JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS VOL. 27, NO. 3, SEPTEMBER 1992

The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns

Charles J. Corrado and Terry L. Zivney*

Abstract

This paper evaluates a nonparametric sign test for abnormal security price performance in event studies. The sign test statistic examined here does not require a symmetrical distribution of security excess returns for correct specification. Sign test performance is compared to a parametric t-test and a nonparametric rank test. Simulations with daily security return data show that the sign test is better specified under the null hypothesis and often more powerful under the alternative hypothesis than a t-test. The performance of the sign test is dominated by the performance of a rank test, however, indicating that the rank test is preferable to the sign test in obtaining nonparametric inferences concerning abnormal security price performance in event studies.

I. Introduction

In financial event studies, a sign test is commonly used to specify statistical significance independently of an assumption concerning the distribution of the excess return population from which data are collected. Seemingly a completely nonparametric procedure, nevertheless, a sign test can be misspecified if an incorrect assumption about the data is imposed. For example, Jain ((1986), p. 88) reports that 53 percent of a sample of five million excess returns is negative. Consequently, Brown and Warner (1980), (1985) and Berry, Gallinger, and Henderson (1990) demonstrate that a sign test assuming an excess return median of zero is misspecified.

This paper studies the specification and power of a sign test that does not assume a median of zero, but instead uses a sample excess return median to calculate the sign of an event date excess return. This version of the sign test is expected to be correctly specified no matter how skewed the distribution of security excess returns and may be efficient compared to a t-test for distributions

*College of Business and Public Administration, University of Missouri, 214 Middlebush Hall, Columbia, MO 65211, and School of Business Administration, University of Tennessee, 615 McCallie Avenue, Chattanooga, TN 37403, respectively. The authors thank **JFQA** referee Robert Connolly for helpful comments.

with heavier tailweights than the normal distribution.[1] Sign test performance is compared with two other event study procedures: a parametric t-test and a nonparametric rank test. These provide benchmarks against which the sign test is compared.

In addition to a comparative evaluation of the three test procedures discussed above, this paper examines two relevant and related issues in event study methodology. The first issue is the effect of short estimation periods on the performance of test statistics commonly used in event studies. This may be important when market model parameter instability is suspected and, as a result, a researcher prefers to use a relatively short estimation period. The second, and related issue, is the effect of a prediction error variance correction on test statistic performance. A prediction error correction may become important as the estimation period shortens; we examine how important the correction procedure might become.[2]

The organization of the paper is as follows. Section II presents the test procedures and test statistics used in this study. Sample construction is described in Section III. Empirical results are presented in Section IV. Section V provides a summary and conclusion.

II. Test Procedures and Test Statistics

We examine tests of the null hypothesis that the shift in the distribution of event date excess returns is zero. Simulation experiments, as in Brown and Warner (1985), are used where the excess return measure is the residual from the standard market model.[3] In each experiment, securities and event dates are randomly selected and a portfolio is formed. A 250-day sample period surrounds each event date. An event date is defined as day 0, and days −244 through −6 comprise a 239-day estimation period from which market model parameters are obtained.

In addition to simulations using 250-day sample periods, we also report the effect on test statistic performance from using shorter sample periods of 100 days and 50 days, where days −94 through −6 and days −44 through −6 form 89-day and 39-day estimation periods, respectively.

We compare the performance of a sign test with two alternative tests: a t-test and a rank test. These two benchmark statistics are T_2 and T_3, respectively, as specified in Corrado (1989). We first review test statistic construction for T_2 and T_3 and, immediately following, we specify construction of the sign test statistic, denoted by T_4.

[1]The efficiency of the sign test in the presence of normal and nonnormal distributions is discussed in Hettmansperger (1984) and Lehmann (1975), (1986). Zivney and Thompson (1989) suggest that a properly specified sign test may provide a more powerful test for abnormal security price performance in event studies than the t-test.

[2]The prediction error correction procedure is discussed in many statistics and econometrics texts. A suggested reference is Maddala (1977).

[3]Brown and Warner ((1985), pp. 210–211) discuss the importance of simulations in evaluating event study methods.

A. t-test (T_2)

Let A_{it} represent the excess return of security i on day t. Each excess return is divided by its estimated standard deviation to yield a standardized excess return,

$$(1) \qquad A'_{it} = A_{it}/S(A_i),$$

where the standard deviation is calculated as [4]

$$(2) \qquad S(A_i) = \sqrt{\frac{1}{238}\sum_{t=-244}^{-6} A_{it}^2}.$$

The day 0 test statistic is given by

$$(3) \qquad T_2 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} A'_{i0},$$

where N is the number of securities in the sample portfolio.

