A Powerful Testing Procedure of Abnormal Stock Returns in Long-Horizon Event Studies

Johan Knif, Department of Finance and Statistics, Hanken School of Economics, FIN-65101 Vaasa, Finland, e-mail: JKnif@hanken.fi

James W. Kolari, Texas A&M University, Department of Finance, College Station, TX 77843-4218 Email: j-kolari@tamu.edu Office phone: 979-845-4803 Fax: 979-845-3884

Seppo Pynnonen[1], Department of Mathematics and Statistics, University of Vaasa, P.O.Box 700, FI-65101, Vaasa, Finland Email: sjp@uwasa.fi Office phone: +358-6-3248259 Fax: +358-6-3248557

First draft: Please do not quote without permission of the authors

Abstract. This paper introduces a new approach based on Sharpe ratios for testing long-run abnormal stock returns in event studies. The proposed test has superior power and size properties over the existing statistics. Normally, we find that cross-correlation is not a major problem in most long horizon event study tests. If it is a problem, we derive cross-correlation adjusted versions of the proposed tests. Based on simulations with actual return data, comparative results show that the proposed test is better specified than other tests under the null hypothesis of no event effects. Furthermore, the power of the proposed test outperforms existing tests.

JEL Classification: C1; G1

Keywords: Abnormal returns; Long-run event study; Standardized returns

[1]Corresponding author

A Robust Test of Long-Run Abnormal Stock Returns in Event Studies

Abstract. This paper introduces a new approach based on Sharpe ratios for testing long-run abnormal stock returns in event studies. The proposed test has superior power and size properties over the existing statistics. Using our approach, cross-correlation is not normally a major problem in most long horizon event study tests. If it is a problem, we derive cross-correlation adjusted versions of the proposed tests. Based on simulations with actual return data, comparative results show that the proposed test is better specified than other tests under the null hypothesis of no event effects. Furthermore, the power of the proposed test outperforms existing tests.

JEL Classification: C1; G1

1 Introduction

Long-horizon event studies of abnormal stock returns deal typically with event windows of several months to years. Unlike short-run event studies using daily stock returns, sample sizes are usually in the hundreds rather than tens of stock return series. Also, in contrast to short-run event studies, there is no separate estimation period in long-run event studies. As noted by Fama [12], Lyon, Barber, and Tsai [26] (henceforth LBT), Kothari and Warner [22], and others, cross-correlation bias and bad model problems tend to plague tests of long-run abnormal returns. The latter problem of an appropriate expected return model is an unresolved asset pricing issue. The Fama-French [13] three-factor model is a frequently used approach in short-run studies. Due to the lack of a reliable mean model in long run event studies, many researchers (e.g, Mitchell and Stafford [28], Eberhart and Siddique [11], Boehme and Sorescu [2], Gombers and Lerner [15], Byun and Rozeff [8], and others) employ a non-model approach popularized by Lyon, Barber, and Tsai (LBT) [26] that utilizes carefully chosen reference portfolios or reference stocks.

Extending the non-model LBT approach, we propose two major innovations that materially improve the size and power properties of statistical tests in long horizon event studies. First, we construct reference portfolios from stocks that form the population from which the event study stocks are sampled. This innovation is consistent with LBT's approach but implies that reference portfolio stocks need not be as closely matched to event study stocks as previously believed, which simplifies the testing process. Second, we use Sharpe ratios to make returns more comparable with respect to each other based on their risk-reward qualities. Departing from extant tests, we define abnormal returns in terms of differences between event study stocks' Sharpe ratios and the market average of the Sharpe ratio in the sample period. Essentially Sharpe ratios are standardized returns (returns scaled by their standard deviations), the use of which in long-horizon event studies is suggested by Fama [12]. He notes that the attraction of using standardized returns stems from the fact that in the derived test statistics the abnormal returns become weighted by their statistical precision "...which seems like the right way to capture the increased information due to the event bunching." [12, p 296]. In this respect, we should note that we do not use standardized abnormal returns per se but differences in the Sharpe ratios of test assets' standardized abnormal returns with respect to market mean Sharpe ratios. Interestingly, this approach can be interpreted in an economic sense to be more consonant with the model approach to event studies that benchmarks abnormal returns relative to common market return and risk.

One by-product of our non-model approach is that it solves in certain cases potential cross-correlation problems that are perceived to plague long-horizon event study results (see, Kothari and Warner [22, Ch. 4], Brav [4, Sec. D], Fama [12, Sec. 4.2.2]). Indeed, our results show that, using appropriate reference portfolios in the abnormal return definition, the cross-correlation problem is virtually eliminated in many cases.

We want to stress that, as discussed in Kothari and Warner [22, Ch. 4], the choice of the mean model is extremely important in long-run horizon event studies. It not only affects the accuracy of expected return estimates but the amount of remaining cross-correlation in abnormal returns also. As shown by Kolari and Pynnonen [19] for short-horizon event studies with clustered event days, it is crucial to choose a mean model that extracts (as much as possible) common cross-correlation in order to improve the power of the tests. In long-horizon event studies, properly defined reference portfolios (or reference stocks) turns out to be a viable method in this respect.

For long-horizon event study methods, a remedy for potential cross-correlation problems is suggested by Jegadeesh and Karceski [18]. However, a shortfall of this approach is that the small sample critical values deviate substantially from the the theoretical thresholds, which causes severe size distortion in the tests.[2] In those relatively extreme cases where the cross-correlation is a problem, the present paper shows that the short-run approach in Kolari and Pynnonen [19] can be adapted to long-horizon event studies to efficiently capture cross-correlation bias even in these cases, such that the size of the proposed test is reasonably close to the intended size and its power outperforms other popular tests.

Other biasing effects in long-horizon even study analyses, discussed among others in Barber and Lyon [1] and Kothari and Warner [21], are new listing or survivor bias and rebalancing bias (see also, Ritter [30]). Survivor bias is inherent in long-run event studies due to the introduction of new companies or delistings in the reference index during the event period. Sampling procedures can be used to control for this bias. Also, as proven by Lyon, Barber, and Tsai [26], rebalancing bias in monthly reference index returns can be avoided by using buy-and-hold abnormal returns (BHARs). Furthermore, unlike short-horizon event studies, inferences of long-horizon event studies may lead to different end results depending on the return metrics employed. This important issue is discussed further in the next section.

[2]We should note that this size distortion can be mitigated to tolerable levels in the case of conventional t-tests with reference stock benchmarks (see, Jegadeesh and Karceski [18, Sec. 3.2]). However, as is shown in this paper, the reference stock approach leads to considerable loss of power in the test statistics.

2 Robust test statistics

In addition to the BHAR measure defined above, average monthly abnormal returns (AARs or CARs) are the other popular return metric applied in long-horizon studies. However, as stated by Fama [12, p. 293], these two metrics can lead to different inferences about event effects.

Due to improved power properties, parametric tests by Patell [29] and Boehmer, Musumeci, and Poulsen (BMP) [3] based on standardized abnormal returns have gained popularity in *short-horizon* event studies. Their advantage stems from the fact that, by scaling the original returns with the standard deviation, abnormal returns become weighted by their statistical precision. Empirical studies (e.g., Patell [29] and Kolari and Pynnonen [19], and others) document that short-run tests using standardized returns have superior power compared to those based on unstandardized returns (e,g, Patell [29] and Kolari and Pynnonen [19], and others).

3 Return metrics

3.1 BHAR

In long-run event studies, as developed by Ikenberry, Lakonishok, and Vermalen [16], Barber and Lyon [1], and Lyon, Barber, and Tsai (LBT) [26], a characteristic-based portfolio matching approach is widely used to estimate buy-and-hold abnormal returns (BHAR). Before the formal definition of BHAR, let us order the sample companies according to their secular event times and denote $t = 1$ the event time of the company with earliest calendar time event in the sample. Next, let k_i denote the number of months associated with the event of company i from the first event in the sample. Thus, $k_1 = 0$ and $k_i \geq 0$ for $i = 2, \ldots n$. For example, if the first event in the sample is in March 1980 and for the ith firm in June 1985, then $k_i = 64$ such that for an $h = 60$ months event period these companies do not have any overlapping event months.

Given these notations, the h-month BHAR for stock i is defined as

$$\mathtt{BHAR}_i(h) = \prod_{t=1}^{h}(1 + R_{i,k_i+t}) - \prod_{t=1}^{h}(1 + R_{b,k_i+t}), \tag{1}$$

where R_{i,k_i+t} is the month $k_i + t$ simple return of the ith stock, and R_{b,k_i+t} is the corresponding return for the benchmark portfolio, $t = 1, \ldots, h$, $i = 1, \ldots, n$.

Using (1) the (average) buy-and-hold abnormal return for a sample of n firms is defined as

$$\overline{\mathtt{BHAR}}(h) = \frac{1}{n}\sum_{i=1}^{n}\mathtt{BHAR}_i(h) \tag{2}$$

with $\mathtt{BHAR}_i(h)$ defined in (1).

It is notable that we can write $\overline{\mathtt{BHAR}}(h)$ in (2) equivalently as

$$\overline{\mathtt{BHAR}}(h) = R_n(h) - R_{bn}(h), \tag{3}$$

where

$$1 + R_n(h) = \frac{1}{n}\sum_{i=1}^{n}(1 + R_i(h)) \tag{4}$$

and

$$1 + R_{bn}(h) = \frac{1}{n}\sum_{i=1}^{n}(1 + R_{b,i}(h)) \tag{5}$$

are the equally weighted h-period (cross) simple returns of the sample and the reference portfolios, respectively, with

$$1 + R_i(h) = \prod_{i=1}^{h}(1 + R_{i,k_i+t}) \tag{6}$$

and

$$1 + R_{b,i}(h) = \prod_{t=1}^{h}(1 + R_{b,k_i+t}). \tag{7}$$

Under the null hypothesis of no event effect, the expected value of $\mathtt{BHAR}_i(h)$ is zero. This hypothesis is tested in the literature by a conventional t-statistic (see Lyon, Barber, ant Tsai [26, p. 173])

$$t_{\mathtt{bhar}} = \frac{\overline{\mathtt{BHAR}}(h)\sqrt{n}}{s_{\mathtt{bhar}}}, \tag{8}$$

or its skweness adjusted version

$$t_{\mathtt{bharsa}} = t_{\mathtt{bhar}} + \frac{\hat{\gamma}}{3\sqrt{n}}\left(t^2_{\mathtt{bhar}} + \frac{1}{2}\right) \tag{9}$$

where $\overline{\mathtt{BHAR}}(h)$ is the sample mean of $\mathtt{BHAR}_i(h)$s, $s_{\mathtt{bhar}}$ is the (cross-sectional) sample standard deviation of $\mathtt{BHAR}_i(h)$ over the sample of n firms, and $\hat{\gamma}$ is an estimate of the skewness coefficient of $\mathtt{BHAR}_i(h)$ (see Lyon, Barber, and Tsai [26, p. 174] for details). We call these statistic in the following as BHAR-T and BHAR-TSA, respectively.

By definition BHAR is bounded below but not above, which implies right skewness in the distribution of these returns (see Barber and Lyon [1] and Lyon, Barber, and Tsai [26], and Kothari and Warner [22, Sec. 4.4.1]). In addition, as observed by Mitchell and Stafford [28, p. 296] and Fama [12, p. 294, property (iii)], BHAR is not a reliable measure long-run abnormal returns due to the compounding effect. That is, if in, say, a 5-year study the first year is abnormal and the subsequent years are normal, BHAR indicates abnormal behavior for all the years after the first one. As a consequence, correct inference about the duration of the adjustment requires short-term analysis. In this example one should be able to test separately the last four years' effect. Whereas this is not a major problem with the BHAR methodology when applied to individual stocks, re-weighting may become an issue in portfolios.

An additional potential problem for BHAR is that simple returns are closer to the log-normal than the normal distribution. For a log-normal random variable, the mean depends both on the location parameter and the variance. That is, the expected value is $\mathrm{E}[R_t(h)] = e^{(\mu+\sigma^2/2)h} - 1$, where $\mu = \mathrm{E}[\log(1 + R_t)]$ and $\sigma^2 = \mathrm{var}[\log(1 + R_t)]$ are the expected value and variance of the single period log-returns. Thus, for example, suppose that the mean annual return of a stock is 12% and (annual) volatility is 20%. Then the five year expected return is $E[R_t(5)] = e^{(0.12+0.5\times0.20^2)\times5} - 1 \approx 1$, or 100%. Assume next that the event changes the risk structure of the firm by increasing its volatility to 0.30. If the risk increase is purely firm specific (as it plausibly would be in firm specific events like seasoned equity offerings or share repurchases), such that it can be diversified away, it would not affect the μ-parameter. However, the volatility increase affects the expected return by raising it to 1.28, or 128%, which would indicate a 28% abnormal return for BHAR. Thus, this methodology would mislead to infer a continuing adjustment in returns due to increased volatility, which is a part of the expected value of a log-normal random variable. It is notable that the median return due to the volatility change remains intact at $e^{\mu} - 1$ (p.a). Although this is a stylized example, it clearly demonstrates that there is potentially also a serious identification problem in the BHAR methodology, which in particular is likely to bias long-horizon inferences.

In addition to the aforementioned issues, Mitchell and Stafford [28, Sec. III] point out serious statistical problems with BHARs that plague reliable statistical inferences. Strong positive skewness and the fact that BHARs generally do not have a zero mean have prompted attempts to overcome these issues by bootstrapping (Brock, Lakonishok, and LeBaron [6], Ikenberry, Lakonishok, and Vermalen [16], Lyon, Barber, and Tsai [26], Jegadeesh and Karceski [18]). However, Mitchell and Stafford [28, Sec. IV.B] and Kothari and Warner [22, Sec. 4] observe that in most cases there is also a cross-correlation problem that boostrapping does not address. As demonstrated in Michell and Stafford [28] as well as short-run event studies with clustered event days by Kolari and Pynnonen [19], even a seemingly trivial (average) cross-correlation substantially changes the t-statistics, their asymptotic distribution, and power of the tests (see Kolari and Pynnonen [19]). Fama [12, p. 295] suggests as a remedy to use the rolling portfolio method of Jaffe [17] and Mandelker [27] with monthly returns.

Overall, the BHAR methodology has many serious issues that can plague inferences in long-run event studies. Barber and Lyon [1] and Lyon, Barber and Tsai [26] motivate the use of BHAR because it ”precisely measures investor experience.” While investor return experience is definitely of interest, it remains an open question whether the BHAR methodology adequately captures economic significance. For example, a key component of investor experience is risk, which BHAR methodology takes into account by means of cross-sectional variance (assuming cross-sectional independence of long-horizon returns). But under this approach, as the number of cross-sectional returns increases, the variance or risk of the equal-(or value-) weighted portfolio converges to zero by the law of large numbers. The approach ignores the cross-correlation, which factually is ubiquitous to virtually all long-horizon event studies. Thus, because returns are more or less cross-correlated, a non-zero limiting variance is implied that equals the market variance for value-weighted portfolios. In view of these serious issues, Mitchell and Stafford [28, p. 307] conclude that "... BHAR methodology, in its traditional form, should not be used for statistical inferences."

3.2 Calendar time returns

Unlike BHARs that focus on holding period returns and base inferences solely on cross-sectional analysis, the calendar time approach tracks event effects of the sample portfolio over the event period months. Fama [12, Sec. 4.2.1] states that if the market efficiency is tested: "... the model of equilibrium jointly tested with market efficiency specifies the unit of time for returns." Referring to Jaffe [17] and Mandleker [27], Fama [12] and Mitchell and Stafford [28] define event portfolios to investigate (for example) three years ($T = 36$ months) event effects such that month t return of the portfolio, or $R_{p,t}$, is the equal-weighted (EW) (or alternatively value-weighted (VW)) sum of returns of those stocks that had the event in previous T periods. The portfolio is rebalanced monthly to drop all companies that reach the end of T months period and to add all companies that have just executed the event. The portfolio excess returns are regressed on the

Fama and French [13] factors

$$R_{p,t} - R_{f,t} = a_p + b_p(R_{m,t} - R_{f,t}) + s_p\texttt{SMB}_t + h_p\texttt{HML}_t + e_{p,t}, \tag{10}$$

where $R_{p,t}$ is the month t simple return of the portfolio for n_t event, $R_{f,t}$ is the three-month T-bill rate, $R_{m,t}$ is the CRSP value-weighted market return, SMB is the small minus big stocks zero investment portfolio, and HML is the high BE/ME minus low BE/ME portfolio (see Fama and French [13]). If model (10) correctly captures the equilibrium price of the portfolio, the intercept a_p measures the average monthly abnormal return on the portfolio of event firms. Under the null hypothesis of no event effect, $a_p = 0$. Because a_p indicates the Jensen's alpha, this method is also called the Jensen-alpha approach (for example) by Kothari and Warner [22].

The advantage of the portfolio method is that it accounts for the possible cross-correlation of returns. However, varying the number of returns in each month imposes some heteroscedasticity in the residuals. Mitchell and Stafford [28] have alleviated this effect by including only those months that have at least 10 event series. The major problem, however, is the bad model problem as discussed in Fama [12], which remains unresolved. In terms of the Fama-French three-factor model in (10), this shows up in the fact that a_p tends to be non-zero in non-event portfolios also. As noted by Mitchell and Stafford [28, p. 315], because the composition of the portfolio changes monthly, the constancy of the slope coefficients is unlikely. An additional biasing effect is the equal weighting of calendar months, so that months with heavy event activity are treated the same as low activity months (Mitchell and Stafford [28, p. 316] and Loughran and Ritter [24, pp. 362–363]). For an extensive discussion of additional problems of using existing factor models like the Fama-French model or reference portfolios, see Loughran and Ritter

[24]. In spite of these obvious shortcomings compared to the BHAR methodology, Mitchell and Stafford [28, p. 326] conclude, "... we strongly advocate a methodology that accounts for the dependence of event-firm abnormal returns, such as the calendar time portfolio approach."

3.3 Cumulative Abnormal Returns

Cumulative abnormal returns are defined as sums of daily abnormal returns over the period of interest. Let R_{it} denote the month t (simple) return of company i, $\mathrm{E}[R_{it}]$ the expected return, and $\mathtt{AR}_{it} = R_{it} - \mathrm{E}[R_{it}]$ the abnormal return of company i in month t, such that the cumulative abnormal return (CAR) over h periods is defined as

$$\mathtt{CAR}_i(h) = \sum_{t=1}^{h} \mathtt{AR}_{i,t}. \tag{11}$$

An obvious advantage of CARs over BHARs and calendar time returns is that it is easy to investigate sub-periods as well as the total event period. In spite of this obvious advantage, the BHAR and calendar time approaches are much more popular than the CAR approach. From economic interpretation standpoint, a disadvantage of CARs is that, when based on simple returns, they do not accurately measure the h period return. This is because the h months return is the product of monthly returns, not the sum, as in BHAR. Barber and Lyon [1, Sec. 2.1] demonstrate empirically the difference utilizing CRSP data. In terms of one year holding period they show that typically the difference $\mathtt{CAR} - \mathtt{BHAR}$ is positive except for high $\mathtt{BHAR}$s where the difference becomes increasingly negative.

