical approach presumes that the event study data is generated by the process modeled, while the simulation approach uses the security return data generated by the market. Simulation provides a tractable means of dealing with situations where an analytical approach may yield results suggesting direction but not magnitude or where such techniques are unusually cumbersome. Both limitations face us in this study.

Brown and Warner [1980] use monthly data to examine the comparative abilities of several combinations of return-generating models and

* Cornell University; † University of Oregon; †† Duke University. We acknowledge the valuable comments of Professors John Elliott and Dale Morse of Cornell, the doctoral students and the Accounting Workshop at Cornell University, Professors Robert Holthausen, Richard Leftwich, and Katherine Schipper of the University of Chicago, Professor James Patell of Stanford University, Professor Pat O'Brien of the University of Rochester, and Professor Larry Dann of the University of Oregon. Remaining errors are the authors' responsibility.

testing procedures to identify the presence of abnormal returns under a variety of event conditions. Event conditions include event-date uncertainty, portfolio size, the magnitude of abnormal returns, event-date clustering, and risk clustering.

Extension of the Brown and Warner findings to daily data [1984] is not clear-cut. Fama [1976] shows that daily returns are more nonnormal than monthly returns. Further, Scholes and Williams [1977] argue forcefully that nonsynchronous trading produces severe problems in estimating beta using daily data.

Brown and Warner [1984] extend their 1980 analysis to daily data. They find that for random samples and short time periods, the market model is not systematically better at identifying the presence of abnormal returns than models that do not incorporate market-wide factors and firm-specific risk. Also, the nonnormality of daily returns has no substantive impact on the power of their test for the various methods studied. However, the choice of the variance-estimation procedure for the "t" tests proves critical.[1] The authors examine serial dependence of excess returns, cross-sectional dependence of excess returns, and the increase in variance of returns around events for the effects on the interpretation of hypothesis tests. Of these issues, the third appears to be the most problematical.

Our study examines the interaction of portfolio size, event-date uncertainty, and the magnitude of abnormal performance over ranges of all three variables simultaneously to determine if these factors have a significant effect on the researcher's ability to detect abnormal performance. This examination is conducted using a naive model as well as with market and risk-adjusted models. Although our study was conducted independently of Brown and Warner [1984], there are similarities in the approaches and in the conclusions. Portfolio size, event-date uncertainty, and the magnitude of abnormal performance are simultaneously manipulated in both studies, although different levels of these variables are examined. Both studies reach similar conclusions about the impact of nonnormality in excess returns on the statistical tests. No improvement in either the specification or the power of the tests is reported in either study when the Scholes-Williams or the Dimson method of estimating beta is used to allow for the nonsynchronous trading problem. This result is, at least in our work, partly due to the sample and data selection process.

[1] Brown and Warner [1984] use three different variance estimation procedures. First, the standard deviation is estimated from the time series of mean excess portfolio returns over the estimation period. This technique takes into account any cross-sectional dependence in the security-specific excess returns. In their second approach, Brown and Warner use a variance estimation procedure which assumes cross-sectional independence. They standardize daily firm residuals using a time-series variance computed in the estimation period. Finally, Brown and Warner estimate the variance of the mean excess return using only cross-sectional event-period excess returns.

There is also the question of whether the residuals should be standardized. Patell [1976] standardizes residuals using the variance calculated over the estimation period. We first establish the level of abnormal returns and then standardize the firm residuals using the standard deviation calculated over a 120-day test period surrounding the announcement. This procedure is a time-series approach, but it is unlike any of those used by Brown and Warner [1984]. It is comparable to Patell's [1976] use of the estimation period to calculate the variance. Inducing a constant level of abnormal performance to each firm is equivalent to adding a constant to the return random variable. Thus our use of a 120-day test period to estimate the variance does not reflect any increase in variance around the event date caused by the information event. (See Beaver [1968] and Patell and Wolfson [1979; 1981; 1982] for evidence suggesting such an increase occurs for certain types of events.) Hence, in our research, the variance is independent of the level of induced abnormal performance across simulations.

The presence of cross-sectional correlation in the specific security excess returns (Brown and Warner [1980], Beaver [1981], and Collins and Dent [1984]) can affect the power of statistical tests if the method of variance estimation assumes cross-sectional independence of returns. If positive cross-sectional correlation is present, such tests result in inappropriately high rejection rates of the null hypothesis whether the null is true or false. Brown and Warner [1984, p. 23] note that "when there is no clustering of event dates, the gains from assuming independence are substantial [which was not the case using monthly returns]." They find, for example, rejection rates approximately equal to the significance level of the test when no abnormal performance is present. The results for our sample are described in section 5.

Both Brown and Warner [1984] and this study examine event-date uncertainty. The basic method in both studies randomly selects the event date from the uncertain event period. This paper extends the empirical work by considering an analytically justifiable alternative approach. Both methods assume that the probability of the event falling on a particular day is uniformly distributed over the uncertain event period. However, under the first method, abnormal returns affect the test statistic only if they are selected to coincide with the event date. Under the second method, the residuals are accumulated over the uncertain event period and the test statistic always reflects the information event, although it is diluted by the inclusion of nonevent residuals. Our results suggest that accumulating residuals has an advantage when uncertainty exists about the event date. Third, we consider the effect of estimating the market model using a multiday estimation period rather than a single day. We find the multiday market model gives results that are essentially equivalent to using a daily beta and accumulating residuals.

In section 6 we consider the impact of event clustering by industry and by time. Section 7 investigates the normality assumption. Since daily residuals are not normally distributed, the simulation techniques used by Brown and Warner are extended to estimate the magnitude and direction of the bias resulting from the use of the "t" test.

2. Return-Generating Models

The empirical results from this study are based on the following model of the prediction error in returns:

$$\epsilon_{it} = R_{it} - (\alpha_i + \beta_i R_{mt}) \tag{1}$$

where:

$R_{it} \equiv$ the actual return on security i for day t,

$R_{mt} \equiv$ the market return for day t,

$\epsilon_{it} \equiv$ prediction error for security i for day t, and

α_i, $\beta_i \equiv$ firm-specific constants.

R_{it}, R_{mt}, and ϵ_{it} are random variables.

Two alternative parameter estimation approaches are used. The first estimation approach uses daily returns to estimate the parameters. The second approach uses the return over a consecutive three-day period to estimate the model parameters.

In this paper we consider five return models based on expression (1). The first three are the same three models used by Brown and Warner. Models using raw returns are not considered here.

Model (1): Mean-Adjusted Returns Model: This is labeled the naive model insofar as market-wide factors and risk are not accounted for explicitly. The predicted return for a security is equal to a constant, estimated by averaging a series of past returns. In equation (1) α is set equal to the average return over the estimation period and β is set equal to zero.

Model (2): Market-Adjusted Return Model: The expected firm return is equal to the market return for that period. Expected returns are constant across securities but not across time. In other words, α is set equal to zero and β is set equal to one.

Model (3): Market Model: In this model the expected firm return is a linear function of the market return using an *OLS* beta. In this case α and β are calculated over the estimation period using *OLS*.

For all three models, the prediction error for security i on day t is defined by equation (1) as the difference between the actual return and the expected return. We then simulate abnormal returns by adding an additional return (in ways to be described) to the prediction error. The three models are compared to determine if explicit recognition of market-wide factors or risk has important implications for measuring abnormal security returns in event studies using daily data. Since the method of estimating risk may need to allow for the nonsynchronous trading problem inherent in models using daily returns data, two extensions of the Market Model are also evaluated for their ability (relative to the Market Model) to detect excess returns. These two extensions are:

Model (4): Scholes-Williams Beta Model: Variation of the Market Model, (3), using the Scholes-Williams [1977] method of estimating beta.

Model (5): Dimson Beta Model: Variation of the Market Model, (3), using the Dimson [1979] method of estimating beta.

3. *Event Conditions*

Three event condition variables which may have an impact on the ability to detect abnormal performance using daily data are portfolio size, event-date uncertainty, and the magnitude of abnormal returns. For example, larger excess returns are easier to detect than smaller excess returns. Similarly, we should detect the presence of abnormal performance more often using larger portfolios which diversify away the influence of firm-specific factors. Further, the likelihood of detecting abnormal performance should increase as the specification of the event date improves.

We examine the ability of the three return-generating models to detect abnormal performance under various combinations of these event conditions. The specific values chosen for these variables are:

(*a*) portfolio size: 10, 20, 30, 40, 50, 75, and 100 firms
(*b*) event-date uncertainty:[2] 1, 2, 3, 4, and 5 days
(*c*) induced abnormal performance, identical for each firm: 0%, 1%, 2%, 3%, 4%, and 5%[3] added to the firm's prediction error.

