Alternative Methods for Robust Analysis in Event Study Applications†

Advances in Investment Analysis and Portfolio Management,
C.F. Lee, ed., Volume 8, 109-132, Elsevier Science Ltd., 2001

by

Lisa A. Kramer*

† This version corrects an omission in the published version by reinforcing the point made by Kramer [1998] that the methodology developed in this paper is useful for studying *non-clustered* events only.

* Rotman School of Management, University of Toronto, 105 St. George St., Toronto Ontario, M5S 3E6, Canada. Email: Lkramer@chass.utoronto.ca. I am grateful to Brad Barber, Glen Donaldson, Espen Eckbo, Wayne Ferson, Burton Hollifield, Mark Kamstra, Maurice Levi, Felice Martinello, Brendan McCabe, Nagpurnanand Prabhala, Ken White, and seminar participants at Simon Fraser University, the University of British Columbia, the University of Toronto, and participants of the Western Finance Association meetings. All remaining errors are my own. I thank the editor and an anonymous referee for helpful comments. This work was financially supported by the Social Sciences and Humanities Research Council of Canada.

Alternative Methods for Robust Analysis in Event Study Applications

Abstract

A variety of test statistics have been employed in the finance and accounting literatures for the purpose of conducting hypothesis tests in event studies. This paper begins by formally deriving the result that these statistics do not follow their conventionally assumed asymptotic distribution even for large samples of firms. Test statistics exhibit a statistically significant bias to size in practice, a result that I document extensively. This bias arises partially due to commonly observed stock return traits which violate conditions underlying event study methods. In this paper, I develop two alternatives. The first involves a simple normalization of conventional test statistics and allows for the statistics to follow an asymptotic standard normal distribution. The second approach augments the simple normalization with bootstrap resampling. These alternatives demonstrate remarkable robustness to heteroskedasticity, autocorrelation, non-normality, and event-period model changes, even in small samples, and they are useful for event studies with non-clustered events.

1 Introduction

This paper focuses on test statistics underlying financial event studies. In such studies, one analyzes the information content of corporate events, making use of stock returns for a collection of firms. The goal is to determine whether a particular financial event, such as an equity issue, debt offering, merger, or regulation change, had a significant effect on firms' returns. Evidence of a significant effect may shed light on various questions including the impact on shareholder welfare, market efficiency, or the effectiveness of government intervention. Techniques have evolved from the seminal event study of stock splits by Fama, Fisher, Jensen, and Roll [1969] to the point that event study methods are a common source of "stylized facts" which influence policy decisions and the direction of research. Recent surveys of the state of the art of event study methods are provided by MacKinlay [1997] and Binder [1998].

Common characteristics of returns data can seriously compromise inference based on event study methods, a point that has not gone unnoticed in the literature. De Jong, Kemna, and Kloek [1992, page 29] report that "results obtained under the usual assumptions on the error process (homoskedastic, normal distribution) shows that ignoring the fat tails and the heteroskedasticity may lead to spurious results." Campbell and Wasley [1993, page 74] find that with daily NASDAQ returns, conventional test statistics "depart from their theoretical unit normal distribution under the null hypothesis." Several studies have provided modifications to conventional techniques which successfully address some of the concerns that arise in practice.[1] This study study contributes to the advancement of event study methodology by providing a feasible means of effectively dealing with a large set of problems under a very wide range of conditions.

The main components of the paper are as follows. In Section 2, I provide a simple correction to conventional test statistics that ensures they follow their assumed asymptotic distribution. In Section 3, I demonstrate that in finite samples, the characteristics of returns data lead to bias in the conventional test statistics, compromising inference. The corrected test statistics show little or no bias. In Section 4, I present a two-step bootstrap procedure that also yields test statistics robust to the problems that plague conventional methods. It is useful for applications where events are not clustered in calendar time. Conclusions follow.

[1] For example, Boehmer, Musumeci, and Poulsen [1991] propose an alternative event study test statistic to account for event-induced variance, Brockett, Chen, and Garven [1994] suggest that event study regression models should account for ARCH and stochastic parameters, and Corhay and Tourani Rad [1996] recommend accounting for GARCH effects. Nonparametric alternatives to conventional methods have also been proposed. For example Marais [1984] uses the bootstrap, Larsen and Resnick [1999] use the bootstrap with cross-sectional stochastic dominance analysis, and Dombrow, Rodriguez, and Sirmans [2000] apply Theil's non-parametric regression in the estimation of abnormal returns.

2 Event Study Methods: Conventional and New

There are two broad goals in conducting financial event studies: testing for a significant *information effect* in stock returns at the time of the event announcement (examples include Schipper and Thompson [1983], Malatesta and Thompson [1985] and Eckbo [1992]) and identifying factors which *determine* the information effect (see, for example, work by Eckbo, Maksimovic, and Williams [1990] and Prabhala [1997]). The first of these, testing for an information effect, is the focus of this study.

In testing for an information effect, one collects a series of consecutive stock returns for a sample of firms of interest along with the corresponding returns on a market index portfolio. A market model is then estimated for each of the firms, and tests are conducted to see if there is evidence of an impact on firms' stock returns at the time of the event. There are several possible approaches available for formulating test statistics to detect the information effect. The most flexible and most commonly adopted of these is based on a multivariate regression model including a dummy variable to pick up the event. The methodology is clearly laid out by Thompson [1985] and is briefly summarized as follows.[2]

Define N as the number of firms being considered. R_{it} is the return on firm i's share where $i = (1, \cdots, N)$, M_{it} is the return on the market index portfolio, and ϵ_{it} is an error term. The point in time at which the event announcement potentially impacts the firms' returns is denoted $t = +1$, hence a dummy variable is defined to equal 1 for $t = +1$ and zero otherwise.[3] The dummy variable picks up the unanticipated portion of the return, *i.e.* the event effect. For each of the N firms being considered, a market model is estimated over a time period of length T such as $t = (-130, \cdots, +10)$, *including* the date of the event and several days following the event:[4] [5]

$$R_{it} = \beta_{i0} + \beta_{i1} M_{it} + \beta_{iD} D_{it} + \epsilon_{it}, \qquad i = (1, \cdots, N). \tag{1}$$

[2]Kramer [1998] considers additional event study methods such as the cross-sectional approach and the Boehmer, Musumeci, and Poulsen [1991] standardized cross-sectional method. Experiment results indicated that except for the standardized cross-sectional method, all methods exhibit the high degree of problems documented in this paper.

[3]I assume single-day event periods here for notational ease. To test the significance of *cumulative* effects over *multiple* event days one simply defines the dummy variable accordingly.

[4]Though more complex alternatives are available, in this paper I consider the relatively simple market model used frequently in the literature. See for example, Brown and Warner [1985], MacKinlay [1997] and Binder [1998]. This permits me to determine how robust various event study methods are to unintentionally neglected features of actual data. Even with more complex models, the problems I document in this paper can arise because the true data generating process underlying stock returns is unknown to the researcher. In practice, one should always use a model that attempts to avoid mis-specification. In conjunction with a well-specified model, the methods I propose below offer a high degree of robustness.

