Short-Horizon Event Study Estimation with a STAR Model and Real Contaminated Events

Panayiotis C. Andreou[a,b], Christodoulos Louca[a,b] and Christos S. Savva[a,c]

[a] Department of Commerce, Finance and Shipping, Cyprus University of Technology, 3036 Lemesos, Cyprus
[b] Durham University Business School, Mill Hill Lane, DH1 3LB, Durham, UK
[c] Centre for Growth and Business Cycle Research, University of Manchester, Manchester, United Kingdom

This version: February 2015

Abstract

We propose a test statistic for nonzero mean abnormal returns based on a Smooth Transition Auto Regressive (STAR) model specification. Estimation of STAR takes into account the probability of contaminated events that could otherwise bias the parameters of the market model and thus the specification and power of the test statistic. Using both simulated and real stock returns data from mergers and acquisitions, we find that the STAR test statistic is robust to contaminated events occurring in the estimation window and in the presence of event-induced increase in return variance. Under the STAR test statistic the true null hypothesis is rejected at appropriate levels. Moreover, it exhibits the highest levels of power when compared with other test statistics that are widely and routinely applied in short-horizon event studies.

JEL classification: G14; G34
Keywords: Event studies, test statistics, contaminated events, Markov switching regression model, Smooth Transition Auto Regressive model.

Corresponding author:
Panayiotis Andreou
Email: benz@pandreou.com
Cyprus University of Technology
Dept. of Commerce, Finance and Shipping
115, Spyrou Araouzou Street
3036 Lemesos, Cyprus
Tel.: +357 25002286
Fax.: +357 25002816

1. Introduction

The short-horizon event study method, introduced in the seminal work of Fama, Fisher, Jensen and Roll (1969), has been one of the cornerstones of financial economics and accounting in the last few decades. Ever since, numerous research papers have endeavoured improvements in the basic empirical methodology.[1] The most notable recent attempts focus on the introduction of test statistics for nonzero mean abnormal returns that are robust to event-induced increase in return variance (see, for instance, Harrington and Shrider 2007; Kolari and Pennönen 2010 and references therein).[2] More recently, Aktas et al. (2007a) emphasize the need to consider event study methods that mitigate the effect of contaminated (unrelated) events arising from corporate actions and announcements that may occur during the estimation window. It is reasonable to expect that company press releases or leakage of private information occurring in the estimation window could create cross-sectional variation in the abnormal returns. This would inevitably bias the estimation of the (true) return-generating process parameters, in particular, the estimated variance of the parameters which could deteriorate the detection of abnormal performance in the event window.

In this study, we sought to make a dual contribution to the empirical corporate finance research. First, to tackle the estimation window contaminating-event problem, we estimate the widely applied event study market model as introduced by Sharpe (1963) by relying to regime switching approaches as a general method, highlighting at the same time the

[1] There is a vast amount of theoretical and empirical studies in the research realm of this topic such the ones of Ball and Torous (1988), Corrado (1989), Boehmer et al. (1991), Salinger (1992), Savickas (2003), Dombrow et al. (2000), Cyree and DeGennaro (2002), Harrington and Shrider (2007), Ahern (2009), Campbell et al. (2010) and Kolari and Pynnönen (2010), among others. The landmark work in this topic is that by Brown and Warner (1980, 1985) who investigate the specification (Type I error – rejecting the null when it is true) and power (Type II error – failing to reject the null when the alternative hypothesis is true) of several modifications of the short-horizon event study by assuming that abnormal returns are intertemporally uncorrelated and there is no (significant) impact of event-induced variance.

[2] Abnormal return (*AR*) is defined to be the difference between the *actual* return that is observed during the event day(s) (namely, the event window) and the *expected* return which is provided from a return-generating model estimated using stock returns data that precede the event (namely, the estimation window). Event-induced increase in return variance occurs when variance in the event window exceeds the variance over the estimation window; as a consequence, test statistics that ignore plausible implications of unexplained variation in true abnormal returns for the structure of heteroskedasticity may fail to detect event-related abnormal performance.

importance of the Smooth Transition Auto Regressive (STAR) specification; this is the first time that the STAR model is utilized to compute an event study test statistic for the detection of nonzero mean abnormal returns. STAR can be viewed as a statistical method that filters out firm-specific events that could otherwise induce unduly variance in the model's generated returns (see Hansen 2011 for a detailed review in threshold autoregressive models and their applications). STAR is a regime-switching model that allows for two regimes, associated with extreme values of the transition function, where the transition from one regime to the other is done in a smooth or abrupt manner (see Terasvirta and Anderson 1992; Terasvirta 1994, among others). We employ the STAR method to better model the stock returns data generating mechanism by taking into account the probability of the occurrence of unrelated firm events. In this respect, estimated parameters of the market model should be less subject to the influence of contaminating events in the estimation period compared to more traditional rival choices. Further, the findings of this study highly support the application of the STAR specification in the event study framework since it fits the empirical stock returns data generating process much better than any other rival method. Thus, the STAR event study test statistic could allow the researcher to conduct valid large scale statistical analysis of abnormal returns that are much better specified, as well as more powerful in detecting the (true) size of the abnormal returns around a particular corporate announcement.

