Banking on Event Studies: Statistical Problems, A Bootstrap Solution, and An Application to Failed-Bank Acquisitions

by

Lisa Andria Kramer
B.B.A. (Honours), Simon Fraser University, 1991

A thesis submitted in partial fulfillment of
the requirements for the degree of
doctor of philosophy

in

The Faculty of Graduate Studies
The Faculty of Commerce and Business Administration

We accept this thesis as conforming
to the required standard

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Commerce and
Faculty of Business Administration

The University of British Columbia
Vancouver, Canada

Date January 21, 1998

DE-6 (2/88)

Abstract

A variety of both parametric and nonparametric test statistics have been employed in the finance literature for the purpose of conducting hypothesis tests in event studies. This thesis begins by formally deriving the result that these statistics may not follow their conventionally assumed distribution in finite samples and in some cases even asymptotically. Thus, standard event study test statistics can exhibit a statistically significant bias to size in practice, a result which I document extensively. The bias typically arises due to commonly observed stock return traits, including non-normality, which violate basic assumptions underlying the event study test statistics. In this thesis, I develop an unbiased and powerful alternative: conventional test statistics are normalized in a straightforward manner, then their distribution is estimated using the bootstrap. This bootstrap approach allows researchers to conduct powerful and unbiased event study inference. I adopt the approach in an event study which makes use of a unique data set of failed-bank acquirers in the United States. By employing the bootstrap approach, instead of more conventional and potentially misleading event study techniques, I overturn the past finding of significant gains to failed-bank acquirers. This casts doubt on the common belief that the federal deposit insurance agency's failed-bank auction procedures over-subsidize the acquisition of failed banks.

Contents

List of Tables

List of Figures

Acknowledgements

While the Ph.D. thesis acknowledgement are traditionally intended to bring forth the names of those who helped make the completion of the research possible, I also wish pay tribute here to important people who have helped me deal with a much more difficult challenge: living day by day with a cancer diagnosis.

I cannot imagine having survived this long without the patience and loving support of my husband Mark Kamstra. He is my most important reason for fighting to live, and I am profoundly lucky to enjoy the unconditional love of this extraordinary man who is, among many other things, one of the world's most brilliant econometricians. (And I'm totally unbiased!)

This thesis would have been totally impossible without my mother Randi Kramer's assistance which began more than 29 years ago. I know how much she loves me, and I'm grateful for every moment we spend together. My father, Bela Kramer, also played a crucial role in the creation of this thesis, though he died the year before I commenced my Ph.D. studies. I know this thesis would have been a better contribution had he lived to make comments and help shape its evolution. I would like to think he would have been proud of me right now. I still think of my brother Trevor Kramer as my little curly-haired, blond, two year old shadow, but now that he is in his twenties, I should try to let go of that image! I thank him greatly for his continued love and support. Lots of other relatives have also provided me with strength and love, including the Michaelsens, Carol and George Kramer (and the rest of the Kramer gang), and the Johansens.

Many friends came to our rescue during difficult times since my cancer diagnosis, including Mary Kelly, Keith Freeland, Nathalie Moyen, Martin Boileau, Jim Storey, Louise Yako, Jan Kiss, Shannon Linegar, and many, many, many others. (I would need as many more pages as the length of this thesis to fully acknowledge all the people who have given something helpful.) I feel blessed to know such love and support.

I wish to acknowledge the medical experts, support workers, volunteers, and fellow cancer-fighters who have been helping me deal with Hodgkin's lymphoma. I am grateful for the contributions of tens of thousands of individuals who willingly suffered or died in clinical trials, leading to the development of viable forms of treatment for the cancer with which I live.

Specific medical doctors who have helped me live this long include Dr. Barbara Melosky, Dr. Tamara Shenkier, Dr. Janet Franiek, and Dr. Ken Evans. The chemotherapy nurses on the sixth floor of the BC Cancer agency are true angels – they include Libby, Iris, Elizabeth, Wilkie, Barb, Cathy, and many others whose names have faded into the (temporary!) chemotherapy-induced numbness of my brain. The amazing support staff of the BC Cancer Agency are actually skilled enough to make the cancer-surviving experience bearable. They include Lis Smith (who has since moved on to do important work elsewhere), Katherine Nicholson, Sarah Sample, Mary Jean Ellis, and others. The volunteers and participants of the BC Cancer Agency relaxation circle have helped me feel like an important member of a supportive community. The unconditional love in that group is profound. I have had the pleasure of getting to know many people in that community, and they have all touched me deeply. Naming each of them would, again, create a volume longer than this thesis, but I

1 Introduction

This thesis focuses on the implementation of financial event studies. In an event study, 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, indicative of a gain or loss to shareholders. Techniques have evolved considerably since the seminal event study of stock splits by Fama, Fisher, Jensen, and Roll [1969]. Today, the use of event study methods is quite mainstream, and event study results are a common source of "stylized facts" which influence policy decisions and the direction of research.

In this thesis, I consider the necessary conditions which underlie hypothesis tests in event studies. If conditions laid out in this thesis are satisfied, then most common event study test statistics asymptotically follow a standard normal distribution. However, it is typically the case that several of the underlying conditions are violated in practice, invalidating use of the distribution assumed to be appropriate for hypothesis testing. Unless tests are conducted with critical values from the *appropriate* distribution, erroneous conclusions may be reached.

Such issues have not gone without notice in the literature. Brown and Warner [1985, page 14], for example, state that when evaluating their event study test statistic in the presence of non-normally distributed data, "stated significance levels should not be taken literally" in some cases. Likewise, 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 recent studies have provided modifications to conventional techniques which successfully address *some* concerns that arise in practice.[1] Others have suggested nonparametric alternatives to conventional methods.[2] My study, however, aims to provide a feasible means of effectively dealing with *all* the problems under a much wider range of conditions than previously considered.

The main components of the thesis are as follows. In the remainder of this section, I present some conventional event study test statistics and discuss the conditions under which they follow their assumed distribution. I establish that violating these basic conditions

---

[1] For example, Boehmer, Musumeci, and Poulsen [1991] propose an alternative event study test statistic. 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.

[2] For example, Brown and Warner [1980, 1985] discuss the sign test, Corrado [1989] introduces the rank test, and Marais [1984] uses the bootstrap.

can render event study hypothesis tests invalid in practice. I also provide evidence that in practice these conditions are often violated, resulting in a significant bias to the statistical size of test statistics and invalidating inference. I argue that this bias has typically been under-estimated until now due to the use of actual CRSP data in conducting the statistical size tests. Therefore, generated data is used in this study, highlighting effects which can be obfuscated in studies using actual data. I document the bias in Section 3 where I provide results of extensive Monte Carlo experiments. Then, in Section 4, I discuss nonparametric alternatives to conventional event study methods, including the sign test, the rank test, and what I call the bootstrap approach. I argue that the only approach which achieves correct statistical size is the bootstrap approach. It involves *(a)* a normalization of conventional test statistics and *(b)* use of a bootstrap-based resampling exercise to empirically estimate the normalized test statistic's distribution. Inference based on this procedure has desirable size and power properties, even in situations where some of the conditions underlying conventional event study approaches are grossly violated. Thus, this bootstrap approach is robust to problems which plague conventional event study methods. I provide details for implementing the new technique in practice, and I document its impressive performance relative to other techniques. In Section 5, I conduct an actual event study which considers returns to the acquirers of failed banks in the United States. The typical finding in such event studies is that of abnormally large gains at the time of the acquisition. Based on this finding, many researchers have suggested the presence of a wealth transfer, arguing that failed-bank acquirers gain at the expense of federal regulators (hence taxpayers). Using the bootstrap approach, I find little evidence of significant abnormal returns, overturning the established result. Conclusions follow.

2 Event Study Methods

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 Patell [1976], Schipper and Thompson [1983], Dimson and Marsh [1985], and Brown and Warner [1980, 1985]) 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, the conventional approach is to collect a series of consecutive stock returns for a sample of firms of interest along with the corresponding returns on a market portfolio. A simple 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

2

at the time of the event. One can choose among many possible approaches in formulating test statistics to detect the information effect. Three of the most common event study test statistics are presented below, although I focus primarily on one – the Dummy Variable Approach – which is most frequently employed and most likely to behave well in practice (due to its relatively better flexibility).[3,4] I provide a detailed presentation of the test statistics and a discussion of how each of the statistics depends on certain key conditions. I argue that the test statistics do not follow their assumed distribution unless these necessary conditions are satisfied.

2.1 Some Common Event Study Methods

The Dummy Variable Approach

The first event study model I present is that which is most commonly adopted and most flexible in implementation. The methodology is clearly laid out by Thompson [1985], and the application is demonstrated by many, including Schipper and Thompson [1983], Eckbo [1985], Malatesta and Thompson [1985], Sefcik and Thompson [1986], and Eckbo [1992]. The Dummy Variable Approach involves estimating a market model by regressing returns for each of N firms being considered on the appropriate set of explanatory variables, including a dummy variable to pick up the event effect.

Define R_{it} as the return on firm i's share where $i = (1, \cdots, N)$, M_{it} as the return on the market portfolio, and ϵ_{it} as 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. (In addition to testing for single-day-event effects, it is also possible to test the significance of *cumulative* effects over *multiple* event days. Methods for doing so are discussed in Appendix A.) This dummy variable effectively allows for a change in the intercept at the time of the event, and hence can be thought of as picking up any unusual event effect in excess of the mean return. 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:[5]

---

[3]An earlier version of this thesis included the Cross-Sectional Approach and the Boehmer, Musumeci, and Poulsen [1991] Standardized Cross-Sectional Method. Preliminary experiment results indicated their performance is qualitatively similar to the performance of the approaches documented below. Since this thesis is meant to be a consideration of some of the most common techniques and not an exhaustive summary of all existing techniques, several methods have been excluded from the investigation in the interest of brevity.

[4]Characteristics of returns data may render one approach more suitable than others in a particular application. For example, a null hypothesis which includes a change in variance at the time of the event should be tested with a technique that accounts for such occurrences.

[5]The notation T_i would be used to allow for different length estimation periods across firms. In most

$$R_{it} = \beta_{i0} + \beta_{i1}M_{it} + \beta_{iD}D_{it} + \epsilon_{it}. \tag{1}$$

Note that I consider this relatively simple market model specification purely for the purposes of demonstration and investigation. By specifying models which intentionally neglect actual features of the data generating process (DGP), I am able to fully explore the implications of various degrees of model misspecification. In practice, the model chosen to estimate should allow for all suspected features of the data, such as autoregressive conditional heteroskedasticity in the disturbances or event-period changes in parameters other than the intercept. In practice, however, it is improbable that a researcher knows the true DGP underlying stock returns. Thus any model selected is likely to be misspecified to some extent. I explore the quantitative repercussions of such oversights, and in so doing I gain a better understanding of the robustness of the different approaches to various oversights. For example, while conventional test statistics are invalidated by neglecting to model even fairly modest (and therefore difficult to detect and easy to overlook) increases in event period variance, the bootstrap approach remains valid. If I were to restrict my study to "state of the art" models, this important distinction would be missed. In studies using test statistics to investigate the significance of a particular event – unlike the study reported in this section which aims to consider the robustness of test statistics themselves – one should consider only *sophisticated* models that attempt to capture all features of the data. (In fact, I consider more elaborate models in Section 5 where an actual event study is presented.) Relatively simplistic models are considered in this section to examine the effect of inadvertent and sometimes unavoidable oversights in model specification.

There is a t-statistic, denoted t_i, associated with each of the N estimated dummy variable coefficients $\hat{\beta}_{1D}$. These are used for testing the null hypothesis of no abnormal event day returns:

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

Typically, the Z^D statistic is assumed to be distributed as standard normal. However, Z^D follows this distribution only under certain conditions. I now present the three cases under which the standard normal distribution may apply. Implicitly underlying these cases is

---

cases, T_i is constant across firms, in which case the use of T is unambiguous.

4

the assumption that the market model shown in Equation (1) is well specified and that requirements of the classical linear regression model are satisfied. A fourth case is also presented – the one which applies to most event studies conducted in practice – for which Z^D cannot be expected to follow the standard normal distribution.

Case A: Small T, Large N

Consider the case where T, the length of the times series, may be short, but the number of firms being considered, N, is large. (Note that if T is *too small*, the market model may not be well specified.) In that case, Z^D will be asymptotically normally distributed, but Z^D will not necessarily be distributed as *standard normal*, even for very large N.

Consider the t_i statistics. If the t_i are identically distributed as Student t, then each t_i has zero mean and a variance of $\frac{T-k}{T-k-2}$, where k denotes the number of regressors in the market model ($k = 3$ for the model shown in Equation (1)). Calculating Z^D by summing the t_i and dividing by $\sqrt{N}$ yields an asymptotically normal random variable (its distributions denoted as $\mathcal{N}(0, \frac{T-k}{T-k-2})$) by central limit theory, provided the t_i are independent and identically Student t-distributed. While Z^D is asymptotically *normal* in this case, it is not *standard normal*. For asymptotic standard normality, the required divisor in calculating Z^D would be the square root of the sum of the variances for each t_i, that is $\sqrt{N\frac{T-k}{T-k-3}}$ instead of $\sqrt{N}$. I must emphasize that the Z^D statistic, as conventionally defined, does not follow its assumed standard normal distribution, *even asymptotically.*

Now if the t_i are not in fact identically distributed as Student t, one can still establish asymptotic normality, provided the t_i are independent. Asymptotic standard normality requires that the denominator of Z^D reflect the non-identical variances of the t_i, hence if each t_i had variance σ_i^2, then $\sqrt{\sum_{i=1}^{N} \sigma_{t_i}^2}$ would be used in the denominator of Equation (2) instead of $\sqrt{N}$. Once again, the Z^D statistic as conventionally defined will not follow its assumed standard normal distribution, *even asymptotically.*

To summarize the case of small T and large N, the following conditions are required for standard normality of Z^D.

(A1) The market model is well specified

(A2) The t_i are independent across firms

(A3) The denominator of Z^D is appropriately defined as the square root of the summed variances of the individual t_i

5

Case B: Large T, Small N

Consider the case of a long time series for each firm, but a small collection of firms. (Note that a time series which is *too long* may lead to an invalid market model specification unless changes in parameters are correctly modeled over time.) Disturbances that are independent and normally distributed with constant variance over time can be denoted

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I_T),$$

where σ^2 is the variance of the disturbances and I_T is a $T \times T$ identity matrix. If the disturbances satisfy this condition, then the t_i statistics are identically distributed as Student t. For large T, the distribution of the t_i asymptotically approaches the standard normal, assuming no individual observation among the regressors contributes excessively to the overall variance of the regressors.[6] Calculating Z^D according to Equation (2) involves summing the asymptotically normal independent t_i and dividing by $\sqrt{N}$. The sum of independent normals is normal, hence Z^D is itself approximately standard normal in this case.

The conditions required for approximate standard normality of Z^D for the case of large T and small N can be summarized as follows.

(B1) The market model is well specified

(B2) The t_i are independent across firms

(B3) The disturbances are independent and normally distributed with constant variance over time:
$$\epsilon \sim \mathcal{N}(0, \sigma^2 I_T)$$

(B4) The Lindeberg condition is satisfied, *i.e.* no observation among the regressors contributes excessively to the overall variance of the regressors

Case C: Large T, Large N

Consider the case where both the time series is long and the number of firms is large. (Once again, note that a time series which is *too long* may lead to an invalid market model specification unless changes in parameters are correctly modeled over time.) This case follows closely the argument presented for Case B above with one minor exception. Summing the

---

[6]This assumption is simply the Lindeberg condition. It ensures that the average contribution of the extreme tails to the variance of the regressors is zero in the limit. Consider for example, the market return, M_{it}. The Lindeberg Condition requires that $\max_t \dfrac{M_{it}^2}{\sum_{t=1}^{T} M_{it}} \to 0$ as $T \to \infty$.

6

asymptotically normal t_i and dividing by $\sqrt{N}$ will yield a Z^D statistic which more closely follows the standard normal: instead of an *approximately* standard normal Z^D, central limit theory leads to the standard normal *asymptotically* since N is large.

The conditions required for standard normality of Z^D for the case of large T and large N are as follows.

(C1) The market model is well specified

(C2) The t_i are asymptotically independent across firms

(C3) The disturbances are independent and normally distributed with constant variance over time:

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I_T)$$

(C4) The Lindeberg condition is satisfied, *i.e.* no observation among the regressors contributes excessively to the overall variance of the regressors

Case D: Small T, Small N

With both a short time series and a small sample size, Z^D cannot be expected to follow a standard normal distribution. Even if the t_i are independent and identically Student t distributed, summing them and dividing by $\sqrt{N}$ does not generally yield a statistic which follows the standard normal. Large N or large T plus a variety of other conditions are required for standard normality of Z^D, as shown in Cases A, B, and C above.

Notice that in none of the cases presented above was *finite sample* standard normality established. *Asymptotic* or *approximate* normality was established, but only under certain conditions. Nonetheless, it is conventional to assume that Z^D follows the standard normal distribution in practice – justified or not. In Section 2.2, I examine the degree to which properties of data used in event studies lead to violations of the various necessary conditions which underly standard normality, and in Section 3, I demonstrate the serious repercussions of violating the conditions in practice.

The Standardized Residual Approach

Another common event study approach is that of Patell [1976]. This procedure has been adopted in many studies, including investigations of corporate acquisitions by Bradley, Desai, and Kim [1988] and Gupta, LeCompte, and Misra [1993]. As before, R_{it} is the return on firm i's share where $i = (1, \cdots, N)$, M_{it} is the return on the market portfolio, and ϵ_{it}

is a disturbance term. The point in time at which the event announcement potentially impacts the firms' returns is denoted $t = +1$. For each firm in the sample, a market model is independently estimated for each firm over some period *prior* to the event, such as $t = (-130, \cdots, -11)$. Unlike the case of the Dummy Variable Approach, the event date is not included in the market model estimation period.

$$R_{it} = \beta_{i0} + \beta_{i1}M_{it} + \epsilon_{it} \tag{3}$$

For each firm, daily out-of-sample forecast errors are computed for the event period, for example $t = (-10, \cdots, +10)$, by making use of the market model estimates for β_{i0} and β_{i1} in conjunction with actual data for R_i and M_{it}:

$$\hat{\epsilon}_{it} = R_{it} - (\hat{\beta}_{i0} + \hat{\beta}_{i1}M_{it}). \tag{4}$$

These forecast errors are then adjusted to incorporate out-of-sample forecast variability. As I explain below, this yields a series of standardized residuals, $\tilde{e}_{it}$, for each of the N firms in the sample. Define the following variables: T_i is the number of observations for firm i in the market model estimation period,[7] $\hat{\sigma}^2_{\hat{\epsilon}_i} = \frac{\sum_{t=1}^{T_i} \hat{\epsilon}^2_{it}}{T_i - 2}$ is the variance of residuals from the market model regression for firm i, M_{it} is the market return for day t of the event period for firm i, M_{ik} is the market return for day k of the market model estimation period for firm i, and $\overline{M}_i$ is the mean market return over the market model estimation period for firm i.

Then firm i's estimated forecast variance, S^2_{it}, is given by the following expression:

$$S^2_{it} = \hat{\sigma}^2_{\hat{\epsilon}_i} \left[1 + \frac{1}{T_i} + \frac{(M_{it} - \overline{M}_i)^2}{\sum_{k=1}^{T_i}(M_{ik} - \overline{M}_i)^2} \right]. \tag{5}$$

The standardized residual for the each of the N firms in the sample is defined for each date $t = (-10, \cdots, +10)$ in the event period as the ratio of the forecast error to the square root of the estimated forecast variance:

$$\tilde{e}_{it} = \frac{\hat{\epsilon}_{it}}{\sqrt{S^2_{it}}}. \tag{6}$$

---

[7]Notice that T_i need not be the same across firms. However, it is typically the case that $T_i = T$ for all i.

The null hypothesis of no abnormal event effect at date $t = +1$ is tested by using Z^{SR}.[8]

$$Z^{SR} = \frac{\sum_{i=1}^{N} \tilde{e}_{i1}}{\sqrt{\sum_{i=1}^{N} \frac{T_i - 2}{T_i - 4}}} \qquad (7)$$

The distribution of the Z^{SR} statistic is conventionally assumed to be standard normal. However, Z^{SR} follows the standard normal only under conditions similar to those discussed for the Dummy Variable Model above. Z^{SR} may be *asymptotically* standard normal or *approximately* standard normal if T and/or N is large and conditions analogous to (A1) – (A3), (B1) – (B4), or (C1) – (C4) are satisfied. In practice, however, this is not typically the case.

The Traditional Approach

Consider now an event study approach like that outlined by Brown and Warner [1980, 1985]. Many have adopted this procedure, including Kalay and Loewenstein [1985] in a study of dividend announcements, James and Wier [1987] for the case of failed-bank acquisitions, and O'Hara and Shaw [1990] in a study of deposit insurance. As before, R_{it} is the return on firm i's share, M_{it} is the return on the market portfolio, and ϵ_{it} is an error term. The point in time at which the news of the event is thought to impact the firms' returns is denoted $t = +1$. As with the Standardized Residual Approach, the following market model is estimated for each firm over some period *prior* to the event, such as $t = (-130, \cdots, -11)$:

$$R_{it} = \beta_{i0} + \beta_{i1} M_{it} + \epsilon_{it}. \qquad (3)$$

The number of days in the estimation period for each firm is denoted by T; $T = 120$ in this example.[9]

The forecast errors are computed, making use of the market model estimates for β_{i0} and β_{i1} and actual data for R_i and M_{it}:

$$\hat{\epsilon}_{it} = R_{it} - (\hat{\beta}_{i0} + \hat{\beta}_{i1} M_{it}). \qquad (4)$$

---

[8]This test statistic pertains only to the case of testing for a single-day effect. Refer to Appendix A for the test statistic used to investigate the significance of cumulative returns over multiple event days.