[5]The notation T_i would be used to allow for different length estimation periods across firms. In most cases, T_i is constant across firms, in which case the use of T is unambiguous.

Estimating the market model produces a t-statistic, t_i, for each of the N estimated dummy variable coefficients $\hat{\beta}_{iD}$. These are used for testing the null hypothesis of no abnormal event day returns:

$$Z = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N}}. \tag{2}$$

Typically, the Z statistic is assumed to be distributed as standard normal. In fact, standard normality would follow only if the t_i were themselves independent and identically distributed as normal with mean zero and unit variance, which they are not in practice. At best, the t_i are Student t-distributed with a variance that differs from unity.[6] As a result, Z does not approach standard normal, even as the number of firms, N, is increased.

To achieve asymptotic standard normality, Z must be normalized by the theoretical standard deviation of the t_i. In practice, with returns displaying even a small degree of heteroskedasticity, non-normality, *etc.*, the t_i have a standard deviation that differs from the theoretical standard deviation. Thus the sample standard deviation of the t_i must be used to normalize Z. The standard deviation of the t_i is defined as:

$$\hat{\sigma}_N = \sqrt{\frac{\sum_{i=1}^{N}(t_i - \bar{t})^2}{N-1}}$$

where $\bar{t}$ is the mean of the t_i. Then the normalized test statistic is obtained by dividing Z by $\hat{\sigma}_N$:

$$\tilde{Z} = \frac{Z}{\hat{\sigma}_N}. \tag{3}$$

This normalized test statistic, $\tilde{Z}$, follows an asymptotical standard normal distribution (as $N \to \infty$). In practice, $\tilde{Z}$ is very robust relative to Z. This is detailed directly below.

[6] If the t_i were distributed as Student t, then the theoretical variance of the t_i would be $\frac{T-k}{T-k-2}$ (where k is the number of parameters estimated in Equation (1); $k = 3$ in this case). With Student t-distributed t_i, the distribution of Z would approach standard normal as the length of the time series, T, is increased (assuming changes in the market model parameters can be correctly specified which can prove difficult or impossible). In practice, characteristics of returns data are likely to lead to Z with an unknown distribution.

3 Impact of Returns Characteristics on Test Statistic Properties

In this section, I quantify the extent to which many well-documented characteristics of the data used in event studies compromise the use of event study test statistics. As I show below, neglecting features of the data such as heteroskedasticity, autocorrelation, non-normality, and changes in event-period variance can lead to test statistics which do not follow their assumed distribution, even as N grows quite large. That is, the properties of returns data compromise finite sample inference. Experiments investigating the size of test statistics are reported immediately below, and the power experiments follow in Section 4.

Examining the validity of standard event study techniques under circumstances commonly encountered in practice requires the use of data that are generated to display properties that closely match those of actual returns data. The use of simulated data in establishing the properties of test statistics is conventional and profuse. In the event study literature, simulated data is employed by authors including Acharya [1993] and Prabhala [1997]. Thus I first employ simulated data in this paper, which facilitates a clear analysis of the marginal impact of many features of the data. For comparative purposes, at the end of Section 4, I also provide results from experiments employing actual CRSP data as used by Brown and Warner [1980, 1985] and others.

In investigating size, *i.e.* the test statistic's behavior when there is no event present, I consider a sample of N = 30, 50, 100, or 200 firms estimated over a time period of T=141 days. The experimental design is as follows.

1. Disturbances, market returns, and model parameters were generated for each firm, and then each of the N firms' returns were generated according to a basic market model:

   $$R_{it} = \beta_{i0} + \beta_{i1} M_{it} + \epsilon_{it}$$

   where $t = (-130, \cdots, +10)$ and $i = (1, \cdots, N)$.

   The properties ascribed to the generated data closely matched those of actual financial returns data. Values were chosen to conservatively mimic what is observed in actual returns as documented in Appendix A. The unanticipated return series, ϵ_{it}, was generated with a standard deviation of 0.77, skewness of 0.15, kurtosis of 6.2, an event period jump in variance of 100%. The coefficient on market returns was generated with an event period increase of 100% and firms' autocorrelation coefficient was generated to be 0.1.

2. OLS was used to estimate each firm's market model, yielding N event-day dummy variable t-statistics. Then Z was computed according to Equation (2) and $\tilde{Z}$ was computed according to Equation (3).

$$Z = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N}} \tag{2}$$

$$\tilde{Z} = \frac{Z}{\hat{\sigma}_N} \tag{3}$$

Recall that $\hat{\sigma}_N$ is the standard deviation of the t_i statistics.

3. Steps 1 and 2 were repeated a total of 1000 times. Each of these 1000 replications can be thought of as a single event study, each generating the Z and $\tilde{Z}$ statistics. By simulating many event studies, strong statements can be made about the behavior of the event study tests.

4. The statistical size of the Z and $\tilde{Z}$ statistics was then evaluated. First, actual rejection rates were computed at various levels of significance, then each actual rejection rate was compared to the assumed "nominal" rejection rate (size) for the test. For example, for a hypothesis test conducted at the $\alpha = 1\%$ level of significance, under the null hypothesis of no abnormal return, the test statistic should indicate rejection 1% of the time if the statistical size of the test is correct. Thus, the actual rejection rate for the 1% test would be compared to its nominal size of 0.01. Standard confidence intervals were constructed around the nominal size to see if actual rejection rates differ significantly from those expected under the null hypothesis. With a large number of replications, strong conclusions can be reached because the confidence intervals around the nominal size values for each statistic are quite small.

Panel A of Table 1 contains results of this experiment for a various sample sizes. The first column lists the α level for conventional tests of significance, 1%, 5%, and 10%. The remaining columns contain results for samples of 30, 50, 100, and 200 firms respectively. The value listed in each cell is the actual rejection rate for the Z statistic of Equation (2) at the particular significance level. Rejection rates which differ significantly from their nominal size are denoted with asterisks.

Two things are remarkable about these results. First, the conventional Z statistic dramatically over-rejects for any significance level, α, and for any number of firms considered. The rejection rates are anywhere from 60% to 530% higher than they are expected under the null which, in practice, could cause a researcher to conclude that an event effect is present when there is none. Second, the degree of over-rejection is not reduced by increasing the sample size, N. This is not surprising given the Z statistic requires a normalization to follow an asymptotic (in N) standard normal distribution, as discussed in Section 2.

Panel B of Table 1 contains details on the experiment for the case of $\tilde{Z}$. Examining this portion of the table reveals that implementing the normalization eliminates the bias relative to the traditional method. For example, when Z was compared to the assumed standard normal distribution for the case of 50 firms, a rejection rate of 9.8% was observed although a rejection rate of 5% was expected under the null hypothesis. Under similar conditions, the $\tilde{Z}$ statistic rejection rate is 4.5% (which is not statistically different from 5%). The difference in performance of the Z and $\tilde{Z}$ statistics arises because of the normalization which permits $\tilde{Z}$ to follow an asymptotic standard normal distribution.

