Generalized Rank Test for Testing Cumulative Abnormal Returns in Event Studies

Professor James Kolari, Chase Professor of Finance, Texas A&M University, TAMU - 4218, Finance Dept., College Station, TX 77843-4218, Email: jkolari@mays.tamu.edu, Office phone: 979-845-4803, Fax: 979-845-3884

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

Abstract. Corrado's (1989) rank test and its modification in Corrado and Zivney (1992) that accounts for possible volatility changes due to the event effect appear to have good (empirical) power properties against the parametric tests of Patell (1976) and Boehmer, Musumeci and Poulsen (BMP) (1991). However, the Corrado-Zivney test is derived for an one-day event window. The ranks of abnormal returns are dependent by construction, which introduces incremental bias in the standard error in the denominator of the simple CAR t-statistic of ranks as the accumulation period grows. This paper proposes a generalized rank test that can be used both for testing cumulative abnormal returns as well as one-day abnormal returns. Empirical properties of the test statistics are studied with simulations using CRSP returns. The results show that the some popular test statistics, including the ordinary t-test and adjusted Corrado-Zivney test with cumulated ranks tend to under-reject the null hypothesis as the CAR period increases. In addition, the power of the cumulated ranks Corrado-Zivney test seems to suffer when the abnormal return is randomly assigned to a single day within the event window. The proposed generalized rank test is robust against these problems. Furthermore, it is robust to abnormal return serial correlation, event-induced volatility, and cross-correlation due to event day clustering, with competitive (empirical) power relative to the standard parametric tests of Patell and BMP.

JEL Classification: G14; C10; C15
EFM Classification: 350; 760

[1]Presenting author

Generalized Rank Test for Testing Cumulative Abnormal Returns in Event Studies

Abstract. Corrado's (1989) rank test and its modification in Corrado and Zivney (1992) that accounts for possible volatility changes due to the event effect appear to have good (empirical) power properties against the parametric tests of Patell (1976) and Boehmer, Musumeci and Poulsen (BMP) (1991). However, the Corrado-Zivney test is derived for an one-day event window. The ranks of abnormal returns are dependent by construction, which introduces incremental bias in the standard error in the denominator of the simple CAR t-statistic of ranks as the accumulation period grows. This paper proposes a generalized rank test that can be used both for testing cumulative abnormal returns as well as one-day abnormal returns. Empirical properties of the test statistics are studied with simulations using CRSP returns. The results show that the some popular test statistics, including the ordinary t-test and adjusted Corrado-Zivney test with cumulated ranks tend to under-reject the null hypothesis as the CAR period increases. In addition, the power of the cumulated ranks Corrado-Zivney test seems to suffer when the abnormal return is randomly assigned to a single day within the event window. The proposed generalized rank test is robust against these problems. Furthermore, it is robust to abnormal return serial correlation, event-induced volatility, and cross-correlation due to event day clustering, with competitive (empirical) power relative to the standard parametric tests of Patell and BMP.

JEL Classification: G14; C10; C15
EFM Classification: 350; 760

1. Introduction

Standardized event study tests by Patell (1976) and Boehmer, Musumeci, and Poulsen (BMP) (1991) have gained popularity over conventional non-standardized tests in testing abnormal security price performance due to their better power properties. Harrington and Shrider (2007) find that in short-horizon tests on mean abnormal returns one should always use tests that are robust against cross-sectional variation in the *true* abnormal return [see Harrington and Shrider (2007)]. They find that that the BMP t-statistic is a good candidate for a robust, parametric test in conventional event studies.[2] Corrado (1989) [and Corrado and Zivney (1992)] introduced a nonparametric rank test based on standardized returns, which has proven to have very competitive, often superior, (empirical) power properties over parametric tests when testing for one-day abnormal returns [e.g. Corrado (1989), Corrado and Zivney (1992), Cowan (1992), Campbell and Wasley (1993), Kolari and Pynnonen (2008)]. Furthermore, the rank test of Corrado

[2] We define conventional event studies as those focusing only on mean stock price effects. Other types of event studies include (for example) the examination of return variance effects [Beaver (1968) and Patell (1976)], trading volume [Beaver (1968) and Campbell and Wasley (1996)], accounting performance [Barber and Lyon (1996)], and earnings management procedures [Dechow, Sloan, and Sweeney (1995) and Kothari, Leone, and Wasley (2005)].

and Zivney (1992) based on event day re-standardized returns has proven to be both robust against event-induced volatility [Campbell and Wasley (1992)] and cross-correlation due to event day clustering [Kolari and Pynnonen (2008)].

The Patell and BMP parametric tests can be rapidly applied to testing cumulative abnormal returns (CARs) over multiple day windows.[3] However, application of the rank test for testing CARs is not straightforward due to the fact that the ranks of the (standardized) abnormal returns over different days are dependent by construction. Thus, as shown in Luoma and Pynnonen (2008), although over short windows the dependence might be negligible, in longer windows the dependence accumulates and will bias downwards the rejection rates of the simple CAR rank t-test. Luoma and Pynnonen (2008) derived exact standard errors of the cumulative ranks and proposed easy-to-apply corrections to the existing cumulative rank t-statistic. The current paper contributes to existing literature by suggesting a generalized rank test procedure which can be used for testing both CARs as well as one-day abnormal returns. The Corrado and Zivney (1991) test, which is developed for testing one-day abnormal returns, is a special case of the procedure proposed

[3]With the correction suggested in Kolari and Pynnonen (2008), these tests are useful also in the case of clustered event days.

in this paper. The proposed procedure is shown to have several advantages over existing tests. First, it is robust to the event-induced volatility. Second, it is robust to cross-correlation due to event day clustering. Third, it proves to have competitive, often superior (empirical) power properties compared to popular parametric tests. Fourth it avoids the under-rejection symptom of the Corrado-Zivney rank test and some parametric tests as the CAR period increases. Fifth, and last, it is fairly robust to autocorrelation of abnormal returns.[4] Robustness with respect to event-induced volatility and power properties of the generalized rank test are demonstrated with simulation studies based on real returns on CRSP stock returns.