However, it should be borne in mind that from a statistical testing point of view, in testing the null hypothesis of no event effect, the interest is solely in two things:

the size and power of the test. That is, the behavior of the test method under the null hypothesis and the power of rejecting the null when the alternative is true. As a consequence, the question whether CARs and BHARs measure different things under the alternative hypothesis is irrelevant. This is because in statistical testing the question is to find a method that has correct null behavior and detects the deviation from the null most efficiently.

Thus, if CAR and BHAR measure the same thing under the null hypothesis, they are equally good initial candidates. The one which results to a more sensitive test statistic with correct null behavior is preferred. As a consequence, selection between BHAR and CAR based on the empirical properties in Barber and Lyon [1, Sec. 2.1] is misleading from a statistical theory point of view: "Though particular sample means of CAR and BHAR are unbiased with respect to zero, CARs are biased estimates of BHARs ..." (Barbet and Lyon [1, p. 346] The unbiasedness under the null of no event effect makes them equally good in the beginning. Economic significance can only be evaluated when the event signal has been detected as statistically significant. Perhaps this can be best illustrated in terms of non-parametric statistics. For example, replacing returns by the rank numbers in short-run event studies has led to powerful test procedures (see e.g. Kolari and Pynnonen [20]). Rank numbers per se do not convey much economic content, but economic importance can be judged on the basis of the returns themselves after statistical significance has been established.

3.4 Continuously compounded returns

Because simple returns are bounded below but not above, a major statistical problem with the simple return is that its distribution becomes by definition skewed. Even though the central limit theorem guarantees normal approximation, the convergence to the limiting distribution slows down materially for skewed distributions and makes normal approximation inaccurate even for moderately large sample sizes such as 200 (Sutton [31], Lyon, Barber and Tsai [26]).

In order to alleviate the skewness symptom, we use log returns, or $r_i(h) = \log(1 + R_i(h))$. It is, however, important to notice the difference between simple and log returns when dealing with portfolios. As is well known (e.g., Campbell, Lo, and MacKinlay [9, p. 11] and exemplified by Barber and Lyon [1, Sec. 2.3], log returns do not aggregate cross-sectionally to portfolio returns. This is not surprising because logarithm is a non-linear function, such that generally $\log(x+y) \neq \log(x)+\log(y)$. As a consequence, if x and y are (cross) simple returns of stocks $\log(x+y)$ and $\log(x)+\log(y)$ measure completely different things. The former measures the continuously compounded return of an equal-weighted portfolio, whereas the latter, if divided by two, measures the average of continuously compounded returns of individual assets. Thus, in order to measure consistently the same characteristics of returns, it is critical to apply logarithms of the same order to the sample stocks and reference portfolios. Basically there are two options, of which one will be discussed next in detail and used in subsequent empirical simulations with actual return data.

3.4.1 Continuously compounded abnormal returns ($\texttt{CCAR}$)

An option in applying continuously compounded returns in a consistent manner to test event mean effects is to test mean differences of average log returns of the test assets and the average log-returns of stocks in the reference portfolio.

More precisely, to preserve consistency with sample and reference stocks, we define an h-period continuously compounded abnormal return (CCAR)

$$\texttt{CCAR}_i(h) = r_i(h) - E[r_i(h)], \tag{12}$$

where

$$r_i(h) = \sum_{t=1}^{h} r_{i,k_i+t} \tag{13}$$

with $r_{i,k_i+t} = \log(1 + R_{i,k_i+t})$ the single period log-return, and R_{i,k_i+t} the single period (month $k_i + t$) simple return of stock i.

The sample average of log returns is

$$\bar{r}_n(h) = \frac{1}{n}\sum_{i=1}^{n} \log(1 + R_i(h)) \tag{14}$$

with $1 + R_i(h)$ defined in equation (6).

The average of the h-period log returns in the reference portfolio of stock i with event period starting in the relative times at month k_i is

$$\bar{r}_{N_i}(h) = \frac{1}{N_i}\sum_{j=1}^{N_i} \log(1 + R_{j,i}(h)), \tag{15}$$

where $1 + R_{j,i}(h) = \prod_{t=1}^{h}(1 + R_{j,k_i+t})$ is the h period cross simple return of the

jth stock in the benchmark portfolio of the sample stock i, and N_i is the number of stocks in the benchmark in the beginning of the event period of stock i.

Subsequently, the continuously compounded abnormal return of stock i is

$$\mathtt{CCAR}_i(h) = r_i(h) - r_{N_i}(h) \tag{16}$$

such that average CCAR becomes

$$\overline{\mathtt{CCAR}}(h) = \frac{1}{n}\sum_{i=1}^{n}\mathtt{CCAR}_i(h). \tag{17}$$

It is notable that $\overline{\mathtt{CCAR}}(h)$ is completely additive. That is

$$\overline{\mathtt{CCAR}}(h) = \frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{h}\mathtt{CCAR}_{i,k_i+t}, \tag{18}$$

where

$$\mathtt{CCAR}_{i,k_i+t} = r_{i,k_i+t} - \bar{r}_{N_i,k_i+t} \tag{19}$$

is the single period continuously compounded abnormal return with

$$\bar{r}_{N_i,k_i+t} = \frac{1}{N_i}\sum_{j=1}^{N_i} r_{j,k_i+t}, \tag{20}$$

the average of month s log-returns of the reference portfolio of the sample stock with event month t, $t \leq s \leq t+h-1$. The additivity allows us to define for any sub-period $1 \leq h_1 \leq h_2 \leq h$

$$\mathtt{CCAR}_i(h_1, h_2) = \sum_{t=h_1}^{h_2}\mathtt{CCAR}_{i,k_i+t}, \tag{21}$$

such that

$$\overline{\texttt{CCAR}}(h_1, h_2) = \frac{1}{n} \sum_{i=1}^{n} \texttt{CCAR}_i(h_1, h_2), \tag{22}$$

which makes it straightforward to test the significance of the event effect in any sub-period with the t-ratio (CCAR-T)

$$t_{\texttt{ccar}} = \frac{\overline{\texttt{CCAR}}(h_1, h_2)\sqrt{n}}{s_{\texttt{CCAR}(h_1,h_2)}}, \tag{23}$$

where $s_{\texttt{CCAR}(h_1,h_2)}$ is the square root of the sample variance of $\texttt{CCAR}_i(h_1, h_2)$

$$s^2_{\texttt{ccar}}(h_1, h_2) = \frac{1}{n-1} \sum_{i=1}^{n} \left(\texttt{CCAR}_i(h_1, h_2) - \overline{\texttt{CCAR}}(h_1, h_2)\right)^2. \tag{24}$$

If $\texttt{CCAR}_i(h_1, h_2)$ are cross-sectionally independent, the central limit theorem implies that the limiting distribution of equation (23) is standard normal including the heteroscedastic case of varying individual variances $\sigma_i^2(h_1, h_2) = \mathrm{var}\left[\texttt{CCAR}_i(h_1, h_2)\right]$. Again, if the event days are clustered, the cross-sectional variance in (24) underestimates the true variance implying over-rejection of the null hypothesis. We will discuss estimation of the cross-correlation in Subsection 3.5.1 below.

3.4.2 Sharpe ratio methodology

The importance of risk adjustment is critical in evaluation of abnormal returns, as even a small error in risk adjustment can accumulate to become economically large in long-horizon calculations (see, Kothari and Warner [22, Sec. 4.2]). For example, in the Jensen's alpha methodology discussed in Section 3.3, in addition to the bad model problem, errors and potential biases in estimating the factor regression coefficient can impose material errors in the intercept (Jensen's alpha measure).

Here we propose to employ the Sharpe ratio for the purpose of risk adjustment in estimating abnormal returns. We compute the Sharpe ratio as

$$\mathtt{sr}_i(h) = \frac{r_i(h) - r_{f,i}(h)}{s_i(h)}, \tag{25}$$

where $r_i(h)$ is the h-period log-return of stock i defined in equation (13), $r_{f,i}(h)$ is the stock i related h-period log-return of the riskless asset, and $s_i(h) = \sqrt{h}s_i$ in which

$$s_i = \sqrt{\frac{1}{h-1}\sum_{t=1}^{h}(r_{i,k_i+t} - \bar{r}_i)^2} \tag{26}$$

is the event period standard deviation of the log-returns r_{i,k_i+t} with $\bar{r}_i$ the event period average of returns r_{i,k_i+t}.

We define abnormal returns, or *abnormal Sharpe ratios* ($\mathtt{ASR}$), as

$$\mathtt{ASR}_i(h) = \mathtt{sr}_i(h) - \overline{\mathtt{sr}}_i(h), \tag{27}$$

where

$$\overline{\mathtt{sr}}_i(h) = \frac{1}{N_i}\sum_{j=1}^{N_i}\mathtt{sr}_j(h) \tag{28}$$

is the average h-period Sharpe-ratio for reference portfolio stocks.

Again, because $r_i(h) = \sum_{t=1}^{h} r_{i,k_i+t}$ and $r_{f,i}(h) = \sum_{t=1}^{h} r_{f,k_i+t}$, it is straightforward to deal with sub-periods by defining in an obvious manner

$$\mathtt{ASR}_i(h_1, h_2) = \mathtt{sr}_i(h_1, h_2) - \overline{\mathtt{sr}}_i(h_1, h_2). \tag{29}$$

Thus, an event effect on the sub-period $1 \le h_1 \le h_2 \le h$ can be tested in terms of the Sharpe-ratios with the t-ratio (SHARPE-T)

$$t_{\texttt{sharpe}} = \frac{\overline{\texttt{ASR}}(h_1, h_2)\sqrt{n}}{s_{\texttt{asr}}(h_1, h_2)}, \tag{30}$$

where

$$\overline{\texttt{ASR}}(h_1, h_2) = \frac{1}{n}\sum_{i=1}^{n} \texttt{ASR}_i(h_1, h_2) \tag{31}$$

is the sample mean of $\texttt{ASR}_i(h_1, h_2)$s, and $s_{\texttt{ASR}}(h_1, h_2)$ is the square root of the cross-sectional sample variance

$$s^2_{\texttt{asr}}(h_1, h_2) = \frac{1}{n-1}\sum_{i=1}^{n} \left(\texttt{ASR}_i(h_1, h_2) - \overline{\texttt{ASR}}(h_1, h_2)\right)^2 \tag{32}$$

of the abnormal Sharpe ratios $\texttt{ASR}_i(h_1, h_2)$. Under the independence of $\texttt{ASR}_i(h_1, h_2)$s, the asymptotic null-distribution of (30) is standard normal.

The Sharpe-ratio has several attractive economic and statistical features. The major economic features are that the Sharpe ratio has an acknowledged economic interpretation as a reward-to-risk ratio. Also, it accounts for the possible change of the return-to-risk ratio of the company due to the event. For example, a seasoned equity offering (SEO) is likely to change the debt-to-equity ratio and hence the leverage of the company which should have an impact on its expected return. If this return effect is not balanced by the changed risk of the company, it may be incorrectly viewed as an abnormal return in an event study. In this regard, the Sharpe ratio accounts for the changed return-to-risk ratio.

Statistically, Sharpe-ratios are essentially standardized returns. Thus, scaling returns by their standard deviation stabilizes the variance of the cross-sectional observations. More importantly, standardized returns have gained popularity in short-run event studies due to their superior power in tests over those based on non-standardized returns (see, Patell [29], Boehmer, Musumeci, and Poulsen [3], Kolari and Pynnonen [19, 20], among others). As previously discussed, this power advantage arises from the fact that in the derived test statistics the abnormal returns become weighted by their statistical precision. Our simulation results below with actual return data confirm the power of scaling in detecting event effects. Next we discuss the estimation of the variances of abnormal returns.

3.5 Estimation of the variance of abnormal returns

Unbiased estimation of the abnormal return standard deviation is critical to the correct size of event effect testing. The size problem in particular in the BHAR approach is a well documented problem (Barber and Lyon [1], Lyon, Barber, and Tsai [26], Mitchell and Stafford [28], and Jegadeesh amd Karceski [18]). Probably the most appropriate approach to control size in the BHAR-methodology seems to be the use of a single reference stock. This, however, becomes costly due to severe loss of power. Jegadeesh and Karceski [18] suggest an approach that is robust to cross-correlation and hetersocedastisity, but unfortunately the cost is considerable loss in power (see Jegadeesh and Karceski [18, p. 110]). As discussed earlier, the major problem with the BHAR methodology is skewness as well as the multiplicative nature of the return definitions which make their statistical properties difficult to discern. Thus, it is not possible to correct the biases and at the same time preserve power.

Log-returns have the great advantage of being sums of shorter term log-returns which makes them much more tractable in a statistical sense. This makes them also much more convenient for construction of suitable test methods to detect possible changes in the return generating process due to event effects. It is notable that, if the event effect has been spotted, its economic significance need not be based on log-returns. It can be based on simple returns and related measures. The role of the statistical tests is to detect the presence of a potential signal. Given the signal, the next step is to evaluate its practical relevance.

3.5.1 Variances of CCAR and abnormal Sharpe ratios

The cross-sectional variances defined by (24) and (32) of the continuously compounded abnormal return CCAR and abnormal Sharpe ration ASR, respectively are logically equivalent. As a consequence the solution of the potential cross-correlation problem is similar in both cases and can be solved in similar techniques.

In the most general case, where the returns are allowed to be of autocorrelated, cross-autocorrelated, and cross-sectional correlated, the estimation of the variances become intractable because of the too complex dependencies. However, by assuming some degree of market efficiency we can rule out some of the serial dependency. Then the cross-sectional dependency related to the remaining autocorrelation is affected by the amount the event periods are overlapping.

Using the portfolio variance formula an the results in Kolari and Pynnonen [19, Sec. 5.3], the variance of $\overline{\mathtt{CCAR}}(h_1, h_2)$ can be written as

$$\mathrm{var}[\overline{\mathtt{CCAR}}(h_1, h_2)] = \frac{1}{n}\bar{\nu}\left(1 + (n-1)\frac{\bar{\nu}_{ij}}{\bar{\nu}}\right), \tag{33}$$

where $\bar{\nu}$ is the average of the variances of $\mathtt{CCAR}_i(h_1, h_2)$s and $\bar{\nu}_{ij}$ is the average of the covariances of $\texttt{CCAR}_i(h_1, h_2)$ and $\texttt{CCAR}_j(h_1, h_2)$, $i, j = 1, \ldots, n$, $i \neq j$.

As pointed out e.g. in Kolari and Pynnonen [19], the expected value of the sample variance $s^2_{\texttt{ccar}}(h_1, h_2)$ defined in equation (24) is

$$\mathrm{E}\left[s^2_{\texttt{ccar}}(h_1, h_2)\right] = \bar{\nu}\left(1 - \frac{\bar{\nu}_{ij}}{\bar{\nu}}\right). \tag{34}$$

Comparing this with (42) we observe that the variance estimator $s^2_{\texttt{ccar}}(h_1, h_2)/n$, utilized in the t-ratio (23), is in the presence of cross-sectional correlation a biased estimator of the population variance. If the ratio of the average covariances and average variances, denoted as $\theta = \bar{\nu}_{ij}/\bar{\nu}$, were known

$$\tilde{s}^2_{\texttt{ccar}}(h_1, h_2) = \frac{1 + (n-1)\theta}{1 - \theta} s^2_{\texttt{ccar}}(h_1, h_2) \tag{35}$$

would be an unbiased estimator of the variance $\overline{\texttt{CCAR}}(h_1, h_2)$. Because the θ-ratio is unknown, it must be estimated from data. We will next discuss approaches to estimate θ.

Estimation of the covariance-variance ratio θ

There are three main instances that affect θ estimation. One is where the event periods are completely separate such that none of the sample stocks have overlapping event months. In such a case the θ parameter is zero because the cross-correlations are zero. This case, however, is quite implausible at least for longer event windows. The second case is the opposite, where the all event months are completely clustered with all stocks sharing the same event month. The third option is in between with partially overlapping event periods. We will deal with these two last cases in detail below.

(a) *Completely overlapping event windows*

In order to deal with the cross-correlation we assume first that all firms share the same event month, such that in $\texttt{CCAR}_{i,k_i+t}$ defined in equation (19) we can set $k_i = 0$. With this convention in order to simplify notations that follow, we denote

$$u_{it} = \texttt{CCAR}_{i,t} \tag{36}$$

such that the the cumulative $\texttt{CCAR}_i(h_1, h_2)$ in (21) becomes

$$\texttt{CCAR}_i(h_1, h_2) = \gamma' u_i, \tag{37}$$

where γ is a h-vector with ones in positions from h_1 to h_2 and zeros elsewhere, $u_i = (u_{i1}, \ldots, u_{ih})'$ is the h-vector of $u_{it} = \texttt{CCAR}_{i,t}$, $t = 1, \ldots, h$, and the prime denotes the transposition. With these notations we make the following assumption:

Assumption 1 *It is assumed that u_{it} are conditionally unpredictable in mean by other returns such that,*

$$\mathrm{E}\left[u_{i,t+k} | u_{1,t}, u_{2,t}, \ldots, u_{n,t}\right] = \mathrm{E}\left[u_{i,t+k} | u_{i,t}\right], \tag{38}$$

for all $i = 1, \ldots, n$, $k = 1, 2, \ldots$

This assumption allows both autocorrelation and cross-autocorrelations, but rules out lead-lag relationships and thus Granger-causality between the series. Accordingly, all observed cross-autocorrelations are due to contemporaneous cross-correlations and (own) serial correlations of the series.

