11-2010

Multi-Country Event Study Methods

Cynthia J. Campbell
Iowa State University, campcj@iastate.edu

Arnold R. Cowan
Iowa State University, arnie@iastate.edu

Valentina Salotti
Iowa State University, vsalotti@iastate.edu

Multi-Country Event Study Methods

Abstract
We provide the first simulation evidence of event-study test performance in multi-country non-US samples. The nonparametric rank and generalized sign tests are more powerful than two common parametric tests, especially in multi-day windows. The two nonparametric tests are mostly well specified, but neither is perfectly specified in all situations. The parametric standardized cross-sectional test can provide a useful robustness check but is less powerful than the nonparametric tests and rejects too often in single-market samples and when firm-specific events affect the market index. Local-currency market-model abnormal returns using national market indexes are sufficient.

Keywords
Event-study methodology, Datastream, Stock-price reaction, International finance, Market-moving events

Disciplines
Finance and Financial Management

Comments
This is a post-print of an article from *Journal of Banking & Finance*, 34, no. 12 (2010): 3078–3090, doi:10.1016/j.jbankfin.2010.07.016.

Multi-Country Event Study Methods

Cynthia J. Campbell
Iowa State University

Arnold R. Cowan
Iowa State University

Valentina Salotti*
Iowa State University

This draft: July 9, 2010

We provide the first simulation evidence of event-study test performance in multi-country non-U.S. samples. The nonparametric rank and generalized sign tests are more powerful than two common parametric tests, especially in multi-day windows. The two nonparametric tests are mostly well specified, but neither is perfectly specified in all situations. The parametric standardized cross-sectional test can provide a useful robustness check but is less powerful than the nonparametric tests and rejects too often in single-market samples and when firm-specific events affect the market index. Local-currency market model abnormal returns using national market indexes are sufficient.

The authors thank the anonymous referee as well as Göran Anderson, Christina Atanasova, Jedrzej Bailkowski, Ginka Borisova, Jens Forssbaeck, Imre Karafiath, James McCown, Greg Nagel, Walt Pohl, session participants at the 2007 INFINITI, 2008 Australasian Finance and Banking, Midwest Finance Association, CRSP Forum, Eastern Finance Association and FMA conferences, and seminar participants at Brock University, Colorado State University, Concordia University, Florida International University, Iowa State University, University of Bologna, and University of Lund for helpful comments. Arnie Cowan gratefully acknowledges support from the Wells Fargo Professorship in Finance at Iowa State University.

*Corresponding author, Department of Finance, Iowa State University, Ames, Iowa 50010–4813, United States; +1 515.294.4201, vsalotti@iastate.edu.

1. Introduction

Researchers use event-study methods to gauge the effects of information arrival on stock prices. The hypothesis tested is that information affects the value of stocks, on average, across firms with similar information arrival. Conclusions regarding the performance of event-study tests that appear in the methodological literature are based on simulations using data from single markets, especially the U.S., but the application of event-study methods to multi-country samples is growing rapidly. The suitability of specific event-study methods when applied to multi-country non-U.S. samples has not been established in the methodological literature. This paper provides simulation evidence of the performance of several methods in such samples.

Stock markets differ on many dimensions, e.g., size, liquidity, trading volume, market-making mechanisms, accounting standards, securities regulation, investor protection, ownership concentration, and corporate governance. Market characteristics can affect the statistical properties of stock returns (see Cole, Moshirian, and Wu, 2008 and Hutson, Kearney, and Lynch, 2008 as examples).We find that return distributions in non-U.S. multi-country samples are non-normal, even at the portfolio level, to a greater degree than U.S.-based studies report. In multi-country samples, where a mixture of distributions is present, the applicability of existing simulation evidence is an unexplored empirical question.

Examining recent journal articles that report event studies on multi-country samples, we find that researchers tend to use simple methods for identifying a benchmark or "normal" return, primarily the single-index market model, with the market-adjusted return method also appearing repeatedly. For testing whether the average abnormal return differs from zero, the "crude dependence adjustment" (CDA) test by Brown and Warner (1980, 1985) is often used (see Bailey, Karolyi, and Salva, 2006 and Aktas, de Bodt, and Roll, 2007 as examples). A parametric test based on standardized abnormal returns, introduced by Patell (1976) and Mikkelson and Partch (1986) and modified by Boehmer, Musumeci, and Poulsen (1991) is also common. Several papers report nonparametric tests such as the rank test (Corrado, 1989) and the generalized sign test (Cowan, 1992), especially in conjunction with a parametric test (as in Harvey, Lins, and Roper, 2004 and Behr and Güttler, 2008, among others). Nonparametric tests are naturally appealing for ill-behaved data, but in the absence of evidence cannot be assumed to be powerful and well specified. When a parametric and a nonparametric test are both reported in an article, they frequently lead to different inferences.

Using the simulation approach pioneered by Brown and Warner (1980, 1985), we investigate the accuracy and power of statistical tests applied to market-model abnormal returns. Overall, we find that the generalized sign test (Cowan, 1992) and rank test (Corrado, 1989) are more powerful in simulation than the two commonly used parametric tests. The parametric tests also are well specified but less powerful than the nonparametric tests. In the presence of a large return variance increase on the event date, the nonparametric tests tend to reject too often, but their specification is better under a more moderate variance increase. The standardized cross-sectional test is well specified under a variance increase and is more powerful than the CDA test.

We also examine test performance in samples that are potentially problematic for test specification or power. These include single-market samples, samples from the most concentrated national markets, and markets with the most non-normally distributed returns. The two nonparametric tests remain mostly well specified and powerful in these settings. The standardized cross-sectional test is less consistently well specified in single-market samples than in multi-country samples.

We also examine the ability of tests to detect abnormal returns when the affected securities are potential "market movers." This is when a stock can make up such a large fraction of its national market's capitalization that the individual price effects of firm-specific information arrivals exert a significant influence on the market index. Thus, abnormal return calculations that use the national market index would deduct the part of the information effect included in the index return from the total information effect in the stock return, potentially reducing power. When we simulate such effects, we find that the rank and generalized sign tests continue to exhibit correct specification and good power. The standardized cross-sectional test, which uses the index return in estimating a security's abnormal return variance, is not as reliably well specified in this situation.

Aspects of multi-country event-study design, other than the selection of a test statistic, are also potentially important. First, many markets are characterized by high frequencies of missing returns due to non-trading. Our results show that a corrective procedure proposed in the literature, treating missing returns as zero returns, sometimes called the "lumped returns" procedure, produces somewhat worse event-study test performance compared to the more standard "trade to trade" method. The latter involves omitting missing-price days from calculations while accounting for the corresponding market-index returns when the stock eventually trades. Second, our results indicate that the use of a national market index, without incorporating an international or U.S. index, is sufficient to produce well-specified and powerful tests of average stock-price effects. Third, the results suggest that for the types of stock-price reaction tests that we investigate, there is no need to convert returns from different markets into a common currency.

2. Data and methods

2.1. Data

We use Datastream to obtain daily data for over 50,000 non-U.S. stocks over 1988–2006. We download prices, dividends, and volume for active and delisted stock codes based on numerous lists compiled by Datastream. We limit the initial data set to equities that meet the following criteria:

- The beginning date of data on Datastream is not missing and is before July 1, 2004. This criterion limits the data set to equities that potentially have adequate data for the random selection and simulation procedures.
- There is a time series of prices available for a minimum of 300 consecutive trading days in 1988–2006. In making this determination, we do not exclude missing prices. However, the criterion requires some judgment because Datastream does not report an ending date for an individual security. We designate the last date of a reported non-missing price as the ending date for each security. If fewer than 300 trading days exist between the reported beginning date or the first trading day of 1988, whichever is later, and the inferred ending date, we exclude the security.

- The security name record on Datastream does not include one of the codes (listed in Appendix A) that indicate the security is not an ordinary share (common stock in U.S. terms).
- The security is not traded in the U.S.

We also download the Datastream Global total market index that corresponds to each equity issue. This is a series of value-weighted national market indexes in local currency that is also called the “level one” Datastream Global index series. Despite their labeling by Datastream as “total market” indexes, Datastream’s online help indicates that the level one indexes “do not include all companies in a market” but consist of “the most important companies by market value.”

Because different markets are characterized by different trading frequencies, excluding stocks from the simulations based on a moderate absolute number of non-missing returns, regardless of the market, could result in an overrepresentation of thickly traded stocks and stocks in more heavily traded markets. Therefore, we adopt a conservative approach to excluding stocks due to missing returns. First, in constructing the data set from which we draw simulation samples, we exclude stocks that are in the quartile of each market in each year having the lowest frequency of non-missing returns (in effect, the quartile of the market with the fewest trading days in that year). Second, we require each randomly selected security-event to have a minimum of 24 non-missing stock-return (and corresponding market-index return) observations in its 251-day estimation period (further described in Section 2.3) and to have a non-missing return on the designated day zero.

2.2. Return and abnormal return calculations

2.2.1. Returns

We calculate individual stock returns from prices and dividends to avoid the rounding problem with Datastream returns reported by Ince and Porter (2006). Each daily stock return is calculated from the previous day with a non-missing price to the current day, including dividends. We use Datastream price data type P, which the database delivers already adjusted for stock splits and other capital events.