The rest of the paper is organized as follows. Section 2 introduces the generalized rank test. Section 3 describes the simulation design and summarizes the most popular test statistics used in event studies against which the generalized rank test is compared with. The results are presented in Section 4. Section 5 concludes.

[4]As will be shown later, the robustness stems from the re-standardization of the abnormal returns, which implies that the BMP-procedure should share the autocorrelation robustness as well. This will be documented formally elsewhere.

2. Generalized rank test

Corrado (1989) and Corrado and Zivney (1992) introduce the rank test for testing an one-day event abnormal return. Cowan (1992) and Campbell and Wasley (1993) use Corrado's rank test for testing cumulative abnormal returns by accumulating the respective ranks. However, this approach has some potential shortcomings. An obvious one is the case where the cumulative abnormal return is clustered on a random single day within the event window as described in the simulation study by Brown and Warner (1985, Section 4.3.2). This situation occurs in practice when the event day is not exactly known, in which case returns are cumulated over an interval to cover the actual event day. In these circumstances cumulative abnormal return tests based on ranks may face problems in detection of the abnormal behavior, especially for longer event windows. The reason is simply that, when all the returns are transformed to rank numbers, they do not account for the magnitudes of returns exceptvia the relative rank. Thus, if one large return is randomly assigned to one day within the event window independently for each stock, there is only one potentially outstanding rank for each stock that is randomly scattered across the window. This is likely to average largely out in the cumulative rank sum, and consequently, result in poor power properties of the test. This problem has been addressed in Cowan (1992), which our simulations confirm.It is notable, that parametric tests, such as the Patell and BMP tests, do not suffer from this shortcoming, as they are based on the accumulated returns over the event window, or sums, such that it does not matter how large abnormal returns are scattered across the window.

In order to develop a non-parametric test that can be used both for testing one-day and cumulative abnormal returns, we next introduce some necessary notations and concepts[see Campbell, Lo, and MacKinlay (1997, Chapter 4) for an excellent discussion of event study methodology]. In forthcoming theoretical derivations we make the following explicit assumption:

Assumption 1 *Stock returns r_{it} are weak white noise continuous random variables with*

$$\begin{aligned} E[r_{it}] &= \mu_i \textit{ for all } t \\ \mathrm{var}[r_{it}] &= \sigma_i^2 \textit{ for all } t \\ \mathrm{cov}[r_{it}, r_{iu}] &= 0 \textit{ for all } t \neq u. \end{aligned} \tag{1}$$

$i = 1, \ldots, n$, $t = 1, \ldots, T$.

Let day $t = 0$ indicate the event day, days $T_0 + 1, T_0 + 2 \ldots, T_1$ are the estimation period days relative to the event day, $T_1{+}1, T_1{+}2, \ldots, T_2$ are event window days, again relative to the event day $(= 0)$. We define $L_1 = T_1 - T_0$ as the estimation period length, $L_2 = T_2 - T_1$ as the event period length, and the combined estimation period and event period length as $T = L_1 + L_2$. Given that AR_{it} denotes the excess return of security i on day t after extracting the market factors used in the abnormal return model in the event study, standardized abnormal returns are defined as

$$\mathrm{SAR}_{it} = \frac{\mathrm{AR}_{it}}{S_{\mathrm{AR}_i}}, \tag{2}$$

where S_{AR_i} is the standard deviation of the residuals from the factor model adjusted for the forecast error. The cumulative abnormal return of security i over τ event days is defined as

$$\mathrm{CAR}_{i\tau} = \sum_{t=t_1+1}^{t_1+\tau} \mathrm{AR}_{it}, \tag{3}$$

with $T_1 \leq t_1 \leq T_2 - \tau$, $1 \leq \tau \leq L_2$. The corresponding standardized cumulative abnormal return (SCAR) is defined as

$$\mathrm{SCAR}_{i\tau} = \frac{\mathrm{CAR}_{i\tau}}{S_{\mathrm{CAR}_{i\tau}}}, \tag{4}$$

where $S_{\mathrm{CAR}_{i\tau}}$ is the standard deviation of the cumulative abnormal returns adjusted for forecast error [see Cambell, Lo, and MacKinlay (1977), Section 4.4.3].

Under the null hypothesis of no event effect both SAR_{it} and $\mathrm{SCAR}_{i\tau}$ are distributed with mean zero and (approximately) unit variance. We can utilize this fact in defining our generalized rank test.

In order to account for the possible event-induced volatility Boehmer, Mucumeci, and Poulsen (1991) re-standardize the SCARs with the cross-sectional standard deviation to get re-standardized SCAR

$$\mathrm{SCAR}^*_{i\tau} = \frac{\mathrm{SCAR}_{i\tau}}{S_{\mathrm{SCAR}}} \tag{5}$$

where

$$S_{\mathrm{SCAR}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\mathrm{SCAR}_{i\tau} - \overline{\mathrm{SCAR}}_{\tau}\right)^2}. \tag{6}$$

is the cross-sectional standard deviation of $\mathrm{SCAR}_{i\tau}$s[5] and

$$\overline{\mathrm{SCAR}}_{\tau} = \frac{1}{n}\sum_{i=1}^{n}\mathrm{SCAR}_{i\tau}. \tag{7}$$

Again, $\mathrm{SCAR}^*_{i\tau}$ is a zero mean and unit variance random variable like other abnormal returns. Thus, we use $\mathrm{SCAR}^*_{i\tau}$ as an abnormal return and define Generalized Standardized Abnormal Returns (GSAR) as follows:

Definition 1 *The generalized standardized abnormal returns (GSAR) is defined as*

$$\mathrm{GSAR}_{it} = \begin{cases} \mathrm{SCAR}^*_{i\tau}, & \textit{for } t_1 + 1 \le t \le t_1 + \tau, \\ \mathrm{SAR}_{it}, & \textit{for } t = T_0 + 1, \ldots, t_1, t_1 + \tau + 1, \ldots, T_2 \end{cases} \tag{8}$$

where $\mathrm{SCAR}^*_{i\tau}$ *is defined in equation* (5) *and* SAR_{it} *defined in equation* (2).