All of the previously discussed results are for the case where data simultaneously display a variety of commonly observed characteristics. An advantage of using simulated data is the ability to examine the marginal impact of particular features of the data individually and at various levels of intensity (for example, with different degrees of skewness and/or kurtosis, with higher or lower levels of event-period variance changes, with characteristics that vary across firms, *etc.*). Kramer [1998] reports the outcome of such experiments. The results indicate that conventional test statistics are significantly biased even when the characteristics of the data are less extreme relative to the data considered above. Event-period variance changes, autocorrelation, non-normality, and other common features of the data are each individually capable of invalidating inference, even when present at very low levels relative to what is conventionally observed in the actual returns data used in event studies. Significantly biased test statistics are observed under a wide range of circumstances, even when very large samples of firms are employed.

It is important to emphasize that the severe problems plaguing the Z statistic are not solved by increasing the number of firms in the sample. It has been shown above that in order to achieve asymptotic standard normality, the Z statistics require a normalization for non-unit variance. Without this normalization, test statistics display no improved performance as N, the number of firms, is increased.

4 The Two-Stage Bootstrap Approach

The aim of the above Monte Carlo analysis has been to demonstrate the practical repercussions of using a conventional event study test statistic. The significant bias of the conventional approach was eliminated by using a normalized version of the commonly employed test statistic. In this section, I present another alternative method which involves two stages.

The first stage involves normalizing the conventional Z statistics as shown in the previous section:

$$\tilde{Z} = \frac{Z}{\hat{\sigma}_N} \tag{3}$$

where $\hat{\sigma}_N$ is the standard deviation of the t_i statistics. The second stage involves using bootstrap resampling to generate the empirical distribution of the normalized test statistic. This alternative procedure will also be shown to demonstrate improved performance relative to traditional methods.

The Bootstrap

Use of the bootstrap involves repeatedly sampling from the actual data in order to empirically estimate the true distribution of a test statistic. This method was initially introduced by Efron [1979] as a robust procedure for estimating the distribution of independent and identically distributed data. Since its inception, the bootstrap's performance under a variety of conditions has been examined in depth in the statistics literature. Work by Liu [1988] establishes the suitability of adopting the bootstrap under conditions most applicable to this setting: that of independent but not necessarily identically distributed observations. Provided the random observations are drawn from distributions with similar means (but not necessarily identical variances) and provided the first two moments are bounded, use of the bootstrap is valid. Several books and articles on the subject provide a good overview of the use of the bootstrap for empirical work in many fields, including Hall [1992], LePage and Billard [1992], Efron and Tibshirani [1993], Hjorth [1994], Li and Maddala [1996], and Manly [1998].

Some of the many applications in finance include Malliaropulos' [1999] paper on mean reversion, a study of mutual fund performance by Cai, Chan, and Yamada [1997], Kothari and Shanken's [1997] study of expected real stock market returns, an investigation of long-horizon security price performance by Kothari and Warner [1997], Malliaropulos' [1996] study of the predictability of long-horizon stock returns using the bootstrap, Liang, Myer and Webb's [1996] bootstrap estimation of the efficient frontier for mixed-asset portfolios, and an investigation of long-horizon predictability of exchange rates by Mark [1995]. In the context of event studies, Marais [1984] uses bootstrapped p-values to conduct inference in conjunction with the standardized residual approach, and Larsen and Resnick [1999] use the bootstrap with cross-sectional stochastic dominance analysis. Given the abundant use of the bootstrap in finance, economics, and beyond, I consider the marginal impact of the bootstrap in the context of event studies.

Implementation

In Figure 1, I provide a diagrammatic representation of the two-stage bootstrap approach for conducting inference in event studies. The procedure is also discussed in detail below. While the explanation makes use of the Z statistic which emerges from a regression framework employing a dummy variable, the steps can be modified in a straightforward manner to enable inference based on any common event study test statistic, as shown by Kramer [1998].

1a. Estimate the appropriate event study market model for each of the N firms in the sample. The most simple possibility is as follows:

$$R_{it} = \beta_{i0} + \beta_{i1} M_{it} + \beta_{iD} D_{it} + \epsilon_{it}, \qquad i = (1, \cdots, N). \tag{1}$$

The estimation yields N t-statistics: one for each firm's estimated dummy variable coefficient. As shown in Figure 1, this collection of t-statistics forms the pool of data upon which the conventional Z statistic is based.

$$Z = \frac{\sum_{i=1}^{N} t_i}{\sqrt{N}} \tag{2}$$

A researcher interested in conducting conventional inference would stop at this point and compare the value of the Z statistic to a critical value from the assumed standard normal distribution. As indicated earlier, that distribution does not apply, even asymptotically as N is increased.

1b. Normalize the Z statistic obtained in Step 1a to account for the fact that its variance differs from unity in practice. First, compute the sample standard deviation of the t_i:

$$\hat{\sigma}_N = \sqrt{\frac{\sum_{i=1}^{N} (t_i - \bar{t})^2}{N - 1}}.$$

where $\bar{t}$ is the mean of the t_i statistics. Then, divide Z by $\hat{\sigma}_N$ to yield the normalized version of Z:

$$\tilde{Z} = \frac{Z}{\hat{\sigma}_N} \tag{3}$$

In the next stage of this method, the empirical distribution of $\tilde{Z}$ will be constructed, facilitating reliable inference.

2a. Under the null hypothesis, the distribution of $\tilde{Z}$ is centered about zero, and hence the empirical distribution must be constructed such that it is also centered about zero. Notice that the N actual t-statistics calculated in Step 1a have a mean, $\bar{t}$, which generally differs from zero. If these t-statistics were used directly to build the empirical distribution, the result would be a distribution centered about the actual value of $\tilde{Z}$. This would occur because in the absence of imposing the null distribution, the distribution of the sample would be replicated, with the sample mean exactly in the middle. Therefore, prior to constructing the empirical distribution, the t-statistics must be adjusted to impose the null hypothesis of no event day abnormal returns (*i.e.* zero mean). Accordingly, a collection of *mean-adjusted* t-statistics, denoted t_i^* is assembled by deducting $\bar{t}$ from each of the individual t-statistics: $t_i^* = t_i - \bar{t}$.

The N mean-adjusted t-statistics are, of course, mean zero, and they constitute the collection of statistics – the population – from which bootstrap samples are drawn in the next step. Having mean-adjusted the t-statistics, the empirical distribution will be centered about zero, allowing one to conduct inference under the null hypothesis of no event.

2b. The mean-adjusted data are used to construct an empirical distribution for $\tilde{Z}$ under the null. This involves drawing many random samples, called "bootstrap samples," from the population of t_i^* statistics. As shown in Figure 1, a single bootstrap sample is constructed by randomly drawing with replacement N observations from the collection of t_i^* statistics. A total of 1000 such bootstrap samples, individually denoted $b = (1, \cdots, 1000)$, are constructed, with each bootstrap sample containing N observations.[7] The particular composition of each bootstrap sample varies randomly. For example, the first sample might contain duplicate occurrences of some of the t_i^* statistics and no occurrences of other t_i^* statistics, the second sample might contain duplicate occurrences of some different t_i^* statistics, etc. The make-up of each of the 1000 constructed bootstrap samples is determined purely by chance.