Event-date uncertainty is introduced in two ways. The first simulates the researcher's attempt to capture the event date by choosing one day at random using a uniform probability distribution from the minimum number of days known to contain the event date. For example, if it is known that the event occurred on one of three consecutive days, the days are identified as −1, 0, 1, with day zero denoted the simulated event date. A day is randomly selected from this uncertain event period. If the day selected is day zero, abnormal performance is introduced. In this case, the probability of correctly specifying the event date is one third.[4]

---

[2] Event-date uncertainty is interpreted as the number of days over which the event could have occurred, i.e., "one day" implies the day on which the event occurred is known; "two days" implies the event could have occurred on either of two days; etc. Equal probability is assigned to each day.

[3] Zero abnormal performance is only needed when there is no event-date uncertainty. Therefore, there are 182 combinations of the three variables. (7 portfolio sizes × 5 event-date uncertainty levels × 5 abnormal performance levels) + (1 abnormal performance level (0%) × 7 portfolio sizes × 1 event-date uncertainty level) = $7 \times 5 \times 5 + 7 = 182$.

[4] Brown and Warner [1984] follow the same procedure in analyzing event-date uncertainty except they use the period −5 to +5 for all their tests. A five-day versus an eleven-day period increases the chance of rejecting the null when it is false.

The second method accumulates firm residuals over the uncertain event period. Therefore, the firm residual that reflects the event is included with certainty, but its effect is diluted by the inclusion of nonevent residuals (two nonevent days in the example above).

The three-day accumulation period suggests using the three-day market-model parameter-estimation approach mentioned in section 2. Together the two alternative parameter-estimation approaches of section 2 and the two means for handling event-date uncertainty lead to three methods by which event-date uncertainty is considered in these simulations. These three methods are:

Method 1: Single-day estimation approach with abnormal returns considered only on the event day (Parameter Estimation Approach 1; means 1 of handling event-date uncertainty).

Method 2: Single-day estimation approach with abnormal returns accumulated over uncertain event period (Parameter Estimation Approach 1; means 2 of handling event-date uncertainty).

Method 3: Multiday (three-day) estimation approach with abnormal returns accumulated over the multiday (three-day) period (Parameter Estimation Approach 2; means 2 of handling event-date uncertainty).

Methods 2 and 3 yield firm residuals with variances greater than the variance of daily residuals. If, for example, there is no serial correlation, the variance of residuals accumulated over three days is three times the variance of the daily residuals.[5] The increased variance makes it more difficult to detect abnormal performance in the portfolio compared to using a single day (if the correct day is known). But if uncertainty about the distribution of the event date exists, the better method for testing for a mean shift is not obvious.

Using a single induced abnormal performance percentage across firms in a given simulation is not realistic. While an event study would encounter some average level of abnormal performance, the specific level would vary across firms. Such differences would be expected if investors have different beliefs or use different valuation models across firms. We could simulate the effect of alternative distributions of abnormal performance but have elected not to do so given the arbitrariness of selecting from the wide range of options and because, as Brown and Warner [1980, p. 212] observe, the current approach "represents a simple case which enables us to focus on the detection of mean shifts when an event takes place, holding constant the conditional variance. The detection of mean shifts is the relevant phenomenon to study when investigating how well

[5] Daily excess returns tend to exhibit serial correlation, possibly due to nonsynchronous trading. The *OLS* estimates of beta are both biased and inconsistent and excess returns more variable, leading to weaker (lower power) tests.

different methods pick up the impact of an event on the value of the firm."[6]

Our study is restricted to random samples with positively induced excess returns due to "good news." However, in the context of a single event study, an earnings announcement could be "good news" for one firm and "bad news" for another. Our methods are applicable to the more general case only if the magnitude of the effect, the uncertainty surrounding the announcement date, and other factors are independent of whether the news is good or bad. Patell and Wolfson [1982] show that this is not necessarily the case. If the distribution of prediction errors is symmetric and expected firm returns are unbiased estimates of actual returns, the sign of the induced abnormal performance (assuming it is the same across all firms) should have no effect on the ability to detect excess returns. Brown and Warner [1980] simulated this condition and found no appreciable difference in results compared to using only positive induced abnormal performance. The argument also presumes the researcher can correctly specify the sign of the expected abnormal return.

4. *Experimental Design*

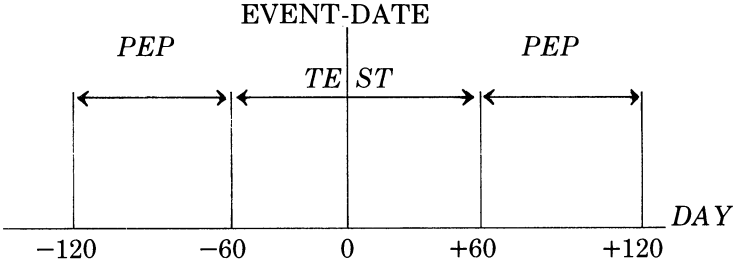
This section describes the sample selection and data generation process.

Step 1: The May 1, 1974–August 31, 1979 period is used and a hypothetical event date is randomly selected for each firm listed on the *CRSP* daily return files. Firm returns for a period of 241 days centered on the event date are extracted. Firms with missing data over the 241-day period are deleted, resulting in 2,069 sample firms.

Step 2: Parameters for each of the five models are estimated using the data for days −120 to −60 and +60 to +120. In the diagram below *PEP* represents the parameter-estimation period [−120:−60] and [60:120]. *TEST* represents the test period [−59:59]. Thus, for Model (1), the expected firm returns for days −59 to +59 are the average of the firms' returns during the *PEP* period. The return is assumed constant across time for any given security. For Model (2), the expected returns are equal to the market returns for days −59 to +59. For Models (3), (4), and (5), the firm and market returns during the parameter-estimation period are used to calculate the appropriate regression parameters. These param-

[6] Morse [1984] shows analytically that the result of uncertainty in the magnitude of the information effect on the conditional variance increases the variance of the mean abnormal return. Any reduction in the variance (from that implied by a uniform distribution) would decrease the variance of the mean abnormal return and increase the ability of the methodologies to detect abnormal performance. In interpreting this work one needs to keep in mind the world Morse models. For example, Morse implicitly assumes that the cross-sectional variances in the estimation period are identical. Our simulation procedure, on the other hand, standardizes the residuals on a firm-by-firm basis.

eters, in conjunction with the market returns for days −59 to +59, are used to generate expected returns for the test period.



Step 3: The simulation procedure replicates portfolios 250 times. With a portfolio size of 50, for example, the number of prediction errors required is 50 × 250 = 12,500. Since there are only 2,069 firms in the sample, ten hypothetical event dates and their prediction errors are extracted from days −59 to +59. These are chosen at 12-day intervals to minimize any serial correlation of prediction errors and, in essence, represent ten independent event-date prediction errors for each firm. Recall that a random date is first chosen for each firm, therefore there is no systematic relationship of prediction errors between firms. We also calculate the standard deviation of each firm's prediction errors for days −59 to +59. When introducing event-date uncertainty by Method 2 in section 3, we accumulate three (or five) consecutive prediction errors centered on day zero at 12-day intervals, again using ten hypothetical event dates for each firm.

Step 4: The next step is to form portfolios of event-date residuals reflecting the various combinations of event conditions previously described.

The procedure is illustrated in the following example for Method 1, and then the alternate steps needed for Methods 2 and 3 in which the residuals are cumulated are given.

EXAMPLE

1. Event conditions (selected by researcher):
   portfolio size: 10
   event-date uncertainty: 3 days
   induced abnormal performance: 3%
2. Select a firm at random with replacement from the population of 2,069 firms. (Each has ten associated event dates.)
3. Randomly select an event date and its associated prediction error with replacement from the ten available.
4. Choose a simulated event date at random; −1, 0, 1 where 0 is the actual event date. (If a four-day event uncertain period is used, the days used for this simulation are −2, −1, 0 and +1. For an even number of days, the extra day is always included before day zero.)

5. If and only if the event date chosen is day 0, add .03 to the firm's prediction error and divide by the firm's standard deviation.[7] This is the case where, under three-day uncertainty, the abnormal return falls on the event day. Probabilistically this occurs one-third of the time when we are testing the standardized abnormal returns for significance.

6. If the event date chosen is not day 0, divide the firm's prediction error (selected in step 3 of this example) by the firm's standard deviation.