[9]Note that this method requires each firm's estimation period to be the same length (*i.e.* $T_i = T$ for all i). It is unclear what the appropriate degrees of freedom would be for the test statistic in the case of different length estimation periods across firms.

The daily fitted residuals are averaged over firms for each day in the estimation period:

$$\bar{\epsilon}_t = \frac{1}{N} \sum_{i=1}^{N} \hat{\epsilon}_{it}.$$

Then the variance of the $\bar{\epsilon}_t$ over the estimation period is denoted $\hat{\sigma}_{\bar{\epsilon}}^2$. Under the null hypothesis of no abnormal returns at the time of the event, the expected value of $\bar{\epsilon}_t$ is zero at $t = +1$. This hypothesis is tested using a Z statistic for date $t = +1$ as follows:[10],[11]

$$
\begin{aligned}
Z^{TRAD} &= \frac{\frac{1}{N} \sum_{i=1}^{N} \hat{\epsilon}_{i1}}{\hat{\sigma}_{\bar{\epsilon}}} \\
&= \frac{\bar{\epsilon}_1}{\hat{\sigma}_{\bar{\epsilon}}}
\end{aligned}
\tag{8}
$$

where

$$\hat{\sigma}_{\bar{\epsilon}} = \sqrt{\frac{1}{T-1} \sum_{j=1}^{T} \left(\bar{\epsilon}_j - \frac{1}{T} \sum_{k=1}^{T} \bar{\epsilon}_k \right)^2}.$$

The Z^{TRAD} statistic is conventionally assumed to be distributed as Student t with $T - 1$ degrees of freedom. Unfortunately, Z^{TRAD} does *not* generally follow the Student t-distribution, even if conditions such as (A1) – (A3) or (B1) – (B4) shown for the Dummy Variable Approach are satisfied. A known distribution for Z^{TRAD} is obtained only if conditions like (C1) – (C4) are satisfied and the number of observations in the time series identically equals the number of firms being considered: $T = N$.[12] Under such circumstances, Z^{TRAD} is *asymptotically standard normal.* In practice, it is inappropriate to assume Z^{TRAD} generally follows the Student t-distribution – there is no finite sample case when this distribution applies.

2.2 Condition Violations

As discussed above, central limit theory can be used to establish asymptotic normality or approximate normality of the Z statistics under certain conditions. Thus, it may be tempting

---

[10]The variance in the denominator is calculated by some researchers, such as Brown and Warner [1980, 1985], over the estimation period – the variance of $\hat{\epsilon}_{it}$ over $t = (-130, \cdots, -11)$ in this case. Others, such as James and Wier [1987], calculate them over the forecasting period – the variance of $\hat{\epsilon}_{it}$ over $t = (-10, \cdots, +10)$ in this case. As a further complication, those who use the latter method sometimes exclude dates $t = 0$ and $t = +1$ from the calculation.

[11]See Appendix A for the case of testing the significance of cumulative returns over multiple event days.

[12]This result is most clearly highlighted by considering the fact that a Student t-ratio with $T-1$ degrees of freedom is defined by a standard normal random variable divided by the ratio of an independent chi-square random variable to its $T - 1$ degrees of freedom. In order for the components of Z^{TRAD} to satisfy the requirements for defining a Student t ratio, it is necessary that $T = N$.

to conclude that provided a large sample of firms or a long time series is employed for an event study, use of the standard normal distribution for the Z statistics is a reasonable approach to take. Unfortunately, it is not clear how large a sample is "large enough," hence use of the standard normal distribution may not be appropriate in *any* finite sample. Furthermore, the standard normal distribution may not apply *even asymptotically* if particular conditions are violated. For example, unless the Z^D statistic is calculated with the appropriate standard deviation in the denominator, in accordance with condition (A3), Z^D may not be standard normal. Finite sample consequences of violating the underlying conditions have not been formalized in the literature, and therefore it is not clear to what degree erroneous conclusions may be reached through use of the assumed approximate distribution (or the asymptotic distribution) when underlying conditions for a Z statistic are violated.

In the next section of this thesis, I quantify the extent to which condition violations invalidate inference based on use of the assumed distributions for the Z statistics, *even for fairly large samples*. Violations of factors other than independence are the main focus of investigation – that is, factors other than (A2), (B2), and (C2) – since independence is largely satisfied in studies of events which are not clustered in time. When independence is satisfied, I demonstrate that the distribution of test statistics can be estimated by adopting the bootstrap. Procedures for doing so are outlined in Section 4 below. A detailed consideration of the implications of violating independence (for studies of events which are clustered in time) is postponed for future study. For this purpose, the bootstrap with moving blocks re-sampling may prove to be useful.[13]

Many well-documented features of financial returns data suggest that the conditions underlying event study hypothesis testing are violated in practice. For example, returns data are known to be non-normally distributed, market model parameters have been found to undergo changes around the time of events, time-varying conditional heteroskedasticity is typically observed in returns data, and events of interest often occur at dates that coincide in time across firms. I discuss each of these in some detail below, with particular attention to explaining how they indicate violations of conditions (A1) – (A3), (B1) – (B4), and (C1) – (C4). In practice, these conditions are violated on many counts. Here I consider only a subset of the possible infractions.

---

[13]A moving block bootstrap is based on random re-samples of blocks of data (*i.e. dependently related strings of observations*) instead of re-samples of individual observations. Work including that of Liu and Singh [1992] adapts the bootstrap for use with weakly dependent data in this way. Such an approach requires an exact model for the form of the dependence, something which may be difficult or impossible to motivate across firms in an event study.

- *Non-normality*

There is considerable evidence that returns data are not normally distributed, violating conditions (B3) and (C3). 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 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."

The work of Brown and Warner [1980, 1985] is perhaps best known among studies which consider the performance of event study test statistics. Among the issues they consider is non-normality with respect to a *particular* Z statistic, Z^{TRAD} presented above. By simulating 250 individual event studies examining samples of 5 to 50 firms, they investigate the performance of Z^{TRAD} when returns data are not normally distributed. They argue that for samples of 50 firms, the test statistic is well-behaved. However, they indicate that the test statistic may *not* be well-behaved for smaller samples – an important result which is often overlooked. In this thesis, some aspects of Brown and Warner's work are extended.[14]

- *Autoregressive Conditional Heteroskedasticity* (ARCH)

Time-varying conditional heteroskedasticity is a well documented empirical regularity of stock returns data, as evidenced by the voluminous literature on the ARCH family (much of which is cited in the survey by Bollerslev, Chou, and Kroner [1992]). Neglecting the time-varying aspect of variance may lead to a violation of the requirement of

---

[14]First, computer generated data are used instead of actual CRSP data. The merits of doing so are fully discussed in Section 3.1 below. Second, in addition to the Z^{TRAD} statistic considered by Brown and Warner, the performance of other common test statistics is also evaluated. Under ideal conditions (for example with (C1) – (C4) satisfied and large $T = N$) it might be the case that none of the common event study test statistics deviate significantly from their assumed distributions, and hence, they may all lead to similar conclusions in hypothesis tests in that setting. However, in the case of violating the necessary underlying conditions, the various test statistics may actually behave differently from one another in practice. Thus, even if the Z^{TRAD} statistic performs well for some sample sizes in the presence of non-normally distributed data, other commonly utilized statistics like Z^D and Z^{SR} may *not* follow their assumed distributions under such conditions. By analyzing the behavior of several common Z statistics under a variety of conditions, fairly general conclusions can be reached. Finally, advances in computer technology facilitate a larger number of replications, permitting tighter confidence bounds and allowing for a rigorous performance test of the Z statistics.

constant variance implicit in conditions (B3) and (C3). The result will be a Z statistic which may not follow its assumed distribution, even for large samples. As discussed by Bollerslev, Chou, and Kroner, non-normality – also implicit in (B3) and (C3) – can arise by failing to account for ARCH in returns data.

- *Changes in variance around the time of the event*

  There is considerable evidence that the variance of the disturbances can undergo changes around the time of an event. For example, Boehmer, Musumeci, and Poulsen [1991] and Donaldson and Hathaway [1994] recognize the importance of modeling changes in the variance during an event period. For the firms they consider, they find variances can rise up to 680% around the time of events. Failure to model such changes violates the model specification conditions – (A1), (B1), and (C1) – and also violates the constant variance requirement in conditions (B3) and (C3). Effects such as these can lead to Z statistics which do not follow the standard normal distribution, even asymptotically.

  Furthermore, estimates of the variance used in the various Z statistics embed different assumptions about the behavior of the disturbances. This may allow for different conclusions to be reached with the use of different statistics, even when employing the same data. Consider an example which highlights the impact of variations on commonly employed Z statistics. If returns become more volatile during the time surrounding the event, then the version of Z^{TRAD} used by Brown and Warner [1980, 1985], which uses *estimation period variance* in the denominator, may be considerably larger than a version of Z^{TRAD} which employs the *event-period variance* in the denominator. (See footnote 10 above for a discussion of the distinction between these two variants.) Then clearly it is possible to reach different conclusions, depending on which of the many available Z statistics is adopted.

- *Changes in market model coefficients during the event period*

  Several researchers have established that market model parameters can undergo changes around the time of the event or follow some distribution or some pattern over time. Donaldson and Hathaway [1994] demonstrate the importance of allowing for changes in the intercept and market return coefficient at the time of the event. 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), and Brockett, Chen, and Garven [1994] argue for the importance of allowing a time-varying stochastic market return coefficient. Failure to

13

model such effects violates the condition that the market model be well-specified –
(A1), (B1), and (C1) – resulting in Z statistics which do not follow their assumed
distribution, even asymptotically.

For example, suppose that for some firms in the sample the true coefficient on the
market return is equal to one during $t = (-130, \cdots, -11)$, then equal to two dur-
ing the event period, $t = (-10, \cdots, +10)$. For the Traditional Approach and the
Standardized Residual Approach, the market model in Equation (3) is estimated over
$t = (-130, \cdots, -11)$, the period during which time the true coefficient is one for all
the firms. Then the estimated coefficients are used to forecast into the event period,
$t = (-10, \cdots, +10)$, during which the true coefficients have actually doubled for some
firms. Failing to account for the increase in the true coefficient will invalidate inference
based on the Z statistics.

- *Clustering*

Clustering arises whenever events occur simultaneously in *calendar* time for the firms
in a sample. This type of cross-sectional dependence arises in many studies, including
those 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.
Such studies often find the residuals are correlated across firms, perhaps due to the
omission of some market-wide explanatory variable. This correlation across firms would
suggest a violation of independence – conditions (A2), (B2), and (C2). A study of the
utility of the bootstrap in this setting is postponed for future consideration. Various
past studies have already conducted some investigation of the impact of cross-sectional
dependence on the performance of conventional event study test statistics. Brown and
Warner [1985] document the fact that test statistics often do not follow their assumed
distribution in the presence of clustering. However, they find that adjusting for the
dependence across firms can introduce a substantial loss of power.

3 Demonstrating Significant Bias

3.1 Experiment Design

It is clear from Section 2.1 that the small sample distribution of common event study statis-
tics relies critically on basic underlying conditions. It is also clear from Section 2.2 that these
conditions are violated in practice. To date, the validity of standard event study techniques
under circumstances commonly encountered has not been fully explored. Hence, I have con-
ducted extensive experiments to investigate further the performance of the Z statistics in

practice. The objective is to quantify the degree to which use of critical values from a conventionally assumed distribution can lead a researcher to false conclusions when conditions underlying a Z statistic are violated. As briefly mentioned in the previous section, past researchers have typically employed *actual returns data* for investigating the performance of test statistics. In this study, instead of using actual returns data, I adopt *generated* data with properties to match closely the characteristics of actual data. Simulated data such as these have been used extensively in the event study literature; see Acharya [1993] and Prabhala [1997] for recent examples.

There are several reasons to prefer the use of generated data for evaluating the statistical size of test statistics. First, because actual returns data incorporate true event effects (some of which may even be unknown to the researcher), and because the timing and magnitude of some of these event effects often cannot be determined by the researcher, it is not advisable to employ actual returns for conducting size investigations. Investigations of size must be done under the null hypothesis, in the absence of both known and unknown event effects. Hence, I employ data generated without event effects, permitting true investigations of statistical size.

Second, if the conditions which underlie valid hypothesis tests are violated on several counts by the characteristics of actual data, there is a potential for the violations to *appear* to counteract one another in a given instance, leading to no *perceivable* evidence of significant bias. Thus, in using actual returns data, one might conclude a test statistic is valid even though it may not be valid in general. For example, positive skewness in returns would tend to lead to overstated test statistics, while a decrease in event-period variance could lead to understated test statistics. If these two factors closely offset one another in a specific set of CRSP data, one might fail to detect what might in fact be a general problem. Observing that counteracting characteristics in the data seem to "cancel each other out" in some settings – *i.e.* finding what seems to be a lack of bias for some particular sample of actual CRSP data – is not sufficient evidence for the general validity of event study test statistics.

If one were to attempt to use actual returns data to evaluate statistical size, a variety of the condition-violating factors would be simultaneously present. One would have to decompose the data and extract each of the factors individually in order to examine marginal effects of particular factors in isolation. Such a decomposition would be immensely complex and unreliable without prior knowledge of the true DGP. By employing generated data in this study, I can easily investigate both the *marginal and collective* impacts of factors which are each individually capable of invalidating event study hypothesis tests, without undertaking complicated and unreliable extractions.

Third, sensitivity analysis is facilitated by the use of simulated data – for example, the

marginal impact of various factors can be examined by simply changing the nature of ascribed properties. This would not be feasible with actual data.

In summary, although many previous Monte Carlo studies on event study test statistics have employed actual CRSP data, correctly specified size and power experiments require the use of *generated* data. Consequently, much of the analysis in this study is based on data generated with properties to match those of actual data.

In my Monte Carlo experiments, I know the true DGP, including the distribution of stock returns, the degree and form of ARCH, the complete set of relevant explanatory variables, *etc.* By intentionally specifying models which ignore true aspects of the DGP, I can quantitatively determine the performance of event studies test statistics in circumstances where their underlying conditions are violated, mimicking the type of analysis undertaken in actual event studies. The Monte Carlo framework of this study allows the relative performance of the various Z statistics to be investigated under a *variety* of realistic conditions.

For example, to evaluate the performance of the Z statistics for a sample of 100 firms, I took the following steps:

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

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

The properties ascribed to the generated data closely matched those of actual financial returns data, leading to violations of a variety of the conditions (A1) – (A3), (B1) – (B4), and (C1) – (C4). In Appendix B, I document the collection of studies I consulted in choosing how to generate the data for the experiments. I also discuss the steps taken to verify others' reported parameter values, and I outline the algorithm chosen to generate the data.

Although the data were generated to violate conditions underlying test statistics, the null hypothesis is *true* to the extent that there is *no event effect* to be detected. I am interested in the statistical size of the Z statistics: the probability of rejecting the null hypothesis of no event-day abnormal return when it is true. It may also be of interest how the statistics behave when there *is* an event effect. In this case, the power of the statistics becomes of interest: the probability of rejecting the null when it is false. Questions of power are examined in Section 4 where the performance of the conventional Z statistics is compared to that of the bootstrap approach.

2. OLS was used to estimate each firm's market model, and Z^D, Z^{SR}, and Z^{TRAD} were computed.

16

3. Steps 1 and 2 were repeated a total of 1000 times in this set of experiments. Each of these 1000 replications can be thought of as a single event study, each generating the three conventional Z statistics, Z^D, Z^{SR}, and Z^{TRAD}. By simulating many event studies, fairly strong statistical statements can be made.

4. The statistical size of each type of Z statistic was then evaluated. First, actual rejection rates were computed at various levels of significance by dividing the number of times the null hypothesis was rejected at a given significance level by the total number of replications. Then each actual rejection rate was compared to the assumed ("nominal") size for the test. 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. Similarly, for a test conducted at the $\alpha = 5\%$ level of significance, the null hypothesis would be rejected 5% of the time if there is no bias to size, and hence the actual rejection rate for the 5% test should be compared to 0.05. As explained in Appendix C, one can construct standard confidence intervals 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, fairly strong conclusions can be reached because the confidence intervals around the nominal size values for each statistic are quite small.

Results for different sample sizes can be explored by adjusting the number of firms for which data are generated. Data can be generated to allow the study of the effect of actual observed returns properties simultaneously, in subsets, or in isolation. That is, I can examine the marginal impact of one observed feature alone, or I can examine the total impact of many features incorporated in data together. Furthermore, the impact of more or less serious degrees of condition violations can be investigated simply by adjusting the properties assigned to the generated data. In fact, I have conducted several sets of experiments in this study, making use of data generated to reflect various violating characteristics together and in isolation.

The main set of experiments, results of which are presented immediately below in Section 3.2, considers the case of returns data which are skewed and leptokurtic and which undergo changes in the DGP during the event period. All of these factors were simultaneously incorporated in the data used for this set of experiments, and the particular parameter values (*e.g.* the exact degree of excess kurtosis) were chosen to mimic what is observed in actual returns. In Appendix D, I report results of experiments where each of the factors was studied in isolation and to various degrees, once again with properties assigned to match

17

those of actual returns data. Collectively, the results of Section 3.2 and Appendix D establish that *common event study test statistics exhibit a large and significant bias to size.* The bias arises due to characteristics displayed by actual returns data – characteristics which violate necessary conditions like those presented in Section 2.1 above. Furthermore, the bias is maintained even when the condition violations are considerably less severe than is empirically observed. On account of this bias, one cannot generally rely upon inference based on the conventional Z statistics in practice. Therefore, in Section 4, I present alternative nonparametric test statistics used in event studies. I show the existing nonparametric methods to be prone to some of the same problems as the Z statistics. However, the nonparametric bootstrap approach I introduce exhibits no bias to size while still maintaining impressive power properties. Due to its good performance, the bootstrap approach can be adopted as a reliable alternative to conventional methods.

3.2 Results

Recall that actual returns data display properties that violate key conditions underlying conventional event study Z statistics. We know that, *in theory*, violating these assumptions can lead to Z statistics which do not follow their assumed distributions. The fact that this is the case *in practice* is now demonstrated: standard Z statistics exhibit significant bias to size, invalidating inference.

Figures 1 – 3 illustrate results for the case of 1000 replications for a sample of 100 firms, with data generated with properties to match actual returns, and with no event effects present (*i.e.* the null hypothesis is true). Properties were chosen by consulting several empirical studies, as documented in Appendix B. The skewness coefficient was 0.15, the kurtosis coefficient was 6.2, the coefficient on M_{it} rose 100% during the event period, and the event-period variance rose by 500%. Keep in mind that the null hypothesis is true – there are no event effects in the data.

Consider Figure 1, the case of Z^D. Along the horizontal axis is the significance level of the hypothesis tests: the conventional range of testing levels, $\alpha = (0.01, \cdots, 0.10)$, is presented. Along the vertical axis is the actual rejection rate for hypothesis tests. The solid line in the graph denotes the nominal size of tests: for a test at the $\alpha = 1\%$ level one should reject 1% of the time, for a test at the $\alpha = 5\%$ level one should reject 5% of the time, *etc.* The dotted lines surrounding this solid line represent the upper and lower 95% confidence bounds around the nominal size, calculated in the standard manner as described in Appendix C. The black dots represent the actual rejection rates for tests conducted at particular significance levels, appearing at intervals of 0.01 on the horizontal axis. For example, for the test at the $\alpha = 5\%$ level, the dot lies close to 0.20, considerably above the upper confidence bound, indicating

18

the actual size differs significantly from the nominal size. When one conducts a test at the 5% level, the rejection rate should be about 0.05 under the null hypothesis. However, what we observe in this figure is an actual rejection rate considerably larger than what is expected. Since the rejection rate lies outside the plotted 95% confidence interval, we conclude the test statistic is significantly biased. The dots for tests at all conventional significance levels are quite clearly outside the 95% confidence bounds – in no cases are the rejection rates within the range expected under the null. The implication is that with violations of the underlying conditions, the actual size of Z^D is significantly biased. In using this statistic under these conditions – conditions which match those observed in practice – the elevated rejection rates might lead one to conclude at a high level of significance that an event effect is present when in fact there may be none. Figures 2 and 3 present similar results for Z^{SR} and Z^{TRAD}. At all conventional testing levels, the rejection rates greatly exceed what would be expected under the null, implying the statistical size of these test statistics is also significantly biased. In fact, they seem to be even more biased than Z^D – rejection rates for Z^{SR} and Z^{TRAD} seem to be greater than those observed for Z^D.

For comparison, Figures 4 – 6 present results for Z statistics computed when the data are well behaved, without skewness, fat tails, changes in event-period variance, or shifts in market model parameters around the time of the event. Notice that for all the testing significance levels, the rejection rates are within a 95% confidence interval. Evidently, the Z statistics are unbiased under these conditions. This result may be somewhat surprising, given the fact that the Z statistics are expected to follow their assumed distributions only *asymptotically*. These experiments establish that the asymptotic distributions apply – even for fairly small samples – provided the data are fairly well behaved.

While Figures 1 – 6 provide a diagrammatic representation of the performance of actual rejection rates relative to nominal size, Tables 1 and 2 contain actual statistics for various sample sizes. Consider first Table 1 which contains rejection rates for skewed and leptokurtic data incorporating shifts in both the market return coefficient and the variance around the event time. The first column, labeled "Size", denotes the significance level of tests conducted. Since it corresponds to the horizontal axes in Figures 1 – 6, the values range increasingly from 0.01 to 0.10. The first set of three columns pertain to Z^D, Z^{SR}, and Z^{TRAD} for the case of 30 firms. The remaining sets of three columns contain results for samples of 50, 100, and 200 firms respectively. The top value listed in each cell is the actual rejection rate for the Z statistic at the particular significance level. The bottom value in each cell is the right-tail probability value (henceforth denoted "p-value") associated with the Z statistic at the particular significance level. The p-value indicates the probability of obtaining a rejection rate *greater than that actually observed*, and it is calculated using the conventionally assumed

19

distribution for a given Z statistic (standard normal for Z^D and Z^{SR}; Student t for Z^{TRAD}).