In particular, if the normality of the returns holds, $\mathrm{E}\left[u_{i,t+k}|u_{i,t}\right]$ is a linear function of the form $\mathrm{E}\left[u_{i,t+k}|u_{i,t}\right] = \varrho_i(k)u_{i,t}$ (assuming $\mathrm{E}[u_{it}] = 0$), where $\varrho_i(k)$ is the autocorrelation of $u_{i,t+k}$ and $u_{i,t}$, or the kth order autocorrelation of the series u_{it}. Then $\mathrm{cov}\left[\tilde{u}_{i,t+k}, u_{j,t}\right] = 0$, for all $i \neq j$, where $\tilde{u}_{i,t+k} = u_{i,t+k} - \varrho_{ik}u_{i,t}$, which implies

$$\mathrm{cov}\left[u_{i,t+k}, u_{j,t}\right] = \sigma_{ij}\varrho_i(k), \tag{39}$$

in which $\sigma_{ij} = \mathrm{cov}\left[u_{i,t}, u_{j,t}\right]$ is the cross-sectional covariance of $u_{i,t}$ and $u_{j,t}$. Using these, the covariance matrix of u_i and u_j becomes

$$\mathrm{cov}\left[u_i, u_j\right] = \sigma_{ij}\mathbf{\Omega}_{ij} \tag{40}$$

where

$$\mathbf{\Omega}_{ij} = \begin{pmatrix} 1 & \varrho_j(1) & \varrho_j(2) & \varrho_j(3) & \cdots & \varrho_j(h) \\ \varrho_i(1) & 1 & \varrho_j(1) & \varrho_j(2) & \cdots & \varrho_j(h-1) \\ \varrho_i(2) & \varrho_i(1) & 1 & \varrho_j(1) & \cdots & \varrho_j(h-2) \\ \varrho_i(3) & \varrho_i(2) & \varrho_i(1) & 1 & & \varrho_j(h-3) \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ \varrho_i(h) & \varrho_i(h-1) & \varrho_i(h-2) & \varrho_i(h-3) & \cdots & 1 \end{pmatrix} \tag{41}$$

Noting that $\gamma'\mathbf{\Omega}_{ij}\gamma = \gamma'\mathbf{\Omega}_{ji}\gamma$ and $\mathbf{\Omega}_i + \mathbf{\Omega}_j = \mathbf{\Omega}_{ij} + \mathbf{\Omega}_{ji}$, where $\mathbf{\Omega}_i \equiv \mathbf{\Omega}_{ii}$ and $\mathbf{\Omega}_j \equiv \mathbf{\Omega}_{jj}$ are the symmetric autocorrelation matrices of u_i and u_j, respectively, the variance of $\overline{\mathtt{CCAR}}(h_1, h_2)$ can be presented (after some algebra) in these notations as

$$\begin{aligned}\operatorname{var}\left[\overline{\texttt{CCAR}}(h_1,h_2)\right] &= \frac{1}{n^2}\left(\sum_{i=1}^{n}\operatorname{var}\left[\gamma' u_i\right]+\sum_{i\neq j}\operatorname{cov}\left[\gamma' u_i,\gamma' u_j\right]\right) \qquad (42)\\ &= \frac{1}{n^2}\sum_{i=1}^{n}(\sigma_i^2+(n-1)\bar{\sigma}_i)\gamma'\mathbf{\Omega}_i\gamma,\end{aligned}$$

where σ_i^2 is the one period (monthly) variance of asset i and $\bar{\sigma}_i=\sum_{j\neq i}\sigma_{ij}/(n-1)$ is the average one period cross-section covariance of asset i with the other assets $j=1,\ldots,n,\ j\neq i$. Assuming that the variances σ_i^2 and average cross-sectional covariances $\bar{\sigma}_i$ are 'uncorrelated' with the quadratic forms of the autocorrelations $\gamma'\mathbf{\Omega}_i\gamma$, use of the covariance identity $\sum_i(x_i-\bar{x}_i)(y_i-\bar{y})=\sum_i x_iy_i-n\bar{x}\bar{y}$ implies under the zero covariance assumption that $\sum_i x_iy_i = n\bar{x}\bar{y}$, or in terms of the variance in (42)

$$\operatorname{var}\left[\overline{\texttt{CCAR}}(h_1,h_2)\right]=\frac{1}{n}\left(\bar{\sigma}^2+(n-1)\bar{\sigma}\right)\gamma'\bar{\mathbf{\Omega}}\gamma. \qquad (43)$$

In equation (43), $\bar{\sigma}^2=\sum_{i=1}^{n}\sigma_i^2/n$, $\bar{\sigma}=\sum_{i\neq j}\sigma_{ij}/[(n(n-1)]$, and $\bar{\mathbf{\Omega}}=\sum_{i=1}^{n}\mathbf{\Omega}_i/n$ are the averages of the variances, σ_i^2, cross-sectional covariances, σ_{ij}, and autocorrelation matrices $\mathbf{\Omega}_{ij}$, respectively. The bottom line is that we can estimate $\bar{\sigma}^2$ as the average of sample variances of the stocks calculated from the monthly returns and the average covariance $\bar{\sigma}$ can be efficiently computed from the equally weighted sample portfolio variance equation

$$\begin{aligned} s_p^2 &= \frac{1}{h-1}\sum_{t=1}^{h}(\texttt{ar}_t-\overline{\texttt{ar}}_t)^2 \qquad (44)\\ &= \frac{1}{n^2}\left(\sum_{i=1}^{n}s_i^2+\sum_{i\neq j}c_{ij}\right)\\ &= \frac{1}{n}(\bar{s}^2+(n-1)\bar{c}),\end{aligned}$$

where

$$s_i^2 = \frac{1}{h-1}\sum_{t=1}^{h}(\mathtt{ar}_{it} - \overline{\mathtt{ar}}_i)^2 \tag{45}$$

is the monthly sample variance of the abnormal return series i,

$$c_{ij} = \frac{1}{h-1}\sum_{t=1}^{h}(\mathtt{ar}_{it} - \overline{\mathtt{ar}}_i)(\mathtt{ar}_{jt} - \overline{\mathtt{ar}}_j) \tag{46}$$

is the monthly sample covariance of the abnormal return series i and j, and $\bar{s}^2 = \sum_i s_i^2/n$ and $\bar{c} = \sum_{i \neq j} c_{ij}/[n(n-1)]$ are respectively the averages of the variances and covariances. Utilizing the last line of equation (44) the average covariance can be efficiently computed as

$$\bar{c} = \frac{ns_p^2 - \bar{s}^2}{n-1}. \tag{47}$$

Thus, an estimator for the θ ratio in (35) is

$$\hat{\theta} = \frac{\bar{c}}{\bar{s}^2} = \frac{ns_p^2 - \bar{s}^2}{(n-1)\bar{s}^2} = \frac{ns_p^2/\bar{s}^2 - 1}{n-1}. \tag{48}$$

This is computationally an efficient way to estimate the covariance variance ratio, because instead of computing $n(n-1)/2$ covariances and n variances to the ratio one needs only to compute $n+1$ variances, i.e., variances of the n sample series and the equally weighted portfolio variance. For example, if $n = 400$, one needs only to compute 401 variances compared to $400 \times 401/2 = 80,200$ variances and covariances. This approach gives also a handy method of computing the average correlation by replacing the original observations by the standardized returns $z_{it} = (\mathtt{ar}_{it} - \overline{\mathtt{ar}}_i)/s_i$, defining the equally weighted index $z_t = \sum_i z_{it}/n$, its sample variance s_z^2, such that the average correlation using equation (47) can be computed simply as $\bar{r} = (ns_z^2 - 1)/(n-1)$. Thus, again only $n+1$ variances are need to compute instead of $n(n-1)/2$ covariances and n variances.

Finally, it may be noted that in principle we could use the portfolio variance s_p^2 defined in equation (44) to estimate the variance of $\overline{\texttt{CCAR}}(h_1, h_2)$. However, typically in long-run event studies the number of firms n is much larger than the number of months h, which implies that estimating the variance from the cross-sectional observations is much more efficient. In addition, the possible autocorrelation becomes implicitly estimated by the cross-sectional variance $s^2_{\texttt{ccar}}(h_1, h_2)$ defined in (24). Also, it must be noted that the estimator of the average covariance is based essentially on only the $h-1$ degrees of freedom, which implies that in short event windows random noise would be material in individual covariance estimates, but fortunately can be expected to be reduced in the average. However, the importance of the selection of a mean model that eliminates as much as possible of the cross-correlation is crucial to improve the power of the tests (c.f. Kolari and Pynnonen [19, p. 4016]).

(b) *Partially overlapping event windows*

If the returns are autocorrelated, unfortunately the solution of partially overlapped event periods does not solve as easily as in the above completely overlapping case. However, if we can assume that the serial correlations are zero or negligible the cross-correlation problem can be solved in a straightforward manner. We make first explicitly the zero autocorrelation assumption:

Assumption 2 *The autocorrelations and cross-autocorrelations of the return series are zero.*

In the notations above, suppose that returns in u_i have an overlapping event window with returns in u_j starting from month k, $1 < k \leq h$. That is, there are

$h-k+1$ overlapping event months. Then the covariance matrix between u_i and u_j is of the form

$$\operatorname{cov}[u_i, u_j] = \sigma_{ij}\tilde{\mathbf{I}}(k) \tag{49}$$

where $\tilde{\mathbf{I}}(k)$ is an $h \times h$ matrix with ones in positions $i, k+i-1$, $i = 1, \ldots, h-k+1$ and zeros elsewhere. I.e., $\tilde{\mathbf{I}}(k)$ is a matrix of the form

$$\tilde{\mathbf{I}}(k) = \begin{pmatrix} 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & & & \ddots & \\ 0 & \cdots & 0 & 0 & \cdots & 1 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{pmatrix}. \tag{50}$$

Consider the cumulative abnormal returns over the whole event window such that in (37) $\gamma = (1, \ldots, 1)'$ is an h-vector of ones. Then the cross-covariance of the cumulative abnormal returns $\mathtt{CCAR}_i(1, h) = \gamma' u_i$ and $\mathtt{CCAR}_j(1, h) = \gamma' u_j$ becomes

$$\operatorname{cov}[\gamma' u_i, \gamma' u_j] = \sigma_{ij}\gamma'\tilde{\mathbf{I}}(k)\gamma = (h-k+1)\sigma_{ij}. \tag{51}$$

The covariance (51) is invariant under the permutation of the elements in the vectors u_i and u_j. Thus, for convenience we can arrange the vectors such that the coordinates of the returns that have the same same calendar months are matched and the rest of the returns with non-overlapping months fill the rest of the vectors. For example, we arrange the overlapping calendar month returns to the end part of the vectors and the non-overlapping returns to the beginning, the $\tilde{\mathbf{I}}(k)$ matrix takes the format in which the last $(h-k+1)$ main diagonal elements are ones and all other elements are zeros.

Matching the returns in the above manner and computing the sample covariance by c_{ij} defined in equation (46), the expected value of it is

$$\mathrm{E}\left[c_{ij}\right] = \frac{h-k-1}{h}\sigma_{ij}. \tag{52}$$

Thus, hc_{ij} is an unbiased estimator of the covariance (51) of the cumulative abnormal returns.

Using the above matching convention we can derive an estimator of the covariance-variance ratio θ that is based on unbiased covariances and variances. Utilizing the portfolio method as in equation (48) gives again a computationally efficient approach. This can be accomplished as follows:

1. Arrange the firms in the order according to their event month in calendar time such that, the return series of the first company has the earliest event month, the second company has the second earliest event month, and so on.

2. Match the coordinate of the first overlapping event months of firm i with firm $i-1$, fill the remaining coordinates up to h in calendar order and move the leftover returns in calendar order to the beginning of the vector, $i = 2, \ldots, n$. If there are no overlapping months in the return vectors of firm i and firm $i-1$, no reordering is needed.

3. Given the reordered series, compute the equally weighted portfolio of the reordered series and compute its variance by formula (44). Compute the individual variances of each series using formula (45) and compute the average of the variances.

4. With the portfolio variance and the average of the variances, compute an estimator for the θ-ratio using equation (48).

It is notable that this approach can be extended also to estimate the θ-ratio for the variance in testing arbitrary sub-periods by constructing the portfolio such that it accurately reflects the cross-correlation of the sub-period returns. However, we do not elaborate the analysis to this direction further.

Let $\hat{\theta}_{\texttt{ccar}}$ denote the estimated θ-ratio by the above procedure using continuously compounded abnormal returns (CCARs), then cross-correlation adjusted CCAR-T (CCAR-TCA), where CCAR-T is defined by equation 23, for testing CCAR becomes

$$t_{\texttt{ccarca}} = \frac{\overline{\texttt{CCAR}}(1,h)\sqrt{n}}{\hat{s}_{\texttt{ccar}}(1,h)}\sqrt{\frac{1-\hat{\theta}}{1+(n-1)\hat{\theta}}}, \tag{53}$$

where

$$s_{\texttt{ccar}}(1,h) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\texttt{CCAR}_i(1,h) - \overline{\texttt{CCAR}}(1,h)\right)^2} \tag{54}$$

is the cross-sectional variance of the h-period CCARs.

In the same manner, let $\hat{\theta}_{\texttt{asr}}$ denote the θ-ratio estimated by the above procedure using abnormal Sharpe-ratios (ASR), a cross-correlation adjusted SHARPE-T (SHARPE-TCA), where SHARPE-T is defined in equation (30), becomes

$$t_{\texttt{sharpeca}} = \frac{\overline{\texttt{ASR}}(1,h)\sqrt{n}}{\hat{s}_{\texttt{asr}}(1,h)}\sqrt{\frac{1-\hat{\theta}_{\texttt{asr}}}{1+(n-1)\hat{\theta}_{\texttt{asr}}}}, \tag{55}$$

where

$$s_{\texttt{asr}}(1,h) = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\texttt{ASR}_i(1,h) - \overline{\texttt{ASR}}(1,h)\right)^2} \tag{56}$$

is the cross-sectional variance of the h-period ASRs.

4 Test statistics, data, and simulation designs

All the test statistics are variants of the t-ratio, i.e., mean abnormal return divided by its standard error. Table 1 summarizes the used t-ratios and their key characteristics.

[Table 1]

4.1 Reference portfolios

We construct three types of reference portfolios, or benchmarks: (a) Size/book-to-market matched portfolios, (b) Market-wide portfolios, and (c) Size/book-to-market matched reference stocks. Each of these are discussed in detail below.

(a) *Size/book-to-market matched portfolios*

The event study methodology recommended in Lyon, Barber, and Tsai [26] has become popular in long horizon event studies (see Jegadeesh and Karceski [18] and references therein). Here we follow their methodology in constructing 70 size and book-to-market reference portfolios. Our sample is comprised of NYSE, AMEX, and NASDAQ stocks in the sample period from July 1973 to December 2009. ADRs, closed-end funds, and REITs are excluded. Monthly returns and market capitalization are obtained from CRSP, and book value data are gathered from Compustat (i.e, item 60).

In constructing the 70 size and book-to-market reference portfolios, first stocks are assigned to size deciles based on the market value of the equity at the end of July each year using CRSP end-of-month prices and total shares outstanding. The NYSE size decile break points are then used to assign firms into the size categories. AMEX and NASDAQ firms are placed into appropriate NYSE size categories according to their end-of-June market value. As in Lyon, Barber, and Tsai [26], because many small firms listed in NASDAQ fall in the small firm decile, the decile is further partitioned into quintiles using all NYSE, AMEX, and NASDAQ firms in the decile. Thus, we have a total of 14 size categories. Each of these 14 size categories are further subdivided into book-to-market ratio quintiles based on the book-to-market (BM) value of the firm as of the end of the previous calender year. The most recent book value of equity is used as of December in a particular year, and the market value of equity at the end of December are used to compute BM. Each of the 14 market size categories are evenly divided into five portfolios based on the previous December's BM. This yields a total of 70 size/BM reference portfolios.

(b) *Market-wide portfolios*

In order to highlight the need of the fairly complicated strategy of constructing reference portfolios for proper risk adjustment of the abnormal returns, we report for comparison purposes results where the risk adjustments are worked out in terms of reference portfolios that are simply market averages of all stocks. More precisely, the market-wide portfolios is formed using the stocks of the universe of benchmark design (a) by constructing each in month the holding period (12, 36, or 60 months) buy-and-hold portfolio from all those stocks that have complete return series over the whole holding period of interest (i.e., 12, 36, or 60 months) from the starting month forwards. For the buy-and-hold abnormal returns (`BHAR`) these portfolios are equally weighted portfolios and for the continuously compounded abnormal returns (CCAR) and abnormal Sharpe ratios (ASR) the portfolios are simply cross-sectional monthly averages of the monthly log-returns (in CCARs) and Sharpe-ratios (in ASRs).

(c) *Size/book-to-market matched reference stock*

A size and book-to-market matched reference stock is a stock of a firm that belongs in the beginning month of the holding period to the same size and book-to-market category (constructed in design (a)) as the sampled firm. From these, firms with market value (in the beginning of the event period) between 70 percent and 130 percent of the sampled firm are identified and the one with book-to-market ratio closest to the sampled firm is selected as the reference firm whose stock is the reference stock.

4.2 Simulations designs

In our simulations, like Lyon, Barber, and Tsai [26], we utilize two different *random* sample designs and various different *non-random* sample designs to choose $n = 200$ firms. In the first random sample design, $n = 200$ firms are randomly selected *with* replacement and 200 event months are randomly selected also *with* replacement, however, such that if a firm becomes sampled multiple times it cannot have twice the same event period. We refer to this design as RANDOM. In the second random design we sample in each simulation round randomly one event day and $n = 200$ firms *without* replacement. This design is referred to as RANDOM CLUSTER to indicate completely clustered event months with high cross-correlation between the returns. In each design we repeat the procedure $1,000$ times to get the empirical distributions of the investigated test statistics.

The reference portfolios are constructed such that only stocks that have complete return series in the holding period are included. Thus, stocks delisted during the holding period h and those listed after the event day are excluded from the reference portfolio. This procedure should make the characteristics of the firms in the reference portfolio closely match those of the sampled firms during the holding period. Each of the above sampling designs is simulated 1,000 times to produce empirical distributions of the test statistics.

5 Simulation results

5.1 Random samples

In this section we report results of the RANDOM design for non-clustered and clustered event days. In the clustered event case we deal with the special case where in each simulation round the sampled 200 stocks all share the same event month. On the basis of the empirical results we will see that in fact the complete clustering is not an issue from which we can deduce that a lower level clustering, where some of the event months are the same and others are more or less close to each other, does not cause particular problems in long run event study testing. Indeed, the empirical simulation results suggest that clustering in fact improves the power of some test procedures.

5.1.1 Non-clustered event months

This section reports results of the random designs (RANDOM) for 12, 36, and 60 months event period. Table 2 reports sample statistics of the tests of this design when there is no event effect. It is notable that in particular `BHAR-T` with buy-and-hold size/book-to-market portfolio adjusted returns is most negative. The skewness adjustment (`BHAR-TSA`) improves the situation. Also if the reference stock adjusted abnormal returns are used the average `BHAR-T` is quite close to zero. The means of `CCAR-T` and `SHARP-T` for abnormal returns defined with respect to the market-wide references are positive and relatively large in the shortest window of 12 months. In the longer windows (36 and 60 months) they are close to zero. The standard deviations of all test statistics are fairly close to the theoretical value of unity. It is notable that in terms of the sample statistics there are no material differences whether the abnormal returns are defined by using the size and book-to-market matched portfolios or the market portfolio.

[Table 2]

Size

Table 3 reports empirical rejection rates at the nominal significance levels of 1%, 5%, and 10% on both tails separately as well as in the two-sided case under the null hypothesis of no event effect. The respective cut-off values corresponding the significance levels are from the normal distribution ± 2.58, ± 1.96, and ± 1.65.