[7]Sources such as Efron and Tibshirani [1993] indicate that 1000 bootstrap samples are sufficient for constructing confidence intervals. I verified this result through extensive experimentation. Increasing the number of bootstrap samples to 10 000 leads to no marked change in results.

2c. The construction of the normalized test statistic for each bootstrap sample is parallel to that of the $\tilde{Z}$ statistic shown earlier. Compute Z_b for each of the 1000 bootstrap samples, where the $b = (1, \cdots, 1000)$ subscript is used to specify the particular bootstrap sample and the $j = (1, \cdots, N)$ subscript distinguishes between particular observations within a given bootstrap sample:

$$Z_b = \frac{\sum_{j=1}^{N} t^*_{b,j}}{\sqrt{N}}. \tag{4}$$

Define the mean of the $t^*_{b,j}$ statistics in a particular bootstrap sample as $\bar{t}^*_b$. The standard deviation of the $t^*_{b,j}$ for each bootstrap sample is:

$$\hat{\sigma}_{N,b} = \sqrt{\frac{\sum_{j=1}^{N}(t^*_{b,j} - \bar{t}^*_b)^2}{N-1}}.$$

Each of the 1000 Z_b statistics of Equation (4) is then normalized by the corresponding $\hat{\sigma}_{N,b}$, yielding 1000 $\tilde{Z}_b$ statistics:

$$\tilde{Z}_b = \frac{Z_b}{\hat{\sigma}_{N,b}}. \tag{5}$$

2d. Ordering the collection of 1000 $\tilde{Z}_b$ statistics from smallest to largest defines the empirical distribution. The histogram depicted at the bottom of Figure 1 is an example of such an empirical distribution. Inference is conducted by comparing the $\tilde{Z}$ statistic from Step 1b to critical values from the empirical distribution. For example, with 1000 bootstrap samples, a 5% left-tail critical value, $\tilde{Z}^{.05}$, is the 50th largest value of the $\tilde{Z}_b$ statistics and a 5% right-tail critical value, $\tilde{Z}^{.95}$, is the 950th largest of the $\tilde{Z}_b$ statistics. If the value of the $\tilde{Z}$ statistic happens to be larger than 95% of the bootstrap $\tilde{Z}_b$ statistics (*i.e.* exceeding $\tilde{Z}^{.95}$) or smaller than 5% of the bootstrap $\tilde{Z}_b$ statistics (*i.e.* falling below $\tilde{Z}^{.05}$), one rejects at the 10% level of significance the two-sided null hypothesis of no abnormal returns.

To summarize, applying Steps 1a - 2d of the two-stage bootstrap approach based on the conventional Z statistic basically involves computing Z using the actual t_i statistics and normalizing it with the variance of the t_i to impose unit variance. This yields $\tilde{Z}$. The t_i statistics are then mean-adjusted to form a population of statistics, the t^*_i, from which random re-samples are drawn. 1000 bootstrap samples are formed, each containing

N observations, and $\tilde{Z}_b$ is computed for each bootstrap sample. The collection of all the $\tilde{Z}_b$ statistics defines the empirical distribution. Finally, event study inference is conducted by comparing $\tilde{Z}$ to critical values from the empirical distribution.[8]

Size

Using the experimental design of Section 3, I now explore the performance of the two-stage approach. As before, data were generated with skewness of 0.15, kurtosis of 6.2, an increase in event period variance of 100%, an increase in the event period market return coefficient of 100%, and an autocorrelation coefficient of 0.1. 1000 replications were conducted for computing the rejection rates for the test statistics, and 1000 bootstrap samples were drawn. Results indicate the two-stage approach eliminates the bias of conventional test statistics shown in Section 3, and the statistical power of the two-stage approach compares favorably with conventional methods.

Panel C of Table 1 contains the results for the $\tilde{Z}$ statistic using bootstrapped critical values. As with the normalized test statistic shown in Panel B, virtually no bias remains. Extensive sensitivity analysis (unreported) indicates that the performance of the technique is robust to perturbations in the properties ascribed to the generated data. The bootstrap approach maintains its unbiased size when applied to data displaying a very wide range of commonly observed properties.

Another alternative method one might consider is to simply take the conventional Z statistic and compare it a distribution constructed using the bootstrap procedure in the spirit of Marais [1984] (without first normalizing for non-unit variance). Results from these experiments, shown in Panel D of Table 1 indicate such an approach has properties similar to the other alternative methods discussed in this paper.

Power

The power of the event study test statistics is evaluated by employing data generated to have a positive abnormal return at the time of the event. Abnormal returns of 0.5%, 0.7%, and 0.9% and a sample size of 50 firms are considered here. According to convention, I document size-adjusted power.[9] Results are shown in Table 2.

Column A contains the rejection rates based on comparing the Z statistic to its commonly assumed standard normal distribution, Column B contains rejection rates based on

[8] It is worth emphasising that use of the bootstrap in this setting requires that the t_i statistics be *independently* distributed. For applications where cross-firm correlation may be present (or for applications where time-series correlation may be present in the case of a multi-day event-period), use of the moving sample bootstrap may be advisable. See Liu and Singh [1992] for details.

[9] See, for example, Davidson and MacKinnon [1993] for a description of size-adjusted test statistics.

comparing the Z statistic to the standard normal distribution, and Column C contains rejection rates based on comparing the $\tilde{Z}$ statistic to its bootstrapped distribution. Column D pertains to the case of comparing the conventional Z statistic to a bootstrap distribution. The first case considered is that of adding abnormal returns of 0.9% on the event day. As shown in the top panel of Table 2, all of the test statistics reject almost 100% of the time in this case. When abnormal returns of 0.7% are added on the event day, shown in the middle panel of the table, results are qualitatively similar. Rejection rates for conventional inference and the inference based on the alternative approach remain very close to 100%. With abnormal returns of 0.5%, shown in the bottom panel, the conventional rejection rates and the rejection rates for the alternative approaches fall slightly, but the performance is qualitatively similar across methods. The overall conclusion to draw is that power under the alternative approaches is very comparable to that of conventional methods. Rejection rates are almost identical for all cases considered. Similar results are obtained under various conditions.

Other Nonparametric Methods

In Kramer [1998], I evaluated the properties of nonparametric methods including the rank test of Corrado [1989] and the sign test. While both the rank test and the sign test demonstrated less bias than conventional Z test statistics, significant bias remained, even for samples with large number of firms. A sample size of 1000 firms was required before the majority of the rejection rates were insignificantly different from what would be expected from an unbiased test statistic.

Thus, for inference in samples of fewer than 1000 firms (and in general), I advocate use of one of the alternative approaches documented in this paper.[10] The outstanding performance of these approach applied to data exhibiting a wide range of the characteristics of actual returns is testament to their robustness. Valid and powerful inference is facilitated, even in situations where conventional methods fail.

