Event Study Testing with Cross-sectional Correlation of Abnormal Returns

James W. Kolari
Texas A&M University

Seppo Pynnönen
University of Vaasa, Finland

This article examines the issue of cross-sectional correlation in event studies. When there is event-date clustering, we find that even relatively low cross-correlation among abnormal returns is serious in terms of over-rejecting the null hypothesis of zero average abnormal returns. We propose a new test statistic that modifies the t-statistic of Boehmer, Musumeci, and Poulsen (1991) to take into account cross-correlation and show that it performs well in competition with others, including the portfolio approach, which is less powerful than other alternatives under study. Also, our statistic is readily useable to test multiple-day cumulative abnormal returns. (*JEL* G14, C10, C15)

It is well known that event studies are prone to cross-sectional correlation among abnormal returns when the event day is the same for sample firms. For this reason, test statistics cannot assume independence of abnormal returns. This article shows that, even when cross-correlation is relatively low, event-date clustering is serious in terms of over-rejecting the null hypothesis of zero average abnormal returns when it is true. To address this problem, we utilize scaled (or standardized) abnormal returns and propose a new t-test statistic that takes into account both cross-correlation and inflation of event-date variance. Extensive simulations are conducted to examine the size and power of several test statistics for random samples of stocks taken from both the whole market and a single industry in the period 1990 to 2005. Experiments with alternative methods of computing abnormal returns that employ different levels of cross-correlation as well as different levels of event-induced volatility reveal that our

The main ideas in this paper were developed while Professor Pynnonen was a Visiting Faculty Fellow at the Mays Business School, Texas A&M University under a sabbatical leave funded by a senior scientist post from the Academy of Finland. The hospitality of the Finance Department and Center for International Business Studies in the Mays Business School and the generous funding of the Academy of Finland and financial support from OP-Group research foundation are gratefully acknowledged. We have benefited from comments by participants at the 2006 annual meetings of the Financial Management Association International, including Ekkehart Boehmer, Jim Musumeci, and Catherine Shenoy. Also, we are grateful for helpful comments from Jaap Bos, Paige Fields, Donald Fraser, Johan Knif, and Scott Lee, the referees, and Joel Hasbrouck (editor). Send correspondence to Seppo Pynnönen, Department of Mathematics and Statistics, University of Vaasa, P. O. Box 700, Vaasa, Finland; telephone: +358-6-3248259.

doi:10.1093/rfs/hhq072 Advance Access publication September 21, 2010

new test statistic is the only parametric method that is robust to these event-date clustering issues. Also, when testing *cumulative abnormal returns* (CARs) in multiple-day windows, our test statistic increasingly dominates nonparametric tests as the window is lengthened.

The next section gives a brief review of relevant literature. Section 2 discusses biases in the presence of cross-correlation. Section 3 reviews previous test statistics, proposes new test statistics for cross-correlated abnormal returns, and discusses some estimation and asymptotic properties of the proposed test statistics. Section 4 overviews the simulation design, and Section 5 presents the simulation results. Section 6 concludes.

1. Literature Review

Ignoring contemporaneous correlations in event-date clustered analyses of returns may introduce considerable downward bias in the standard deviation and thereby overstate the t-statistic, which leads to over-rejection of the null hypothesis (Collins and Dent 1984; Bernard 1987; Salinger 1992; Aktas, de Bodt, and Roll 2004; Kothari and Warner 2007). The problem with clustered event days is similar to a panel data setup. As discussed by Petersen (2008), there are two general forms of dependencies in these data sets (viz., autocorrelation and cross-correlation). Both of these sources tend to cause substantial bias in the standard errors of ordinary least squares (OLS) regression estimates if not properly accounted for, thereby resulting in unreliable inferences. We discuss briefly the autocorrelation problem after our simulation results and show that under fairly general conditions our proposed test statistic (viz., an adjusted Boehmer, Musumeci, and Poulsen [ADJ-BMP] statistic) is robust in this respect. Our main focus is on potential bias due to cross-correlation. In this respect, the usual approach to account for cross-correlation between returns is the portfolio method suggested by Jaffe (1974). This approach aggregates firm returns into an equally weighted portfolio and investigates the abnormal returns of the portfolio. While this method captures contemporaneous dependency between the returns, it is generally suboptimal due to lower power than other alternatives.

There have been several other attempts in the literature to resolve the contemporaneous correlation problem (Kothari and Warner 2007). For example, the multivariate regression method (MVRM) with generalized least squares (GLS) is known to be optimal under certain assumptions. Unfortunately, it requires accurate estimation of the covariance matrix of the residual returns, which is not normally possible in finite samples, especially as the number of firms increases. Also, because every covariance plays an eminent role in GLS through matrix inversion, estimation errors in individual covariances introduce more sampling error into the standard errors than they eliminate, thereby making the test results even worse (Malatesta 1986). Chandra and Balachandran (1990) further argue that GLS is highly sensitive to model misspecification, which may lead to inefficient test results even if the covariance matrix is known.

They conclude that GLS should be avoided in event studies because the correct model specification is rarely known for certain. Consequently, these authors recommend the use of nongeneralized least squares that would yield different weighted portfolio methods.

Some non-parametric tests have proven to be useful in testing for event effects due to well-documented (theoretical) robustness results in statistical literature (e.g., see Lehmann 1975; Bickel and Doksum 1977; Conover 1999). For example, the Wilcoxon (1945) rank-sum test has relatively higher power compared with parametric tests, particularly for fat-tailed distributions. Based on these ideas, Corrado (1989) and Corrado and Zivney (1992) recommend non-parametric rank and sign tests that are expected to be robust against event-induced volatility and cross-correlation. The most popular approach for testing CARs with these methods is a cumulated ranks test. Over a small number of periods, this cumulative rank test is able to detect abnormal behavior (Cowan 1992; Campbell and Wasley 1993, 1996). However, for longer-period CARs, it is prone to incremental misspecification in the standard deviation of the related t-ratio. This misspecification stems from the technical serial correlation of the ranks, with the effect increasing as the number of cumulated periods grows. Furthermore, the test rapidly loses power in CARs if the event effect is randomly assigned to one day within the event window (Cowan 1992, p. 14; Kolari and Pynnonen 2008).

Particularly relevant to the present study, parametric tests based on scaled abnormal returns methods have been found to be superior in terms of power over those based on non-scaled returns. Scaled (or standardized) abnormal returns are defined as abnormal returns divided by the standard deviation of estimation-period residuals corrected by the prediction error. The most widely used scaled tests are the t-statistics of Patell (1976) and Boehmer, Musumeci, and Poulsen (1991). Savickas (2003) has modified the BMP statistic by scaling the abnormal returns with conditional standard deviations estimated by a generalized autoregressive conditional heteroscedasticity (GARCH) model. The advantage of the BMP test over the Patell test is that BMP accounts for potential event-induced volatility, in addition to potential return autocorrelation, as discussed in Section 5.3. While an increase in event-period variability is considered to be intuitive (see Collins and Dent 1984; Brown and Warner 1980, 1985; Kothari and Warner 2007), Harrington and Shrider (2007) show that cross-sectional variation in the effects of events *always* produces event-induced variance.

To better understand how researchers deal with cross-correlation and event-induced variance inflation, we surveyed all event studies with potential event-date clustering published in the *Journal of Finance*, the *Journal of Financial Economics*, the *Journal of Financial and Quantitative Analysis*, and the *Review of Financial Studies* in the period 1980 to mid-2007. Online Appendix A summarizes our findings for 76 studies. Some studies focus on a common date (e.g., regulatory, government, and legal events); however, we also include numerous studies with possible event-date clustering due to the density of event dates surrounding takeover, merger and acquisition, bankruptcy and financial distress, newly listed and delisted stock, securities markets, and other events. Many other event studies likely have some degree of event-date clustering. Because the efficient scaled Patell and BMP statistics are not applicable to cross-correlated returns, studies deal with the correlation and variance inflation in the following ways (with number of studies in parentheses): sign and rank nonparametric tests (27), GLS (16), and other (4). Sign or rank nonparametric tests invariably employ Wilcoxon (1945) tests rather than the more recent Corrado (1989) and Corrado and Zivney (1992) tests. Correlation is most often taken into account by using the portfolio method (55 studies), which not only is relatively inefficient but does not account for possible event-induced variance. Only one study, by Aktas, de Bodt, and Roll (2004), employs the scaled BMP test by bootstrapping the p-values in order to correct them for the cross-correlation effect. We infer from these findings that, while many event studies attempt to adjust for variance inflation and cross-correlation, they avoid efficient scaled tests due to the cross-correlation effect.

2. Biases in the Presence of Correlation

In forthcoming theoretical derivations, we make the conventional assumption that asset returns $r_{1t}, r_{2t}, \ldots, r_{nt}$ of n firms for calendar time period t are serially independently multivariate normally distributed random variables with constant mean and constant covariance matrix for all t (Campbell, Lo, and MacKinlay 1997, Section 4.3). We consider the problem induced by cross-correlation in the simple setting of testing for zero-mean abnormal returns with a t-ratio on a single common event day. Given this setup, we demonstrate that scaled returns reduce the implied cross-correlation problem into the single number of average correlation, which provides a novel way to account for the misspecification of standard deviation in the scaled tests relying on independent cross-sectional returns. The key to the simplification is that all scaled abnormal returns have the same variance, which we denote generally as σ_A^2. Also, let σ_{ij} denote the population covariance of scaled abnormal returns for securities i and j. Then, using simple algebra, the variance of the mean of the scaled abnormal returns over n firms is

$$\sigma_{\bar{A}}^2 = \frac{1}{n}\sigma_A^2 + \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}\sigma_{ij}. \tag{1}$$

Because the variances are the same for all scaled abnormal returns, i.e., $\sigma_i^2 = \sigma_j^2 = \sigma_A^2$, the covariances can be written as

$$\sigma_{ij} = \sigma_i\sigma_j\rho_{ij} = \sigma_A^2\rho_{ij}, \tag{2}$$

where ρ_{ij} is the correlation of the abnormal returns of stocks i and j. As such, we can write Equation (1) as

$$\sigma_{\bar{A}}^2 = \sigma_A^2 \left(\frac{1}{n} + \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j \neq i} \boldsymbol{\rho}_{ij} \right) = \frac{\sigma_A^2}{n} \left(1 + (n-1)\bar{\rho}\right), \tag{3}$$

where $\bar{\rho}$ is the average correlation of the abnormal returns.[1] Thus, the entire cross-correlation problem in the standard error of the mean of scaled abnormal returns reduces to the single number of mean correlation. The advantage of this procedure becomes evident in particular in the estimation, wherein averaging tends to cancel out individual estimation errors in single correlations. This is a substantial advantage over, for example, GLS-based approaches in which every estimated correlation plays a crucial role. At the extreme, if a single correlation becomes badly estimated close to unity, the covariance matrix becomes ill-conditioned and GLS fails. In the averaging, each correlation has weight $2/n(n-1)$, where n is the number of series, and thus cannot have such a dominating effect. It is also interesting to note from Equation (3) that, in order to keep the correlation matrix positive definite, the return correlations cannot be highly negative on average.

Assuming that the event does not change the residual correlation, the average correlation of the abnormal returns can be estimated by averaging the sample correlations of the estimation-period residuals. Because residual correlation on average is positive, it is obvious that assuming zero residual correlation will underestimate the variance by the factor $(1 + (n-1)\bar{\rho})$ and lead to overstated rejection rates of the t-statistics. According to Equation (3), the severity of cross-sectional correlation in the subsequent t-statistics is a function of both the average correlation and the number of firms.