In particular `BHAR-T` suffers size distortion by over-rejecting if the abnormal returns are defined with respect to the matched reference portfolio or market-wide portfolio. Even the skewness adjustment (`BHAR-TSA`) does not materially improve the situation. The tendency is that these statistics over-reject negative abnormal returns while positive abnormal returns tend to be under-rejected. These results are consistent with earlier findings (e.g., Lyon, Barber, and Tsai [26, Table III]). The skewness of the return distribution of `BHAR` is the reason for these results such that a sample of size 200 observations is not yet large enough to warrant reliable Central Limit Theorem based large sample approximation to the distribution of these statistics. As Table 3 shows, reliable size results by `BHAR-T` are obtained only if the abnormal returns are defined in terms of the reference stock. However, as will be seen below, this leads to material power losses.

Also `CCAR-T` and `SHARPE-T` suffer some size distortion in the shortest window of 12 months in particular in the case of the market-wide benchmark. In the longer windows of 36 and 60 months `SHARPE-T` is rejecting generally close to the nominal rates when abnormal behavior is defined with respect to market-wide benchmark or a reference stock. `CCAR-T` is rejecting close to the nominal in the 36 months window for all benchmarks.

Generally even in the cases with statistically significant deviations from the nominal rejection rate, size distortions of `CCAR-T` and `SHARPE-T` are numerically quite small, in particular compared to those of `BHAR-T` and `BHAR-TSA`.

[Table 3]

Power

Tables 4 and 6 with with graphical depicts collected to Figure 1 report the powers of the test statistics for the RANDOM design. There are three outstanding results.

First, although `BHAR-T` and `BHAR-TCA` tend to over-reject the null hypothesis of no event effect (the ± 0 row in the panels), their powers on all event windows are far inferior to `CCAR-T` and in particular `SHARPE-T`.

[Tables 4, 5, and 6]

Second, comparing powers of test based on abnormal returns defined with respect to portfolio (market or matched) in Panels A, B, D, E, G, and H with tests based on abnormal returns defined with respect to reference stock in Panels C, F, and I in Figure 1, one sees the dramatic power loss due to the matched *stock* method (Panels C, F, I) compared to those of the matched *portfolio* method (Panels A, D, G) or the *market-wide* reference portfolio method (Panels B, E, H). It is also notable that comparing the results of `SHARPE-T` in Table 6 with the `BHAR-T` and `BHAR-TSA` in Table 4 or Table 5, we see that the power of `SHARPE-T` is even with the matched reference stock method (Table 6) more powerful than the `BHAR-T` and `BHAR-TSA` with abnormal returns defined with respect to reference portfolio (matched or market-wide). For example, in the three years event window, Panels B of *matched portfolio* method in Table 4 shows that a 12% abnormal return is detected by `BHAR-T` with probability 0.300 and by `BHAR-TSA` with probability 0.473, while in terms of `SHARPE-T` with *matched stock* method it is detected with probability 0.559 (Table 6, Panel B). With reference portfolio method based abnormal returns a 12% abnormal return is detected by `SHARPE-T` with probability as high as 0.831 (Panel B of Table 4), which is more than two times better than that of `BHAR-T` and almost two times better than that of `BHAR-TSA`, even though these two statistics tend to over-reject the null-hypothesis.

[Figure 1]

Third, as Tables 4 and 5, and Figure 1 (Panels A, B, D, E, G, H) show it does not make a big difference in terms of size and power of the tests whether the abnormal returns are defined with respect to the matched portfolios or with respect to the market-wide portfolio. The latter is far easier to construct.

Thus, in all the results show the benefit of scaling the returns by the standard deviation and empirically confirm the advantages of adjusting returns by their statistical accuracy, which naturally also lead to economically well defined risk adjusted returns (Sharpe ratios). Later we deal with empirical properties of the Sharpe-ratio based approach in non-random samples. Before that we investigate size and power properties the tests in the case of completely clustered event months, i.e., when the event month is the same for all stocks in the sample.

5.1.2 Clustered event months

Size

Similar to Table 3, Table 7 reports the rejection rates of the test statistics in different event windows and at the nominal significance levels of 1%, 5%, and 10% when the event month is the same for all firms in a sample (RANDOM CLUSTER design, see Section 4.2). Generally, all test statistics are approximately equally well specified in the RANDOM CLUSTERED as in the RANDOM design.

At first glance this might seem surprising, because clustered event months imply cross-sectionally correlated returns, which biases standard error estimates (see e.g., Kolari and Pynnonen [19]). Indeed, on average the returns are cross-sectionally correlated in our samples, such that in our $1,000$ simulations of $n = 200$ stock samples the average monthly (log) return cross-correlations are 0.165, 0.183, and 0.186 for the 12, 36, and 60 months event windows. However, the average cross-correlations of the abnormal returns turned out to be generally about zero. This implies that cross-correlation becomes eliminated from the abnormal returns on average and thus does not impose size distortion to the test statistics. This is consistent with earlier findings (e.g., Lyon, Barber, and Tsai [26, p. 188]).

An explanation of the disappearance of the cross-correlation can sketched in terms of a factor model. That is, in general terms, let r_i denote the return of stock i, $i = 1, \ldots,$ that obey a factor model

$$r_i = \beta_i' \mathbf{f} + u_i, \tag{57}$$

where β_i is a vector of factor coefficients conforming the dimension of the common factors in vector $\mathbf{f}$ and the prime denotes transposition. The common factors capture the cross-correlation of the returns such that the u_is are uncorrelated with each other and with the common factors. Let $r_m = \beta_m' \mathbf{f}$ denote the return of the equally weighted market portfolio (in which the idiosyncratic component cancels out). Assume we have a sample n stocks in our event study. Then the average abnormal return becomes

$$\mathtt{ar} = \bar{r} - r_m = (\bar{\beta}_n - \beta_m)' \mathbf{f} + \bar{u}, \tag{58}$$

where $\bar{r}$, $\bar{\beta}_n$, and $\bar{u}$ are the arithmetic means of r_i, β_i, and u_i. The variance of $\mathtt{ar}$ becomes

$$\mathrm{var}[\mathtt{ar}] = (\bar{\beta}_n - \beta_m)' \Sigma_f (\bar{\beta}_n - \beta_m) + \mathrm{var}[\bar{u}], \tag{59}$$

where Σ_f is the covariance matrix of the common factors. Now, as the sample size increases $\bar{\beta}_n \to \beta_m$ such that the covariance term $(\bar{\beta}_n - \beta_m)'\Sigma_f(\bar{\beta}_n - \beta_m)$ disappears and hence, the variance of the abnormal return becomes independent of the cross-correlations. Typically in long horizon event studies the sample sizes are relatively large (hundreds) such that the limiting behavior becomes a good approximation for the reality.

Power

To conserve space, we report only graphical results regarding the power results. Analogous to Section 5.1.1, we add different magnitudes of abnormal returns, indicated by the x-axes in the Panels of Figure 2, randomly across the event months. Again, `SHARPE-T` has superior power over the other methods. Also, it is notable that the complete clustering of the event days does not weaken the power of tests. In fact, it seems to even improve the symmetry of `BHAR-T` and `BHAR-TSA`. We have demonstrated these results in terms of the 36 months event window in Figure 3, in which the power results from the RANDOM and RANDOM CLUSTER designs are depicted in the same graph for each test statistic. As the figures shows the power functions are almost identical in each case.

[Figure 2]

[Figure 3]

5.2 Non-random samples

In this section we report statistical results of the investigated test statistics (Table 1) under different types of sampling designs that typically are termed as non-random samples (Lyon, Barber, and Tsai [26], Mitchell and Stafford [28], Jegadeesh and Karceski [18]). The investigated designs according to which the samples are constructed are: (1) Overlapping event periods, (2) Firm size, (3) Book-to-market, (4) Pre-event return performance, and (5) industry clusterings. In each design we draw $1,000$ samples of 200 event samples to asses the performance of selected test statistics.

5.2.1 Overlapping event periods

In particular for the longer end event periods of five years the likelihood of multiple events increases. An example referred to by Lyon, Barber, Tsai [26, p. 190] is Microsoft's common stock split in April 1990, and June 1992. In this particular case three and five years event windows would have respectively 11 and 35 overlapping months. Clearly these two series are not independent. In fact, for example in the five-year window assuming zero autocorrelation, equation (51) implies that the theoretical covariance of the five year log-returns of these series equals $35\sigma_i^2$, where σ_i^2 is the monthly variance. Unlike for clustered event days discussed in Section 5.1.2, reference portfolio or reference stock does not eliminate the cross-correlation. However, as our simulation results show, moderate number of stocks with multiple event within the same event window do not materially bias `CCAR-T` and `SHARPE-T` tests even if the cross-correlation is not accounted for.

[Table 8]

Table 8 reports results of samples in which 10%, 50%, and 100% of firms are sampled twice with overlapping event event periods. The overlapping cases are generated such that for the same firm two separate event months are randomly selected such that the difference between the months are at most $h-1$, where h is the length of the event period. Thus, in the 10% overlapping design we select in each of the 200 event samples first randomly 10 firms and two random event months (i.e., altogether 20 months) and then the rest 180 firms are sampled such that they do not have overlapping event months. In the 50% case 50 firms have two event months and finally in the 100% case 100, i.e., all firms in have two events such that the event periods share at least one common month. Each design $1,000$ respective samples are drawn to asses the empirical behavior of the test statistics.

The results show two major results. First, although multiple events with overlapping event periods generate cross-correlation, even a relatively high such as 10% of cases share the overlapping symptom it does not materially affect the performance of the test statistics. In particular `CCAR-T` and `SHARPE-T` reject close to the nominal rates in each event period horizons whichever benchmark is used. `BHAR-T` is well specified with reference stock while `BAHR-TSA` tend to over-reject on the lower tail and in two-sided testing in three and five year horizons. Thus, the cross-correlation is not a serious problem for a moderate number of double events within the same event period for a firm. In fact, while the average cross-correlation of the return series in our simulation samples was e.g., in the 60 months window 0.035, the average cross-correlation of the abnormal Sharpe-ratio was only 0.00018.

The second major result is that when the share of stocks with two overlapping events increases the cross-correlation starts to bias in particular the statistics that do not account for it (`BHAR-T`, `BHAR-TSA`, `CCAR-T`, and `SHARPE-T`). In the 50% case the rejection rates of these statistics are typically more than one and half times the nominal rate and more than double when all firms have two overlapping event periods (the 100% case). In addition use of the market as a reference increases miss-specification in almost all cases. While the average return cross-correlation of 0.038 for example in the 60 months window is not much higher than in the 10% case, the average cross-correlation of the abnormal Sharpe ratio increases to 0.0024. This biases by equation (55) the standard deviation on average by a factor $\sqrt{1 + (200 - 1)0.0024} \approx 1.22$, or 22%. The biasing effect explains partially the miss-specification. In the above 10% case the standard deviation of e.g., `BHAR-T`, `BHAR-TSA`, `CCAR-T`, and `SHARPE-T` are respectively 1.11, 1.12, 1.01, and 1.01 for the matched market portfolio in the 60 months event period. The corresponding standard deviations in the 100% case are 1.25, 1.27, 1.21, and 1.18, while for cross-correlation adjusted statistics `CCAR-TCA` and `SHARPE-TCA` they are 1.00 and 0.99, i.e., virtually the same as the theoretical standard deviations. However, the means are respectively 0.65 and 0.48 which deviate highly significantly from zero. Thus, while the adjustment of the standard deviations account the cross-correlation bias, return adjustment by the reference portfolios or even by the matched reference stock do not reflect properly the average abnormal return if there is a very large number of the multiple event stocks with overlapping event windows. However, as the above results show, there is no discernible bias if the number of these cases is moderate.

5.2.2 Firm size

For example Boehme and Sorescu [2] find long horizon price drift following dividend initiations and resumptions only for small stocks. To investigate the impact of firm size based sampling on the test statistics, we draw separately from each of the smallest and largest size deciles $1,000$ samples of 200 events. The deciles are constructed according to NYSE market sizes as described in Section 4.1.

Table 9 report the simulation results. It turned out that in these sub-populations the overall market portfolio failed to work because of the general tendency of small firms having average returns higher than the market average and large firms having on average lower returns than the market average, a phenomenon reported also in Lyon, Barber, and Tsai [26, Table I]. Accordingly, we report only results using the size/book-to-market matched portfolios and size/book-to-market matched stocks as benchmarks.

Generally, with these benchmarks all test statistics are reasonably well specified in the large firms (Panel A of Table 9). In small firm samples (Panel B) major miss-specifications for all tests, except `BHAR-T` with size/book-to-market matched reference portfolio, are on the upper 5% tail in the three-year event horizon. Our `CCAR-T` and `SHARPE-T` are typically, however, better specified than `BHAR` tests. The results of `BHAR-T` and `BHAR-TSA` are consistent with those reported in Lyon, Barber, and Tsai [26, Sec. B.1.].

5.2.3 Book-to-market ratio

To asses the impact of book-to-market based sampling on the test statistics, we draw separately from the smallest and largest book-to-market deciles $1,000$ samples of 200 events. The book-to-market deciles are formed in terms of the NYSE stocks with AMEX and NASDAQ firms placed into appropriate deciles. The book-to-market values of are defined according to the December of previous year with market value on the last day of the month.

[Table 10]

Table 10 reports the empirical sizes from the simulations. In the high book-to-market samples `CHAR-T` and `SHARPE-T` reject generally statistically at the nominal 5% rate in the one year and three year horizons. In the five year window all tests over-reject in the lower tail and two-tailed tests with the size/book-to-market reference portfolio. Reference stock abnormal returns produce generally reasonably well specified results in all tests. In the low book-to-market samples `BHAR-T` rejection is correctly specified with the reference stock abnormal, while `CCAR-T` and `SHARPE-T` over-reject to some extend on the lower tail in the three-year horizon. Reference portfolio based abnormal returns yield reasonably close rejection rates only in the one-year horizon with `CCAR-T` and `SHARPE-T`. `BHAR-T` over-rejects on the lower tail and under-rejects on the upper tail in all event horizons with reference portfolio abnormal returns.

The general conclusion is that the size/book-to-market based reference portfolios do not fully reflect the return characteristics when the population of firms is restricted to these extreme book-to-market size deciles. Thus, if in practical applications it is known that the sample is biased specially to small book-to-market firms the reference portfolio should also be constructed in terms of these firms. The lowest decile is populated more heavily with NASDAQ stocks (NASDAQ 67%, Amex 9%, and NYSE 24% in our samples) while in the highest book-to-market decile the percentage are NASDAQ (55%), Amex (21%), and NYSE (24%). Our next experiment demonstrates more concretely the importance of matching the reference portfolio in line with the population characteristics from which the stocks are sampled.

5.2.4 Pre-event return performance

Recent stock price appreciation is considered to be an important factor to motivate issue equity (Graham and Harvey [14], Campello and Graham [10]). Similarly share repurchases are generally preceded by a period of low stock returns (Lyon, Barber and Tsai [26]). Equity and debt issuers tend to underperform matching non-issuers in the post-issue years (Lyanders, Sun, and Zhang [25]).

To asses the impact of recent price performance on the test statistics, we follow the strategy implemented in Lyon, Barber, and Tsai [26, Sec B.3.] and calculate the preceding six-month buy-and-hold return for all firms in each month from July 1973 through December 2009 and rank all firms in deciles on the basis of their six-month performance preceding the ranking month. We draw separately $1,000$ sets of 200 firms sampled with replacement from the lowest and highest return deciles. In computing abnormal returns, we use in addition to the size/book-to-market reference portfolios, an equally weighted reference portfolio including all stocks in the highest decile and a similarly constructed portfolio of the lowest decile stocks. In the reference stock method the stock which has the closest book-to-market value with the sample stock in the decile is used as the reference.

[Table 11]

Results are reported in Table 11. The table clearly shows the importance of accounting the pre-event characteristics of the subset from which the stocks are sampled. The portfolio method which matches the size and book-to-market characteristics with the sample stock yields badly badly biased rejection rates in particular in the one-year event in all test statistics. The bias is due to the miss-match of the average stock returns and reference portfolio returns. The average buy-and-hold returns over the simulations are in the 1-year, 3-year, and 5-year windows for the high (low) deciles are 33.9% (5.0%), 78.0% (59.4%), and 141.3% (130.0%), while the corresponding average buy-and-hold reference portfolio returns are 17.7% (15.6%), 64.8% (64.8%), and 130.0% (130.9%). Thus, in particular in the 1-year window the average stock returns differ substantially from the benchmark and hence point out the need of finding better matching in constructing a benchmark.

Using the decile based reference largely solves the problem in particular with `CCAR-T` and `SHARPE-T`. In terms of these statistics, the corresponding decile based reference portfolio and reference stock in all horizons rejection rates that statistically coincide with the nominal rate (0.05) in both tails as well as in two-tailed testing. Of these the reference portfolio is far more preferable because of the material power gains as demonstrated in of the three-year event window in Figure 4. Panels A and C plot the powers for `BHAR-T`, `CCAR-T`, and `SHARPE-T` with pre-event decile matched portfolios and Panels B and D the corresponding powers with pre-event size decile matched firms. The figure also demonstrates the superior power of `SHARPE-T` over `CCAR-T` and in particular over `BHAR-T` and `BHAR-TSA` which ever benchmark is used.

[Figure 4]

5.2.5 Industry clustering

To be included.

5.3 Calendar time portfolio method

Loughran and Ritter [23], Brav and Gombers [5], and Boehme and Sorescu [2], among others, employ the Fama-French factor-type models to analyze long-horizon event effects. The performance and properties of Fama-French three-factor based calendar time is widely investigated in Mitchell and Stafford [28]. A major problem with the approach is the "bad model" problem discussed in Fama [12] which stems from the fact the the specified factors do not capture all the common return characteristics of stocks.

In order to alleviate this problem Boehme and Sorecsu [2] (see also Mitchell and Stafford [28, Sec V] utilize a model they call an adjusted Fama-French model. In this approach the difference of the test asset and its size/book-to-market matched control stock is regressed on the Fama-French factors. Given that the reference stock has similar characteristics as the test asset, this approach has the potential to eliminate all unknown common factors from the resulting abnormal return. We follow this practice and form calendar time portfolios

$$(R_{\texttt{test}} - R_{\texttt{control}})_{p,t} = \alpha_p + \beta_p(R_{m,t} - R_{f,t}) + s_p\texttt{SMB}_t + h_p\texttt{HML}_t + e_{p,t}, \qquad (60)$$

where $(R_{\texttt{test}} - R_{\texttt{control}})_{p,t}$ is the monthly portfolio return (equally weighted or value weighted) of the difference of the simple returns of the test asset and its size/book-to-market matched control firm, $R_{m,t}$ is the monthly simple return of the value weighted market index, $R_{f,t}$ is the monthly return of the three month Treasury bills, SMB is the monthly Fama-French small-minus-big factor return, and HML is the monthly Fama-French high-minus-low factor return. In month t the portfolio $(R_{\texttt{test}} - R_{\texttt{control}})_{p,t}$ includes all those stocks whose event period includes the month. Thus, the number of stocks, n_t, can vary monthly from zero to the total number of sampled stock. The month index t is running from the earliest month to the latest month of the event periods of the stocks in the sample and months with $n_t = 0$ are discarded from the analysis.