B. Rank Test (T_3)

Let K_{it} denote the rank of the excess return A_{it} in security i's 250-day time series of excess returns,

$$(4) \qquad K_{it} = rank\ (A_{it}), \qquad t = -244, \ldots, +5.$$

To allow for missing returns, ranks are standardized by dividing by one plus the number of nonmissing returns in each firm's excess returns time series,

$$(5) \qquad U_{it} = K_{it}/(1 + M_i),$$

where M_i is the number of nonmissing returns for security i. This yields order statistics for the uniform distribution with an expected value of one-half.[5] The rank test statistic substitutes $(U_{it} - 1/2)$ for the excess return A_{it}, yielding this day 0 test statistic,[6]

$$(6) \qquad T_3 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(U_{i0} - 1/2)/S(U).$$

[4]For estimation periods of 89 days and 39 days, the standard deviation $S(A_i)$ is calculated as

$$\sqrt{\frac{1}{88}\sum_{t=-94}^{-6} A_{it}^2} \quad \text{and} \quad \sqrt{\frac{1}{38}\sum_{t=-44}^{-6} A_{it}^2}, \quad \text{respectively,}$$

[5]A discussion of order statistics is contained in Lehmann (1986).

[6]If there are no missing returns, i.e., M_i the same for all i, then this statistic is identical to T_3 as specified in Corrado ((1989), p. 388). Without an adjustment for missing returns, the rank test may be misspecified. Corrado ((1989), p. 389) avoids this problem by restricting event date selection to allow a 250-day sample for each security.

The standard deviation $S(U)$ is calculated using the entire sample period,

$$(7) \qquad S(U) = \sqrt{\frac{1}{250}\sum_{t=-244}^{+5}\left(\frac{1}{\sqrt{N_t}}\sum_{i=1}^{N_t}(U_{it}-1/2)\right)^2},$$

where N_t represents the number of nonmissing returns in the cross-section of N-firms on day t in event time.

C. Sign Test (T_4)

Let the median excess return in security i's time series of excess returns be denoted by *median* (A_i). For each day in the sample period, the sign of each excess return is calculated as

$$(8) \qquad G_{it} = sign(A_{it} - median\ (A_i)) \qquad t = -244, \ldots, +5,$$

where $sign(x)$ is equal to $+1$, -1, or 0 as x is positive, negative, or zero, respectively. From the signs G_{it}, this day 0 test statistic is constructed

$$(9) \qquad T_4 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} G_{i0} / S(G).$$

The standard deviation $S(G)$ is calculated using the entire sample period,[7]

$$(10) \qquad S(G) = \sqrt{\frac{1}{250}\sum_{t=-244}^{+5}\left(\frac{1}{\sqrt{N_t}}\sum_{i=1}^{N_t} G_{it}\right)^2},$$

where N_t is the number of nonmissing returns in the cross-section of N-firms on day t in event time.

The sign test procedure specified above transforms security excess returns into sign values where the probability of a value of $+1$ is equal to the probability of a value of -1 regardless of any asymmetry in the original distribution. This procedure precludes the misspecification documented by Brown and Warner (1980), (1985) and Berry, Gallinger, and Henderson (1990) of a sign test that assumes an excess return median of zero.

III. Sample Construction

From the Center for Research in Security Prices (CRSP) Daily Return Files, we obtain daily return data for 600 firms. All firms are listed over the entire period from July 1962 through December 1986 and none have more than ten missing returns. From this data base, we construct 1,200 portfolios each of 10 and 50 securities. Each time a security is selected for inclusion in a portfolio, a hypothetical event date is randomly generated. Securities and event dates are randomly selected with replacement.

[7]For sample periods of 100 days and 50 days, the standard deviations $S(U_i)$ and $S(G_i)$ are calculated by summation from day -94 through day $+5$, and day -44 through day $+5$, respectively.

IV. Empirical Results

A. Test Statistics with No Abnormal Performance

1. Test Statistics without a Variance Correction for Prediction Error

For sufficiently large sample sizes, the Central Limit Theorem implies that the distribution of each test statistic will converge to normality. We examine the completeness of this convergence for portfolio sizes of 10 and 50 securities. Table 1 summarizes the empirical distributions of the three test statistics without a variance correction for prediction error by reporting the first four sample moments and the studentized range of each statistic for portfolio sizes of 10 and 50 securities and sample periods of 50, 100, and 250 days.