Suppose next that there are blocks of firms such that cross-correlation is zero between the blocks. In such a case, estimation of the average cross-correlation can be improved by taking into account the blocking. Thus, assume that there are q clusterings or groupings of multiple event days and that in each group the event day is the same for the corresponding firms. For clustered common event days, the correlations of the non-overlapping event-day-groups are zero, and the covariance matrix of the scaled abnormal returns is block-diagonal. Equivalently, we can form groups across q industries (i.e., firms from the same industry), where all have the same event day but

[1] Scaled abnormal returns are the prediction errors divided by the estimated residual standard deviation of the factor model used to define returns. Under the normality assumption, the scaled returns are t-distributed with $m-p-1$ degrees of freedom, where m is the length of the estimation period, and p is the number of explanatory variables in the factor model. Thus, the common variance of the abnormal returns is $\sigma_A^2 = (m-p-1)/(m-p-3)$. If the estimation periods are different for each firm, then $\sigma_j^2 = (m_j-p-1)/(m_j-p-3)$, where m_j is the number of observations in the estimation period for the jth firm. In this case, Equation (3) holds only approximately. Nevertheless, if the estimation periods are reasonably long, such that $\sigma_i^2 = (m_i-p-1)/(m_i-p-3) \approx (m_j-p-1)/(m_j-p-3) = \sigma_j^2$, then substituting σ_i^2 and σ_j^2 with $\sigma_A^2 = \frac{1}{n}\sum_{j=1}^{n}(m_j-p-1)/(m_j-p-3)$ (Patell 1976) in Equation (2) introduces only negligible bias into the standard error of the mean abnormal return.

it is assumed that the between-industry correlations are zero. In both cases, the kth block corresponds to the covariance matrix of the firms belonging to the kth group with covariance matrix Σ_k, $k = 1, \ldots, q$. The average scaled abnormal return is $\bar{A} = \frac{1}{n}\sum_{k=1}^{q} n_k \bar{A}_k$, where $\bar{A}_k$ is the average scaled abnormal return in subgroup k, and n_k is the number of firms in subgroup k. Again assuming that within each subgroup the variances of the scaled returns are the same, the variance of $\bar{A}_k$ is of the form shown in Equation (1). Consequently, the variance of the average abnormal return over all firms becomes

$$\sigma^2_{\bar{A}} = \frac{1}{n^2}\sum_{k=1}^{q} n_k^2 \sigma^2_{\bar{A}_k} = \frac{1}{n^2}\sum_{k=1}^{q} n_k \sigma_k^2 (1 + (n_k - 1)\bar{\rho}_k), \tag{4}$$

where $\sigma^2_{\bar{A}_k}$ is the variance of the average abnormal returns in group k, σ_k^2 is the variance of the scaled abnormal return in group k, and $\bar{\rho}_k$ is the average abnormal return correlation in group k.[2] Again, ignoring residual correlation tends to understate variance and upwardly bias rejection rates of test statistics.

3. Test Statistics

Brown and Warner (1980) define abnormal returns of a security for three different event-study portfolio models of the process generating ex ante expected returns: MEAN, MARKET, and OLS MODEL. These portfolio models are defined in Table 1 for easy reference. In order to reduce the residual cross-correlation to a minimum, Table 1 defines three additional models: the Fama and French (1992, 1993) three-factor model (FF MODEL), the Fama-French three-factor model augmented with an industry portfolio (FF INDUSTRY MODEL), and the industry adjusted model (INDUSTRY). The latter two models are used in single-industry analyses available in the online appendix.

Relevant to the present article, Patell (1976) argues that scaled abnormal returns (SARs) should be used for statistical tests. SARs are defined as

$$A_{it} = \frac{AR_{it}}{s_i\sqrt{1 + d_t}}, \tag{5}$$

where s_i is the regression residual standard deviation, and d_t is the correction term of the form $x_t'(X'X)^{-1}x_t$ due to the estimation of the regression

[2] In the more general case where m_{jk} is the length of the estimation period of firm j in group k, $j = 1, \ldots, n_k$, $k = 1, \ldots, q$, the variances are $\sigma_j^2(k) = (m_{jk} - p - 1)/(m_{jk} - p - 3)$. If the estimation periods have a reasonable number of observations, then $\sigma_i^2(k) = (m_{ik} - p - 1)/(m_{ik} - p - 3) \approx (m_{jk} - p - 1)/(m_{jk} - p - 3) = \sigma_j^2(k)$ and approximating individual variances $\sigma_i^2(k)$ with the average $\sigma_k^2 = \frac{1}{n_k}\sum_{j=1}^{n_k}(m_{jk} - p - 1)/(m_{jk} - p - 3)$ again introduces only negligible bias into the standard error of the average abnormal return in Equation (4).

Table 1
Ex ante abnormal return generating models

Model label	Abnormal return model
MEAN	$AR_{it} = r_{it} - \bar{r}_i$
MARKET	$AR_{it} = r_{it} - r_{mt}$
OLS MODEL	$AR_{it} = r^e_{it} - \alpha_i - \beta_i r^e_{mt}$
FF MODEL	$AR_{it} = r^e_{it} - \alpha_i - \beta_{im} r^e_{mt} - \beta_{i,smb} SMB - \beta_{i,hml} HML$
FF INDUSTRY MODEL[a]	$AR_{it} = r^e_{it} - \alpha_i - \beta_{im} r^e_{mt} - \beta_{i,smb} SMB - \beta_{i,hml} HML - \beta_{i,I} I^e_t$
INDUSTRY[a]	$AR_{it} = r_{it} - I_t$

Notation is defined as follows: r_{it} is the observed return, $\bar{r}_i$ is the sample mean return estimated from the estimation period, r_{mt} is the market return, r^e_{it} and r^e_{mt} are stock i and market excess returns over the Treasury bill rate, respectively; α_i and β_i are estimated OLS coefficients from the estimation period; *SMB* is the small-minus-big market capitalization portfolio return; and *HML* is the high-minus-low book equity/market equity portfolio return.
[a]These models are used in the industry simulations that are reported in Online Appendix C, where I^e_t is the industry excess return, and I_t is the industry raw return.

parameters in the estimation period, with vector x_t of explanatory variable values (including the constant) on event day t, and matrix X of explanatory variable values in the estimation period. The basic intuition is that standardization weights individual observations by the inverse of the standard deviation, which implies that more volatile (i.e., more noisy) observations get less weight in the averaging than the less volatile or more reliable observations. While scaled abnormal returns are more difficult to interpret than raw returns, they have been proven to exhibit better statistical properties. Thus, scaled returns should be used only for statistical testing purposes as signal detection devices of the event effect, while raw returns carry the economic information for interpretation purposes when a signal is detected.

3.1 Scaled test statistics

Scaled abnormal returns are used by BMP (1991) to define the following t-statistic:

$$t_B = \frac{\bar{A}\sqrt{n}}{s}, \tag{6}$$

where s is the (cross-sectional) standard deviation of the event-day scaled abnormal returns defined as the square root of the sample variance, or

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(A_i - \bar{A})^2. \tag{7}$$

It can be easily shown that[3]

$$E\left[s^2\right] = (1-\bar{\rho})\sigma_A^2. \tag{8}$$

3 Proof available from the authors upon request. See also Sefcik and Thompson (1986, p. 327).

Thus, Equation (7) is a biased estimator of the variance σ_A^2 when contemporaneous return correlations are nonzero. Normally, because $\bar{\rho}$ is positive, s^2 understates the true cross-sectional variance. Because $E[s^2/(1-\bar{\rho})] = \sigma_A^2$, a feasible estimator of the variance σ_A^2 is

$$s_A^2 = \frac{s^2}{1-\bar{r}}, \tag{9}$$

where $\bar{r}$ is the average of the sample cross-correlations of the estimation-period residuals. Therefore, the variance of the mean abnormal return $\bar{A}$ can be estimated by replacing the parameters in Equation (3) with their respective estimators so that

$$s_{\bar{A}}^2 = \frac{s_A^2}{n}(1+(n-1)\bar{r}). \tag{10}$$

Using this variance in the t-ratio results in a scaled test statistic (hereafter ADJ-BMP) that accounts for both cross-correlation and event-induced volatility in testing for the mean event effect:

$$t_{AB} = \frac{\bar{A}}{s_{\bar{A}}} = \frac{\bar{A}\sqrt{n}}{s_A\sqrt{1+(n-1)\bar{r}}}. \tag{11}$$

Hence, the BMP t-statistic can be adjusted as follows: $t_{AB} = t_B\sqrt{\frac{1-\bar{r}}{1+(n-1)\bar{r}}}$. If the average return cross-correlation is zero, the ADJ-BMP and BMP statistics yield the same results even if the event days are clustered with cross-correlated returns.

It should be noted that, under the assumption of the square-root rule of the standard deviation of returns over different return periods, the correlation factor $\sqrt{\frac{1-\bar{r}}{1+(n-1)\bar{r}}}$ for the ADJ-BMP t-statistic remains the same when testing CARs with these tests.[4] Thus, the t-ratio in Equation (11) is readily available for testing CARs by replacing the mean scaled abnormal return $\bar{A}$ with the mean scaled cumulative abnormal return (SCAR) and the standard deviation s_A with the cross-sectional standard deviation of SCARs.

The Patell (1976) statistic (PATELL) is based on scaled abnormal returns also. It is defined as

$$t_P = \frac{\bar{A}\sqrt{n}}{\sqrt{(m-p-1)/(m-p-3)}} = \bar{A}\sqrt{\frac{n\times(m-p-3)}{m-p-1}}, \tag{12}$$

[4] This is due to the fact that, if we let $r_t(\tau) = \log(P_t) - \log(P_{t-\tau})$ denote the τ-period log return of a security, then under the assumptions of independence and homogeneity of one-period return variances, the τ-period standard deviation is $\sigma(\tau) = \sqrt{\tau}\,\sigma$, where σ is the one-period standard deviation of returns. This result also applies to the covariances (as variance is a special case of covariance), such that $\sigma_{ij}(\tau) = \tau\,\sigma_{ij}$, where σ_{ij} is the one-period cross-sectional covariance between securities i and j. Thus, $\rho_{ij}(\tau) = \sigma_{ij}(\tau)/\sigma_i(\tau)\sigma_j(\tau) = \sigma_{ij}/\sigma_i\sigma_j = \rho_{ij}$, which implies that the τ-period cross-sectional correlation is the same as the one-period correlation.

where p is the number of explanatory variables in the expected return regression. Using Equation (3) and noting that under the assumption of no event-induced variance, the variance of the scaled abnormal return is $\sigma_A^2 = (m-p-1)/(m-p-3)$, a cross-correlation adjusted scaled t-ratio (hereafter ADJ-PATELL) can be defined as

$$t_{AP} = \frac{\bar{A}\sqrt{n}}{\sqrt{(m-p-1)/(m-p-3)}\sqrt{1+(n-1)\bar{r}}}, \tag{13}$$

where again $\bar{r}$ is the average of the sample correlations of estimation-period residuals. The PATELL statistic can be adjusted as follows: $t_{AP} = t_P/\sqrt{1+(n-1)\bar{r}}$. Statistic (13) is essentially the same as Brown and Warner's (1985) portfolio method t-statistic with scaled residual returns. As before, the correction factor $1/\sqrt{1+(n-1)\bar{r}}$ is unchanged in CAR tests.