7. Either step 5 or step 6 above yields the standardized return for the first firm in the portfolio.

8. Repeat steps 2 through 7 until a portfolio of ten firm observations is formed.

9. Sum the standardized returns and divide by the square root of the portfolio size to obtain the portfolio t statistic.

10. Repeat steps 8 and 9 250 times and observe the percentage of portfolio returns indicating the presence of abnormal performance. We assume that firm residuals are distributed normally. Therefore, portfolio residuals are approximated by a t-distribution with the appropriate degrees of freedom. In this case, the presence of abnormal performance can be observed by counting the portfolio residuals greater than the 5% critical value (i.e., greater than 1.833 for this example).

11. Repeat step 10 for all other combinations of event conditions.

Steps 5 through 7 differ for Methods 2 and 3 which use cumulative residuals. The alternate steps for Method 2 are:

1*a*. Cumulate the abnormal returns for the firm consistent with the equation $\sum_{it} e_{it}$ where the sum is taken over the number of days in the uncertain event period. In the three-day uncertainty example $t = -1, 0, 1$.

2*a*. Add the abnormal performance (.03) to the sum of the three days of abnormal returns.

3*a*. Divide by the square root of three times the firm's variance. This yields the cumulative standardized residual for the first firm in the sample.

The alternate steps for Method 3 are:

1*b*. Add .03 to the firm's residual. This residual is computed using the three-day market model.[8]

---

[7] Adding 3% to the firm's residual is equivalent to adding 3% to the firm's return and then subtracting the expected return computed using the appropriate return-generating function. Adding a constant to a firm's return and then standardizing the resulting residual causes a smaller impact on more volatile firms. This may not be realistic in some cases but seems to be more appropriate than adding a standardized constant. In fact, the shock impact is itself a random variable whose distribution across firms is unknown. Again this simple case allows us to examine the effect of hypothetical information events on mean shifts of firm returns using different methods.

[8] The parameter estimation procedure for the three-day Market Model utilizes three-day returns. These returns are calculated consistent with the equation $\prod_t (1 + R_{it}) - 1$, where the product is taken over days -1 to $+1$. The test period three-day residual is the difference between the actual three-day return as computed using this equation and the expected return based on the three-day Market Model. The standard deviation is calculated using the 39 three-day residuals in the original 118-day test period.

2*b*. Divide the augmented residual by the firm's standard deviation. This standard deviation is based on three-day residuals in the test period. This procedure yields the cumulative standardized residual for the first firm.

5. *Discussion of Results*

One of our objectives is to quantify the effects of portfolio size, magnitude of abnormal performance, and event-date uncertainty on the ability to detect excess returns. In order to make the analysis manageable, we look at the effects in paired fashion. We first consider Method 1 for introducing event-date uncertainty and later compare the results with Methods 2 and 3. The three pairs are:

(*a*) Portfolio size and event-date uncertainty
(*b*) Abnormal performance magnitude and portfolio size
(*c*) Event-date uncertainty and abnormal performance magnitude

Additionally, we compare results across the three return-generating models: the Mean-Adjusted Returns Model (Model 1), the Market-Adjusted Returns Model (Model 2), and the Market Model (Model 3). The "t" test is used to identify the presence of abnormal performance and we defer until section 7 an analysis of size and direction of bias in the "t" test due to the nonnormality of daily return residuals. Also, we compare the ability to detect abnormal performance using the three methods detailed in section 3 which allow for event-date uncertainty. Finally, we examine any benefits which result from the use of the Scholes-Williams and Dimson methods of estimating beta.

EFFECTS OF PORTFOLIO SIZE, EVENT-DATE UNCERTAINTY, AND ABNORMAL PERFORMANCE MAGNITUDE

Table 1 gives the percentage of portfolios rejected at the 0.05 level when there is no abnormal performance. The numbers represent the percentage of the 250 portfolio replications that indicate abnormal performance, using a 5% t-test when, in fact, none is present. On average, 5% of the portfolios would be expected to show abnormal performance. For 250 replications, the 95% confidence limits for the percentage of rejections are 2% and 8%. The observations for all three models lie within this range, although they appear to be slightly biased downward. This bias is at least partially explained by the fact that the population of residuals for the 2,069 firms is slightly skewed to the left: the mean is −.005 while the median is −.04. Furthermore, the tendency to reject a true null too often due to cross-sectional correlation seems to be unimportant, as suggested in Brown and Warner [1984]. We conclude that tests using each of the three models are well specified in rejecting the existence of abnormal performance when it is not present.

Table 2 isolates the event-date uncertainty, portfolio size effects for the Market Model (Model 3) using Method 1 to handle event-date

TABLE 1

Percentage of Portfolios Indicating the Presence of Abnormal Performance[a]
Abnormal Performance = 0%
Method 1

Portfolio Size	Model		
	Mean-Adjusted Return	Market-Adjusted Return	Market Model
10	4	5	3
20	4	4	5
30	4	4	4
40	5	3	6
50	4	6	4
75	4	5	3
100	2	4	4
Average	3.9	4.4	4.1

[a] Based on 250 replications for each portfolio size. The 95% confidence interval about the expected percentage of 5 under the null hypothesis is from 2% to 8% for each entry in the exhibit. None of the values in the exhibit is, therefore, significant.

TABLE 2

Percentage of Portfolios Indicating the Presence of Abnormal Performance:[a]
Market Model
Induced Abnormal Performance = 3%
Method 1

Event-Date Uncertainty Days	Portfolio Size (No. of Firms): 10	20	30	40	50	75	100
5	22	39	48	56	63	75	86
4	32	44	58	65	75	91	98
3	42	62	82	88	91	98	100
2	66	92	97	100	100	100	100
1	99	100	100	100	100	100	100

[a] 250 portfolios per cell.

uncertainty. The number in the cells represent the percentage of portfolios that indicate the presence of abnormal performance at the 0.05 probability level when an abnormal performance level of 3% is actually present. As expected, more uncertainty about the event date makes it more difficult to detect abnormal performance across all portfolio sizes. Also, for a given level of event-date uncertainty, larger portfolios more accurately detect the presence of abnormal performance. The importance of the interaction of the two factors is striking. For instance, with a portfolio size of ten, the probability of detecting abnormal performance drops from 0.99, almost certainty, to only 0.22 when event-date uncertainty increases from one to five days. The results underline the importance of establishing the precise event date, as discussed by Dodd et al. [1984]. Similarly, with five days uncertainty about the event date, increasing portfolio size from 10 to 100 triples (0.26 to 0.86) the probability of detecting abnormal performance. Increasing portfolio size mitigates the problem of event-date uncertainty. Similar conclusions, not reported here, result for Model (1), Model (2), and for other levels (0.01, 0.02, 0.04, and 0.05) of induced abnormal performance.

Table 3 isolates the abnormal performance, portfolio size effects for the Market Model. The trends are as expected. Larger portfolios are better able to detect abnormal performance and abnormal performance is more easily detected, the greater its magnitude. We also observe the inability of the Market Model to reject a false null properly when the true level of abnormal performance is low, particularly with small portfolio sizes. Indeed, abnormal performance below 1% with even minimal event-date uncertainty will often go undetected even with relatively large portfolios. Results similar to those described in this paragraph are observed for other levels of event-date uncertainty. Further, all three models have essentially the same ability to detect abnormal returns.

Table 4 examines the abnormal performance, event-date uncertainty effects for the Market Model. Portfolio size is fixed at 50 firms. As expected, larger magnitudes of induced abnormal performance are more easily detected and increases in event-date uncertainty decrease the likelihood of detection. Once again, all models seem to perform about equally well. Low levels of abnormal performance coupled with event-date uncertainty will often go undetected even with reasonably large sample sizes.

The previous exhibits portray only part of the picture since one of the variables is always held constant. Figure 1 synthesizes the paired effects demonstrated in the previous exhibits and shows the interaction of the individual effects as well as summarizing additional results not previously reported. The numbers in the cells represent the minimum magnitude of abnormal performance detectable for various combinations of event-date

TABLE 3

Percentage of Portfolios Indicating the Presence of Abnormal Performance:[a]
Market Model
Event-Date Uncertainty = 3 Days
Method 1

Induced Abnormal Performance (Percent)								
	5	66	90	98	97	100	99	100
	4	56	74	93	97	98	100	100
	3	42	62	82	88	92	98	100
	2	24	45	58	62	68	87	94
	1	12	14	16	27	29	38	51
Portfolio Size (No. of Firms)		10	20	30	40	50	75	100

[a] 250 portfolios per cell.

TABLE 4

Percentage of Portfolios Indicating the Presence of Abnormal Performance:[a]
Market Model
Portfolio Size = 50 Firms
Method 1

Induced Abnormal Performance (Percent)	Event-Date Uncertainty (Days): 1	2	3	4	5
5	100	100	100	95	88
4	100	100	98	88	84
3	100	100	91	75	63
2	100	98	68	54	37
1	96	48	29	26	16

[a] 250 portfolios per cell.