Second, we consider biases arising in short-horizon event study of test statistics for nonzero mean abnormal returns using *real* data from Mergers and Acquisitions (M&As) for the period 1980-2010. The majority of prior literature focuses on providing analytical and empirical evidence of the resulting test statistics biases using *randomly* selected firms with *simulated* induced contaminated events during the estimation period (e.g., Aktas et al. 2007a; Harrington and Shrider 2007). Nevertheless, by carrying out specification and power tests on estimation periods using simulated returns data may not always be representative enough of the resulting cross-sectional variation in abnormal returns emerging from event-contamination that is taking place in real situations. In the context of M&As, which by selection reflects a *non-random* sample (e.g., Fuller et al. 2002; Bhagat et al. 2005; Guo and Petmezas 2012), it is quite frequent for bidding firms to engage in several other unrelated corporate activities (e.g., earnings announcement, changes to dividend policy, etc) in the period preceding the deal announcement (see also, Bhagat et al. 2005; Aktas et al. 2009). In a vast amount of M&A deals, the number of major corporate events that may emerge in the estimation window is rather high: for example, Fuller et al. (2002) study a M&As sample where bidding firms complete bids for five or more targets within a three-year window. Moreover, the *nature* and *duration* of an event may not be captured accordingly with simulated contaminated stock-return series. Company press releases or leakage of private information usually happens rather close to the event announcement (e.g., Harrington and Shrider 2007; Kolari and Pynnönen 2010; Guo and Petmezas 2012). Therefore, any contaminated events are more likely to cluster in a non-random manner in the period just before the announcement day (implying, for instance, the need of a right-skewed distribution for capturing the arrival of news in the market). Lastly, the widely-adopted simulation approach of Brown and Warner (1980, 1985) does not consider the possible endogeneity of announcement decisions in the presence of private or market information. Harford (2005), for example, documents that M&As cluster in time due to the market timing of industry shocks. There is also evidence to support that market and firm-specific news-releases (Délèze and Hussain 2014; Laopodis 2010) and contagion effects caused by financial crises (Kenourgios et al. 2013) cause simultaneous price movements to different asset classes across different markets. This increases significantly the likelihood to observe great overlap on the event dates which could introduce contemporaneous correlations in the abnormal returns leading to incorrect inferences regarding the detection of abnormal returns.[3] All abovementioned cases may be rather compelling to be modelled properly under the simulated data set environment, especially when corporate event announcements occur in extreme market conditions where, for instance, firm-specific mean stock return or volatility are particularly high. Therefore assessing any aberration between *simulated* and *real* events that could impact the specification and power of event study test statistics remains an open question which we investigate in this paper.

From the stand-point of both the academic researcher and the investment practitioner, it is crucial to know which cross-sectional abnormal return test statistic(s) should be employed to make inferences; this is especially important for many corporate events such as M&As that represent huge deals and associate with enormous market dollar values.[4] To better answer this question under a general viewpoint, we undertake a horserace of return-generating process and statistical tests. We identify the one(s) that are best suited using both, simulated stock returns in accordance to the traditional approach (Brown and Warner 1980 and 1985; Boehmer et al. 1991), as well as real stock returns coming from M&A deals. We focus on the most prominent cases of methods proposed so far in the literature. First, we include approaches that have been extensively used in prior empirical studies like the standardized cross-sectional test proposed by Boehmer et al. (1991), RANK test proposed by Corrado (1989) and GARCH test proposed by Savickas (2003). Second, we consider methods having greater flexibility to mitigate the presence of firm-specific (unrelated) events that precede corporate announcements like the BETA-1 approach, the two-state Markov-switching market model test (TSMM) proposed by Aktas et al. (2007a) and the STAR test

[3] Cam and Ramiah (2014) also discuss the possibility that researchers may reach different results depending on the financial econometrics adjustments and asset pricing model used when calculating expected returns.

[4] Although we focus our analysis on M&As as a major corporate event, our inferences could easily be generalized for any other corporate decisions that exhibit similar market performance, such as season equity offerings, share repurchase, goodwill write offs, cross-listings, etc.

statistic introduced in this study. For the TSMM, we do not only consider two-state variance regimes as in Aktas et al. (2007a), but we also investigate two-state regimes in the market model parameters (i.e., stock return mean regression equation).[5] Unlike their peers, the regime-switching family comprised by the TSMM and STAR models, postulate an adaptation of the event study methodology that automatically takes into consideration the probability of contaminated events that could otherwise bias the estimated parameters of the market model (affecting in this way the specification and power of test statistics). Therefore, they should exhibit superior robustness to event-induced increase in return variance caused by the cross-sectional variation in the effects of a firm-specific event occurring in the estimation window. In this respect, our null expectation is that test statistics computed from the regime-switching family, in particular from the STAR model, would perform significantly better than their rival methods.

Our results show that the traditional test statistics employed by prior researchers in the short-horizon event study are mis-specified and exhibit weak statistical power in the presence of contaminating events in the estimation period, especially in the presence of event-induced increase in return variance. On the contrary, the STAR methodology introduced in this study is the best choice since it is resilient in any type (simulated or real) of firm-specific contamination that may occur; moreover, its statistical power is high even under severe event-induced increase in return variance. The Markov switching regression models as proposed in Aktas et al (2007a) are found to be the second best choice, although not too behind from the performance of the STAR model. Nonetheless, with real stock returns sub-samples which present either extreme mean stock returns or extreme stock return volatility, we find a clear superiority of the STAR method all over other rival ones. Previous research

[5] Letting the market model parameters being regime-dependent allow a more realistic representation of the return-generating mechanism since prior empirical research has revealed a significant time-variation in the slope parameter which depends on rising and falling market conditions (Hays and Upton 1986; Klein and Rosenfeld 1987; Chang and Weiss 1991; Chiang et al. 2013).

has largely overlooked the possibility that there may be significant aberrations between simulated and real events that could induce unduly variance in the estimation window that inevitably could impact the power of the test statistics.[6] Therefore, the overall observed superiority of the STAR event study test statistic, especially under extreme market conditions, is evidence to support the use of this method in future empirical work relying on short-horizon event studies.

The paper is organized as follows. Section 2 presents our experimental design including the data we use, the simulation setup we follow regarding the event study return-generating process and relevant test statistics that we investigate. Section 3 discusses the performance results from the simulated and the real stock-returns data. Section 4 concludes the study.