Implicitly, the PATELL test assumes that scaled abnormal returns have the same variance, while they may change in the BMP test due to the event effect. If there is no volatility effect due to the event, all scaled abnormal returns would have roughly a unit variance and lead to the PATELL test. If there is a volatility effect, the BMP test adjusts for event-induced variance inflation by estimating the average event-day-volatility cross-sectionally with the usual sample standard deviation. However, when the event day is the same for all firms, the scaled abnormal returns are potentially correlated, which implies a need for further adjustments to the volatility in both tests to correct for potential bias.[5]

Table 2 demonstrates numerically the inflation effects of cross-correlation. Even with low average correlation, problems begin to emerge already at a small sample size of $n = 10$ firms. Based on an average correlation of only 0.05, the true rejection probabilities at the nominal 5% level in two-sided tests in a sample of 10 firms are 0.11 for the BMP test and 0.10 for the PATELL test. In a larger sample of 100 firms, the true rejection probabilities with average correlation of 0.05 exceed 0.40. That is, if cross-correlation is not taken into account, instead of a 5% rejection rate, the true Type I error would be higher than 40%.

It is evident that cross-correlation can seriously bias rejection rates. However, even if a testing procedure properly accounts the correlation, it is important to seek abnormal return methods that reduce the cross-correlation to a minimum. This is because cross-correlation always weakens a testing procedure by decreasing power, as correlation implies overlapping information. Due to correlation, unlike the case of independence, the amount of information does

[5] The intuition is that the distributions of the test statistics of the PATELL and BMP under the null hypothesis are not what they are supposed to be, i.e., approximately N(0,1). With cross-correlation, for example, the PATELL statistic is approximately distributed as $N(0, 1+(n-1)\bar{\rho})$, which implies that under the null hypothesis, the variance of the test statistic increases when $\bar{\rho} > 0$. This means that more probability mass lies farther away in the tails as the sample size increases and, hence, the rejection rate (or probability that the test statistic exceeds, say, the 5% threshold 1.96) increases as the sample size increases. If $\bar{\rho} = 0$, the null distribution is again $N(0, 1)$ for all sample sizes, and no size problems emerge.

Table 2
True rejection probabilities at the nominal 5% level for two-tailed BMP and PATELL *t*-tests of average abnormal returns when residuals are cross-sectionally correlated

Number of	Average correlation ($\bar{\rho}$)					
firms (n)	0.00	0.01	0.05	0.10	0.15	0.20
			Panel A. BMP Test			
5	0.05	0.06	0.08	0.12	0.15	0.19
10	0.05	0.06	0.11	0.18	0.24	0.29
20	0.05	0.07	0.17	0.27	0.36	0.42
30	0.05	0.09	0.22	0.35	0.43	0.50
50	0.05	0.11	0.30	0.44	0.53	0.59
100	0.05	0.17	0.43	0.57	0.65	0.70
200	0.05	0.26	0.56	0.68	0.74	0.78
			Panel B. PATELL Test			
5	0.05	0.05	0.07	0.10	0.12	0.14
10	0.05	0.06	0.10	0.16	0.20	0.24
20	0.05	0.07	0.16	0.25	0.32	0.37
30	0.05	0.08	0.21	0.32	0.40	0.45
50	0.05	0.11	0.29	0.42	0.50	0.55
100	0.05	0.16	0.42	0.55	0.62	0.67
200	0.05	0.26	0.55	0.67	0.72	0.76

not increase at the same rate as the number of observations. This can be easily demonstrated in terms of our otherwise powerful new test statistics ADJ-BMP and ADJ-PATELL, defined in Equations (11) and (13), respectively. For the sake of simplicity, we use the ADJ-PATELL statistic. Let $E[AR_j] = \mu_j$ denote the expected abnormal return of the firm j on the event day, such that the null hypothesis is $\mu_j = 0$. For convenience, assume that the error standard deviations σ_j and the average cross-sectional correlation $\bar{\rho}$ between the return series are known. As such, the scaled abnormal returns have unit variances. The ADJ-PATELL statistic in Equation (13) can be written after some straightforward algebra as

$$t_{AP} = z + \frac{\sqrt{n}}{\sqrt{1+(n-1)\bar{\rho}}}\left(\frac{1}{n}\sum_{j=1}^{n}\frac{\mu_j}{\sigma_j}\right), \tag{14}$$

where $z = (\bar{A} - \bar{\mu})\sqrt{n}/\sqrt{1-(n-1)\bar{\rho}}$ is a standard normal random variable with $\bar{\mu} = \sum_{j=1}^{n} \mu_j/n$ the average of the population abnormal returns of the n firms. After simplification,[6] Equation (14) reduces to

$$t_{AP} = z + a_n \bar{\mu}\bar{\nu}, \tag{15}$$

where $a_n = \sqrt{n}/\sqrt{1+(n-1)\bar{\rho}}$, and $\bar{\nu} = \sum_j \frac{\nu_j}{n}$ with $\nu_j = \frac{1}{\sigma_j}$. Hence, the power function for the ADJ-PATELL test at the significance level α in a

[6] Denoting $\nu_j = 1/\sigma_j$, and using properties of the sample covariance, or $cov(x, y) = \sum(x_j - \bar{x})(y_j - \bar{y})/n = \sum x_j y_j/n - \bar{x}\,\bar{y}$, the right-hand term in parentheses in Equation (14) can be written as $\sum \mu_j \nu_j/n = cov(\mu, \nu) + \bar{\mu}\bar{\nu}$. Note that $\bar{\nu}$ is the inverse of the harmonic mean of the volatilities σ_j. If the event does not affect the volatilities σ_j, then $cov(\mu, \nu) = 0$, which is particularly true when the event generates a constant mean effect, such that $\mu_j = \mu$ for all j.

one-tailed test with the alternative hypothesis $H_1 : \mu > 0$ can be defined as

$$\pi_\alpha(\bar{\mu}) = \Phi(z_{1-\alpha} + a_n \bar{\mu}\bar{\nu}), \tag{16}$$

where $z_{1-\alpha}$ is the $1-\alpha$ fractile in the standard normal distribution. Using similar derivations for the ADJ-BMP statistic, we can rewrite the formula in Equation (11) as

$$t_{AB} = z + b_n \bar{\mu}\bar{\nu}, \tag{17}$$

where z is a standard normal random variable, $b_n = (\sqrt{n}/\sigma_A)\sqrt{(1-\bar{\rho})/(1+(n-1)\bar{\rho})}$, and the power function is again of the form of Equation (16). It is immediately obvious that both a_n and b_n decrease as the average correlation increases, which by Equation (16) leads to reduction of the power of the tests. Hence, it is advantageous to reduce cross-correlation in abnormal returns to a minimum. This can be achieved by using factor models, such as the FF MODEL, that extract as much as possible of the common cross-sectional correlation from the residuals. By contrast, MEAN adjusted returns should be avoided due to higher cross-sectional correlation (Brown and Warner 1985). Nonetheless, as demonstrated by Table 2, even a small residual correlation should not be overlooked by using statistics that are based on the assumption of cross-sectional independence.

3.2 Other test statistics

Here we overview non-scaled and nonparametric event study tests. First, the cross-sectional t-statistic (UNADJ) is

$$t_{cs} = \frac{AAR\sqrt{n}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n} s_i^2(1+d_i)}}, \tag{18}$$

where AAR is the average (non-scaled) abnormal return, s_i^2 is the regression model residual variance of security i, and d_i is the sampling error component similar to that in Equation (5). Second, based on the portfolio method, the t-statistic (PORT) is

$$t_{pf} = \frac{AR_0}{s\sqrt{1+d}}, \tag{19}$$

where AR_0 is the (equally weighted) portfolio abnormal return on the event day ($t = 0$), and

$$s = \sqrt{\frac{1}{m-p}\sum_{t=t_1}^{t_2-1}(R_t - \hat{R}_t)^2}, \tag{20}$$

where $R_t = \frac{1}{n}\sum_{i=1}^{n} R_{i,t}$ is the equally weighted portfolio return of n securities, $\hat{R}_t$ is the prediction of the portfolio return, m is the number of observations in the estimation period, p is the number of estimated parameters, t_1 is the start of the estimation period, $t_2 - 1$ is the end of the estimation period (t_2 is the start of the event period), and d again is the sampling error component. The PORT test implicitly accounts for cross-sectional correlation; however, if event-induced volatility is present, this method becomes misspecified. Under the assumption of constant event-induced volatility, a simple correction is to use the feasible cross-sectional variance estimator in Equation (9) of the BMP statistic, such that an event-induced-volatility-robust standard error of the event-day abnormal return AR_0 of the equally weighted portfolio is $s_A s\sqrt{1+d}$. Thus, a volatility robust version for the PORT test is

$$t_{Apf} = \frac{AR_0}{s_A s\sqrt{1+d}} = \frac{t_{pf}}{s_A}. \tag{21}$$

The derivation of this result is straightforward and is available from the authors on request. Due to its standard application in the case of clustered event days, we use the PORT test in forthcoming simulations. Third, and last, we report the results of Corrado and Zivney's (1992) modification of the nonparametric rank test (RANK) of Corrado (1989):

$$t_r = \frac{(\bar{U}_0 - 1/2)\sqrt{n}}{s_u}, \tag{22}$$

where

$$s_u = \sqrt{\frac{1}{M}\sum_{t=t_1}^{t_3} n_t(\bar{U}_t - 1/2)^2}, \tag{23}$$

such that $s_u/\sqrt{n}$ is the standard error of $\bar{U}_0$, $\bar{U}_t = \frac{1}{n_t}\sum_{i=1}^{n_t} U_{it}$ are the average scaled (standardized) ranks for $t = t_1, t_1+1, \ldots, t_3$ with t_3 the end of the event period, M is the number of observations in the combined estimation and event period, and $t = 0$ is the event day. The scaled ranks are defined as

$$U_{it} = \text{rank}(A^*_{it})/(M_i + 1), \tag{24}$$

where M_i is the number of non-missing returns in the combined estimation and event period for the ith series, and

$$A^*_{it} = \begin{cases} A_{it}, & \text{if } t = t_1, \ldots, -1, +1, \ldots, t_3 \\ A_{i0}/s, & \text{if } t = 0, \end{cases} \tag{25}$$

where A_{it} is the scaled factor model residual for $t \neq 0$, and A_{i0}/s is the event-day scaled residual rescaled with the event-day cross-sectional standard deviation, s, defined in Equation (7).[7] The RANK test implicitly accounts for cross-sectional correlation, as the standard deviation s_u is based on the averages $\bar{U}_t$ that are equally weighted portfolios of the scaled ranks. In addition, unlike the original Corrado (1989) rank test, the Corrado and Zivney (1992) RANK test is robust to event-induced volatility due to the rescaling of the event-day return in Equation (25) by the cross-sectional standard deviation.

3.3 Asymptotic distributions of the test statistics

In this section, we briefly discuss asymptotic distributions (i.e., large sample properties) of the test statistics accounting for cross-correlation, including PORT, ADJ-PATELL, ADJ-BMP, and RANK. Online Appendix B provides the technical derivations.