[5]If the event days are clustered, equation (6) should be further adjusted for cross-sectional correlation as Kolari and Pynnonen (2008)

That is, the cumulated period is considered as one time point in which the generalized standardized abnormal return, GSAR, equals the re-standardized cumulative abnormal return defined in equation (5) and for other time points GSAR equals the usual standardized abnormal returns defined in equation (2).

In the next step we redefine the time indexing such that the cumulative abnormal return period (CAR-period of length τ) as a whole is squeezed into one observation with time index $t = 0$, called the *cumulative event day*. In this new indexing approach the day index for the first observation before the cumulative period becomes $t = -1$ and the day after (if any) becomes $t = +1$, and so forth. However, in order to simplify notations we continue to use $T_0 + 1$ and T_2, respectively, as generic symbols for the first and last observations relative to the cumulative event day $t = 0$. The total number of the combined event and estimation period observations reduces in the ranking to $T' = T - \tau + 1$, which we call the adjusted number of observations.

With these conventions, the standardized ranks are defined as:

Definition 2 *The standardzed ranks of the generalized standardized abnor-*

mal returns are defined as

$$K_{it} = \text{Rank}(\text{GSAR}_{it})/(T' + 1), \tag{9}$$

$i = 1, \ldots, n, \quad t = T_0 + 1, \ldots, T_2,$ *where* $T_0 + 1$ *and* T_2 *are the first and last observation relative to the cumulative event day* $t = 0$, $T' = T_2 - T_0$ $(= T - \tau + 1)$ *is the adjusted number of observations equal to the number of generalized standardized abnormal returns defined in equation* (8)*, and* τ *is the CAR-period length.*

Given that K_{i0} indicates the standardized rank related to the cumulative abnormal return, under the null hypothesis of no mean event effect

$$E\left[K_{i0}\right] = \frac{1}{2}. \tag{10}$$

With these results, we can define a single t-ratio that can be used for testing either cumulative abnormal returns or a single day abnormal returns.

Definition 3 *Given the null hypothesis of no mean event effect*

$$H_0 : \mu_\tau = 0, \tag{11}$$

where $\mu_\tau = E\left[\text{CAR}_\tau\right]$ *is the expected value of the cumulative abnormal returns over the period of length* τ*, the generalized rank* (GRANK) t*-statistic* *is defined as*

$$t_{\texttt{grank}} = \frac{\bar{K}_0 - 1/2}{S_K}, \tag{12}$$

where

$$S_K = \sqrt{\frac{1}{T'} \sum_{t=T_0+1}^{T_2} \frac{n_t}{n} \left(\bar{K}_t - \frac{1}{2} \right)^2}, \tag{13}$$

$$\bar{K}_t = \frac{1}{n_t} \sum_{i=1}^{n_t} K_{it} \tag{14}$$

with n_t the number of valid generalized standardized abnormal returns, $GSAR_{it}$ available at time point t, $t = T_0+1, \ldots, T_2$, $T' = T_2 - T_0$ is the adjusted number of observations in the combined estimation and event period, and $\bar{K}_0$ is the mean $\bar{K}_t$ for $t = 0$, the CAR*-period standardized rank.*

Due to the Central Limit Theorem, the t-statistic in equation (12) is asymptotically $N(0,1)$ distributed under the null hypothesis of (11).

The ranking procedure based on the GSAR-returns, defined in equation (8), with the rank t-ratio in equation (12) gives a novel non-parametric testing procedure for cumulative abnormal returns with several desirable properties. First, as noted above, the procedure captures simultaneously testing for single abnormal returns and cumulative abnormal returns. Second, because the test is based on an equally weighted portfolio of standardized ranks, it is robust against cross-sectional correlation, which is an issue with clustered event days [e.g. Kolari and Pynnonen (2008) and references therein]. Third, the test procedure is robust against event-induced volatility due to the re-standardization of the GSARs by the cross-sectional variance on the event days. Fourth, the procedure takes into account the random event day within the event window, which is an obvious problem with the existing rank tests. Fifth, the rank statistic is less sensitive to single outliers which may badly obscure parametric t-tests. Sixth, and last, the rank test is robust to autocorrelation.[6]

Remark 1 *The generalized cumulative abnormal returns* (GSAR) *can be used to extend the sign test in Corrado and Zivney* (1992) *for testing cumulative abnormal returns. This can be simply achieved by defining*

$$G_{it} = \mathrm{sign}\left[\mathrm{GSAR}_{it} - \mathrm{median}(\mathrm{GSAR}_{it})\right] \tag{15}$$

where $\mathrm{sign}(x)$ *is equal to* $+1$, 0, -1 *as* x *is* > 0, $= 0$, *or* < 0, *and then using these signed values to define the Corrado-Zivney sign test statistic* T_4.

We next comparatively examine the the empirical behavior of our generalized rank test procedure relative to the most popular parametric tests of Patell

[6] Assuming that all return series have equally many observations, it is straightforward to verify that $\mathrm{E}[S_K^2] = \mathrm{var}[\bar{K}_0]$. Thus, there is no autocorrelation bias in the standard error estimate S_K of of $\bar{K}_0$ (see also Remark 2 below).

(1987) and BMP (1991), as well as the non-parametric rank test of Corrado-Zivney (1992) with which our test coincides in the case of testing single abnormal returns.

3. Simulation Design

In this section we employ a simulation methods using actual return data to investigate the empirical behavior of the generalized rank test introduced in Section 2 for testing cumulative abnormal return and compare the results with the the most popular parametric tests, including the ordinary t-test, Patell (1987) t-test, BMP (1991) t-test, and the Corrado-Zivney (1992) rank test.

3.1 *Abnormal Return Model*

The abnormal behavior of security returns can be estimated via the market

$$r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it}, \tag{16}$$

where r_i is the return of stock i, r_m is the index return of value-weighted CRSP stocks, and ϵ_i is is a white noise random component, uncorrelated with r_m.

The resulting abnormal returns are obtained as differences of realized and predicted returns on day t in the event period,

$$\mathrm{AR}_{it} = r_{it} - \left(\hat{\alpha}_i + \hat{\beta}_i r_{mt}\right) \tag{17}$$

where the parameters are estimated from the estimation period with ordinary least squares.