2. Methods and experimental design

Following the seminal work by Brown and Warner (1985), previous researchers generate simulated random shocks in the estimation window in two steps. First, firms are randomly picked from the universe of stocks included in the Center for Research in Security Prices (CRSP) database. Second, an extraordinary event is simulated for each firm and randomly introduced in the real stock price series in the estimation window. Such a contaminated event is necessary to artificially induce variance in the estimation parameters of the market model.

[6] Campbell et al. (2001) document a noticeable increase in firm-level volatility relative to market volatility over the period from 1962 to 1997 which is associated with a decline in the explanatory power of the market model (see also empirical evidence in Aktas et al. 2007b, as well as Arora et al. 2009 for emerging markets). Kothari and Warner (2007) note that this is highly relevant on the implication behind the event study because it suggests a time-variation to the power of test statistics to detect abnormal performance for certain events. Studies that rely on purely simulated variance-induced events may fail to properly capture such stylized (structured) patterns in the returns-generating mechanism.

Our empirical analysis utilizes two sources of data return-generating processes to investigate the specification and power of test statistics. First, we follow the traditional event study approach proposed by Brown and Warner (1980, 1985) to simulate abnormal performance and variance increase event days. Since then, a notable amount of other research studies have adopted this approach (see, for instance, Corrado 1989; Corrado and Zivney 1992; Boehmer et al. 1991; Savickas 2003; Aktas et al. 2009; and Kolari and Pynnönen 2010). The traditional approach relies solely on simulated data to construct the event-induced variance increase in the return-generating process. In addition, we follow the notion postulated by Aktas et al. (2007a) to contaminate the event study estimation window.

Second, we repeat all specification and power tests using a large sample of M&As to investigate the impact of estimation period contamination under the influence of unrelated events and of event-induced increase in return variance that may emerge naturally for firms that engage in this particular corporate activity. It is of vital importance to investigate whether assertions regarding the specification and power test statistics obtained with simulated contaminated returns also remain unchanged with real returns on which firm-specific events emerge naturally and reflect news arrival in a purely stochastic manner (i.e., in accordance with the efficient market hypothesis). M&A announcements reflect real-event contamination that naturally induce variance in the estimation window. In this respect, we inherently take into account both, the cross-sectional variation relating to the underlying economic effects of a real event, as well as any structure in heteroskedasticity arising from these events.[7] Hence, by investigating the performance of test statistics using a large data set of M&As we avoid any (unrealistic) assumptions imposed when contamination is done in an artificial manner.

[7] Both of these elements are deemed important since Harrington and Shrider (2007) identify them to be "troubling features" of the statistical tests reported in many prior studies. In addition, the real return-generating process would allow many different sorts of unrelated events to affect the estimation period, revealing which tests are robust when employed with non-simulated data.

2.1. Event study method

2.1.1. Return generating processes

A classical approach for abnormal returns is based on the market model introduced by Sharpe (1963):

$$R_{j,t} = a_j + \beta_j R_{m,t} + \varepsilon_{j,t} \quad (1)$$

where,

– α_j and β_j are the coefficient estimates for firm j;

– R_j and R_m are vectors of returns for firm j and for a market portfolio proxy m on day t, respectively; and the residuals are supposed to be independent and identically distributed and capture the abnormal behaviour.

An alternative way to obtain event-day predictions and standard errors is by employing the market model augmented by a dummy variable to capture the effects of the event (see, Karafiath 1988; Salinger 1992):

$$R_{j,t} = a_j + \beta_j R_{m,t} + \gamma_j D_{j,t} + \varepsilon_{j,t} \quad (2)$$

where,

– D_j is a dummy variable equal to 1 at the event window for firm j, and 0 otherwise;

The coefficient γ captures the abnormal return which in essence is the forecast error of what is expected to observe using a normal return-generating model compared to what is really observed during the event window. This model however, as shown by Aktas et al. (2007a), is problematic and its OLS solution overestimates the standard error of an individual firm's abnormal returns when the true return-generating process has two-states (see also discussions in Salinger 1992). This is mainly attributed to the fact that the variance covariance matrix is rather a state dependent and no longer homoskedastic. Therefore they suggest a model that captures a low and high variance regime. Hence, building on Aktas et al.

(2007a), we incorporate regime dependent intercepts and slope coefficients in the mean specification as follows:

$$\begin{aligned} &R_{j,t} = a_{j,S} + \beta_{j,S} R_{m,t} + \gamma_j D_{j,t} + \varepsilon_{j,S,t} \\ &\varepsilon_{j,S,t} \sim N\left(0, \sigma^2_{j,S}\right) \end{aligned}, \qquad (3)$$

where S is a state variable, with S = 1 for the low regime state and S = 2 for the high regime state. More specifically, parameters $\alpha_{j,S}$ and $\beta_{j,S}$ allow to explicitly incorporate the presence of contaminated events into the mean specification of the model. Similar intuition applies for the variance where the high variance state is greater than the low variance state ($\sigma^2_{j,2} > \sigma^2_{j,1}$).[8]

As for the way the transition between the two regimes is governed, we follow methodology and notations as in Hamilton (1994). More specifically the transition between the two regimes is governed the by a Markov chain of order 1, for which the transition matrix is given by:

$$P = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix} \qquad (4)$$

where $p_{kl}=p(S_t = k|S_{t-1} = l)$ corresponds to the probability of changing from state l to state k with the unconditional probability of the regime given by:

$$\begin{aligned} p\left(S_t = 1\right) &= \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \\ p\left(S_t = 2\right) &= \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \end{aligned} \qquad (5)$$

In addition, we introduce a STAR model specification that has the capacity to capture the state dependent generating process of stock returns. This model can be thought as an extension of the autoregressive models allowing for changes in parameters according to the