TABLE 1

Summary Measures of the Distribution of Test Statistics with No Abnormal Performance

	OLS Market Model Adjusted Excess Returns				
Test Statistic	Mean	Standard Deviation	Skewness	Kurtosis	Studentized Range
250-Day Sample Period, Portfolio Size = 50					
T_2	−0.022	1.077	0.133	3.634*	7.375
T_3	−0.028	1.015	−0.016	2.979	6.488
T_4	−0.014	1.008	−0.064	2.887	6 259
Portfolio Size = 10					
T_2	−0.013	1.053	0.261*	3 732*	8.129*
T_3	−0.037	1.007	−0.006	2.917	6.341
T_4	−0.049	1.012	−0.049	2.776	6.167
100-Day Sample Period, Portfolio Size = 50					
T_2	−0.035	1.098	0.226*	3.936*	8.949*
T_3	−0.046	1.009	−0.014	2.931	6.540
T_4	−0.051	1.008	−0.074	2.852	6.531
Portfolio Size = 10					
T_2	−0.028	1.060	0.231*	3.701*	7 919*
T_3	−0.042	0.996	0.008	2.852	6.671
T_4	−0.052	1 008	−0.034	2.688	5.800
50-Day Sample Period, Portfolio Size = 50					
T_2	−0.029	1 136	0 276*	4.441*	9 949*
T_3	−0.033	1.015	−0.016	2.847	6.199
T_4	−0.016	0.992	−0.079	2.866	6.464
Portfolio Size = 10					
T_2	−0.039	1.091	0.305*	4.152*	8.991*
T_3	−0.052	0.992	−0.004	2.736	5.844
T_4	−0.066	1.004	−0.039	2.590	5.703

Test statistic distribution measures based on 1,200 simulation experiments: portfolio sizes = 10, 50 securities, sample periods = 50, 100, 250 days. Randomly selected daily stock returns and event dates over the period 1962 through 1986. T_2 is a parametric *t*-test statistic, T_3 is a nonparametric rank test statistic, and T_4 is a nonparametric sign test statistic.

*Significant at 99-percent confidence level.

The distributions of the *t*-test statistic and the rank test statistic reported in Table 1 for a 250-day sample period are similar to those reported in Brown and

Warner (1985) and Corrado (1989). Compared to a standard normal distribution, the t-test statistic distribution is significantly positively skewed with a coefficient of skewness of 0.261 for portfolio sizes of 10 securities, but somewhat less skewed with portfolio sizes of 50 securities where the coefficient of skewness is 0.133. The distribution of the t-test statistic is significantly kurtotic relative to a normal distribution for all portfolio sizes and sample periods where the smallest reported coefficient of kurtosis is 3.634.[8] By contrast, the rank test statistic and the sign test statistic sample moments are all close to those expected from a standard normal population.

The significant kurtosis reported for the t-test suggests that the distribution of the t-test statistic deviates from normality in the tails of the distribution. Since statistical inferences are based on tailweight probabilities for a normal distribution, the t-test might yield biased inferences in event studies using daily stock returns. To assess the potential severity of this bias, Table 2 summarizes the tailweight specification of the three test statistics by reporting rejection rates in nominal 5-percent level and 1-percent level, upper-tail and lower-tail tests for portfolio sizes of 10 and 50 securities and sample periods of 50, 100, and 250 days.

TABLE 2

Rejection Rates of Null Hypothesis with No Abnormal Performance

	Nominal Test Levels							
	Portfolio Size = 50				Portfolio Size = 10			
	Lower Tail		Upper Tail		Lower Tail		Upper Tail	
	5%	1%	5%	1%	5%	1%	5%	1%
250-Day Sample Period								
T_2	5.6%	1.3%	6.1%	2.1%*	5.4%	1.4%	6.5%	2.0%*
T_3	5.8	1.3	4.8	1.2	5.8	1.0	4.3	1.2
T_4	6.4	1.3	5.3	0.7	6.6	1.6	4.8	0.9
100-Day Sample Period								
T_2	6.5%	1.3%	6.1%	2.3%*	5.3%	1.7%	6.6%	1.8%*
T_3	5.7	1.0	4.4	1.1	5.3	0.9	4.0	1.1
T_4	6.3	1.3	3.8	0.6	6.3	1.0	5.0	0.8
50-Day Sample Period								
T_2	7.2%*	1.6%	6.8%*	2.3%*	6.5%	1.8%*	6.3%	2.1%*
T_3	6.1	1.2	4.5	1.1	5.9	1.1	4.3	0.8
T_4	5.0	1.1	4.2	0.6	6.2	1.1	4.5	0.3

Rejection rates based on 1,200 simulation experiments: portfolio sizes = 10, 50 securities, sample periods = 50, 100, 250 days. Randomly selected securities and event dates over the period 1962 through 1986. T_2 is a parametric t-test statistic, T_3 is a nonparametric rank test statistic, and T_4 is a nonparametric sign test statistic.