To take into account different methods of handling the non-trading of stocks, we calculate both trade-to-trade and lumped daily returns (Maynes and Rumsey, 1993). Trade-to-trade returns are simply the calculated returns from non-missing price days; the return on a missing price day is missing. For a stock with a missing price, the corresponding market-index return is added to the next non-missing price day's index return for a trade-to-trade abnormal return calculation. Lumped returns consist of trade-to-trade returns on non-missing price days and zero on missing price days. The market-index return adjustment for missing trade-to-trade returns is not performed for lumped returns because the lumped return calculation produces no missing returns. Maynes and Rumsey suggest that lumped returns, by increasing the number of return observations, can improve the efficiency of estimators and test statistics used in event studies.

2.2.2. Abnormal returns

We use market-model abnormal returns for the simulations.[1] The abnormal return is:

$$u_{it} = R_{it} - (\alpha + \beta R_{mt}), \qquad (1)$$

[1] The conclusions are similar using market-adjusted returns (details not reported).

where R_{it} is the return of security *i* on day *t*, $\hat{\alpha}$ and $\hat{\beta}$ are ordinary least squares estimates of market model parameters, and $R_{m\,t}$ is the national value-weighted market index return.[2]

Researchers using event-study methods commonly examine multi-day windows to account for potential imprecision in dating the event or the availability of information about it to market participants, or for uncertainty about the speed of the event's effects on security prices. Multi-day windows can be particularly useful in multi-country samples where time zones and holidays affect the dates on which information can be impounded in stock prices. We examine windows of three and 11 trading days centered on the event date. The cumulative abnormal return for stock *i* over the event window of days T_1 through T_2 is:

$$CAR_i \;\; T_1, T_2 \;\; = \sum_{t=} \tag{2}$$

The cumulative average abnormal return for a sample of *N* stocks is:

$$CAAR \;\; T_1, T_2 \;\; = \frac{1}{N} \sum_{i=}^{N} \quad . \tag{3}$$

2.3. Simulation method

We adopt the simulation approach pioneered by Brown and Warner (1980, 1985) and used in several subsequent methodological studies (e.g., Campbell and Wasley, 1993; Corrado, 1989; Cowan, 1992; Cowan and Sergeant, 1996; Savickas, 2003). The approach resembles a Monte Carlo simulation, but instead of drawing from a theoretical probability distribution, observations are randomly drawn from actual data. To simulate an event study, the researcher

[2] The Datastream Global level one index for each national market is value- (capitalization-) weighted; the database provides no equal weighted version.

randomly selects a stock and an event date, and repeats the process to create multiple samples. Historical stock and market-index return data for the randomly selected security-events are used to estimate parameters and calculate test statistics. To evaluate the ability of a test to detect a stock-price reaction to an event, the researcher artificially induces or "seeds" an abnormal return by adding a constant to the actual return. Repetition across multiple samples provides a picture of the specification and the power of the test.

In this study, we create 1,000 samples, each containing 250 security-events. To allow for losses of randomly selected security-events due to inadequate data, we initially select 625,000 stocks with replacement using a uniform random-number generator. On each draw, each stock in the data set has a probability of being selected that is proportional to the number of trading days for which it has a price field (which may or may not contain a non-missing price) on Datastream during the sample period. For each stock selection, we randomly draw an event date (day zero) using a uniform distribution over the period from 259 trading days after the first recorded trading day for the stock to 35 days before the last recorded trading day.[3]

Trading days -256 through -6 are designated as the estimation period for market model parameters, standard deviations, fractions of abnormal returns with positive or negative signs, and ranks. Trading days -5 through +5 are designated as the event period, from which we separately examine day zero and three-day and 11-day windows centered on day zero. To simu-

[3] The specific choices of 259 and 35 days are arbitrary but are motivated by our interest in avoiding the inclusion of the initial and final trading days in the estimation and event periods and allowing the option of using longer event windows.

late abnormal returns, we add the following seeds to the event-day return: -0.03, -0.01, -0.005, 0, 0.005, 0.01, and 0.03.

2.4. Event-study tests

We examine four alternative statistical tests from the literature, two parametric and two nonparametric. The Patell (1976) Z statistic is among the most common methods for testing the null hypothesis of zero abnormal returns. Other studies are frequently cited for an identical or nearly identical test, particularly Dodd and Warner (1983) and Mikkelson and Partch (1986). Brown and Warner (1980, 1985) point out that a distinguishing feature of the test is that it assumes independence of returns across security-events. This assumption can improve power but can lead to misspecification when departures from the assumption are substantial.[4]

Boehmer, Musumeci, and Poulsen (1991) introduce a variance-change corrected version of the Patell test, called the standardized cross-sectional test. Because the variance correction does not harm the performance of the test when there is no variance change, we use the standardized cross-sectional test instead of the Patell test regardless of whether we simulate a variance change.

[4] Campbell and Wasley (1993) report that the Patell test rejects a true null hypothesis too often with Nasdaq samples due to the non-normality of Nasdaq returns, particularly lower priced and less liquid securities. Cowan and Sergeant (1996) report such excessive rejections in Nasdaq samples in upper-tailed but not lower-tailed tests. Maynes and Rumsey (1993) report a similar misspecification of the test using the most thinly traded one-third of Toronto Stock Exchange stocks.

The standardized cross-sectional test statistic for day t is:

$$Z_t = \frac{N^{-} \sum}{\sqrt{\frac{1}{N-} \sum}}, \qquad (4)$$

where

$$s_{it} = \sqrt{\left(\quad \sum \quad \right) \left[\quad \sum \quad \right]} \qquad (5)$$

M_i. is the number of non-missing estimation period returns for security i, and $\bar{R}_{m\_Est}$ is the mean daily national market index return in the estimation period. For multi-day windows, we use a test statistic that incorporates a correction for the serial dependence that exists by construction in successive prediction errors that are based on the same parameter estimates. The serial dependence correction is not discussed by Boehmer, Musumeci, and Poulsen (1991), but was introduced for the Patell test by Mikkelson and Partch (1988). Cowan (1993) reports that the correction performs well in simulation. The multi-day test statistic calculation starts with the standardized cumulative abnormal return,

$$SCAR_i \; T_1, T_2 = \quad , T_2 \; / s_{CAR_i \; T_1, T_2} \; , \qquad (6)$$

where T_1 and T_2 are the beginning and ending days of the eleven-day (three-day) event window. The estimated standard deviation of each $CAR_i \; T_1,T_2$ that incorporates the serial dependence correction is:

$$s_{CAR_i\ T_1,T_2} = \left(\qquad \sum \qquad \right)^{1/2} \left\{ \left[\qquad \left(\sum \qquad \right) \qquad \sum \qquad \right] \right\}^{1/2}, \tag{7}$$

and W_i is the number of non-missing daily returns for security i in the event window.

The standardized cross-sectional statistic for the window is:

$$Z_t = \sum \frac{\qquad}{\sqrt{N}\, s_{SCAR}}, \tag{8}$$

where

$$s_{SCAR} = \left[\qquad \sum \left(\qquad \sum \qquad \right) \right]^{1/2} \tag{9}$$

The second parametric test incorporates the portfolio (sample) time-series standard deviation; Brown and Warner (1980, 1985) describe the test as featuring a "crude dependence adjustment." That is, the test compensates for potential dependence of returns across security-events by estimating the standard deviation using the time series of sample (portfolio) mean returns from the estimation period. Therefore, we refer to this as the CDA test. The CDA test statistic for day zero is:

$$t_{CDA} = \tag{10}$$

where $\bar{u}_t$ is the equal weighted portfolio mean abnormal return on day t, i.e.,

$\bar{u}_t = \qquad \sum \qquad$, and the standard deviation of $\bar{u}_t$ for all t is:

$$s(\overline{u}) = \sqrt{\quad \sum_{t=-}\quad} \tag{11}$$

where $\overline{\overline{u}} = \quad \sum \quad$. The standard deviation estimated using portfolio-level time-series data from the estimation period reflects all the pairwise correlations between abnormal returns, thereby addressing cross-sectional dependence. If the u_{it} are normal, independent, and identically distributed, this test statistic is distributed Student t and is approximately standard normal under the null hypothesis. For the three- and 11-day event windows, the CDA test statistic is:

$$t_{CDA(T_1,T_2)} = \quad T_2 \;/ \left[\quad \right], \tag{12}$$

where T_1 and T_2 are the beginning and ending days of the event window.

The first nonparametric test is the generalized sign test analyzed by Cowan (1992). Cowan reports the test to be well specified and powerful in random samples of NYSE-AMEX and Nasdaq stocks; given the 1972–1990 data period, his Nasdaq sample is thinly traded on average.