[8] Salinger (1992) also discusses deficiencies of the traditional approach on the estimation of the abnormal returns variance when the market model parameters are not stable and which could lead to incorrect inferences about the detection of abnormal returns.

value of a transition variable. More specifically, the market STAR model specification can be presented as follows:

$$R_{j,t} = (\alpha_j^{(1)} + \beta_j^{(1)} R_{m,t}) G(z_t, \zeta, c) + (\alpha_j^{(2)} + \beta_j^{(2)} R_{m,t})(1 - G(z_t, \zeta, c)) + \gamma_j D_{j,t} + \varepsilon_{j,t}^{(i)}, \quad \varepsilon_{j,t}^{(i)} \sim N(0, \sigma_j^{2,(i)}) \tag{6}$$

where i = 1 for the low state of the market i = 2 for the high state of the market. To assess whether the effects on the returns vary with the state of the market, we employ a continuous transition function $G(z_t, \zeta, c)$, which changes smoothly from 0 to 1, as the transition variable z_t increases. A popular choice is the logistic function:[9]

$$G(z_t, \zeta, c) = \frac{1}{1 + \exp(-\zeta (z_t - c))}. \tag{7}$$

In practice the appropriate transition variable z_t is unknown, however a good choice is to use lagged endogenous variables (in our case, $R_{j,t-1}$). This is also supported by the LM-type statistic (see Terasvirta 1994, for further details) which was conducted during the analysis and supports that the first lag of the dependent variable is the best choice. Therefore, at any given point in time the evolution of $R_{j,t}$ is determined by a weighted average of two different regression models. The weights assigned to the two models depend on the value taken by the transition variable z_t. For small (large) values of z_t, $G(z_t, \zeta, c)$ is approximately equal to zero (one) and, hence, almost all weight is put on the first (second) part of the model.

The parameter c, denotes the threshold between the two regimes corresponding to $G(z_t, \zeta, c)$=0 and $G(z_t, \zeta, c)$=1, in the sense that the logistic function changes monotonically from 0 to 1 as z_t increases.[10]

[9] This logistic form has been widely used for smooth transition models. For further details we refer to Terasvirta and Anderson (1992), Terasvirta (1994), and van Dijk and Franses (1999).

[10] The starting values of ζ_i and c_i (with ζ_i >0) are determined by a grid search and are estimated in one step by maximizing the likelihood function while the threshold point between the states is estimated by the model.

The parameter ζ determines the speed at which the weights between the two parts of the specification change as z_t increases; the higher ζ, the faster is this change. If $\zeta \rightarrow 0$, the weights become constant (equal to 0.5) and the model becomes linear, whereas, if $\zeta \rightarrow \infty$, the logistic function approaches a Heaviside function, taking the value of 0 for $z_t < c$ or 1 for $z_t > c$.

Although the STAR and Markov specifications belong to the family of switching models, there is an important conceptual difference between them. As noted by Deschamps (2008), the STAR model incorporates strong prior knowledge on the factors determining the onset of transitions between regimes (through the transition variable), while in the Markov switching model, such prior knowledge only consists in a flexible evolution equation. Therefore, the choice of an appropriate transition variable allows STAR to make better use of available information to deliver better results.

2.1.2. Statistical tests of significance

During the past decades, many studies contributed test statistics to the area of the event study methodology. These tests are the BMP (Boehmer et al. 1991), BETA-1, RANK (Corrado 1989), GARCH (Savickas 2003), TSMM (Aktas et al. 2007a) and its mean return regime dependent extensions.

To introduce the different tests, we follow the notation used by Boehmer et al. (1991) and Aktas et al. (2007a): For each test, we consider the null hypothesis of no cross-sectional average (cumulative) abnormal returns around the event date.[11]

- N: number of firms in the sample;
- AR_{jE}: abnormal return of firm j on the event date (following Eq. (2));

[11] These tests are analyzed very briefly. For further information about the tests we refer the reader to the original contributions made by the authors of each test.

- AR_{jt}: abnormal return of firm *j* on date *t*;
- *T*: number of days within the estimation period;
- *TE*: number of days within the event period;
- $\bar{R}_m$: average return on the market portfolio during the estimation period;
- $R_{m,E}$: market return on the event date
- $R_{m,t}$: market return on date *t*
- $\hat{S}_j$: standard deviation of firm *j*'s *AR* during the estimation period;
- SR_{jE}: standardized *AR* of firm *j* on the event date, calculated as:

$$SR_{jE} = \frac{AR_{jE}}{\left[\hat{S}_j \sqrt{1 + \frac{1}{T} + \frac{\left(R_{m,E} - \bar{R}_m\right)^2}{\sum_{t=1}^{T}\left(R_{m,t} - \bar{R}_m\right)^2}}\right]} \tag{8}$$

BMP test

The BMP test employs a cross-sectional approach that relies on the use of standardised abnormal returns and it was introduced to deal with event-induced increase in return variance. The BMP test takes the following form:

$$Z_{BMP} = \frac{\frac{1}{N}\sum_{j=1}^{N} SR_{jE}}{\sqrt{\frac{1}{N(N-1)}\sum_{j=1}^{N}\left(SR_{jE} - \sum_{j=1}^{N}\frac{SR_{jE}}{N}\right)^2}} \tag{9}$$

BETA-1 test

The BETA-1 test is a simplification of the BMP test (where the restrictions $\beta = 1$ and $\alpha = 0$ transform Eq. (1) to the market-adjusted model). The test is based on cross-sectional estimates of the standard deviation of the event-day abnormal returns, AR_E. The BETA-1 test takes the following form:

$$Z_{BETA-1} = \frac{\frac{1}{N}\sum_{j=1}^{N} AR_{jE}}{\sqrt{\frac{1}{N(N-1)}\sum_{j=1}^{N}\left(AR_{jE} - \sum_{j=1}^{N}\frac{AR_{jE}}{N}\right)^2}} \qquad (10)$$

This test does not rely on estimating the unconditional expected return with stock returns data prior to the event window. Due to this feature, it has been employed by several empirical studies to investigate wealth effects of M&As (e.g., Fuller et al. 2002; Moeller et al. 2004) in an effort to isolate the event window abnormal returns from any unrelated events that could have been observed in the estimation window prior to the announcement.