3.2 *Test Statistics*

The ordinary t-test (ORDIN) is defined as

$$t_{\texttt{ordin}} = \frac{\overline{\mathrm{CAR}}_\tau}{\mathrm{s.e}(\overline{\mathrm{CAR}}_\tau)}, \tag{18}$$

where

$$\overline{\mathrm{CAR}}_\tau = \frac{1}{n}\sum_{i=1}^{n} \mathrm{CAR}_{i\tau}, \tag{19}$$

and $\mathrm{s.e}(\overline{\mathrm{CAR}}_\tau)$ is the standard error of the average cumulative abnormal return $\overline{\mathrm{CAR}}_\tau$, which is adjusted with the prediction error [see e.g. Campbell, Lo and MacKinlay (1997), Sec 4.4.3].

The Patell (1987) test statistic (PATELL) is

$$t_{\texttt{patell}} = \sqrt{\frac{n \times (L_1 - p - 3)}{L_1 - p - 1}}\,\overline{\mathrm{SCAR}}_\tau, \tag{20}$$

where n is the number of cross-section observations, L_1 is the length of the estimation period, p is the number of explanatory variables in the abnormal return regression, and $\overline{\mathrm{SCAR}}_\tau$ is the average standardized cumulative abnormal return defined in equation (7).

The Boehmer, Mucumesi, and Poulsen (1992) test statistic (BMP) is

$$t_{\texttt{bmp}} = \frac{\overline{\text{SCAR}}_\tau \sqrt{n}}{S_{\text{SCAR}}}, \tag{21}$$

where S_{SCAR} is the cross-sectional standard deviation of SCARs defined in equation (6).

The Corrado-Zivney (1992) test statistic is based on the standardized ranks

$$U_{it} = \text{rank}(\text{SAR}^*_{it})/(T_i + 1), \tag{22}$$

in which T_i is the length of the time series of return series i and

$$\text{SAR}^*_{it} = \begin{cases} \text{SAR}_{it}, & \text{for } t \neq 0 \\ \dfrac{\text{SAR}_{it}}{S_{\text{SAR}}}, & \text{for } t = 0, \end{cases} \tag{23}$$

with

$$S_{\text{SAR}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\text{SAR}_{i0} - \overline{\text{SAR}}\right)^2} \tag{24}$$

the standard deviation of the event day standardized abnormal returns, and $\overline{\text{SAR}}$ is the average standardized abnormal return on the event day.

The Corrado and Zivney (1992) test statistic (CZ) test is

$$t_{\texttt{cz}} = \frac{\bar{U}_0 - 1/2}{S_{\bar{U}}}, \tag{25}$$

where

$$\bar{U}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} U_{it}, \tag{26}$$

$\bar{U}_0$ is the event day average ($t = 0$), and

$$S_{\bar{U}} = \sqrt{\frac{1}{T}\sum_{t=T_0+1}^{T_2} \frac{n_t}{n}\left(\bar{U}_t - \frac{1}{2}\right)^2} \tag{27}$$

is the standard error of $\bar{U}_0$, and n_t number of valid returns on day t. It is notable that in the CZ test the standard deviation uses both the estimation period and event period observations.

Corrado and Zivney (1992) introduce their statistic only for testing a single event day return. In order to test mean return on event windows longer than one day Cowan (1992) and Campbell and Wasley (1993) suggest simply to aggregate the ranks over the window and use the t-ratio (CUM RANK)

$$t_{\texttt{cumrank},\tau} = \frac{\left(\bar{U}_\tau - \tau/2\right)}{\sqrt{\tau} S_{\bar{U}}}, \tag{28}$$

where

$$\bar{U}_\tau = \sum_{t=t_1+1}^{t_1+\tau} \bar{U}_t \tag{29}$$

is the sum of average ranks over the CAR-period $T_1 + 1 \le t_1 \le t_1 + \tau \le T_2$ with the indexing conventions of Section 2. As discussed above, the major shortcoming of this test is poor power properties when the abnormal return is randomly assigned to one day within the event event window [see also Cowan (1992), pp. 14–15].[7]

[7] Campbell and Wasley (1993) use the Corrado's (1989) non-standardized ranks.

Remark 2 *Unlike GRANK defined in equation* (12)*, CUM RANK is biased by autocorrelation, part of which comes from the technical negative autocorrelation due to ranking. The biasedness can be seen as follows. Assuming again that the returns are cross-sectionally uncorrelated and that there are equally many observation in each return series (*$n_t = n$ *for all* t*), then using straightforward algebra*

$$\mathrm{var}[\bar{U}_\tau] = \frac{\sigma_U^2}{n}\iota_\tau'\Omega_\tau\iota_\tau, \tag{30}$$

where $\sigma_U^2 = \mathrm{var}[U_{it}]$,[8] ι_τ *is a* τ*-vector of ones, and* Ω_τ *is the* $\tau \times \tau$ *average autocorrelation matrix of the individual ranks* $U_{i,t_1+1}, U_{i,t_1+2}, \ldots, U_{i,t_1+\tau}$ *over the CAR-period,* $i = 1, \ldots, n$*. On the other hand, again using straightforward algebra, the expected value of the variance estimator* $\tau S_{\bar{U}}^2$ *of* $\mathrm{var}[\bar{U}_\tau]$ *derived from equation* (27) *and utilized in the CUM RANK* t*-statistic* (28) *becomes*

$$\tau \mathrm{E}[S_{\bar{U}}^2] = \frac{\sigma_U^2}{n}\tau, \tag{31}$$

which coincides with equation (30) *if* Ω_τ *is an identity matrix, i.e., when there is no autocorrelation. Thus,* $\tau S_{\bar{U}}^2$ *is a biased estimator of* $\mathrm{var}[\bar{U}_\tau]$*. In particular, negative autocorrelation implies under-estimation of the true variance and hence under-rejection of the null hypothesis and loss of power.*

[8] Under the null hypothesis of no event effect, the (standardized) ranks are discrete uniform distributed. The variance of the non-standardized ranks is $(T^2 - 1)/12$, and hence, the variance of the standardized ranks becomes $\sigma_U^2 = (T-1)/[12(T+1)] \approx 1/12$. The crucial point here is that they are the same for all series.