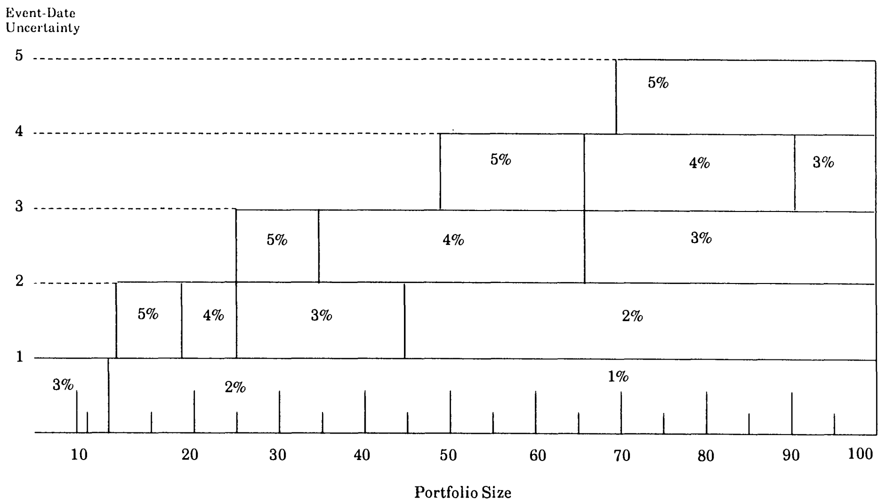


FIG. 1.—Amount of abnormal performance detectable 95% of the time using the market model (method 1). For example, using method 1 and a portfolio size of about 35 or larger, the researcher can expect to detect an abnormal performance of 4% with three days uncertainty about when the event took place 95% of the time. If the portfolio increases to 68 or more, the researcher can detect an abnormal performance level of 3% with the same confidence. (The portfolio sizes when the returns are accumulated over the three-day period drop to below 20 for 3% abnormal performance and to about 10 for 4%. These values are not in the exhibit.)

uncertainty and portfolio size. The presence of abnormal performance in a portfolio is again measured using the "t" test and a 5% significance level. The criterion for successfully detecting the existence of abnormal performance across many portfolios is that at least 95% of the portfolios have to indicate the presence of abnormal performance. For example, 95% of the time a portfolio of size 20 or greater will detect 4% abnormal performance with two days uncertainty about the event date. For portfolio sizes greater than 20 or abnormal performance above 4%, the probability of detecting abnormal performance with two-day event-date uncertainty increases above 95%. Similarly, 95% of the time, a portfolio of size 50 detects 5% abnormal performance with four days uncertainty about the event date. In all cases, larger portfolio sizes or greater magnitudes of abnormal performance or less event-date uncertainty make it easier to detect the presence of abnormal performance. Figure 1 serves as a guide for designing experiments and evaluating the results of event studies involving random samples when the researcher can judge the likelihood of an event generating the specific level of abnormal performance. The trade-offs are also illustrated in figure 2.

Brown and Warner [1984] arrive at similar conclusions. However, the results are not directly comparable due to differences in experimental design. The closest comparable values are given below for the percentage of samples of size 50 indicating abnormal performance using a 0.05 probability level and a "t" test:

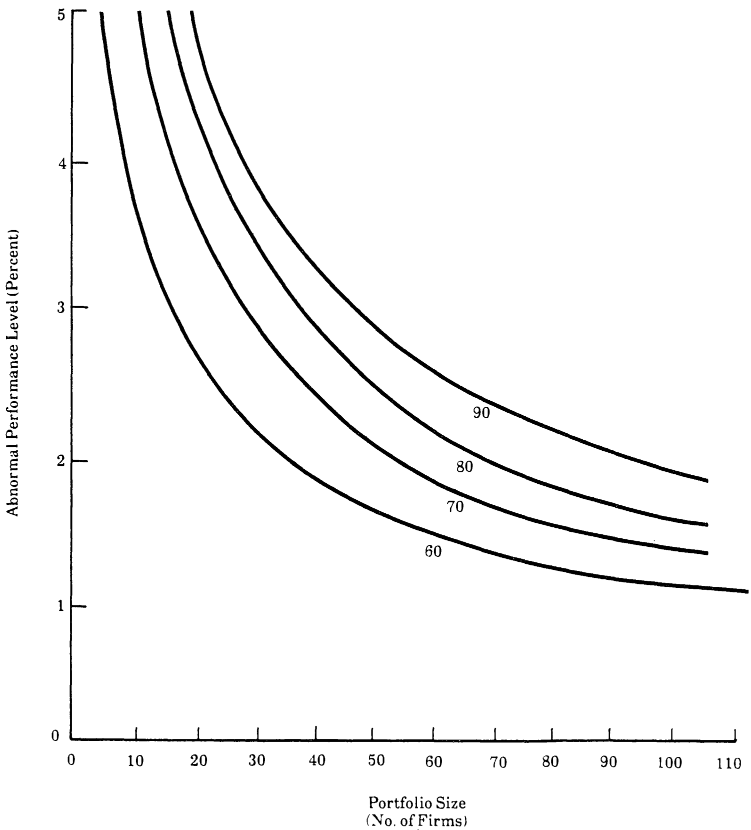


FIG. 2.—Percentage of portfolios indicating the presence of abnormal performance: market model. Event-date uncertainty = 3 days (method 1). Plot based on table 3.

Case	B & W	DP & S
a. No abnormal performance:		
Mean-Adjusted Return	5.2	4
Market-Adjusted Return	4.8	6
Market Model	4.8	4
	B & W	DP & S
b. 1% abnormal performance:	no event-date	uncertainty
Mean-Adjusted Return	96.0	97
Market Model	97.6	96

In "*b*" Brown and Warner use time-series variance estimation assuming independence which is most comparable to what is used in this study.

COMPARISON OF RETURN-GENERATING MODELS

Table 5 indicates the relative performance of the models in percentage terms when applied to random samples. These comparisons reflect all combinations of event conditions, not just the ones presented in tables 2, 3, and 4. The numbers reflecting 0% abnormal performance are not included in the comparisons of table 5 since larger rejection proportions are not better in that instance. Model 1 "better" than Model 2 means that for a combination of three event conditions (event-day uncertainty, portfolio size, and the induced level of uncertainty) Model 1 has a higher rejection frequency (is more powerful) than Model 2. Table 5 indicates this is true in one third of the combinations.

Table 5 depicts the percentage of times (out of $175 = 7 \times 5 \times 5$ event conditions) that one model performs "better than, equal to, or worse than" another model. Binomial tests on these data (omitting ties) suggest that there is no significant difference between Models 1 and 2. However, tests on the data in panels 2 and 3 indicate Model 3, the Market Model, performs significantly better (at the 5% level) than either the Mean-Adjusted or Market-Adjusted returns models. While Model 3 is significantly better, the difference does not appear to be important since the rejection percentages are quite close in all cases. On average, the differences come to less than two percent for any combination of event conditions. A sufficiently large sample will show such significance regardless of its importance (given a sufficiently small variance). We caution that our statistical results are based on a 120-day estimation period. Other estimation period durations are not simulated in this study.

COMPARISON OF METHODS FOR DEALING WITH EVENT-DATE UNCERTAINTY

Event-date uncertainty is an important problem in studies of the type examined here (see Dodd et al. [1984]). If the event date is known with certainty, tests which incorporate residuals over longer periods are less powerful in detecting abnormal performance than tests specific to the day of the announcement. Our treatment of this issue reflects the relative strength of three approaches when event-date uncertainty is present. The comparison allows the researcher to estimate the trade-off in selecting a short event period which might not include the event date for a substantive subset of the sample versus selecting a longer event period that has a large probability of containing the event date.

Our tests compare the following:

Method 1: Abnormal performance is added to a firm's residual return only if the randomly selected event day corresponds to day zero.

Methods 2 and 3: Abnormal performance is added to each firm's day zero residual return. The residual returns for each firm are then accumulated over the uncertain event-day period. Method 2 uses a single-day Market Model, while Method 3 uses a multi-day Market Model.

Our results compare tests conducted using Method 1 with results using Method 2 or 3. The procedure implicitly assumes the researcher knows the timing of the uncertain period.