Generally, in event studies, sample information can be expanded by increasing the length of the estimation period and the number of return series. Under the assumption of multivariate normality stated in Section 2, scaled returns converge in probability to standard normal random variables as the length of the estimation period increases, which implies that the ADJ-PATELL and PORT statistics are asymptotically normally distributed. The ADJ-BMP and RANK tests are dependent on cross-sectional information to adjust for event-induced volatility. In our case, the main concern is cross-sectional correlation. According to Lehmann (1999, p. 107), the crucial point in the limiting normality of the average abnormal return $\bar{A}$ is that $\lim_{n\to\infty} var[\sqrt{n}\bar{A}]$ is finite. In the present context, this implies that $n\bar{\rho}_n \to \gamma$, where γ is some finite constant, and the subscript n in the average (residual) cross-correlation indicates that it depends on the number of observations. Thus, the average correlation must converge to zero at the rate n. Given residual cross-correlations after extracting common factors from the returns, one plausible assumption in financial markets is that residual returns are zero between industries and the number of industries grows with the number of firms. As such, convergence of the mean residual correlation to zero is readily achieved. Within one industry, asymptotic normality requires that the factor model of abnormal returns captures the common correlations such that the residual cross-correlation vanishes in the limit as the number of firms increases.

3.4 Estimation of the average cross-correlation from the event period

The above test statistics typically assume that the model parameters used to estimate abnormal returns do not change from the estimation period to the

[7] In the case of the event-day rescaled return $A^*_{i0} = A_{i0}/s$, one should use the feasible estimator in Equation (9), or $s_A = s/\sqrt{1-\bar{r}}$, to account for cross-sectional correlation. However, this substitution usually does not have a material impact on the results, as the average correlation is relatively small and the correction is not dependent on the number of cross-section observations like the correction terms in the ADJ-PATELL and ADJ-BMP t-statistics.

event period. In the BMP test, variance changes are mitigated by estimating the cross-sectional variances on the event days. In the same manner, if the cross-correlations change (Bittlingmayer and Hazlett 2000), or estimation period observations are not available, it is possible to compute the average correlation from the event-period returns provided that the event period is sufficiently long to enable estimation. The fact that only the mean cross-correlation is needed in our procedures alleviates the latter problem substantially, as estimation errors in individual correlations can again be expected to average out in the mean. For comparative purposes, simulations with MARKET adjusted returns use only the event period to estimate the individual correlations.

4. Simulation Design

Simulation procedures with real returns are used to study and compare empirically Type I and Type II errors for different test statistics and adjusted return models with event-date clustering. We follow the design setup in Brown and Warner (1985) and construct 1,000 independently drawn portfolios of sizes $n = 50$, 30, or 10 securities. To conserve space, the results for only $n = 50$ are reported in most cases. Online Appendix C reports further simulations focusing on a single industry, where cross-correlations can be expected to be higher than in the marketwide case. Simulations for both the whole market and a single industry are conducted for CRSP daily returns from January 1990 to December 2005. Two-sided tests are reported for different abnormal returns, including FF MODEL, OLS MODEL, and MARKET[8] adjusted returns defined in Table 1. These models extract in decreasing order common cross-sectional correlation from the abnormal returns and thereby allow us to investigate the robustness of Type I and Type II errors of the test statistics to different levels of cross-correlation. Online Appendix C augments the latter model with industry index excess returns (FF INDUSTRY MODEL) to further reduce the residual cross-correlations in single industry simulations. Daily value-weighted CRSP returns are used to proxy the market portfolio, and excess returns are computed using one-month Treasury bill rates.

In each round of simulation, a common randomly drawn event day is selected, which is set as date "0," and then a sample of n securities is selected *without* replacement. The event period is (-10 to $+10$) days around the event day. Similar to earlier simulation studies by Brown and Warner (1985) and

[8] In some cases, there may not be enough observations for estimation of the factor loadings in the FF MODEL. In this situation, a common practice is to use MARKET adjusted returns. However, average cross-correlations as well as standard deviations must be estimated. A drawback of this approach is difficulty in accurately estimating the standard deviations of returns, which are needed to calculate the scaled returns. If only the event-period data are available, there is no other option than to use this period to estimate the standard deviations and cross-correlations. We therefore investigate this situation by reporting simulation results in which only the event period is used with MARKET adjusted returns to estimate the needed parameters.

Boehmer, Musumeci, and Poulsen (1991), each security in the sample must have at least 50 returns in the estimation period (-249 to -11) and no missing returns in the last 30 days (-19 to $+10$).

Type I errors (i.e., rejecting the null when it is true) are investigated for different magnitudes of event-induced volatility by multiplying the day 0 residual returns by a factor $\sqrt{c}$. Charest (1978), Mikkelson (1981), Penman (1982), and Rosenstein and Wyatt (1990) have generally found that the event-period standard deviation is about 1.2 to 1.5 times the estimation-period standard deviation. One choice would be to select constant factors; but, to add more realism, we use a random factor such that, given the abnormal return, the effect of the volatility increment can be investigated by drawing c for each security from uniform distributions $U[1, 2]$, $U[1.5, 2.5]$, and $U[2.5, 3.5]$, respectively. On average, these represent increments in the variances and covariances by factors of 1.5, 2.0, and 3.0, respectively, and therefore increase standard deviations by factors on average from $\sqrt{1.5} \approx 1.2$ to $\sqrt{3} \approx 1.7$, which covers the range cited in the above papers. No event-induced volatility corresponds to a fixed $c = 1$. This procedure increases the volatility of each return on the event day by a random amount, holding the correlations unchanged.

Additionally, Type II errors (i.e., failing to reject the null when the alternative is true) are examined in terms of the powers of the tests by incrementally adding (subtracting) a constant ranging from 0% to ±3% to the day 0 abnormal return and estimating power functions for different adjusted-return methods. These event-induced mean effects are investigated with and without variance inflation.

Finally, size and power properties of the cross-correlation adjusted ADJ-PATELL and ADJ-BMP tests as well as RANK tests are examined for CARs of CAR(-1, $+1$), CAR(-2, $+2$), CAR(-5, $+5$), and CAR(-10, $+10$) (i.e., windows of lengths 3, 5, 11, and 21 days, respectively). Abnormal returns are randomly assigned to a single day within the CAR window. The estimation period is the same in each case. For the RANK test, as discussed in Corrado (1989, p. 395) and Campbell and Wasley (1993, footnote 4), the estimation period and event period are divided into multiple-day intervals to compute the needed returns. For example, in the case of CAR(-5, $+5$), we divide the 239-day estimation period into 11-day periods and compute 11-day returns in these periods. Since an estimation period of 239 days includes 21 full 11-day periods with 8 days left over, we discard the first 8 days. In the 21-day event period, we compute the 11-day CAR from day -5 to day $+5$, and discard days -10 to -6 and $+6$ to $+10$. Thus, altogether there are 22 multiple-day returns. Because this multiple-day procedure rapidly reduces the number of observations, we require in these experiments that each security in the sample has no missing returns in the last 150 days (i.e., in the estimation period from -139 to -11 and event period from -10 to $+10$). CARs for all CRSP securities are estimated using the FF MODEL.

5. Simulation Results

5.1 Single-event-day results

In one-day analyses of the event-day abnormal return, we investigate Types I and II error rejection rates for different test statistics. Sample statistics from 1,000 simulations for all CRSP stocks in the sample period for FF MODEL, OLS MODEL, and MARKET adjusted returns are provided in Table 3.

The overall average of the return cross-correlations in the simulations is about 0.085 compared with only 0.022 average residual cross-correlation for the FF MODEL, 0.030 for the OLS MODEL, and 0.031 for MARKET

Table 3
Marketwide (all CRSP stocks) sample statistics for event-study test statistics using different abnormal-return definitions based on 1,000 random portfolios of $n = 50$ securities with no event effect when residual returns are correlated

	Mean	Median	Std dev	Skewness	Excess kurtosis	Min	Max
Panel A. FF MODEL Adjusted Returns: $AR_{it} = r_{it}^e - \alpha_i - \beta_{im} r_{mt}^e - \beta_{i,smb} SMB_t - \beta_{i,hml} HML_t$							
Average return cross-correlation	0.085*	0.083*	0.032	0.388*	−0.239	0.014	0.192
Average residual cross-correlation	0.022*	0.003	0.038	1.627*	1.379*	−0.006	0.176
UNADJ test	−0.047	−0.059	1.365	0.064	2.444*	−5.851	6.990
PATELL test	−0.055	−0.093	1.601	−0.151	3.823*	−9.045	8.443
BMP test	−0.069	−0.091	1.588	−0.190*	2.081*	−6.752	5.737
PORT test	−0.040	−0.046	1.012	0.016	0.328*	−4.151	3.321
ADJ-PATELL test	−0.045	−0.082	1.072	−0.043	0.344*	−3.731	3.452
ADJ-BMP test	−0.055	−0.075	1.032	−0.047	−0.333*	−3.154	2.684
RANK test	−0.037	−0.066	1.041	−0.040	−0.234	−3.527	3.104
Panel B. OLS MODEL Adjusted Returns: $AR_{it} = r_{it}^e - \alpha_i - \beta_{im} r_{mt}^e$							
Average return cross-correlation	0.083*	0.081*	0.031	0.352*	−0.449*	0.026	0.180
Average residual cross-correlation	0.030*	0.015*	0.033	1.654*	1.566*	0.000	0.157
UNADJ test	−0.046	0.014	1.426	−0.090	1.413*	−6.158	6.736
PATELL test	−0.068	−0.016	1.655	0.013	3.355*	−8.277	10.771
BMP test	−0.100	−0.017	1.684	−0.273*	1.801*	−7.639	7.424
PORT test	−0.019	0.001	1.019	−0.001	0.949*	−4.438	3.563
ADJ-PATELL test	−0.033	−0.010	1.064	0.061	1.008*	−3.700	4.742
ADJ-BMP test	−0.054	−0.012	1.055	−0.303*	0.717*	−5.457	3.127
RANK test	−0.036	−0.009	1.074	−0.214*	0.240	−4.624	2.987
Panel C. MARKET Adjusted Returns: $AR_{it} = r_{it} - r_{mt}$							
Average return cross-correlation	0.064*	0.052*	0.050	1.629*	3.416*	−0.006	0.331
Average residual cross-correlation	0.031*	0.015*	0.059	4.308*	21.663*	−0.011	0.490
UNADJ test	−0.059	−0.026	1.268	0.079	1.666*	−5.942	6.211
PATELL test	−0.074	−0.079	1.458	0.096	2.620*	−6.499	8.157
BMP test	−0.085	−0.093	1.694	−0.137	3.107*	−9.401	8.095
PORT test	−0.053	−0.028	1.006	0.066	0.576*	−3.402	4.063
ADJ-PATELL test	−0.045	−0.053	0.960	0.077	0.671*	−3.045	3.483
ADJ-BMP test	−0.047	−0.061	1.067	0.040	0.493*	−3.719	4.433
RANK test	−0.043	−0.012	0.948	0.020	−0.294	−2.892	2.890

The sample period covers January 3, 1990, through December 31, 2005, with daily returns for all CRSP stocks (i.e., a total of 17,878 return series). The abnormal return models are provided in Table 1. Notation is defined as follows: r_{it}^e is the excess return of stock i, r_{mt}^e is the value-weighted CRSP market excess return, r_{it} and r_{mt} are the corresponding raw returns, *SMB* is the small-minus-big market capitalization factor, and *HML* is the high-minus-low book equity/market equity factor (see Kenneth French's website). Asterisks indicate significant differences from zero at the 5% level or smaller. In Panels A and B, the parameters are estimated from the 239-day estimation period. In Panel C, only the event-period (21 days) observations are used in the estimation.