3.3 *Sample constructions*

We follow the simulation design introduced in Brown and Warner (1985). From the CRSP database we select 1,000 samples of $n = 50$ return series with replacement. A random event day is assigned for each sample. The database includes 17,878 daily return series in the sample period January 2, 1990 to December 31, 2005.

The event day is denoted as day "0", and the event window consists ± 10 days around the day "0". We report the results for event day $t = 0$ abnormal return AR(0) and for cumulative abnormal returns $\text{CAR}(-1, +1)$, $\text{CAR}(-5, +5)$, and $\text{CAR}(-10, +10)$.

The estimation period is comprised of 239 days prior to the event period (i.e., days -249 to -11). In order for a return series to be included, no missing returns are allowed in the last 30 days from -19 to $+10$.

We also investigate event-induced volatility effects on the test statistics. Charest (1978), Mikkelson (1981), Penman (1982), and Rosenstein and Wyatt (1990) have found that the event period standard deviation is about 1.2 to 1.5 times the estimation period standard deviation (i.e., 20 to 50 percent increased volatility). Accordingly, we indroduce increased volatility by multiplying the cumulated event period returns by a factor $\sqrt{c}$, with values $c = 1$ for no event induced volatility, $c = 1.5$ for an approximate 20 percent increased volatility (i.e., $\sqrt{1.5} \approx 1.2$), $c = 2.0$ for an approximate 40 percent increased volatility (i.e., $\sqrt{2.0} \approx 1.4$), and $c = 3.0$ for an approximate 70 percent increased volatility (i.e., $\sqrt{3} \approx 1.7$) due to the event effect. To add more realism we generate the volatility factors c for each stock based on the following uniform distributions $U[1, 2]$, $U[1.5, 2.5]$, or $U[2.5, 3.5]$, respectively, yielding on average the variance effects of 1.5, 2.0, and 3.0. For the no volatility effect experiment we fix $c = 1.0$

The power properties of the tests are investigated empirically by using a single random event day and assigning a fixed abnormal return to the event window return. Depending on the length of the accumulation period in the CAR, we generate fixed abnormal returns as follows. In the case of a single abnormal return, we generate abnormal returns for the event day $t = 0$ of sizes $\mu_1 = 0.5$, $\mu_2 = 1.0$, $\mu_3 = 1.5$, and $\mu_4 = 2.0$ percents, which we refer to as the base abnormal returns. For the cumulative abnormal periods we use these base abnormal returns and generate the cumulative event effects according to the length of the period, such that the cumulative abnormal return is $(1 + \log \tau)\mu_i$, $i = 1, \ldots, 4$, where τ is the number of days over which the appropriate cumulative return is calculated and log is the natural logarithm. Thus, for CAR$(-1, +1)$ we have $\tau = 3$ and $1 + \log 3 \approx 2$, whereas for CAR$(-5, +5)$ we have $\tau = 11$ and $1 + \log 11 \approx 3.4$, and for CAR$(-10, +10)$ we have $\tau = 21$ and $1 + \log 21 \approx 4$.

4. Results

4.1 *Rejection rates under the null hypothesis*

This section discusses the rejection rates (Type I errors) of the null hypothesis when there is no event effect. The rejection rates indicate the fractions by which the test statistics exceed in 1,000 simulations the nominal cutoffs ± 1.96 at the 5 percent level in two-sided tests.

Sample Statistics

[Table 1]

Table 1 reports sample statistics for the test statistics under the null hypothesis of no event effect. Under the null hypothesis all the test statistics should be approximately $N(0, 1)$-distributed. For the single abnormal return AR(0) the means of all t-statistics are statistically close to zero. For example, the

BMP statistic has a sample mean of -0.044, which is the largest in absolute value, is only 1.43 standard errors below zero. However, all the CARs are slightly negative and statistically significant, which implies that all the test statistics are slightly negative also. The distributions of the test statistics are generally symmetric for the most part and do not exhibit fat tails. Standard deviations are close to their theoretical values of unity.

Rejection rates

Table 2 reports the two-sided rejection rates (Type I errors) at the 5 percent level under the null hypothesis of no event mean effect. The second column shows the results with no event induced volatility. All rejection rates are close to the nominal rate of 0.05 for short CAR-windows of AR(0) and CAR$(-1,+1)$. The PATELL, BMP, and CUM RANK statistics also reject close to the nominal rate for the longer CAR-windows of CAR$(-5,+5)$ and $CAR(-10,+10)$, where all rates are within the approximate 99 percent confidence interval of $[0.032, 0.068]$. The ORDIN and CUM RANK statisitcs under-reject the null hypothesis for these longer CAR- windows. The general tendency for these test statistics is that under-rejection worsens as the length of the CAR-window increases. For the CUM RANK statistics a partial explanation is that it does not account for the technical negative autocorrelation implied by the rank transformation.

[Table 2]

Columns 3, 4, and 5 report the rejection rates under the null hypothesis when there is event- induced variance present. The ORDIN and PATEL statistics over-reject as the variance increases, which is a well known outcome. At the highest variance factor of $c = 3$, which corresponds to an increase in volatility by a factor of 1.7, the Type I error rate for the ORDIN statistics is over 0.20 and for the PATELL statisitic over 0.25. The BMP, CUM RANK, and GRANK statistics are robust to the volatililty increase.

4.2 *Power of the tests*

Power results of the tests are shown in Panels A to D of Table 3 and graphically depicted in corresponding panels of Figure 1. The zero abnormal return line (bold face) in each panel indicates the Type I error rates and replicate the second column of Table 2 (i.e., no event-induced volatility). The remainder of Table 3 indicates the rejection rates for the respective (cumulative) abnormal return in the first column.

