

# Another Look At Theoretical And Empirical Issues In Event Study Methodology

Thomas P. McWilliams, (tmcwilliams@drexel.edu), Drexel University  
Victoria B. McWilliams, (victoria.mcwilliams@villanova.edu), Villanova University

## Abstract

*The goal of this manuscript is to help to improve the integrity of research that uses event study methodology. We discuss issues related to correctly performing event studies and, in some cases, provide alternatives to a variety of recommendations made by McWilliams and Siegel (1997) regarding the application of event study methodology. While McWilliams and Siegel provide a good starting point for providing guidance in the use of event study methodology, our revised recommendations add additional value beyond McWilliams and Siegel by being more consistent with statistical theory, existing research results, and accepted practice. These recommendations, along with those found in McWilliams and Siegel, should lead to higher quality research, regardless of the discipline to which event study methodology is applied.*

## Introduction

Event study methodology is used frequently to identify the stock price reaction of a specific event and, based on the reaction, allows the researcher to conclude whether the event was detrimental or beneficial to the firm's shareholders. As the result of Brown and Warner's (1980, 1985) examination of event studies and their underlying assumptions, empirical work in finance and accounting has made extensive use of this methodology. As observed by McWilliams and Siegel (1997, p. 626), the field of management has also used event study methodology in a variety of ways to identify organizational and public policy implications of both endogenous and exogenous corporate events. As a result of the increased popularity of event study methodology in the management literature, McWilliams and Siegel (M&S)

studied theoretical and empirical issues surrounding the methodology and, based on their findings, made recommendations with the intent of improving the integrity of management and other studies that rely on the methodology. While M&S focus on research in management, the issues raised are generally not management specific and their and our recommendations apply to other disciplines as well.

After carefully examining the M&S analysis and recommendations, we have concluded that several guidelines provided by M&S which are related to statistical methodology and interpretation of results require additional attention. It is our intention to further the M&S objective of providing guidance that will lead to defensible and reliable inferences based on event study methodology by providing a more complete view of some of the issues raised by M&S. In particular we will address the definition of the event

---

*Readers with comments or questions are encouraged to contact the authors via e-mail.*

study test statistic, sample size issues and the assumption of normality, the use of a bootstrap testing procedure, and nonparametric test procedures. In the interest of brevity, we will not touch on issues where we agree with the recommendations of M&S, such as the need to select defensible event windows or the need to control for confounding events that may make interpretation of results questionable. The information and guidance provided here along with information found in M&S will allow researchers to use event study methodology with confidence that their empirical results provide meaningful insight into the effect of specific organizational events.

**Recommended Test Statistics**

M&S (1997, pp. 628-630) discuss the underlying assumptions and research design issues relevant to event studies. They present a commonly used test statistic similar to those recommended by Dodd and Warner (1983) or Patell (1976) for testing the null hypothesis that the event has a zero stock price effect. The development of this statistic is outlined below, and we also present an alternative statistic that we prefer to use, especially when working with longer event windows.

As a starting point we assume that for each firm in the study a market model estimation period and an event window have been identified. Guidance regarding this process can be found in M&S, and, for ease of comparison, we use notation similar to theirs to define the event study model:

- $R_{it}$  = the rate of return on the share price of firm  $i$  on day  $t$
- $R_{mt}$  = the rate of return on the market portfolio on day  $t$
- $AR_{it}$  = the abnormal return, based on the market model discussed in M&S, for firm  $i$  on day  $t$ :  $AR_{it} = R_{it} - (a_i + b_i R_{mt})$  where  $a_i$  and  $b_i$  are the ordinary least squares (OLS) estimators, calculated over the estimation period, for the market model for firm  $i$

- $S_i$  = the standard error of the OLS market model regression for firm  $i$
- $T$  = the number of days in the estimation period
- $k$  = the number of days in the event window

To simplify and clarify formulas we use  $\sum_{t \in EP}$  to signify a summation taken over values of the time index  $t$  that are in the estimation period, and  $\sum_{t \in EW}$  to signify a summation taken over values of  $t$  that are in the event window.