Use of Actual Returns Data in Monte Carlo Experiments

The experiments documented earlier in this paper make use of simulated data. Some past researchers have employed *actual* returns data for investigating the performance of event study test statistics. Thus, for comparative purposes, I replicated the above experiments using actual returns data in place of simulated data. I followed the approach laid out by

[10]For the two-stage bootstrap approach, the computations for a single event study require little additional CPU time relative to conventional methods, and they can be undertaken with any standard statistical package. Several statistical analysis packages, including Shazam, have the bootstrap method built right in.

Brown and Warner [1985] and Boehmer, Musumeci, and Poulsen [1991], creating portfolios of firms by randomly resampling from actual CRSP data. The results from these experiments are qualitatively identical to those presented above.

For example, in exploring event-induced variance, I once again considered the impact of a 100% increase with various sample sizes. Results appear in Table 3. As before, Panel A applies to the conventional Z statistic using a standard normal distribution, Panel B is for the $\tilde{Z}$ statistic compared to a standard normal, Panel C is for the $\tilde{Z}$ statistic using the bootstrap distribution, and Panel D is for the case of the Z statistic compared to a bootstrap. Samples of 30, 50, 100, and 200 firms were considered, and 1000 replications were conducted. Results shown in the table indicate that the conventional Z statistic is significantly biased while the three alternative methods are mostly unbiased. Unreported experiment results based on other commonly encountered levels of increased variance (50% and 200%) and considering other typical characteristics of the data were qualitatively similar.[11]

5 Conclusions

There are several different analytic techniques commonly used in event studies – both parametric and nonparametric. These approaches differ in market model specification and estimation or differ in the calculation of the statistics used for hypothesis testing. A common feature of all of the approaches, however, is that basic underlying conditions must be satisfied for test statistics to have their assumed distribution. These conditions are typically violated in practice, invalidating inference. Monte Carlo results presented above indicate the statistical size of commonly employed test statistics is significantly biased when data exhibit characteristics identical to those observed in actual stock returns.

Researchers attempting to conduct event studies with small samples typically recognize that conventionally assumed distributions may be inappropriate for conducting hypothesis tests, and hence they may attempt to collect data for a larger set of firms. (Event studies on samples of 15 or 20 firms are, nonetheless, fairly common.) In the past, the rationale for such an increase in sample size may have been based on appeals to asymptotic theory. In this paper, however, I have argued that even with large numbers of firms, the asymptotic distribution of conventional test statistics is not valid.

As a solution to such problems, I have presented alternative testing procedures based on normalizing conventional Z statistics and/or empirically estimating their distribution with the bootstrap. I presented evidence to establish the validity of using such approaches on

[11] Conventional test statistics generally showed highly significant bias while the alternative methods showed little or no bias.

data with properties like those of actual stock returns. The techniques are not prone to exhibit biased size in common situations which render bias in conventional techniques – with no sacrifice of power – hence I recommend their use for hypothesis testing in event studies where events are not clustered in calendar time. In situations where the incidence of non-normality or heteroskedasticity in the data is particularly extreme, it may be prudent to adopt the two-stage bootstrap approach, as in practice it will be most robust to the presence such features of the data.

For studies where events are clustered in calendar time, the bootstrap approach developed here is not applicable, for instance in studies which aim to determine the impact upon a collection of firms following a change in a particular government policy or a change in regulation pertaining to an industry. With clustered events, the residuals are highly correlated across firms. Brown and Warner [1985] document the fact that conventional test statistics often do not follow their assumed distribution in the presence of clustering, and they find that adjusting for the dependence across firms can introduce a substantial loss of power. A study of the utility of a refined version of the bootstrap approach in the context of clustered events is postponed for future consideration.

References

Acharya, S., 1993, "Value of latent information: Alternative event study methods," *Journal of Finance*, 48, 363-385.

Binder, John J. 1998, "The Event Study Methodology since 1969," *Review of Quantitative Finance and Accounting*, 11(2), September, 111-137.

Boehmer, E., J. Musumeci, and A. B. Poulsen, 1991, "Event-study methodology under conditions of event-induced variance," *Journal of Financial Economics*, 30, 253-273.

Bollerslev, T., R.Y. Chou, and K.F. Kroner, 1992, "ARCH modeling in finance," *Journal of Econometrics*, 52, 5-59.

Boudoukh, Jacob, Matthew P. Richardson, and Robert F. Whitelaw, 1994, "A tale of three schools: Insights on autocorrelations of short-horizon stock returns," *Review of Financial Studies*, 7(3), 539-573.

Brockett, P.L, H.M. Chen, and J.R. Garven, 1999, "A new stochastically flexible event methodology with application to proposition 103," *Insurance Mathematics and Economics*, 25(2), 197-217.

Brown, S.J. and J.B. Warner, 1980, "Measuring security price performance," *Journal of Financial Economics*, 8, 205-258.

Brown, S.J. and J.B. Warner, 1985, "Using daily stock returns: The case of event studies," *Journal of Financial Economics*, 14, 3-31.

Cai, Jun, K.C. Chan, and Takeshi Yamada, 1997, "The performance of Japanese mutual funds," *Review of Financial Studies*, 10(2), 237-273.

Campbell, Cynthia J. and Charles E. Wasley, 1993, "Measuring security price performance using daily NASDAQ returns," *Journal of Financial Economics*, 33, 73-92.

Corhay, A. and A. Tourani Rad, 1996, "Conditional heteroskedasticity adjusted market model and an event study," *The Quarterly Journal of Economics and Finance*, 36(4), 529-538.

Corrado, Charles J., 1989, "A nonparametric test for abnormal security-price performance in event studies," *Journal of Financial Economics*, 23, 385-395.

Davidson, Russell and James G. MacKinnon, 1993. *Estimation and inference in econometrics* (Oxford University Press, New York).

de Jong, F., A. Kemna, and T. Kloek, 1992, "A contribution to event study methodology with an application to the Dutch stock market," *Journal of Banking and Finance*, 16, 11-36.

Dombrow, Jonathan, Mauricio Rodriguez, and C.F. Sirmans, 2000, "A complete nonparametric event study approach," *Review of Quantitative Finance and Accounting*, 14(4), June, 361-380.

Donaldson, R. Glen, and Frank Hatheway, 1994, "An expanded approach to the empirical analysis of illegal insider trading and stock prices," University of British Columbia manuscript.

Donaldson, R. Glen, and Mark Kamstra, 1997, "An artificial neural network-GARCH model for international stock return volatility, " *Journal of Empirical Finance*, 4, 17-46.

Eckbo, B. Espen, 1992, "Mergers and the value of antitrust deterrence," *Journal of Finance*, 47, 1005-1029.

Eckbo, B. Espen, Vojislav Maksimovic, and Joseph. Williams, 1990, "Consistent estimation of cross-sectional models in event studies," *The Review of Financial Studies*, 3, 343-365.

Efron, B., 1979, "Bootstrap methods: Another look at the jackknife," *Annals of Statistics*, 7, 1-26.

Efron, B. and R.J. Tibshirani, 1993. *An introduction to the bootstrap* (Chapman & Hall, New York).