There are two outstanding results. First, at all magnitudes of abnormal returns (positive or negative), test statistics based on standardized abnormal returns generally have superior power over the ordinary test statistic ORDIN, which is based on the non-standardized returns. The only exception is the CUM RANK statistic for longer CAR-windows, especially for $CAR(-10, +10)$. This result is consistent with the discussion in Section 2. That is, when the cumulative abnormal return is randomly assigned to single days within the cumulative window, they do not show up as high in the summation. The conservative rejection rate of the CUM RANK statistic under the null hypothesis tends to increase this effect. It should be noted, however, that in unreported simulations we also distributed the cumulative abnormal return evenly across the CAR-window. In this case the CUM RANK statistic has power equal to the PATELL, BMP, and GRANK statistics.

Second, like the PATELL and BMP statistics, the generalized rank test statistic GRANK is immune to the way the cumulative abnormal return is distributed across the (cumulated) event window, as the ranking procedure relies on the cumulative standardized abnormal returns. In addition, as shown in Table 3 and Figure 1, the power of the GRANK statistic tends to be consistently higher than any of the parametric test statistics.

In sum, consistent with earlier studies, the CUM RANK statistics is very competitive and robust relative to the parametric tests for single event days [Corrado (1989), Corrado and Zivney (1992), Cowan (1992), Campbell and Wasley (1993), Kolari and Pynnonen (2008)]. However, as the above simulations and the results in Cowan (1992) indicate, its usefulness for testing cumulative abnormal returns can be problematic. Our suggested GRANK statistic fills this gap and extends the usage of robust non-parametric rank tests to testing cumulative abnormal returns.

5. Conclusions

This paper has introduced a generalized rank test based on generalized standardized abnormal returns. This procedure can be used to test both single abnormal returns as well as cumulative abnormal returns. The test is robust to abnormal return serial correlation and event-induced volatility. Also, it implicitly accounts the possible cross-correlation of abnormal returns, which is an issue if event days are clustered. Simulation results show that the test has good and often superior (empirical) power relative to popular parametric tests at all event window lengths.

Earlier studies have shown the rank test of Corrado (1989) and Corrado and Ziwney (1993) is very competitive and robust relative to the parametric tests for single event days [Corrado (1989), Corrado and Zivney (1992), Cowan (1992), Cambell and Wasley (1993), Kolari and Pynnonen (2008)]. However, its usefulness for testing cumulative abnormal returns is questionable when the abnormal return is randomly assigned within the event window. Our suggested GRANK test avoids these problems and provides a robust nonparametric rank test that can be used for testing cumulative abnormal returns with equally high power as single abnormal returns.

References

Beaver, William, H. (1968). The information content of annual announcements, *Journal of Accounting research Supplement* 6, 67–92.

Barber, Brad, M. and John D. Lyon (1997). Detecting long-run abnormal returns: The empirical power and specification test statistics, *Journal of Financial Economics* 43, 341–372.

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.

Brown, Stephen, J. and Jerold B. Warner (1985). Using daily stock returns. case of event studies, *Journal of Financial Economics* 14, 3–31.

Campbell, Cynthia, J. and Charles E. Wasley (1993). Measuring Security performance using daily NASDAQ returns, *Journal of Financial Economics* 33, 73–92.

Campbell, John, Y., Andrew W. Lo, and A. Craig MacKinlay (1997). *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

Charest, Guy (1978) Dividend information, stock returns, and market efficiency, *Journal of Financial Economics* 6, 265–296.

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

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

Cowan, Arnold, R. (1992). Nonparametric event study tests, *Review of Quantitative Finance and Accounting* 2, 343-358.

Dechow, P., R. Sloan, and A. Sweeney (1995). Detecting earnings management, *Accounting Review* 70, 3–42.

Harrington, Scott, E. and David G. Shrider (2007). All events induce variance: Analyzing abnormal returns when effects vary across firms, *Journal of Financial and Quantitative Analysis* 42, 229–256.

Kolari, James, W. and Seppo Pynnonen (2008). Correction for cross-sectional correlation in testing for abnormal returns in event studies, *Submitted*.

Kothari, S.P., A. Leone, and C. Wasley (2005). Performance-matched discretionary accruals, *Journal of Accounting and Economics* 39, 163–197.

Luoma, Terhi and Seppo Pynnonen (2008). Testing for Cumulative Abnormal Returns with the Rank Test (*Work in progress*).

Mikkelson, Wayne, H. (1981) Convertible calls and security returns, *Journal of Financial Economics* 9, 237-264.

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

Rosenstein, Stuart and Jeffrey G. Wyatt (1990). Outside directors, board independence, and shareholders wealth, *Journal of Financial Economics* 26, 175-191.

Table 1. Sample statistics in event tests for 1,000 random portfolios of $n = 50$ securities from CRSP data base.

Test statistics	Fama-French 3-factor adjusted abnormal returns.						
Panel A: AR(0)	Mean	Median	Std	Skewness	Excess Kurtosis	Min	Max
Abnormal Return	-0.020	-0.012	0.579	-0.472^c	3.754^c	-4.265	1.926
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	-0.033	-0.020	0.997	-0.475^c	3.361^c	-7.068	3.447
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	-0.030	-0.030	1.026	-0.103	0.532	-4.368	3.808
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	-0.044	-0.030	0.985	-0.065	0.156	-3.486	4.252
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (28)]	-0.009	0.014	0.977	-0.016	0.256	-3.265	3.953
GRANK $t_{\texttt{grank}}$ [Eq. (12)]	-0.009	0.014	0.977	-0.016	0.256	-3.265	3.953
Panel B: CAR($-1,+1$)	Mean	Median	Std	Skewness	Excess Kurtosis	Min	Max
CAR($-1,+1$)	-0.042	-0.051	0.943	-0.171^b	1.458^c	-4.784	3.517
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	-0.043	-0.050	0.949	-0.207	1.664^c	-5.606	3.367
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	-0.104^c	-0.101^c	1.033	0.088	0.582^c	-3.595	4.056
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	-0.110^c	-0.104^c	1.026	0.015	0.147	-3.416	3.570
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (28)]	-0.056	-0.038	0.957	-0.011	0.211	-3.261	3.895
GRANK $t_{\texttt{gcar}}$ [Eq. (12)]	-0.053	-0.010	0.996	-0.027	0.196	-3.612	3.590
Panel C: CAR($-5,+5$)	Mean	Median	Std	Skewness	Excess Kurtosis	Min	Max
CAR($-5,+5$)	-0.211^c	-0.220^c	1.697	0.060	0.262	-6.201	6.231
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	-0.115^c	-0.110^c	0.871	-0.008	0.107	-2.993	3.072
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	-0.172^c	-0.146^c	0.984	-0.106	0.166	-3.392	3.403
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	-0.168^c	-0.144^c	1.043	-0.017	0.229	-4.270	3.927
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (28)]	-0.077^c	-0.053	0.887	-0.044	0.040	-2.713	3.037
GRANK $t_{\texttt{grank}}$ [Eq. (12)]	-0.063	0.006	1.038	-0.091	0.206	-3.792	3.592
Panel D: CAR($-10,+10$)	Mean	Median	Std	Skewness	Excess Kurtosis	Min	Max
CAR($-10,+10$)	-0.390^c	-0.485^c	2.426	0.138	-0.027	-7.146	8.234
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	-0.151^c	-0.167^c	0.876	0.057	-0.233	-2.636	2.462
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	-0.227^c	-0.186^c	0.936	0.067	-0.127	-2.649	3.134
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	-0.236^c	-0.194^c	1.014	0.114	-0.010	-2.852	3.750
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (28)]	-0.150^c	-0.154^c	0.860	0.149^a	0.032	-2.663	2.944
GRANK $t_{\texttt{grank}}$ [Eq. (12)]	-0.098^c	-0.091^b	1.047	0.046	-0.058	-3.034	3.654