The statistics which correspond directly to Figure 1 (*i.e.* Z^D for the case of 100 firms with non-normal data undergoing DGP changes) appear in the Table 1 column labeled "100 Firms" and "Z^D". The values in the columns labeled "Z^{SR}" and "Z^{TRAD}" under the "100 Firms" heading correspond to Figures 2 and 3. To evaluate the statistical size of the 5% test for the Z^D statistic, for example, the relevant information is found along the row labeled "0.05." For this statistic, the actual rejection rate for the 5% test is 0.187, a value which is certainly qualitatively larger than the rate of 0.05 expected under the null. More importantly from a statistical point of view, the p-value associated with such a rejection rate is indistinguishable from zero at three decimal places. If a given p-value is below 0.025 or above 0.975, then the associated rejection rate lies outside of the 95% confidence interval shown in the Figure 1.[15] Thus, the rejection rate for a 5% test for Z^D with a sample of 100 firms exceeds the upper cut-off value shown in Figure 1. The rejection rates for the remaining tests at $\alpha = (0.01, \cdots, 0.10)$ are also outside any reasonable confidence interval, indicative of a significant bias to the size of the Z^D statistic.

Inspection of Table 1 reveals that significant over-rejection is also observed for all three of the Z statistics with samples of 30 or 50 firms (as might be expected given the over-rejection with 100 firms). For tests at *any* common significance level, the rejection rates are anywhere from 100% to 2000% higher than they are expected to be under the null. In fact, even with a fairly large sample of 200 firms, the degree of over-rejection is qualitatively unchanged. That is, the significant bias documented in this thesis is not necessarily eliminated by gathering data for more firms. As argued in Section 2 above, there are cases where the conventional Z statistics follow a *non-standard normal* distribution asymptotically – *i.e.* the asymptotic distribution may be normal, but the variance may not be unity *even in the limit.*

Consider now Table 2 which reports statistics for the case where data are normally distributed with no changes in DGP, as shown in Figures 4 – 6. The layout of this table is identical to that of Table 1. Notice that the rejection rates for all of the statistics with normally distributed data are not qualitatively different from what would be expected under the null. Furthermore, all the p-values are within the 95% confidence interval. Thus, with well-behaved data, the Z statistics are all unbiased, even in small samples.

Discussion

The experiment results presented above and in Appendix D provide striking evidence that valid inference cannot be conducted by applying standard event study methods to

---

[15]Notice that in considering the p-values in Table 1, one is not restricted to a 95% confidence interval around the nominal size. For example, p-values below 0.005 or above 0.995 would indicate an actual rejection rate which lies outside a 99% confidence interval.

data which exhibit commonly observed characteristics. Significant over-rejection takes place under a wide range of circumstances, even when very large samples are employed. This over-rejection can be explained as follows.

First, when the data exhibit excess kurtosis, the normality conditions embedded in the null hypothesis are violated – conditions (B3) and (C3). This is indicative of disturbances which are distributed with greater mass in both the upper and lower tails than would be observed under the null. The incidence of positive skew places even greater mass in the upper tail. A Z statistic based on such data is of course drawn from a distribution which has a fatter right tail than would be observed under the null hypothesis. Comparing such a Z statistic to the conventional critical value (a critical value which implicitly assumes there is no skew or excess kurtosis), often results in rejection of the null hypothesis. If the Z statistic were instead compared to a critical value from the *appropriate* distribution, rejection of the null hypothesis would not be indicated as frequently.

Furthermore, when the data incorporate an increase in the market return coefficient during the event period, but the market model does not account for this factor, conditions underlying the conventionally assumed distribution are violated – in particular the well specified model conditions are violated – (A1), (B1), and (C1). The event-period disturbance in such a model incorporates the part of the market return which should have been picked up by a larger coefficient. Because market returns are themselves skewed and leptokurtic, the result is a Z statistic drawn from a distribution with increased right-tail mass relative to the standard normal. As explained above, comparing such a Z statistic to the conventional critical value can lead to an overstated rejection of the null hypothesis.

Finally, when there is an unmodeled rise in variance during the event period, the conventionally assumed distribution can be further invalidated, due to a violation of the conditions that the market model is well specified – conditions (A1), (B1), and (C1) – and that the variance is constant – conditions (B3) and (C3). Failure to model a change in variance makes it more likely that the null hypothesis will be rejected. That is, an event day disturbance drawn from a distribution with higher variance is more likely to result in a Z statistic which exceeds a critical value based on the assumption of constant variance.

All of these effects individually imply over-rejection. When they are combined, the tendency to over-reject is quite dramatic, as shown in Figures 1 – 3. Experiments documented in Appendix D investigate these factors individually to demonstrate that the statistical size of Z statistics can be significantly biased even in very simple cases.

The experiments have focused on a few specific categories of violations of the conditions underlying event studies. Other types of violations, such as those involving ARCH or the omission of relevant explanatory variables, may also be of interest. However, they have not

21

been examined directly in this study. Note that the experiments have also focused on factors which tended to elicit a *positive* bias to size, particularly at low significance levels. That is, excess kurtosis, positive skew, event-period variance increases, and increases in event-period coefficients all tend to lead to *over-rejection* of the null hypothesis. By introducing counteracting effects which *individually* lead to biases of different sign, for example negative skewness and a decrease in event-period variance simultaneously incorporated in the data with excess kurtosis, the *overall* sign of the bias to size can become ambiguous. However it is important to keep in mind that there is no reason to believe the effects will "cancel each other out". That is, ambiguity in the sign of the bias does *not* imply that conclusions based on event study hypothesis tests are reliable. Inability to determine the bias direction would just imply that a researcher would not know whether he is over-rejecting, under-rejecting, or rejecting appropriately.

It is important to emphasize that many of the problems documented above are not addressed by simply increasing the size of the sample. When the essential underlying conditions are all satisfied, the Z statistics should be *asymptotically* standard normal in distribution. The approximation tends to be reasonable for a variety of sample sizes when the data are well behaved. However, it is not typically the case that all the necessary conditions are satisfied when we do an event study. In some cases, depending on the conditions violated, central limit theory may establish the *asymptotic normality* of the Z statistics. However the statistics may not be *asymptotically standard normal*. The conventional test statistics are incorrectly defined with an inappropriate standard deviation in the denominator in some contexts. Thus, even if one considered samples as large as 10 000 or 100 000, large and significant size biases would still be maintained. Hence assuming the standard normal distribution applies for any particular sample size – even a very large sample size – can easily lead to invalid conclusions.

4   Nonparametric Event Study Approaches

The aim of the above Monte Carlo analysis has been to demonstrate that common event study Z statistics often fail to follow their assumed distributions in practice. In light of this fact, there is clearly a need for an alternative, generally unbiased procedure. Since the conditions underlying the conventional *parametric* tests are often violated in practice, it is conceivable that *nonparametric* testing procedures might perform comparatively better. In this section I discuss existing nonparametric tests used in event studies, and I also propose a new approach which I argue has better statistical properties than any existing parametric or nonparametric approach.

22

The new method, which I call the bootstrap approach, is quite simple to adopt. It involves *(a)* a normalization of a conventional Z test statistic and *(b)* use of a bootstrap-based resampling exercise to empirically estimate the normalized test statistic's distribution.[16] Later in this section, detailed steps are laid out for implementing the new procedure in event studies. Its unbiased size is documented, and its power is shown to be comparable to that of standard techniques.

4.1 Existing Nonparametric Methods

Two nonparametric tests which have been used occasionally in event studies are the rank test, introduced by Corrado [1989], and the sign test. Both of these tests require that a market model such as that shown in Equation (3) above is estimated for each firm, over a period which precedes the time of the event such as $t = (-130, \ldots, -11)$. Abnormal returns are forecasted into the event period, making use of the market model estimates, yielding $\hat{\epsilon}_{it}$ for an event period such as $t = (-10, \ldots, +10)$, as shown in Equation (4) above. A single time series of abnormal returns is then formed for each firm by combining the non-event-period residuals, $\hat{\epsilon}_{it}$ for $t = (-130, \ldots, -11)$, with the forecasted event-period abnormal returns, $\hat{\epsilon}_{it}$ for $t = (-10, \ldots, +10)$. The total number of non-event-period observations (T) plus event-period observations is denoted T'.

Rank Test

For each firm, the time series of event-period and non-event-period abnormal returns is ranked by magnitude, from smallest to largest. This yields the times series of ranks, denoted K_{it}, for each firm, where $t = (-130, \ldots, +10)$.

$$K_{it} = rank(\hat{\epsilon}_{it}) \tag{9}$$

The following standard deviation is calculated using the K_{it} time series for of all the firms collectively.

$$S(K) = \sqrt{\frac{1}{T'} \sum_{t=-130}^{+10} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{K_{it}}{1 + T'} - \frac{1}{2} \right) \right)^2} \tag{10}$$

---

[16]Recent research on event studies, such as that of Kothari and Warner [1995], indicates that conventional tests on *long-horizon* (multiyear, for example) abnormal returns may demonstrate significant bias. In order to avoid bias, Ikenberry, Lakonishok, and Vermaelen [1995] conduct inference in a long-horizon event study using an approach based on the bootstrap. The Ikenberry *et. al.* method is quite different from the bootstrap approach I propose in this thesis, and it may exhibit biased size due to the fact that actual CRSP data for a sample of matched firms is used to build an empirical distribution, rather than using the sample data directly. (See the discussion in Section 3.1 regarding problems which may plague the use of actual data.)

Then the event-day rank test statistic is defined as follows:

$$T^{RANK} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left(\frac{K_{i1}}{1+T'} - \frac{1}{2} \right) / S(K) \tag{11}$$

The rank test statistic T^{RANK} is typically assumed to follow the standard normal distribution. In practice, however, this may not be the case.[17] Underlying the rank test is the presumption that each observation in a particular firm's time series of abnormal returns follows the identical distribution as the other abnormal returns for that firm. In Section 2.2 I discussed many well-documented features of stock return data that can violate this requirement, including ARCH, event-period changes in variance, and event-period changes in model coefficients. Asymptotic results imply that the rank test will be well-defined even in the presence of heterogeneities such as these. For finite samples, however, it is unknown to what extent the test statistic might deviate from its assumed standard normal distribution.

To explore the extent to which the rank test might differ from its assumed distribution, I conducted experiments similar in nature to those presented for the Z statistics in Section 3.1 above. Table 3 presents the results for 1000 replications with various sample sizes. Data were generated with properties identical to those used for the Z test experiments shown in Table 1 above, with properties chosen to match those of actual returns and with no abnormal event-period returns present (i.e. the null hypothesis of no event effect is true). The skewness coefficient was 0.15, the kurtosis coefficient was 6.2, the coefficient on M_{it} rose 100% during the event period, and the event-period variance rose by 500%. While the rank test demonstrates less bias than the parametric Z statistics under these conditions, significant bias is nonetheless observed. Hence, one cannot reasonably expect the rank test to follow its asymptotic standard normal distribution in practice. For example, examining the first row of Table 3 reveals that for a test at the 5% level with a sample of 30 firms, the rank test rejects 10.9% of the time. (For comparison, with the same sample size and with data displaying similar characteristics, the 5% Z statistics rejected 18.3% to 23.7% of the time, as shown in Table 1.)

---

[17]Studies which have found the rank test to be well-behaved in event studies, such as that by Corrado [1989] and Corrado and Zivney [1992], have used actual CRSP data for conducting statistical size and power experiments. While it is intuitive that the rank test should behave better than its parametric counterparts, the Monte Carlo results of these studies may not be definitive. As discussed in Section 3.1 above, the use of CRSP data can mask the true biases of a test statistic. Hence it is worthwhile to reconsider the rank test in an study that uses generated data and avoids the aforementioned problems inherent in the use of CRSP data.

It is important to observe that the rank test remains significantly biased even for considerably larger sample sizes. In fact, for the data characteristics used in this example, the rank test demonstrates bias until a sample size of 1000 is adopted; 1000 firms is the smallest sample for which the majority of the rejection rates presented are acceptably close to what is expected under the null hypothesis. With this large sample size, rank tests at the 1% level do not reject significantly differently from 1% of the time, rank tests at the 5% level do not reject significantly differently from 5% of the time, and so on. It is very important to keep in mind that data with more extreme characteristics (greater skew, kurtosis, or event-period changes, for example, all of which are commonly exhibited by actual returns data) would require considerably larger sample sizes to ensure the rank test has proper statistical size. Because in practice one does not know to what extent the model is mis-specified, one cannot say with any degree of certainty what sample size is required for the rank test to have proper statistical size in a particular application. The fact that the rank test is significantly biased for even fairly large samples with data displaying fairly modest violating characteristics is cause for concern.

Sign Test

The sign test employed in some event studies is based on the number of firms which have a positive event-day abnormal return, denoted by N^+. The total number of firms is denoted by N.

$$T^{SIGN} = \frac{N^+ - (N/2)}{\sqrt{N/4}} \qquad (12)$$

The sign test statistic is typically assumed to follow a standard normal distribution. Unfortunately, this distribution will apply only in the case of symmetrically distributed returns, and stock returns are known to be skewed to the right. Some researchers have defined sign tests which are claimed to be reliable for asymmetric distributions. While such statistics may be reliable for the case of skewness, they will be invalidated by other factors that also invalidate Z tests and the rank test (such as changes in event-period model parameters).

I conducted experiments to investigate the performance of the sign test under conditions identical to those for which the rank test experiments were conducted. Results are presented in Table 4. In accordance with previous experiments, I conducted 1000 replications for various sample sizes, and the data were generated with properties chosen to match those of actual returns without the additional presence of abnormal event-period returns. Comparing the figures shown in Table 4 for the sign test with figures shown in Table 3 for the rank test indicates that there is little qualitative difference in the performance of these test statistics.

Like the rank test, the sign test shows significant bias for various sample sizes. For example, for a sign test at a 5% level of significance with a sample of 30 firms, the sign test rejects 7.8% of the time (which is significantly outside any reasonable confidence interval). A study by Corrado and Zivney [1992] suggested that under some conditions, sign tests can behave more poorly than the rank test. Results in Tables 3 and 4 reveal some cases where the performance of the sign test is better than that of the rank test and some cases where it is worse. However, given that the sign test can show significant bias (even for very large sample sizes), it may be unreliable for general use. In fact, use of the sign test on data with more extreme characteristics than adopted for this particular exercise could reasonably be expected to lead to even greater bias than shown here.

4.2   The Bootstrap Approach for Event Studies

As an alternative to existing parametric and nonparametric methods, one can adopt a technique based on bootstrap resampling. Marais [1984] uses bootstrap p-values to conduct inference in conjunction with the Standardized Residual Approach. In this thesis, I demonstrate how the bootstrap can be used in conjunction with *any* of the conventional event study methods, and I rigorously demonstrate the improved size properties of the bootstrap approach relative to other methods.

Unlike the bootstrap approach used by Marais, there are two components to the bootstrap approach I outline. The first of these requires the *normalization* of conventional test statistics – a new innovation in the use of the bootstrap. Because the violation of underlying conditions can lead to event study test statistics with variances different from unity (even asymptotically), it is inappropriate to compare conventional test statistics to their assumed distribution. With the appropriate normalization, however, it is possible to obtain test statistics with unit variance. This normalization is quite straightforward, involving nothing more than dividing the conventional test statistic by an appropriate standard deviation – the standard deviation of the N t-statistics for the case of the Dummy Variable Approach.[18] The second component of my bootstrap procedure makes use of the empirical distribution of the normalized test statistic. This involves repeatedly sampling from the actual data in order to empirically estimate the true distribution.

The bootstrap was initially introduced by Efron [1979] as a robust procedure for estimating the distribution of independent and identically distributed data. Since its inception, the

---

[18]In principle one should normalize the Z^D statistic by the *exact* value required to impose unit variance. If the t_i were identically distributed as Student t, this value would be $\sqrt{\frac{T-k}{T-k-2}}$, the standard deviation of a Student t-statistic with $T - k$ degrees of freedom. Because the t_i are unlikely to be exactly Student t-distributed in practice, the normalization I propose is based on the sample standard deviation of the t_i statistics – the best approximation available.

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 essentially similar means (but not necessarily identical variances) and provided the first two moments are bounded, use of the bootstrap is valid.[19] Several recent books on the subject provide a good overview of the bootstrap, including Hall [1992], LePage and Billard [1992], Efron and Tibshirani [1993], and Hjorth [1994].

In recent years, the bootstrap has come into common use for empirical work in many fields. A small sub-set of the many recent applications includes 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, Li and Maddala's [1996] survey of developments in using bootstrap methods for time series models, Mooney's [1996] study of political science applications using the bootstrap, and Bullock's [1995] test of the efficient redistribution hypothesis using the bootstrap.

Implementation

Consider conducting an event study on a sample of N firms. One option would be to follow the conventional Dummy Variable Approach, introduced in Section 2. This would involve estimating the market model in Equation (1) for each firm, yielding N individual Student t-statistics (denoted t_i for $i = (1, \cdots, N)$). Standard practice is to compute $Z^D = \sum_{i=1}^{N} t_i / \sqrt{N}$ and to compare the resulting value to a critical value from the standard normal – a distribution which may not in fact be appropriate. Inference based on the bootstrap requires instead that *(a)* Z^D is normalized by the standard deviation of the t_i, and then *(b)* the empirical distribution is estimated by repeatedly sampling from the original data.

In Figure 10, I provide a diagrammatic representation of the bootstrap approach for conducting inference in event studies. Each of the six steps shown is also discussed in detail below. While the procedure makes use of the conventional Z^D statistic which emerges from the Dummy Variable Approach, the steps can be modified in a straightforward manner to enable inference based on any common event study test statistic. In Appendix E, I explain

---

[19]Technically, the conditions are as stated in Liu's Theorem 1. Let $X_i, \cdots, X_N$ be independent random observations. The mean and variance of X_i are denoted μ_i and σ_i^2 for $i = (1, \cdots, N)$. If *(i)* $\lim_{N \to \infty}(1/N) \sum_{i=1}^{N} (\mu_i - \bar{\mu}_N)^2 = 0$, *(ii)* $\liminf_{N \to \infty} \left((1/N) \sum_{i=1}^{N} \sigma_i^2 \right) > 0$, and *(iii)* $E|X_i|^{2+\delta} \leq K < \infty$ for some $\delta > 0$ and for all i, then $\lim_{N \to \infty} \| P^* \left(\sqrt{N}(\bar{Y}_N - \bar{X}_N) \leq x \right) - P \left(\sqrt{N}(\bar{X}_N - \bar{\mu}_N) \leq x \right) \|_\infty = 0$ *a.s.*, where P^* stands for the bootstrap probability, $\bar{Y}_N$ is the mean of a bootstrap sample $Y_1, \cdots, Y_N$ drawn with iid samples from the empirical distribution based on $X_1, \cdots, X_N$, and $\| \ \|_\infty$ stands for the sup-norm over x. That is, given the above conditions, the bootstrap consistently estimates the distribution of interest.

the modifications required for bootstrap inference based on the Traditional Approach and the Standardized Residual Approach.

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

$$R_{it} = \beta_{i0} + \beta_{i1}M_{it} + \beta_{iD}D_{it} + \epsilon_{it}. \tag{1}$$

The estimation yields N t-statistics: one for each firm's estimated dummy variable coefficient. As shown in Figure 10, this collection of t-statistics forms the pool of data upon which the conventional Z^D statistic is based.

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

A researcher interested in conducting conventional inference would stop at this point and compare the value of Z^D to a critical value from the assumed standard normal distribution. As indicated earlier, such inference may not be valid.

2. Normalize the Z^D statistic obtained in Step 1 to account for the fact that its variance differs from unity in practice. First, compute the standard deviation of the actual t_i, denoted $\hat{\sigma}_t$. Then, divide Z^D by $\hat{\sigma}_t$ to yield the normalized version of Z^D which will be used to conduct event study inference with the aid of the bootstrap. Denote the normalized test statistic with a tilde:

$$
\begin{aligned}
\tilde{Z}^D &= \frac{\frac{\sum_{i=1}^{N} t_i}{N}\sqrt{N}}{\hat{\sigma}_t} \\
&= \frac{\sum_{i=1}^{N} t_i/\sqrt{N}}{\hat{\sigma}_t} \\
&= \frac{Z^D}{\hat{\sigma}_t}
\end{aligned}
\tag{13}
$$

Clearly, the normalized test statistic $\tilde{Z}^D$ is just the original Z^D statistic divided by the standard deviation of the individual firms' t-statistics. In the remaining steps, the empirical distribution of $\tilde{Z}^D$ will be constructed, facilitating reliable inference.

3. Under the null hypothesis, the distribution of $\tilde{Z}^D$ is centered about zero, and hence the empirical distribution is constructed such that it is also centered about zero. Notice that the N actual t-statistics calculated in Step 1 have a mean, $\bar{t} = \sum_{i=1}^{N} t_i/N$, 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 $\tilde{Z}^D$. 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:

$$t_i^* = t_i - \bar{t}. \tag{14}$$

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 in a straightforward manner.

4. The mean-adjusted data are used to construct an empirical distribution for $\tilde{Z}^D$. This involves drawing many random samples, called "bootstrap samples," from the population of t_i^* statistics. As shown in Figure 10, 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.[20] 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; the third sample might contain two occurrences of some statistics, three occurrences of some other statistics, four occurrences of a few statistics, and no occurrences of the remaining statistics. The make-up of each of the 1000 constructed bootstrap samples would be determined purely by random chance.