The problem of event-date uncertainty under a uniform probability distribution and no serial correlation is addressed by Morse [1984] using an analytical approach. He compares the ability to detect excess returns using daily and monthly data. Morse's analytical development, using monthly trading days, can be generalized to returns accumulated over any specified number of days. Under his assumptions, it is preferable to accumulate residuals over the uncertain event period rather than attempt to select a particular event day under uncertainty.[9]

The above analytical result is sustained by our data. Compare columns 3 and 4 of table 6; the former uses Method 1 for capturing event-date uncertainty, while the latter uses Method 2. Accumulating residuals over the three days uncertainty is preferable to choosing one of the three days at random as the event date. (The presence of abnormal performance is detected more frequently using Method 2.) The same results hold for Method 2 with a five-day uncertain period. The power differential is less dramatic for large sample sizes and large levels of abnormal performance. When smaller samples and lower levels of abnormal performance are expected the cumulative procedure is to be preferred.

AN ALTERNATIVE METHOD FOR INTRODUCING EVENT-DATE UNCERTAINTY

A logical alternative to the procedures we use to consider event-date uncertainty is to adopt a multi-day Market Model. In this case the Market Model's beta is estimated on the return over the multiday period.

[9] This result holds under the assumptions of no serial correlation in the residuals and the event day falling with equal probability on one of the days during the uncertain period regardless of the method used. Furthermore, confounding events must not occur during the uncertain period, and the Market Model must be "correctly" estimated.

In section 3 this is denoted Method 3. We simulate this method using a three-day period to estimate the Market Model. Further, our simulations use the three-day Market Model only in combination with three-day event uncertainty.

Dann [1981], Shane and Spicer [1983], and DeAngelo and Rice [1983] all use multi-day models to capture event-date uncertainty. For example, DeAngelo and Rice [1983, p. 349] comment: "We measure abnormal returns in two-day increments because, as Masulis (1980) and Dann (1981) have emphasized, the price impact of new information (here an antitakeover proposal) can at best be isolated to fall within a two-day period."[10] Column 5 of table 6 summarizes the results by abnormal performance percentage and portfolio size for the Market Model estimated by combining consecutive three-day periods and assuming three-day event-date uncertainty. In comparing columns 3, 4, and 5 remember that the residuals used to obtain the results in column 5 are accumulated over three days.

We observe that using the three-day Market Model and accumulating the residuals over the three-day uncertainty period yields results which are essentially equivalent to those produced by Method 2. Method 2 also accumulates residuals over the uncertain event period but uses a return-generating model based on a daily beta. Simulations based on different combinations of the event conditions, not reported in table 6, are consistent. The results indicate that the method of measuring risk, single-day estimation or multi-day estimation, does not affect the ability to detect abnormal performance. While the three-day market model seems to do a little better in detecting abnormal performance (comparing

[10] Analytically, we obtain:

$$R_{it} = \alpha_{it} + \beta_{it} R_{mt} + e_{it} \tag{4}$$

which implies:

$$1 + R_{it} = 1 + \alpha_{it} + \beta_{it} R_{mt} + e_{it} \tag{5}$$

where:

$$\begin{aligned} E(e_{it}, e_{jt}) &= 0; \qquad i \neq j \\ E(e_{it}) &= 0 \\ E(e_{it}, R_{mt}) &= 0 \end{aligned}$$

and for a three-day Market Model:

$$\begin{aligned} 1 + R_{it} &= (1 + R_{1it})(1 + R_{2it})(1 + R_{3it}) \\ &= (1 + R_1)(1 + R_2)(1 + R_3) \\ &= 1 + R_1 + R_2 + R_3 + R_1R_2 \\ &\quad + R_1R_3 + R_2R_3 + R_1R_2R_3. \end{aligned} \tag{6}$$

Since it is possible, analytically, for the four cross-product terms to affect the estimation of the Market Model parameters, we examine the question empirically.

columns 4 and 5 of table 6), the results are not statistically different. Figure 3 provides a visual impression. These results are consistent with the conclusions reached when comparing the Mean-Adjusted, Market-Adjusted, and Market Models. Our results are also consistent with those of Shane and Spicer [1983, p. 531] who report "the results were totally consistent with the results . . . based on one-day returns."

COMPARISON OF METHODS OF RISK ESTIMATION

To examine the problem of risk estimation in the presence of nonsynchronous trading, we compare the performance of the Market Model (using an *OLS* beta) to variations using the Scholes-Williams and Dimson methods of estimating beta. Firms are divided equally into low-, medium-, and high-trading volume groups.[11] The simulation utilizes 3% abnormal performance, three days event-date uncertainty, and a portfolio size of 50 firms. The numbers in table 7 represent the percentage of portfolios that indicated the presence of abnormal performance using the 5% *t*-test. The Scholes-Williams and Dimson methods do not increase the ability to detect abnormal performance on daily returns for the thinly traded stocks.[12] The results also do not suggest a frequency of trading effect within our sample. However, this does not settle the issue for two

TABLE 5
Pair Comparisons of Models[a]

Panel 1	
Model (1) "better" than Model (2)	33%
Model (1) "equal" to Model (2)	25%
Model (1) "worse" than Model (2)	42%
	100%
Panel 2	
Model (1) "better" than Model (3)	20%
Model (1) "equal" to Model (3)	26%
Model (1) "worse" than Model (3)	54%
	100%
Panel 3	
Model (2) "better" than Model (3)	21%
Model (2) "equal" to Model (3)	27%
Model (2) "worse" than Model (3)	52%
	100%

[a] Model 1: Mean-Adjusted Returns.
Model 2: Market-Adjusted Returns.
Model 3: Market Model.

[11] The fraction of shares outstanding that were traded in the same calendar year as the hypothetical event date is used as a proxy for trading frequency. An alternative measure is to compare the percentage of days with no trades for firms in our sample with the percentage of nontrading days for the eliminated firms over some other time period.

[12] The Scholes-Williams instrumental variable approach is viable only when there is no serial correlation in the lag between the last trade and the close of the day. Evidence of such correlation is noted by Morse [1984] and is not unexpected.

TABLE 6

Pair Comparison of Rejection Percentages for Methods 1, 2, and 3[a]
Event-Day Uncertainty = 3 Days

Portfolio Size	Abnormal Performance	Method 1	Method 2	Method 3
10	.01	12	19	19
20	.01	14	34	36
30	.01	16	42	48
40	.01	27	53	62
50	.01	29	62	66
75	.01	38	75	82
100	.01	51	89	91
10	.02	27	47	56
20	.02	45	76	84
30	.02	58	89	97
40	.02	62	96	98
50	.02	68	100	100
75	.02	87	100	100
100	.02	94	100	100
10	.03	42	78	85
20	.03	62	98	100
30	.03	82	100	100
40	.03	88	100	100
50	.03	91	100	100
75	.03	98	100	100
100	.03	100	100	100
10	.04	56	96	96
20	.04	74	100	100
30	.04	93	100	100
40	.04	97	100	100
50	.04	98	100	100
75	.04	100	100	100
100	.04	100	100	100
10	.05	66	98	100
20	.05	90	100	100
30	.05	98	100	100
40	.05	97	100	100
50	.05	100	100	100
75	.05	99	100	100
100	.05	100	100	100

# times Method 1 < Method 2	30
# times Method 1 = Method 2	5
# times Method 1 > Method 2	0
# times Method 1 < Method 3	30
# times Method 1 = Method 3	5
# times Method 1 > Method 3	0
# times Method 2 < Method 3	13
# times Method 2 = Method 3	22
# times Method 2 > Method 3	0

[a] Method 1: Single-day estimation with abnormal return only on event day.
Method 2: Single-day estimation with abnormal return accumulated over three days.
Method 3: Three-day estimation with abnormal returns accumulated over three days.