Rejection rates based on 1,000 simulations for portfolios of size 50 securities with estimation period of 239 days and event period 21 days. The event day is the day 250 denoted as $t = 0$. Cumulative abnormal returns, CAR($-d,+d$), with $d = 0$, 1, 5, and 10 are cumulated around the event day. Securities from CRSP and event dates from period 1990 2005 are randomly selected with replacement. Ordinary t-test, $t_{\texttt{patell}}$, and $t_{\texttt{bmp}}$ are parametric tests, $t_{\texttt{cumrankr}}$, and $t_{\texttt{grank}}$ are nonparametric tests. $a = 10\%$, $b = 5\%$, and $c = 1\%$ significant. GRANK and CUM RANK coincide for the single event day abnormal return AR(0).

Table 2. Rejection rates in two-tailed test at the 5 % level of the null hypothesis of no mean event effects with different event windows and with different levels of event induced volatility.

Test statistics	Market model abnormal returns. Event induced volatility, $\sqrt{c}\,\sigma$ $c = 1.0$	$c = 1.5$	$c = 2.0$	$c = 3.0$
Panel A: AR(0)				
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	0.043	0.094	0.145	0.220
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	0.051	0.116	0.178	0.268
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	0.045	0.042	0.043	0.043
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (28)]	0.045	0.047	0.051	0.042
GRANK $t_{\texttt{grank}}$ [Eq. (12)]	0.045	0.047	0.051	0.042
Panel B: CAR$(-1, +1)$				
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	0.037	0.084	0.128	0.208
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	0.069	0.115	0.161	0.261
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	0.054	0.051	0.056	0.056
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (25)]	0.037	0.038	0.036	0.038
GRANK $t_{\texttt{grank}}$ [Eq. (12)]	0.048	0.047	0.046	0.049
Panel C: CAR$(-5, +5)$				
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	0.029	0.063	0.105	0.195
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	0.057	0.105	0.154	0.250
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	0.057	0.061	0.059	0.062
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (28)]	0.028	0.027	0.028	0.028
GRANK $t_{\texttt{grank}}$ [Eq. (12)]	0.059	0.058	0.055	0.056
Panel D: CAR$(-10, +10)$				
ORDIN $t_{\texttt{ordin}}$ [Eq. (18)]	0.026	0.076	0.121	0.212
PATELL $t_{\texttt{patell}}$ [Eq. (20)]	0.045	0.103	0.151	0.246
BMP $t_{\texttt{bmp}}$ [Eq. (21)]	0.066	0.071	0.064	0.063
CUM RANK $t_{\texttt{cumrank}}$ [Eq. (28)]	0.022	0.023	0.022	0.022
GRANK $t_{\texttt{grank}}$ [Eq. (12)]	0.067	0.066	0.066	0.069

Rejection rates based on 1,000 simulations for portfolios of size 50 securities with estimation period of 239 days and event period 21 days. The event day is day 250 denoted as $t = 0$. Cumulative abnormal returns, CAR$(-d, +d)$, with $d = 0$, 1, 5, and 10 are cumulated around the event day. Securities from CRSP and event dates from period 1990 to 2005 are randomly selected with replacement. The ordinary t-test $t_{\texttt{ordin}}$, the PATELL test $t_{\texttt{patell}}$, and the BMP test $t_{\texttt{bmp}}$ are parametric tests, the CUM RANK test, $t_{\texttt{cumrank}}$, and the GRANK test $t_{\texttt{grank}}$ are nonparametric tests. GRANK and CUM RANK coincide for the single event day abnormal return AR(0). The 95 percent confidence interval around the 0.05 rejection rate is $[0.036, 0.064]$ and the respective 99 percent confidence interval is $[0.032, 0.068]$

Table 3. Two-tailed average rejection rates at the 0.05 significance level for selected test statistics sampled from 1,000 random portfolios of n = 50 securities with randomly assigned (cumulative) abnormal return on one event day within the cumulated window.