5. As shown in Figure 10, a $\tilde{Z}_b^D$ statistic is computed for each of the 1000 bootstrap samples. (The $b = (1, \cdots, 1000)$ subscript is used to specify the particular bootstrap

---

[20]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 above 1000 leads to no marked change in results.

sample.) Unit variance is imposed by a straightforward normalization which employs the standard deviation of the N mean-adjusted t-statistics in each bootstrap sample, denoted $\hat{\sigma}_{t_b^*}$. (The b subscript is used to specify the bootstrap sample – because the particular set of t^* statistics varies by bootstrap sample, the value of $\hat{\sigma}_{t_b^*}$ also varies by bootstrap sample. Note that below, I label the t^* statistics in a particular bootstrap with the subscript b and with the subscript $j = (1, \cdots, N)$ to distinguish between particular observations within a given bootstrap sample.)

A normalized statistic, $\tilde{Z}_b^D$, is calculated for each bootstrap sample as follows:

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

$$
= \frac{\sum_{j=1}^N t_{bj}^*/\sqrt{N}}{\hat{\sigma}_{t_b^*}}
$$

$$
= \frac{Z_b^D}{\hat{\sigma}_{t_b^*}}. \tag{15}
$$

That is, for each of the bootstrap samples, the mean of the N randomly selected t_{bj}^* is multiplied by $\sqrt{N}$ and divided by $\hat{\sigma}_{t_b^*}$, the standard deviation of the N individual t_{bj}^* in the bth bootstrap sample. Notice that computing $\tilde{Z}_b^D$ is equivalent to computing Z_b^D based on the t_{bj}^* in that bootstrap sample and then dividing by $\hat{\sigma}_{t_b^*}$.

6. Ordering the collection of 1000 $\tilde{Z}_b^D$ statistics from smallest to largest defines the empirical distribution. The histogram depicted at the bottom of Figure 10 is an example of such an empirical distribution. Inference is conducted by comparing the $\tilde{Z}^D$ statistic from Step 2 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 $\tilde{Z}_b^D$ statistics and a 5% right-tail critical value, $\tilde{Z}^{.95}$, is the 950th largest of the $\tilde{Z}_b^D$ statistics. If the value of the $\tilde{Z}^D$ statistic happens to be larger than 95% of the bootstrap $\tilde{Z}_b^D$ statistics (i.e. exceeding $\tilde{Z}^{.95}$) or smaller than 5% of the bootstrap $\tilde{Z}_b^D$ 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.[21]

---

[21]This type of confidence interval for the bootstrap is the *percentile interval*, and I apply it using normalized data. Alternative methods of constructing confidence intervals exist, including the *percentile interval* using non-normalized data, the *standard normal interval*, the *bias-corrected and accelerated interval*, and the *approximate bootstrap confidence interval*. Consult Efron and Tibshirani [1993] for a detailed exposition of these alternatives. In this particular application, the basic percentile interval using non-normalized data was found to perform inadequately relative to the percentile interval with normalized data. The bias-corrected

To summarize, applying Steps 1 - 6 of the bootstrap approach based on the Dummy Variable Approach basically involves computing the conventional Z^D using the actual t_i statistics and normalizing it with the variance of the t_i to impose unit variance. This yields $\tilde{Z}^D$. 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 $\tilde{Z}_b^D$ is computed for each bootstrap sample. The collection of all the $\tilde{Z}_b^D$ statistics defines the empirical distribution. Finally, event study inference is conducted by comparing $\tilde{Z}^D$ to critical values from the empirical distribution.[22] [23]

4.3 Performance of the Bootstrap Approach

In this section, I report results of Monte Carlo experiments conducted to compare commonly used event study techniques with procedures based on the bootstrap approach when data exhibit properties matching those of real data. Recall that with existing parametric and nonparametric methods, the statistical size of test statistics was shown to demonstrate significant bias. When the alternative approach is adopted, the bias is eliminated. Fortunately, the elimination of the bias comes at little or no expense. Investigations of statistical power indicate that the normalization and bootstrap approach has power comparable to conventional approaches. Furthermore, 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.

Size

Independent data were generated to match properties of actual financial returns data, with studies documented in Section 2.2 guiding my choice of parameter values. Adopting the steps outlined in Section 4.2 above, 1000 replications were conducted for a variety of sample sizes. Data were generated with skewness of 0.15, kurtosis of 6.2, an increase in event period variance of 500% and an increase in the event period market return coefficient of 100%. Z^D, Z^{SR}, and Z^{TRAD} were computed for each replication, and the statistics

---

and accelerated interval and the approximate bootstrap confidence interval both lead to significantly increased demands on computer resources. In light of the fact that the percentile interval already performs quite well in this setting, alternative intervals are not explored in this study.

[22]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 bootstrap may not be advisable.

[23]The bootstrap procedure has no advantage over conventional methods when applied to a *non-pivotal* test statistic. When applied to a *pivotal* test statistic, the bootstrap achieves second-order asymptotic efficiency like that achieved with the use of Edgeworth expansions. (Edgeworth expansions, however, require substantially more analytic effort, and in fact the analytics of such expansions are sometimes intractible).

were normalized (yielding $\tilde{Z}^D$, $\tilde{Z}^{SR}$, and $\tilde{Z}^{TRAD}$) and compared to critical values from their empirical distributions. I drew 1000 bootstrap samples for each replication, and I employed the percentile interval for conducting inference.

Figures 7 – 9 illustrate the results for the case of 50 firms. Rejection rates for $\tilde{Z}^D$ are shown in Figure 7, while rejection rates for $\tilde{Z}^{SR}$ and $\tilde{Z}^{TRAD}$ appear in Figures 8 and 9 respectively. The striking result to draw from these figures is that all of the rejection rates lie within a 95% confidence interval. Even though the underlying data display characteristics which grossly invalidate *conventional* event study techniques (as shown in Figures 1, 2, and 3), the statistical size of the bootstrap approach shows no significant bias.

The values of the rejection rates and the right-tail p-values which correspond to the figures appear in Table 5 along with results for other sample sizes. The rejection rates lie within a 95% confidence interval in almost every cell.[24] Thus, the bootstrap approach exhibits unbiased size, even for very small samples. Extensive sensitivity analysis (similar to that reported for conventional techniques in Appendix D) indicates that the excellent performance of the technique is robust even to *perturbations* in the properties ascribed to the generated data. The bootstrap approach maintains its unbiased size when applied to data displaying any commonly observed properties.

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. With many replications, the overall rejection rate should approach unity for a powerful test. However, evaluating the power of a *biased* test statistic is uninformative – obviously a test statistic which tends to over-reject will seem more powerful than one which is unbiased. Thus, prior to computing the rejection rates in the presence of a positive abnormal return, all of the event study test statistics must be *size-adjusted*. These adjustments have little impact on the power experiments for the bootstrap approach, since the approach is, of course, unbiased. However, the size-adjustments are absolutely critical for evaluating the power of conventional, biased approaches. Appendix F provides further details on size-adjustments for the power comparisons.

In this case, 1000 replications are conducted for computing the rejection rates for the test statistics. For each replication, identical data are used to evaluate the conventional statistics and the normalized statistics. Hence, differences in rejection rates can arise only randomly or due to differences in the computation of the statistics and the critical values

---

[24]Three of the 120 statistics are outside a 95% confidence interval, however since this is not statistically unusual, there is no indication of biased size.

used to evaluate them. The main result is that inference based on the bootstrap approach is just as powerful as that based on standard methods.

Table 6 presents results for the case of 50 firms, 1000 replications, 1000 bootstrap samples, and data with characteristics which match actual data (skewness of 0.15, kurtosis of 6.2, an event period variance increase of 500%, and an increase in the event period market return coefficient of 100%). An abnormal return of a given magnitude is added to the event day errors in order to facilitate a comparison of power. Abnormal returns of 0.5%, 0.7%, and 0.9% are considered here. Rejection rates of 100% are observed for both conventional techniques and the bootstrap methods with abnormal returns of 1% and greater.

The first set of three columns reports the rejection rates based on comparing Z^D, Z^{SR}, and Z^{TRAD} to their commonly assumed distributions. The next set of three columns reports rejection rates based on comparing $\tilde{Z}^D$, $\tilde{Z}^{SR}$, and $\tilde{Z}^{TRAD}$ to their bootstrap distributions. 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 6, all 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 both conventional inference and the inference based on the bootstrap approach remain very close to 100%. With abnormal returns of 0.5%, shown in the bottom panel, both the conventional rejection rates and the rejection rates for the bootstrap approach fall slightly, but the performance is qualitatively similar across methods. The overall conclusion to draw is that power under the bootstrap approach is quite comparable to that of conventional methods. Rejection rates are almost identical for all cases considered. Similar results are obtained under various conditions.

Discussion

Results of the experiments conducted in this study indicate that the bootstrap approach is very robust to features observed in actual data which invalidate the use of standard event study methods. At least two reasons can be cited to explain the good results. Both reasons allude to use of the appropriate distribution when conducting inference with a given test statistic.

First, recall that the bootstrap approach involves a normalization of the conventional Z statistics – dividing by the standard deviation of the t_i statistics for the case of Z^D. (See Step 2 in Figure 10.) The distribution of the conventional Z^D statistic is assumed to be standard normal; in practice, however, Z^D may have a variance which differs from one, even asymptotically. Normalizing Z^D by the standard deviation of the t_i addresses that problem by imposing unit variance. This is one factor which helps explain the performance of my

new method.

The second reason for the improved performance of this technique relative to conventional approaches has to do with the fact that it involves estimating the distribution of the $\tilde{Z}$ statistic *empirically* rather than imposing some assumed distribution which may apply only asymptotically. In so doing, the actual firms' data are employed. To the extent that the sample data accurately reflect the population, the empirical distribution constructed will be a good approximation of the true distribution followed by $\tilde{Z}$, even in finite samples.

One might wonder whether alternative event study approaches more simple than mine might also successfully address problems in standard event study methods. For example, one reasonable strategy might be simply to adopt the first part of my proposed method, comparing the *normalized Z* statistics to the conventionally assumed distributions. Perhaps adjusting the Z statistic to have unit variance would be sufficient to enable valid inference, thereby making use of the bootstrap unnecessary.

Table 7 provides details on such an experiment for a variety of sample sizes. Experiments were conducted for samples of 30, 50, 100, and 200 firms with data incorporating characteristics to match those of actual returns. The figures in the table reveal that while the simpler approach does reduce the magnitude of the bias relative to the standard approaches, it does not reduce the bias nearly as much as both normalizing and employing the bootstrap. That is, the test statistics under such an approach still exhibit significant bias.

For example, consider the dummy variable approach for a sample of 50 firms. When the conventional Z^D was compared to the assumed standard normal distribution for the case of 50 firms (see Table 1), a rejection rate of 12.3% was observed although a rejection rate of 1% was expected under the null hypothesis. To evaluate the performance of the simpler approach under similar conditions, refer to the 50 firms column of Table 7: comparing the *normalized $\tilde{Z}^D$* to the standard normal distribution leads to a rejection rate of 6.0% when a rejection rate of 1% is expected under the null hypothesis. For comparison, when the normalized $\tilde{Z}^D$ was compared to its bootstrap distribution (instead of the standard normal) a rejection rate of 1.4% was observed when 1% was expected under the null (see Table 5). Note that the first two rejection rates indicate significant bias, while the final one does not. That is, the two-step bootstrap approach is the only one which successfully eliminates significant bias.

Glancing at the rest of Table 7 reveals that the simple approach shows significant bias under all cases considered. The magnitude of the bias appears to diminish for larger sample sizes, but there is no quantitative change in the p-values which reflect the significance of the bias. P-values indistinguishable from zero at three decimal places are observed uniformly throughout the table.

For samples with *very* large numbers of firms – perhaps in the tens of thousands or the

hundreds of thousands – one may wonder whether calculating $\tilde{Z}^D$ and comparing it to the standard normal distribution might be valid, possibly by appealing to Case A outlined in Section 2.1 above. Normalizing Z^D works toward satisfying condition (A3), that the test statistic is defined with the appropriate denominator. Considering non-synchronous events will avoid a violation of (A2), that the firms' t-statistics be cross-sectionally independent. However, whether or not the market model is satisfied – condition (A1) – is difficult to ascertain in practice. Something as seemingly innocuous as a change in model parameters over the span of firms' data sets would be sufficient to violate this condition. Furthermore, whether or not the number of firms in the sample is "large enough" to appeal to Case A is also unknown in practice. In light of the fact that some of the conditions underlying standard normality for Case A may not be met, it is *not advisable* to adopt the strategy of comparing the normalized Z^D to the standard normal distribution rather than estimating its distribution empirically. One faces significant bias in using either conventional test statistics or normalized test statistics with their conventionally assumed distributions.

Another alternative one might consider as a potentially useful simpler approach might take the conventional Z statistics and compare them to distributions constructed using the bootstrap procedure (without conducting any normalizations for non-unit variance). Results from these experiments are shown in Table 8. Once again, experiments were conducted for samples of 30, 50, 100, and 200 firms using data generated with characteristics to match those observed in actual returns. While this second simple approach reduces biases more markedly than the previously discussed simple approach, it still does not reduce the bias nearly as much as both normalizing and employing the bootstrap. That is, the test statistics under this simple approach still exhibit significant bias in most cases. For example, consulting Table 8 reveals that with a sample of 50 firms with a test conducted at the 1% level, the rejection rate for the conventionally computed Z^D statistic is 2.6% when compared to a bootstrap distribution. This appears qualitatively less biased than the rejection rate for comparing the conventional Z^D to a standard normal distribution (12.3%) and for comparing the normalized $\tilde{Z}^D$ to a standard normal distribution (6.0%). However the 2.6% rejection rate for this approach is still significantly biased, with a p-value indistinguishable from zero at three decimal places. Table 8 reveals similar results for other test statistics and other sample sizes. There are some cells in Table 8 where the rejection rates are within a 95% confidence interval around the rejection rate expected under the null hypothesis. However, in some cases the rejection rate is *above* what would be expected under the null, and in other cases it is *below*. There does not appear to be any systematic pattern to these isolated cases. The important result in this table is that severe and significant bias is observed for most cases. While the rejection rates under this approach are certainly closer to the values

35

expected under the null hypothesis than for some other approaches, the best performance yet observed comes from the use of both normalization and the bootstrap.

In general, I advocate the approach based jointly upon normalizing the conventional Z^D statistic and estimating its distribution empirically. The outstanding performance of this approach applied to data exhibiting a wide range of the characteristics of actual returns is testament to its robustness. Valid and powerful inference is facilitated, even in situations where conventional methods fail. In the next section, I demonstrate by example that established financial event study results can be overturned when the bootstrap approach is employed.

5   An Application: Failed-Bank Acquisitions

When a bank fails in the United States, the Federal Deposit Insurance Corporation (FDIC) aims to find the most cost effective resolution possible, while still fulfilling its obligations as a deposit insurer.[25] One means of resolution available to the FDIC is to pay off insured deposits and liquidate the bank. Alternatively, the FDIC may arrange for an auction, known as a purchase and assumption, to find a new financial institution to assume operations of the failed bank. In the event of an auction, no public announcement is made, in the interest of minimizing the potential for panic among investors. Instead, potential bidders are privately contacted by the FDIC and invited to submit bids.[26] Because a failed bank's assets are insufficient to cover its liabilities, there are often substantial subsidies from the FDIC in order to make the takeover worthwhile for the acquirer. The winning bid is selected not only on the basis of bid magnitude, but also by considering characteristics of the individual bidders which might influence the successful operation of the failed bank and which may have effects on competition or soundness of the banking industry. Thus, it is not always the highest bid that wins. In light of this fact, some researchers have speculated that regulators over-compensate in the failed-bank acquisition process.

The consensus of past *empirical* research on failed-bank acquisitions is indeed that banking regulators provide more financial assistance to acquiring firms than is strictly required. Several researchers have made this suggestion, including Pettway and Trifts [1985], James

---

[25]In the past, the Federal Savings and Loans Insurance Corporation (FSLIC) and the FDIC jointly insured depositors' holding at US financial institutions: FSLIC insured deposits at thrifts while the FDIC insured holdings at commercial banks. Under legislation introduced in 1989, FSLIC was essentially dismantled, with the FDIC assuming their former responsibilities.

[26]The invited banks are selected on the basis of meeting certain requirements: they must be rated as low-risk banks, they must be at least double the size of the failed bank in terms of total assets, and they must comply with state and federal bank acquisition laws and a few additional FDIC requirements. Also, it has often been a requirement that the invited banks operate in the same county as the failed bank.

and Wier [1987], Bertin, Ghazanfari, and Torabzadeh [1989], Balbirer, Jud, and Lindahl [1992], and Gupta, LeCompte, and Misra [1993]. Common event study methods such as those presented in Section 2 above are adopted in these studies, and the nearly unanimous finding is that failed-bank acquirers earn significantly positive abnormal returns. Therefore, these researchers conclude that wealth transfers routinely occur from regulators to firms that purchase failed banks.[27] [28]

The theoretical and empirical evidence I presented in previous sections suggests conventional event study test statistics are significantly biased, hence concluding a wealth transfer routinely takes place in failed-bank acquisitions may be premature. I therefore re-examine the the question of gains to failed-bank acquirers employing the bootstrap approach. In so doing, I assemble one of the largest collections of data on failed-bank acquirers yet investigated.

5.1 Data

Each year, the *FDIC Annual Report* lists all FDIC-insured banks which have failed. In the case of an auction, the name of the successful bidder (the acquiring firm) is also provided. In order for a failure to be included in my sample, the acquirer must have publicly traded shares. Sometimes, the acquirer listed in the *FDIC Annual Report* is a subsidiary of a larger company, for which shares may or may not be traded. By making use of *Moody's Banking and Finance* manual and *The S&P Security Owner's Stock Guide,* I was able to trace the ownership of the successful bidder firms and isolate those which were themselves publicly traded or which were wholly-owned subsidiaries of other publicly traded banks. Upon finding a publicly traded acquirer, I ensured that the firm's shares were traded for at least 130 days prior to and 10 days following the date of the bank failure, in order to facilitate estimation of the market model for each firm.[29]

---

[27]Those who claim to find evidence of the gains to acquirers offer several explanations to support their findings. Some authors point to evidence of gains to acquirers in takeovers of *non-financial* firms as motivation for the existence of gains to firms that acquire failed banks. For example, Dodd and Ruback [1977], Bradley [1980], and Bradley, Desai, and Kim [1988] have all found empirical evidence that acquiring firms benefit in the event of tender offers. Other researchers refer to legislation in the US designed to reduce the total monetary outlay of federal banking regulators. The Financial Institutions Reform, Recovery, and Enforcement Act (FIRREA) of 1989 and the Federal Deposit Insurance Corporation Improvement Act (FDICIA) of 1991 are two such pieces of legislation. Others still make reference to theoretical models that suggest features of the regulators' auction procedures result in gains to acquirers of failed banks. See the models of Johnson [1979] and French and McCormick [1984], for example.

[28]An investigation by Cochran, Rose, and Fraser [1995] seems to be the only one which does not support the wealth transfer view. Their findings suggest that positive abnormal returns tend to be concentrated among acquirers of large failed banks and that buyers of small failed banks do not realize significantly positive abnormal returns.

[29]Individual share returns and market returns data were obtained from the CRSP *Daily Returns File.*

Past failed-bank takeover studies have typically excluded acquisitions for which the date of the auction announcement could not be found in *The Wall Street Journal Index*. This severely limits the number of firms considered in a failed bank study since a large portion of failed bank auctions are never announced in *The Wall Street Journal* (WSJ). The WSJ tends to announce details related to the affairs of large banks, resulting in a greater likelihood that the results of an auction for a small failed bank will go unreported and thus be excluded from a sample. In fact, with the increased incidence of bank failures over the 1980's, the WSJ adopted an explicit policy of announcing failures of only those banks for which the asset base exceeded a minimum threshold. Hence, a sample that only includes acquisitions announced in the WSJ will be limited in size.

By including all publicly traded acquirers of failed banks in my sample regardless of whether or not announcement in the WSJ took place, and by obtaining the date of the acquisition and the date of the announcement from the original official news releases obtained directly from the FDIC, I am able to construct a much larger sample than other studies of failed banks have considered, making use of a larger proportion of the total population of failed-bank acquirers.[30]

During the period from mid-1989 to the end of 1991, a total of 347 bank failures required FDIC assistance, 208 of which were resolved by auction. Of these, 45 bank closures met my sample selection requirements listed above. As in many event studies, failure to meet the public trading requirement is the factor which most severely restricted the sample size. However, my use of FDIC news releases enabled the inclusion of acquirers which would normally be excluded from a failed-bank event study. Thus, my sample includes a much higher proportion of the population of auctions than has been conventional.[31]

The sample of 45 failed-bank acquirers used in this study is listed in Table 9. The first column, labeled "Firm", lists the number assigned to each acquisition in the sample, based on the chronological order of the failed-banks' closures. The second column lists the date of the closure, while the name and state of the failed bank are listed in the third column. The ticker symbol of the acquirer (or the bank that wholly owns the acquirer) is listed in

---

[30]Note that even in the absence of obtaining the actual press releases, it would still be possible to reasonably approximate the date of the acquisition by using the failed bank closure date information available in the *FDIC Annual Report*. It is *usually* the case that the acquisition takes place within a few days of the bank closure; in fact, the acquisition *almost always* takes place the business day immediately following the closure. Brown and Warner [1985] suggest greater power in event studies is achieved by knowing the event date *with certainty*. For my sample, obtaining the FDIC press releases ensures the event date is known with certainty.