Begin by standardizing each daily abnormal return according to

$$SAR_{it} = AR_{it} / SD_{it} \tag{1}$$

where  $SD_{it}$  is the standard deviation of  $AR_{it}$ :

$$SD_{it} = S_i \sqrt{1 + \frac{1}{T} + \frac{(R_{mt} - \bar{R}_m)^2}{\sum_{j \in EP} (R_{mj} - \bar{R}_m)^2}} \tag{2}$$

where  $\bar{R}_m$  represents the mean return, over the firm's estimation period, of the market portfolio. Under the usual assumptions of simple regression analysis, it can be shown that each  $SAR_{it}$  value follows a  $t$  distribution with  $T-2$  degrees of freedom. By accumulating and standardizing over time and across firms according to:

$$CAR_i = \frac{1}{\sqrt{k}} \sum_{t=1}^k SAR_{it} \tag{3}$$

$$ACAR = \frac{1}{n} \sqrt{\frac{T-4}{T-2}} \sum_{i=1}^n CAR_i \tag{4}$$

$$Z = ACAR \times \sqrt{n} \tag{5}$$

we arrive at our test statistic  $Z$ . If we assume that the  $CAR_i$  are independent, which is the typical assumption in studies involving firm specific rather than common event dates, then  $Z$  has, ap-

proximately, a standard normal distribution under the null hypothesis of no stock price effect.<sup>1</sup>

While this approach is heavily used in the literature, we believe that an approach used, for example, by Meznar, Nigh, and Kwok (1994) and by Mikkelsen and Partch (1988) is more appropriate. For each firm calculate the cumulative abnormal return over the event window:

$$CAR_i = \sum_{t \in EW} AR_{it} \quad (6)$$

which has standard deviation

$$SD_i = S_i \sqrt{k + \frac{k^2}{T} + \frac{\sum_{t \in EW} R_{mt} - k(\bar{R}_m)^2}{\sum_{t \in EP} (R_{mt} - \bar{R}_m)^2}} \quad (7)$$

Each firm's cumulative abnormal return is then standardized according to

$$SCAR_i = CAR_i / SD_i \quad (8)$$

Finally, a Z-statistic is calculated according to:

$$Z = \frac{1}{\sqrt{n}} \sum_{i=1}^n SCAR_i \quad (9)$$

Under the null hypothesis this statistic will also have, approximately, a standard normal distribution. We prefer this approach because the derivation of  $SD_i$  accounts for the serial correlation of daily event-period abnormal returns for the same firm, while the approach presented by M&S ignores this correlation, treating the abnormal returns as being independent.<sup>2</sup> This "correction" for serial correlation is in most cases unlikely to have a major impact on the value of the test statistic, but a study by Coutts, Mills, and Roberts (1995) concluded that it can make a substantive difference for longer event windows. In addition, it is no more difficult to calculate than the M&S recommended approach.

### The Assumption of Normality

In the discussion on sample size, M&S cite several studies that use relatively small sample portfolios ( $N = 2, 18, 20,$  and  $22$ ). They state "Clearly in many instances, imposing normality assumptions would indeed be quite heroic" (1997, p. 635) and suggest that a bootstrap technique, which we will discuss later, be used as a supplemental approach to testing the research hypothesis.

We agree that, at  $N = 2$ , the assumption of normality is unwarranted: we would view an event study based on two firms as anecdotal in nature, regardless of the statistical methodology used. However, with respect to the other sample sizes cited, M&S's statement regarding the assumption of a normally distributed test statistic is presented without justification or reference and is in direct conflict with the results of several published studies. Brown and Warner (1985); Dyckman, Philbrick, and Stephan (1985); and Corrado and Zivney (1992) have all conducted independent large scale empirical studies regarding the behavior of commonly used event study test statistics based on daily stock returns data from the New York and American stock exchanges. Sample sizes used were as small as  $N = 5$  for the Brown and Warner study and  $N = 10$  for the other two studies. The studies are consistent in their conclusions, which were based on goodness of fit tests and empirical rejection rates: while firm-specific daily abnormal returns tend to follow skewed distributions, the normal distribution was a good fit for the aggregate Z-statistics used for hypothesis testing. Therefore, we disagree with the statement regarding normality made by M&S and are comfortable recommending the use of the normal distribution for smaller sample sizes.