		Test statistic				
Panel A:	AR(0)	ORDIN	PATELL	BMP	CUM RANK	GRANK
	−3.0	0.998	1.000	0.999	1.000	1.000
	−2.0	0.939	1.000	0.995	0.999	0.999
	−1.0	0.433	0.855	0.822	0.876	0.876
	−0.5	0.129	0.324	0.348	0.378	0.378
	0.0	**0.043**	**0.051**	**0.045**	**0.045**	**0.045**
	0.5	0.120	0.324	0.314	0.384	0.384
	1.0	0.405	0.852	0.824	0.900	0.900
	2.0	0.935	0.997	0.992	0.998	0.998
	3.0	0.995	1.000	0.999	1.000	1.000
Panel B:	CAR(−1, +1)	ORDIN	PATELL	BMP	CUM RANK	GRANK
	−4.2	0.989	1.000	0.998	0.998	0.999
	−2.1	0.599	0.956	0.936	0.909	0.967
	−1.0	0.173	0.498	0.514	0.475	0.544
	0.0	**0.037**	**0.069**	**0.054**	**0.037**	**0.048**
	1.0	0.161	0.405	0.439	0.471	0.506
	2.1	0.552	0.935	0.910	0.929	0.968
	4.2	0.986	1.000	0.996	1.000	1.000
Panel C:	CAR(−5, +5)	ORDIN	PATELL	BMP	CUM RANK	GRANK
	−8.5	0.993	1.000	1.000	0.846	1.000
	−5.1	0.793	0.996	0.995	0.742	0.995
	−2.5	0.266	0.684	0.699	0.463	0.716
	−0.8	0.056	0.155	0.182	0.119	0.169
	0.0	**0.029**	**0.057**	**0.057**	**0.028**	**0.059**
	0.8	0.030	0.078	0.105	0.083	0.138
	2.5	0.206	0.566	0.605	0.457	0.686
	5.1	0.731	0.990	0.970	0.833	0.994
	8.5	0.992	1.000	0.998	0.967	0.999
Panel D:	CAR(−10, +10)	ORDIN	PATELL	BMP	CUM RANK	GRANK
	−10.1	0.967	1.000	0.999	0.505	0.999
	−6.1	0.659	0.980	0.982	0.482	0.976
	−4.0	0.362	0.795	0.819	0.390	0.811
	−2.0	0.114	0.329	0.376	0.205	0.347
	−1.0	0.058	0.132	0.163	0.094	0.146
	0.0	**0.026**	**0.045**	**0.066**	**0.022**	**0.067**
	1.0	0.024	0.047	0.070	0.040	0.096
	2.0	0.059	0.172	0.213	0.133	0.291
	4.0	0.247	0.647	0.667	0.352	0.759
	6.1	0.527	0.953	0.946	0.524	0.977
	10.1	0.958	1.000	1.000	0.728	1.000

GRANK and CUM RANK coincide for the single event day abnormal return AR(0).

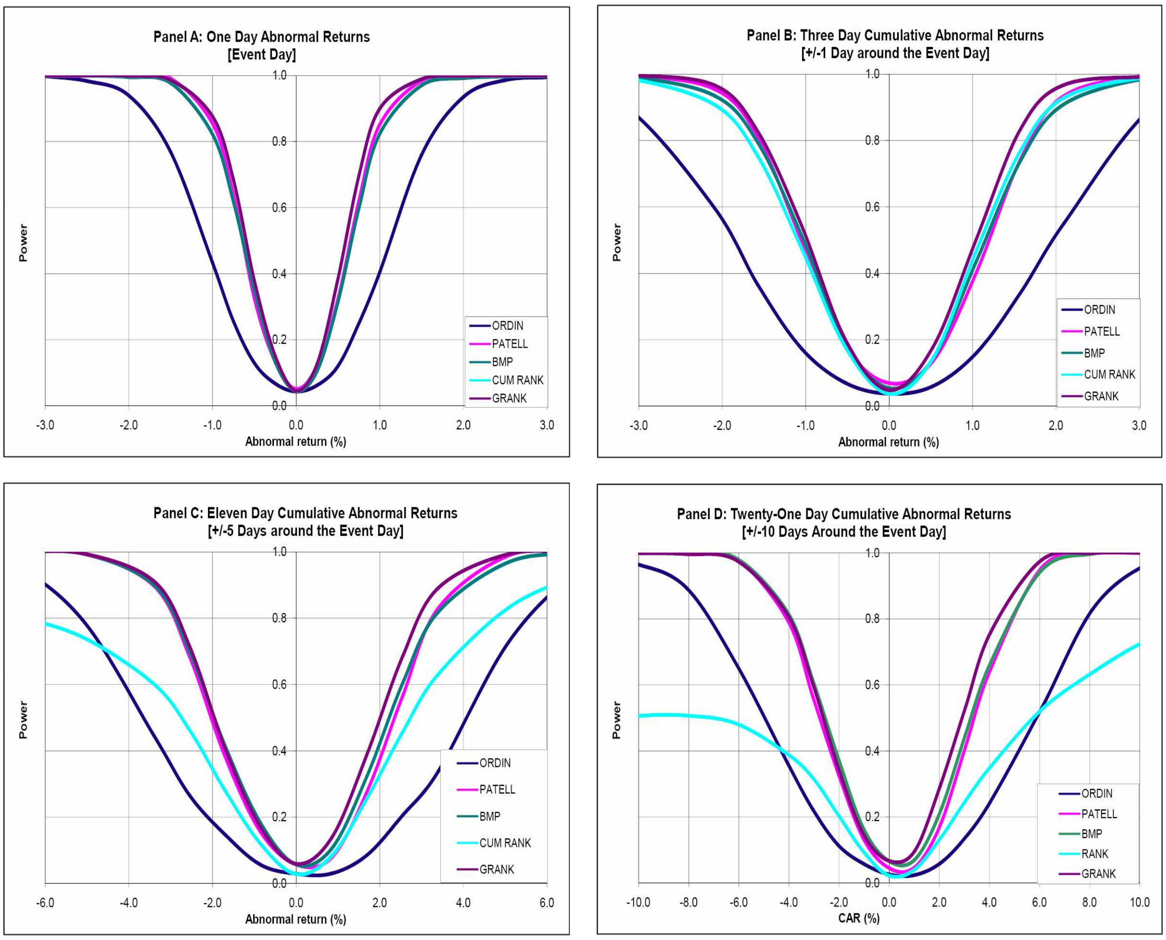

Figure 1. Estimated power functions with different CAR-windows for ORDIN, PATELL, BMP, CUM RANK, and GEN RANK tests based on 1,000 samples of $n = 50$ security portfolios from the CRSP database: Significance level is 0.05, two-sided tests, and no event-induced variance.

GRANK and CUM RANK coincide for the single event day abnormal return AR(0).