* Significant at 99-percent confidence level.

[8]Normal population skewness and kurtosis coefficients are 0 and 3, respectively. Pearson and Hartley ((1966), pp. 207–208) provide 99-percent critical values of 0.165 and 3.37 for skewness and kurtosis coefficients, respectively, for samples of size 1200 from a normal population. The studentized range 99-percent critical value of 7.80 is obtained from Fama ((1976), p. 40).

The results reported in Table 2 indicate that the t-test is well specified in upper-tail and lower-tail 5-percent level tests with sample periods of 100 days and 250 days; the largest deviation from a correct 5-percent rejection rate is a rate of 6.6 percent obtained with a portfolio size of 10 and a 100-day sample period. With a short 50-day sample period (39-day estimation period), t-test specification deteriorates slightly with a tendency to reject the null hypothesis too often.[9] Since sample periods in event studies are usually longer than 100 days, the parametric t-test provides reliable test specification in 5-percent level tests even with portfolios of as few as 10 securities.[10]

The adverse effects of the significant skewness and kurtosis (reported in Table 1) on t-test specification only became apparent in 1-percent level tests. The t-test yields biased inferences in upper-tail 1-percent level tests where it typically rejects the null hypothesis at rates of 2 percent or more.[11] In lower-tail 1-percent level tests, the t-test is somewhat better specified where the largest rejection rate is 1.8 percent. By contrast, both the rank test and the sign test are well specified in upper-tail and lower-tail 5-percent level and 1-percent level tests for all portfolio sizes and sample periods examined here.

2. Test Statistics with a Variance Correction for Prediction Error

In event studies using market model excess returns, the day 0 excess returns variance is sometimes corrected for prediction error. Although theoretically correct, the importance of a prediction correction declines as the estimation period lengthens. Using 239-day estimation periods, Brown and Warner ((1985), p. 8) and Corrado ((1989), p. 387) both report that a variance correction for prediction error did not discernibly affect the results of their simulation experiments. To assess the impact of a prediction correction with shorter estimation periods, in Table 3, we report test statistic specification with no abnormal performance for portfolios of 10 securities using 89-day and 39-day estimation periods, corresponding to 100-day and 50-day sample periods, respectively.[12]

Comparing test specification with a prediction error variance correction reported in Table 3 with test specification without a prediction correction reported in Tables 1 and 2, we see that t-test specification is improved slightly by the correction procedure; the coefficients of skewness and kurtosis are smaller, although they still deviate significantly from the expected coefficients for a normal distribution. Also, rejection rates for the t-test are closer to correct levels with a prediction correction. The improvement is discernible in experiments using a

[9]Significant deviations from correct specification can be ascertained by the critical values

$$c = \alpha \pm 2.575\sqrt{\alpha(1-\alpha)/n},$$

where α is the correctly specified rejection rate, and 2.575 is the 99.5th percentile of the standard normal distribution. With a sample size of $n = 1200$, for $\alpha = 5$ percent, these critical values are 5 percent ± 1.62 percent, and, for $\alpha = 1$ percent, they are 1 percent ± 0.74 percent.

[10]Using 239-day estimation periods, Brown and Warner ((1985), p. 14) report reliable t-test specification in 5-percent level tests for portfolios containing as few as 5 securities.

[11]Brown and Warner ((1985), p. 14) caution that because the empirical distribution of the t-test statistic is kurtotic relative to a normal distribution, ". . . significance levels should not be taken literally."

[12]The effect of a prediction error correction for portfolios of 50 securities was also examined; we found similar, but less noticeable effects.