### The "Bootstrap" Test

Based on the perception of a problem with normality when dealing with small portfolios, M&S (1997, pp. 634-635) recommend the use of a bootstrap testing procedure, citing Barclay &

Litzenberger (1988) as a precedent. This recommendation surprises us, as we are not aware of any published event study based on daily returns data (the predominant application) which has used a bootstrap approach. Barclay and Litzenberger work with *intraday* returns data, such as stock price movements measured over 15 minute intervals, and state (1988, p. 79) that "Little is known about the distribution of intraday stock returns..." as justification for using the bootstrap. Therefore, Barclay and Litzenberger's (B&L) reasoning does not provide a motivation for the use of the bootstrap with daily returns data whose distribution, as noted in the previous section, has been studied extensively.

We have a number of additional comments on B&L and on the bootstrap presented by M&S. Because the bootstrap is a fairly complex statistical technique that has seen little use in the event study literature, we begin with some background information that may help the reader to follow our arguments.

The bootstrap is a "resampling" procedure (Efron and Tibshirani, 1986) where a sample (say  $X_1, X_2, \dots, X_n$ ) is obtained and a statistic (say  $U(X_1, X_2, \dots, X_n)$ ) is calculated as a function of the sample. To make an inference regarding the underlying population, knowledge of the sampling distribution of  $U$  is required, and it is assumed that initially no information is available. This problem is addressed by obtaining empirical information about the distribution through a resampling process: a large number ( $m$ ) of samples of size  $n$  are taken with replacement from the original sample, and a new value for  $U$  is calculated for each of these samples. The values of  $U$  obtained by resampling constitute an empirical sampling distribution that can be used to make inferences.

A major focus of research regarding the bootstrap (Hinkley, 1988) has been to find methods for generating confidence intervals for an unknown parameter, and this application is straightforward. However, as noted by Hinkley and by Hall & Wilson (1991), the use of boot-

strap methods to perform hypothesis tests is a more complex problem. The basic difficulty is that, to obtain the p-value required by a hypothesis test, we need to calculate a probability under the null ( $H_0$ ) distribution of the test statistic, which requires knowledge of the sampling distribution of the test statistic when  $H_0$  is true. This information cannot necessarily be obtained through a simple resampling procedure, as suggested by M&S, since we do not know whether the sample came from an  $H_0$  or  $H_1$  distribution, as that is of course the research question under investigation.

Articles by Hinkley and by Hall and Wilson (H&W) both contain suggestions for dealing with this problem, such as the use of bootstrap pivoting, which produces a statistic whose asymptotic distribution does not depend on any unknown parameters. The reader is referred to H&W for details.

Now consider the "bootstrap" tests presented by B&L and by M&S. Of the two test statistics given by M&S, the following comments specifically relate to the statistic based on abnormal return ( $AR$ ) values. Their proposed approach based on the  $PRNEG$  statistic has similar problems. Our concerns regarding the B&L and M&S approaches are as follows:

- The bootstrap methodology presented by B&L does not address the  $H_0$  sampling distribution problem discussed above and does not give a meaningful p-value. Therefore, it does not provide a viable precedent for use in event studies.
- To test the null hypothesis of a zero stock price effect, we require a test statistic that aggregates abnormal returns across all firms in the sample portfolio and across all days in the event period. The  $AR$  statistic used by M&S is a *daily* average abnormal return, where the average is taken over all firms in the sample. No guidance is given regarding how to aggregate over the event window when constructing the empirical sampling

distribution or the test statistic, so M&S have not defined the test statistic to be bootstrapped.