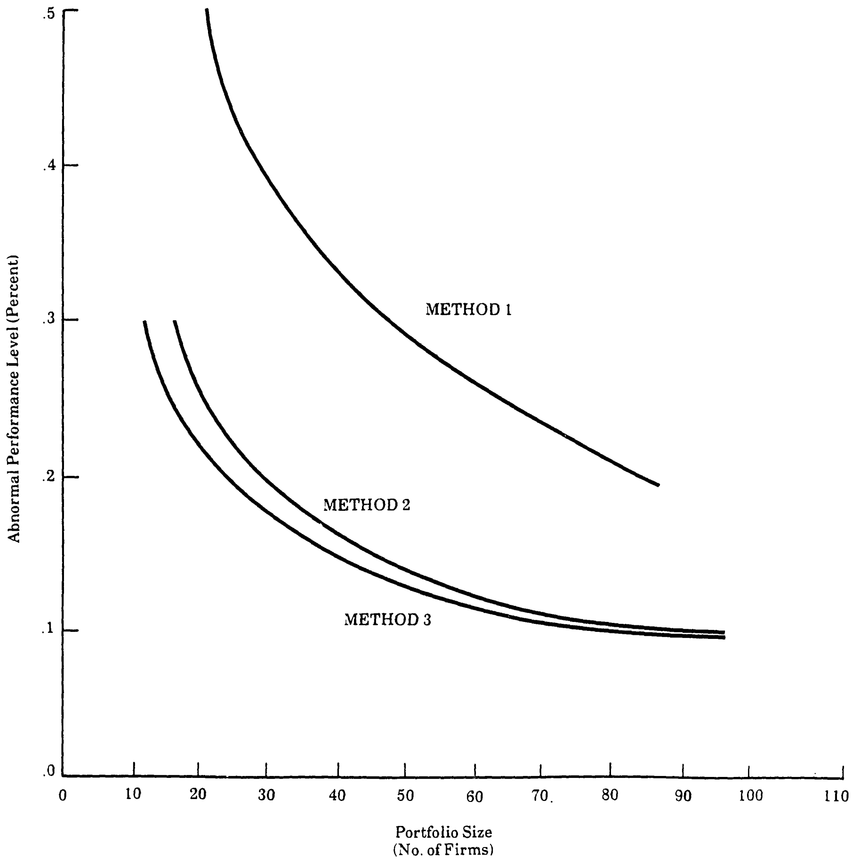


FIG. 3.—Portfolios indicating the presence of abnormal performance 90% of the time: market model. Event-date uncertainty = 3 days (methods 1, 2, and 3). Plot based on table 6.

TABLE 7

Percentage of Portfolios Indicating the Presence of Abnormal Performance[a]
Abnormal Performance = 3%
Event-Date Uncertainty = 3 Days
Portfolio Size = 50 Firms

Trading Frequency	Models		
	(4) Scholes-Williams	(5) Dimson	(3) *OLS*
Low	94	93	94
Medium	94	94	96
High	90	90	90
Average	93	93	93

[a] The numbers give the percentage of portfolios that indicate the presence of abnormal performance using the 5% probability level and a "t" test. Each is based on 250 portfolios.

reasons. First, our sample is based initially on firms included in the *CRSP* daily return file, themselves biased toward more frequently traded stocks. Second, firms were excluded if they had missing data, which reduces the impact of less frequently traded issues. The deletion of firms with missing data may introduce an active-firm, large-firm trading bias. We compared the size of firms in our sample with firms which had been excluded from our sample due to missing data. The two variables used as measures of size are total assets and sales. Our results indicate that the firms in our sample have, on average, three times the total assets and three times the sales of firms that were excluded. Therefore, the results of this study may not be generalizable to smaller firms.

The results are consistent but not directly comparable with Brown and Warner [1984] since they used up to 1% abnormal performance, no event-date uncertainty, portfolio sizes of 20, and a different proxy (exchange listing) for trading frequency. Our simulations suggest these results are representative of all combinations of event conditions subject to the sample composition limitations mentioned in the preceding paragraph.

To summarize the results of this section, we have: (1) shown that the likelihood of detecting abnormal returns increases with larger portfolios, smaller event-date uncertainty, and increasing abnormal returns; (2) found that the Market Model may offer more powerful tests than the Mean-Adjusted Returns Model and the Market-Adjusted Returns Model in detecting abnormal performance; (3) demonstrated the advantage of using a longer "event" period when that "event" period accurately captures the uncertainty in the event date (i.e., the duration of the event-uncertainty period is known ex ante); (4) documented that the use of a Market Model based on three-day returns does not alter the ability to detect abnormal performance (this model combines the returns over three consecutive days, subject to using the same estimation period as is used for estimating the Market-Model parameters using one-day returns); (5) illustrated that the Scholes-Williams and Dimson methods of estimating risk do not enhance the ability to detect abnormal performance using daily data for firms on the *CRSP* daily tapes with no missing data points (the *OLS* method leads to detectable biases in estimates of beta that are directly correlated with trading frequency); (6) established combinations of event conditions that lead to the likelihood of finding statistically significant results, as summarized in figure 1.

The conclusions above rest on the assumptions made and the procedures used in conducting the simulation tests. These restrictions include the sample selection process which, subject to data availability, presumes randomness. A bias away from less frequently traded firms would tend to reduce the importance of allowing for nonsynchronous trading. Missing data led us to omit a number of *CRSP* firms. Table 8 indicates that the firms remaining in our sample are more likely to be among those on

TABLE 8

Trading Frequency of Compustat at Firms on CRSP Tapes Classified by Inclusion or Exclusion in Our Sample

Total number of firms we used	2069
Total number of firms not used	2062
Total number on *CRSP* tapes	4131
Of the 2069 firms:	
Number on *Compustat*	1416
Of the 2062 firms:	
Number on *Compustat*	626

Trading frequency	Number of firms we used	Number of firms not used
low third	472	297
medium third	472	158
high third	472	171
	1416	626

The data are consistent with a trading frequency bias due to sample selection. The nature of the bias is not clear-cut since, among the unused firms, the relationship to trading frequency does not monotonically decrease.

Compustat. Further, the firms excluded from our sample trade significantly less frequently, on average, than those in our sample. Additionally, a 120-day estimation period divided into two equal parts surrounds a 120-day event period. Further, event-date uncertainty is simulated using a uniform probability distribution, and we have imposed cross-sectionally homogeneous abnormal returns.

6. *Event Clustering*

The results so far have assumed that the securities in the data base are a random selection among those for which adequate data exist. Yet event studies typically do not involve random samples.

Events are often clustered in accounting empirical studies. The clustering may be by industry, by time, or both. Studies of *LIFO* (Sunder [1973]) and depreciation changes (Kaplan and Roll [1972]) and the mandated switch to full-cost accounting in the oil and gas industry (Dyckman and Smith [1979], Collins and Dent [1979], and Lev [1979]) are examples. Our data provide an opportunity to examine the question of the impact of event clustering on the sensitivity of testing procedures.

Brown and Warner [1980] examined the clustering issue in the context of monthly returns. As they point out [1980, p. 232] in discussing time clustering, "the general impact of clustering is to lower the numbers of securities whose month '0' behavior is independent.... If performance measures such as the deviation from historical mean returns or market model residuals are positively correlated across securities in calendar time, then such clustering will increase the variance of the performance measures (e.g., the average residual) and hence lower the power of the tests." Brown and Warner [1980] also investigate the impact of risk clustering, which may proxy for industry clustering. They argue [1980, p. 236]: "it seems reasonable to expect that tests for abnormal performance will be more powerful for low risk securities than for high risk securities; the intuition is simply that a given level of abnormal performance should be easier to detect when 'normal' fluctuations in sample security returns (and the standard errors of parameter estimates such as β) are small rather than large."

To investigate industry clustering, we select four industry groupings, each one of which uses the firms which are in our sample and which were used by Dyckman and Smith [1979] and Collins and Dent [1979] to study the impact of *SFAS No. 19* on the oil and gas industry. The four industry groupings are: Aerospace, Building, Machinery, and Oil and Gas. The results are provided in table 9.

In three of the four industry-clustering cases, we find a substantially lower probability of detecting abnormal returns using the single set of event conditions (3% abnormal performances, three days event-date uncertainty, and a portfolio size of 50) than when clustering is not present. For these three industries, the reduction in the ability to detect abnormal performance is significant at the 0.001 level.

These results are surprising since one would not expect a systematic relationship among residuals for firms in a particular industry, if the residuals are sufficiently diversified over time. It turns out, however, that the average betas for firms in the three previously mentioned industries are substantially higher than the average betas for our population as a whole (approximately 1.4 vs. 1.0). Higher beta firms tend to have more volatile returns and if the Market Model explains, on average, the same percentage of variation for these firms as for the population, then the higher beta firms (industries) will have higher variance residuals.

TABLE 9

Percentage of Portfolios Indicating the Presence of a 2% Level of Abnormal Performance Industry and Time Clustering

Item	Portfolio Size[a]	Percentage of Portfolios: Clustering	Percentage of Portfolios: No Clustering[b]
Industry Clustering Only			
Aerospace**	20	34	45
Building**	30	44	58
Machinery	20	48	45
Oil & Gas**	40	52	62
Time Clustering Only*	40	56	62

[a] Maximum portfolio size given sample.
[b] From table 3.
Event Conditions:
Event-Date Uncertainty = 3 days
Abnormal Performance = 2%
Market Model.
* Significant at the .01 probability level (normal approximation to the binomial).
** Significant at the .001 probability level.