TABLE 3

Summary Measures of Test Statistic Performance with Variance Correction for Prediction Error; No Abnormal Performance

Test Statistic	OLS Market-Model Adjusted Excess Returns: Mean	Standard Deviation	Skewness	Kurtosis	Studentized Range
100-Day Sample Period, Portfolio Size = 10					
T_2	−0.026	1.045	0.221*	3.664*	8.019*
T_3	−0.042	0.996	0.009	2.844	5.811
T_4	−0.052	1.006	−0.038	2.677	6.635
50-Day Sample Period, Portfolio Size = 10					
T_2	−0.036	1.056	0.229*	3.717*	8.054*
T_3	−0.049	0.991	−0.050	2.735	5.926
T_4	−0.067	1.007	−0.006	2.575	5.685

Nominal Test Levels; Portfolio Size = 10

	50-Day Sample Period: Lower Tail 5%	Lower Tail 1%	Upper Tail 5%	Upper Tail 1%	100-Day Sample Period: Lower Tail 5%	Lower Tail 1%	Upper Tail 5%	Upper Tail 1%
T_2	5.2%	1.3%	6.0%	1.8%*	5.2%	1.3%	6.3%	1.8%*
T_3	5.3	0.9	4.3	0.3	5.3	0.9	4.1	1.1
T_4	6.3	1.0	4.3	0.8	6.3	1.0	5.0	0.8

Test statistic distribution measures based on 1,200 simulation experiments: portfolio size = 10 securities; sample periods = 50, 100 days. Randomly selected daily stock returns and event dates over the period 1962 through 1986. T_2 is a parametric t-test statistic, T_3 is a nonparametric rank test statistic, and T_4 is a nonparametric sign test statistic.

*Significant at 99-percent confidence level.

50-day sample period, but quite small in experiments using a 100-day sample period. The specification of the rank test statistic and the sign test statistic are not noticeably affected by the prediction error correction procedure.

In simulation results not reported here, the power of the parametric t-test in detecting abnormal security price performance was typically slightly reduced by a prediction error variance correction, but only by trivial magnitudes. We conclude that for sample periods longer than 100 days, the presence or absence of a variance correction for prediction error does not materially affect statistical inferences in event studies.

B. Test Statistics with Abnormal Performance

We now assess the ability of the three test statistics to detect abnormal performance in the day 0 excess returns distribution. As in Brown and Warner (1985), abnormal security price performance is simulated by adding a constant to the day 0 return of each security.[13]

[13]Brown and Warner ((1980), p. 212) argue that detecting mean shifts is the relevant phenomenon when comparing event study test procedures.

Table 4 reports rejection rates of the null hypothesis at the 5-percent and 1-percent test levels for abnormal performance levels of ±1/2 percent and ±1 percent, portfolio sizes of 10 and 50 securities, and sample periods of 50 days, 100 days, and 250 days.

TABLE 4

Rejection Rates of Null Hypothesis with Abnormal Performance of ± 1/2 percent and ± 1 percent

	Abnormal Performance							
	−1%		−1/2%		+1/2%		+1%	
Test Levels	5%	1%	5%	1%	5%	1%	5%	1%
250-Day Sample Period, Portfolio Size = 50								
T_2	98.3%	94.1%	65.0%	39.0%	63.5%	36.4%	98.6%	93.8%
T_3	99.3	97.2	76.0	50.3	75.6	48.7	99.8	97.7
T_4	98.3	90.3	72.3	43.9	68.2	39.9	98.5	91.7
Portfolio Size = 10								
T_2	57.4%	30.4%	23.3%	8.2%	21.6%	8.8%	53.5%	28.9%
T_3	64.1	36.6	29.3	9.6	25.9	7.8	63.3	33.3
T_4	56.1	25.3	29.4	8.8	25.8	8.1	53.2	24.7
100-Day Sample Period, Portfolio Size = 50								
T_2	98.5%	94.9%	66.8%	41.8%	63.0%	37.8%	98.5%	94.8%
T_3	99.6	97.2	75.9	50.8	76.3	49.2	99.8	97.8
T_4	98.3	89.6	72.3	43.8	67.9	38.7	98.8	91.6
Portfolio Size = 10								
T_2	59.6%	32.3%	24.9%	8.3%	22.3%	9.0%	53.7%	29.5%
T_3	63.7	35.3	30.3	9.4	27.3	8.0	63.2	32.5
T_4	55.5	22.4	28.8	7.9	25.1	6.8	52.8	21.3
50-Day Sample Period, Portfolio Size = 50								
T_2	98.7%	95.0%	68.9%	43.1%	65.7%	40.6%	98.9%	94.8%
T_3	99.4	96.6	75.1	48.3	76.9	47.0	99.5	96.6
T_4	98.0	87.9	69.3	39.6	67.2	37.4	97.6	89.3
Portfolio Size = 10								
T_2	61.5%	34.8%	27.0%	9.2%	23.2%	9.7%	55.0%	31.5%
T_3	63.4	32.6	28.8	9.7	27.5	7.0	61.0	30.8
T_4	53.1	19.8	27.4	6.6	24.9	4.7	51.2	17.2

Rejection rates based on 1,200 simulation experiments: portfolio sizes = 10, 50 securities; sample periods = 50, 100, 250 days. Randomly selected securities and event dates over the period 1962 through 1986. T_2 is a parametric t-test statistic, T_3 is a nonparametric rank test statistic, and T_4 is a nonparametric sign test statistic.