The parameter α_p in (60) is the abnormal return parameter. The corresponding regression t-ratio gives the statistic for testing the abnormal return. We refer to this statistic as `CALENDAR-T`, which is defined as

$$t_{\texttt{calendar}} = \frac{\hat{\alpha}_p}{s(\hat{\alpha}_p)}, \tag{61}$$

where $\hat{\alpha}_p$ is an estimator of α_p and $s(\hat{\alpha}_p)$ is the corresponding standard error of $\hat{\alpha}_p$.

Empirical rejection rates at the nominal 5% levels under different sample designs discussed in Sections 5.1 and 5.2 are reported in Table 12. The results are based on the $1,000$ samples of 200 event months according to the sample designs described in Sub-sections 5.1.1 through 5.2.4. For each sample the calendar time portfolio is abnormal returns according to model (60) are computed for equally weighted and value weighted portfolio returns by ordinary least squares (OLS). The varying number of stocks in different calendar months causes heteroscedasticity into the residuals. The heteroscedasticity can be accounted by utilizing weighted least squares (WLS). The results, however, proved to be qualitatively the same as those of OLS. Accordingly we report only these results.

[Table 12]

The calendar time portfolio method is relatively well specified in all designs for the equally weighted as well as value weighted portfolios. The results are in Table 12. In the case of the equally weighted portfolios reported in Panel A of the table, most of the significantly deviating rejection rates are because of under-rejection, i.e., the empirical rejection rate is significantly lower that the nominal level. The most sever under-rejection is in the two-tailed testing in the "Low pre-event returns" design for the 3 years holding period, in which the empirical rejection rate is 0.18, i.e., about one third of the nominal rate 0.05. In this regard the results are consistent with those of Lyon, Barber, and Tsai [26, Table X], who report only one-sided (over-rejection) binomial test significances in their table. However, visual inspection of Panel A of Table X in Lyon, Barber, and Tsai [26] indicates a similar tendency of under-rejection as our explicit testing shows. Compared with our results, the biggest difference is in the "Low book-to-market ratio" design, in which Lyon, Barber, and Tsai report severe over-rejection in the lower tail ([26, Panel A of Table X]), whereas our results do not indicate such symptom.

For the value weighted results, reported in Panel B of Table 12, most of the significant deviations from the nominal levels are over-rejections, which in most cases are not terribly bad. Only in the designs "Small firms" and "Low pre-event returns" the over-rejections are severe, being at worst from two to five times the nominal rate of 0.05.

Thus, in all the calendar time testing in terms of Fama-French factors regressed on the return differences of a test stock and its size/book-to-market matched control firm stock return yields reasonably well specified test results. However, as Figure 5 demonstrates, the calendar time method suffers from a very weak power compared to the Sharpe-ratio based methods suggested in this paper. For example an abnormal return of size 10% is detected by `CALENDAR-T` only by about 0.15

probability, while `SAHRPE-T` detects it with reference stock method with probability 0.45 and with reference stock method with probability of 0.75. Thus, the calendar time approach turns out to be very weak in detecting abnormal return behavior compared in particular to the Sharpe-ratio based tests.

[Figure 5]

6 Conclusions

This paper proposes an innovative testing approach based on standardized Sharpe ratios to test for abnormal returns in long-horizon (12-to-60 months) event effects. Simulations with actual returns demonstrate that the suggested test statistics do not have material size distortion and that the power of the statistics based on Sharpe ratios are superior to the conventional t-tests popular in long-horizon event studies. Also, the results show that event day clustering in random samples does not distort the power or size of test statistics. Finally, the reference portfolio turns out to be central in the definition of the abnormal return in order to keep the size of the test under control. In terms of the size of the tests, the choice of size/book-to-market matched reference stocks is often a good candidate, but unfortunately this procedure leads to a material power loss in testing. The calendar time approach is another event study popular methodology. It is recommended in particular due to its implicit control of potential cross-sectional correlation in abnormal returns. However, contrary to this common belief, we show that cross-correlation is very rarely a problem in long-horizon event study testing. Because of its very weak power, calendar time based testing is not recommended as a useful approach in long-horizon event study testing.

References

[1] Barber, Brad M. and john D. Lyon, 1997, Detecting long-run abnormal stock returns: The empirical power and specification of test statistics, *Journal of Financial Economics* 43, 341–372.

[2] Boehme, Rodney D. and Sorin M. Sorescu, 2002, The long-run performance following dividend initiations and resumption: underreaction or product chance? *Journal of Finance* 57, 871–900.

[3] Boehmer, Ekkehart, Jim Musumeci, and Anette B. Poulsen, 1991, Event-study methodology under conditions of event-induced variance, *Journal of Financial Economics* 30, 253–272.

[4] Brav, Alon, Inference in long-horizon event studies: A Bayesian approach with application to initial public offerings, *Journal of Finance* 55, 1979–2016.

[5] Brav, Alon and Paul A. Gombers, 1997, Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture and non-venture capital-backed companies, *Journal of Finance* 52, 1791–1821.

[6] Brock, William, Josef Lakonishok, and Blake LeBaron, 1992, Simple trading rules and the stochastic properties of stock returns, *Journal of Finance* 47, 1731–1764.

[7] Brown, Stephen J., and Jerold B. Warner, 1985, Using daily stock returns: The case of event studies, *Journal of Financial Economics* 14, 3–31.

[8] Buyn, Jinho and Michael S. Rozeff, 2003, Long run performance after stock splits: 1927 to 1996, *Journal of Finance* 58, 1063–1085.

[9] Campbell, John Y., Andrew W. Lo, and A. Graig MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton: Princeton University Press.

[10] Campello, Murillo and John R. Graham 2013, Do stock prices influence corporate decisions? Evidence from the technology bubble, *Journal of Financial Economics* 107, 89–110.

[11] Eberhart, Allan C. and Akhtar Siddique, 2002, The long-term performance of corporate bonds (and stocks) following seasoned equity offers, *Review of Financial Studies* 15, 1385–1406.

[12] Fama, Eugene F., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283–306.

[13] Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

[14] Grham, John R. and Campbell R. Harvey, 2001, The theory and practice of corporate finance: evidence from the field, *Journal of Financial Economics* 60, 187–243.

[15] Gombers, Paul A. and Josh Lerner, 2003, The really long run performance of initial public offerings: the pre-Nasdaq evidence, *Journal of Finance* 58, 1355–1392.

[16] Ikenberry, David, Josef Lakonishok, and Theo Vermaelen, 1995, Market underreaction to open market share repurchases, *Journal of Financial Economics* 39, 181–208.

[17] Jaffe, Jeffrey, 1974, Special information and insider trading, *Journal of Business* 47, 411-428.

[18] Jegadeesh, Narasimhan and Jason Karceski, 2009, Long-run performance evaluation: Correlation and heteroscedasticity-consistent tests, *Journal of Empirical Finance* 16, 101–111.

[19] Kolari, James W. and Seppo Pynnonen, 2010, Event study testing with cross-sectional correlation of abnormal returns, *Review of Financial Studies* 23, 3996–4025.

[20] Kolari, James W. and Seppo Pynnonen, 2011, Nonparametric rank tests for event studies, *Journal of Empirical Finance* 18, 953–971. (Also, SSRN Working Papers http://ssrn.com/abstract=1254022.)

[21] Kothari, S. P. and Jeroid B. Warner, 1997, Measuring long-horizon security price performance, *Journal of Financial Economics* 43, 301–340.

[22] Kothari, S. P. and Jeroid B. Warner, 2006, Econometrics of event studies, in: B. E. Eckbo, ed., *Handbooks of Corporate Finance: Empirical Corporate Finance*, Chapter 1 (Amsterdam: Elsevier/North-Holland).

[23] Loughran, Tim and Jay R. Ritter, 1995, The new issues puzzle, *Journal of Finance* 50, 23–52.

[24] Loughran, Tim and Jay R. Ritter, 2000, Uniformly least powerful tests of market efficiency, *Journal of Financial Economics* 55, 361–389.

[25] Lyenders, Evgeny, Le Sun, and Lu Zhang, 2008, The new issues puzzle: Testing the investment-based explanation, *Review of Financial Studies* 21, 2825–2855.

[26] Lyon, Johan D., Brad M. Barber, and Chih-Ling Tsai, 1999, Improved methods for tests of long-run abnormal stock returns, *Journal of Finance* 54, 165–201.

[27] Mandelker, Gershon, 1974, Risk and return: The case of merging firms, *Journal of Financial Economics* 1, 305–335.

[28] Mitchell, Mark L. and Erik Stafford, 2000, Managerial decisions and long-run stock price performance, *Journal of Business* 73, 287–329.

[29] Patell, James A., 1976, Corporate forecasts of earnings per share and stock price behavior: Empirical test, *Journal of Accounting Research* 14, 246–276.

[30] Ritter, Jay R., 1991, The long-run performance of initial public offerings, *Journal of Finance* 46, 3–27.

[31] Sutton, Clifton D., 1993, Computer-intensive methods for tests about the mean of an asymmetrical distribution, *Journal of the American Statistical Associtation* 88, 802–810.

Table 1: Summary of statistical tests

Test statistic	Return type	Return scaling	Cross-correlation	Mathematical form
`BHAR-T`: eq. (8)	Simple	Non-standardized	Non-robust	$t_{\mathtt{bhar}} = \frac{\overline{\mathtt{BHAR}}\sqrt{n}}{s_{\mathtt{bhar}}}$
`BHAR-TSA`: eq. (9)	Simple	Non-standardized	Non-robust	$t_{\mathtt{bharsa}} = t_{\mathtt{bhar}} + \frac{\hat{\gamma}}{3\sqrt{n}}\left(t^2_{\mathtt{bhar}} + \frac{1}{2}\right)$
`CCAR-T`: eq. (23)	Log-return	Non-standardized	Non-robust	$t_{\mathtt{ccar}} = \frac{\overline{\mathtt{CCAR}}(h_1,h_2)\sqrt{n}}{\tilde{s}_{\mathtt{ccar}}(h_1,h_2)}$
`SHARPE-T`: eq. (30)	Log-return	Standardized	Non-robust	$t_{\mathtt{sharpe}} = \frac{\overline{\mathtt{ASR}}(h_1,h_2)}{\tilde{s}_{\mathtt{asr}}(h_1,h_2)}$
`CCAR-TCA`: eq. (53)	Log-return	Non-standardized	Robust	$t_{\mathtt{accar}} = t_{\mathtt{ccar}} \times \sqrt{\frac{1-\hat{\theta}_{\mathtt{ccar}}}{1+(n-1)\hat{\theta}_{\mathtt{ccar}}}}$
`SHARPE-TCA`: eq. (55)	Log-return	Standardized	Robust	$t_{\mathtt{sharpeca}} = t_{\mathtt{asr}} \times \sqrt{\frac{1-\hat{\theta}_{\mathtt{sharpe}}}{1+(n-1)\hat{\theta}_{\mathtt{asr}}}}$

In `CCAR-T` and `SHARPE-T` the $\hat{\theta}_{\mathtt{ccar}}$ and $\hat{\theta}_{\mathtt{asr}}$ are estimators of the covariance-variance ratios using the procedure discussed in detail in Section 3.5.1. In all tests except `BHAR-TSA` the performance is investigated for three types of benchmarks: (a) buy-and-hold size/book-to-market matched portfolios, (b) market-wide portfolios, and (c) size/book-to-market matched reference stocks. Construction of the benchmarks are discussed in Section 4.1.

Table 2: Sample statistics of 1,000 simulations of the RANDOM design for $n = 200$ event firms with event periods of 12, 36, and 60 months.

Statistic	Benchmark	Mean	Median	Min	Max	Std	Skew	Kurt
Panel A: Annual (12 months) abnormal returns		%	%	%	%	%		
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	−0.15	−0.01	−3.81	3.00	1.07	−0.51	0.25
BHAR-T	Buy-and-hold equally weighted market portfolio	−0.21	−0.07	−4.20	3.14	1.08	−0.56	0.45
BHAR-T	Size/book-to-market matched reference stock	0.02	0.02	−3.02	2.88	0.98	−0.06	−0.21
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	−0.05	0.03	−3.58	3.75	1.10	−0.23	0.21
BHAR-TSA	Buy-and-hold equally weighted market portfolio	−0.11	−0.03	−3.89	4.10	1.10	−0.24	0.39
CCAR-T	Average log-returns of size/book-to-market stocks	0.06	0.08	−2.86	3.94	1.02	−0.02	0.22
CCAR-T	Market-wide average of log-returns	0.34	0.33	−2.83	4.63	1.02	0.08	0.33
CCAR-T	Size/book-to-market matched reference stock (log ret)	−0.09	−0.06	−2.87	3.39	1.00	0.01	−0.04
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	−0.02	−0.02	−3.16	3.24	1.04	−0.02	0.01
SHARPE-T	Market-wide average of Sharpe ratios	0.36	0.34	−2.88	3.71	1.04	0.00	−0.07
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	−0.11	−0.13	−3.33	3.80	0.99	0.03	0.29
Panel B: Three-year (36 months) abnormal returns								
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	−0.36	−0.25	−5.06	2.54	1.10	−0.64	0.57
BHAR-T	Buy-and-hold equally weighted market portfolio	−0.03	−0.08	−3.02	3.10	1.02	0.09	−0.24
BHAR-T	Size/book-to-market matched reference stock	−0.36	−0.23	−5.03	2.57	1.11	−0.65	0.53
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	−0.23	−0.20	−4.67	3.00	1.09	−0.33	0.30
BHAR-TSA	Buy-and-hold equally weighted market portfolio	−0.22	−0.18	−3.93	3.03	1.08	−0.25	0.07
CCAR-T	Average log-returns of size/book-to-market stocks	−0.09	−0.10	−3.54	3.38	0.98	0.01	0.19
CCAR-T	Market-wide average of log-returns	−0.09	−0.12	−3.38	3.30	0.99	0.05	0.07
CCAR-T	Size/book-to-market matched reference stock (log ret)	0.00	0.00	−3.12	3.52	1.01	−0.02	0.06
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	−0.11	−0.12	−3.56	3.10	0.98	−0.06	0.24
SHARPE-T	Market-wide average of Sharpe ratios	−0.12	−0.11	−3.53	2.77	0.99	−0.09	0.04
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	−0.01	0.00	−3.60	3.16	1.01	0.01	0.09

Continues

Table 2 Continues.

Statistic	Benchmark	Mean	Median	Min	Max	Std	Skew	Kurt
Panel C: Five-year (60 months) abnormal returns		%	%	%	%	%		
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	−0.29	−0.10	−4.91	2.34	1.17	−0.69	0.58
BHAR-T	Buy-and-hold equally weighted market portfolio	−0.04	−0.04	−3.20	2.82	1.01	−0.01	−0.24
BHAR-T	Size/book-to-market matched reference stock	−0.31	−0.12	−4.96	2.55	1.19	−0.72	0.68
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	−0.14	−0.05	−4.41	3.17	1.16	−0.35	0.33
BHAR-TSA	Buy-and-hold equally weighted market portfolio	−0.14	−0.06	−4.17	3.50	1.14	−0.22	0.15
CCAR-T	Average log-returns of size/book-to-market stocks	0.09	0.07	−2.94	2.94	1.06	0.02	−0.18
CCAR-T	Market-wide average of log-returns	0.11	0.08	−3.03	3.50	1.06	0.06	−0.16
CCAR-T	Size/book-to-market matched reference stock (log ret)	−0.06	−0.06	−3.57	2.96	1.00	−0.04	−0.10
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	0.04	0.05	−3.21	3.14	1.05	−0.03	−0.16
SHARPE-T	Market-wide average of Sharpe ratios	0.04	0.08	−3.17	3.35	1.04	−0.06	−0.29
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	−0.10	−0.11	−4.02	3.06	1.01	−0.09	−0.12

BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from the sample period from July, 1973 through December 2009.

Table 3: Empirical rejection rates of the test statistics under the null hypothesis of no event effect in the RANDOM design at the nominal 1%, 5%, and 10% significance levels from 1,000 random portfolios of $n = 200$ securities from CRSP monthly data base sampled from the time period from July, 1997 to December, 2009.

Statistic	Benchmark	Lower 0.005	Upper 0.005	2-tail 0.01	Lower 0.025	Upper 0.025	2-tail 0.05	Lower 0.05	Upper 0.05	2-tail 0.10
Panel A: Annual (12 months) abnormal returns										
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	0.025^c	0.001^a	0.026^c	0.061^c	0.012^c	0.073^c	0.094^c	0.027^c	0.121^b
BHAR-T	Buy-and-hold equally weighted market portfolio	0.030^c	0.001^a	0.031^c	0.066^c	0.012^c	0.078^c	0.106^c	0.022^c	0.128^c
BHAR-T	Size/book-to-market matched reference stock	0.005	0.002	0.007	0.019	0.022	0.041	0.040	0.043	0.083^a
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	0.019^c	0.007	0.026^c	0.051^c	0.028	0.079^c	0.080^c	0.049	0.129^c
BHAR-TSA	Buy-and-hold equally weighted market portfolio	0.022^c	0.007	0.029^c	0.053^c	0.026	0.079^c	0.086^c	0.044	0.130^c
CCAR-T	Average log-returns of size/book-to-market stocks	0.005	0.009^a	0.014	0.025	0.032	0.057	0.048	0.054	0.102
CCAR-T	Market-wide average of log-returns	0.004	0.019^c	0.023^c	0.011^c	0.052^c	0.063^a	0.031^c	0.100^c	0.131^c
CCAR-T	Size/book-to-market matched reference stock (log-ret)	0.006	0.005	0.011	0.031	0.021	0.052	0.070^c	0.041	0.111
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	0.008	0.008	0.016^a	0.033	0.032	0.065^b	0.059	0.055	0.114
SHARPE-T	Market-wide average of Sharpe ratios	0.002	0.018^c	0.020^c	0.014^b	0.061^c	0.075^c	0.026^c	0.107^c	0.133^c
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	0.007	0.004	0.011	0.036^b	0.022	0.058	0.058	0.033^b	0.091
Panel B: Three-year (36 months) abnormal returns										
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	0.034^c	0.000^b	0.034^c	0.086^c	0.002^c	0.088^c	0.124^c	0.013^c	0.137^c
BHAR-T	Buy-and-hold equally weighted market portfolio	0.038^c	0.000^b	0.038^c	0.090^c	0.004^c	0.094^c	0.121^c	0.012^c	0.133^c
BHAR-T	Size/book-to-market matched reference stock	0.004	0.007	0.011	0.030	0.024	0.054	0.059	0.049	0.108
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	0.020^c	0.001^a	0.021^c	0.063^c	0.017	0.080^c	0.091^c	0.026^c	0.117^a
BHAR-TSA	Buy-and-hold equally weighted market portfolio	0.020^c	0.002	0.022^c	0.058^c	0.015^b	0.073^c	0.094^c	0.027^c	0.121^b
CCAR-T	Average log-returns of size/book-to-market stocks	0.007	0.005	0.012	0.025	0.017	0.042	0.053	0.033^b	0.086
CCAR-T	Market-wide average of log-returns	0.007	0.004	0.011	0.026	0.019	0.045	0.050	0.041	0.091
CCAR-T	Size/book-to-market matched reference stock (log-ret)	0.007	0.006	0.013	0.032	0.022	0.054	0.058	0.048	0.106
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	0.009^a	0.002	0.011	0.036^b	0.019	0.055	0.051	0.037^a	0.088
SHARPE-T	Market-wide average of Sharpe ratios	0.011^c	0.002	0.013	0.030	0.012^c	0.042	0.052	0.040	0.092
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	0.004	0.007	0.011	0.028	0.025	0.053	0.052	0.045	0.097

Continues

Table 3 Continues.