RANK test

This is a non-parametric test based on the ranks of abnormal returns proposed by Corrado in 1989 (see also Corrado and Zivney 1992). The RANK test merges the estimation and event windows in a single time series. Abnormal returns are sorted and a rank is assigned to each day. If K_{jt} is the rank assigned to firm *j*'s abnormal return on day *t*, then the RANK test is given by:

$$T_{RANK} = \frac{\frac{1}{N}\sum_{j=1}^{N}\left(K_{jE} - \bar{K}\right)}{S(K)} \qquad (11)$$

where $\bar{K}$ is the average rank and $S(K)$ is the standard error, calculated as:

$$S(K) = \sqrt{\frac{1}{T+TE}\sum_{t=1}^{T+TE}\left(\frac{1}{N}\sum_{j=1}^{N}\left(K_{jt} - \bar{K}\right)\right)} \qquad (12)$$

The use of ranks neutralizes the impact of the shape of the abnormal returns distribution (e.g., its skewness and kurtosis and the presence of outliers). It should therefore represent an attractive alternative way of neutralizing contaminating events within the estimation window that may cause event-induced increase in variance and cross-correlation.

GARCH test

This test assumes that the variance of the error term of Eq. (2) is a time varying process. By adopting a GARCH approach Savickas (2003) suggested the use of the following return-generating process:

$$\begin{aligned} &\varepsilon_{j,t} \sim N\left(0, h_{j,t}\right) \\ &h_{j,t} = \omega_j + \varphi_j \varepsilon_{j,t-1}^2 + \theta_j h_{j,t-1} + d_j D_{j,t} \end{aligned} \quad , \tag{13}$$

where $h_{j,t}$ is the conditional time-varying variance and ω_j, φ_j, θ_j and d_j are the coefficients of a GARCH(1,1) specification. Due to its time-varying nature, the GARCH model has the ability to control for the time-varying variance of *AR* and the event-induced increase in return variance.

The conditional variance $h_{j,t}$ provides a natural estimator of the *AR* variance. Savickas (2003) used it to standardize the *AR* before proceeding with the BMP test. In this setting, *AR* is captured in the γ estimate of Eq. (3) and Eq. (8) is replaced by:

$$SR_{jE}^{*} = \hat{\gamma} \Big/ \sqrt{\hat{h}_{j,E}} \tag{14}$$

Mean-Variance two-state market model test (MV-TSMM)

The MV-TSMM test has been proposed by Aktas et al. (2007a) and utilizes the restricted version of Eq. (3) where only the variance is assumed to be state dependent (hereafter V-TSMM). The idea of this test is based on the fact that the presence of contaminated events within the estimation window has an impact on the (ex-post) estimation of the abnormal return variance forcing in a way traditional tests to overestimate the variance of the residuals during the estimation period. To deal with this bias, the V-TSMM test relies on the Markov switching regression framework developed by Hamilton (1989, 1994) and as in the previous GARCH-test, the standard error of the γ estimate is used to standardize the AR as in Eq. (14). The test can then be constructed using the same approach as for Z_{BMP}:

$$Z_{V-TSMM} = \frac{\frac{1}{N}\sum_{j=1}^{N} SR_{jE}^{*}}{\sqrt{\frac{1}{N(N-1)}\sum_{j=1}^{N}\left(SR_{jE}^{*} - \sum_{j=1}^{N}\frac{SR_{jE}^{*}}{N}\right)^{2}}} \tag{15}$$

Prior empirical research has revealed a significant time-variation in the slope parameter which depends on market conditions (Hays and Upton 1986; Klein and Rosenfeld 1987; Chang and Weiss 1991; Chiang et al. 2013). Therefore, neglecting the state effects in the mean equation may lead to misleading inferences. To avoid this problem, we suggest using the unrestricted version of Eq. (3) and proceed using the same steps as before.[12] We name this test $Z_{MV\text{-}TSMM}$.

STAR test

Finally, the STAR test statistic, utilizes Eq. (6) were mean and variance are assumed to be state dependent.[13] These states change according to the behaviour of the transition variable (in our case of the one-period lagged returns) filtering out firm-specific contaminating events that could otherwise influence the mean and variance in the model's estimations.[14] This model shares similar benefits to the rest of the regime specifications and has the advantage that it endogenously determines (and quantifies) the level of change from one regime to the other.

[12] Estimation is based on the Maximum-likelihood method using the MSVAR library in the GAUSS software.
[13] Programming code for the STAR event study model is freely available from the authors' websites.
[14] The one-period lagged return was proved to be the most appropriate transition variable for more than 90% of the firms in our sample (based on the LM-type test).