- The M&S test procedure is actually *not* a bootstrap. M&S recommend generating an empirical distribution based on abnormal returns calculated over the market model estimation period (i.e. the residuals from the OLS regression). Then, a daily average abnormal return is calculated for the event window (again, the issue of aggregating over the window is not addressed) and compared to this empirical distribution to obtain a p-value. The resampling process is not based on the abnormal return values that are used to calculate the test statistic, so this process should not be referred to as a bootstrap.
- If we assume that, under the null hypothesis of a zero stock price effect, the distribution of abnormal returns during the event period will be the same as the distribution of abnormal returns (regression residuals) during the estimation period, then the M&S approach is a viable way to empirically determine the  $H_0$  distribution of the test statistic. However, there are two problems with this assumption. The first is that, even when a regression model is well behaved, OLS residuals will have less variance than future prediction errors,<sup>3</sup> so the M&S approach underestimates variability in the test statistic's sampling distribution, which leads to overstating statistical significance. The second is that a possible result of an event is to increase variability in abnormal returns during the event window without shifting the mean (Brown and Warner, 1985). While traditional nonparametric procedures such as the sign test or the Wilcoxon signed rank test will not be affected by this increase, the M&S approach would once again tend to overstate statistical significance due to underestimating variability in the sampling distribution.

- As a final area of concern, consider the determination of the bootstrap sample size. The example bootstrap application presented by M&S (1997, pp. 634-635) uses a 15 firm sample portfolio and a 200-day market model estimation period. M&S recommend using  $m = 15 \times 20 = 3,000$  repetitions to generate the bootstrap distribution of the test statistic. The logic behind this formula is not stated or referenced, and we see no reason why the bootstrap sample size should increase with either the number of firms or the length of the estimation period. In fact, the central limit theorem leads us to expect test statistics to have smoother, more symmetric sampling distributions when these statistics represent aggregations over larger portfolios or longer estimation periods, so we do not understand why empirical approximation of these smoother distributions would require larger samples. Note that we are not aware of any specific formula for a bootstrap sample size, but that, as a precedent, Hinkley (1988) uses  $m = 1,000$  for both confidence interval and hypothesis testing applications. We believe that a value of this magnitude is sufficient.

To summarize, M&S recommend a new procedure that has a number of potential problems or pitfalls, and suggest that it be considered the new standard approach in event study methodology applications. This suggestion is made in the absence of any research results which demonstrate that the M&S approach is superior to, or even as good as, other accepted nonparametric approaches such as those discussed in the following sections.

#### **Nonparametric Tests, Outliers, and Influential Values**

M&S state that "The test statistics employed in event studies tend to be quite sensitive to outliers....," "One important control for outliers is for researchers to report nonparametric test statistics.", and "...it is clear that researchers should adjust the event study technique, or be

especially careful to identify outliers, when dealing with small samples." (1997, p. 635). However, they present no discussion regarding what constitutes an outlier in this application, how to identify outliers, why nonparametric procedures provide an effective control for outliers, or how to interpret results from these procedures. We support the use of nonparametric tests, as discussed below, but do not believe that reporting the values of nonparametric test statistics constitutes any sort of control for outliers.

It is common in event studies to perform a nonparametric test, such as a sign test or a signed rank test, in conjunction with the parametric test based on the  $Z$ -statistic. If the test results are consistent (which is in itself a subjective issue) with respect to statistical significance, the resulting inferences are strengthened. If they are not consistent, a closer examination of the distribution of the data is warranted. Test results might be inconsistent due to extreme observations in the data set which are influential<sup>4</sup> with respect to the  $Z$ -statistic but have a lesser impact on the more robust nonparametric statistics, but this is not a foregone conclusion as there are other possibilities. For example, the  $Z$ -test is generally the more powerful test when its underlying assumptions hold, so it is more likely to detect an actual stock price effect than, say, a sign test. On the other hand, if a stock price effect exists and the effect is to cause an unusually high percentage of abnormal returns to be positive but the magnitude of the returns is small, then the sign or signed rank test would be more likely to detect the effect than the  $Z$ -test. The researcher should also keep in mind that the tests are not actually testing the same hypothesis, as the  $Z$ -statistic is used to test a statement about a population mean while the sign and signed rank tests are used to test a statement about a population median. When dealing with skewed distributions, as is the case for abnormal returns,<sup>5</sup> this could lead to inconsistent results.