Statistic	Benchmark	Lower 0.005	Upper 0.005	2-tail 0.01	Lower 0.025	Upper 0.025	2-tail 0.05	Lower 0.05	Upper 0.05	2-tail 0.10
Panel C: Five-year (60 months) abnormal returns										
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	0.039^c	0.000^b	0.039^c	0.084^c	0.007^c	0.091^c	0.127^c	0.021^c	0.148^c
BHAR-T	Buy-and-hold equally weighted market portfolio	0.049^c	0.000^b	0.049^c	0.090^c	0.007^c	0.097^c	0.127^c	0.018^c	0.145^c
BHAR-T	Size/book-to-market matched reference stock	0.007	0.002	0.009	0.022	0.021	0.043	0.053	0.052	0.105
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	0.028^c	0.005	0.033^c	0.058^c	0.026	0.084^c	0.095^c	0.049	0.144^c
BHAR-TSA	Buy-and-hold equally weighted market portfolio	0.026^c	0.006	0.032^c	0.056^c	0.025	0.081^c	0.088^c	0.052	0.140^c
CCAR-T	Average log-returns of size/book-to-market stocks	0.004	0.008	0.012	0.030	0.044^c	0.074^c	0.054	0.076^c	0.130^c
CCAR-T	Market-wide average of log-returns	0.004	0.009^a	0.013	0.026	0.040^c	0.066^b	0.047	0.084^c	0.131^c
CCAR-T	Size/book-to-market matched reference stock (log-ret)	0.007	0.001^a	0.008	0.026	0.020	0.046	0.051	0.046	0.097
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	0.005	0.004	0.009	0.035^b	0.038^c	0.073^c	0.063^a	0.060	0.123^b
SHARPE-T	Market-wide average of Sharpe ratios	0.003	0.004	0.007	0.028	0.033	0.061	0.057	0.061	0.118^a
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	0.008	0.002	0.010	0.027	0.018	0.045	0.065^b	0.035^b	0.100

BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from the sample period from July, 1973 through December 2009 using the RANDOM design described in Section 4.2.
$a = 10\%$, $b = 5\%$, and $c = 1\%$ significant deviation from the nominal rejection rate indicated by the column header.

Table 4: Empirical rejection rates at the 5% level two-sided testing of different magnitudes of abnormal returns defined with respect to *size/book-to-market matched portfolios* in the RANDOM design with $1,000$ random portfolios of $n = 200$ securities for annual (PANEL A), three-year (PANEL B), and five-year (PANEL C) event windows.

Abn ret (%)	BHAR-T	BHAR-TCA	CCAR-T	SHARPE-T
Panel A: Annual (12 months) abnormal returns				
−10	0.755	0.653	0.879	0.994
−6	0.439	0.363	0.431	0.788
−4	0.273	0.229	0.212	0.447
−2	0.158	0.132	0.098	0.166
±0	0.073	0.079	0.057	0.065
+2	0.059	0.095	0.099	0.144
+4	0.133	0.207	0.272	0.439
+6	0.313	0.445	0.497	0.738
+10	0.783	0.851	0.880	0.986
Panel B: Three-year (36 months) abnormal returns				
−24	0.914	0.777	0.998	1.000
−12	0.607	0.474	0.679	0.907
−6	0.308	0.234	0.227	0.367
−3	0.189	0.135	0.096	0.137
±0	0.088	0.080	0.042	0.055
+3	0.043	0.059	0.076	0.095
+6	0.058	0.125	0.185	0.283
+12	0.300	0.473	0.581	0.831
+24	0.959	0.983	0.989	1.000
Panel C: Five-year (60 months) abnormal returns				
−25	0.842	0.660	0.974	0.998
−15	0.555	0.431	0.644	0.856
−10	0.380	0.289	0.333	0.530
−5	0.222	0.161	0.123	0.177
±0	0.091	0.084	0.074	0.073
+5	0.059	0.112	0.161	0.196
+10	0.134	0.257	0.397	0.534
+15	0.337	0.507	0.702	0.849
+25	0.834	0.918	0.971	1.000

BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from the sample period from July, 1973 through December 2009.

Table 5: Empirical rejection rates at the 5% level two-sided testing of different magnitudes of abnormal returns defined with respect to *market-wide portfolio* in the RANDOM design with 1,000 random portfolios of $n = 200$ securities for annual (PANEL A), three-year (PANEL B), and five-year (PANEL C) event windows.

Abn ret (%)	BHAR-T	BHAR-TCA	CCAR-T	SHARPE-T
Panel A: Annual (12 months) abnormal returns				
-10	0.769	0.655	0.787	0.982
-6	0.454	0.375	0.300	0.637
-4	0.285	0.235	0.130	0.309
-2	0.161	0.134	0.071	0.098
± 0	0.078	0.079	0.063	0.075
$+2$	0.060	0.088	0.159	0.248
$+4$	0.106	0.183	0.348	0.531
$+6$	0.266	0.394	0.571	0.821
$+10$	0.751	0.826	0.896	0.991
Panel B: Three-year (36 months) abnormal returns				
-24	0.910	0.742	0.995	1.000
-12	0.597	0.453	0.636	0.891
-6	0.314	0.222	0.208	0.370
-3	0.179	0.121	0.090	0.136
± 0	0.094	0.073	0.045	0.042
$+3$	0.049	0.060	0.083	0.096
$+6$	0.056	0.126	0.175	0.254
$+12$	0.279	0.451	0.548	0.769
$+24$	0.958	0.985	0.981	1.000
Panel C: Five-year (60 months) abnormal returns				
-25	0.835	0.600	0.965	1.000
-15	0.567	0.400	0.598	0.822
-10	0.390	0.274	0.315	0.478
-5	0.227	0.151	0.123	0.175
± 0	0.097	0.081	0.066	0.061
$+5$	0.064	0.108	0.164	0.183
$+10$	0.129	0.255	0.385	0.509
$+15$	0.315	0.504	0.664	0.797
$+25$	0.826	0.913	0.961	0.998

BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from the sample period from July, 1973 through December 2009.

Table 6: Empirical rejection rates at the 5% level two-sided testing of different magnitudes of abnormal returns defined with respect to *size/book-to-market matched reference stocks* in the RANDOM design with 1,000 random portfolios of $n = 200$ securities for annual (PANEL A), three-year (PANEL B), and five-year (PANEL C) event windows.

Abn ret (%)	BHAR-T	BHAR-TCA	CCAR-T	SHARPE-T
Panel A: Annual (12 months) abnormal returns				
-8	0.347	0.387	0.482	0.763
-6	0.219	0.261	0.298	0.528
-4	0.119	0.152	0.175	0.285
-2	0.067	0.095	0.092	0.117
± 0	0.046	0.070	0.060	0.056
$+2$	0.055	0.085	0.059	0.082
$+4$	0.112	0.139	0.106	0.210
$+6$	0.181	0.214	0.207	0.406
$+8$	0.299	0.334	0.351	0.645
Panel B: Three-year (36 months) abnormal returns				
-24	0.730	0.746	0.898	0.985
-12	0.281	0.328	0.359	0.581
-6	0.113	0.148	0.124	0.198
-3	0.069	0.103	0.081	0.096
± 0	0.054	0.089	0.054	0.053
$+3$	0.060	0.102	0.072	0.080
$+6$	0.103	0.141	0.125	0.197
$+12$	0.252	0.312	0.346	0.559
$+24$	0.698	0.712	0.884	0.985
Panel C: Five-year (60 months) abnormal returns				
-25	0.586	0.618	0.808	0.964
-15	0.280	0.334	0.421	0.621
-10	0.151	0.204	0.232	0.342
-5	0.075	0.122	0.085	0.137
± 0	0.043	0.084	0.046	0.045
$+5$	0.073	0.116	0.090	0.105
$+10$	0.143	0.199	0.202	0.279
$+15$	0.241	0.302	0.373	0.515
$+25$	0.576	0.611	0.791	0.923

BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from the sample period from July, 1973 through December 2009.

Table 7: Empirical rejection rates of the test statistics under the null hypothesis of no event effect in the RANDOM CLUSTER design at the nominal 1%, 5%, and 10% significance levels from 1,000 random portfolios of $n = 200$ securities from CRSP monthly data base sampled from the time period from July, 1997 to December, 2009.

Statistic	Benchmark	Lower 0.005	Upper 0.005	2-tail 0.01	Lower 0.025	Upper 0.025	2-tail 0.05	Lower 0.05	Upper 0.05	2-tail 0.10
Panel A: Annual (12 months) abnormal returns										
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	0.013^c	0.001^a	0.014	0.050^c	0.012^c	0.062^a	0.085^c	0.028^c	0.113
BHAR-T	Buy-and-hold equally weighted market portfolio	0.017^c	0.001^a	0.018^b	0.048^c	0.011^c	0.059	0.075^c	0.025^c	0.100
BHAR-T	Size/book-to-market matched reference stock	0.007	0.005	0.012	0.029	0.022	0.051	0.057	0.039	0.096
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	0.007	0.006	0.013	0.034^a	0.026	0.060	0.066^b	0.050	0.116^a
BHAR-TSA	Buy-and-hold equally weighted market portfolio	0.009^a	0.005	0.014	0.028	0.022	0.050	0.057	0.043	0.100
CCAR-T	Average log-returns of size/book-to-market stocks	0.004	0.005	0.009	0.021	0.028	0.049	0.047	0.056	0.103
CCAR-T	Market-wide average of log-returns	0.003	0.007	0.010	0.021	0.028	0.049	0.046	0.059	0.105
CCAR-T	Size/book-to-market matched reference stock (log-ret)	0.007	0.003	0.010	0.038^c	0.018	0.056	0.071^c	0.037^a	0.108
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	0.008	0.005	0.013	0.023	0.025	0.048	0.048	0.051	0.099
SHARPE-T	Market-wide average of Sharpe ratio	0.008	0.006	0.014	0.028	0.025	0.053	0.053	0.048	0.101
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	0.014^c	0.007	0.021^c	0.042^c	0.022	0.064^b	0.070^c	0.041	0.111
Panel B: Three-year (36 months) abnormal returns										
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	0.028^c	0.000^b	0.028^c	0.056^c	0.006^c	0.062^a	0.094^c	0.019^c	0.113
BHAR-T	Buy-and-hold equally weighted market portfolio	0.023^c	0.001^a	0.024^c	0.052^c	0.007^c	0.059	0.093^c	0.017^c	0.110
BHAR-T	Size/book-to-market matched reference stock	0.006	0.005	0.011	0.032	0.019	0.051	0.061	0.040	0.101
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	0.016^c	0.005	0.021^c	0.046^c	0.021	0.067^b	0.065^b	0.043	0.108
BHAR-TSA	Buy-and-hold equally weighted market portfolio	0.010^b	0.004	0.014	0.038^c	0.022	0.060	0.060	0.040	0.100
CCAR-T	Average log-returns of size/book-to-market stocks	0.003	0.001^a	0.004^a	0.022	0.014^b	0.036^b	0.043	0.047	0.090
CCAR-T	Market-wide average of log-returns	0.002	0.006	0.008	0.019	0.017	0.036^b	0.041	0.049	0.090
CCAR-T	Size/book-to-market matched reference stock (log-ret)	0.008	0.001^a	0.009	0.036^b	0.015^b	0.051	0.069^c	0.032^c	0.101
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	0.004	0.000^b	0.004^a	0.035^b	0.014^b	0.049	0.053	0.044	0.097
SHARPE-T	Market-wide average of Sharpe ratios	0.001^a	0.003	0.004^a	0.031	0.017	0.048	0.053	0.038^a	0.091
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	0.006	0.004	0.010	0.040^c	0.014^b	0.054	0.071^c	0.029^c	0.100

Continues

Table 7 Continues.

Statistic	Benchmark	Lower 0.005	Upper 0.005	2-tail 0.01	Lower 0.025	Upper 0.025	2-tail 0.05	Lower 0.05	Upper 0.05	2-tail 0.10
Panel C: Five-year (60 months) abnormal returns										
BHAR-T	Buy-and-hold size/book-to-market matched portfolio	0.022^c	0.000^b	0.022^c	0.060^c	0.002^c	0.062^a	0.101^c	0.017^c	0.118^a
BHAR-T	Buy-and-hold equally weighted market portfolio	0.028^c	0.000^b	0.028^c	0.068^c	0.003^c	0.071^c	0.113^c	0.015^c	0.128^c
BHAR-T	Size/book-to-market matched reference stock	0.002	0.003	0.005	0.028	0.019	0.047	0.055	0.047	0.102
BHAR-TSA	Buy-and-hold size/book-to-market matched portfolio	0.011^c	0.003	0.014	0.036^b	0.019	0.055	0.069^c	0.048	0.117^a
BHAR-TSA	Buy-and-hold equally weighted market portfolio	0.006	0.003	0.009	0.038^c	0.016^a	0.054	0.073^c	0.044	0.117^a
CCAR-T	Average log-returns of size/book-to-market stocks	0.004	0.003	0.007	0.021	0.018	0.039	0.044	0.057	0.101
CCAR-T	Market-wide average of log-returns	0.003	0.004	0.007	0.019	0.024	0.043	0.036^b	0.053	0.089
CCAR-T	Size/book-to-market matched reference stock (log-ret)	0.010^b	0.003	0.013	0.033	0.019	0.052	0.062^a	0.035^b	0.097
SHARPE-T	Size/book-to-market matched average Sharpe-ratio	0.003	0.003	0.006	0.019	0.021	0.040	0.045	0.044	0.089
SHARPE-T	Market-wide average of Sharpe ratios	0.005	0.005	0.010	0.021	0.023	0.044	0.046	0.049	0.095
SHARPE-T	Size/book-to-market reference stock Sharpe-ratio	0.006	0.001^a	0.007	0.034^a	0.013^b	0.047	0.056	0.025^c	0.081^b

BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from the sample period from July, 1973 through December 2009 using the RANDOM CLUSTER design described in Section 4.2.

$a = 10\%$, $b = 5\%$, and $c = 1\%$ significant deviation from the nominal rejection rate indicated by the column header.

Table 8: Empirical rejection rates of the null hypothesis of no event effect at the 5% nominal level in one-sided (cut off point -1.68 on the lower tail and $+1.68$ on the upped tail) and two-sided (cut off points ± 1.96) tests from non-random samples of sampling twice the same firm with partially overlapping event periods.

Panel A: Annual (12 months) event period

		Percent of stocks with overlapping event period								
		10%			50%			100%		
Stat	Benchm	Lower	Upper	2-tail	Lower	Upper	2-tail	Lower	Upper	2-tail
BHAR-T	Portf	0.096^c	0.028^c	0.066^b	0.113^c	0.041	0.086^c	0.118^c	0.063	0.118^c
BHAR-T	Market	0.093^c	0.025^c	0.067^b	0.107^c	0.042	0.090^c	0.120^c	0.056	0.122^c
BHAR-T	Stock	0.051	0.032^b	0.040	0.069^b	0.065^a	0.069^b	0.068^b	0.076^c	0.086^c
BHAR-TSA	Portf	0.077^c	0.053	0.062	0.092^c	0.074^c	0.096^c	0.101^c	0.088^c	0.125^c
BHAR-TSA	Market	0.073^c	0.061	0.062	0.086^c	0.070^b	0.107^c	0.104^c	0.088^c	0.127^c
CCAR-T	Portf	0.049	0.061	0.048	0.054	0.097^c	0.083^c	0.065^a	0.124^c	0.136^c
CCAR-T	Market	0.040	0.066^b	0.046	0.049	0.101^c	0.081^c	0.053	0.137^c	0.117^c
CCAR-T	Stock	0.058	0.044	0.047	0.085^c	0.064^a	0.087^c	0.076^c	0.084^c	0.091^c
CCAR-TCA	Portf	0.056	0.073^c	0.073^c	0.044	0.084^c	0.081^c	0.053	0.101^c	0.089^c
CCAR-TCA	Market	0.054	0.075^c	0.071^c	0.039	0.093^c	0.083^c	0.043	0.105^c	0.085^c
CCAR-TCA	Stock	0.061	0.051	0.077^c	0.082^c	0.060	0.097^c	0.064^a	0.062	0.077^c
SHARPE-T	Portf	0.050	0.054	0.054	0.060	0.072^c	0.072^c	0.087^c	0.115^c	0.118^c
SHARPE-T	Market	0.047	0.050	0.056	0.061	0.069^b	0.067^b	0.081^c	0.121^c	0.113^c
SHARPE-T	Stock	0.044	0.044	0.040	0.091^c	0.053	0.077^c	0.083^c	0.079^c	0.092^c
SHARPE-TCA	Portf	0.053	0.066^b	0.071^c	0.060	0.065^a	0.072^c	0.063	0.076^c	0.083^c
SHARPE-TCA	Market	0.055	0.061	0.076^c	0.048	0.064^a	0.071^c	0.048	0.081^c	0.079^c
SHARPE-TCA	Stock	0.061	0.054	0.068^b	0.095^c	0.048	0.084^c	0.077^c	0.056	0.075^c

Continues ...

Table 8 continues.