The null hypothesis of the generalized sign test is that the fraction of day zero abnormal returns having a particular sign is equal to the fraction expected to have that sign. For negative seeds, we test the null of a non-negative sign; for positive seeds, we test the null of a non-positive sign. The fraction of abnormal returns expected to have a given sign based on the 251-day estimation period for a sample of N security-events is:

$$\hat{p} = \frac{1}{N} \sum_{i=}^{N} \quad \sum \tag{13}$$

where $M_i \leq$ is the number of non-missing returns in the estimation period for security-event *i*. For an upper-tail alternative hypothesis,

$$S_{it} = \begin{cases} \quad > \\ \quad \text{wise} \end{cases} . \tag{14}$$

The test statistic uses the normal approximation of a binomial distribution with parameter $\hat{p}$. The generalized sign test statistic is:

$$Z_G = \frac{}{[N\hat{p}(1-} , \tag{15}$$

where for an upper-tail alternative hypothesis, *w* is the number of stocks on the event date or in the event window for which the abnormal return or cumulative abnormal return is positive. For a lower-tail alternative hypothesis, substitute negative for positive in the definitions of S_{it} and *w*.

The second nonparametric test is Corrado's (1989) rank test. Unlike the generalized sign test, which relies on the frequency of positive or negative returns, the rank test transforms each security's time series of abnormal returns into their respective ranks. The rank statistic for day zero is:

$$t_{rank} = \left[\sum \right] \tag{16}$$

where k_{i0} is the rank of security-event *i*'s day zero abnormal return in security-event *i*'s combined 251-day estimation period and 11-day event period time series, $\bar{k}$ is the expected rank defined below, and s_k is the time-series standard deviation of the sample mean abnormal return ranks.

Corrado (1989) does not allow for missing observations in the return time series and therefore assumes the expected rank to be constant across securities. To allow for missing returns, we use a procedure equivalent to that of Corrado and Zivney (1992). We rank each security-event's non-missing returns with the lowest rank being zero. If there are missing returns, we transform the security-event's raw ranks to a scale of 0–261 by multiplying the raw rank by a scaling factor (262 divided by one plus the number of non-missing returns) and truncating to the integer part. The expected rank is the empirical mean of the transformed ranks, $\bar{k} = \frac{1}{261} \sum_{j=-} \sum_{=}^{N}$ where N_t is the number of non-missing returns on day t. The standard deviation, s_k, is estimated at the portfolio level from the combined 251-day estimation and 11-day event periods as:

$$s_k = \left\{ \quad \sum \quad \sum \quad \right\}^{1/} . \tag{17}$$

The rank statistic converges to standard normal as the number of securities in the portfolio increases (Corrado, 1989).

Corrado (1989) applies the rank test only to day zero. Similar to Cowan (1992), we apply the rank test to a multi-day window CAAR by substituting security-event *i*'s mean rank across the three or 11 days that make up the window, in place of k_{i0} in equation (17), and dividing s_k by the square root of three or 11.

Corrado (1989) reports the rank test to be well specified and powerful for NYSE stocks. Campbell and Wasley (1993) find similar results for this test for Nasdaq stocks, even in small portfolios and infrequently traded low priced securities.

3. Results

3.1. Statistical properties of returns

Table 1 reports statistics of the 54 countries' equity returns represented in the sample before random selection (and before dropping the least often traded quartile of each market). Statistics for the U.S. market (NYSE, Nasdaq, and Amex), which is not in the simulation sample, are shown for comparison. Large developed markets such as Canada, Japan, and the U.K. are heavily represented, but markets that individually have less than 5% of the stock return-days in the sample collectively make up 53.4% of all return-days.

[Insert Table 1 here.]

The descriptive statistics of returns in Table 1 are averages of statistics calculated at the individual security level. For most markets, the average of stocks' median returns is close to zero. However, there is wide variation in the average of mean, standard deviation, and percentage of returns equal to zero. Many average means appear to be distorted by outliers. The trimmed means (dropping the most extreme ½% of individual stock means in each tail) are more reasonable but still appear to be outlier-driven compared to the medians, consistent with non-normality. The average skewness and excess kurtosis of returns in the overall data set and for most markets are markedly greater than zero, suggesting that non-normal returns are pervasive. The overall average standard deviation, skewness, and excess kurtosis are several times the corresponding statistics for the U.S. The results in Table 1 indicate that individual equity returns in multi-country, non-U.S.-dominated samples generally are more volatile and diverge from a normal distribution substantially farther than in U.S. samples.

Table 2 reports the properties of event-day abnormal returns for the 250,000 randomly selected security-events (Panels A and B) and for the portfolios of 250 security-events each (Panel C and D) in the final sample when no abnormal performance is introduced. The results reflect the exclusion of stocks with large numbers of missing returns described in Section 2.1. The abnormal returns are positively skewed and fat-tailed. For example, individual day zero trade-to-trade returns have skewness of 7.798 and excess kurtosis of 41,570. Portfolio abnormal returns are markedly less skewed and fat-tailed than individual abnormal returns, but they still diverge from a normal distribution. Properties differ little between trade-to-trade and lumped abnormal returns. The Jarque-Bera test rejects the null hypothesis of normality for both individual and portfolio abnormal returns.

[Insert Table 2 here.]

The average skewness and excess kurtosis for trade-to-trade returns of individual non-U.S. stocks in Table 2, Panel A far exceed the corresponding results in the literature for U.S. stocks. For example, Brown and Warner (1985) report that market-model abnormal returns of individual NYSE-AMEX stocks over 1962–1979 have skewness of 1.01 and kurtosis of 6.80 (excess kurtosis thus being 3.80). Brown and Warner also report that 50-stock portfolios have skewness of 0.10 and excess kurtosis of 0.10. For Nasdaq securities, Campbell and Wasley (1993) report that the average skewness and kurtosis of market model abnormal returns of individual stocks are 0.96 and 16.98 from December 14, 1973 through December 20, 1987.

Cowan and Sergeant (1996) report that market-model abnormal returns in the most thinly traded Nasdaq sample in 1983–1993 have average skewness of 0.68 and excess kurtosis of 26.51. We conclude that random event-study samples of non-U.S. stocks exhibit far more severe departures from normal return distributions than U.S. stocks. [5]

3.2. Simulations with multi-country random samples

Table 3 presents the simulation results for a one-day event window. Because the sign of the seeded abnormal return is known, we report one-tailed test results. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% binomial success probability, are 3.3% to 6.8%. The one-day results in Panel A of Table 3 show that using trade-to-trade returns, none of the tests excessively rejects a true null hypothesis. In Panel B using lumped returns, the overall patterns across tests under the null do not differ greatly from Panel A.

[Insert Table 3 here.]

For the one-day event window, the choice of method lies mostly with the relative power of the tests. The most powerful tests in Table 3 are the generalized sign test (GST) and the rank test, both of which virtually always reject the null hypothesis with a seeded abnormal return. The CDA test is the worst in terms of power. The standardized cross-sectional test is

[5] The kurtosis statistics in Table 2 for individual security abnormal returns are particularly extreme in comparison to U.S. samples. In the statistics literature, Schmid and Trede (2003) report that the sample kurtosis statistic is extremely sensitive to outliers. We find this statement to be applicable to our data set. For example, when we re-calculate the excess kurtosis of individual security abnormal trade-to-trade returns excluding observations below the first percentile or above the 99th percentile, the excess kurtosis drops from 41,750 to 5.23. More pronounced outliers and kurtosis than have been reported in U.S. data are consistent with extreme price movements and illiquidity in many of the markets covered by Datastream.

well specified but somewhat less powerful than the nonparametric tests. We conclude that for testing the one-day stock-price reaction, the nonparametric tests dominate.

[Insert Table 4 here.]

Table 4 shows that, using the three-day event window (–1, +1), the conclusions are similar, though the power of all tests is reduced as the average daily impact of the seeded abnormal return is smaller. The rank test is slightly more powerful than the GST with a negative seeded abnormal return of less than 3% absolute value, whereas the GST is slightly more powerful with positive seeds of less than 3%. Similar to the day zero results, Table 4 reports that the CDA test is the least powerful in three-day windows. Campbell and Wasley (1993) similarly find the CDA test to be substantially less powerful than the Patell and rank tests in multi-day windows for Nasdaq samples. We conclude that for testing the three-day stock-price reaction, the nonparametric tests dominate in terms of power and specification.

[Insert Table 5 here.]

The eleven-day window results in Table 5 are again qualitatively similar. Of the two parametric tests, the CDA test again is by far the weaker. Brown and Warner (1985) point out that the distinguishing feature of the CDA test, estimating the variance at the portfolio level to adjust for cross-sectional dependence, "can actually be harmful compared to procedures which assume independence," as the standardized cross-sectional test assumes. Brown and Warner explain that even if the independence assumption is only approximately true, it offers greater precision in estimating the variance of abnormal returns and therefore "can make it easier to detect abnormal performance when it is present." Multi-country samples are likely to be highly heterogeneous, magnifying the imprecision in the CDA test's variance estimate.

We conclude that for multi-day windows, nonparametric tests are likely to be the best choices. Test performance in Tables 3, 4, and 5 is slightly better with trade-to-trade as opposed to lumped returns. Therefore, we conduct the remaining simulations on trade-to-trade returns only.

3.3. Simulations using random samples with a variance increase on the event date

Brown and Warner (1985) report that a return variance increase on the event date adversely affects the specification of tests that use variance estimates from outside the event window. When the variance increases on the event date, using a time-series of non-event period data to estimate the variance of the mean abnormal return can result in too many rejections of the null hypothesis. We use the method of Brown and Warner (1985) to simulate a temporary doubling of the stock-return variance on day zero by adding the day –6 return to the day 0 return and subtracting the average estimation period return.