[31]For example, while James and Wier [1987] assemble 19 acquirers over 11 years (representing about 13% of the population of auctions during that time), my sample consists of 45 acquirers over two and a half years (representing about 22% of the population). Any attempts to further increase the population of acquirers would be hindered by the fact that most of the acquiring firms are not publicly traded.

the final column.[32] For each acquirer, the date of the failed bank closure reported in the *FDIC Annual Report* is taken as $t = 0$ in event time. The FDIC news release is made after markets close on that date, hence $t = +1$ is the time when announcement is expected to have an impact on the acquirer's share price.

5.2 Analysis of Gains to Acquirers

In this section I present evidence that use of the bootstrap event study procedure overturns past findings of gains to failed-bank acquirers. In order to motivate the fact that my results do not arise due to idiosyncratic features of my data relative to data used by others, I discuss results of both conventional event study approaches and the bootstrap approach. As explained below, I explored a variety of sensible specifications for the market model.

I first estimated a market model independently for each of the 45 firms in my sample. One of the market models adopted is the most basic Dummy Variable model discussed extensively above:

$$R_{it} = \beta_{i0} + \beta_{i1}M_{it} + \beta_{iD}D_{it} + \epsilon_{it}. \tag{1}$$

Then I collected the 45 t_i statistics, the test statistics on the individual firms' estimated dummy variable coefficients $\hat{\beta}_{iD}$. Using these statistics, I calculated the conventional Z^D statistic:

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

A researcher adopting conventional event study methods would stop here and compare the value of Z^D to the assumed standard normal distribution to determine whether or not the failed-bank acquirers realized significant gains. Following this approach requires that the

---

[32]Note that in Table 9 multiple entries appear for some of the acquiring firms. Multiple acquisitions by a single firm may be included in the sample provided the timing of one acquisition does not fall in the market model estimation period for another acquisition by the same firm. In cases where a firm made several failed bank acquisitions, I ensured that the acquisitions included in the sample were sufficiently separated in time to ensure the estimation periods for the events did not overlap. In order to avoid estimation complications, I excluded from my sample all cases in which the estimation periods overlapped for a firm undertaking more than one FDIC-assisted acquisition – unless the acquisitions took place on the same calendar date (which I treated as a single large acquisition). Four multiple-acquirers remain in the sample of 45 firms, each having successfully acquired either two or three failed banks at sufficiently spaced dates during the years 1989 to 1991.

necessary conditions for Cases A, B, or C be satisfied by the data – conditions (A1) – (A3), (B1) – (B4), and (C1) – (C4) presented in Section 2.

For this particular application, Cases A and C are ruled out since the sample size is small. If the length of the time series, $T = 130$, is considered small, then Case B is also ruled out. That would suggest Case D applies: Z^D would not be expected to follow the standard normal distribution and conventional analysis would be invalidated.

If one were to presume $T = 130$ is sufficiently large to invoke asymptotic theory, then conditions (B1) – (B4) must be satisfied to enable valid inference. Specification tests (for example, Lagrange Multiplier tests for normality and for ARCH) using the residuals from estimating each firm's market model provided statistically significant evidence that these necessary conditions are in fact violated here. In light of the presence of a variety of factors discussed in Section 2.2, the conditions underlying conventional analysis are unlikely to be satisfied in this application.

Thus, I adopted my alternative procedure, following the steps shown in Figure 10. This procedure does not require the data to satisfy conditions which are in fact violated by the data. The first step, calculation of Z^D has already been completed. Thus, I proceeded to compute the value of $\tilde{Z}^D$, dividing the conventional Z^D statistic by $\hat{\sigma}_t$, the standard deviation of the original t_i statistics:

$$
\begin{aligned}
\tilde{Z}^D &= \frac{\frac{\sum_{i=1}^{N} t_i}{N}\sqrt{N}}{\hat{\sigma}_t} \\
&= \frac{\sum_{i=1}^{N} t_i/\sqrt{N}}{\hat{\sigma}_t} \\
&= \frac{Z^D}{\hat{\sigma}_t}.
\end{aligned}
$$

The value of $\tilde{Z}^D$ differs from Z^D to the extent that $\hat{\sigma}_t$ differs from unity. In the present application, the value of $\hat{\sigma}_t$ is 1.61.

Next, I mean-adjusted the N individual t_i statistics obtained from estimating the market model. These N mean-adjusted statistics, denoted t_i^*, were obtained by deducting the mean $\bar{t}$ from each of the t_i. This imposed the null hypothesis of no event-day gains to acquirers and ensured the empirical distribution would be centered about zero. I proceeded to draw 1000 bootstrap samples from this collection of mean-adjusted data. For each bootstrap sample, I randomly drew with replacement 45 times from the collection of mean-adjusted statistics. The t_i^* statistic associated with a particular firm might have appeared in a given bootstrap sample many times, once, or not at all, as determined by random chance. The 45 mean-adjusted statistics drawn for a particular bootstrap sample are denoted t_{bj}^*, with

$b = (1, \cdots, 1000)$ designating the bootstrap sample with which they are associated. The standard deviation of the 45 values of t^*_{bj} for a given bootstrap sample is denoted $\hat{\sigma}_{t^*_b}$.

For each bootstrap sample, I then computed a $\tilde{Z}^D_b$ statistic:

$$
\begin{aligned}
\tilde{Z}^D_b &= \frac{\frac{\sum_{j=1}^N t^*_{bj}}{N} \sqrt{N}}{\hat{\sigma}_{t^*_b}} \\
&= \frac{\sum_{j=1}^N t^*_{bj}/\sqrt{N}}{\hat{\sigma}_{t^*_b}} \\
&= \frac{Z^D_b}{\hat{\sigma}_{t^*_b}}.
\end{aligned}
$$

Ordering the collection of all B of the $\tilde{Z}^D_b$ statistics from smallest to largest then defined the empirical distribution used to conduct inference. The p-value of a particular $\tilde{Z}^D$ statistic was given by the area of the empirical distribution located to its right, i.e., the proportion of the 1000 $\tilde{Z}^D_b$ statistics greater than $\tilde{Z}^D$.

Much recent work suggests that more parameterized models lead to more reliable inference in event studies. For example, Brockett, Chen, and Garven [1994] allow for stochastic betas and account for GARCH, and they claim this removes bias in conventional event study tests. Adopting a more sophisticated model that attempts to account for known aspects of the DGP is less likely to violate the necessary conditions which underlie the conventional methods, improving the likelihood that inference will be valid. Because the Dummy Variable market model specified in Equation (1) is rather simplistic, I investigated a variety of enhancements to explore differences in results obtained using commonly employed methods versus the bootstrap approach. For example, I explored inclusion of dummy variables to pick up changes in the market return coefficient around the time of the event or anticipation effects prior to the event. I also investigated factors such as allowing for variance changes around the event time and accounting for ARCH.

For example, when the basic Dummy Variable of Equation (1) was adopted, the conventional Z^D statistic indicated strong rejection of the null hypothesis of no abnormal gains to acquirers – the p-value based on the assumed standard normal was 0.007.[33] In contrast, comparing the $\tilde{Z}^D$ statistic to the empirical distribution failed to indicate rejection of the null at any conventional significance level.

With more sophisticated market model specifications, the right-tail p-values for both conventional approaches and the bootstrap approach became somewhat larger. However, p-values for the bootstrap approach were consistently insignificant (under some specifications,

---

[33]Note that p-values in this range were also obtained when Z^{TRAD} and Z^{SR} were calculated. The conventional Z statistics all led to similar conclusions in this study.

the p-value for $\tilde{Z}^D$ was as high as 0.15), while the standard approach p-values remained significant almost without exception. That is, regardless of model specification, a researcher adopting conventional methods would conclude there was evidence of gains to acquirers, while one adopting the bootstrap approach would not.

There is clearly merit to specifying the event study market model with attention to economic forces driving the return generating process. For example, if there is reason to believe risk changes around the time of the event, an allowance for changes in the market return coefficient or the variance should be considered. However, one can never be sure that a chosen specification is *close enough* to the truth. Even a highly parameterized model will inevitably invalidate some of the underlying conditions given what we know to be true about the properties of actual returns. An advantage of the bootstrap approach is that it permits inference without requiring that all of these conditions be satisfied. Thus, while one should certainly aim to specify the most sensible market model possible, an inadvertent failure to capture some aspect of the data generating process does not invalidate inference under the bootstrap event study approach.

Discussion

Results of experiments documented in the above sections establish that the deviation from the standard normal approximation for Z^D can be highly significant in finite samples – over-rejection in the order of hundreds of percent are not uncommon. The data used in this study are quite typical of financial returns data, exhibiting several of the commonly observed features extensively documented in Section 2.2. These features lead to violations of the conditions which underlie conventional analysis, hence simply employing the conventional Z^D and comparing the resulting value to the standard normal distribution may be quite misleading.

When instead one adopts the bootstrap method, the necessary conditions which underlie conventional approaches need not be satisfied. Valid inference is facilitated whether or not the conditions are satisfied. In carrying out analysis, the individual t_i statistics form a population from which random re-samples are drawn. The particular composition of these bootstrap samples will vary randomly, thus some samples will contain the more extreme values (indeed, some of the firms may be represented in a given bootstrap sample several times) while other samples may not. When analysis is based on the empirical distribution, the extreme nature of some t_i statistics relative to others is accounted for, *without making inappropriate distributional assumptions.* Hence, an accurate estimate of the appropriate distribution for $\tilde{Z}^D$ is achieved. To the extent that the empirical distribution differs from the assumed distribution, conflicting results can be obtained using the bootstrap approach

versus conventional methods, and indeed are obtained in this example.

6 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 based on conventional analysis. 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 thesis, however, I have argued that for the asymptotic distribution of the test statistics to be valid, characteristics of the data must satisfy basic necessary conditions which are commonly violated by the data used in event studies. That is, test statistics may be biased *regardless of how large a sample is assembled.*

As a solution to this problem, I have proposed an alternative testing procedure based on normalizing conventional Z statistics and empirically estimating their distribution with the bootstrap. I presented theoretical evidence to establish the validity of using this bootstrap approach on data with properties like those of actual stock returns, and I demonstrated the empirical performance of the bootstrap approach. The technique is not prone to exhibit biased size in common situations which render bias in conventional techniques – with no sacrifice of power – hence I recommend its use for hypothesis testing in event studies.

When I conducted an actual event study using stock returns data for real firms, the difference between approaches became even more evident. While conventional event study test statistics implied that failed-bank acquirers gained at the expense of the FDIC, that result was overturned by adopting the bootstrap approach. This brings to question the validity of other well-known event study results in finance, and suggests a very broad range of topics for future research.

References

Acharya, S., 1988, "A generalized econometric model and tests of a signalling hypothesis with two discrete signals," *Journal of Finance*, 43, 413-429.

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

Balbirer, S.D., G.D. Jud, and F.W. Lindahl, 1992, "Regulation, competition, and abnormal returns in the market for failed thrifts," *Journal of Financial Economics*, 31, 107-131.

Bertin, W.J., F. Ghazanfari, and K.M. Torabzadeh, 1989, "Failed bank acquisitions and successful bidders' returns," *Financial Management*, Summer, 93-100.

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.

Bradley, M., 1980, "Interfirm tender offers and the market for corporate control," *Journal of Business*, 53, 345-376.

Bradley, M., A. Desai, and E.H. Kim, 1988, "Synergistic gains from corporate acquisitions and their division between the stockholders of target and acquiring firms," *Journal of Financial Economics*, 21, 3-40.

Brockett, P.L, H.M. Chen, and J.R. Garven, 1994, "Event study methodology: A new and stochastically flexible approach," University of Texas manuscript.

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.

Bullock, D.S., 1995, "Are government transfers efficient? An alternative test of the efficient redistribution hypothesis," *Journal of Political Economy*, 103, 1236 - 1274.

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

Cochran, B., L.C. Rose, and D.R. Fraser, 1995, "A market evaluation of FDIC assisted transactions," *Journal of Banking and Finance*, 19, 261-279.

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.

Corrado, Charles J. and Terry L. Zivney, 1992, "The specification and power of the sign test in event study hypothesis tests using daily stock returns," *Journal of Financial and Quantitative Analysis*, 27, 465-478.

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.

Dimson, E. and P. Marsh, 1985, "Event study methodology and the size effect: The case of UK press recommendations," *Journal of Financial Economics*, 17, 113-142.

Dodd, P., and R. Ruback, 1977, "Tender offers and stockholder returns," *Journal of Financial Economics*, 5, 351-373.

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

Eckbo, B.E., 1985, "Mergers and the market concentration doctrine: Evidence from the capital market," *Journal of Business*, 58, 325-349.

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

Eckbo, B.E., V. Maksimovic, and J. 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.

Fama, E.F., and K.R. French, 1993, "Common risk factors in the returns on stocks and bonds," *Journal of Financial Economics*, 33, 3-56.

French, K. and R. McCormick, 1984, "Sealed bids, sunk costs and the process of competition," *Journal of Business*, 57, 417-441.

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).

Ikenberry, D., J. Lakonishok, and T. Vermaelen, 1995, "Market underreaction to open market share repurchases," *Journal of Financial Economics*, 39, 181-208.

James, C. and P. Wier, 1987, "An analysis of FDIC failed bank auctions," *Journal of Monetary Economics*, 20, 141-153.

Johnson, 1979, "Auction markets, bid preparation costs and entrance fees," *Journal of Law and Economics*, 55, 313-318.

Kalay, A. and U. Loewenstein, 1985, "Predictable events and excess returns: The case of dividend announcements," *Journal of Financial Economics*, 14, 423-449.

King, B.F., 1966, "Market and industry factors in stock price behavior," *Journal of Business*, 39, Part 2, 139-190.

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

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

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

Larsen, R.J. and M.L. Marx, 1981, *An Introduction to Mathematical Statistics and its Applications,* (Prentice-Hall, Inc., New Jersey).

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).

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., 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.

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.

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

Mooney, C.Z., 1996, "Bootstrap statistical inference: examples and evaluations for political science," *American Journal of Political Science*, 40, 570-602.

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.

O'Hara, M. and W. Shaw, 1990, "Deposit insurance and wealth effects: The value of being 'Too Big to Fail'," *Journal of Finance*, 45, 1587-1600.

Patell, J.M., 1976, "Corporate forecasts of earnings per share and stock price behavior: empirical tests," *Journal of Accounting Research*, 14, 246-276.

Pettway, R.H. and J.W. Trifts, 1985, "Do banks overbid when acquiring failed banks?" *Financial Management*, Summer, 5-15.

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.

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.

Sefcik, S.E. and R. Thompson, 1986, "An approach to statistical inference in cross-sectional models with security abnormal returns as dependent variables," *Journal of Accounting Research*, 24, 316-334.

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 Testing for Cumulative Effects

The main text focuses on single-day event study test statistics. One may also be interested in tests designed to detect significant *cumulative* effects over *multiple* event days. For example, instead of testing for a significant effect at $t = +1$ which may have resulted directly from a particular announcement, one may wish to test for an effect over a multiple-day period like $t = -3$ through $t = -1$, perhaps in order to determine whether or not there was an information leak prior to the announcement of the event. All three of the common event study test statistics discussed in the main body of the text can be simply modified to allow for multiple day tests.

A.1 The Dummy Variable Approach

Instead of using a single dummy variable in each firm's market model, one would define a dummy variable for each day in the multiple-day period of interest, denoting the first of the multiple days as a and the last of the multiple days as b. For example, in the case of testing for information leak during $t = -3$ through $t = -1$, one would set $a = -3$ and $b = -1$. Then (making use of the index variable τ to facilitate summation over the multiple-event-day testing period) the market model for firm i would become:

$$R_{it} = \beta_{i0} + \beta_{i1}M_{it} + \sum_{\tau=a}^{b} \beta_{iD\tau}D\tau_{it} + \epsilon_{it}. \tag{16}$$

In testing for a cumulative effect over the period starting at $t = a$ and ending at $t = b$, each firm's market model would have $b - a + 1$ dummy variables: one for every day in the multiple-event-day period of interest. Thus, $D\tau_{it}$ would be set to equal one for $t = \tau$ and zero otherwise.

There would be a t-statistic for each of the dummy variable coefficients for each firm, and all of these would be used to test the significance of cumulative effects over the multiple event days starting with $t = a$ and ending with $t = b$:

$$Z^D = \frac{\sum_{t=a}^{t=b} \sum_{i=1}^{N} t_i}{\sqrt{b - a + 1}\sqrt{N}}. \tag{17}$$

Of course, just as the distribution of the single-day Z^D test statistic depends on basic assumptions discussed in Section 2.1 of the main text, the multiple-day Z_{cum}^D test statistic's

48

distribution would depend upon satisfying similar assumptions. Note that a well-specified market model is a critical component of these assumptions. This includes, among other features, the absence of autocorrelation in the disturbances for a firm's market model.

A.2   The Standardized Residual Approach

A Z^{SR} statistic can also be defined to test for cumulative effects over multiple days, starting with $t = a$ and ending with $t = b$:

$$Z^{SR}_{cum} = \frac{\sum_{t=a}^{b} \sum_{i=1}^{N} \tilde{e}_{it}}{\sqrt{b - a + 1}\sqrt{\sum_{i=1}^{N} \frac{T_i - 2}{T_i - 4}}}. \tag{18}$$

Of course, the distribution of this test statistic would rely critically on assumptions like those laid out in Section 2.1 above.

A.3   The Traditional Approach

To test for cumulative effects over multiple days within the event period, starting with $t = a$ and ending with $t = b$, Z^{TRAD} would be defined as follows:

$$Z^{TRAD}_{cum} = \frac{\sum_{t=a}^{b} \bar{\epsilon}_t}{\sqrt{(b - a + 1)\hat{\sigma}^2_{\bar{\epsilon}}}}$$

The distribution of this test statistic would rely upon conditions like those discussed in Section 2.1 of the text.

B   Further Details on Experiment Design

For the 4-step Monte Carlo approach outlined in Section 3.1 above, non-normality and DGP changes were incorporated in the generated data for each firm. Several empirical studies 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*

    As discussed in Section 2.2, there is considerable evidence that market returns and individual firms' returns are highly skewed and leptokurtic. Thus, past studies such as those by Kon [1984] and Lamoureux and Lastrapes [1990] were consulted regarding the

49

statistical properties of actual stock returns and market returns data,[34] 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 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 data for firms' disturbances.

- *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 Hathaway [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%. Likewise, De Jong, Kemna and Kloek [1992] Brockett, Chen, and Garven [1994], and Donaldson and Hathaway [1994] show that the coefficient on market returns in the market model is not necessarily constant over time. Donaldson and Hathaway 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. For the experiments conducted in this thesis, the values chosen are conservative. The event period variance can increase by as much as 500% and the event period market return coefficient can rise by as much as 100%.

---

[34]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.

The change in event period variance was incorporated in the data by re-scaling event period disturbances to have a variance up to 500% larger than that of non-event period disturbances. 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.[35]

- *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 thesis, 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. In this set of experiments, disturbance terms for all firms were conservatively generated with kurtosis of

---

[35]In these experiments, data were generated for all firms assuming a market model intercept of zero and a baseline market return coefficient of one. Allowing different firms to have different parameter values or selecting parameter values to match values observed in practice would leave the experiment results completely unchanged. (Marais [1984] refers to this fact on page 42 of his study.) The explanation for this invariance with respect to choice of parameter value relies on the fact that OLS is unbiased. As explained in Section 3.1, 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.

6.2, skewness of 0.15, standard deviation of 0.77, and zero mean. Notice that the properties given to the *disturbances* were set in accord with moments estimated on actual *returns*. I verified that the properties of market model residuals are in fact indistinguishable from those of firms' returns by estimating the market model for a collection of 45 financial firms. Since the value of R^2 in the estimation of the market model is typically less than 10% and is occasionally even less than 1%, the fact that the large residual component displays similar characteristics to the returns is not surprising. Thus, for the Monte Carlo experiments it is entirely reasonable to ascribe the variance of firms' *returns* to the generated *disturbances*.[36]

C   Confidence Intervals for the Monte Carlos

The confidence intervals around the nominal size values are constructed as follows. Each replication can be considered a binomial experiment under the null hypothesis, where rejecting the null is considered a success and failing to reject is considered a failure. Define n as the number of replications, p_0 as the nominal size or the theoretical rejection rate under the null hypothesis (*e.g.* 0.05 for a 5% level of significance), and Y as the actual number of rejections. With n independent replications resulting in Y rejections and hence an actual rejection rate of $\frac{Y}{n}$, we can construct a confidence interval around the assumed probability of rejection p_0. Under the null hypothesis, the actual rejection rate $\frac{Y}{n}$ is distributed with mean p_0 and variance $\frac{1}{n}p_0(1 - p_0)$. Thus, the expression

$$\frac{\frac{Y}{n} - p_0}{\sqrt{\frac{1}{n}p_0(1 - p_0)}} \tag{19}$$

behaves like a standard normal. For a given significance level, α, it is the case that

$$\Pr\left[-z_{\alpha/2} < \frac{\frac{Y}{n} - p_0}{\sqrt{\frac{1}{n}p_0(1 - p_0)}} < +z_{\alpha/2} \right] \approx 1 - \alpha, \tag{20}$$

where $z_{\alpha/2}$ denotes the appropriate critical value from the normal distribution for a two-sided hypothesis test (*i.e.* $\Pr(Z \leq -z_{\alpha/2}) = \Pr(Z \geq +z_{\alpha/2}) = \alpha/2$).