Since we induce a constant level of abnormal performance across firms, the relative impact on firms with high variance residuals will be lower. This is consistent, in turn, with the fact that the probability of detecting abnormal performance for the machinery industry sample was not lower. This sample has the lowest average beta of the four.[13] The result is consistent with Brown and Warner's observation, quoted earlier, suggesting that abnormal performance is more easily detected in low-risk portfolios. The results suggest that the extent of the impact (and possibly even the direction) of industry clustering is a function of the riskiness of the portfolio. However, we have examined only four industry samples and can offer no definitive conclusions on this issue.

The results reported in table 9 are consistent with those reported by Brown and Warner [1984]. Brown and Warner argue that variance is a better risk measure than beta. They use exchange listing (NYSE vs. AMEX) as a proxy for variance since the residual standard deviations for NYSE stocks are lower, averaging 60% of those of AMEX stocks.[14] When 1% abnormal performance is added, Brown and Warner are able to reject the null hypothesis of no abnormal performance more frequently for NYSE (low risk) than for AMEX (high risk) securities.

We are also able to test for the effect of time clustering (but not combined with industry clustering) with our data. The clustering analysis is based on *CRSP* days 3314 to 3354, which yield a sample of 117 firms. The result given by the entry on the last line in table 6 is based on these 117 firms. Time clustering reduces the power of the test, as is illustrated by the lower percentage of times the test detects abnormal performance. The results are consistent across event conditions, but this across-event consistency offers little additional evidence due to a lack of independence in the data.[15]

Our time-clustering results are consistent with those anticipated by Brown and Warner but apparently not with their findings for either monthly or daily data. While finding lower rejection rates using mean adjusted returns and monthly data, the authors [1980, p. 235] conclude, "when abnormal performance is present, the rejection rates when there

[13] The average betas for the firms by industry are:

Aerospace 1.43
Building 1.35
Machinery 1.09
Oil & Gas 1.44

[14] Brown and Warner [1984] dichotomize their sample with NYSE and AMEX stocks. Rejection frequencies for all three methods of beta estimation are higher for NYSE stocks than for AMEX stocks due to the lower variance of NYSE stock returns.

[15] While we have not explicitly examined the issue, Brown and Warner [1984, p. 20] observe that: "For hypothesis tests over intervals of more than a day, the failure to take into account autocorrelation in estimating the variance of the cumulative mean excess return [potentially introduced by nonsynchronous trading] could result in misspecification [in the test results]." Brown and Warner go on to say: "However, . . . it would appear that autocorrelation plays a minor role."

is [time] clustering are not markedly different from those when there is no clustering." When using daily data, Brown and Warner [1984, p. 16] similarly conclude, "the results on [test] specification are not radically altered in experiments where there is clustering in event dates and hence nonindependence of the excess return measures."

Event studies tend to select securities from a limited number of industries, often a single one. When, in addition, there is time clustering, the power of the statistical tests may be further eroded as both problems reinforce one another. Our results suggest that the riskiness of the securities may also influence the test results. We are, unfortunately, unable to investigate this issue empirically with our data. Nevertheless, the concerns expressed by Brown and Warner [1980, p. 235] in particular may be more important than their tests suggest.[16]

Given that we have examined only four industries and a single set of event conditions, caution should be exercised in extrapolating these results to other situations. Methods which begin to address such difficulties (such as Schipper and Thompson [1983] and Collins, Rozeff, and Salatka [1982]) are in order.

A limited set of event conditions was applied to yet another nonrandom sample of firms. This sample consists of firms which repurchased common stock via cash tender offers.[17] The calendar event dates are spread from 1962 to 1976. This sample does not represent a case of time or industry clustering. However, the sample is nonrandom in the sense that all firms initiated a common stock repurchase. In his 1981 study, Dann observed that these firms display an average common-stock, two-day return of approximately 15% at the time of the repurchase announcement.

Due to the small size of the sample, only a portfolio size of ten is used in our event–condition combinations. All five levels of abnormal performance are induced and all five levels of event-date uncertainty are examined. As would be expected, around the event dates themselves, the presence of abnormal performance is detected almost 100% of the time. (The lowest percentage we obtained is 98.) Given the 15% return already present on these days, the result is to be expected.

In addition, we align the sample in event time on days other than the specific event date. For example, portfolios were formed so that the return for each firm in the portfolio is representative of the day ten days before the event date. The test procedure is performed on ten of the 118 days in the test period. Since the performance on each of the ten days appears to be similar, the results across all ten days are averaged.

---

[16] Thus the corrections for cross-sectional dependence in security-specific returns to estimate the variance used in the "t" test (see Beaver [1981] and Collins and Dent [1984]) may be important when clustering is present. The issue is an empirical one.

[17] This sample was provided by Professor Larry Dann. Details on the composition of this sample may be found in his 1981 article in the *Journal of Financial Economics*.

Therefore, the percentage in each cell in table 10 is based on 2,500 (250 repetitions per day × 10 days) sample points. The results, when compared to data such as reported in table 4 but for portfolios of size ten, indicate that induced abnormal performance is detected less frequently for the nonrandom sample of stock repurchase firms than for the sample of 2,069 randomly selected firms.

Characteristics such as stock repurchases may unexpectedly pervade what otherwise appears to the researcher to be a random sample. In such cases the indicated power of the statistical tests developed for random samples is not likely to be appropriate.

7. *The Normality of Excess Returns*

Fama [1976] provides evidence that the distribution of daily returns is leptokurtic. Brown and Warner [1984] give data that indicate this is also the case for excess returns based on daily data. However, this fact need not necessarily bias hypothesis tests toward a Type I error since such tests are based on mean excess returns.

This section provides evidence that the t-test is an accurate test for the presence of abnormal performance despite the nonnormality of the distribution of daily residuals.

Table 11 illustrates the departure from normality of daily return residuals for individual securities. Panel 1 represents the fraction of observations expected for a normally distributed variable. Panel 2 represents the actual fraction (based on all 20,690 residuals) of Market Model standardized residuals falling in the corresponding range.

The distribution of actual residuals is leptokurtic with a negative median. A chi-square goodness-of-fit test on the actual residuals indicates it is unlikely that the residuals are drawn from a normal distribution (the chi-square statistic is 20.16; the critical 5% value is 16.92). This

TABLE 10

Percentage of Portfolios Indicating the Presence of Abnormal Performance:[a]
Market Model
Portfolio Size = 10 Firms
Method 1

Induced Abnormal Performance (Percent)		Event-Date Uncertainty (Days): 1	2	3	4	5
	5	100	77	51	38	27
	4	99	62	37	25	18
	3	96	43	24	16	13
	2	69	21	13	9	8
	1	20	8	7	5	5

[a] Based on a nonrandom sample of firms. The total sample from Dann [1981] consists of 108 firms which experienced a stock repurchase.

TABLE 11

Comparison of Normality of Sample Residuals for Ten Event Dates and Each of 2,069 Firms (20,690 Observations) with the Normal Distribution

Range of Standardized Normal Deviate	Panel 1 Normal	Panel 2 Actual	Difference
−00 to −3.5	0%	.2%	.2%
−3.499 to −3.0	.1%	.2%	.1%
−2.999 to −2.5	.5%	.5%	0%
−2.499 to −2.0	1.7%	1.4%	−.3%
−1.999 to −1.5	4.4%	3.1%	−1.3%
−1.499 to −1.0	9.1%	8.0%	−1.1%
−.999 to −.5	14.9%	15.7%	.8%
−.499 to 0	19.3%	23.6%	4.3%
0 to .499	19.3%	20.7%	1.4%
.5 to .999	14.9%	13.2%	−1.7%
1.0 to 1.499	9.1%	6.9%	−2.2%
1.5 to 1.999	4.4%	3.4%	−1.0%
2.0 to 2.499	1.7%	1.6%	−.1%
2.5 to 2.999	.5%	.8%	.3%
3.0 to 3.499	.1%	.3%	.2%
3.5 to +.00	0%	.4%	.4%

TABLE 12

Percentage of Portfolios with Standardized Residuals > 5% Critical Value:[a]
Normally Distributed Residuals
Abnormal Performance = 3%

Event-Date Uncertainty (Days)	Portfolio Size (No. of Firms): 10	20	30	40	50	75	100
5	22	39	49	57	64	76	88
4	32	44	59	67	78	91	98
3	44	64	83	88	91	98	100
2	66	93	97	100	100	100	100
1	99	100	100	100	100	100	100

[a] 250 portfolios per cell.

result tends to increase the likelihood of rejecting the null hypothesis when it is true, although the effect is mitigated as the portfolio size used in the test increases.

Our results for excess returns support those of Brown and Warner [1984]. However, we also consider whether the detection of abnormal performance would change appreciably if the initial residuals were normally distributed before inducing an abnormal performance level. We adopt the following method for examining this issue.