1. ±1-Percent Abnormal Performance, 5-Percent Level Tests

With ±1-percent abnormal performance introduced, the rank test is more powerful than the t-test, and the t-test is more powerful than the sign test. With +1-percent abnormal performance, across 50-day, 100-day, and 250-day sample periods, in upper-tail tests for portfolios of size 10, rejection rate averages are 54.1 percent, 62.5 percent, and 52.4 percent for the t-test, rank test, and sign test, respectively.[14] With −1-percent abnormal performance, in lower-tail tests across the three sample periods for portfolios of size 10, rejection rate averages are 59.5 percent, 63.7 percent, and 54.9 percent for the t-test, rank test, and sign test, respectively.

2. ±1/2-Percent Abnormal Performance, 5-Percent Level Tests

With ±1/2-percent abnormal performance introduced, the rank test dominates the sign test, and the sign test dominates the t-test. With +1/2-percent abnormal performance, across 50-day, 100-day, and 250-day sample periods, in upper-tail tests for portfolios of size 50, the rejection rate averages are 64.1 percent, 76.3 percent, and 67.8 percent for the t-test, rank test, and sign test, respectively. With −1/2-percent abnormal performance introduced, across the three sample periods, in lower-tail tests for portfolios of size 50, rejection rate averages are 66.9 percent, 75.7 percent, and 71.3 percent for the t-test, rank test, and sign test, respectively.

C. Test Statistics with a Day 0 Variance Increase

Brown and Warner (1985) and Corrado (1989) show that the parametric t-test statistic is vulnerable to misspecification caused by an increase in the variance of the distribution of event date excess returns. To compare the specification and power of the three test statistics in the presence of a day 0 variance increase, we follow Brown and Warner's (1985) procedure of transforming each security's day 0 excess return by summing a day 0 excess return and an excess return from outside the sample period,

$$A_{i0}^{*} = A_{i0} + A_{i6}. \qquad (11)$$

The transformed excess return A_{i0}^{*} replaces the original excess return A_{i0} when computing test statistics with a simulated variance increase.

1. Test Statistics with a Cross-Sectional Variance Adjustment

Brown and Warner (1985) and Corrado (1989) show that a t-test procedure that solely relies on a cross-sectional standard deviation is not very powerful in detecting abnormal security price performance.[15] Recently, however, Boehmer, Musumeci, and Poulsen (1991) and Sanders and Robins (1991) independently demonstrate that a simple cross-sectional variance adjustment, applied after controlling for cross-sectional heteroskedasticity, yields a well-specified t-test when there actually is an event date variance increase, and importantly, does not noticeably affect t-test power when there is no variance increase. The cross-sectionally adjusted t-test statistic has the following form,

$$T_2 \text{ (adjusted)} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} A'_{i0} / S(A'_0), \qquad (12)$$

[14]For example, the 54.1-percent average for the t-test is obtained as the average of 53.5 percent, 53.7 percent, and 55.0 percent, for sample periods of 250, 100, and 50 days, respectively.

[15]This statistic is described in Brown and Warner ((1985), pp. 7–8) and Corrado ((1989), pp. 386–387), and is T_1 in Corrado's notation.

where the standardized returns A'_{it} are defined in Equation (1) and the day 0 cross-sectional standard deviation is calculated as

$$(13) \qquad S(A'_0) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(A'_{i0} - \overline{A}'_0)^2}.$$

We shall here examine the impact of a cross-sectional variance adjustment on the specification and power of the sign test and the rank test, with the adjusted t-test as a benchmark. For the sign test and the rank test, a cross-sectional variance adjustment may be implemented as described immediately below.

Based on the standardized excess returns A'_{it}, we define the following standardized excess returns series,

$$(14) \qquad X_{it} = \begin{cases} A'_{it} & t \neq 0 \\ A'_{it}/S(A'_0) & t = 0, \end{cases}$$

where the day 0 cross-sectional standard deviation $S(A'_0)$ is defined in Equation (13).