As before, standard error of the γ estimate is used to standardize the *AR* as in Eq. (14). The test is constructed using the same approach as for $Z_{BMP.}$ We denote this test Z_{STAR}:

$$Z_{STAR}=\frac{\frac{1}{N}\sum_{j=1}^{N}SR_{jE}^{*}}{\sqrt{\frac{1}{N\left(N-1\right)}\sum_{j=1}^{N}\left(SR_{jE}^{*}-\sum_{j=1}^{N}\frac{SR_{jE}^{*}}{N}\right)^{2}}} \qquad (16)$$

2.2. Data and sample construction

2.2.1. Simulated return generating processes

To generate the theoretical *event-free* return-generating process, we consider all stocks reported in the CRSP daily returns file from January 1980 to December 2010. Contaminating-events in each stock time-series are generated by random sampling from a uniform distribution, whereas for each stock we safeguard that the resulting return time-series (estimation period and event window) does not belong in the M&As data set used in this study. The estimation window is going from -255 to -30 days relative to the event date for the traditional tests estimated using the market model following Eq. (1). For the set of tests employing the dummy-based market model following Eq. (2), estimation is done from -255 to +5 days relative to the event date. Following Fuller et al. (2002), we choose in all tests to measure cumulative abnormal returns (CARs) in the 11-day window [-5,+5] around the event announcement date. All stocks and event dates were randomly chosen with replacement such that each stock/date combination had an equal chance of being chosen at each selection. In the spirit of previous studies (e.g., Savickas et al. 2003; Aktas et al. 2007a; Harrington and Shrider 2007), we exclude securities with missing information on the event day and securities with less than 100 nonzero returns over the estimation window and no missing prices in the 11-day window surrounding the event-announcement. The latter treatment is to avoid observations where the security had recently been added to the CRSP and to limit stocks that are not actively traded in the market. Following the norm in similar studies, for each replication we construct 1000 samples of 50 stocks.[15]

We investigate the specification and power of the test statistics by deliberately contaminating the data in the estimation window. We simulate significant events by introducing abnormal returns (of random sign) into the estimation window on randomly selected points in time which are twice the standard deviation of the actual stock. To generate stochastic shocks, we follow the method proposed of Brown and Warner (1985) by adding another two demeaned returns randomly drawn from the estimation window. The sign of the simulated abnormal return is determined by random sampling from a Bernoulli distribution. Allowing either a positive or a negative sign for the abnormal return merely reflects the unknown type of the events that may emerge in the estimation window. In statistical terms, the return for stock *j* on a contaminated date *t*, denoted $R^{*}_{j,t}$ is generated as follows:

$$R^{*}_{j,t} = R_{j,t} \pm 2\sigma_{R_J} + (R_{j,X} - \overline{R}_j) + (R_{j,Y} - \overline{R}_j) \qquad (17)$$

where $R_{j,t}$ is the actual return, $R_{j,X}$ and $R_{j,Y}$ are the returns randomly selected from the estimation period and $\overline{R}_j$ and σ_{R_J} are the mean return and standard deviation in the estimation window.

The number and nature of the events during the estimation window is determined in two steps. First, a random sample is drawn from a Poisson distribution with a mean of 2 which captures the number of events during the estimation window. Events were then randomly assigned to specific days in the estimation window by random sampling from a

[15] We choose to report results using a portfolio size of 50 stocks to maintain conformity with notable previous studies (e.g., Brown and Warner 1980; Savickas 2003; Aktas et al. 2007a; Harrington and Shrider 2007; Kolari and Pynnönen 2010 etc). Nevertheless, some recent studies such as the ones by Ahern (2009) and Campbell et al. (2010) simulate larger stock portfolio sizes. We have repeated the whole analysis with 1000 samples of either 100 or 250 stocks each to find that our results/inferences remain unchanged.

uniform distribution. Second, the length (in number of days) of each event was again randomly sampled from a Poisson distribution, this time with a mean of 4.

On the simulated event day, the abnormal performance is 0% for the specification analysis and +1% for the power analysis.[16] To capture the event-induced increase in return variance, similarly to Aktas et al. (2007a) each stock's day 0 return, $R_{j,0}$, was transformed to triple its variance by adding 2 demeaned returns randomly drawn from the estimation window. The event-day transformed return was again following the nature of Eq. (17). In all cases, our market portfolio is taken to be the CRSP value weighted index.[17]

2.2.2. Real data return generating processes

This paper also strives to empirically validate the robustness of each event study approach to the return-generating mechanism by detecting abnormal performance in real data. It is intriguing from a practical perspective to investigate whether the results from simulations are also obtained when dealing with a real sample of corporate event announcements such as M&As. Therefore, we depart from the previous literature and instead of simulating abnormal returns to deliberately contaminate the estimation window with some significant events we randomly choose stock return time-series from a sample of M&A deals.[18] In this manner, contaminated (unrelated) events due to other corporate actions that precede the acquisition announcement emerge naturally in the estimation window. This treatment allows us to be more realistic with respect to the characteristics of the contaminating events, which instead of

[16] We reach qualitatively similar results for any other abnormal performance above 1%.
[17] All results are robust when we instead use the CRSP equally weighted index.
[18] The samples for the tests are completed as follows. For the 10% rejection rates: to construct the 1000 portfolio samples, each 50-firm sample is formed by randomly picking 45 firms from the universe of CRSP stocks (event-free sample) and 5 firms from the M&As data set (contaminated sample). Likewise, for the 5% rejection rates, in the first (second) 500 samples we randomly pick 3 (2) firms from the M&As data set and 47 (48) firms from the universe of CRSP stocks. For the 1% rejection rates, in the first (second) 500 samples we randomly pick 1 (0) firms from the M&As data set and 49 (50) firms from the universe of CRSP stocks.

being artificially generated using some *pre-determined* nuisance distributional parameters, are taken from stock returns of firms that undergo a real corporate action.

The analysis includes all merger and acquisition announcements involving U.S. targets and taking place in the period 1980–2010, extracted from the Securities Data Corporation (SDC) database. We require that firms are listed in NYSE, AMEX or NASDAQ with available CRSP data, the outcome of the deal is known (either completed or withdrawn), and the deal value is over 1 million USD. We exclude deals whose value represents less than 1% of the bidder's market capitalization. The final sample includes 4421 bidders and 5928 targets which constitute the universe of our real stock returns data.