First consider the standardized residuals for all firms in the sample (ten residuals for each of 2,069 firms). The standard deviation for each of these 20,690 residuals is equal to one by definition. We rank these residuals from smallest to largest and then replace them with their respective expected order statistic. For example, if 20,690 observations are drawn from a standard normal distribution, the expected value of the smallest observation is x_1, the expected value of the second smallest is x_2, etc. If this expected order statistic is multiplied by the firm's standard deviation of residuals, the result is a firm residual that can be manipulated (in the sense of adding abnormal performance under the various event conditions) in the same manner as has been done previously. We elected to use expected order statistics rather than sampling from a standard normal distribution to avoid departures from normality due to sampling error. Also, by retaining the identity of each firm through the ranking process, we avoided the need to assign each firm standard deviation randomly to ten residuals. Partial results are shown in table 12. The numbers in the cells represent the percentage of time that portfolio residuals indicated the presence of abnormal performance. Comparing table 12 to table 2 (portfolios are formed for the same firms and residuals) indicates that the results are essentially identical.

Of the 175 event conditions when the null is false, we find that using normally distributed residuals gives slightly higher rejection frequencies 86 times, lower rejection frequencies in 16 instances, and identical rejection frequencies 73 times. Furthermore, the magnitude of the bias is quite small and declines as the portfolio size increases. Of the 86 cases where the rejection frequency is higher, the average bias is 1.4 percentage points. Of the 16 cases where the rejection frequency is lower, the average bias is 0.7 percentage points. Apparently the use of portfolios compensates quite well for any nonnormality present in the return residuals.

8. *Conclusions*

This study examines some important issues in using daily-returns data to perform event studies. The effects of portfolio size, event-date uncertainty, and magnitude of abnormal performance are quantified over a range of each variable (using the Brown and Warner [1980] method of introducing event-date uncertainty). Portfolio simulations are used to investigate the interactive effects of these variables. The results are subject to the following limitations, described in more detail in the paper: (1) randomness of the sample, subject to a large-firm bias resulting from our sample selection procedures; (2) a single 120-day estimation period; (3) simulation of event-date uncertainty using a uniform distribution over a set period; and (4) a constant imposed abnormal return percentage across firms for a given simulation.

The results can be summarized as follows.

(1) The abilities of the three models (Mean-Adjusted Returns Model, Market-Adjusted Returns Model, and Market Model) to detect correctly the presence of abnormal performance are similar, although we find a

slight preference for the Market Model. While this difference is statistically significant, it does not appear important.

(2) When the day of the event is uncertain, the researcher, in selecting among the three methods for considering event-date uncertainty, is better off accumulating the residuals using either a one-day or multi-day Market Model (Method 2 or 3), under the assumptions of uniform uncertainty for each day in the known event period and freedom from serial correlation. Furthermore, the uncertain event-date period must be free of confounding events to avoid either "event smearing" (Brown and Warner [1980]) or event reinforcement. In the latter case, an effect may be found, but the cause remains ambiguous.

(3) The use of Scholes-Williams and Dimson methods of estimating risk to mitigate the effects of the nonsynchronous trading problem did not improve the ability to detect correctly the presence of abnormal performance for the *CRSP* firms used in this study which had no missing data points. However, a trading bias exists in our sample toward the exclusion of less frequently traded firms for which the nonsynchronous trading issue is more important.

(4) Event-date clustering by industry or time appears generally to reduce the ability of traditional methods to detect abnormal performance. However, the effect may be dependent on the portfolio's risk.

(5) The nonnormality of individual-security daily-return residuals has little effect on the inferences drawn from the use of the t-test applied to portfolios.

We anticipate that the issues addressed in this study will receive further attention by others. First, we encourage researchers to avoid the large-firm bias in studying the impact of nonsynchronous trading on detecting abnormal performance. Second, the choice of a data base that would permit an in-depth investigation of the joint clustering by time and industry would be beneficial. A third area of study could be devoted to examining actual distributions of abnormal return levels across firms for specified events. If commonalities were found, say as to average levels and variances for events with related characteristics, the results could be used in establishing whether available techniques would be likely to reveal an information effect against background noise. Similar work might be attempted with the distribution of event-date uncertainty. However, we do not believe that the necessary data are likely to be available or, even if they are, that the commonalities will be easily established. Simulations of other randomly selected distributions do not seem to us to be particularly useful. Finally, where there is an established theory or, perhaps, some suggestive empirical results, judgment samples may provide additional insights into the added test power for special cases where attention can be focused based on the prior work.

REFERENCES

BEAVER, W. "The Information Content of Annual Earnings Announcements." *Journal of Accounting Research* (Supplement 1968): 67–92.

———. "Econometric Properties of Alternative Security Return Methods." *Journal of Accounting Research* (Spring 1981): 163–84.

BROWN, S., AND J. WARNER. "Measuring Security Price Performance." *Journal of Financial Economics* (June 1980): 205–58.

———. "Using Daily Stock Returns: The Case of Event Studies." Working paper, University of Rochester, March 1984.

COLLINS, D., AND W. DENT. "The Proposed Elimination of Full Cost Accounting in the Extractive Petroleum Industry: An Empirical Assessment of Market Consequences." *Journal of Accounting and Economics* (March 1979): 3–44.

———. "A Comparison of Alternative Testing Methodologies Used in Capital Market Research." *Journal of Accounting Research* (Spring 1984): 48–84.

COLLINS, D., M. ROZEFF, AND W. SLATKA. "The SEC's Rejection of SFAS No. 19: Tests of Market Price Reversal." *The Accounting Review* (January 1982): 1–17.

DANN, L. "Common Stock Repurchase: An Analysis of Returns to Bondholders and Stockholders." *Journal of Financial Economics* (June 1981): 113–38.

DEANGELO, H., AND E. RICE. "Antitakeover Charter Amendments and Stockholder Wealth." *Journal of Financial Economics* (April 1983): 329–60.

DIMSON, E. "Risk Measurement When Shares Are Subject to Infrequent Trading." *Journal of Financial Economics* (June 1979): 197–226.

DODD, P., ET AL. "Qualified Audit Opinions and Stock Prices: Information Content, Announcement Dates and Concurrent Disclosures." *Journal of Accounting and Economics* (April 1984): 3–38.

DYCKMAN, T., AND A. SMITH. "Finance Accounting and Reporting by Oil and Gas Producing Companies: A Study of Information Effects." *Journal of Accounting and Economics* (March 1979): 45–74.

FAMA, E. *Foundation of Finance.* New York: Basic Books, 1976.

KAPLAN, R., AND R. ROLL. "Investor Evaluation of Accounting Information: Some Empirical Evidence." *Journal of Business* (April 1972): 225–57.

LEV, B. "The Impact of Accounting Regulation on the Stock Market: The Case of Oil and Gas Companies." *The Accounting Review* (July 1979): 485–503.

MASULIS, R. "The Effects of Capital Structure Change on Security Prices: A Study of Exchange Offers." *Journal of Financial Economics* (June 1980): 139–78.

MORSE, D. "An Econometric Approach to the Choice of Using Daily versus Monthly Returns in Testing for Information Effects on Returns." *Journal of Accounting Research* (Autumn 1984): 605–23.

PATELL, J. "Corporate Forecasts of Earnings per Share and Stock Price Behavior: Empirical Tests." *Journal of Accounting Research* (Autumn 1976): 246–76.

———, AND M. WOLFSON. "Anticipated Information Releases Reflected in Call Option Prices." *Journal of Accounting and Economics* (August 1979): 117–40.

———, AND M. WOLFSON. "The Ex Ante and Ex Post Price Effects of Quarterly Earnings Announcements Reflected in Options and Stock Prices." *Journal of Accounting Research* (Autumn 1981): 434–58.

———, AND M. WOLFSON. "Good News, Bad News, and the Intraday Timing of Corporate Disclosures." *The Accounting Review* (July 1982): 509–27.

SCHIPPER, K., AND R. THOMPSON. "The Impact of Merger-Related Regulations on the Shareholders of Acquiring Firms." *Journal of Accounting Research* (Spring 1983): 184–221.

SCHOLES, M., AND J. WILLIAMS. "Estimating Betas from Nonsynchronous Data." *Journal of Financial Economics* (December 1977): 309–27.

SHANE, P., AND B. SPICER. "Market Response to Environmental Information Produced Outside the Firm." *The Accounting Review* (July 1983): 521–38.

SUNDER, S. "Relation Between Accounting Changes and Stock Prices: Problems of Measurement and Some Empirical Evidence." *Journal of Accounting Research* (Supplement 1973): 1–45.