Corrado and Zivney (1992) present a version of the rank test that is adjusted for variance increases by standardizing the abnormal return on the event date only. In results not reported in a table, we find this test to be misspecified in multi-country samples with a simulated variance increase. Because ranks are based on the combined estimation and event period, and standardized abnormal returns in multi-country samples are more likely to exhibit extreme values, standardizing only on the event date could distort the ranks. We therefore introduce a further variant of the rank test in which abnormal returns are standardized each day of the estimation and event periods before ranking.[6]

The results are in Table 6. The results for the standardized cross-sectional and CDA tests resemble those in Tables 3 through 5. However, the GST is now misspecified in the upper tail. From a detailed examination of the data (not reported in a table), we find that the misspecification is due to negative abnormal returns tending to cluster closer to zero than positive abnormal returns. When the net amount added to induce a variance increase is positive and the underlying abnormal return is negative, the sign of the abnormal return is more often reversed than when the added quantity is negative and the underlying abnormal return is positive. It is difficult to say whether such sign reversals would occur in the case of an actual event where the effect on the mean return is zero and the variance increases, but the result warrants caution in interpreting a significantly positive GST result when a large variance increase is suspected, especially when the point estimate of abnormal return is small in absolute value. While the rank test is severely misspecified in the lower tail, the standardized rank test shows good specification and power.

[Insert Table 6 here.]

In Table 6 overall, the standardized cross-sectional and standardized rank tests appear to perform best. The standardized rank test is the more powerful of the two for day zero and in the three-day window, and the standardized cross-sectional test is the more powerful in the

[6] We do not evaluate the standardized rank test in other sections of the paper because the version in which the ranks are standardized every day is computationally burdensome, increasing the length of a simulation run by an order of magnitude or more, and the simple rank test performs well under most conditions.

11-day window. However, given the prevalence of outliers in our data, the Brown and Warner (1985) method of simulating a doubling of variance, by drawing from a stock's own return distribution, could exaggerate the potential event-driven deformation of stocks' usual return distributions. To check the robustness of the conclusions drawn from Table 6, we use a different method and simulate a more modest increase in the abnormal return variance on day 0.

Following Boehmer, Musumeci, and Poulsen (1991), we generate a random value based on a normal distribution with a mean of zero and an appropriate variance. We add the random value to the day 0 abnormal return to increase the variance by 50%. The results (not displayed) show that the generalized sign test is misspecified in the upper tail for day 0 but correctly specified for the multi-day windows, and the standardized cross-sectional test and rank test (without standardization) are correctly specified for day 0 and the multi-day windows. The generalized sign and rank tests are more powerful than the standardized cross-sectional test, especially in multi-day windows. While vigilance is in order, the generalized sign and rank tests appear to be fairly reliable under modest event-induced variance increases. The standardized cross-sectional test provides a good robustness check when a variance increase is suspected.

3.4. Simulations with country-clustered samples

Although the main focus of this paper is the use of event study methods on multi-country samples, studies conducted on non-U.S. single country samples (see for instance Dutta and Jog, 2009) also appear frequently. The small populations and limited trading history of many markets in the data set raises the potential concern that a sample from a single market or a few markets could suffer from extensive cross-correlation, which the literature (e.g., Brown and Warner 1980, 1985) shows can cause various tests to be misspecified. Therefore, we repeat the main simulations using country-clustered samples. That is, each of the 1,000 samples contains 250 security-events that are from a single market, but the markets vary across the 1,000 samples. To create the samples, we start with the set of security-events described in Section 2.3, but this time we sort the data set by market, and by order of random drawing within each market, before forming samples. We use a number of samples from each market that is proportional to the number of stock return-days (the sum of the number of available days for each stock) for each market.

[Insert Table 7 here.]

The results are in Table 7. For all lengths of the event window, the CDA and nonparametric tests are well specified. The GST and the rank test continue to dominate the CDA in terms of power. However, there is a noticeable improvement in the power of the CDA compared to the simulation results using multi-country samples. Apparently, estimating a single variance for the sample is more accurate when the sample is more homogeneous, consistent with Brown and Warner's (1985) argument that the CDA test's not assuming independence across securities is disadvantageous for power when the abnormal return variance differs across securities.

Table 7, Panel C shows that the standardized cross-sectional test is misspecified in 11-day windows, rejecting a true null hypothesis about 8% of the time versus the nominal significance level of 5%. Like the Patell test from which it derives, the standardized cross-sectional test is based on assuming independence of abnormal returns across security-events. The results are consistent with the assumption being more often violated in single-country non-U.S. samples than in multi-country samples.

A caveat to interpreting the results of this section is that our sample formation method for Table 7 forces the number of samples to be proportional to the markets' representation in the data set of daily stock returns from which we draw, resulting in more samples from larger markets with longer histories.

The results so far indicate that the generalized sign, rank, and standardized cross-sectional tests mostly perform well in non-U.S., multi-country, and single-country samples. The CDA test is always substantially less powerful than the other tests so we do not examine it in further simulations.

3.5. Simulations with samples from the most concentrated markets

Some markets in the data set are long established as relatively large, developed, integrated markets in countries with equity-oriented financial systems. Others are only getting started in the latter years of our sample period, and still others are at various stages of development in various years that we study. In this section, we investigate whether the main results hold up in samples restricted to less advanced markets. To gauge a market's degree of development, we use the extent to which trading is concentrated in a few issues.[7] To measure trading concentration while allowing for changing market characteristics over time, we divide the data into an initial four-year period and five subsequent three-year periods. We calculate each stock's daily market value traded by multiplying its volume by the closing price the same day.

[7] Trading concentration is important also because of the potential effects on other stocks of dominant issues' trading. For example, Braun and Larrain (2009) report that large IPOs can alter the return distributions of other stocks in emerging markets.

Our empirical proxy for a market's concentration is a Herfindahl index calculated using the median daily market value traded in the four- or three-year period.

[Insert Table 8 here.]

We restrict the simulation samples in each period to the ten markets with the largest concentration proxy in the period, excluding any market having fewer than 20 issues with data in the period. The results are in Table 8. The standardized cross-sectional test rejects marginally more often than 5% in the lower tail for day 0. Otherwise, the standardized cross-sectional, rank, and generalized sign tests exhibit proper specification and power similar to the main simulations. We conclude that the good performance of the three tests, and the superior power of the two nonparametric tests, is robust to trading concentration.

3.6. Samples from markets with the most non-normal returns

One could argue that analyzing the most concentrated markets still does not guarantee that the results are not driven by markets where returns depart less dramatically from normality than in other markets. While normality per se is not an issue for nonparametric tests, the tests still could be sensitive to return irregularities that lead to non-normality, such as the frequent large outliers that characterize extremely fat-tailed distributions.

Table 9 reports simulations on the ten markets with the most non-normally distributed equity returns in each three- to four-year period. The power of the parametric standardized cross-sectional test is not harmed by increased non-normality. Power improves slightly on day 0 and in three-day windows relative to the main results (Tables 3 and 4) and deteriorates slightly in 11-day windows (compared to Table 5). The lack of much change in the power is not necessarily unexpected, as the usual concern about applying parametric tests when distributional assumptions are violated relates not to power but to specification under the null hypothesis. The specification of the test also holds up well in the most non-normal samples. This is surprising in light of the finding in Campbell and Wasley (1993) that a closely related parametric test, that likewise assumes independence, is misspecified in early Nasdaq samples, which are characterized by non-normality. The standardized cross-sectional test differs in that it adjusts for time-series heteroskedasticity, which tends to penalize outliers on the event date. We speculate that this additional variance adjustment is important in controlling the tendency of parametric tests that assume independence to reject a true null hypothesis too often in non-normally distributed samples.

[Insert Table 9 here.]

Table 9 also shows that the generalized sign test continues to perform well, with uniformly better power than the standardized cross-sectional test; its power in 11-day windows is improved relative to random samples. The rank test also shows a power improvement in the 11-day window, but this is accompanied by an upper-tail rejection rate under the null of almost double the nominal 5%.

3.7. Samples from the most concentrated markets in the case of market-moving events

In concentrated markets, some stocks could be a large enough component of national market indexes that events affecting the stocks also affect the market indexes, making it difficult to detect abnormal performance by adjusting the stock return using the national market index. To investigate this possibility, we again use the concentrated market samples from Table 8, with the following modification to the return-generating process. We multiply each stock's seeded return by the stock's fraction of its market's capitalization and add the product to the market index before calculating abnormal returns. The results in Table 10 show that the standardized cross-sectional test is misspecified in the lower tail for day 0. The generalized sign and rank tests continue to be well specified and powerful.[8]

[Insert Table 10 here.]

4. Conclusions

We examine the performance of event-study statistical tests applied to market-model abnormal trade-to-trade and lumped returns in simulations using actual return data on 48,258 ordinary share issues from 54 non-U.S. markets over 1986–2006. In random samples, security abnormal returns, and even portfolio abnormal returns for 250-stock samples, depart substantially from a normal distribution. The simulation results show that four common tests tend to be well specified under most test conditions that we simulate. Two nonparametric tests, the generalized sign and rank tests, are the most powerful. The parametric standardized cross-sectional test is less powerful, and the other parametric test, based on Brown and Warner's (1980, 1985) "crude dependence adjustment", tends to be quite weak, especially in longer event windows, although its power increases somewhat in single-country samples.