Our research design with the real data closely follows the one we implement with the simulated stock returns. The aim of this analysis yet, is to validate the power of the event study when the estimation window is *naturally* contaminated by disruptions in the normal return-generating process, of the size and amplitude we might expect from various real-world corporate event announcements. Therefore, unlike the simulation case, various contaminating-events are assumed to inherently exist in the estimation window. Estimation and event window lengths are taken to be the same as with the simulated paradigm. To be able to study the rejection rates of the test statistics under the no event-induced increase in variance case, we enforce a neutral abnormal performance of 0% on the real event date by de-meaning the real stock returns in the 11-day period [-5,+5] around the event date.[19] In this way the actual event market reaction is neutralized and the event window with actual data has similar behaviour to the simulated data. Thereafter, to capture the event-induced increase in variance phenomenon each security's day 0 return, $R_{j,0}$, was transformed to triple its variance by adding 2 demeaned returns randomly drawn from the estimation window. The event-day

[19] Unreported findings suggest that our results are robust to longer or shorter event windows (e.g. [-20, +20] and [-1, +1].

transformed return was again following the nature of Eq. (17). Finally, to study the power analysis of each method, we introduce an 1% abnormal return on the real event date.

3. Discussion of results

3.1. Simulated contaminated events

Following previous studies, we first provide analytical and empirical evidence of the resulting biases using randomly selected firms with artificially induced abnormal events during the estimation period. Our results are presented in Tables 1 to 4, which show rejection and power rates. In particular, Tables 1 and 3 provide analysis on specification tests (Type I errors) under the null hypothesis of no event effect on the abnormal returns, while Tables 2 and 4 report power tests (Type II errors) under the alternative hypothesis of nonzero mean abnormal returns. Tables 3 and 4 capture event-induced increase in return variance generated by stochastic shocks to comfort to recent theoretical and empirical evidence that documents that all events induce variance (see, Harrington and Shrider 2007). Nevertheless, to facilitate comparisons with prior literature, Tables 1 and 2 report specification and power analysis under the no event-induced increase in variance case. Results are presented for the BMP, BETA-1, RANK, STAR, GARCH, V-TSMM and MV-TSMM tests.

Prior to examining the performance of various statistical tests of significance, we investigate the in-sample performance of the models used in this study. We rely on two widely used loss functions, namely the mean square error (MSE) and mean absolute error (MAE). In particular, we find that the STAR model specification delivers the overall best results with MSE (MAE) equal to 0.98 (0.77), followed by MV-TSMM with loss values of 1.00 (0.81) and V-TSMM with 1.02 (0.82). The rest two model specifications, namely the GARCH and market model specifications deliver the worst fitting performance results with much higher loss values. The in-sample performance of the alternative market models provides empirical support for the application of the STAR specification in the event study framework since it fits the returns data over the estimation window much better than any other rival method. Therefore, in-sample modelling performance of empirical stock returns generating process renders the use of the STAR method more desirable.

3.1.1. Tests with no change in return variance

Table 1 presents rejection rates for test statistics when an event creates no abnormal returns and no increase in variance in the event window. Panel A of Table 1 shows that if contaminating events are not present in the estimation window, all tests perform relatively well with BMP and TSMM to look the most attractive ones. Similar results are obtained in the case of contaminating events in the estimation window since as it can be seen in Panel B rejection rates are similar to the ones presented in Panel A.

[Insert Table 1 about here]

Table 2 presents power analysis in the case where an event creates an increase in the event window returns but no increase in variance. Panel A provides the analysis when the estimation window is not contaminated while in Panel B the same analysis is presented in the case of contaminating events. In Panel A of Table 2 we observe that RANK is the most powerful test. The less powerful tests are the BETA-1 and GARCH. When the significance level is 1%, STAR and V-TSMM provide similar results, while in the case of 5% and 10% levels, we find similar results for STAR and MV-TSMM. In the presence of contaminating events as shown in Panel B of Table 2, BETA-1 and GARCH are still the less powerful tests. In the case of 1% and 5% significance levels, RANK is the best approach and MV-TSMM the second best one. When the significance level is 10%, MV-TSMM performs slightly better than RANK. It is worth mentioning that while the power of the RANK test in Panels A and B seems to be the same, we do not observe the same behavior for the MV-TSMM. Particularly for the MV-TSMM we notice an increase in the power when contaminating events are present. Furthermore, we observe that the STAR, V-TSMM and BMP allow the presence of contaminating events to be controlled much better for, without severe changes in power. This argument is not valid in the case of GARCH since a reduction in the power of the test is observed in the presence of contaminating events.

[Insert Table 2 about here]

Taken all evidence together from Tables 1 and 2, the traditional BMP and BETA-1 approaches seem to perform well in the presence of contaminating events. Yet, for BETA-1 this comes at a significant cost when someone considers the reduction in its power. Likewise Aktas et al. (2007a) and Kolari and Pennönen (2010), among others, our analysis reveals that under the simulated setting, the RANK test appears to be a well attractive alternative since is seems resilient to the presence of contaminating events and preserves the highest levels of power among the candidate tests in the absence of event-induced increase in variance. The more elaborated tests of the regime-switching family (STAR, V-TSMM and MV-TSMM) do not seem to outperform overall the (less complex) traditional tests in the absence of event-induced increase in variance. Nonetheless, Harrington and Shrider (2007) provide theoretical and empirical evidence to support the notion that all events induce variance. Therefore, the following section compares the statistical performance of the battery of tests we employ in our analysis in a more realistic fashion where we simulate also a variance increase on the event date.