adjusted returns. Hence, extraction of the Fama-French factors reduces the cross-correlation by almost 75%, while the other models reduce it by about 65%. Unlike the average return cross-correlations, the distribution of the average residual cross-correlations in the simulations is highly positively skewed, especially for the FF MODEL, in which the median is only about 0.003 (i.e., one-seventh of the mean), while the median for the return cross-correlations is about 0.083. For all abnormal return definitions, the average t-statistics are close to zero, as expected under the null hypothesis of no event effect being imposed. An important feature of these results is that, although the average residual correlation is fairly low, its effect is quite dramatic on the distributional properties of the unadjusted t-statistics. The standard deviations of the UNADJ, PATELL, and BMP t-statistics are from 1.4 to 1.7 times the theoretical value of one. The average standard deviations of the test statistics accounting for cross-correlation, including the PORT, ADJ-PATELL, ADJ-BMP, and RANK tests, are in line with the theoretical value of unity. However, the distributions of these statistics, like the other tests, exhibit some excess kurtosis.[9]

5.1.1 Type I error rates. Rejection rates for the test statistics under the null hypothesis of no event effect based on different abnormal return definitions, including FF MODEL, OLS MODEL, and MARKET adjusted returns, are reported in Table 4. The models extract common cross-correlation in descending order. The rejection rates indicate the fractions by which the test statistics exceed in 1,000 simulations the nominal cutoffs at the 5% level (i.e., 1.96 in absolute value for two-tailed tests). For example, with no event-induced variance after extracting the three Fama-French factors from returns in Panel A, the estimated over-rejection rates are 0.124, 0.156, and 0.153 for the UNADJ, PATELL, and BMP tests, respectively. By contrast, the rejection rates for the PORT, ADJ-PATELL, ADJ-BMP, and RANK tests are close to the nominal rate of 0.05 when no event-induced variance ($c = 1$) is present (i.e., in 1,000 simulations, the rejection rates are within the 95% confidence interval of [0.036, 0.064]). These results demonstrate the importance of accounting for cross-correlation to avoid material over-rejection of the null hypothesis. Event-induced variance increases the over-rejection rates for the UNADJ, PATELL, and BMP statistics as well as the PORT and ADJ-PATELL statistics (see columns 3, 4, and 5 in Panel A). Importantly, the ADJ-BMP and RANK tests have rejection rates close to the nominal rate and, hence, are robust with respect to both cross-correlation and event-induced variance. The results for OLS MODEL and MARKET adjusted returns in Panels B and C of Table 4 confirm these findings.

[9] The 5% (large sample) threshold in 1,000 simulations for skewness is $\sqrt{6 \times \text{chi}^{-1}(0.05, 1)/1000} = 0.152$ and for kurtosis is $\sqrt{24 \times \text{chi}^{-1}(0.05, 1)/1000} = 0.304$, where $\text{chi}^{-1}(\cdot)$ is the inverse chi-square distribution function.

Table 4
Marketwide (all CRSP stocks) two-tailed average rejection rates for different test statistics at the 5% significance level for the null hypothesis of no event mean effect in the presence of event-induced variance-covariance based on 1,000 random portfolios of $n = 50$

	Average event-induced variance-covariance factor c			
	$c = 1.0$	$c = 1.5$	$c = 2.0$	$c = 3.0$
Panel A. FF MODEL: $AR_{it} = r^e_{it} - \alpha_i - \beta_{i,m} r^e_{mt} - \beta_{i,smb} SMB_t - \beta_{i,hml} HML_t$ (averge residual cross-correlation of 0.022)				
UNADJ	0.124	0.190	0.238	0.343
PATELL	0.156	0.217	0.271	0.375
BMP	0.153	0.150	0.150	0.156
PORT	0.055	0.105	0.159	0.255
ADJ-PATELL	0.063	0.124	0.174	0.285
ADJ-BMP	0.051	0.047	0.047	0.051
RANK	0.055	0.053	0.056	0.055
Panel B. OLS MODEL: $AR_{it} = r^e_{it} - \alpha_i - \beta_{im} r^e_{mt}$ (average residual cross-correlation of 0.030)				
UNADJ	0.147	0.226	0.282	0.385
PATELL	0.177	0.255	0.320	0.406
BMP	0.206	0.199	0.206	0.203
PORT	0.061	0.111	0.156	0.239
ADJ-PATELL	0.065	0.120	0.174	0.249
ADJ-BMP	0.052	0.050	0.055	0.055
RANK	0.062	0.063	0.065	0.062
Panel C. MARKET Adjusted Returns: $AR_{it} = r_{it} - r_{mt}$ (avg residual cross-correlation of 0.031)				
UNADJ	0.105	0.174	0.232	0.320
PATELL	0.146	0.207	0.257	0.318
BMP	0.197	0.200	0.205	0.201
PORT	0.055	0.104	0.156	0.249
ADJ-PATELL	0.043	0.083	0.118	0.165
ADJ-BMP	0.064	0.063	0.063	0.068
RANK	0.034	0.036	0.034	0.034

The sample period covers January 3, 1990, through December 31, 2005, with daily returns for all CRSP stocks (total of 17,878 return series). Different adjusted returns are defined in the table, where r^e_{it} is the excess return of stock i, r^e_{mt} is the value-weighted market excess return, *SMB* is the small-minus-big market capitalization factor, and *HML* is the high-minus-low book equity/market equity factor (see Kenneth French's website). The rejection rates indicate the fractions by which the test statistics exceed in 1,000 simulations the nominal cutoff at the 5% level (i.e., 1.96 in the two-tailed test). The variances (covariances) are increased according to the magnitudes of different volatility-increasing designs. The no-volatility effect is when the factor c is a constant equal to 1. In the three other designs, each event-day (day 0) return $r_{i,0}$ is multiplied by $\sqrt{c}$, where c are random deviates drawn from the appropriate uniform distribution, $U(1, 2)$, $U(1.5, 2.5)$, or $U(2.5, 3.5)$, with means 1.5, 2.0, and 3.0, respectively, depending on the design. Thus, the highest volatility with c drawn from $U(2.5, 3.5)$ corresponds to an average variance that is 3 times the non-event variance, or $\sqrt{3} \approx 1.7$ times the non-event standard deviation. The correlations of the returns remain unchanged. With a true rejection rate of 0.05, the 95% CI for the average rates in 1,000 replicates is [0.036, 0.064]. In Panels A and B, the parameters are estimated from the 239-day estimation period. In Panel C, only the event-period (21 days) observations are used in the estimation.

5.1.2 Type II error rates. In this subsection, rejection rates for the test statistics under the alternative hypothesis of non-zero mean effects (i.e., power analyses) for FF MODEL adjusted returns are provided. So far, the evidence has indicated that the rejection rates for the PORT, ADJ-PATELL, ADJ PORT, and RANK tests are robust to cross-correlation under the null hypothesis of no event effect.

Power results for no event-induced variance in the abnormal return models of FF MODEL, OLS MODEL, MARKET, and MEAN are shown in Panels A

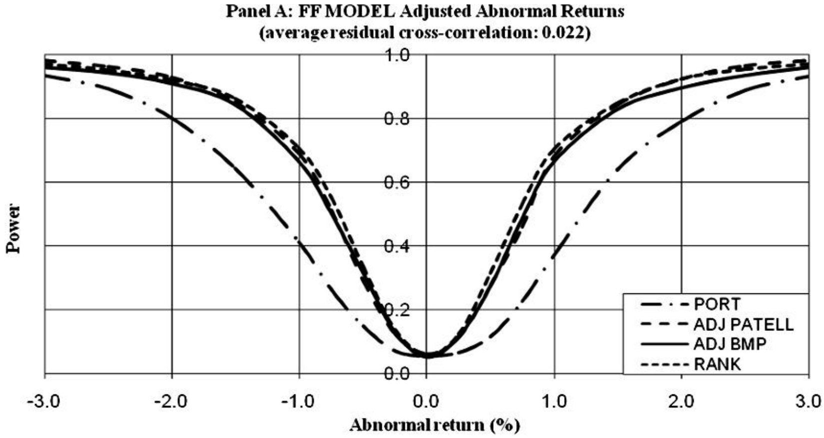
Table 5
Marketwide (all CRSP stocks) two-tailed average rejection rates at the 0.05 significance level for selected test statistics sampled from 1,000 random portfolios of $n = 50$ securities with abnormal return ranging from −3.0 to +3.0% in different abnormal-return models

Abnormal return (%)	PORT	ADJ-PATELL	ADJ-BMP	RANK
Panel A: FF MODEL (average residual cross-correlation of 0.022)				
−3.0	0.935	0.981	0.960	0.970
−2.0	0.803	0.929	0.909	0.920
−1.0	0.413	0.691	0.665	0.708
−0.5	0.148	0.292	0.315	0.334
0.0	**0.055**	**0.063**	**0.051**	**0.057**
+0.5	0.117	0.276	0.273	0.316
+1.0	0.372	0.682	0.665	0.705
+2.0	0.790	0.923	0.896	0.925
+3.0	0.931	0.981	0.959	0.972
Panel B: OLS MODEL (average residual cross-correlation of 0.030)				
−3.0	0.933	0.987	0.968	0.974
−2.0	0.729	0.926	0.890	0.912
−1.0	0.292	0.552	0.555	0.601
−0.5	0.107	0.212	0.234	0.255
0.0	**0.061**	**0.065**	**0.052**	**0.062**
+0.5	0.113	0.186	0.217	0.248
+1.0	0.277	0.522	0.528	0.609
+2.0	0.723	0.912	0.877	0.913
+3.0	0.916	0.986	0.963	0.968
Panel C: MARKET Adjusted Returns (average residual cross-correlation of 0.031)				
−3.0	0.933	0.957	0.960	0.920
−2.0	0.785	0.895	0.900	0.839
−1.0	0.327	0.546	0.606	0.487
−0.5	0.129	0.191	0.258	0.182
0.0	**0.055**	**0.043**	**0.064**	**0.034**
+0.5	0.115	0.164	0.233	0.159
+1.0	0.290	0.514	0.577	0.471
+2.0	0.762	0.870	0.886	0.831
+3.0	0.926	0.960	0.959	0.924
Panel D: MEAN Adjusted Returns (average residual cross-correlation of 0.080)				
−3.0	0.810	0.947	0.894	0.913
−2.0	0.517	0.736	0.667	0.736
−1.0	0.162	0.255	0.253	0.304
−0.5	0.082	0.101	0.100	0.128
0.0	**0.057**	**0.066**	**0.049**	**0.052**
+0.5	0.085	0.097	0.088	0.123
+1.0	0.181	0.295	0.280	0.334
+2.0	0.527	0.747	0.684	0.756
+3.0	0.872	0.942	0.889	0.926

The abnormal-return models are summarized in Table 1. In Panels A, B, and D, the parameters are estimated from the 239-day estimation period. In Panel C, only the event-period (21 days) observations are used in the estimation.

to D of Table 5. Figure 1 graphically displays the results. The zero abnormal return line (boldface) in each panel of the table indicates the Type I error rates, which, as noted already above, are all reasonably close to the nominal error rate of 0.05. In general, the results demonstrate that, regardless of the abnormal-return model, all test statistics have good power to detect large abnormal returns of ±3%. As the event-induced mean effect is reduced, the PORT test is substantially less powerful than the other tests, which confirms this well-known shortfall. Also, the RANK test is slightly more powerful than the

ADJ-PATELL and ADJ-BMP tests, which detect the false null hypothesis at about the same rate. Figure 1 shows that, consistent with the theory in Section 3.1, cross-correlation rapidly weakens the powers of the tests. As shown in Panel A of Figure 1, the highest power in each abnormal return level is attained when the correlation is smallest (i.e., 0.022) using FF MODEL adjusted returns. OLS MODEL and MARKET adjusted return results in Panels B and C have virtually the same cross-correlations (i.e., about 0.030) and, hence, produce similar powers for the test statistics. MEAN adjusted returns have the highest cross-correlation (i.e., equal to the return cross-correlation of 0.080), which leads to substantial power reduction. Since MARKET adjusted returns in Panel C use 21 versus 239 daily observations for MEAN adjusted returns in Panel D, we infer that cross-correlation explains most of this power reduction rather than the number of time-series observations.