If parametric and nonparametric tests yield inconsistent results, the researcher seeking further understanding should begin by simply look-

ing at the distribution of the  $SCAR_i$  values defined in equation (8). This action, combined with an understanding of the underlying assumptions, power properties, and robustness properties of the tests being used will hopefully shed some light on why the tests disagree and which test is more credible.

Suppose that an examination of the distribution of the  $SCAR_i$  indicates that one or a few values are "extreme" relative to the rest of the data (we hesitate to call such values outliers, as discussed below). As these are standardized values, we might for example label an  $SCAR_i$  value greater than 2.0 or 2.5 as being extreme. A logical next step is to determine whether these values are influential, i.e. is there a significant impact on the test statistic if they are deleted, either singly or as a group? If an observation or small group of observations is influential, the researcher needs to deal with justifying the deletion or retention of these observations, and we can give no specific advice on this issue.

In our discussion we have deliberately avoided the use of the term "outlier." M&S use this term without discussing what constitutes an outlier in an event study analysis. In statistical analyses in general, the term outlier is used to refer to a data value that is not well fitted (or well "explained") by the statistical model of interest. For example, in regression analysis an outlier is a data point that is seen, *after fitting a regression equation*, to have an unusually large residual (Neter, Kutner, Nachtsheim, and Wasserman, 1996, pp. 103-104).

In an event study analysis, if an actual stock price effect exists, then we recognize the possibility of heterogeneous stock price reactions within the sample portfolio. This is the motivation for fitting, as recommended by M&S in Step 9 (1997, p. 653) an econometric model, such as the commonly used multiple regression model, to explain this cross-sectional variation in abnormal returns. Such models typically include, as independent variables, firm specific variables such as firm size, firm industry, or the degree of

managerial share ownership. Until this model is fitted we do not feel that it is appropriate to label an observation as an outlier, as an  $SCAR_t$  value which initially appears to be extreme may be well explained by the independent variables used in the regression model. Once the regression analysis has been performed, both influential values (with respect to the regression analysis) and outlying  $SCAR_t$  values can be identified using common measures such as Cook's distance and studentized residuals. Once again, the researcher's judgement comes into play in deciding how to deal with these values.

To summarize, we support the use of non-parametric tests, but believe that these tests plus a more in-depth analysis is required to identify and deal with influential or outlying observations.

#### The Exact and Approximate Sign Test

Regarding a specific nonparametric test, M&S (1997, p. 636) recommend – particularly in the case of a small sample size – the use of sign test. They report the “binomial Z statistic”:

$$Z_p = (PRNEG_t - p^*) / [p^*(1 - p^*) / N]^{1/2} \tag{10}$$

where  $PRNEG_t$  = the proportion of excess returns on day  $t$ ;  $p^*$  = the expected proportion under the null hypothesis; and  $N$  = the number of firms in the sample. Note that in this formula it is not logical to subscript  $PRNEG$  with the time index ( $t$ ), as this suggests that a separate statistic is calculated for each day in the event window. The hypothesis is about the stock price effect over a possibly multi-day event window, and a sign test statistic that covers the entire period is required. The test statistic should be based on the collection of  $N$  firm-by-firm cumulative abnormal returns ( $CAR_t$  values), where the accumulation is across the event window.