Panel B: Three year (36 months) event period

		Percent of stocks with overlapping event period								
		10%			50%			100%		
Stat	Benchm	Lower	Upper	2-tail	Lower	Upper	2-tail	Lower	Upper	2-tail
BHAR-T	Portf	0.124^c	0.024^c	0.086^c	0.117^c	0.032^b	0.086^c	0.117^c	0.055	0.107^c
BHAR-T	Market	0.126^c	0.022^c	0.095^c	0.112^c	0.031^c	0.089^c	0.133^c	0.062	0.108^c
BHAR-T	Stock	0.064^a	0.045	0.050	0.060	0.055	0.058	0.060	0.080^c	0.078^c
BHAR-TSA	Portf	0.096^c	0.057	0.084^c	0.089^c	0.079^c	0.085^c	0.096^c	0.097^c	0.120^c
BHAR-TSA	Market	0.093^c	0.047	0.084^c	0.084^c	0.058	0.080^c	0.098^c	0.114^c	0.124^c
CCAR-T	Portf	0.043	0.056	0.044	0.035^a	0.111^c	0.087^c	0.030^c	0.207^c	0.164^c
CCAR-T	Market	0.041	0.061	0.049	0.023^c	0.120^c	0.088^c	0.025^c	0.254^c	0.205^c
CCAR-T	Stock	0.063	0.035^a	0.050	0.069^b	0.064^a	0.070^b	0.041	0.119^c	0.094^c
CCAR-TCA	Portf	0.037	0.059	0.050	0.026^c	0.084^c	0.070^b	0.003^c	0.05	0.023^c
CCAR-TCA	Market	0.046	0.066^b	0.048	0.019^c	0.096^c	0.070^b	0.017^c	0.181^c	0.120^c
CCAR-TCA	Stock	0.066^b	0.044	0.058	0.060	0.054	0.055	0.031^c	0.081^c	0.061
SHARPE-T	Portf	0.052	0.055	0.051	0.050	0.074^c	0.071^c	0.067^b	0.158^c	0.134^c
SHARPE-T	Market	0.051	0.061	0.059	0.039	0.096^c	0.079^c	0.047	0.198^c	0.167^c
SHARPE-T	Stock	0.059	0.036^a	0.058	0.062	0.048	0.057	0.062	0.089^c	0.087^c
SHARPE-TCA	Portf	0.052	0.061	0.056	0.044	0.059	0.058	0.041	0.089^c	0.080^c
SHARPE-TCA	Market	0.049	0.061	0.059	0.032^b	0.073^c	0.063	0.029^c	0.131^c	0.093^c
SHARPE-TCA	Stock	0.057	0.045	0.064^a	0.058	0.047	0.051	0.044	0.067^b	0.062

Continues ...

Table 8 continues.

Panel C: Five year (60 months) event period

		Percent of stocks with overlapping event period								
		10%			50%			100%		
Stat	Benchm	Lower	Upper	2-tail	Lower	Upper	2-tail	Lower	Upper	2-tail
BHAR-T	Portf	0.104^c	0.024^c	0.074^c	0.101^c	0.03^c	0.081^c	0.109^c	0.055	0.098^c
BHAR-T	Market	0.113^c	0.017^c	0.081^c	0.091^c	0.030^c	0.077^c	0.102^c	0.064^a	0.096^c
BHAR-T	Stock	0.041	0.060	0.047	0.051	0.062	0.054	0.050	0.101^c	0.085^c
BHAR-TSA	Portf	0.076^c	0.062	0.082^c	0.078^c	0.067^b	0.083^c	0.091^c	0.107^c	0.122^c
BHAR-TSA	Market	0.081^c	0.058	0.083^c	0.074^c	0.076^c	0.080^c	0.080^c	0.114^c	0.122^c
CCAR-T	Portf	0.035^a	0.084^c	0.061	0.033^b	0.135^c	0.101^c	0.016^c	0.237^c	0.173^c
CCAR-T	Market	0.035^a	0.081^c	0.054	0.024^c	0.177^c	0.129^c	0.007^c	0.338^c	0.256^c
CCAR-T	Stock	0.053	0.050	0.044	0.047	0.079^c	0.073^c	0.030^c	0.143^c	0.111^c
CCAR-TCA	Portf	0.036^a	0.085^c	0.060	0.023^c	0.108^c	0.077^c	0.009^c	0.158^c	0.099^c
CCAR-TCA	Market	0.037	0.087^c	0.063	0.019^c	0.137^c	0.086^c	0.004^c	0.236^c	0.146^c
CCAR-TCA	Stock	0.053	0.051	0.043	0.037	0.067^b	0.058	0.020^c	0.108^c	0.077^c
SHARPE-T	Portf	0.037	0.067^b	0.049	0.046	0.12^c	0.097^c	0.026^c	0.187^c	0.146^c
SHARPE-T	Market	0.029^c	0.068^b	0.042	0.030^c	0.155^c	0.115^c	0.014^c	0.284^c	0.222^c
SHARPE-T	Stock	0.058	0.052	0.048	0.055	0.077^c	0.070^b	0.045	0.126^c	0.090^c
SHARPE-TCA	Portf	0.034^b	0.061	0.056	0.030^c	0.095^c	0.069^b	0.013^c	0.126^c	0.081^c
SHARPE-TCA	Market	0.036^a	0.072^c	0.057	0.025^c	0.116^c	0.078^c	0.006^c	0.209^c	0.133^c
SHARPE-TCA	Stock	0.056	0.058	0.054	0.043	0.061	0.059	0.030^c	0.086^c	0.062

Significances: $a = 0.10$, $b = 0.05$, $c = 0.01$ by the binomial test for two-tailed testing (i.e., whether the test procedure under rejects or over rejects the null hypothesis of no event effect).

Table 9: Rejection rates of alternative tests in size-based samples

		Event Horizon								
		1 Year			3 Years			5 Years		
Stat	Benchm	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%
Panel A: Large firms										
BHAR-T	Portf	0.059	0.046	0.051	0.068^b	0.042	0.051	0.088^c	0.034^b	0.058
BHAR-T	Stock	0.052	0.052	0.056	0.040	0.061	0.051	0.031^c	0.067^b	0.053
BHAR-TSA	Portf	0.052	0.056	0.062	0.058	0.057	0.066^b	0.071^c	0.057	0.065^a
CCAR-T	Portf	0.041	0.066^b	0.058	0.037	0.062	0.053	0.055	0.059	0.052
CCAR-T	Stock	0.052	0.061	0.056	0.028^c	0.067^b	0.053	0.036^a	0.070^b	0.057
SHARPE-T	Portf	0.049	0.065^a	0.053	0.044	0.060	0.058	0.056	0.054	0.057
SHARPE-T	Stock	0.059	0.050	0.063	0.042	0.065^a	0.064^a	0.042	0.068^b	0.056
Panel B: Small firms										
BHAR-T	Portf	0.074^c	0.026^c	0.052	0.046	0.045	0.050	0.115^c	0.013^c	0.079^c
BHAR-T	Stock	0.036^a	0.060	0.054	0.025^c	0.105^c	0.057	0.047	0.053	0.044
BHAR-TSA	Portf	0.058	0.067^b	0.067^b	0.037	0.111^c	0.077^c	0.080^c	0.042	0.066^b
CCAR-T	Portf	0.037	0.061	0.044	0.014^c	0.131^c	0.080^c	0.047	0.046	0.044
CCAR-T	Stock	0.046	0.050	0.054	0.029^c	0.095^c	0.061	0.064^a	0.040	0.046
SHARPE-T	Portf	0.033^b	0.050	0.042	0.025^c	0.093^c	0.068^b	0.049	0.039	0.040
SHARPE-T	Stock	0.037	0.043	0.045	0.041	0.072^c	0.062	0.051	0.043	0.051

Significances: $a = 0.10$, $b = 0.05$, $c = 0.01$ by the binomial test.

Table 10: Rejection rates of alternative tests in high book-to-market and low-book-to-market decile samples samples

		Event Horizon								
		1 Year			3 Years			5 Years		
Stat	Benchm	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%
Panel A: High book-to-market samples										
BHAR-T	Portf	0.076^c	0.030^c	0.061	0.108^c	0.022^c	0.082^c	0.161^c	0.005^c	0.116^c
BHAR-T	Stock	0.033^b	0.049	0.041	0.043	0.047	0.045	0.050	0.042	0.043
BHAR-TSA	Portf	0.059	0.067^b	0.067^b	0.082^c	0.061	0.077^c	0.129^c	0.025^c	0.088^c
CCAR-T	Portf	0.050	0.053	0.049	0.035^a	0.064^a	0.049	0.124^c	0.013^c	0.081^c
CCAR-T	Stock	0.051	0.044	0.044	0.037	0.058	0.054	0.057	0.043	0.053
SHARPE-T	Portf	0.055	0.044	0.051	0.054	0.054	0.048	0.128^c	0.022^c	0.082^c
SHARPE-T	Stock	0.037	0.036^a	0.037	0.047	0.051	0.057	0.065^a	0.044	0.056
Panel B: Low book-to-market samples										
BHAR-T	Portf	0.091^c	0.028^c	0.057	0.150^c	0.016^c	0.104^c	0.161^c	0.005^c	0.116^c
BHAR-T	Stock	0.052	0.048	0.038	0.061	0.041	0.045	0.050	0.042	0.043
BHAR-TSA	Portf	0.073^c	0.055	0.066^b	0.107^c	0.042	0.079^c	0.129^c	0.025^c	0.088^c
CCAR-T	Portf	0.069^b	0.036^a	0.049	0.113^c	0.023^c	0.077^c	0.124^c	0.013^c	0.081^c
CCAR-T	Stock	0.068^b	0.044	0.053	0.077^c	0.037	0.050	0.057	0.043	0.053
SHARPE-T	Portf	0.073^c	0.041	0.064^a	0.122^c	0.018^c	0.077^c	0.128^c	0.022^c	0.082^c
SHARPE-T	Stock	0.062	0.038	0.042	0.071^c	0.032^b	0.058	0.065^a	0.044	0.056

Significances: $a = 0.10$, $b = 0.05$, $c = 0.01$ by the binomial test.

Table 11: Rejection rates of alternative tests in highest and lowest pre-event return performance samples

		Event Horizon								
		1 Year			3 Years			5 Years		
Stat	Benchm	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%
Panel A: High pre-event returns										
BHAR-T	Portf	0.000^c	0.876^c	0.767^c	0.010^c	0.179^c	0.098^c	0.049	0.064^a	0.054
BHAR-T	Pre-event	0.090^c	0.023^c	0.063	0.096^c	0.018^c	0.070^b	0.146^c	0.012^c	0.106^c
BHAR-T	Stock	0.061	0.047	0.056	0.051	0.049	0.043	0.053	0.042	0.040
BHAR-TSA	Portf	0.000^c	0.920^c	0.853^c	0.009^c	0.287^c	0.198^c	0.035^a	0.126^c	0.088^c
BHAR-TSA	Pre-event	0.072^c	0.042	0.055	0.074^c	0.050	0.055	0.099^c	0.041	0.086^c
CCAR-T	Portf	0.001^c	0.798^c	0.687^c	0.030^c	0.080^c	0.055	0.124^c	0.018^c	0.073^c
CCAR-T	Pre-event	0.050	0.053	0.041	0.036^a	0.058	0.047	0.056	0.045	0.054
CCAR-T	Stock	0.059	0.047	0.058	0.055	0.048	0.048	0.055	0.045	0.047
SHARPE-T	Portf	0.000^c	0.775^c	0.664^c	0.053	0.050	0.056	0.231^c	0.006^c	0.152^c
SHARPE-T	Pre-event	0.054	0.051	0.058	0.048	0.059	0.059	0.056	0.044	0.055
SHARPE-T	Stock	0.054	0.049	0.054	0.044	0.055	0.054	0.059	0.046	0.042
Panel B: Low pre-event returns										
BHAR-T	Portf	0.599^c	0.000^c	0.517^c	0.223^c	0.001^c	0.173^c	0.125^c	0.013^c	0.090^c
BHAR-T	Pre-event	0.120^c	0.018^c	0.090^c	0.133^c	0.006^c	0.103^c	0.121^c	0.013^c	0.097^c
BHAR-T	Stock	0.050	0.046	0.045	0.051	0.047	0.043	0.044	0.050	0.043
BHAR-TSA	Portf	0.486^c	0.000^c	0.390^c	0.175^c	0.011^c	0.123^c	0.082^c	0.036^a	0.080^c
BHAR-TSA	Pre-event	0.101^c	0.045	0.083^c	0.101^c	0.045	0.070^b	0.087^c	0.037	0.072^c
CCAR-T	Portf	0.977^c	0.000^c	0.961^c	0.713^c	0.000^c	0.601^c	0.539^c	0.001^c	0.400^c
CCAR-T	Pre-event	0.052	0.052	0.059	0.042	0.045	0.045	0.041	0.058	0.056
CCAR-T	Stock	0.049	0.056	0.053	0.064^a	0.052	0.048	0.061	0.054	0.056
SHARPE-T	Portf	0.990^c	0.000^c	0.977^c	0.897^c	0.000^c	0.833^c	0.822^c	0.000^c	0.715^c
SHARPE-T	Pre-event	0.048	0.055	0.050	0.047	0.052	0.046	0.046	0.053	0.052
SHARPE-T	Stock	0.057	0.050	0.057	0.058	0.048	0.058	0.049	0.050	0.054

Significances: $a = 0.10$, $b = 0.05$, $c = 0.01$ by the binomial test.

Table 12: Rejection rates of abnormal returns of the Fama-French three factor model of portfolios of return differences between test assets and size/book-to-market matched control stocks.

	Event Horizon								
	1 Year			3 Years			5 Years		
Sample design	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%	Lower 5%	Upper 5%	2-tail 5%
Panel A: Equally weighted calendar time portfolios									
Random samples	0.053	0.045	0.042	0.040	0.043	0.040	0.026^c	0.035^a	0.026^c
Clutered event months	0.069^b	0.053	0.076^c	0.041	0.031^c	0.036^a	0.043	0.031^c	0.038
Overlapping returns	0.045	0.051	0.041	0.040	0.061	0.053	0.044	0.052	0.042
Large Firms	0.051	0.060	0.060	0.039	0.076^c	0.054	0.047	0.083^c	0.057
Small firms	0.039	0.036^a	0.038	0.014^c	0.054	0.037	0.027^c	0.029^c	0.023^c
High book-to-market ratio	0.027^c	0.041	0.027^c	0.030^c	0.031^c	0.027^c	0.021^c	0.030^c	0.020^c
Low book-to-market ratio	0.040	0.045	0.034^b	0.046	0.021^c	0.025^c	0.039	0.030^c	0.038
High pre-event returns	0.052	0.043	0.050	0.037	0.032^b	0.036^a	0.025^c	0.035^a	0.029^c
Low pre-event returns	0.032^b	0.028^c	0.028^c	0.021^c	0.026^c	0.018^c	0.020^c	0.022^c	0.019^c
Panel B: Value weighted calendar time portfolios									
Random samples	0.062	0.038	0.055	0.070^b	0.042	0.057	0.074^c	0.042	0.059
Clutered event months	0.091^c	0.037	0.076^c	0.088^c	0.047	0.078^c	0.084^c	0.041	0.069^b
Overlapping returns	0.060	0.043	0.060	0.061	0.058	0.056	0.059	0.041	0.056
Large Firms	0.064^a	0.045	0.056	0.053	0.056	0.062	0.070^b	0.071^c	0.077^c
Small firms	0.101^c	0.014^c	0.062	0.107^c	0.014^c	0.070^b	0.168^c	0.003^c	0.104^c
High book-to-market ratio	0.082^c	0.030^c	0.063	0.081^c	0.038	0.056	0.077^c	0.028^c	0.055
Low book-to-market ratio	0.082^c	0.032^b	0.061	0.069^b	0.049	0.070^b	0.051	0.062	0.054
High pre-event returns	0.096^c	0.024^c	0.060	0.088^c	0.028^c	0.066^b	0.083^c	0.022^c	0.057
Low pre-event returns	0.200^c	0.002^c	0.113^c	0.249^c	0.001^c	0.149^c	0.247^c	0.002^c	0.152^c

Significances: $a = 0.10$, $b = 0.05$, $c = 0.01$ by the two-tailed binomial test.

continues ...

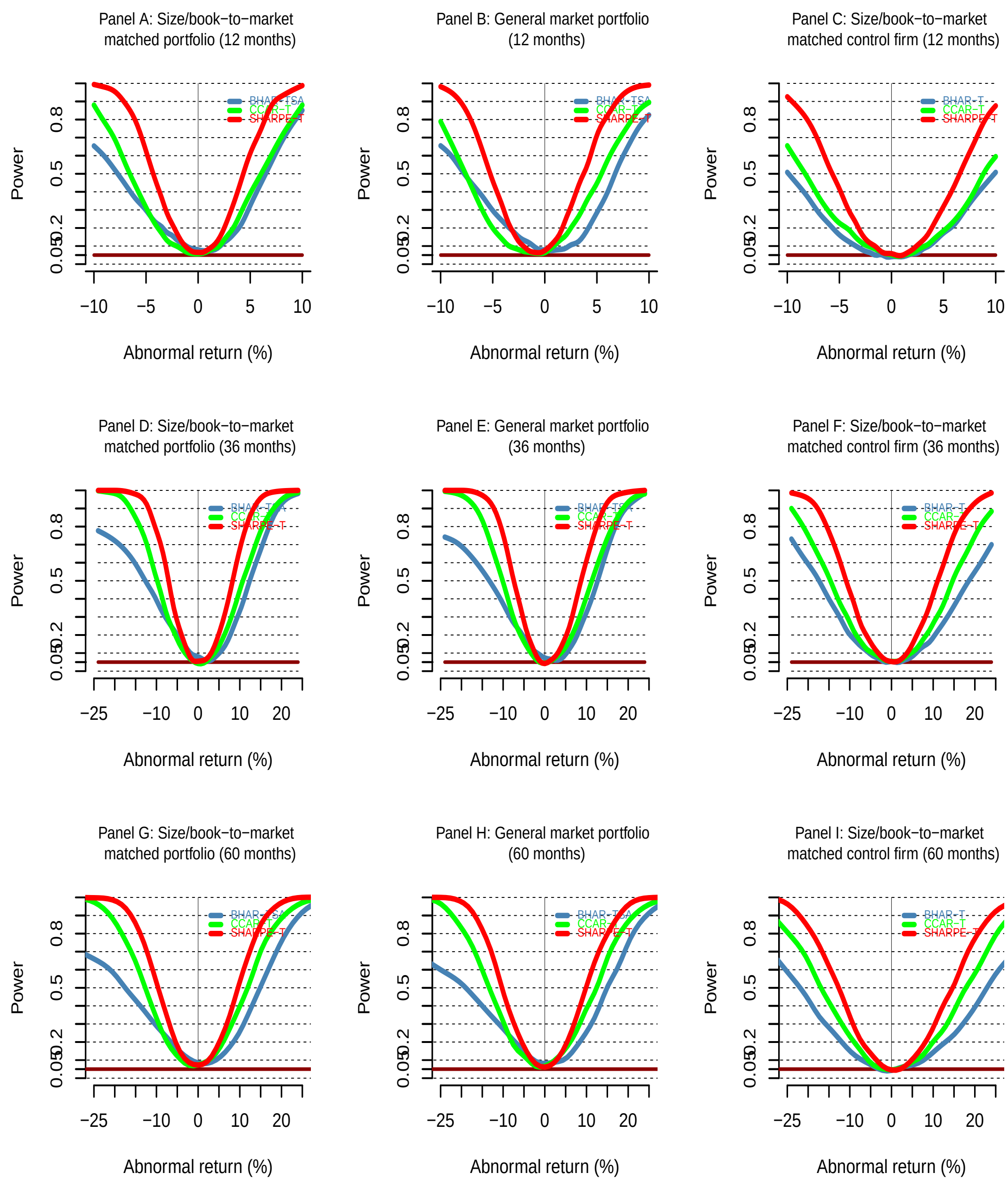
The results of the table are based on $1,000$ random samples of 200 months in each of which an event stock is randomly assigned. The random samples are separately constructed for 12, 36, and 60 months event periods. The abnormal return is the intercept of the regression

$$(R_{t,\mathtt{test}} - R_{t,\mathtt{control}})_{p,t} = \alpha_p + \beta_p(R_{m,t} - R_{f,t}) + s_p\mathtt{SMB}_t + h_p\mathtt{HML}_t + r_{p,t},$$

where $R_{t,\mathtt{test}} - R_{t,\mathtt{control}})_{p,t}$ is the portfolio (equally weighted or value weighted) of the return differences of the simple returns $R_{i,t,\mathtt{test}}$ of the test assets and simple returns $R_{i,t,\mathtt{control}}$ of the size/book-to-market matched control firms, $R_{m,t}$ is the market value weighted index return, R_{ft} is the return of the three month Treasury bills, $\mathtt{SMB}_t$ is the Fama-French small-minus-big portfolio return, and $\mathtt{HML}_t$ is the Fama-French high-minus-low book-to-market value factor return. The intercept parameter α_p is the abnormal return parameter. The numbers in the table indicate the percentage of t-values of the estimated α-coefficients in the $1,000$ that exceed the theoretical 5% cut off points in lower tail (-1.64), upper tail ($+1.64$), and two-tail (± 1.96) tests. The first sample design is non-clustered random sample described in Section 5.1.1, the second design is clustered random sample described in Section 5.1.2, and the rest of the designs are non-random resigns discussed in Sections 5.2.1 through 5.2.4.