Fama, E.F., 1965, "The behavior of stock market prices," *Journal of Business*, 34, 420-429.

Fama, E.F., L. Fisher, M. Jensen, and R. Roll, 1969, "The adjustment of stock prices to new information," *International Economic Review*, 10, 1-21.

Glosten, L.R., R. Jagannathan, and D.E. Runkle, 1993, "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779-1801.

Gupta, A., R.L.B. LeCompte, and L. Misra, 1993, "FSLIC assistance and the wealth effects of savings and loans acquisitions," *Journal of Monetary Economics*, 31, 117-128.

Hall, P., 1992. *The bootstrap and Edgeworth expansion* (Springer-Verlag, New York).

Hjorth, J.S.U., 1994. *Computer intensive statistical methods: Validation, model selection and bootstrap* (Chapman & Hall, New York).

Kon, S.J., 1984, "Models of stock returns – a comparison," *Journal of Finance*, 34, 147-165.

Kothari, S.P. and Jay Shanken, 1997, "Book-to-market, dividend yield, and expected market returns: a time-series analysis," *Journal of Financial Economics*, 44(2), May, 169-203.

Kothari, S.P. and Jerold B. Warner, 1997, "Measuring long-horizon security price performance," *Journal of Financial Economics*, 43, March, 301-340.

Kramer, Lisa A., 1998, "Banking on event studies: statistical problems, a new bootstrap solution and an application to failed bank acquisitions," PhD Dissertation, University of British Columbia.

Lamoureux, C.G. and W.D. Lastrapes, 1990, "Heteroskedasticity in stock return data: volume versus GARCH effects," *Journal of Finance*, 45, 221-229.

Larsen, Glen A., Jr. and Bruce Resnick, 1999, "A performance comparison between cross-sectional stochastic dominance and traditional event study methodologies," *Review of Quantitative Finance and Accounting*, 12, 103-112.

LePage, R. and L. Billard, 1992. *Exploring the limits of bootstrap* (John Wiley & Sons, Inc., New York).

Li, H., and G.S. Maddala, 1996, "Bootstrapping time series models," *Econometric Reviews*, 15, 115-158.

Liang, Y., F.C.N. Myer, and J.R. Webb, 1996, "The bootstrap efficient frontier for mixed-asset portfolios," *Real Estate Economics*, 24, 247-256.

Liu, R.Y., 1988, "Bootstrap procedures under some non-iid models," *Annals of Statistics*, 16, 1696-1708.

Liu, R.Y. and K. Singh, 1992, "Moving blocks jackknife and bootstrap capture weak dependence," in R. LePage and L. Billard, eds, *Exploring the Limits of Bootstrap* (John Wiley & Sons, Inc., New York).

MacKinlay, Craig A., 1997, "Event studies in economics and finance," *Journal of Economic Literature*, 34, March, 13-39.

Malatesta, P.H. and R. Thompson, 1985, "Partially anticipated events: A model of stock price reactions with an application to corporate acquisitions," *Journal of Financial Economics*, 14, 237-250.

Malliaropulos, D., 1999, "Mean reversion in southeast Asian stock markets," *Journal of Empirical Finance*, 6(4), 355-384.

Malliaropulos, D., 1996, "Are long-horizon stock returns predictable? A bootstrap analysis," *Journal of Business Finance and Accounting*, 23, 93-106.

Mandelbrot, B., 1963, "The variation of certain speculative prices," *Journal of Business*, 36, 394-419.

Manly, Bryan F. J., 1998, *Randomization, bootstrap and Monte Carlo methods in biology*, (Chapman & Hall, New York.)

Marais, M. Laurentius, 1984, "An application of the bootstrap method to the analysis of squared standardized market model prediction errors," *Journal of Accounting Research*, 22 Supplement, 34-54.

Mark, Nelson C., 1995, "Exchange rates and fundamentals: Evidence on long-horizon predictability," *American Economic Review*, 85(1), March, 201-218.

McCabe, B., 1989, "Misspecification tests in econometrics based on ranks," *Journal of Econometrics*, 40, 261-278.

Nelson, D.B., 1990, "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, 59, 347-370.

Officer, R.R., 1967, "The distribution of stock returns," *Journal of the American Statistical Association*, 67, 807-812.

Prabhala, N.R., 1997, "Conditional methods in event-studies and an equilibrium justification for standard event-study procedures," *The Review of Financial Studies*, Spring, 1-38.

Ramberg, J.S., P.R. Tadikamalla, E.J. Dudewicz, and E.F. Mykytka, 1979, "A probability distribution and its uses in fitting data," *Technometrics*, 21, May, 78-82.

Ramberg, J.S. and B.W. Schmeiser, 1972, "An approximate method for generating symmetric random variables," *Communications of the Association for Computing Machinery*, 15, 987-990.

Ramberg, J.S. and B.W. Schmeiser, 1974, "An approximate method for generating symmetric random variables," *Communications of the Association for Computing Machinery*, 17, 78-82.

Schipper, K. and R. Thompson, 1983, "The impact of merger-related regulations on the shareholders of acquiring firms," *Journal of Accounting Research*, 21, 184-221.

Thompson, R., 1985, "Conditioning the return-generating process on firm-specific events: A discussion of event study methods," *Journal of Financial and Quantitative Analysis*, 20, 151-168.

A Further Details on Experiment Design

For the Monte Carlo approach outlined in Section 3 above, non-normality and changes in the Data Generating Process (DGP) were incorporated in the generated data for each firm. Several empirical studies and actual data were consulted in order to choose parameter values that would accurately reflect the properties of actual returns data. The choice of particular parameter values for generating data and the choice of the algorithms employed are motivated below.

- *Non-Normality and Heteroskedasticity*

  There is considerable evidence that returns data are not normally distributed. In particular, skewed and fat-tailed distributions have been documented extensively. For some early evidence of these characteristics in firms' return data, see Mandelbrot [1963], Fama [1965], and Officer [1967]. An investigation by Kon [1984] reveals, almost without exception, significant skewness and excess kurtosis (fat tails) among daily returns for 30 individual stocks and 3 standard market indexes. Bollerslev, Chou, and Kroner [1992, page 11] state it is "widely recognized that the unconditional price or return distributions tend to have fatter tails than the normal distribution." They also observe that even by accounting for Autoregressive Conditional Heteroskedasticity (ARCH), one may fail to capture all of the non-normality in returns data: "standardized residuals from the estimated models ... often appear to be leptokurtic." ARCH is a well documented empirical regularity of stock returns data, as evidenced by its voluminous literature (much of which is cited in the survey by Bollerslev, Chou, and Kroner [1992]). Neglecting the time-varying variance returns data can result in non-normality of the disturbances in a market model and hence Z statistics which do not follow their assumed distributions, even for large samples.