Although correctly specified in random samples, none of the three relatively powerful tests perfectly conforms to the nominal 5% significance level across various test conditions

[8] Stocks trading in concentrated markets could be more correlated with world stock returns than local returns due to limited local information production. To address this possibility, in a robustness check not reported in a table, we calculate abnormal returns using a two-index market model with both local and U.S. level one market indexes from Datastream. Following Jin and Myers (2006), we introduce two leads and lags for the local and U.S. indexes. The specification and power of tests using the expanded model do not differ significantly from the single-factor, local-index market model. Aktas, de Bodt, and Roll (2004) report that their event-study inferences do not change as a result of using local indexes or converting returns to U.S. currency.

designed to check robustness. The standardized cross-sectional test tends to reject a true null hypothesis too often for longer windows in two situations: country-clustered samples and when stock-specific events move the national market index. The generalized sign test and the rank test (unless standardized) are sensitive to a large return variance increase on the event date, but may perform well under more moderate variance increases. The rank test is sensitive to sample-wide extreme non-normality when testing a longer event window. However, most results under the null hypothesis show correct specification. Therefore, we recommend that at least two of the three tests be used and that any disagreement be interpreted with caution. The rank and generalized sign tests would be logical to use for balanced power and correct specification. If the researcher prefers to sacrifice some power for the sake of conservatism, the standardized cross-sectional test would provide a useful robustness check except where high cross-sectional correlation is likely (as in country clustering) and where firm-specific events are likely to move the market index.

Apart from the selection of a test statistic, the results suggest that trade-to-trade returns and simple market-model methods of calculating abnormal returns with national market indexes, without converting to a common currency, work well. More elaborate methods do not improve test specification or power in the settings that we examine.

References

Aktas, N., de Bodt, E., Roll, R., 2004. Market response to European regulation of business combinations. Journal of Financial and Quantitative Analysis 39(4), 731-757.

Aktas, N., de Bodt, E., Roll, R., 2007. Is European M&A regulation protectionist? The Economic Journal 117(522), 1096-1121.

Behr, P., Güttler, A., 2008. The informational content of unsolicited ratings. Journal of Banking and Finance 32, 587-599.

Bailey, W., Karolyi, G.A., Salva, C., 2006. The economic consequences of increased disclosure: Evidence from international cross-listing. Journal of Financial Economics 81, 175-213.

Boehmer, E., Musumeci, J., Poulsen, A.B., 1991. Event-study methodology under conditions of event-induced variance. Journal of Financial Economics 30(2), 253-272.

Braun, M., Larrain, B., 2009. Do IPOs affect the prices of other stocks? Evidence from emerging markets. Review of Financial Studies 22(4), 1505-1544.

Brown, S.J., Warner, J.B., 1980. Measuring security price performance. Journal of Financial Economics 8(3), 205-258.

Brown, S.J., Warner, J.B., 1985. Using daily stock returns: The case of event studies. Journal of Financial Economics 14(1), 3-31.

Cole, R.A., Moshirian, F., Wu, Q., 2008. Bank stock returns and economic growth. Journal of Banking and Finance 32, 995-1007.

Campbell, C.J., Wasley, C.E., 1993. Measuring security price performance using daily NASDAQ returns. Journal of Financial Economics 33(1), 73-92.

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

Cowan, A.R., 1992. Nonparametric event study tests. Review of Quantitative Finance and Accounting 1(4), 343-358.

Cowan, A.R., 1993. Tests for cumulative abnormal returns over long periods: Simulation evidence. International Review of Financial Analysis 2(1), 51-68.

Cowan, A.R., Sergeant, A.M.A., 1996. Trading frequency and event study test specification. Journal of Banking and Finance 20(10, Dec), 1731-1757.

Dodd, P., Warner, J.B., 1983. On corporate governance: A study of proxy contests. Journal of Financial Economics 11, 401-438.

Dutta, S., Jog, V., 2009. The long-term performance of acquiring firms: A re-examination of an anomaly. Journal of Banking and Finance 33, 1400-1412.

Harvey, C.R., Lins, K.V., Roper, A.H., 2004. The effect of capital structure when expected agency costs are extreme. Journal of Financial Economics 74, 3-30.

Hutson, E., Kearney, C., Lynch, M., 2008. Volume and skewness in international equity markets. Journal of Banking and Finance 32, 1255-1268.

Ince, O.S., Porter, R.B. 2006. Individual equity return data from Thomson Datastream: Handle with care! Journal of Financial Research 29(4), 463-479.

Jin, L., Myers, S.C., 2006. R^2 around the world: New theory and new tests. Journal of Financial Economics 79, 257-292.

Maynes, E., Rumsey, J., 1993. Conducting event studies with thinly traded stocks. Journal of Banking and Finance 17, 145-157.

Mikkelson, W.H., Partch, M.M., 1986. Valuation effects of security offerings and the issuance process. Journal of Financial Economics 15, 31-60.

Mikkelson, W.H., Partch, M.M., 1988. Errata: Withdrawn security offerings. Journal of Financial and Quantitative Analysis 23(4), 487.

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

Savickas, R., 2003. Event-induced volatility and tests for abnormal performance. Journal of Financial Research 26, 165-178.

Schmid, F., Trede, M., 2003. Simple tests for peakedness, fat tails and leptokurtosis based on quantiles. Computational Statistics and Data Analysis 43, 1-12.

Wuertz, D., Katzgraber, H., 2009. Precise finite-sample quantiles of the Jarque-Bera adjusted Lagrange multiplier test. Working paper, Swiss Federal Institute of Technology, Zurich.

Appendix A. Sample selection details

This appendix provides more details of the data selection procedure in Section 3.1. We exclude a security if the name record on Datastream includes one of the following codes that indicates it is not an ordinary share issue: CV, CONV, CVT, FD, OPCVM, PREF, PF, PFD, PFC, PFCL, RIGHTS, RTS, UNIT, UNITS, WTS, WT, WARR, WARRANT, and WARRANTS. To avoid using securities traded in the U.S., we exclude a security if it has a Datastream mnemonic beginning with U: or @, or an exchange code of NYS, ASE, NAS, XBQ, BOS, CHI, MID, NMS, OTC, PBT, PHL, PSE, or XNT. The mnemonic is usually in the format market code:ticker, with market code: omitted for U.K. stocks. As tickers are recycled within markets, mnemonics do not uniquely identify stocks within Datastream.

Datastream includes a field for each equity issue that identifies the "associated" level one market index. At the time we downloaded much of the data, late 2004 and early 2005, the field for dead stocks was essentially always filled with TOTMKUK, the code for the U.K. level one index, regardless of the market on which the stock traded while alive. This appears to be largely corrected in downloads from 2007 or later. To ensure that we use the correct index for dead stocks, we identify dead stocks by searching the name field for the codes DEAD, SUSP, DELIST, EXPD, DEL, DELEST, DELISTED, and DEF. We use the market code portion of the mnemonic to identify the market and select the corresponding market index.

One of the frustrations of dealing with Datastream is that the market code portion of the security mnemonic, the exchange code, and the market portion of the level one Datastream Global index mnemonic are different. To select level one market indexes for dead stocks, we use the following pairings of security-mnemonic market code (level one market index mnemonic):

A	TOTMKAU	H	TOTMKNL	PH	TOTMKPH
AG	TOTMKAR	ID	TOTMKID	PK	TOTMKPK
B	TOTMKBG	I	TOTMKIT	PO	TOTMKPO
BN	TOTMKBN	IN	TOTMKIN	Q	TOTMKTH
BR	TOTMKBR	IS	TOTMKIS	R	TOTMKSA
C	TOTMKCN	J	TOTMKJP	RS	TOTMKRS
CB	TOTMKCB	K	TOTMKHK	S	TOTMKS
CL	TOTMKCL	KN	TOTMKKN	SL	TOTMKCY
CN	TOTMKCH	KO	TOTMKKO	T	TOTMKSG
CP	TOTMKCP	L	TOTMKMY	TK	TOTMKTK
CZ	TOTMKCZ	LX	TOTMKLX	TW	TOTMKTA
D	TOTMKBD	M	TOTMKFN	U	TOTMKUS
E	TOTMKES	MC	TOTMKMC	V	TOTMKVE
ED	TOTMKED	MX	TOTMKMX	W	TOTMKSD
EG	TOTMKEY	N	TOTMKN	Z	TOTMKNZ
F	TOTMKFR	O	TOTMKOE	ZI	TOTMKZI
G	TOTMKGR	P	TOTMKPT		
GD	TOTMKPH	PE	TOTMKPE		

If the associated index field is empty and the stock is not dead, or if the stock is dead and we cannot identify a level one market index corresponding to its market, we drop the stock from the data set.