M&S state (1997, p. 636 Footnote 4) that “The idea behind this test is that, if the event has

no significant effect on shareholder returns, then abnormal returns will be normally distributed—that is, half the companies will experience positive abnormal returns and the other half, negative abnormal returns.” This statement is inaccurate and misleading for several reasons. First, as discussed earlier, studies regarding the behavior of daily abnormal returns consistently conclude that they have skewed distributions – we are not aware of any research that suggests that they are normally distributed. Second, the usual motivation for *any* nonparametric test is that the researcher is unwilling to assume a specific underlying distribution such as the normal: the benefit or value of a test such as the sign test is that it can be used under these circumstances. Finally, it does not follow that if abnormal returns are normally distributed then “half the companies will experience positive abnormal returns”. This is a statement about the median of the distribution of abnormal returns, which is not linked to normality. The correct logic behind the use of the sign test is that, if the event has no effect on shareholder returns, then the proportion of positive and negative cumulative abnormal returns observed during the event window should be the same as in non-event periods. This proportion may or not equal 1/2 under the null hypothesis, which we will discuss later.

Furthermore, the statistic recommended by M&S is the large sample normal approximation of the traditional sign test statistic. Since M&S are concerned with small sample applications, it would be more appropriate to use the exact statistic which, for the one-tailed test considered by M&S, is simply the number of observed negative cumulative abnormal returns (see Newbold, 1995, pp. 386-389). Under the null hypothesis, this statistic has an exact binomial distribution with parameters  $N$  and  $p$ . As an example of the magnitude of error which might be induced by the use of the approximate rather than the exact statistic, suppose that  $N = 15$  and we wish to test  $H_0: p \geq 0.5$  vs.  $H_1: p < 0.5$ . If we observe five negative  $CAR_t$  values, the exact probability value ( $p$ -value) is 0.1509 while the approximation yields 0.0984 for an error of 35%.<sup>6</sup> In this

case the approximation might lead a researcher to falsely conclude that the result is significant at the 0.10 level. Even larger errors could occur with smaller sample sizes.

Another issue faced by the researcher is the choice of  $p^*$ . It is common in published event studies to use  $p^* = 1/2$  as the null value of  $p$ , so the implicit assumption is that 1/2 of the cumulative abnormal returns will be negative when no price effect is present; i.e. the median cumulative abnormal return is zero. While this is accepted practice, this approach ignores the aforementioned skewness in the distribution of abnormal returns, so 1/2 is not the appropriate null value. This misspecification is documented in Brown and Warner (1980, 1985) and in Berry, Gallinger, and Henderson (1990). Both Corrado and Zivney (1992) and Cowan (1992) present more general sign test statistics which do not use  $p^* = 1/2$  but rather estimate the value of  $p$  from estimation period abnormal returns. The logic behind these tests is not unlike that of the second bootstrap procedure (based on *PRNEG*) presented by M&S. However, we recommend the Corrado and Zivney or Cowan procedures over the M&S bootstrap, as properties such as accuracy of model specification and power have been studied and found acceptable for the former tests while the performance of the bootstrap procedure in event studies, in comparison with these tests, is an open research question.

### **Recommendations and Conclusions**

A researcher planning to use event study methodology should do so in a statistically valid, logically defensible manner. Toward that objective, M&S (1997, pps. 651-653) recommend a step-by-step procedure for event studies. We agree with their first six steps, and because we agree with these recommendations, we simply summarize them as follows.

It is important for the researcher to identify the event to be studied which theoretically provides new information to the market (step 1), as well as to specify the theory justifying the mar-

ket's reaction to this new information (step 2). The researcher must also identify the sample firms experiencing the event and the date of the event occurrence (step 3), and choose a defensible event window (step 4). It is crucial for the researcher to control for confounding events in order to ensure that the stock price reaction that is measured is reflecting a reaction to the event in question (step 5). Step 6 requires that the researcher compute abnormal returns during the event window and test for their statistical significance. We agree that this step is crucial; however, we suggest that the researcher use a test statistic that corrects for serial correlation, especially in the situation where a longer event window is justified.

We are less comfortable with steps 7 through 10 of M&S, as there are many areas of concern. Step 7 requires nonparametric testing of event study results. We support the use of nonparametric tests, but not as specified by M&S. We also have a broader view than M&S regarding how to identify outliers and influential values and how to interpret discrepancies in the results of parametric and nonparametric tests. Additionally, because M&S are concerned with small sample applications (1997, pp. 635-636), it is more appropriate to use the exact, rather than the approximate, sign test statistic.