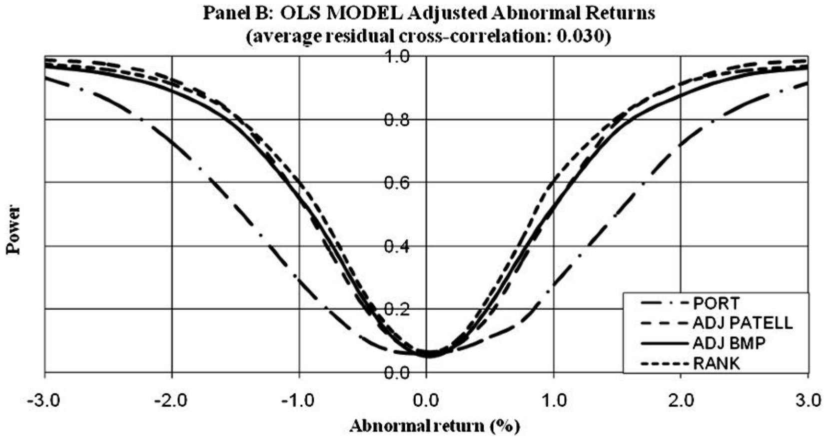
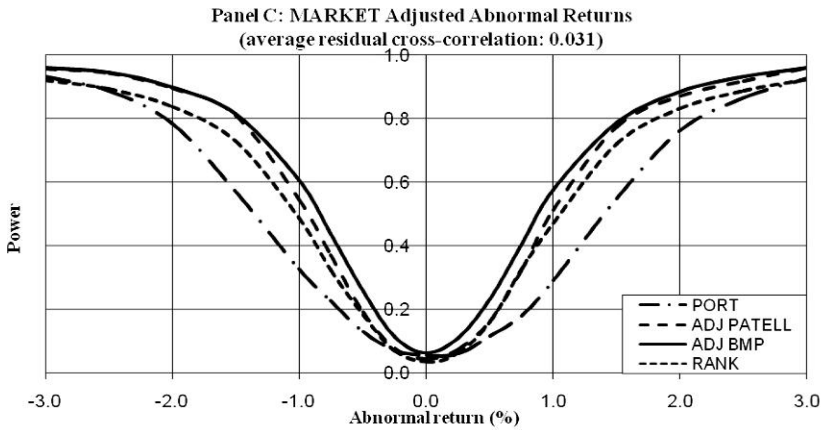


Figure 1
Continued



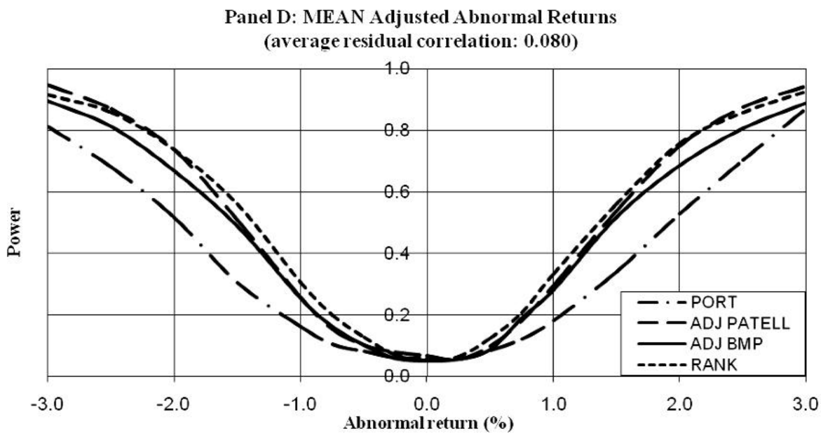


Figure 1

Estimated power functions with different abnormal-return definitions for PORT, ADJ-PATELL, ADJ-BMP, and RANK tests based on 1,000 samples of $n = 50$ security portfolios from the CRSP database: Two-sided tests, significance level 0.05, and no event-induced variance

The sample period covers January 3, 1990, through December 31, 2005, with daily returns for all CRSP stocks (i.e., a total of 17,878 return series). The abnormal returns are generated by adding a constant ranging from 0% to 3.0% to the abnormal returns. Panel A contains the results for FF MODEL adjusted returns: $AR_{it} = r^e_{it} - \alpha_i - \beta_{im}r^e_{mt} - \beta_{i,smb}SMB_t - \beta_{i,hml}HML_t$, where r^e_{it} is the excess return of stock i, r^e_{mt} is the value-weighted market excess return, *SMB* is the small-minus-big market capitalization factor, and *HML* is the high-minus-low book equity/market equity factor (see Kenneth French's website). Panel B employs OLS MODEL adjusted returns: $AR_{it} = r^e_{it} - \alpha_i - \beta_{im}r^e_{mt}$. Panel C, utilizes MARKET adjusted returns: $AR_{it} = r_{it} - r_{mt}$. Panel D employs MEAN adjusted returns: $AR_{it} = r_{it} - \bar{r}_i$, where r_{it} is the stock return, and $\bar{r}_i$ is the sample mean of the return in the estimation period. In Panel C, the needed parameters are estimated from the 21-day event period, while for the other panels the estimation period contains 239 days prior to the event period.

In sum, these results show that it is important to account for even small cross-correlation to prevent over-rejection. Also, it is advantageous to use abnormal-return definitions that reduce cross-correlation to a minimum to maximize the power of the test statistics.

In unreported results, we repeated the power analyses with different levels of event-induced volatility inflation. The most relevant results are those for the ADJ-BMP and RANK tests, as they prove to be robust to event-induced volatility. Increasing volatility exacerbates detection of the mean event effect, which showed up in the results as lower power at all levels of abnormal returns. However, even when there is high event-induced volatility inflation, the ADJ-BMP and RANK tests have power similar to the PORT test under no event-induced volatility inflation. Again the RANK test tends to have slightly higher power than the ADJ-BMP test. This result might seem surprising, as observations containing magnitude information are replaced by their corresponding rank number with only ordering information. Nevertheless, consistent with statistical literature cited in Section 1, relative ordering may be an advantage in certain circumstances, such as fat-tailed distributions. For instance, under the null hypothesis of no event effect, the ordering should be symmetric in the sense that the rank number of the event-day returns will be equally likely above or below the average rank, even in the case when the original distribution of the observations is asymmetric. Relevant to financial applications, high-frequency stock returns are known to have fat-tailed distributions. In such circumstances, using outlier-insensitive ranking information for testing the location parameter may be an advantage over parametric t-tests.[10]

We infer from these results that, due to their robustness with respect to both cross-correlation and event-induced volatility, the ADJ-BMP and RANK tests are recommended in single-event-day testing. Moreover, these tests have superior power properties compared with the commonly used PORT test.

5.2 Cumulative abnormal returns

Most event studies examine CAR behavior over multiple days around the event date. CARs allow for the possibility that the event date is not exactly known and capture post-event market reaction. This section reports rejection rates and power results for testing CARs using the most efficient statistics ADJ-PATELL, ADJ-BMP, and RANK. We use simulations with FF MODEL adjusted abnormal returns for samples of $n = 50$ stocks.

As discussed in Section 3.1, return cross-correlations remain invariant under accumulation. Thus, testing CARs is straightforward with the ADJ-PATELL and ADJ-BMP statisitcs. The ADJ-PATELL statistic t_{AP} defined in Equation (13) is adapted for testing CARs by replacing the average standardized abnormal return with the average SCAR (Campbell, Lo, and MacKinlay 1997, p. 162). In the same manner, the ADJ-BMP statistic t_{AB} defined in Equation (11) for testing CARs is obtained by replacing the standardized average

[10] The potential power loss of such t-tests in fat-tailed outlier-prone distributions is due to the sensitivity of the standard deviation to outlying observations that magnify the standard deviation in the denominator of the t-statistic. A higher standard deviation deflates the t-value and thereby decreases the power of the test. Our empirical results are consistent with simulation results in Corrado (1989) and Corrado and Zivney (1992).

abnormal return with the average SCAR scaled by its cross-sectional standard deviation. Also, the RANK test is applied to ranks of multiple-day returns instead of the common practice of accumulating the single-day ranks (Campbell and Wasley 1993), which has some known issues documented (e.g., Cowan 1992).

Our simulation results reveal that average CARs become increasingly negative as the CAR window is lengthened. Consequently, the means of all test statistics in the simulations are highly statistically significant and negative under the null hypothesis of no event effect. One explanation for this might be returns' skewness. However, a more plausible explanation for the negative CARs under the null of no mean event effect stems from the systematic prediction error of log returns. Under geometric Brownian motion (GBM),

$$P_{t+dt} = P_t e^{(\mu - \frac{1}{2}\sigma^2)dt + \sigma dW_t}, \tag{26}$$

where μ and σ are the instantaneous mean return and volatility of the stock price, respectively, and W_t is the standard Brownian motion, the expected log-return (with $dt = 1$) becomes

$$E\left[r_{t+1}\right] = \mu - \frac{1}{2}\sigma^2, \tag{27}$$

where $r_{t+1} = \log(P_{t+1}/P_t)$ is the one-period log-return. Thus, predicting the instantaneous mean return μ introduces a small positive bias into the predictions. Subtraction of one-half of the instantaneous variance from the prediction should correct this bias. Note that this correction is nominal in single-day returns (e.g., for a stock with an annual volatility of 40% over 250 trading days, the correction would equal 0.032%, or only 3.2 basis points). However, as the CAR window is lengthened, the correction is increasingly important.

Applying this correction, Table 6 reports sample statistics for CARs as well as the test statistics for different CAR windows. In shorter windows of up to 5 days, the average CARs are statistically zero. In 11- and 21-day windows, the average CARs remain slightly negative, with t-values reduced tenfold from the uncorrected results. Thus, the proposed correction eliminates the negative average CAR issue for the most part.