Another problem has to do with the trading volume data we use as part of our market-concentration measure. A small amount of volume data is misreported in the data set we downloaded for our simulations. Specifically, 61 of the originally downloaded volume figures are negative. As of mid-2008, Datastream apparently changed the negative volumes to zero or missing. Our spot checking uncovers no changes to volume figures that were not negative in our original download.

Table 1

Descriptive statistics of daily trade-to-trade returns of individual equities in 54 sample countries, 1988-2006

The sample includes non-U.S. stocks (ordinary shares) that have Datastream price data starting before 2004 and ending no earlier than 1988. The inclusion criteria are based on the trading history in the Datastream database, not necessarily a stock's entire history as a public issue. We calculate returns using Datastream split-adjusted prices and dividends. The ½% trimmed mean column reports the trimmed mean (a robust estimator of location) across stocks, of the untrimmed mean daily return, where the trimming removes the ½% most extreme observations in each tail of the sample. The U.S. (NYSE, Nasdaq, and Amex) is shown for comparison; it is not included in any other results in the paper.

				Mean across stocks of:						
Country	Number of stocks	Mean no. of returns per stock	% of the overall sample	Mean	Mean (½% trimmed)	Median	Standard deviation	Skewness	Excess kurtosis	Percent of zero returns
U.S.	15,482	1,836	NA	0.008	0.001	0.000	0.080	1.517	42.259	22.7%
Overall	48,258	1,665	100.00%	0.077	0.008	0.001	2.696	4.891	229.823	20.7%
Argentina	135	1,350	0.20%	0.171	0.081	0.000	2.847	3.015	88.436	14.2%
Australia	2,263	1,369	3.90%	0.005	0.003	0.000	0.109	2.773	128.021	18.2%
Austria	228	1,646	0.50%	0.002	0.003	-0.001	0.073	5.961	164.610	19.1%
Belgium	886	1,130	1.20%	0.184	0.024	-0.001	4.866	7.033	220.496	12.1%
Brazil	798	820	0.80%	0.122	0.027	0.003	1.821	3.744	111.260	11.5%
Canada	6,786	1,644	13.90%	0.016	0.011	0.000	0.319	5.614	206.772	24.8%
Chile	259	1,362	0.40%	0.007	0.007	0.001	0.064	2.091	58.124	13.2%
China	1,435	1,894	3.40%	0.000	0.000	0.000	0.029	0.230	18.577	5.0%
Colombia	156	375	0.10%	0.052	0.021	-0.004	0.401	2.009	51.534	3.9%
Cyprus	140	1,124	0.20%	0.003	0.003	0.000	0.098	5.856	142.557	19.9%
Czech Rep.	32	2,061	0.10%	0.000	0.000	0.000	0.026	0.397	11.363	32.0%
Denmark	379	1,567	0.70%	0.046	0.006	0.000	1.201	2.081	116.856	12.0%
Ecuador	3	4	0.00%	-0.008	-0.008	-0.001	0.053	-2.914	9.073	0.1%
Finland	266	1,787	0.60%	0.001	0.001	0.000	0.041	1.778	58.237	22.2%
France	2,094	1,542	4.00%	0.012	0.004	0.001	0.356	3.616	152.092	13.4%
Germany	6,306	1,016	8.00%	0.023	0.003	0.009	0.295	3.780	170.197	26.8%
Greece	472	2,092	1.20%	0.015	0.014	0.000	0.371	22.056	764.598	11.9%
Hong Kong	1,150	1,875	2.70%	0.004	0.002	0.000	0.149	4.491	216.371	14.9%
Hungary	47	1,549	0.10%	0.006	0.004	0.000	0.102	1.980	49.765	9.9%
India	1,315	1,966	3.20%	0.004	0.003	0.000	0.079	1.813	67.243	7.7%
Indonesia	415	1,394	0.70%	0.004	0.003	0.000	0.081	3.144	79.857	20.6%
International	89	1,308	0.10%	0.001	0.001	0.000	0.103	4.209	507.578	3.2%
Ireland	138	2,268	0.40%	0.002	0.001	0.000	0.055	2.883	255.134	52.7%
Israel	762	1,485	1.40%	0.010	0.002	0.000	0.076	2.531	96.376	16.6%
Italy	565	2,436	1.70%	0.464	0.192	0.000	17.352	20.866	731.872	15.1%

Table 1 continued

Country	Number of stocks	Mean no. of returns per stock	% of the overall sample	Mean across stocks of: Mean	Mean (½% trimmed)	Median	Standard deviation	Skewness	Excess kurtosis	Percent of zero returns
Japan	3,715	2,663	12.30%	0.382	0.025	0.000	19.914	8.989	511.614	12.4%
Luxembourg	113	1,046	0.00%	0.003	0.003	0.000	0.066	6.001	179.580	16.1%
Malaysia	1,004	2,294	0.03%	0.001	0.001	0.000	0.043	2.169	47.749	20.0%
Mexico	327	982	0.00%	0.013	0.007	0.001	0.129	2.100	93.828	9.3%
Morocco	12	513	0.00%	0.002	0.002	0.000	0.036	11.781	280.527	70.2%
Netherlands	591	1,863	0.01%	0.034	0.014	0.000	1.020	5.069	492.191	28.5%
New Zealand	339	1,430	0.01%	0.011	0.003	0.008	0.093	3.981	221.934	27.1%
Norway	430	1,194	0.01%	0.192	0.058	0.000	2.340	1.993	54.138	13.2%
Pakistan	293	1,264	0.01%	0.008	0.006	0.001	0.090	3.928	114.510	7.0%
Peru	193	634	0.00%	0.010	0.009	0.003	0.125	2.091	47.971	7.7%
Philippines	296	1,565	0.01%	0.047	0.006	-0.001	0.776	5.190	189.201	20.8%
Poland	278	1,371	0.01%	0.001	0.001	0.000	0.041	1.098	41.389	11.8%
Portugal	222	1,183	0.00%	0.030	0.014	0.002	0.698	11.563	331.729	12.8%
Romania	47	1,571	0.00%	0.075	0.066	0.000	3.016	11.666	608.271	15.8%
Russian Fed.	117	342	0.00%	0.437	0.223	0.002	9.467	3.019	66.736	8.0%
Singapore	853	1,743	1.90%	0.020	0.001	0.016	0.064	1.813	40.001	19.7%
Slovakia	1	47	0.00%	0.000	0.000	0.000	0.000	—	—	1.2%
South Africa	865	1,345	1.40%	0.008	0.007	0.000	0.165	3.459	138.087	20.0%
Spain	261	2,334	0.80%	0.033	0.014	0.000	1.313	14.828	540.357	16.3%
Sri Lanka	272	1,317	0.40%	0.007	0.007	0.000	0.124	4.396	137.067	14.3%
Sweden	942	1,306	1.50%	0.081	0.007	0.000	1.432	2.927	86.111	16.1%
Switzerland	679	1,573	1.30%	0.179	0.052	0.000	5.326	7.614	430.834	14.7%
Taiwan	1,274	1,969	3.10%	0.000	0.000	0.000	0.032	1.011	33.400	9.0%
Thailand	885	1,440	1.60%	0.003	0.003	-0.001	0.169	2.932	144.274	11.1%
Turkey	371	2,561	1.20%	0.004	0.003	0.000	0.116	1.217	168.673	19.4%
UK	5,398	1,847	12.40%	0.141	0.009	0.000	3.678	6.907	461.158	44.4%
Venezuela	64	960	0.10%	0.017	0.009	0.006	0.081	2.254	80.448	13.6%
Zimbabwe	2	444	0.00%	0.029	0.029	0.000	0.401	5.029	139.258	28.1%

Table 2

Properties of day zero abnormal returns with no abnormal performance induced

The combined simulated event-study samples contain 250,000 trading days for ordinary non-U.S. stocks from 1988-2006; price and dividend data come from Datastream. Each daily stock return is calculated from the previous trading day having a non-missing price to the current trading day, including dividends. No return is calculated on a day with a missing price. Trade-to-trade returns consist of calculated returns from non-missing price days. For a stock with a missing price, the corresponding market return is added to the market return on the next non-missing price day for trade-to-trade abnormal return calculation. Lumped returns consist of trade-to-trade returns on non-missing price days and zero on missing price days. The market index for market model abnormal returns is the country-specific Datastream Global Index (level one). The market model is estimated by ordinary least squares. Jarque-Bera statistic tests normality; in every case, the null hypothesis of a normal distribution is rejected at the 0.01% significance level based on either the chi-square with two degrees of freedom or the critical values in Wuertz and Katzgraber (2009).