A cross-sectional variance-adjusted rank test is obtained by first dividing the ranks of the excess returns defined in Equation (14) by one plus the number of nonmissing returns (as specified in Equation (5)),

$$(15) \qquad U_{it} = rank\ (X_{it})/(1 + M_i),$$

and then proceeding to calculate the rank test statistic as specified in Equation (6), which is reproduced here for convenience,

$$(6) \qquad T_3 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(U_{i0} - 1/2)/S(U).$$

A cross-sectional variance-adjusted sign test is obtained by defining the signs of the excess returns in Equation (14) as follows,

$$(16) \qquad G_{it} = sign\ (X_{it} - median\ (X_i)) \qquad t = -244, \ldots, +5,$$

and then proceeding to calculate a sign test statistic as specified in Equation (9), which is reproduced here,

$$(9) \qquad T_4 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} G_{i0}/S(G).$$

To assess the relative performance of the three test statistics, both before and after a cross-sectional variance adjustment, the day 0 variance is doubled by transforming the day 0 excess returns as specified in Equation (11) above. Table 5 reports the results of simulation experiments for portfolios of 50 securities with abnormal performance levels of 0 percent, ±1/2 percent, and ±1 percent with 100-day and 250-day sample periods.

TABLE 5

Rejection Rates with a Day 0 Variance Doubling, with and without a Cross-Sectional Variance Adjustment

Test Statistic	Abnormal Performance	With Adjustment Lower Tail	With Adjustment Upper Tail	Without Adjustment Lower Tail	Without Adjustment Upper Tail
250-Day Sample Period					
T_2	0%	6.4%	3.9%	14.7%	12.1%
T_3		6.8	4.5	12.0	5.4
T_4		6.8	4.3	7.8	3.3
T_2	−1/2% / +1/2%	40.9%	36.4%	60.9%	57.8%
T_3		46.0	47.5	62.2	51.1
T_4		40.5	39.9	44.1	35.6
T_2	−1% / +1%	85.1%	83.1%	94.8%	94.4%
T_3		89.1	90.9	95.8	92.5
T_4		81.8	83.9	84.3	81.3
100-Day Sample Period					
T_2	0%	6.7%	3.5%	15.4%	11.8%
T_3		6.5	4.5	12.7	4.6
T_4		6.5	4.0	8.0	2.5
T_2	−1/2% / +1/2%	43.2%	37.3%	62.4%	57.8%
T_3		46.8	47.2	61.8	51.4
T_4		39.5	38.8	41.6	35.9
T_2	−1% / +1%	85.7%	83.1%	95.2%	93.8%
T_3		88.8	90.6	95.3	92.4
T_4		81.3	84.3	83.4	81.3

Test statistic performance results based on 1,200 simulation experiments with abnormal performance of 0 percent, ±1/2 percent, ±1 percent; portfolio size = 50 securities; sample periods = 100, 250 days. T_2 is a parametric t-test statistic, T_3 is a nonparametric rank test statistic, and T_4 is a nonparametric sign test statistic. Rejection rates with negative abnormal performance are reported for lower-tail 5-percent tests and rejection rates with positive abnormal performance are reported for upper-tail 5-percent level tests.

As shown in Table 5, all three test statistics display some misspecification without a cross-sectional variance adjustment, but the t-test is the most misspecified. The specification of all three test statistics is improved by a day 0 cross-sectional variance adjustment in the presence of a day 0 variance increase. Since the t-test was the most misspecified as a result of an event date variance increase, the improvement is most remarkable for the t-test. For example, with no abnormal performance and a 250-day sample period, the lower-tail rejection rate of 14.7 percent for the t-test without a cross-sectional variance adjustment is reduced to 6.4 percent with the variance adjustment.

With abnormal performance and a doubled day 0 variance, among the three test statistics, the t-test is most affected by a cross-sectional variance adjustment. For example, with +1/2-percent abnormal performance and a 250-day sample period, the t-test rejection rate is 57.8 percent without a cross-sectional variance adjustment and 36.4 percent with the variance adjustment. With the same abnormal performance and sample period, the rank test rejection rate is 51.1 percent without, and 47.5 percent with the variance adjustment. The corresponding sign test rejection rates are 35.6 percent, and 39.9 percent with the variance adjustment.