Making use of these results, a confidence interval for a given α level can be constructed around the nominal size p_0. The $100(1 - \alpha)$% confidence interval is given by:

---

[36]In any case, the magnitude of the variance does not affect the quantitative results in these experiments. Since all of the Z statistics involve a normalization based on variance, the magnitude of the disturbances' variability completely drops out in the absence of event period variance changes. In the presence of event period variance changes, the absolute magnitude of the change is also irrelevant – in such cases it is the *proportional* change in variance that drives results. Sensitivity checks supported these statements.

$$\left(p_0 - z_{\alpha/2} \sqrt{\frac{p_0 (1 - p_0)}{n}}, \ p_0 + z_{\alpha/2} \sqrt{\frac{p_0 (1 - p_0)}{n}} \right). \tag{21}$$

If the value of $\frac{Y}{n}$ falls outside of this interval, then the null hypothesis that the actual size equals the stated size is rejected, indicating a bias to size.

D   Results of Further Experiments

In Section 3, I demonstrated that common event study test statistics exhibit significantly biased size when applied to data with characteristics which violate underlying conditions. These characteristics, carefully chosen to mirror properties exhibited by actual returns, included excess kurtosis, positive skewness, an increase in variance around the event time, and an increase in the market return coefficient around the event time. In this appendix, I demonstrate that the significant bias remains even when the condition violations are modeled to be less severe. The effects of characteristics individually at various levels of intensity are examined, and the impact of allowing different characteristics for different firms is investigated.

D.1   The Marginal Effect of Individual Factors

For the experiments documented in Section 3 above, I generated data to incorporate several violating characteristics simultaneously. This led to significant bias in the statistical size of event study Z statistics. When the violating characteristics are each considered in isolation, hypothesis tests based on the conventional Z statistics continue to over-reject, indicating that even the presence of individual factors in isolation is sufficient for biased size. A consideration of one factor – event period variance changes – is presented below.[37]

Event-induced variance of a much smaller magnitude than found by past researchers – proportional increases as low as 5% – can lead to significant bias in the statistical size of Z statistics. The market returns and the disturbances were generated as standard normal, then non-event period disturbances were given a variance consistent with actual CRSP data.[38] During the event period $t = (-10, \cdots, +10)$, the disturbances were increased by a factor

---

[37]This factor is examined in detail because it has the most profound impact in leading to bias in conventional test statistics. Factors such as skewness, excess kurtosis, or event period changes in the market return coefficient have a smaller role in causing the bias, though their effects are still statistically significant in many cases.

[38]Note that the *magnitude* of the variance does not influence results. It is the proportional difference between event period and non-event period disturbances which drives the result in these experiments.

ranging from 5% to 500%. In all cases, there was considerable over-rejection of the null hypothesis, despite the fact that there was no event effect present.

Table 10 contains the results based on an event study of 50 firms[39] and 1000 replications. The first column presents the nominal size. The remaining columns report information on the cases of 500%, 100%, 20%, and 5% increases in variance respectively. For the 500% case, the rejection rates are large in magnitude, ranging from 11% to 17% when they are expected to be 1%. All of the right-tail p-values for all of the statistics are indistinguishable from zero for this case. For the cases of 100% and 20% proportional rises in variance, all of the Z statistics significantly over-reject once again. The p-values are all indistinguishable from zero at three decimal places. With a 5% increase in variance, all three Z statistics over-reject at low significance levels, but rejection rates are within the confidence bounds for tests conducted at higher significance levels.

The degree of over-rejection documented in Table 10 is quite remarkable given the conservative nature of this experiment. The data was generated with no skewness or excess kurtosis, and firms' true market return coefficients did not undergo any changes during the event period. Furthermore, the values chosen for the variance increases are quite moderate relative to what is actually observed. Recall that much greater increases in event period variance have been documented in past studies such as Donaldson and Hathaway [1994] where variance is observed to rise by as much as 1100%.

As explained in Section 3, the intuition for the incidence of over-rejection is based on the fact that in the presence of event period increases in variance, the event day disturbance is drawn from a wider distribution than assumed under the null. As a result, it is more likely that a Z statistic larger than the critical value will be observed, and hence significant over-rejection takes place. In the case of an unmodeled *decrease* in event period variance, the opposite result may obtain. The event day disturbance will be drawn from a less disperse distribution than assumed under the null, and hence the Z statistics will tend to under-reject.

D.2 Allowing Different DGPs Across Firms

In previous experiments, while data for each firm were generated randomly, overall properties like variance, skewness, kurtosis, and model coefficients were constrained to be identical for each firm in the sample. In this section, the constraint is relaxed, and these properties are permitted to vary across firms. While the results presented below pertain specifically to the

---

[39]With a greater number of firms in the sample, there is still evidence that erroneous conclusions can be reached − rejection rates are statistically larger than they should be. Experiments were also conducted with a 10% proportional increase in variance at the time of the event for samples of 100, and 200 firms. In all cases, almost all of the rejection rates for all the statistics at all levels of significance were outside the 95% confidence interval.

case of changes in event period variance which differ across firms, the results for varying other aspects of the experiment are qualitatively similar. In all cases, the statistical size of event study Z statistics is significantly biased.

Data were initially generated with skewness of 0.15, kurtosis of 6.2, an event period jump of 100% in the market return coefficient, and an increase in event period variance which varies across firms. Rejection rates and right-tail p-values for the experiment appear in Table 11. The first set of three columns presents rejection rates and p-values for the case case where the variance increase is uniformly distributed between 400% and 500% across firms, and the second set of columns is for the case of a uniformly distributed increase between 100% and 500%. Under both these sets of conditions, the rejection rates are considerably larger than they should be under the null. Rejection rates are anywhere from 100% to 700% higher than expected under the null, and all right-tail p-values are indistinguishable from zero at three decimal places.

In order to consider the *marginal* effect of changes in event-period variance which differs across firms, data were also generated without skewness, excess kurtosis, or an increase in the true market return parameter. Even in the absence of these actual features of returns data, event-period variance increases are sufficient to significantly bias the size of test statistics. The final two sets of columns contain these results. Whether the event-period increase is uniformly distributed between 400% and 500% across firms or between 100% and 500% across firms, the over-rejections are highly significant. Right-tail p-values for tests at significance levels 0.01 − 0.10 are indistinguishable from zero.

E   The New Approach Based on Z^{TRAD} or Z^{SR}

In Section 4.2, I presented detailed steps for employing the bootstrap to conduct valid inference based on the conventional test statistic used in the Dummy Variable Approach, Z^D. In this appendix, I suggest simple modifications to those steps to enable bootstrap inference based on the other conventional event study test statistics considered in this thesis, Z^{TRAD} and Z^{SR}. In fact, the steps can be modified to allow the use of the bootstrap with any normalized conventional event study test statistic. The modifications to Steps 1 − 6 are as follows.

1. Instead of estimating the Dummy Variable market model shown in Equation (1), estimate the market model shown in Equation (3). Instead of calculating Z^D, calculate Z^{TRAD} or Z^{SR}, as appropriate.

$$Z^{TRAD} = \frac{\frac{1}{N} \sum_{i=1}^{N} \hat{\epsilon}_{i1}}{\hat{\sigma}_{\bar{\epsilon}}} \tag{8}$$

$$Z^{SR} = \frac{\sum_{i=1}^{N} \tilde{e}_{i1}}{\sqrt{\sum_{i=1}^{N} \frac{T_i - 2}{T_i - 4}}} \tag{7}$$

2. Instead of normalizing the original Z^D statistic, Z^{TRAD} or Z^{SR} is normalized. Recall that for the case of Z^D, the standard deviation of the t-statistics is used to normalize. Neither the Traditional Approach nor the Standardized Residual Approach yield t-statistics, thus each of Z^{TRAD} and Z^{SR} must be normalized by some analogous standard deviation. For the Traditional Approach, one takes the standard deviation of the event day forecast errors divided by $\sqrt{N}\hat{\sigma}_{\bar{\epsilon}}$. For the Standardized Residual Approach, the standard deviation of the event day standardized residuals multiplied by $\sqrt{N}$ and divided by $\sqrt{\frac{T_i - 2}{T_i - 4}}$ is used. These random variables are denoted x_i^{TRAD} and x_i^{SR} respectively:

$$x_i^{TRAD} \equiv \frac{\hat{\epsilon}_{i1}}{\sqrt{N}\hat{\sigma}_{\bar{\epsilon}}}$$

$$x_i^{SR} \equiv \frac{\tilde{e}_{i1}\sqrt{N}}{\sqrt{\sum_{i=1}^{N} \frac{T_i - 2}{T_i - 4}}}$$

Then, defining the standard deviation of the x_i^{TRAD} as $\hat{\sigma}_{x^{TRAD}}$ and the standard deviation of the x_i^{SR} as $\hat{\sigma}_{x^{SR}}$, the normalized Z statistics are computed as follows:

$$\begin{aligned} \tilde{Z}^{TRAD} &= \frac{\frac{\sum_{i=1}^{N} x_i^{TRAD}}{N}\sqrt{N}}{\hat{\sigma}_{x^{TRAD}}} \\ &= \frac{Z^{TRAD}}{\hat{\sigma}_{x^{TRAD}}} \end{aligned} \tag{22}$$

$$\begin{aligned} \tilde{Z}^{SR} &= \frac{\frac{\sum_{i=1}^{N} x_i^{SR}}{N}\sqrt{N}}{\hat{\sigma}_{x^{SR}}} \\ &= \frac{Z^{SR}}{\hat{\sigma}_{x^{SR}}} \end{aligned} \tag{23}$$

56

3. Now, instead of mean-adjusting the actual t-statistics to form the population from which bootstrap samples are randomly drawn, the mean-adjusted x_i^{TRAD} and x_i^{SR} are used. Define the mean of x_i^{TRAD} and x_i^{SR} as $\overline{x}^{TRAD}$ and $\overline{x}^{SR}$ respectively. Then the mean-adjusted statistics from which the bootstrap sample are randomly drawn are defined as follows. For the Traditional Approach

$$x_i^{TRAD*} = x_i^{TRAD} - \overline{x}^{TRAD},$$

and for the Standardized Residual Approach,

$$x_i^{SR*} = x_i^{SR} - \overline{x}^{SR}.$$

4. Instead of randomly re-sampling from the population of t_i^*, the random re-samples are drawn from the x_i^{TRAD*} or the x_i^{SR*} defined above.

5. Instead of forming a normalized version of Z_b^D for each bootstrap sample, a normalized version of Z_b^{TRAD} or Z_b^{SR} is formed. Z_b^{TRAD} is normalized by the standard deviation of the x_{bj}^{TRAD*} in the bootstrap sample, or Z_b^{SR} is normalized by the standard deviation of the x_{bj}^{SR*} in the bootstrap sample.

$$
\begin{aligned}
\tilde{Z}_b^{TRAD} &= \frac{\frac{\sum_{j=1}^{N} x_{bj}^{TRAD*}}{N}\sqrt{N}}{\hat{\sigma}_{x_b^{TRAD*}}} \\
&= \frac{Z_b^{TRAD}}{\hat{\sigma}_{x_b^{TRAD*}}}
\end{aligned}
\tag{24}
$$

$$
\begin{aligned}
\tilde{Z}_b^{SR} &= \frac{\frac{\sum_{j=1}^{N} x_{bj}^{SR*}}{N}\sqrt{N}}{\hat{\sigma}_{x_b^{SR*}}} \\
&= \frac{Z_b^{SR}}{\hat{\sigma}_{x_b^{SR*}}}
\end{aligned}
\tag{25}
$$

6. Finally, critical values from the empirical distribution defined by the $\tilde{Z}_b^{TRAD}$ or the $\tilde{Z}_b^{SR}$ are used to evaluate the normalized statistic computed in Step 2.

F Size-Adjustments for Power Comparisons

As discussed in Section 4.3 above, each test statistic must be size-adjusted prior to conducting the power comparisons. Consider, for example, a biased test statistic which has a rejection rate of 25% for a test at the $\alpha = 5\%$ significance level (like Z^D for a sample of 100 firms, as considered in Figure 1). This rejection rate would have been obtained using a critical p-value of 0.05 for the $\alpha = 5\%$ test. One size-adjusts to determine the actual p-value for which 5% of the statistics are rejected with a test at a 5% significance level. Since using a p-value of 0.05 led to a rejection rate of 25%, the size-adjusted p-value would be *much smaller* than 0.05. Once the size-adjusted 5% critical p-value has been determined (by Monte Carlo simulation), it is used for conducting tests at the 5% significance level in the power experiments. Proper critical values are determined in this fashion for all of the relevant testing levels, 0.01 – 0.10, for a given statistic. Then these size-adjusted critical values are used to evaluate the test statistic in the power experiments.

Unfortunately, the set of actual critical values extracted for a particular test statistic under a particular DGP are not applicable to other test statistics or to settings where different features of the data (such as skewness or event-period variance changes) are observed. To evaluate the power of statistics under different conditions, a new set of appropriate critical values must be extracted. Computer programs to extract the critical values require 1000 or more replications, each replication involving data generation and model estimation for a given number of firms and the computation of event study test statistics, and each replication also requiring 1000 bootstrap re-sampling exercises. After the appropriate critical values are extracted, the power comparisons must be undertaken, an exercise which again involves many replications, computations, and re-samples.

Table 1: Z Statistics: Non-Normal Data with DGP Changes, Using Conventionally Assumed Distributions

Actual Rejection Rates and
Right-Tail Probability Values
for Tests at Common Significance Levels

1000 Replications, Various Numbers of Firms

	30 Firms			50 Firms			100 Firms			200 Firms		
Size	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}
0.01	0.099*	0.152*	0.154*	0.123*	0.168*	0.164*	0.120*	0.163*	0.158*	0.135*	0.184*	0.188*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.02	0.120*	0.182*	0.181*	0.148*	0.195*	0.191*	0.135*	0.192*	0.196*	0.154*	0.220*	0.209*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.03	0.150*	0.203*	0.203*	0.166*	0.202*	0.202*	0.154*	0.215*	0.215*	0.179*	0.235*	0.228*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.04	0.169*	0.222*	0.216*	0.173*	0.214*	0.222*	0.173*	0.235*	0.231*	0.195*	0.251*	0.245*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	0.183*	0.237*	0.231*	0.181*	0.226*	0.230*	0.187*·	0.244*	0.243*	0.210*	0.258*	0.253*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.06	0.198*	0.254*	0.248*	0.197*	0.237*	0.241*	0.204*	0.258*	0.257*	0.235*	0.270*	0.270*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.07	0.208*	0.272*	0.259*	0.210*	0.246*	0.252*	0.212*	0.267*	0.266*	0.241*	0.281*	0.274*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.08	0.215*	0.284*	0.266*	0.216*	0.256*	0.260*	0.224*	0.279*	0.279*	0.248*	0.292*	0.287*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.09	0.233*	0.290*	0.277*	0.224*	0.266*	0.272*	0.238*	0.286*	0.282*	0.260*	0.297*	0.292*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.242*	0.298*	0.288*	0.233*	0.276*	0.282*	0.248*	0.292*	0.291*	0.269*	0.304*	0.299*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

This table presents actual rejection rates and the associated right-tail probability values for Z statistics computed on non-normally distributed data which incorporate changes in the data generating process during the event period. The Ramberg et al. [1979] algorithm was used to generate data with a skewness coefficient of 0.15 and a kurtosis coefficient of 6.2. The true market model coefficient doubles around the time of the event, and the variance of the disturbances jumps by 500% during the event period.

The Z statistics are calculated for tests at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on the various Z statistics for various sample sizes (i.e. number of firms). Z^D is calculated according to Equation (2) and is assumed to follow a standard normal distribution. Z^{SR} is calculated according to Equation (7) and is also assumed to follow a standard normal distribution. Z^{TRAD} is calculated according to Equation (8) and is assumed to follow a Student t-distribution with 119 degrees of freedom.

The first three columns report results for Z statistics computed on a sample of 30 firms. The remaining sets of three columns pertain to sample sizes of 50, 100, and 200 firms respectively. The top value in each cell is the actual rejection rate for a particular Z statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the test statistic's conventionally assumed distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 2: Z Statistics: Normal Data without DGP Changes, Using Conventionally Assumed Distributions

Actual Rejection Rates and
Right-Tail Probability Values
for Tests at Common Significance Levels

1000 Replications, Various Numbers of Firms

	30 Firms			50 Firms			100 Firms			200 Firms		
Size	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}
0.01	0.012	0.011	0.011	0.006	0.006	0.006	0.010	0.010	0.011	0.008	0.006	0.006
	0.263	0.375	0.375	0.898	0.898	0.898	0.500	0.500	0.375	0.737	0.898	0.898
0.02	0.021	0.019	0.020	0.019	0.018	0.022	0.020	0.019	0.020	0.022	0.019	0.019
	0.411	0.589	0.500	0.589	0.674	0.326	0.500	0.589	0.500	0.326	0.589	0.589
0.03	0.030	0.025	0.030	0.037	0.035	0.037	0.028	0.027	0.027	0.028	0.028	0.030
	0.500	0.823	0.500	0.097	0.177	0.097	0.645	0.711	0.711	0.645	0.645	0.500
0.04	0.041	0.038	0.041	0.047	0.046	0.045	0.034	0.033	0.041	0.039	0.038	0.042
	0.436	0.627	0.436	0.129	0.166	0.210	0.834	0.871	0.436	0.564	0.627	0.373
0.05	0.055	0.052	0.052	0.058	0.055	0.055	0.046	0.044	0.053	0.044	0.044	0.047
	0.234	0.386	0.386	0.123	0.234	0.234	0.719	0.808	0.332	0.808	0.808	0.668
0.06	0.062	0.061	0.061	0.065	0.065	0.065	0.058	0.055	0.062	0.054	0.050	0.059
	0.395	0.447	0.447	0.253	0.253	0.253	0.605	0.747	0.395	0.788	0.908	0.553
0.07	0.068	0.067	0.066	0.080	0.073	0.078	0.069	0.070	0.077	0.068	0.067	0.069
	0.598	0.645	0.690	0.108	0.355	0.161	0.549	0.500	0.193	0.598	0.645	0.549
0.08	0.072	0.074	0.078	0.091	0.087	0.093	0.082	0.076	0.086	0.076	0.077	0.080
	0.824	0.758	0.592	0.100	0.207	0.065	0.408	0.679	0.242	0.679	0.637	0.500
0.09	0.085	0.082	0.086	0.101	0.099	0.103	0.092	0.087	0.093	0.089	0.088	0.088
	0.710	0.812	0.671	0.112	0.160	0.075	0.413	0.630	0.370	0.544	0.587	0.587
0.1	0.095	0.090	0.097	0.112	0.111	0.113	0.101	0.099	0.107	0.103	0.101	0.100
	0.701	0.854	0.624	0.103	0.123	0.085	0.458	0.542	0.230	0.376	0.458	0.500

This table presents actual rejection rates and the associated right-tail probability values for Z statistics computed on normally distributed data which do not incorporate any changes in the DGP at the time of the event.

The Z statistics are calculated for tests at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on the various Z statistics for various sample sizes (*i.e.* number of firms). Z^D is calculated according to Equation (2) and is assumed to follow a standard normal distribution. Z^{SR} is calculated according to Equation (7) and is also assumed to follow a standard normal distribution. Z^{TRAD} is calculated according to Equation (8) and is assumed to follow a Student t-distribution with 119 degrees of freedom.

The first three columns report results for Z statistics computed on a sample of 30 firms. The remaining sets of three columns pertain to sample sizes of 50, 100, and 200 firms respectively. The top value in each cell is the actual rejection rate for a particular Z statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the test statistic's conventionally assumed distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 3: Rank Test Statistic: Non-Normal Data with DGP Changes, Using Conventionally Assumed Distributions

Size	30 Firms	50 Firms	100 Firms	200 Firms	400 Firms	500 Firms	600 Firms	800 Firms	1000 Firms
0.01	0.027*	0.022*	0.022*	0.019*	0.023*	0.015	0.017*	0.020*	0.012
	0.000	0.000	0.000	0.002	0.000	0.056	0.013	0.001	0.263
0.02	0.044*	0.033*	0.031*	0.032*	0.036*	0.031*	0.025	0.036*	0.026
	0.000	0.002	0.006	0.003	0.000	0.006	0.129	0.000	0.088
0.03	0.057*	0.047*	0.049*	0.051*	0.044*	0.044*	0.037	0.048*	0.038
	0.000	0.001	0.000	0.000	0.005	0.005	0.097	0.000	0.069
0.04	0.080*	0.057*	0.065*	0.062*	0.057*	0.061*	0.055*	0.056*	0.048
	0.000	0.003	0.000	0.000	0.003	0.000	0.008	0.005	0.098
0.05	0.094*	0.072*	0.080*	0.072*	0.067*	0.074*	0.065*	0.069*	0.055
	0.000	0.001	0.000	0.001	0.007	0.000	0.015	0.003	0.234
0.06	0.102*	0.086*	0.097*	0.086*	0.074	0.079*	0.078*	0.081*	0.063
	0.000	0.000	0.000	0.000	0.031	0.006	0.008	0.003	0.345
0.07	0.114*	0.103*	0.102*	0.094*	0.081	0.093*	0.086*	0.088*	0.070
	0.000	0.000	0.000	0.001	0.086	0.002	0.024	0.013	0.500
0.08	0.130*	0.120*	0.111*	0.112*	0.090	0.098*	0.094	0.098*	0.082
	0.000	0.000	0.000	0.000	0.122	0.018	0.051	0.018	0.408
0.09	0.144*	0.128*	0.115*	0.128*	0.098	0.107	0.103	0.103	0.090
	0.000	0.000	0.003	0.000	0.188	0.030	0.075	0.075	0.500
0.1	0.157*	0.139*	0.122*	0.137*	0.111	0.114	0.116	0.112	0.100
	0.000	0.000	0.010	0.000	0.123	0.070	0.046	0.103	0.500

This table presents actual rejection rates and the associated right-tail probability values for the rank statistic computed on non-normally distributed data which incorporate changes in the data generating process during the event period. The Ramberg *et al.* [1979] algorithm was used to generate data with a skewness coefficient of 0.15 and a kurtosis coefficient of 6.2. The true market model coefficient doubles around the time of the event, and the variance of the disturbances jumps by 500% during the event period.