N	Median	Mean	Standard deviation	Skewness	Excess kurtosis	Jarque-Bera statistic
Panel A: Trade-to-trade returns – individual securities						
250,000	–0.001	-0.002	0.199	7.798	41,570	2.22×10^{13}
Panel B: Lumped returns – individual securities						
250,000	–0.001	-0.002	0.195	14.250	46,116	1.80×10^{13}
Panel C: Trade-to-trade returns – 250-stock portfolios						
1,000	-0.002	-0.002	0.013	0.137	163	11,06,160
Panel D: Lumped returns – 250-stock portfolios						
1,000	-0.002	-0.002	0.012	0.392	178	1,322,059

Table 3
Day zero rejection rates in 1,000 samples of 250 non-U.S. stocks each, 1988-2006

The stocks are ordinary share issues; data come from Datastream. We randomly sample from the available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. When no trade price is reported, the stock's trade-to-trade abnormal return is set to missing and, for abnormal return calculation, the market return is added to the market return on the next non-missing price day. A lumped return is identical to the trade-to-trade return when there is no missing price on the current or previous trading day; when there is a missing price, the lumped return is zero and the market return is not adjusted. To simulate a stock-price reaction, we add a seeded return to the stock return on the selected event date (day zero). The market index is the country-specific "Total Market" index of the Datastream Global series, also called a level one index; the indexes are value weighted. The market model is estimated by OLS; market-model abnormal returns are prediction errors. The estimation period, for signs, standard deviations, and market model parameters, is trading days –256 through –6 relative to the event; ranks for the rank test incorporate days –256 through +6. The null hypothesis of the standardized cross-sectional (Std. csect.) and time-series portfolio standard deviation (CDA) tests is that the mean abnormal return on day 0 is zero. The null hypothesis of the generalized sign test (GST) is that the fraction of day 0 abnormal returns having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank test is that the mean rank of day zero is equal to that of the estimation period. Alternative hypotheses are one tailed. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns based on trade-to-trade returns								
Std. csect.	0.948	0.941	0.874	0.040	0.052	0.916	0.949	0.949
CDA	0.519	0.202	0.081	0.005	0.018	0.094	0.204	0.484
GST	1.000	1.000	1.000	0.053	0.043	1.000	1.000	1.000
Rank	1.000	1.000	1.000	0.046	0.041	1.000	1.000	1.000
Panel B: Market-model abnormal returns based on lumped returns								
Std. csect.	0.948	0.941	0.873	0.042	0.050	0.919	0.949	0.949
CDA	0.500	0.193	0.070	0.006	0.017	0.087	0.186	0.465
GST	1.000	1.000	0.999	0.042	0.068	1.000	1.000	1.000
Rank	1.000	1.000	1.000	0.049	0.042	1.000	1.000	1.000

Table 4

Three-day window rejection rates in 1,000 samples of 250 non-U.S. stocks each, 1988-2006

The stocks are ordinary share issues; data come from Datastream. We randomly sample from the available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. When no trade price is reported, the stock's trade-to-trade abnormal return is set to missing and, for abnormal return calculation, the market return is added to the market return on the next non-missing price day. A lumped return is identical to the trade-to-trade return when there is no missing price on the current or previous trading day; when there is a missing price, the lumped return is zero and the market return is not adjusted. To simulate a stock-price reaction, we add a seeded return to the stock return on the selected event date (day zero). The market index is the country-specific "Total Market" index of the Datastream Global series, also called a level one index; the indexes are value weighted. The market model is estimated by OLS; market-model abnormal returns are prediction errors. The abnormal returns of trading days (–1,+1) are added to create the three-day window cumulative abnormal return (CAR). The estimation period, for signs, standard deviations, and market model parameters, is trading days –256 through –6 relative to the event; ranks for the rank test incorporate days –256 through +6. The null hypothesis of the standardized cross-sectional test (Std. csect.) and time-series portfolio standard deviation (CDA) tests is that the mean CAR is zero. The null hypothesis of the generalized sign test reported as GST is that the fraction of CARs having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank test is that the mean rank in the event window is equal to that of the estimation period. Alternative hypotheses are one tailed. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns based on trade-to-trade returns								
Std. csect.	0.946	0.904	0.603	0.057	0.057	0.668	0.933	0.946
CDA	0.350	0.102	0.025	0.003	0.021	0.038	0.112	0.336
GST	1.000	0.998	0.861	0.036	0.061	0.940	1.000	1.000
Rank	1.000	1.000	0.939	0.051	0.043	0.939	0.999	1.000
Panel B: Market-model abnormal returns based on lumped returns								
Std. csect.	0.946	0.905	0.610	0.057	0.054	0.651	0.931	0.944
CDA	0.340	0.108	0.027	0.002	0.016	0.033	0.099	0.314
GST	1.000	0.996	0.822	0.025	0.078	0.948	1.000	1.000
Rank	1.000	0.999	0.943	0.047	0.046	0.939	0.999	1.000

Table 5

Eleven-day window rejection rates in 1,000 samples of 250 non-U.S. stocks each, 1988-2006

The stocks are ordinary share issues; data come from Datastream. We randomly sample from the available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. When no trade price is reported, the stock's trade-to-trade abnormal return is set to missing and, for abnormal return calculation, the market return is added to the market return on the next non-missing price day. A lumped return is identical to the trade-to-trade return when there is no missing price on the current or previous trading day; when there is a missing price, the lumped return is zero and the market return is not adjusted. To simulate a stock-price reaction, we add a seeded return to the stock return on the selected event date (day zero). The market index is the country-specific "Total Market" index of the Datastream Global series, also called a level one index; the indexes are value weighted. The market model is estimated by OLS; market-model abnormal returns are prediction errors. The abnormal returns of trading days –5 through +5 are added to create the 11-day window cumulative abnormal return (CAR). The estimation period, for signs, standard deviations, and market model parameters, is trading days –256 through –6 relative to the event; ranks for the rank test incorporate days –256 through +6. The null hypothesis of the standardized cross-sectional (Std. csect.) and time-series portfolio standard deviation tests (CDA) is that the mean CAR is zero. The null hypothesis of the generalized sign test reported as GST is that the fraction of CARs having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank test is that the mean rank in the event window is equal to that of the estimation period. Alternative hypotheses are one tailed. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns based on trade-to-trade returns								
Std. csect.	0.927	0.627	0.277	0.044	0.060	0.295	0.687	0.937
CDA	0.213	0.024	0.007	0.000	0.054	0.060	0.071	0.213
GST	1.000	0.744	0.325	0.039	0.058	0.466	0.870	1.000
Rank	1.000	0.818	0.509	0.036	0.036	0.431	0.781	0.998
Panel B: Market-model abnormal returns based on lumped returns								
Std. csect.	0.929	0.650	0.304	0.052	0.054	0.267	0.643	0.938
CDA	0.208	0.045	0.010	0.001	0.053	0.055	0.061	0.200
GST	1.000	0.705	0.289	0.029	0.071	0.490	0.873	1.000
Rank	1.000	0.817	0.516	0.037	0.034	0.421	0.776	0.998

Table 6
Rejection rates in 1,000 samples of 250 non-U.S. stocks each with a stock-return variance increase on day zero, 1988-2006

The stocks are ordinary share issues; data come from Datastream. We randomly sample from the available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. To simulate a stock-price reaction, we add a seeded return to the stock return on the selected event date (day 0). To simulate a doubling of variance on day zero, we use the approach of Brown and Warner (1985). Stock returns are trade-to-trade. The market index is the country-specific total market index (level one index) of the Datastream Global series, which is value weighted. The market model is estimated by OLS; market-model abnormal returns are prediction errors. The abnormal returns of three or 11 trading centered on day zero are added to create window cumulative abnormal returns (CAR). The estimation period, for signs, standard deviations, and market model parameters, is trading days –256 through –6 relative to the event; ranks for the rank test incorporate days –256 through +6. The null hypothesis of the standardized cross-sectional (Std. csect.) and time-series portfolio standard deviation (CDA) tests is that the mean day zero abnormal return or mean CAR is zero. The null hypothesis of the generalized sign test reported as GST is that the fraction of day zero abnormal returns or CARs having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank and standardized rank tests is that the mean rank in the event window is equal to that in the estimation period. Alternative hypotheses are one tailed. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns, event day zero								
Std. csect.	0.948	0.935	0.729	0.055	0.053	0.771	0.943	0.948
CDA	0.539	0.214	0.090	0.033	0.016	0.107	0.210	0.471
GST	1.000	1.000	0.891	0.029	0.098	0.981	1.000	1.000
Rank	1.000	1.000	0.998	0.237	0.017	0.956	1.000	1.000
Std. rank	0.992	0.983	0.921	0.042	0.038	0.967	1.000	1.000
Panel B: Market-model abnormal returns, three-day event window (–1,+1)								
Std. csect.	0.946	0.878	0.510	0.054	0.050	0.547	0.913	0.945
CDA	0.361	0.105	0.033	0.026	0.006	0.050	0.114	0.336
GST	1.000	0.971	0.609	0.014	0.077	0.860	0.998	1.000
Rank	1.000	0.996	0.875	0.140	0.017	0.689	0.980	1.000
Std. rank	0.987	0.879	0.547	0.039	0.044	0.632	0.941	0.998
Panel C: Market-model abnormal returns, eleven-day event window (–5,+5)								
Std. csect.	0.929	0.592	0.268	0.043	0.061	0.273	0.637	0.937
CDA	0.208	0.026	0.009	0.058	0.000	0.063	0.072	0.219
GST	1.000	0.672	0.306	0.034	0.068	0.428	0.832	1.000
Rank	0.995	0.712	0.434	0.064	0.022	0.236	0.553	0.990
Std. rank	0.865	0.402	0.207	0.039	0.056	0.258	0.456	0.923

Table 7

Country clustering: Rejection rates in 1,000 single-country samples of 250 stocks each, 1988-2006