  Thus, in determining properties to assign the data generated for experiments, I consulted past studies such as those by Kon [1984] and Lamoureux and Lastrapes [1990] which document the statistical properties of actual stock returns and market returns data,[12] including mean, variance, skewness, and kurtosis, as well as conditional heteroskedasticity parameters. The intention was to simulate stock returns and market returns data with specific statistical properties that closely match actual data. Kon documents the first four moments over a particular sample period for 30 individual

[12] I verified these researchers' findings myself using actual CRSP data. The reported values appear to be correct with the exception of minor typographical errors. I also verified the magnitude of Kon's reported moment values on other sample periods and found the values to be quite similar with the exception of the time around the crash of 1987.

firms traded on the NYSE and for several market indices. His results for the moments are as follows. Positive skew for individual firms was observed in twenty-eight of the thirty cases considered. Of these positive cases, the skewness ranged between 0.0678 and 0.9080. The median of all thirty cases was about 0.32, and most values were between 0.30 and 0.40. Excess kurtosis was observed in all 30 stocks considered by Kon. The range in the kurtosis coefficient was 4.8022 to 13.9385, with a median of about 6.3, and with most values between 5 and 7. The standard deviation of returns exceeded 0.77 for all firms and was observed as high as 2.89. For the experiments conducted in this study, skewness up to 0.15, kurtosis up to 6.2, and a standard deviation of 0.77 were adopted in generating conservative data for firms' disturbances.

- *Autocorrelation*

  The evidence on autocorrelation in stock returns is voluminous. For example, Donaldson and Kamstra [1997] report autocorrelation coefficients for daily data in several international and domestic market indices ranging from 0.07 to 0.3. Likewise, Boudoukh, Richardson, and Whitelaw [1994] report autocorrelation coefficients for returns on portolios of US firms ranging from below 0.1 to above 0.5. For the experiments conducted in this study, a conservative autocorrelation coefficient of 0.1 is employed in generating market returns and firms' returns.

- *Changes in the DGP*

  There is considerable evidence that the data generating process for returns can change dramatically during the time of an event. For example, Boehmer, Musumeci, and Poulsen [1991] report that most researchers who have investigated event-induced variance changes have found variances can increase anywhere from 44% to 1100% during event periods. Donaldson and Hatheway [1994] also find evidence of variance changes – both increases and decreases – during periods of insider trading. Of their cases where a *rise* in variance is observed during the event period, the amount of the increase ranges from about 4% to 680%. Failure to model such changes can render the market model specification invalid leading to event study test statistics which do not follow their assumed distributions, even asymptotically.[13]

  Likewise, several researchers have established that market model coefficients can also undergo changes around the time of the event (or follow some distribution or some

[13]Furthermore, variance estimates used in variously defined versions of the Z statistic embed different assumptions about the behavior of the disturbances. If the estimated market model does not properly account for the time-varying behavior of the disturbances, then different conclusions may be reached with the use of different test statistics, even when employing the same data.

pattern over time). De Jong, Kemna, and Kloek [1992] find evidence that the coefficient on the market return is not necessarily constant over the estimation period in event studies. They find it often follows a mean-reverting AR process. Donaldson and Hatheway [1994] demonstrate the importance of allowing for changes in the intercept and market return coefficient at the time of the event. They find that the market return coefficient can fall by as much as 106% or rise by as much as 4238% in the collection of firms they consider. Brockett, Chen, and Garven [1999] demonstrate the importance of allowing a time-varying stochastic market return coefficient.

Thus, for the experiments conducted in this paper, The event period variance can increase by as much as 100% and the event period market return coefficient can rise by as much as 100%. The change in event period variance was incorporated in the data by re-scaling event period disturbances to have a variance up to 100% larger than that of non-event period disturbances.

- *Generating the Data*

  There are many reasonable options for generating non-normal returns data, including the following. Returns can be modeled to incorporate excess kurtosis by using a Student t-distribution with low degrees of freedom. (Bollerslev and Wooldridge [1992], for example, use a Student t with 5 degrees of freedom to generate fat-tailed data for their Monte Carlo simulations.) Skewness can be incorporated by making use of asymmetric models of conditional variance, such as the EGARCH model of Nelson [1990] or the Modified GARCH model of Glosten, Jagannathan, and Runkle [1993]. Alternatively, both skewness and excess kurtosis can be simultaneously incorporated by making use of an algorithm described in Ramberg, Dudewicz, Tadikamalla, and Mykytka [1979]. (This algorithm is a generalization of Tukey's lambda distribution, and it was developed by Ramberg and Schmeiser [1974, 1975]. For an application in the context of a simulation study, see McCabe [1989].) Basically, the Ramberg *et al.* algorithm allows one to select particular values for the first four moments of a distribution in generating random variates. For experiments reported in this paper, I adopted the Ramberg *et al.* algorithm to generate returns data with the first four moments matching those of actual data. Results of experiments based on use of the Student t-distribution to generate leptokurtic but symmetric data are qualitatively similar.

In this study, parameters for the Ramberg *et al.* algorithm were selected by consulting various studies (including Kon [1984]) and by examining actual returns data. Disturbance terms for all firms were conservatively generated with kurtosis of 6.2, skewness of 0.15, standard deviation of 0.77, and zero mean. The change in the market return coefficient during the event period was incorporated as follows. During the non-event period, $t = (-130, \cdots, -11)$, the coefficient of market returns was set to equal one, while during the event period, $t = (-10, \cdots, +10)$, the coefficient doubled. A market model intercept of zero was assumed for all firms.

Allowing firms to have different baseline values for the market return coefficient and the market model intercept (matching values observed in practice, for example) does not affect the experiment results. The explanation for this invariance with respect to choice of parameter value relies on the fact that OLS is unbiased. As explained in Section 3, the market returns, M_{it}, and the disturbances, ϵ_{it}, are generated with particular properties to match actual data. Then setting $\beta_{i0} = 0$ and $\beta_{i1} = 1$, firms' returns, R_{it}, are generated according to the basic market model $R_{it} = \beta_{i0} + \beta_{i1} M_{it} + \epsilon_{it}$. When R_{it} is regressed on M_{it} and a constant, the OLS estimators for β_{i0} and $\beta_{i1} = 1$ are unbiased. That is the estimated intercept and market return coefficient are equal to the chosen values on average. Thus, restricting the chosen values to be zero and one for all firms is a harmless simplification. See Marais [1984, page 42] for further details.