We disagree with step 8, which indicates that bootstrap methods should be used for small samples. We are not aware of any empirical studies which demonstrate that bootstrap methods perform well in this application, and the approach suggested by M&S, which we do not regard as a true bootstrap, has not been empirically validated and will not perform as well as other nonparametric procedures in situations where the abnormal return variance is inflated during the event period.

We also disagree with step 9 that *requires* the researcher to specify a theory explaining the cross-sectional variation in abnormal returns resulting from the event, and subsequently testing the theory. While explaining the cross-sectional

variation of the stock price reaction to an event is an important issue, it is a separate research question. As a separate research question, it may or may not be immediately relevant to the issue of determining the stock price reaction per se. Note that we do feel that, if outlying  $SCAR_t$  values are a potential problem, a cross-sectional analysis is a prerequisite to the identification of firms having these values.

Finally, we would add a caveat to step 10, the requirement that the researcher disclose sample firm names and event dates. In some cases, it may require tremendous time and effort to identify firms affected by an event and the related event date. In this situation, we feel that it should be up to the researcher to decide whether it is appropriate to share the sample details. We do, however, believe that the researcher should *always* be required to specify enough detail about *how and from what sources* data are identified so that any other interested party could, with enough effort, replicate the reported study. Failure to provide this level of detail is indefensible.

We encourage any researcher performing event study methodology to pay careful attention to information presented in this manuscript along with McWilliams and Siegel (1997). The combined recommendations of M&S and those presented here will lead to a high quality empirical analysis which will allow the researcher to make justifiable inferences about the effect of events on shareholder wealth as reflected in the firm's stock price. The integrity of these inferences will allow researchers to confidently identify organizational and public policy implications of corporate events.

### Suggestions for Future Research

While we were not comfortable with the bootstrap procedure presented by M&S, we feel that the general concept of basing the hypothesis test on an empirical sampling distribution may have validity and warrants further study. A second area of interest is the development of meas-

ures designed to identify firms that are influential in the sense that their inclusion in the sample has a significant impact on the study's test statistics. Additional research in these two areas will further advance researchers' ability to effectively use event study methodology. □□

### Endnotes

1. The analog to formula (2) presented by M&S is missing a plus sign, and their formula is confusing as the variable  $t$  is used both as a subscript to denote the day, within the event window, for which the standard deviation is being calculated and as a summation index for a sum taken over the estimation period. The M&S analog to formula (4) contains improper subscripts, as  $ACAR$  is subscripted with a time index when in fact it does not depend on time.
2. In the market model it is assumed that day-to-day stock price returns are uncorrelated, so our statement regarding serial correlation of abnormal returns may seem incorrect. We omit a mathematical proof (which we will provide on request) that this correlation exists in favor of the following intuitive argument. Daily abnormal returns for a firm are calculated from a common sample regression equation, so each abnormal return is subject to measurement error due to estimation error in the regression parameters. This common source of error leads to the serial correlation.
3. A proof of this statement can be provided on request. The intuitive argument is that, through the OLS fitting process, the slope and intercept values of the sample regression equation are calculated to optimally fit a specific set of observed data and we cannot expect

future data to fit the sample regression line as closely as this original data set.

4. We use the term "influential" as it is commonly used in regression analysis. An observation is influential if the value of some statistic of interest (for example the sample slope in regression analysis) changes significantly when that observation is removed from the data set. Cook's distance (Neter, Kutner, Nachtsheim, and Wasserman, 1996: 380-382) is an example of a measure of influence used in regression analysis. We are not aware of any published measures of influence that can be used in event study analysis, so the identification of influential observations will involve an element of subjectivity.
5. As discussed in the section titled "The Assumption of Normality," the results of a number of empirical studies suggest that firm-specific abnormal returns will have skewed distributions.
6. In this example, the exact  $p$ -value is the probability of obtaining five or fewer observations under a binomial distribution having parameters  $n = 15$  and  $p = 0.5$ . This cumulative probability is easily found, for example, using a binomial distribution spreadsheet function. The approximate  $p$ -value is obtained through a similar calculation that uses equation (10) and the cumulative standard normal distribution. Modern spreadsheet programs can readily calculate exact  $p$ -values for any sample size likely to be used in practice, including sample sizes in the hundreds, so we see no reason to use the less accurate approximation for small or large samples.