Table 7 reports the size and power results for the ADJ-PATELL, ADJ-BMP, and RANK tests using different CAR windows with abnormal returns ranging from -3 to $+3$% and various levels of event-induced variance. The zero (cumulative) abnormal return corresponds to no event effect. Qualitatively, the results are quite similar to the single-day findings, especially for shorter CAR windows of ± 1 and ± 2. The boldface line of zero mean effect shows that the rejection rates for ADJ-BMP and RANK are fairly close to the nominal size of 0.05 even in the presence of event-induced volatility. The ADJ-PATELL test again tends to over-reject when event-induced volatility is introduced. The power of RANK and ADJ-BMP tests is fairly similar for CAR windows up to

Table 6
CAR window results with marketwide (all CRSP stocks) sample statistics for ADJ-PATELL, ADJ-BMP, and RANK tests based on 1,000 random portfolios of n = 50 securities with no event effect

						t-Values			
	Mean	Median	Std dev	Skew	Excess kurtosis	Mean	Median	Skew	Excess kurtosis
Panel A: 3-Day CAR Window (−1, +1)									
CAR (%)	0.005	0.005	1.234	0.405	5.914	0.12	0.10	5.23	38.18
ADJ-PATELL	−0.036	0.005	1.081	−0.126	0.972	−1.05	0.12	−1.63	6.27
ADJ-BMP	−0.016	0.005	1.024	−0.058	0.055	−0.50	0.12	−0.75	0.36
RANK	0.030	0.063	1.003	−0.076	0.018	0.95	1.58	−0.98	0.11
Panel B: 5-Day CAR Window (−2, +2)									
CAR (%)	−0.041	0.006	1.578	0.109	3.871	−0.83	0.09	1.41	24.99
ADJ-PATELL	−0.095	−0.024	1.071	−0.326	0.580	−2.82	−0.57	−4.21	3.75
ADJ-BMP	−0.072	−0.030	1.043	−0.168	−0.148	−2.18	−0.73	−2.16	−0.95
RANK	−0.016	0.008	1.001	−0.145	−0.291	−0.51	0.21	−1.87	−1.88
Panel C: 11-Day CAR Window (−5, +5)									
CAR (%)	−0.127	−0.078	2.352	0.140	4.993	−1.70	−0.84	1.80	32.23
ADJ-PATELL	−0.116	−0.079	1.026	−0.163	0.424	−3.56	−1.95	−2.11	2.74
ADJ-BMP	−0.103	−0.086	1.048	−0.104	0.021	−3.10	−2.08	−1.34	0.14
RANK	0.026	0.030	0.990	−0.112	−0.210	0.82	0.75	−1.45	−1.36
Panel D: 21-Day CAR Window (−10, +10)									
CAR (%)	−0.279	−0.217	3.570	0.746	8.256	−2.47	−1.54	9.63	53.29
ADJ-PATELL	−0.179	−0.124	1.048	−0.105	0.346	−5.42	−2.99	−1.35	2.24
ADJ-BMP	−0.173	−0.131	1.074	−0.104	0.140	−5.08	−3.07	−1.34	0.90
RANK	−0.035	−0.031	0.971	0.083	−0.235	−1.13	−0.80	1.07	−1.52

The sample period covers January 3, 1990, through December 31, 2005, with daily returns for all CRSP stocks (i.e., a total of 17,878 return series). Abnormal returns are computed using the FF MODEL: $AR_{it} = r^e_{it} - \alpha_i - \beta_{im} r^e_{mt} - \beta_{i,smb} SMB_t - \beta_{i,hml} HML_t$, where r^e_{it} is the excess return of stock i, r^e_{mt} is the value-weighted CRSP market excess return, *SMB* is the small-minus-big market capitalization factor, and *HML* is the high-minus-low book equity/market equity factor (see Kenneth French's website). Statistics for the RANK test are based on the respective multiple-day returns.

±2. For longer windows of ±5 (11 days) and ±10 (21 days), the ADJ-BMP tests tends to be more powerful than the RANK test. However, there may be some mild tendency for the ADJ-BMP test to over-reject the null hypothesis at the longest CAR-window of ±10 days with a p-value on the upper borderline of the 95% confidence interval. The underperformance of the RANK test with respect to the ADJ-BMP test is attributable largely to reduced multiple-day returns available for estimation. For example, in CAR(−10, +10) with a 21-day window, the original 239 estimation period reduces to only 11 eleven-day returns, which decreases accuracy of the FF-MODEL parameter estimates and increases randomness in the rankings. We infer that for shorter CAR windows, the RANK test is competitive with the ADJ-BMP test; however, as the CAR window is lengthened, the ADJ-BMP test increasingly dominates the RANK test.

5.3 Some robustness issues

As suggested by an anonymous referee, all the test statistics dealt with above rely on the assumption of homoscedastic and serially uncorrelated returns. The biasing effect in single-day-event study testing comes from the residual

Table 7
CAR window results for marketwide (all CRSP stocks) two-tailed average rejection rates for different test statistics at the 5% significance level for the null hypothesis of no event mean effect in the presence of event-induced variance-covariance based on 1,000 random portfolios of $n = 50$

	Average event-induced variance-covariance inflation factor, c											
	No event-induced variance			c = 1.5			c = 2.0			c = 3.0		
Abn return	ADJ PAT	ADJ BMP	RANK	ADJ PAT	ADJ BMP	RANK	ADJ PAT	ADJ BMP	RANK	ADJ PAT	ADJ BMP	RANK
	3-Day CAR (−1, +1)											
−3	0.911	0.893	0.915	0.895	0.833	0.854	0.884	0.789	0.807	0.870	0.698	0.702
−2	0.775	0.773	0.777	0.761	0.667	0.677	0.744	0.590	0.606	0.722	0.486	0.478
−1	0.403	0.400	0.408	0.409	0.306	0.306	0.419	0.260	0.244	0.445	0.201	0.168
0	**0.066**	**0.052**	**0.050**	**0.125**	**0.051**	**0.046**	**0.172**	**0.051**	**0.049**	**0.272**	**0.053**	**0.053**
+1	0.378	0.391	0.415	0.402	0.296	0.302	0.430	0.234	0.250	0.459	0.174	0.180
+2	0.785	0.775	0.807	0.770	0.687	0.723	0.760	0.592	0.645	0.735	0.480	0.509
+3	0.917	0.891	0.924	0.907	0.834	0.875	0.881	0.789	0.828	0.881	0.711	0.748
	5-Day CAR (−2, +2)											
−3	0.846	0.819	0.822	0.833	0.752	0.751	0.820	0.688	0.680	0.794	0.578	0.561
−2	0.665	0.656	0.649	0.650	0.546	0.528	0.641	0.463	0.435	0.635	0.366	0.331
−1	0.297	0.300	0.267	0.325	0.224	0.201	0.352	0.189	0.158	0.404	0.155	0.125
0	**0.066**	**0.057**	**0.042**	**0.124**	**0.055**	**0.041**	**0.185**	**0.055**	**0.045**	**0.276**	**0.056**	**0.045**
+1	0.238	0.269	0.262	0.288	0.200	0.187	0.334	0.149	0.140	0.395	0.124	0.107
+2	0.628	0.631	0.650	0.615	0.514	0.521	0.615	0.422	0.438	0.602	0.326	0.327
+3	0.836	0.817	0.843	0.828	0.730	0.759	0.790	0.666	0.677	0.790	0.546	0.558
	11-Day CAR (−5, +5)											
−3	0.678	0.678	0.573	0.674	0.572	0.428	0.663	0.485	0.345	0.646	0.373	0.263
−2	0.443	0.443	0.323	0.458	0.341	0.250	0.472	0.279	0.203	0.483	0.215	0.152
−1	0.184	0.184	0.128	0.237	0.161	0.098	0.280	0.140	0.084	0.334	0.111	0.070
0	**0.060**	**0.060**	**0.046**	**0.119**	**0.057**	**0.046**	**0.166**	**0.059**	**0.047**	**0.260**	**0.057**	**0.046**
+1	0.123	0.123	0.119	0.176	0.117	0.090	0.237	0.099	0.077	0.311	0.086	0.067
+2	0.365	0.365	0.355	0.383	0.288	0.259	0.403	0.239	0.202	0.436	0.179	0.147
+3	0.632	0.632	0.608	0.605	0.503	0.475	0.599	0.413	0.380	0.599	0.320	0.286
	21-Day CAR (−10, +10)											
−3	0.488	0.505	0.301	0.500	0.417	0.225	0.516	0.350	0.189	0.536	0.274	0.135
−2	0.310	0.332	0.167	0.353	0.257	0.132	0.385	0.226	0.106	0.426	0.183	0.083
−1	0.141	0.161	0.072	0.202	0.133	0.062	0.262	0.121	0.056	0.327	0.103	0.051
0	**0.071**	**0.065**	**0.042**	**0.128**	**0.066**	**0.041**	**0.178**	**0.065**	**0.042**	**0.263**	**0.064**	**0.042**
+1	0.076	0.093	0.064	0.135	0.082	0.053	0.191	0.068	0.050	0.273	0.067	0.043
+2	0.203	0.238	0.145	0.248	0.177	0.110	0.280	0.149	0.092	0.340	0.106	0.068
+3	0.382	0.414	0.286	0.397	0.322	0.201	0.442	0.265	0.157	0.442	0.197	0.120

The sample period covers January 3, 1990, through December 31, 2005, with daily returns for all CRSP stocks (total of 17,878 return series). Abnormal returns are computed using the FF MODEL: $AR_{it} = r^e_{it} - \alpha_i - \beta_{im} r^e_{mt} - \beta_{i,smb} SMB_t - \beta_{i,hml} HML_t$, where r^e_{it} is the excess return of stock i, r^e_{mt} is the value-weighted CRSP market excess return, *SMB* is the small-minus-big market capitalization factor, and *HML* is the high-minus-low book equity/market equity factor (see Kenneth French's website). Statistics for RANK test are based on the respective multiple-day returns. Rejection rates indicate the fractions by which the test statistics exceed in 1,000 simulations the nominal cutoff at the 5% level (i.e., 1.96 in the two-tailed test). The variances (covariances) are increased according to the magnitudes of different volatility-increasing designs. In the first design, no volatility effect is imposed. In the three other designs, each CAR is multiplied by $\sqrt{c}$, where c are random deviates drawn from the appropriate uniform distribution, $U(1, 2)$, $U(1.5, 2.5)$, or $U(2.5, 3.5)$, with means 1.5, 2.0, and 3.0, respectively, depending on the design. Thus, the highest volatility with c drawn from $U(2.5, 3.5)$ corresponds to an average variance that is 3 times the non-event variance, or $\sqrt{3} \approx 1.7$ times the non-event standard deviation. The correlations of the returns remain unchanged. With a true rejection rate of 0.05, the 95% CI for the average rates in 1,000 replicates is [0.036, 0.064].

standard deviation estimated from the estimation period. In this regard, the biasing effect of a small-order serial correlation is negligible in stock returns, which is confirmed by our simulation results for all test statistics that account for cross-correlation (viz., PORT, ADJ-PATELL, ADJ-BMP, and RANK). However, under accumulation of returns, autocorrelations add up and may affect the PORT and ADJ-PATELL statistics (Salinger 1992). This problem is substantially alleviated in ADJ-BMP due to rescaling with the cross-sectional standard deviation. Rescaling implies that the same-order autocorrelation in the cumulated returns of the numerator of the ADJ-BMP statistic becomes a part of the denominator also, thereby canceling out (Kolari and Pynnonen 2008, footnote 4 and remark 3). The multi-day RANK test is based on one-period returns, and therefore, as in the case of single-day testing, the biasing effect of a low-order autocorrelation is negligible. In unreported results, the average estimation-period first-order autocorrelation of the residuals for the $1{,}000 \times 50 = 50{,}000$ sampled series across our simulations was only about -0.02, such that potential biasing effects to ADJ-PATELL do not show up in our simulation results in Table 7.