The rank statistic is calculated for tests at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on various sample sizes (*i.e.* number of firms). Equation (11) is used to compute the rank test statistic and it is assumed to follow a standard normal distribution.

The first column reports results for a sample of 30 firms, and remaining columns pertain to successively larger sample sizes. The top value in each cell is the actual rejection rate for the test statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the conventionally assumed standard normal distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 4: Sign Test Statistic: Non-Normal Data with DGP Changes, Using Conventionally Assumed Distributions

Actual Rejection Rates and
Right-Tail Probability Values
for Tests at Common Significance Levels

1000 Replications, Various Numbers of Firms

Size	30 Firms	50 Firms	100 Firms	200 Firms	400 Firms	500 Firms	600 Firms	800 Firms	1000 Firms
0.01	0.036*	0.032*	0.022*	0.021*	0.021*	0.019*	0.012	0.014	0.018*
	0.000	0.000	0.000	0.000	0.000	0.002	0.263	0.102	0.001
0.02	0.045*	0.038*	0.031*	0.030*	0.032*	0.024	0.026	0.025	0.021
	0.000	0.000	0.006	0.012	0.003	0.183	0.088	0.130	0.411
0.03	0.054*	0.047*	0.038	0.049*	0.042*	0.045*	0.044*	0.043*	0.041*
	0.000	0.001	0.069	0.000	0.013	0.003	0.004	0.008	0.021
0.04	0.063*	0.057*	0.059*	0.058*	0.055*	0.053*	0.055*	0.052	0.053*
	0.000	0.003	0.001	0.002	0.008	0.018	0.007	0.026	0.018
0.05	0.078*	0.064*	0.069*	0.061*	0.057	0.065	0.066*	0.064*	0.054
	0.000	0.021	0.002	0.055	0.154	0.148	0.010	0.021	0.281
0.06	0.092*	0.087*	0.081*	0.078*	0.076*	0.076*	0.071	0.073	0.075*
	0.000	0.000	0.003	0.008	0.017	0.017	0.071	0.042	0.023
0.07	0.105*	0.095*	0.088*	0.087*	0.092*	0.089*	0.088*	0.086*	0.085
	0.000	0.001	0.013	0.018	0.003	0.010	0.012	0.024	0.031
0.08	0.104*	0.098*	0.099*	0.100*	0.099*	0.096	0.097*	0.094	0.095
	0.003	0.018	0.013	0.010	0.013	0.031	0.024	0.051	0.040
0.09	0.115*	0.111*	0.110*	0.112*	0.109*	0.107	0.109*	0.108*	0.107
	0.002	0.010	0.014	0.008	0.018	0.030	0.018	0.023	0.030
0.10	0.127*	0.124*	0.120*	0.121*	0.119*	0.122*	0.121*	0.123*	0.117
	0.002	0.006	0.018	0.013	0.023	0.010	0.013	0.008	0.037

This table presents actual rejection rates and the associated right-tail probability values for the sign test statistic computed on non-normally distributed data which incorporate changes in the data generating process during the event period. The Ramberg *et al.* [1979] algorithm was used to generate data with a skewness coefficient of 0.15 and a kurtosis coefficient of 6.2. The true market model coefficient doubles around the time of the event, and the variance of the disturbances jumps by 500% during the event period.

The sign test statistic is calculated for tests at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on various sample sizes (*i.e.* number of firms). Equation (12) is used to compute the rank test statistic and it is assumed to follow a standard normal distribution.

The first column reports results for a sample of 30 firms, and remaining columns pertain to successively larger sample sizes. The top value in each cell is the actual rejection rate for the test statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the conventionally assumed standard normal distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 5: Normalized $\tilde{Z}$ Statistics: Non-Normal Data with DGP Changes, Using the Bootstrap Distribution

Actual Rejection Rates and
Right-Tail Probability Values
for Tests at Common Significance Levels

1000 Replications, Various Numbers of Firms

Size	30 Firms			50 Firms			100 Firms			200 Firms		
	$\tilde{Z}^D$	$\tilde{Z}^{SR}$	$\tilde{Z}^{TRAD}$	$\tilde{Z}^D$	$\tilde{Z}^{SR}$	$\tilde{Z}^{TRAD}$	$\tilde{Z}^D$	$\tilde{Z}^{SR}$	$\tilde{Z}^{TRAD}$	$\tilde{Z}^D$	$\tilde{Z}^{SR}$	$\tilde{Z}^{TRAD}$
0.01	0.014	0.014	0.009	0.014	0.011	0.012	0.013	0.011	0.015	0.017*	0.015	0.015
	0.102	0.102	0.625	0.102	0.375	0.263	0.170	0.375	0.056	0.013	0.056	0.056
0.02	0.024	0.024	0.023	0.028	0.025	0.023	0.024	0.023	0.024	0.023	0.025	0.026
	0.183	0.183	0.249	0.035	0.129	0.249	0.183	0.249	0.183	0.249	0.129	0.088
0.03	0.033	0.029	0.028	0.040	0.037	0.038	0.038	0.033	0.033	0.032	0.035	0.033
	0.289	0.574	0.645	0.032	0.097	0.069	0.069	0.289	0.289	0.355	0.177	0.289
0.04	0.044	0.038	0.036	0.049	0.045	0.047	0.051	0.046	0.040	0.043	0.044	0.041
	0.259	0.627	0.741	0.073	0.210	0.129	0.038	0.166	0.500	0.314	0.259	0.436
0.05	0.055	0.047	0.048	0.057	0.055	0.056	0.061	0.054	0.054	0.054	0.053	0.052
	0.234	0.668	0.614	0.155	0.234	0.192	0.055	0.281	0.281	0.281	0.332	0.386
0.06	0.060	0.060	0.056	0.071	0.069	0.065	0.069	0.067	0.067	0.069	0.061	0.067
	0.500	0.500	0.703	0.071	0.115	0.253	0.115	0.176	0.176	0.115	0.447	0.176
0.07	0.068	0.065	0.064	0.081	0.081	0.076	0.078	0.082	0.077	0.080	0.073	0.078
	0.598	0.732	0.771	0.086	0.086	0.229	0.161	0.068	0.193	0.108	0.355	0.161
0.08	0.077	0.072	0.074	0.089	0.091	0.090	0.096	0.095	0.091	0.089	0.085	0.084
	0.637	0.824	0.758	0.147	0.100	0.122	0.031	0.040	0.100	0.147	0.280	0.321
0.09	0.084	0.082	0.084	0.103	0.100	0.095	0.107	0.104	0.106	0.101	0.095	0.091
	0.746	0.812	0.746	0.075	0.135	0.290	0.030	0.061	0.039	0.112	0.290	0.456
0.1	0.094	0.090	0.090	0.114	0.109	0.104	0.121*	0.118	0.119*	0.108	0.109	0.097
	0.736	0.854	0.854	0.070	0.171	0.337	0.013	0.029	0.023	0.200	0.171	0.624

This table presents actual rejection rates and the associated right-tail probability values for $\bar{Z}$ statistics based on the bootstrap approach. The data are non-normally distributed and incorporate changes in the DGP during the event period. The Ramberg *et al.* [1979] algorithm was used to generate data with a skewness coefficient of 0.15 and a kurtosis coefficient of 6.2. The true market model coefficient doubles around the time of the event, and the variance of the disturbances jumps by 500% during the event period.

The $\bar{Z}$ statistics are calculated for tests at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns reports statistics based on the various $\tilde{Z}$ statistics for various sample sizes (*i.e.* number of firms). $\tilde{Z}^D$ is calculated according to Equation (13), $\tilde{Z}^{SR}$ is calculated according to Equation (23), and $\tilde{Z}^{TRAD}$ is calculated according to Equation (22). Confidence intervals are constructed using Student t intervals and 1 000 bootstrap samples as explained in the text.

The first three columns report results for $\tilde{Z}$ statistics computed on a sample of 30 firms. The remaining sets of three columns pertain to sample sizes of 50, 100, and 200 firms respectively. The top value in each cell is the actual rejection rate for a particular $\bar{Z}$ statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the test statistic's conventionally assumed distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 6: Power Comparisons: Z Statistics and Normalized $\tilde{Z}$ Statistics

**Actual Rejection Rates
for Tests at Common Significance Levels**

1000 Replications, 50 Firms

Abnormal Return	Size	Z^D	Z^{SR}	Z^{TRAD}	$\tilde{Z}^D$	$\tilde{Z}^{SR}$	$\tilde{Z}^{TRAD}$
0.009	0.01	1.000	1.000	1.000	0.998	0.999	0.999
	0.02	1.000	1.000	1.000	1.000	0.999	1.000
	0.03	1.000	1.000	1.000	1.000	1.000	1.000
	0.04	1.000	1.000	1.000	1.000	1.000	1.000
	0.05	1.000	1.000	1.000	1.000	1.000	1.000
	0.06	1.000	1.000	1.000	1.000	1.000	1.000
	0.07	1.000	1.000	1.000	1.000	1.000	1.000
	0.08	1.000	1.000	1.000	1.000	1.000	1.000
	0.09	1.000	1.000	1.000	1.000	1.000	1.000
	0.1	1.000	1.000	1.000	1.000	1.000	1.000
0.007	0.01	0.982	0.987	0.987	0.958	0.969	0.967
	0.02	0.990	0.991	0.990	0.972	0.980	0.988
	0.03	0.996	0.996	0.995	0.984	0.988	0.991
	0.04	0.997	0.998	0.996	0.989	0.993	0.993
	0.05	0.998	0.998	0.998	0.995	0.994	0.994
	0.06	0.998	0.998	0.998	0.997	0.998	0.996
	0.07	0.999	0.999	0.998	0.997	0.998	0.996
	0.08	1.000	0.999	0.998	0.999	0.999	0.996
	0.09	1.000	0.999	0.999	0.999	0.999	0.998
	0.1	1.000	1.000	0.999	0.999	0.999	0.999
0.005	0.01	0.790	0.832	0.829	0.761	0.781	0.781
	0.02	0.854	0.866	0.849	0.829	0.825	0.851
	0.03	0.903	0.913	0.906	0.870	0.872	0.878
	0.04	0.931	0.929	0.926	0.890	0.907	0.900
	0.05	0.943	0.945	0.946	0.914	0.928	0.929
	0.06	0.951	0.951	0.952	0.942	0.942	0.941
	0.07	0.960	0.960	0.958	0.946	0.947	0.952
	0.08	0.968	0.962	0.964	0.955	0.951	0.957
	0.09	0.969	0.971	0.971	0.956	0.958	0.960
	0.1	0.972	0.974	0.974	0.960	0.964	0.963

This table presents actual rejection rates for Z statistics based on conventional approaches and for normalized $\tilde{Z}$ statistics evaluated using the bootstrap. An abnormal return (shown in the far left column) is added at the time of the event in order to facilitate a comparison of power. The data are non-normally distributed and incorporate changes in the DGP during the event period. The Ramberg et al. [1979] algorithm was used to generate data with a skewness coefficient of 0.15 and a kurtosis coefficient of 6.2. The market model slope coefficient increases by 100% around the time of the event, and the variance of the disturbances jumps by 500% during the event period.

The test statistics are calculated for tests at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on the various statistics. Z^D is calculated according to Equation (2) and is assumed to follow a standard normal distribution. Z^{SR} is calculated according to Equation (7) and is also assumed to follow a standard normal distribution. Z^{TRAD} is calculated according to Equation (8) and is assumed to follow a Student t-distribution with 119 degrees of freedom. $\tilde{Z}^D$ is calculated according to Equation (13), $\tilde{Z}^{SR}$ is calculated according to Equation (23), and $\tilde{Z}^{TRAD}$ is calculated according to Equation (22). Bootstrap confidence intervals are constructed using percentile intervals and 1000 bootstrap samples as explained in the text. The value in each of the cells is the actual rejection rates for a particular Z statistic or $\tilde{Z}$ statistic computed at particular significance level.

64

Table 7: Normalized $\tilde{Z}$ Statistics: Using Conventionally Assumed Distributions (without use of Bootstrap)

Actual Rejection Rates and
Right-Tail Probability Values
for Tests at Common Significance Levels

1000 Replications, Various Numbers of Firms

	30 Firms			50 Firms			100 Firms			200 Firms		
Size	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}
0.01	0.077*	0.078*	0.051*	0.079*	0.079*	0.060*	0.077*	0.079*	0.059*	0.060*	0.059*	0.049*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.02	0.087*	0.091*	0.059*	0.087*	0.093*	0.062*	0.088*	0.091*	0.062*	0.078*	0.079*	0.057*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.03	0.091*	0.095*	0.062*	0.092*	0.095*	0.065*	0.093*	0.093*	0.063*	0.086*	0.087*	0.057*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.04	0.103*	0.098*	0.069*	0.105*	0.097*	0.071*	0.100*	0.098*	0.070*	0.097*	0.088*	0.061*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	0.111*	0.110*	0.078*	0.108*	0.107*	0.079*	0.101*	0.101*	0.075*	0.099*	0.097*	0.064*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.021
0.06	0.129*	0.120*	0.093*	0.125*	0.118*	0.092*	0.122*	0.110*	0.092*	0.102*	0.104*	0.083*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
0.07	0.143*	0.141*	0.114*	0.142*	0.138*	0.116*	0.136*	0.129*	0.111*	0.126*	0.119*	0.105*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.08	0.153*	0.150*	0.135*	0.150*	0.147*	0.134*	0.142*	0.143*	0.135*	0.128*	0.123*	0.126*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.09	0.159*	0.161*	0.143*	0.157*	0.157*	0.141*	0.153*	0.151*	0.137*	0.136*	0.135*	0.132*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.183*	0.181*	0.170*	0.180*	0.176*	0.168*	0.174*	0.168*	0.159*	0.146*	0.147*	0.148*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

This table presents actual rejection rates and the associated right-tail probability values for normalized $\tilde{Z}$ statistics computed on non-normally distributed data which incorporate changes in the data generating process during the event period. The Ramberg *et al.* [1979] algorithm was used to generate data with a skewness coefficient of 0.15 and a kurtosis coefficient of 6.2. The true market model coefficient doubles around the time of the event, and the variance of the disturbances jumps by 500% during the event period.

The test statistics are calculated for significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on the various Z statistics for various sample sizes (*i.e.* number of firms). Z^D is calculated according to Equation (2) and is assumed to follow a standard normal distribution. Z^{SR} is calculated according to Equation (7) and is also assumed to follow a standard normal distribution. Z^{TRAD} is calculated according to Equation (8) and is assumed to follow a Student t-distribution with 119 degrees of freedom.

The first three columns report results for $\tilde{Z}$ statistics computed on a sample of 30 firms. The remaining sets of three columns pertain to sample sizes of 50, 100, and 200 firms respectively. The top value in each cell is the actual rejection rate for a particular $\tilde{Z}$ statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the test statistic's conventionally assumed distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 8: Conventional (Non-Normalized) Z Statistics: Using Bootstrap Distributions

Actual Rejection Rates and
Right-Tail Probability Values
for Tests at Common Significance Levels

1000 Replications, Various Numbers of Firms

Size	30 Firms			50 Firms			100 Firms			200 Firms		
	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}
0.01	0.021*	0.021*	0.022*	0.026*	0.029*	0.031*	0.023*	0.021*	0.024*	0.031*	0.028*	0.030*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.02	0.033*	0.031*	0.033*	0.046*	0.040*	0.046*	0.034*	0.034*	0.034*	0.040*	0.038*	0.038*
	0.002	0.006	0.002	0.000	0.000	0.000	0.001	0.001	0.001	0.000	0.000	0.000
0.03	0.046*	0.044*	0.045*	0.060*	0.061*	0.059*	0.048*	0.048*	0.049*	0.047*	0.046*	0.047*
	0.002	0.005	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.001
0.04	0.061*	0.052	0.053*	0.072*	0.068*	0.069*	0.066*	0.059*	0.060*	0.062*	0.057*	0.057*
	0.000	0.026	0.018	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.003
0.05	0.066*	0.064*	0.064*	0.083*	0.081*	0.078*	0.076*	0.067*	0.067*	0.068*	0.066*	0.063
	0.010	0.021	0.021	0.000	0.000	0.000	0.000	0.007	0.007	0.005	0.010	0.030
0.06	0.077*	0.078*	0.077*	0.093*	0.089*	0.092*	0.082*	0.073	0.080*	0.081*	0.080*	0.079*
	0.012	0.008	0.012	0.000	0.000	0.000	0.002	0.042	0.004	0.003	0.004	0.006
0.07	0.087*	0.088*	0.085	0.100*	0.103*	0.104*	0.091*	0.084	0.082	0.090*	0.090*	0.087*
	0.018	0.013	0.032	0.000	0.000	0.000	0.005	0.041	0.068	0.007	0.007	0.018
0.08	0.095	0.099*	0.096	0.112*	0.117*	0.118*	0.105*	0.093	0.093	0.100*	0.102*	0.104*
	0.040	0.013	0.031	0.000	0.000	0.000	0.002	0.065	0.065	0.010	0.005	0.003
0.09	0.103	0.106	0.102	0.124*	0.130*	0.126*	0.115*	0.105	0.103	0.115*	0.114*	0.109*
	0.075	0.039	0.092	0.000	0.000	0.000	0.003	0.049	0.075	0.003	0.004	0.018
0.1	0.112	0.115	0.115	0.138*	0.146*	0.141*	0.123*	0.111	0.116	0.128*	0.121*	0.120*
	0.103	0.057	0.057	0.000	0.000	0.000	0.008	0.123	0.046	0.002	0.013	0.018

This table presents actual rejection rates and the associated right-tail probability values for Z statistics computed on non-normally distributed data which incorporate changes in the data generating process during the event period. The Ramberg *et al.* [1979] algorithm was used to generate data with a skewness coefficient of 0.15 and a kurtosis coefficient of 6.2. The true market model coefficient doubles around the time of the event, and the variance of the disturbances jumps by 500% during the event period.

The test statistics are calculated for significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on the various $\tilde{Z}$ statistics for various sample sizes (*i.e.* number of firms). $\tilde{Z}^D$ is calculated according to Equation (13), $\tilde{Z}^{SR}$ is calculated according to Equation (23), and $\tilde{Z}^{TRAD}$ is calculated according to Equation (22). Confidence intervals are constructed using Student t intervals and 1000 bootstrap samples as explained in the text.

The first three columns report results for Z statistics computed on a sample of 30 firms. The remaining sets of three columns pertain to sample sizes of 50, 100, and 200 firms respectively. The top value in each cell is the actual rejection rate for a particular Z statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the test statistic's conventionally assumed distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 9: Sample of Failed-Bank Acquirers

Firm	Closure Date	Failed Bank, State	Ticker Symbol of Acquirer
1	08/17/89	First State Bank, Texas	FBT
2	10/06/89	First Bank, Colorado	COLC
3	11/09/89	City National Bank of Plano, Texas	CBSS
4	12/08/89	Westheimer Memorial Bank, Texas	HIB
5	12/14/89	Canyon Lake Bank, Texas	VICT
6	01/26/90	Monroe Savings Bank, New York	FES
7	02/01/90	Tyler National Bank, Texas	NCB
8	02/15/90	Gateway National Bank, Arizona	HIB
9	03/22/90	Miami National Bank, Florida	COBA
10	03/30/90	Everman National Bank of Fort Worth, Texas	ONE
11	04/18/90	The Seamen's Bank for Savings, New York	CMB
12	04/26/90	Central Arizona Bank, Arizona	SPC
13	05/03/90	Richardson National Bank, Texas	CMCA
14	05/04/90	Tucker State Bank of Jacksonville, Florida	BBI
15	06/01/90	The Home National Bank of Milford, Mass.	BBNK
16	06/07/90	Texas National Bank, Texas	CHL
17	07/06/90	Milford Savings Bank, Mass.	PEBW
18	07/13/90	Permanent Savings Bank, New York	KEY
19	08/10/90	National Bank of Washington, Washington DC	RIGS
20	08/16/90	Capitol Bank and Trust, Oklahoma	FFB
21	08/30/90	American Bank of Commerce, Texas	DVL
22	10/25/90	First National Bank of Jackson, Tennessee	FTEN
23	01/11/91	Community National Bank, CT	FNG
24	02/22/91	The McKinley Bank, Ohio	OHBC
25	03/08/91	Manilabank, California	USTC
26	03/14/91	Crossroads Bank, Texas	VICT
27	03/28/91	The Landmark Bank, CT	PBCT
28	04/12/91	Arizona Commerce Bank, Arizona	VALC
29	05/10/91	Madison National Bank; Washington DC and Virginia	SBK
30	05/22/91	The First National Bank of Toms River, New Jersey	FFB
31	05/31/91	Goldome, New York	KEY
32	06/07/91	Woburn Five Cents Savings Bank, Mass.	FSCB
33	06/21/91	Beacon Co-operative Bank, Mass.	GROV
34	06/28/91	First Mutual Bank for Savings, Mass.	BKB
35	07/19/91	Community Guardian Bank, New Jersey	ISB
36	07/19/91	Pontchartrain State Bank, Louisiana	FCOM
37	07/26/91	The Housatonic Bank and Trust Company, CT	SSBC
38	08/09/91	Citytrust, CT and Mechanics and Farmers Savings Bank, CT	CMB
39	08/30/91	Lowell Institution for Savings, Mass.	FMLY
40	09/13/91	Valley Bank, Vermont	VFSC
41	09/19/91	Southeast Bank, Florida	FTU
42	10/18/91	Central Bank, CT	CTBX
43	10/25/91	First Hanover Bank, North Carolina	CCBF
44	12/13/91	Merchants National Bank, Mass.	BNKW
45	12/20/91	North Ridge Bank, Florida	ICBK

Table 10: Z Statistics: The Marginal Effect of Higher Event-Period Variance

Actual Rejection Rates and
Right-Tail Probability Values
for Tests at Common Significance Levels

1000 Replications, 50 Firms

	Percentage Increase in Variance During the Event Period											
	500%			100%			20%			5%		
Size	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}
0.01	0.110*	0.168*	0.169*	0.041*	0.048*	0.048*	0.016*	0.017*	0.017*	0.012*	0.012*	0.012*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.017	0.022	0.008
0.02	0.141*	0.198*	0.200*	0.059*	0.067*	0.068*	0.030*	0.030*	0.031*	0.023*	0.023*	0.024*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.016	0.005
0.03	0.162*	0.218*	0.221*	0.075*	0.088*	0.088*	0.041*	0.041*	0.043*	0.033	0.033	0.034*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.027	0.050	0.010
0.04	0.180*	0.233*	0.234*	0.091*	0.102*	0.103*	0.051*	0.053*	0.053*	0.043	0.043	0.044*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.077	0.092	0.012
0.05	0.195*	0.246*	0.247*	0.105*	0.116*	0.117*	0.062*	0.062*	0.064*	0.053	0.053	0.053
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.071	0.099	0.071
0.06	0.210*	0.256*	0.260*	0.118*	0.130*	0.131*	0.072*	0.073*	0.076*	0.062	0.062	0.064
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.177	0.264	0.060
0.07	0.221*	0.267*	0.270*	0.130*	0.142*	0.146*	0.084*	0.086*	0.086*	0.072	0.071	0.073
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.228	0.377	0.128
0.08	0.231*	0.276*	0.278*	0.143*	0.155*	0.159*	0.093*	0.095*	0.096*	0.083	0.083	0.084
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.105	0.169	0.065
0.09	0.240*	0.284*	0.287*	0.155*	0.167*	0.171*	0.103*	0.104*	0.105*	0.093	0.093	0.093
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.182	0.191	0.139
0.1	0.250*	0.293*	0.294*	0.167*	0.179*	0.182*	0.114*	0.114*	0.117*	0.101	0.100	0.102
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.421	0.447	0.242

This table presents actual rejection rates and the associated right-tail probability values for Z statistics computed on normally distributed data which exhibit proportionally higher variance during the event period $t = (-10, \ldots, +10)$. There is no excess kurtosis or skewness incorporated in the data, and the true market model coefficient does not change around the time of the event.