## References

1. Barclay, M. & Litzenberger, R., "Announcement effects of new equity issues and the use of intraday price data," *Journal of Financial Economics*, Vol. 21, No. 1, pp. 71-99, 1988.
2. Berry, M.A., Gallinger, G.W., & Henderson, G.V. Jr., "Using daily stock returns in event studies and the choice of parametric versus nonparametric test statistics", *Quarterly Journal of Business and Economics*, Vol. 29, No. 1, pp. 70-85, 1990.
3. Brown, S. & Warner, J., "Measuring security price performance," *Journal of Financial Economics*, Vol. 8, No. 3, pp. 205-258, 1980.
4. Brown, S.J. & Warner, J.B., "Using daily stock returns: The case of event studies," *Journal of Financial Economics*, Vol. 14, No. 1, pp. 3-31, 1985.
5. Corrado, C.J. & Zivney, T.L., "The specification and power of the sign test in event study hypothesis tests using daily stock returns," *Journal of Financial and Quantitative Analysis*, Vol. 27, No. 3, pp. 465-478, 1992.
6. Coutts, J.A., Mills, T.C., & Roberts, J., "Testing cumulative prediction errors in event study methodology," *Journal of Forecasting*, Vol. 14, No. 2, pp. 107-115, 1995.
7. Cowan, A.R., "Nonparametric event study tests," *Review of Quantitative Finance and Accounting*, Vol. 2, No. 2, pp. 343-358, 1992.
8. Dodd, P. & Warner, J.B., "On corporate governance: A study of proxy contests," *Journal of Financial Economics*, Vol. 11, Nos. 1-4, pp. 401-438, 1983.
9. Dyckman, T., Philbrick, D., & Stephan, J., "A comparison of event study methodologies using daily stock returns: A simulation approach," *Journal of Accounting Research*, Vol. 22 supplement, pp. 1-30, 1985.
10. Efron, B. and Tibshirani, R., "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy," *Statistical Science*, Vol. 1, No. 1,

- pp. 54-75, 1986.
11. Hall, P. and Wilson, S.R., "Two guidelines for bootstrap hypothesis testing," *Biometrics*, Vol. 47, pp. 757-762, 1991.
  12. Hinkley, D.V., "Bootstrap methods," *Journal of the Royal Statistical Society Series B*, Vol. 50, No. 3, pp. 321-337, 1988.
  13. McWilliams, A. & Siegel, D., "Event studies in management research: Theoretical and empirical issues," *Academy of Management Journal*, Vol. 40, No. 3, pp. 626-657, 1997.
  14. Meznar, M., Nigh, D., & Kwok, C., "Effect of announcements of withdrawal from South Africa on stockholder wealth," *Academy of Management Journal*, Vol. 37, No. 6, pp. 1663-1648, 1994.
  15. Mikkelson, W.H. & Partch, M.M., "Withdrawn Security Offerings," *Journal of Financial and Quantitative Analysis*, Vol. 23, No. 2, pp. 119-133, 1988.
  16. Neter, J., Kutner, M.H., Nachtsheim, C.J., and Wasserman, W., *Applied Linear Statistical Models*, 4th Edition, Richard D. Irwin, Inc., Chicago, Illinois, 1996.
  17. Newbold, P., *Statistics for Business & Economics*, Prentice Hall, New Jersey, 1995.
  18. Patell, J.M., "Corporate forecasts of earnings per share and stock price behavior: empirical tests," *Journal of Accounting research*, Vol. 14, No. 2, pp. 246-276, 1976.

Notes

Notes