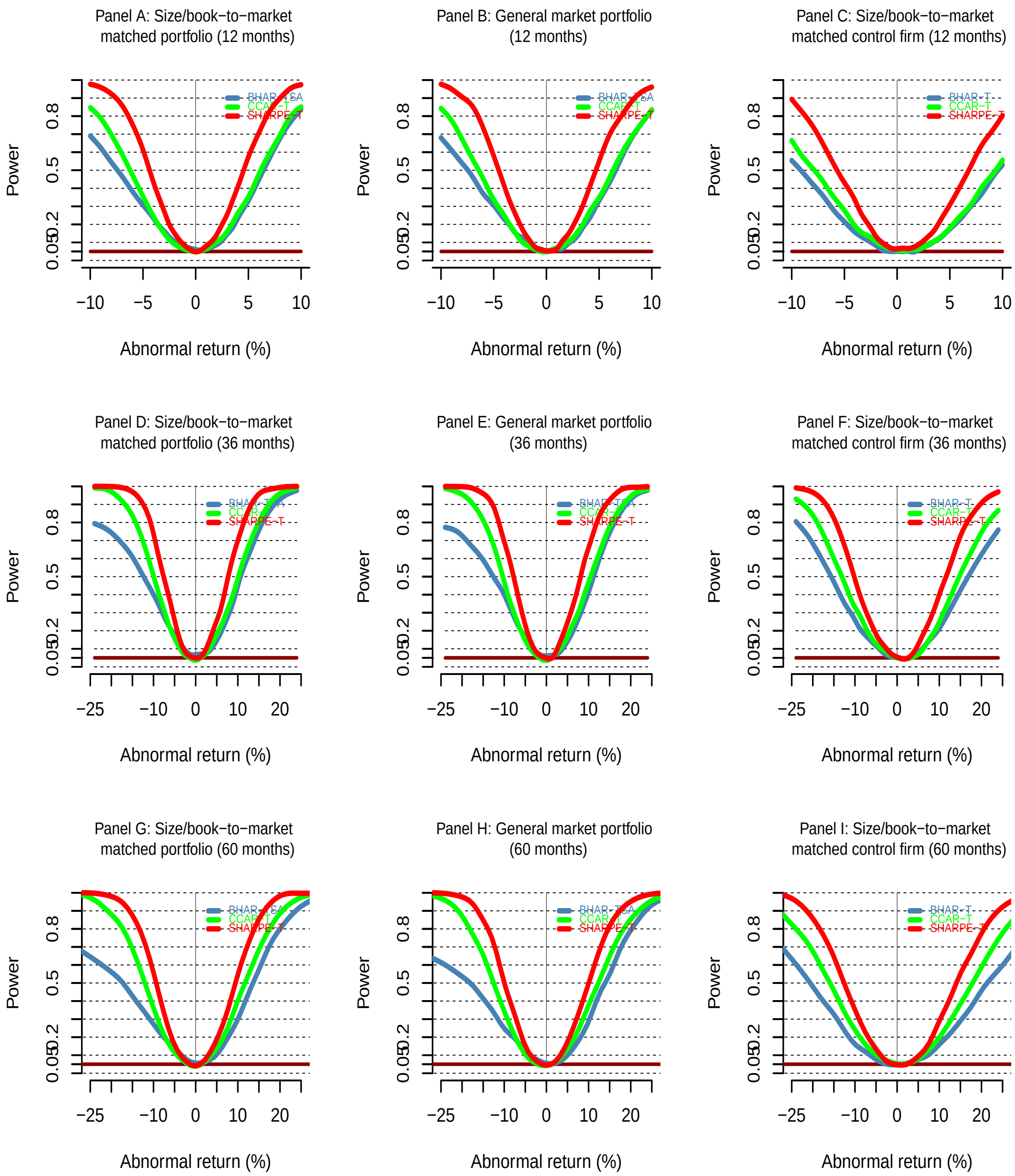
rest of the designs are non-r

Figure 1: Empirical power functions at the 5% level in two-sided testing of abnormal returns defined with respect to *size/book-to-market matched reference portfolios*, *general market portfolio*, and *size/book-to-market matched reference firm* in the RANDOM design with 1,000 random portfolios of $n = 200$ stocks for annual (PANELS A, B, C), three-year (PANELS D, E, F), and five-year (PANELS G, H, I) event windows.



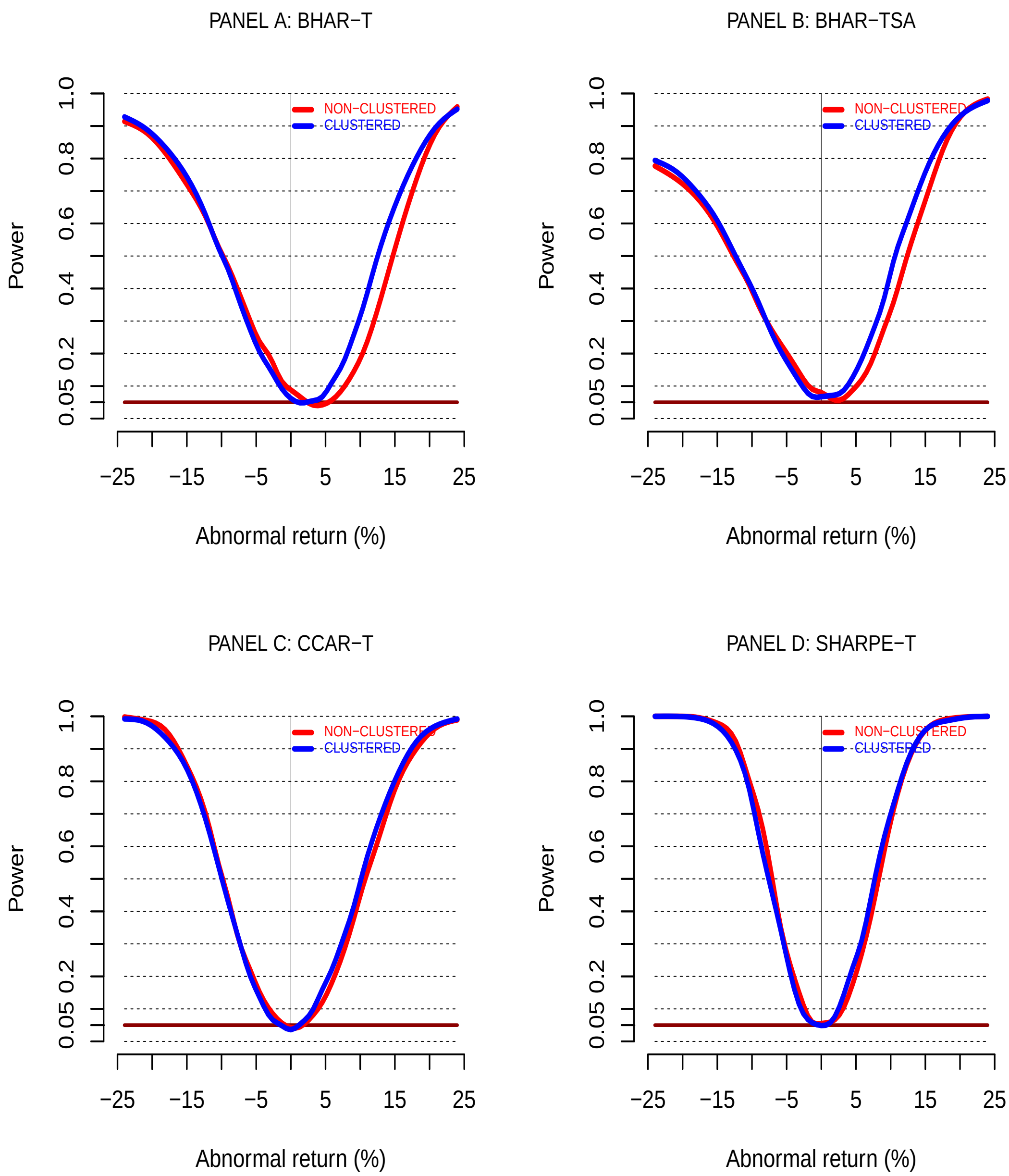
BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from July, 1973 through December 2009.

Figure 2: Empirical power functions at the 5% level two-sided testing of abnormal returns defined with respect to the *size/book-to-market matched reference portfolios general market portfolio*, and *size/book-to-market matched reference firm* in the RANDOM CLUSTER design with 1,000 random portfolios of $n = 200$ stocks for annual (PANELS A, B, C), three-year (PANELS D, E, F), and five-year (PANELS G, H, I) event windows.



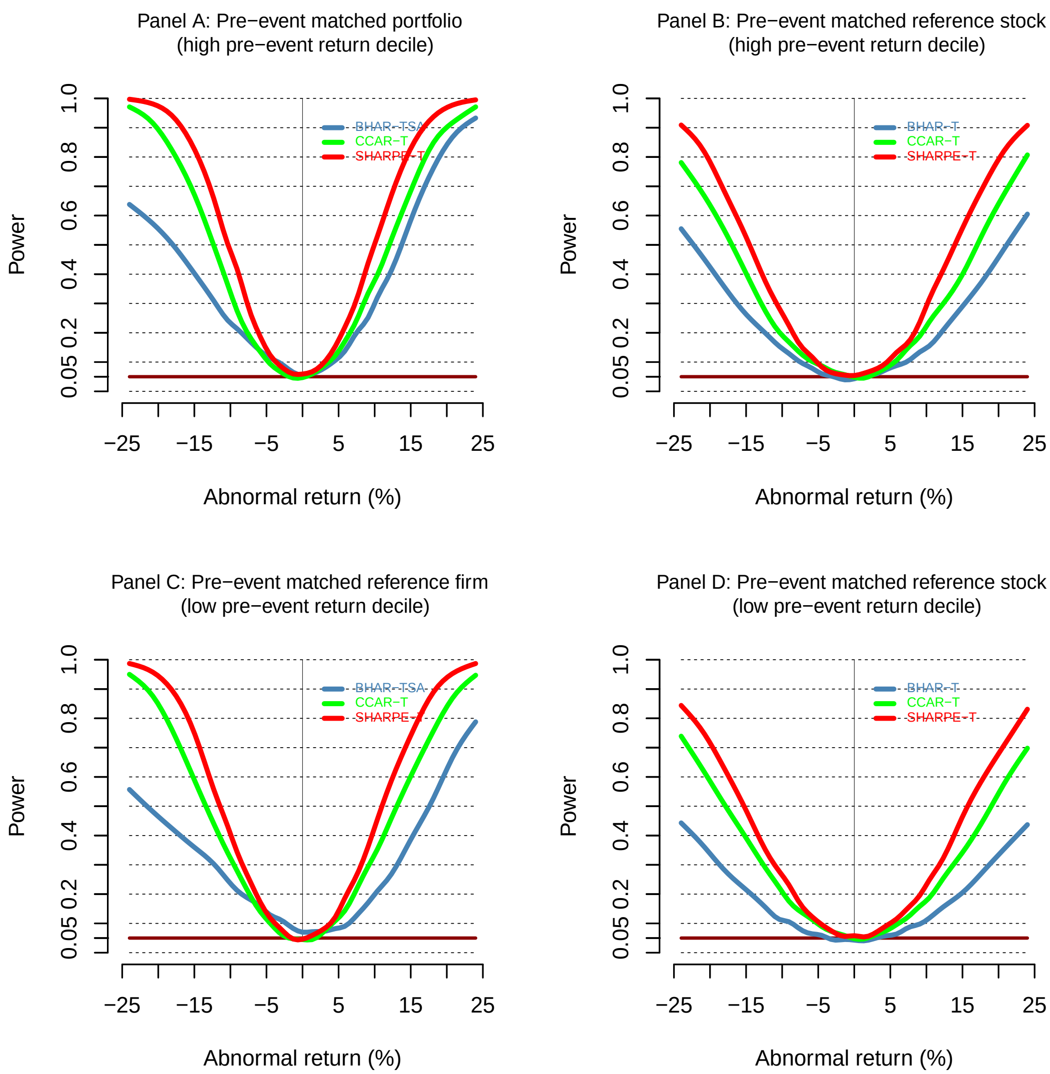
BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from July, 1973 through December 2009.

Figure 3: Empirical power functions of BHAR-T, BHAR-TSA, CCAR-T, and SHARPE-T for the RANDOM design (red line) and RANDOM CLUSTER design (blue line) at the 5% level in two-sided testing of abnormal returns defined with respect to the *size/book-to-market matched portfolio* with $1,000$ random portfolios of $n = 200$ securities for three-year (36 months) event window.



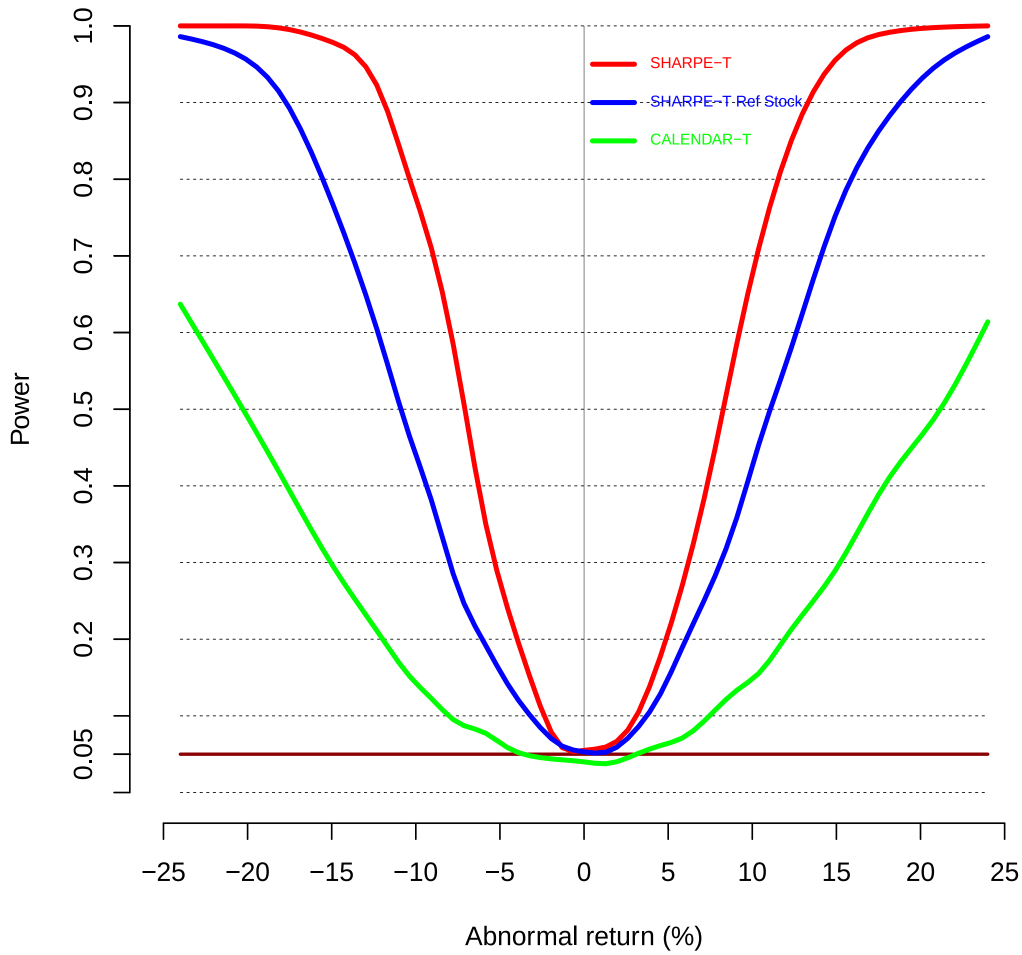
BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from July, 1973 through December 2009.

Figure 4: Empirical power functions of BHAR-T, BHAR-TSA, CCAR-T, and SHARPE-T statistics in samples from high and low pre-event return stocks at the 5% level in two-sided testing of abnormal returns defined with respect to the *decile matched portfolios* (Panels A and C) and with respect to *decile and book-to-market matched firm* (Panels B and C). The plot a based on 1,000 random portfolios of $n = 200$ securities for three-year (36 months) event window.



BHAR-T is Buy-and-hold abnormal return t-statistic defined in equation (8), BHAR-TSA is the skewness adjusted BHAR-T statistic (Lyon, Barber, and Tsay [26]) defined in equation (9), CCAR-T is the continuously compounded abnormal return t-statistic defined in equation (23), and SHARPE-T is the Sharpe ratio based test statistic defined in equation (30). The test statistics and their key properties are summarized in Table 1. Data in the simulations utilizes actual monthly returns from CRSP data base from July, 1973 through December 2009.

Figure 5: Empirical power functions of SHARPE-T (reference portfolio), SHARPE-T (reference stock), and CALENDAR-T statistics in random samples based on 1,000 random portfolios of $n = 200$ securities for three-years (36 months) event window.



SHARPE-T is the Sharpe ratio based test statistic defined in equation (30) of the references portfolio abnormal return, `SHARP-T Ref Stock` is the `SHARPE-T` statistic of testing the size/book-to-market reference stock based abnormal return [Section 4.1 (c)]. The `CALENDAR-T` statistic is defined in equation (61) and the SHARPE-T statistic and its key properties are discussed in Section 3.4.2. Data in the simulations utilizes actual monthly returns from CRSP data base from July, 1973 through December 2009.