The Z statistics are calculated at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on the various Z statistics for various magnitudes of variance changes. Z^D is calculated according to Equation (2) and is assumed to follow a standard normal distribution. Z^{SR} is calculated according to Equation (7) and is also assumed to follow a standard normal distribution. Z^{TRAD} is calculated according to Equation (8) and is assumed to follow a Student t-distribution with 119 degrees of freedom.

The second through fourth columns of the table report results for an increase in event-period variance of 500%. The next set of three columns report results for an increase of 100%. Results for a 20% increase appear next, and results for a 5% increase appear in the final three columns. The top value in each cell is the actual rejection rate for a particular Z statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the test statistic's conventionally assumed distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.

Table 11: Z Statistics: Different Variance Changes Across Firms

**Actual Rejection Rates and
Right-Tail Probability Values
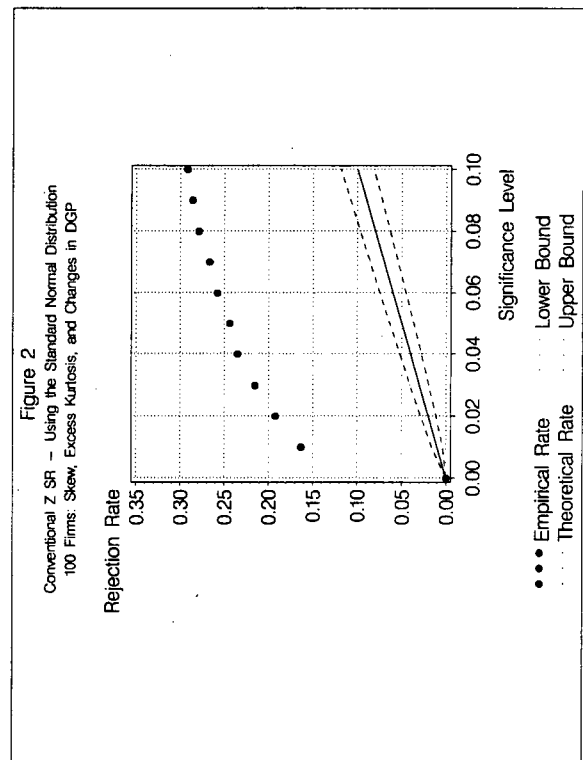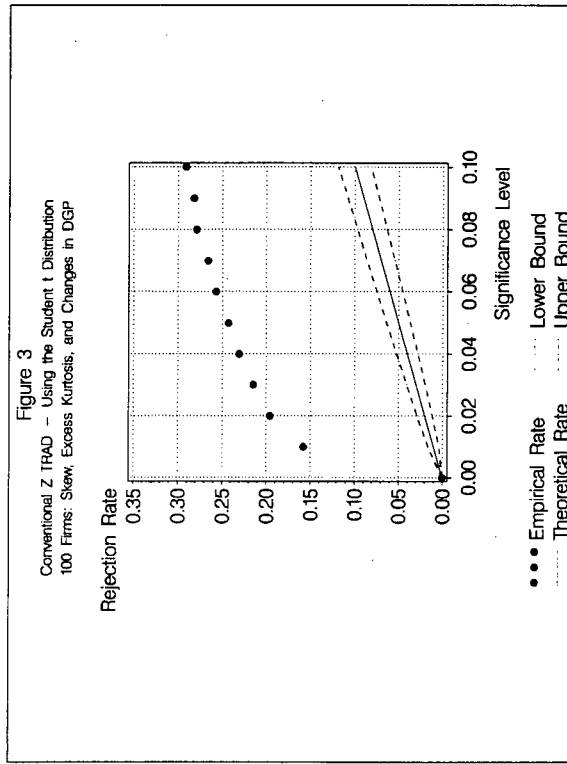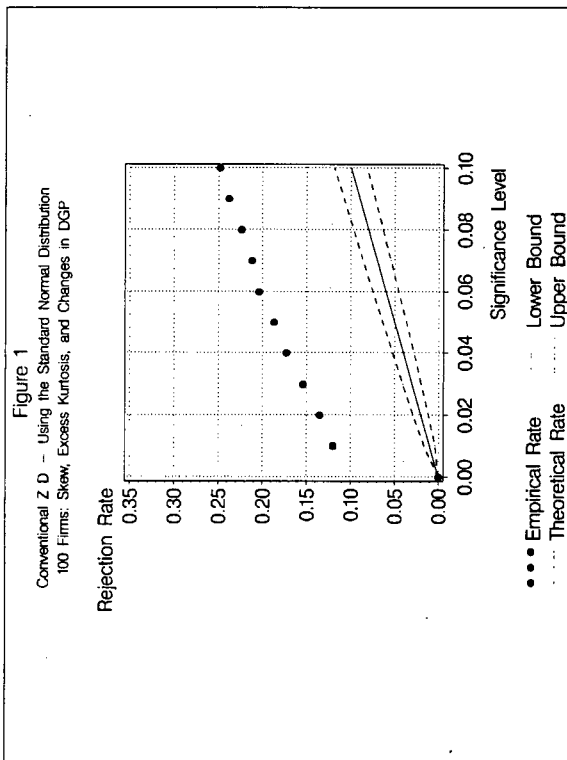for Tests at Common Significance Levels**

1000 Replications, 50 Firms

	Other Effects Present: Skewness, Kurtosis, and Slope Change						No Other Effects Present					
	400% to 500% Jump in Variance			100% to 500% Jump in Variance			400% to 500% Jump in Variance			100% to 500% Jump in Variance		
Size	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}	Z^D	Z^{SR}	Z^{TRAD}
0.01	0.119*	0.129*	0.183*	0.085*	0.088*	0.142*	0.116*	0.162*	0.170*	0.091*	0.129*	0.127*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.02	0.154*	0.161*	0.212*	0.117*	0.118*	0.177*	0.145*	0.190*	0.199*	0.119*	0.155*	0.153*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.03	0.178*	0.181*	0.237*	0.141*	0.136*	0.202*	0.163*	0.218*	0.220*	0.140*	0.176*	0.181*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.04	0.195*	0.203*	0.256*	0.161*	0.161*	0.213*	0.181*	0.229*	0.238*	0.155*	0.190*	0.196*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.05	0.209*	0.221*	0.272*	0.179*	0.182*	0.223*	0.197*	0.247*	0.251*	0.167*	0.204*	0.216*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.06	0.220*	0.238*	0.282*	0.194*	0.200*	0.246*	0.213*	0.261*	0.265*	0.182*	0.214*	0.223*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.07	0.232*	0.251*	0.292*	0.205*	0.212*	0.263*	0.220*	0.269*	0.277*	0.191*	0.228*	0.233*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.08	0.247*	0.264*	0.305*	0.214*	0.221*	0.273*	0.232*	0.282*	0.281*	0.204*	0.240*	0.247*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.09	0.256*	0.270*	0.318*	0.226*	0.230*	0.285*	0.240*	0.293*	0.289*	0.222*	0.248*	0.257*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
0.1	0.263*	0.278*	0.327*	0.236*	0.240*	0.292*	0.254*	0.300*	0.300*	0.228*	0.265*	0.268*
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

This table presents actual rejection rates and the associated right-tail probability values for Z statistics computed on data which incorporate changes in the data generating process during the event period. The degree of change is permitted to vary across firms. A sample size of 50 firms is employed.

The first set of three columns reports results for Z statistics computed on data which incorporate an increase in variance ranging from 400% to 500%. The next set of three columns is for a variance increase ranging from 100% to 500%. Data for both these sets of columns were generated using the Ramberg *et al.* [1979] algorithm. Skewness of 0.15 and kurtosis of 6.2 were incorporated in the data, as was a jump of 100% in the true market model parameter. The final two sets of three columns contain results for the Z statistics computed on data which again incorporate variance increases of 400% to 500% and 100% to 500% respectively, without excess kurtosis, skewness, or a change in the true market model parameter.

The Z statistics are calculated for tests at significance levels ranging from 0.01 to 0.10, as shown in the first column. Each of the remaining columns report statistics based on the various Z statistics for various sample sizes (*i.e.* number of firms). Z^D is calculated according to Equation (2) and is assumed to follow a standard normal distribution. Z^{SR} is calculated according to Equation (7) and is also assumed to follow a standard normal distribution. Z^{TRAD} is calculated according to Equation (8) and is assumed to follow a Student t-distribution with 119 degrees of freedom.

The top value in each cell is the actual rejection rate for a particular Z statistic computed at a particular significance level. The second value is the right-tail probability value (p-value) associated with that rejection rate. The p-value is calculated using the test statistic's conventionally assumed distribution, and it denotes the probability of observing a value greater than the actual rejection rate. A p-value greater than 0.975 or less than 0.025 indicates that the rejection rate is outside a 95% confidence interval and hence suggests a bias to size. Rejection rates which fall outside a 95% confidence interval are marked with an asterisk to denote significance.
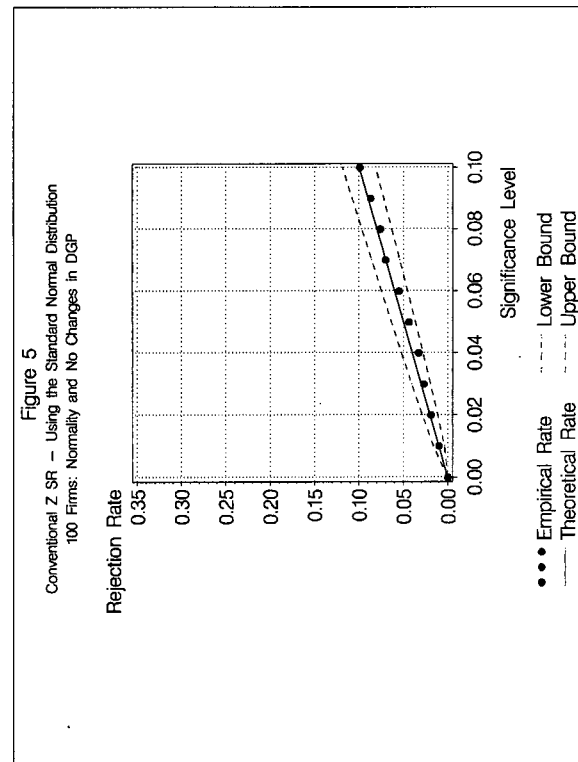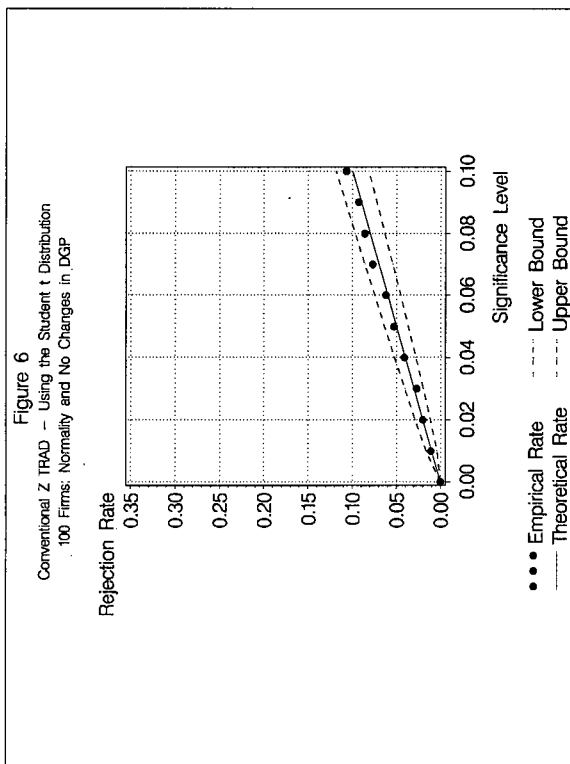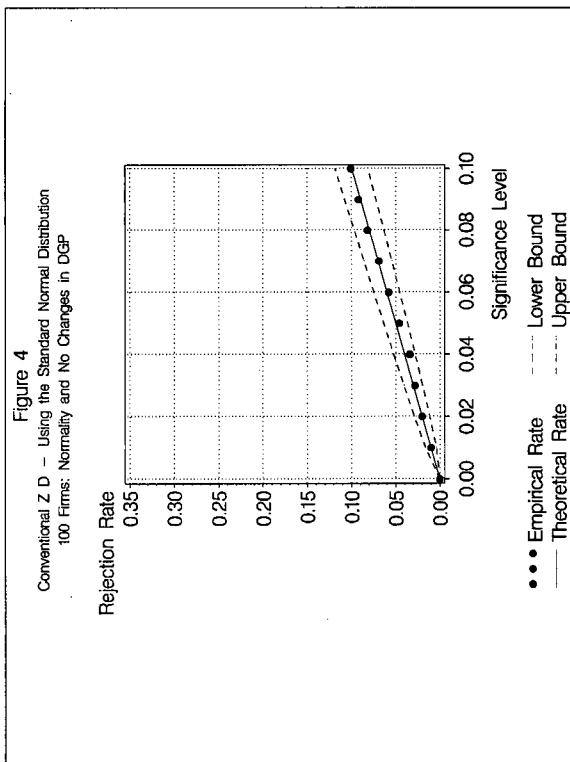
69

Figure 3
Conventional Z TRAD – Using the Student t Distribution
100 Firms: Skew, Excess Kurtosis, and Changes in DGP

Figure 1
Conventional Z D – Using the Standard Normal Distribution
100 Firms: Skew, Excess Kurtosis, and Changes in DGP

Figure 2
Conventional Z SR – Using the Standard Normal Distribution
100 Firms: Skew, Excess Kurtosis, and Changes in DGP

Figure 4
Conventional Z D — Using the Standard Normal Distribution
100 Firms: Normality and No Changes in DGP

Figure 5
Conventional Z SR — Using the Standard Normal Distribution
100 Firms: Normality and No Changes in DGP

Figure 6
Conventional Z TRAD — Using the Student t Distribution
100 Firms: Normality and No Changes in DGP

71

Figure 7
Normalized Z D — Using the Bootstrap Distribution
50 Firms: Skew, Excess Kurtosis, and Changes in DGP
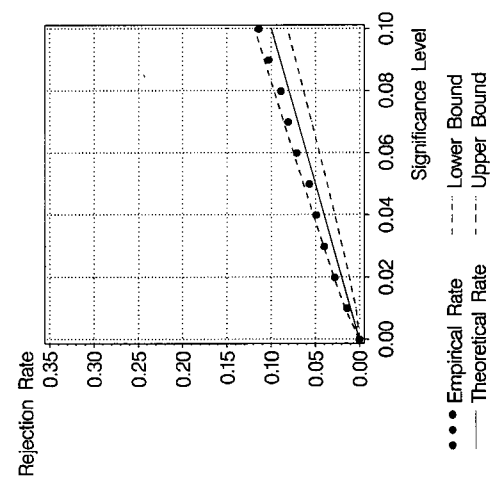
Figure 9
Normalized Z TRAD — Using the Bootstrap Distribution
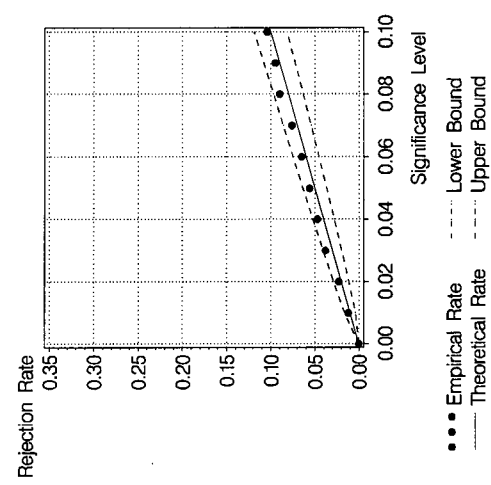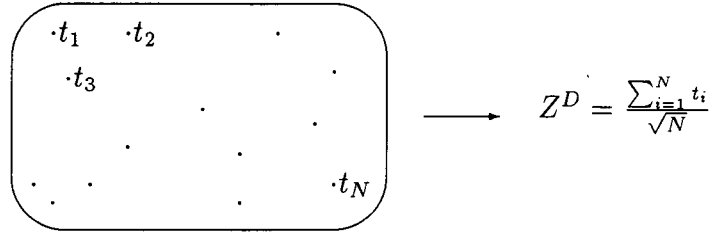50 Firms: Skew, Excess Kurtosis, and Changes in DGP

Figure 8
Normalized Z SR — Using the Bootstrap Distribution
50 Firms: Skew, Excess Kurtosis, and Changes in DGP

FIGURE 10: The Bootstrap Approach

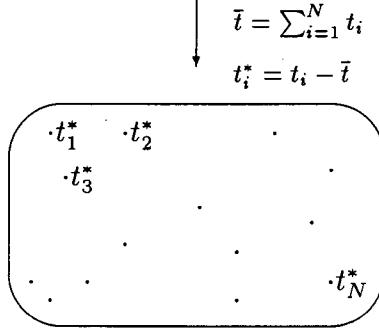1. Compute the conventional Z^D statistic

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

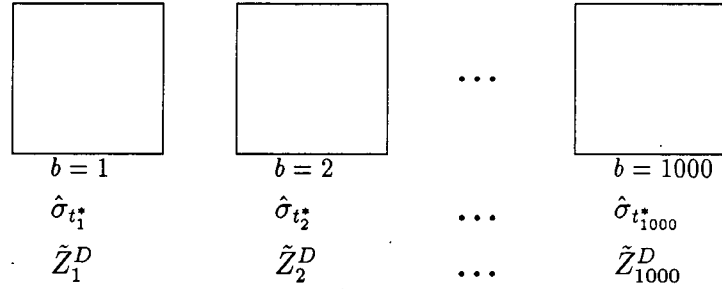2. Normalize the conventional Z^D statistic using
   $\hat{\sigma}_t$ = standard deviation of the t_i

$$\hat{\sigma}_t \longrightarrow \tilde{Z}^D = \frac{Z^D}{\hat{\sigma}_t}$$

3. Mean-adjust the t_i statistics

$$\bar{t} = \sum_{i=1}^{N} t_i$$
$$t_i^* = t_i - \bar{t}$$

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

Randomly resample
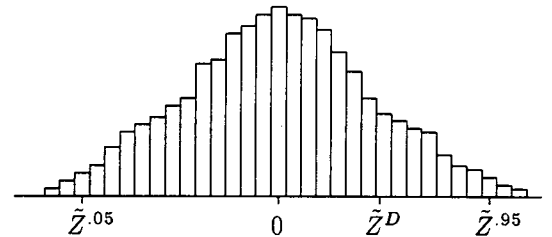N values

$b = 1$ $b = 2$ $\cdots$ $b = 1000$

5. Compute $\tilde{Z}_b^D$ for each sample
   using $\hat{\sigma}_{t_b^*}$:

$$\tilde{Z}_b^D = \frac{\sum_{j=1}^{N} t_{bj}^*}{\hat{\sigma}_{t_b^*}}$$

$\hat{\sigma}_{t_1^*}$ $\hat{\sigma}_{t_2^*}$ $\cdots$ $\hat{\sigma}_{t_{1000}^*}$

$\tilde{Z}_1^D$ $\tilde{Z}_2^D$ $\cdots$ $\tilde{Z}_{1000}^D$

6. Build the empirical distribution from the 1000
   values of $\tilde{Z}_b^D$ and conduct inference

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

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

$\tilde{Z}^{.05}$ 0 $\tilde{Z}^D$ $\tilde{Z}^{.95}$

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

73