Each sample contains stocks (ordinary share issues) from a single non-U.S. market; data come from Datastream. We randomly select a market and randomly sample from its available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. Sampling is with replacement. The null hypothesis of the standardized cross-sectional (Std. csect.) and time-series portfolio standard deviation (CDA) tests is that the mean day zero abnormal return or mean CAR is zero. The null hypothesis of the generalized sign test reported as GST is that the fraction of day zero abnormal returns or CARs having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank test is that the mean rank in the event window is equal to that of the estimation period. The estimation period, for signs, standard deviations, and market model slope and intercept, is trading days –256 through –6 relative to the event. The market index for market model abnormal returns is the country-specific Datastream Global Index (level one). Market model return is stock return minus the market index return. The market model is estimated by ordinary least squares. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% binomial success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns, event day zero								
Std. csect.	0.964	0.949	0.812	0.049	0.048	0.870	0.961	0.965
CDA	0.887	0.740	0.512	0.030	0.046	0.523	0.746	0.887
GST	1.000	1.000	0.964	0.065	0.052	0.986	1.000	1.000
Rank	1.000	1.000	0.985	0.043	0.043	0.987	1.000	1.000
Panel B: Market-model abnormal returns, three-day event window (–1,+1)								
Std. csect.	0.960	0.882	0.586	0.063	0.062	0.636	0.918	0.962
CDA	0.845	0.557	0.225	0.032	0.039	0.266	0.567	0.841
GST	1.000	0.955	0.740	0.046	0.061	0.823	0.990	1.000
Rank	1.000	0.987	0.803	0.044	0.051	0.819	0.988	1.000
Panel C: Market-model abnormal returns, eleven-day event window (–5,+5)								
Std. csect.	0.941	0.616	0.327	0.081	0.079	0.336	0.627	0.953
CDA	0.742	0.280	0.132	0.064	0.046	0.138	0.245	0.707
GST	0.984	0.670	0.374	0.058	0.097	0.470	0.783	1.000
Rank	0.990	0.734	0.459	0.042	0.047	0.428	0.685	0.982

Table 8

Rejection rates in the most concentrated non-U.S. stock markets, 1,000 samples of 250 non-U.S. stocks each.

Each sample contains stocks (ordinary share issues) from the ten most concentrated non-U.S. stock markets in 1988–1991, 1992–1994, 1995–1997, 1998–2000, 2001–2003, and 2004–2006. To determine the most concentrated markets in each period, we calculate a Herfindahl index based on each stock's number of shares traded times closing price from Datastream. We pool the data for the identified markets from all periods and randomly sample, with replacement, from the available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. The null hypothesis of the standardized cross-sectional (Std. csect.) is that the mean day zero abnormal return or mean CAR is zero. The null hypothesis of the generalized sign test reported as GST is that the fraction of day zero abnormal returns or CARs having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank test is that the mean rank in the event window is equal to that of the estimation period. The estimation period, for signs, standard deviations, and market model slope and intercept, is trading days –256 through –6 relative to the event. Returns are trade-to-trade. The market index for market model abnormal returns is the country-specific Datastream Global Index (level one). The market model is estimated by ordinary least squares. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% binomial success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns, event day zero								
Std. csect.	0.877	0847	0.756	0.069	0.038	0.799	0870	0.886
GST	1.000	1.000	0.999	0.044	0.043	1.000	1.000	1.000
Rank	1.000	1.000	1.000	0.044	0.042	1.000	1.000	1.000
Panel B: Market-model abnormal returns, three-day event window (–1,+1)								
Std. csect.	0.819	0.743	0.521	0.059	0.036	0.458	0.797	0.848
GST	1.000	1.000	0.943	0.055	0.029	0.943	1.000	1.000
Rank	1.000	1.000	0.967	0.043	0.044	0.958	1.000	1.000
Panel C: Market-model abnormal returns, eleven-day event window (–5,+5)								
Std. csect.	0.632	0.373	0.195	0.056	0.026	0.181	0.407	0.720
GST	1.000	0.869	0.535	0.051	0.030	0.554	0.890	1.000
Rank	0.993	0.800	0.507	0.051	0.044	0.482	0.770	0.995

Table 9

Rejection rates for markets with the most non-normally distributed returns, 1,000 samples of 250 non-U.S. stocks each

Each sample contains stocks (ordinary share issues) randomly selected with replacement from the ten non-U.S. stock markets where stock return distributions deviate most from normality in 1988–1991, 1992–1994, 1995–1997, 1998–2000, 2001–2003, and 2004–2006. To determine the most non-normal markets, we calculate the Jarque-Bera test statistic for non-normality, J, over each period for each stock that has at least 100 trading days of non-missing returns in the period, and rank markets by median J. We pool the data for the identified markets from all periods and randomly sample, with replacement, from the available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. The null hypothesis of the standardized cross-sectional (Std. csect.) is that the mean day zero abnormal return or mean CAR is zero. The null hypothesis of the generalized sign test reported as GST is that the fraction of day zero abnormal returns or CARs having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank test is that the mean rank in the event window is equal to that of the estimation period. The estimation period, for signs, standard deviations, and market model slope and intercept, is trading days –256 through –6 relative to the event. Returns are trade-to-trade. The market index for market model abnormal returns is the country-specific Datastream Global Index (level one). The market model is estimated by ordinary least squares. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% binomial success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns, event day zero								
Std. csect.	0.959	0.956	0.919	0.051	0.044	0.941	0.959	0.964
GST	1.000	1.000	1.000	0.033	0.053	1.000	1.000	1.000
Rank	1.000	1.000	1.000	0.048	0.053	1.000	1.000	1.000
Panel B: Market-model abnormal returns, three-day event window (–1,+1)								
Std. csect.	0.840	0.572	0.253	0.042	0.053	0.305	0.682	0.888
GST	1.000	1.000	0.994	0.039	0.041	0.999	1.000	1.000
Rank	1.000	1.000	1.000	0.039	0.064	1.000	1.000	1.000
Panel C: Market-model abnormal returns, eleven-day event window (–5,+5)								
Std. csect.	0.943	0.921	0.728	0.057	0.040	0.745	0.930	0.950
GST	1.000	0.902	0.526	0.040	0.052	0.607	0.948	1.000
Rank	0.998	0.924	0.727	0.047	0.098	0.857	0.970	1.000

Table 10

Rejection rates with market-moving events in 1,000 concentrated-market samples of 250 non-U.S. stocks each

Each sample contains stocks (ordinary share issues) from the ten most concentrated non-U.S. stock markets in 1988–1991, 1992–1994, 1995–1997, 1998–2000, 2001–2003, and 2004–2006. To determine the most concentrated markets in each period, we calculate a Herfindahl index based on each stock's number of shares traded times closing price from Datastream. We pool the data for the identified markets from all periods and randomly sample, with replacement, from the available stock-listing day combinations; a listing day is when the market is open and the stock is listed for trading, subject to data availability screens. To simulate market-moving events, we find f_{MV}, the four-week moving average ratio, on day zero, of each stock's market value to the total value of stocks in its market. We multiply the seeded return by the stock's f_{MV} and add the product to the market index return before calculating the stock's abnormal return. The null hypothesis of the standardized cross-sectional (Std. csect.) is that the mean day zero abnormal return or mean CAR is zero. The null hypothesis of the generalized sign test reported as GST is that the fraction of day zero abnormal returns or CARs having a particular sign is equal to the fraction of estimation-period abnormal returns with that sign. The null hypothesis of the rank test is that the mean rank in the event window is equal to that of the estimation period. The estimation period, for signs, standard deviations, and market model slope and intercept, is trading days –256 through –6 relative to the event. Returns are trade-to-trade. The market index for market model abnormal returns is the country-specific Datastream Global Index (level one). The market model is estimated by ordinary least squares. The 95% confidence limits for the normal approximation to 1,000 binomial trials with a 5% probability of success (rejection) on each trial are 3.7% to 6.4%; the 99% limits, still with a 5% binomial success probability, are 3.3% to 6.8%.

	Seeded return							
	–3%	–1%	–0.5%	0%	0%	0.5%	1%	3%
Test	Lower-tailed rejection rates				Upper-tailed rejection rates			
Panel A: Market-model abnormal returns, event day zero								
Std. csect.	0.872	0.850	0.748	0.084	0.026	0.779	0.862	0.879
GST	1.000	1.000	1.000	0.037	0.045	1.000	1.000	1.000
Rank	1.000	1.000	1.000	0.048	0.042	1.000	1.000	1.000
Panel B: Market-model abnormal returns, three-day event window (–1,+1)								
Std. csect.	0.817	0.741	0.498	0.063	0.028	0.457	0.791	0.842
GST	1.000	1.000	0.925	0.064	0.035	0.950	1.000	1.000
Rank	1.000	1.000	0.945	0.049	0.052	0.954	0.998	1.000
Panel C: Market-model abnormal returns, eleven-day event window (–5,+5)								
Std. csect.	0.614	0.367	0.159	0.040	0.036	0.203	0.452	0.693
GST	1.000	0.861	0.537	0.049	0.036	0.557	0.876	1.000
Rank	0.998	0.777	0.475	0.052	0.041	0.467	0.772	0.996