Serial heteroscedasticity has more serious implications. Modeling the volatility process with GARCH models is likely to be a useful approach. In the same manner, modeling the event-induced correlation effects could be worked out utilizing multivariate GARCH. Unfortunately, in practice, multivariate GARCH (with the possible exception of Engle's 2002 dynamic conditional correlation) cannot be estimated with the number of series typically used in event studies. The problem simplifies substantially in our case, as only the average cross-correlation is needed according to Equation (1). Therefore, a possibility is to estimate the average correlation from pairwise correlations utilizing bivariate GARCH models. Although computationally rather expensive, this should not be a major obstacle with modern high-speed computers. A more difficult potential problem even with a bivariate GARCH is the increased number of parameters that must be estimated using highly non-linear equations. This may lead to problems similar to those encountered by theoretically attractive GLS, where the gains become quickly dissipated by estimation and specification errors (see Malatesta 1986 and Chandra and Balachandran 1990). Nonetheless, we believe that there are ways to utilize GARCH-type approaches efficiently to predict the average time-varying correlations needed in the adjusted test statistics. This interesting topic is left for future research. According to our simulations studies in Tables 4 through 7, tests accounting for cross-correlation are fairly well specified under the null hypothesis of no event mean effect. Thus, the potential effect of serial heteroscedasticity and autocorrelation in returns is likely to show up in the powers of tests.

A related and harder problem to solve than serial heteroscedasticity is event effects on not only return variances but also correlations. The scaled test statistics are built upon the assumptions that cross-correlations of abnormal returns do not change in the event period from those of the estimation period and that possible event-induced volatility effects induce the same relative changes in the abnormal return volatilities. Regarding the null distribution of the test statistics, it is notable that these assumptions need only hold under the null hypothesis of no event mean effect. They may change under the alternative hypothesis and affect the powers (or Type II errors) of the tests. Thus, the robustness issue under the null hypothesis is a concern only when there is no mean effect of the event.

Of course, if event-induced volatility and correlation effects are predictable by a GARCH-type or some other model, these problems can be resolved. In practice, however, it may be extremely difficult to estimate individual variances and the event-period correlation matrix. Fortunately, this is not necessary to apply the ADJ-BMP test. It is necessary only to compute the ratio of event-period average covariances to average variances of the scaled abnormal (cumulative) returns. This result follows directly from the variance formula of an equally weighted portfolio. We denote the SCARs over the relevant accumulation period for stocks i and j simply by A_i and A_j and their variances and covariances as ω_i, ω_j, and γ_{ij}, respectively. Then, using the variance of a sum formula (portfolio variance formula) in the same manner as in deriving Equation (1), we can write the variance of the mean SCAR, $\bar{A}$, as

$$Var[\bar{A}] = \frac{\bar{\omega}}{n}(1 + (n-1)\theta), \tag{28}$$

where $\theta = \frac{\bar{\gamma}}{\bar{\omega}}$ is the ratio of average cross-covariance, $\bar{\gamma} = \sum_i \sum_{j \neq i} \frac{\gamma_{ij}}{n(n-1)}$, and average variance, $\bar{\omega} = \sum_i \frac{\omega_i}{n}$, of the SCARs. If the average covariance and variance are known, an appropriate test statistic would be $z = \sqrt{n}\frac{\bar{A}}{\sqrt{\bar{\omega}(1+(n-1)\theta)}}$, which under the null hypothesis of no event effect is (approximately) $N(0,1)$ distributed. In order to apply the test statistic in practice, estimates of the average variance as well as covariance/variance ratio are needed. There is no general-purpose solution to this problem. However, it turns out that eventually the only unknown parameter in the problem is θ. In fact, the expectation of the cross-section variance in Equation (7) (under the null hypothesis of no event mean effect) can be straightforwardly shown to be

$$E[s^2] = \bar{\omega}(1-\theta), \tag{29}$$

such that, given θ, $\frac{s^2}{1-\theta}$ would be an unbiased estimator $\bar{\omega}$. Thus, we see that the whole problem of finding a fairly general solution to volatility and cross-correlation robust efficient testing of mean event effects with powerful scaled test statistics reduces to finding an estimate of one single parameter of θ, the ratio of the event-induced average covariances and average variances. ADJ-BMP is a special operational case of this more general solution, which applies the assumption that under the null hypothesis, the correlations do not change and that volatility effects are relatively (at least approximately) the same across different stocks. Our simulation results show that for CRSP returns, even if the (relative) volatility effects are not the same, the ADJ-

BMP is robust to event-induced volatility. However, accounting for the impact of θ should not be overlooked in applications. If an event affects the ratio such that the average-sample-period correlation becomes a poor approximation of θ, the consequences may be as dramatic as ignoring cross-correlation in event study testing. In this respect, the results in Table 2 can be utilized as guidelines to evaluate biasing effects by simply replacing the average correlations line in the table with $\theta - \bar{\rho}$, i.e., the difference between the mean-covariance/mean-variance ratio and the average cross-correlation. In this regard, it would be worthwhile to find appropriate methods to estimate θ efficiently in future research.

6. Conclusion

In this article, we have demonstrated that even relatively low cross-sectional correlation in an event study with clustered event days can cause serious over-rejection of the null hypothesis of no event mean effect. Our results indicate that low cross-correlations bias tests in both marketwide and industry studies. Subsequently, we proposed cross-correlation and volatility-adjusted (ADJ-BMP) as well as cross-correlation-adjusted (ADJ-PATELL) scaled test statistics. Simulations with real CRSP returns show that, when there is no event-induced volatility increase, both of these adjusted test statistics are approximately equally powerful and reject the null hypothesis at the correct nominal rate when the null hypothesis is true. Also, in terms of power, the proposed statistics are clearly superior to the commonly used portfolio method with clustered event days. While the ADJ-PATELL test is sensitive to event-induced volatility in the presence of low cross-correlation, the ADJ-BMP statistic is robust to both variance changes and cross-correlation. Comparing our statistics with other popular tests, only the Corrado and Zivney (1992) nonparametric rank test shares the robustness and power properties of the ADJ-BMP test for single-day abnormal returns and short CAR windows. However, for longer CAR windows, the ADJ-BMP test outperforms the rank test. Based on these event-date clustering results, we conclude that our ADJ-BMP test statistic adjusts for cross-correlation and variance inflation in both single- and multiple-day event windows.

References

Aktas, N., E. de Bodt, and R. Roll. 2004. Market Response to European Regulation of Business Combinations. *Journal of Financial and Quantitative Analysis* 39:731–57.

Bernard, V. L. 1987. Cross-Sectional Dependence and Problems in Inference in Market-based Accounting Research. *Journal of Accounting Research* 25:1–48.

Bickel, P. J., and K. A. Doksum. 1977. *Mathematical Statistics*: *Basic Ideas and Selected Topics*. San Francisco: Holden-Day.

Bittlingmayer, G., and T. W. Hazlett. 2000. DOS Kapital: Has Antitrust Action Against Microsoft Created Value in the Computer Industry? *Journal of Financial Economics* 55:329–59.

Boehmer, E., J. Musumeci, and A. B. Poulsen. 1991. Event Study Methodology Under Conditions of Event Induced Variance. *Journal of Financial Economics* 30:253–72.

Brown, S. J., and J. B. Warner. 1980. Measuring Security Price Performance. *Journal of Financial Economics* 8:205–58.

———. 1985. Using Daily Stock Returns: The Case of Event Studies. *Journal of Financial Economics* 14: 3–31.

Campbell, C. J., and C. E. Wasley. 1993. Measuring Security Price Performance Using Daily NASDAQ Returns. *Journal of Financial Economics* 33:73–92.

———. 1996. Measuring Abnormal Trading Volume for Samples of NYSE/ASE and NASDAQ Securities Using Parametric and Non-parametric Test Statistics. *Review of Quantitative Finance and Accounting* 6:309–26.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Chandra, R., and B. V. Balachandran. 1990. A Synthesis of Alternative Testing Procedures for Event Studies. *Contemporary Accounting Research* 6:611–40.

Charest, G. 1978. Dividend Information, Stock Returns, and Market Efficiency. *Journal of Financial Economics* 6:265–96.

Collins, D. W., and W. T. Dent. 1984. A Comparison of Alternative Testing Methodologies in Capital Market Research. *Journal of Accounting Research* 22:48–84.

Conover, W. J. 1999. *Practical Nonparametric Statistics* (3rd ed.). Hoboken, NJ: Wiley.

Corrado, C. J. 1989. A Nonparametric Test for Abnormal Security-price Performance in Event Studies. *Journal of Financial Economics* 23:385–95.

Corrado, C. J., and T. L. Zivney. 1992. The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns. *Journal of Financial and Quantitative Analysis* 27:465–78.

Cowan, A. R. 1992. Nonparametric Event Study Tests. *Review of Quantitative Finance and Accounting* 2: 343–58.

Engle, R. F. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroscedasticity Models. *Journal of Business and Economics Statistics* 9: 339–50.

Fama, E. F., and K. R. French. 1992. The Cross-section of Expected Stock Returns. *Journal of Finance* 47: 427–65.

———. 1993. The Cross-section of Expected Returns. *Journal of Financial Economics* 33:3–56.

Harrington, S. E., and D. G. Shrider. 2007. All Events Induce Variance: Analyzing Abnormal Returns When Effects Vary Across Firms. *Journal of Financial and Quantitative Analysis* 42:229–56.

Jaffe, J. F. 1974. The Effect of Regulation Changes on Insider Trading. *Bell Journal of Economics and Management Science* 5:93–121.

Kolari, J., and S. Pynnonen. 2008. A Rank Test of Cumulative Abnormal Returns in Event Studies Based on Generalized Abnormal Returns. 2008 Australian Finance and Banking Conference (SSRN: http://ssrn.com/abstract=1254022).

Kothari, S. P., and J. B. Warner. 2007. Econometrics of Event Studies. In B. E. Eckbo (ed.), *Handbooks of Corporate Finance*: *Empirical Corporate Finance*, Chapter 1. Amsterdam: Elsevier/North-Holland.

Lehmann, E. L. 1975. *Nonparametrics*: *Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

———. 1999. *Elements of Large Sample Theory*. New York: Springer.

Mikkelson, W. 1981. Convertible Calls and Security Returns. *Journal of Financial Economics* 9:237–64.

Malatesta, P. H. 1986. Measuring Abnormal Performance: The Event Parameter Approach Using Joint Generalized Least Squares. *Journal of Financial and Quantitative Analysis* 21:27–38.

Patell, J. 1976. Corporate Forecasts of Earnings Per Share and Stock Price Behavior: Empirical Tests. *Journal of Accounting Research* 14:246–76.

Penman, S. 1982. Insider Trading and the Dissemination of Firms' Forecast Information. *Journal of Business* 55:479–503.

Petersen, M. A. 2008. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. *Review of Financial Studies* 22:435–80.

Rosenstein, S., and J. G. Wyatt. 1990. Outside Directors, Board Independence, and Shareholders' Wealth. *Journal of Financial Economics* 26:175–91.

Salinger, M. 1992. Standard Errors in Event Studies. *Journal of Financial and Quantitative Analysis* 27: 39–53.

Savickas, R. 2003. Event Induced Volatility and Tests for Abnormal Performance. *Journal of Financial Research* 26:165–78.

Sefcik, S. E., and R. Thompson. 1986. An Approach to Statistical Inference in Cross-sectional Models with Security Abnormal Returns as Dependent Variable. *Journal of Accounting Research* 24:316–34.

Wilcoxon, F. 1945. Individual Comparisons by Ranking Methods. *Biometrics* 1:80–83.