When an event date variance increase is likely, correct specification for the t-test requires that a cross-sectional variance adjustment be implemented.[16] For the rank test, in contrast, a variance adjustment appears to be unimportant in tests for positive abnormal performance, but necessary in tests for negative abnormal performance. Sign test specification and power is only slightly improved by a cross-sectional variance adjustment. After the variance correction is applied, sign test and t-test power are comparable, but both are dominated by the rank test.

D. Summary and Conclusions

We study the specification and power of a nonparametric sign test for abnormal security price performance in event studies that does not require a median of zero in the distribution of security excess returns for correct specification. The performance of the sign test is compared with a parametric t-test and a nonparametric rank test. The sign test is shown to be better specified than the t-test under a complete null hypothesis of no abnormal performance and no variance increase. Both the sign test and the rank test are equally well specified under this complete null hypothesis. In the presence of an event date variance increase, all three test statistics display some misspecification, but the misspecification is most severe for the t-test.

When abnormal performance is present, sign test power is greater than that of a t-test in detecting ±1/2-percent abnormal performance, but less than that of a t-test in detecting ±1-percent abnormal performance. The rank test dominates both the sign test and the t-test in detecting both ±1/2-percent and ±1-percent abnormal performance.

The effect on test statistic performance from using short estimation periods to obtain market model parameters was examined. With estimation periods as short as 89 days, the performance of all three test statistics was virtually unaffected. With 39-day estimation periods, only a slight deterioration in test performance was noticed. Similarly, virtually no improvement in test statistic performance resulted from the use of an event date excess return correction for prediction error.

The results of simulation experiments presented here indicate that a sign test based on sample excess return medians provides reliable, well-specified inferences in event studies. This version of the sign test is better specified than the t-test and has a power advantage over the t-test in detecting small (±1/2-percent) levels of abnormal performance. However, both the sign test and the t-test are dominated by the rank test. This suggests that if a researcher wishes to assess statistical significance independently of a parametric assumption concerning the distribution of the data, the rank test is preferred to the sign test.

[16]Rohrbach and Chandra (1989) and Sanders and Robins (1991) provide tests for an event date variance increase in market model residuals.

References

Berry, M. A.; G. W. Gallinger; and G. V. Henderson, Jr. "Using Daily Stock Returns in Event Studies and the Choice of Parametric Versus Nonparametric Test Statistics." *Quarterly Journal of Business and Economics*, 29 (Winter 1990), 70–85.

Boehmer, E.; J. Musumeci; and A. B. Poulsen. "Event Study Methodology under Conditions of Event-Induced Variance." *Journal of Financial Economics*, 30 (Dec. 1991), 253–272.

Brown, S., and J. Warner. "Measuring Security Price Performance." *Journal of Financial Economics*, 8 (March 1980), 205–258.

________________. "Using Daily Stock Returns: The Case of Event Studies." *Journal of Financial Economics*, 14 (Sept. 1985), 3–31.

Corrado, C. J. "A Nonparemetric Test for Abnormal Security-Price Performance in Event Studies. *Journal of Financial Economics*, 23 (Aug. 1989), 385–395.

Fama, E. F. *Foundations of Finance*. New York: Basic Books (1976).

Hettmansperger, T. *Statistical Inference Based on Ranks*. New York: Wiley (1984).

Jain, P. C. "Analyses of the Distribution of Security Market Model Prediction Errors for Daily Returns Data." *Journal of Accounting Research*, 24 (Spring 1986), 76–96.

Lehmann, E. L. *Nonparametrics: Statistical Methods Based on Ranks*. Oakland, CA: Holden-Day (1975).

________________. *Testing Statistical Hypotheses*. New York: Wiley (1986).

Maddala, G. S. *Econometrics*. New York: McGraw-Hill (1977).

Pearson, E. S., and H. O. Hartley. *Biometrika Tables for Statisticians*, Vol. I. Cambridge, UK: Cambridge Univ. Press (1966).

Rohrbach, K., and R. Chandra. "The Power of Beaver's *U* against a Variance Increase in Market Model Residuals." *Journal of Accounting Research*, 27 (Spring 1989), 145–155.

Sanders, R. W., and R. P. Robins. "Discriminating between Wealth and Information Effects in Event Studies." *Review of Quantitative Finance and Accounting*, 1 (July 1991), 307–330.

Zivney, T. L., and D. J. Thompson II. "The Specification and Power of the Sign Test in Measuring Security Price Performance: Comments and Analysis." *The Financial Review*, 24 (Nov. 1989), 581–588.