Table 1: Test Statistic Properties: Non-Normal Data with DGP Changes

Rejection Rates
for Tests at Common Significance Levels
1000 Replications, Various Numbers of Firms

Panel A: Z Statistic
Using the Standard Normal Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.043***	0.040***	0.044***	0.054***
0.05	0.102***	0.098***	0.110***	0.119***
0.10	0.161***	0.158***	0.181***	0.191***

Panel B: $\tilde{Z}$ Statistic
Using the Standard Normal Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.009	0.006	0.010	0.012
0.05	0.054	0.045	0.054	0.059*
0.10	0.104	0.096	0.107	0.113*

Panel C: $\tilde{Z}$ Statistic
Using the Bootstrap Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.014	0.013	0.012	0.012
0.05	0.059*	0.048	0.057	0.059*
0.10	0.109	0.095	0.111	0.117**

Panel D: Z Statistic
Using the Bootstrap Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.011	0.009	0.007	0.011
0.05	0.052	0.046	0.054	0.060*
0.10	0.109	0.095	0.110	0.113*

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

Table 2: Power Comparisons

Rejection Rates
for Tests at Common Significance Levels
1000 Replications, 50 Firms

Ab. Return	Sig. Level (α)	(A) Z Using Std. Normal	(B) $\tilde{Z}$ Using Std. Normal	(C) $\tilde{Z}$ Using Bootstrap	(D) Z Using Bootstrap
0.009	0.01	1.000	0.999	0.999	1.000
	0.05	1.000	1.000	1.000	1.000
	0.1	1.000	1.000	1.000	1.000
0.007	0.01	0.988	0.986	0.962	0.990
	0.05	0.998	0.999	0.996	0.999
	0.1	1.000	0.999	0.999	0.999
0.005	0.01	0.778	0.810	0.737	0.822
	0.05	0.946	0.954	0.929	0.948
	0.1	0.986	0.987	0.976	0.988

Table 3: Test Statistic Properties: Using Actual Data

Rejection Rates
for Tests at Common Significance Levels
1000 Replications
100% Increase in Event Period Variance

Panel A: Z Statistic
Using the Standard Normal Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.033***	0.036***	0.044***	0.048***
0.05	0.106***	0.098***	0.110***	0.117***
0.10	0.172***	0.173***	0.166***	0.172***

Panel B: $\tilde{Z}$ Statistic
Using the Standard Normal Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.011	0.012	0.006	0.015*
0.05	0.051	0.044	0.054	0.048
0.10	0.107	0.096	0.103	0.104

Panel C: $\tilde{Z}$ Statistic
Using the Bootstrap Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.010	0.012	0.006	0.014
0.05	0.054	0.044	0.054	0.051
0.10	0.112	0.096	0.103	0.100

Panel D: Z Statistic
Using the Bootstrap Distribution

Sig. Level (α)	30 Firms	50 Firms	100 Firms	200 Firms
0.01	0.012	0.015*	0.011	0.015*
0.05	0.056	0.052	0.059*	0.054
0.10	0.108	0.106	0.105	0.105

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

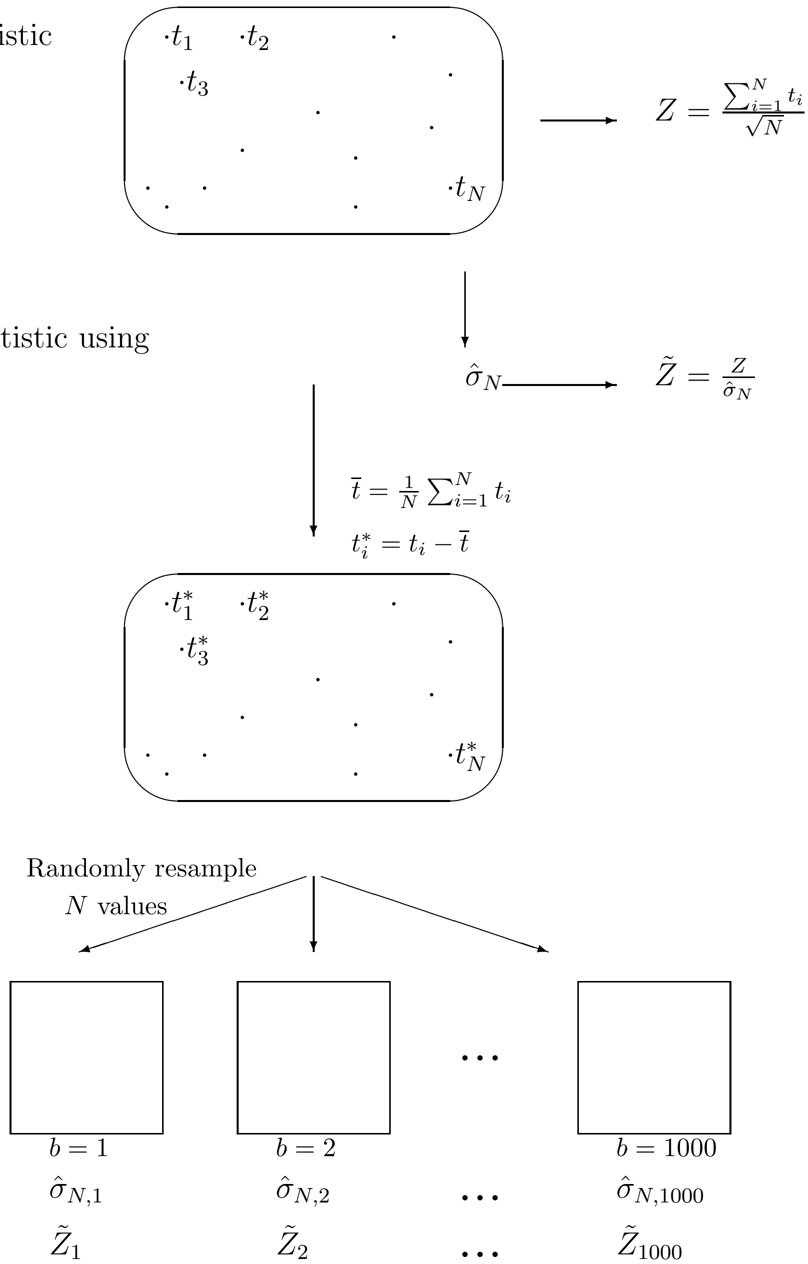
FIGURE 1: The Two-Stage Bootstrap Approach

1a. Compute the conventional Z statistic

1b. Normalize the conventional Z statistic using $\hat{\sigma}_N$ = standard deviation of the t_i

2a. Mean-adjust the t_i statistics

2b. Construct 1000 bootstrap samples denoted $b = (1, \cdots, 1000)$



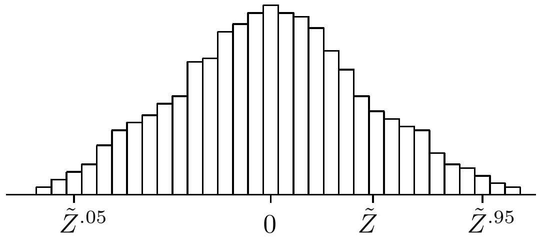
2c. Compute $\tilde{Z}_b$ for each sample using each sample's $\hat{\sigma}_{N,b}$:

$$\tilde{Z}_b = \frac{\sum_{j=1}^{N} t_{bj}^* / \sqrt{N}}{\hat{\sigma}_{N,b}}$$

2d. Build the empirical distribution from the 1000 values of $\tilde{Z}_b$ and conduct inference

$\tilde{Z}^{.05}$ = critical value defined by the 50^{th} largest value of the $\tilde{Z}_b$

$\tilde{Z}^{.95}$ = critical value defined by the 950^{th} largest value of the $\tilde{Z}_b$



If $\tilde{Z} < \tilde{Z}^{.05}$ or $\tilde{Z} > \tilde{Z}^{.95}$, then reject the two-tailed null hypothesis of no abnormal event effect at a 10% level of significance.