3.1.2. Tests with variance increase in the event window

Table 3 reports rejection rates for different cross sectional test statistics when an event creates no abnormal returns but increases variance in the event window. Panel A and Panel B of Table 3 show that the RANK test is poorly specified when an event-induced increase in return variance is present. The traditional BMP continues to exhibit relatively good performance in terms of the specification tests both, in the absence or presence of contaminating events. Furthermore, for the family of the regime-switching models rejection rates are quite close to the expected ones and do not seem to be much different in Panels A and B, proving supportive evidence that these tests are robust in the presence of contaminating events under the induced increase in variance analysis.

[Insert Table 3 about here]

In Table 4 results show the power analysis for different test statistics when an event creates an increase in returns and variance. Comparing the results of Tables 2 and 4 we observe that the presence of an event-induced increase in return variance, drastically affects the power of all tests. In particular, the traditional BMP test statistic exhibits a severe reduction in power and appears to be extremely sensitive to the presence of event-induced increase in variance. In a similar fashion, BETA-1 and GARCH perform even worse than BPM and appear to be extremely weak in detecting abnormal performance in the presence of variance increases events. Contrary to its badly misspecification presented in Table 3, RANK shows the highest power from all traditional test statistics. The regime-switching family of models, STAR, V-TSMM and MV-TSMM, seem however to be the most attractive approaches in this investigation since by comparing the results in Panel A and Panel B of Table 4 there is a significant increase in the power of these three tests.

[Insert Table 4 about here]

Since the performance of the test statistics differs significantly with respect to the specification error (as shown in Table 3), power analysis which is tabulated in Table 4 is not directly comparable across the different tests. Therefore, to further scrutinize our findings, we employ a graphical method proposed by Davidson and Mackinnon (1998), namely the *size-power curves*, which allows the comparison of alternative test statistics that have different size. The results are presented in Figure 1: Panel A depicts the size-power curves in the absence of contaminating events in the estimation window, whereas, Panel B depicts evidence relating to the behavior of the tests in the presence of contaminating events. Consistent with the results in Aktas et al. (2007a), the size-power curves of BETA-1 and GARCH reveal that these are the least powerful tests. The widely applied BMP and RANK tests perform much better than BETA-1 and GARCH, yet their performance is much inferior compared to the regime-switching family models. In particular, the graphical evidence in Figure 1 strongly supports that the STAR event study model provides the overall most powerful test statistics, since its curve dominates all other test statistic curves, both without or with contaminating events. The MV-TSMM and the TSMM test are the second best choices.

[Insert Figure 1 about here]

Overall, empirical evidence so far lends credence to the use of the regime-switching family models and in particular towards the utilization of the STAR event study model which appears to be the most accurate and robust method in the presence of both, event-induced increase in return variance and contaminated events.[20]

3.2. Real sample of M&As data

Undoubtedly, the traditional simulated type event study approach has been routinely utilized to investigate specification and power performance of standard event study methods in the presence of artificially contaminated events and event-induced increase in variance cases. It remains interesting, however, to empirically validate whether similar model rankings

[20] To further guard against erroneous inferences that may arise in the presence of cross-sectional correlation and non-normal stock returns, we also apply the adjusted BMP test (cBMP) of Kolari and Pynnönen (2010) and the generalized RANK test (GRANK) of Kolari and Pynnönen (2011). Overall, performance of cBMP and GRANK is better when compared to their initial counterparts (i.e., the BMP and RANK tests, respectively). Our empirical results and inferences, however, are unchanged regarding the superiority of the regime-switching models and in particular of the STAR event study method over all other test statistics (the same holds true for the analysis that follows in Section 3.2). For the sake of brevity, we omit presenting results of these two tests in the tables; yet for illustration purposes and completeness, we include their size-power curve performance in Figures 1 and 2.

in terms of specification and power tests are also observed when dealing with real-world situations. Therefore, we focus on actual return generating processes coming from M&As where estimation period contamination and event-induced increase in return variance should emerge naturally for firms that engage in this particular corporate activity.

The results from the M&As are presented in Table 5. Panels A and B present the rejection rates without or with the presence of event-induced increase in variance, respectively. Likewise, Panels C and D provide the corresponding power analysis performance. Overall, we observe that the most powerful approaches are again from the regime-switching family, in particular the STAR, V-TSMM and MV-TSMM models. Under the event-induced increase in return variance setting, the models exhibit rejection rates close to the expected ones, as well as the highest levels of power among all rival test statistics. In general, these results are in the same line of reasoning with the ones we observed in the simulated environment under the event-induced increase in return variance case.

[Insert Table 5 about here]

Figure 2 depicts the size-power curves with event-induced return of 1% and an event-induced increase in return variance using the M&As sample. The overall model rankings are almost similar to the ones observed with the simulated data. The size-power curve evidence of BETA-1 and GARCH pinpoint that these are still the least powerful tests with real stock returns. The RANK test performs better than the BETA-1 and GARCH ones, whilst in contrast to results we get with the simulated data, the BMP test is now significantly inferior to the RANK test and only slightly better than BETA-1 and GARCH. Despite the good behavior exhibited by the RANK model, yet, its performance is much inferior compared to the regime-switching family models. In particular, the graphical evidence in Figure 2 supports that the STAR event study model provides the overall most powerful test statistics, since its curve dominates all other test statistic curves. The MV-TSMM and the TSMM test are again the second best choices. A noticeable difference we observed with the real stock returns is that the performance wedge between the regime-switching models and the rest traditional test statistics is much greater when compared to the simulated cases.

[Insert Figure 2 about here]

We further investigate the sensitivity of event study residuals to extreme market conditions. Particularly, from the sample of M&As we pick the 20% deals with either the lowest mean stock return performance or the highest volatility in the estimation window.[21] We draw our motivation from prior literature. For instance, Klein and Rosenfeld (1987) suggest that traditional event study methods may suffer from serious deficiencies due to high autocorrelation that may emerge in the time-series of the resulting abnormal returns if the event days take place during either bull or bear markets (see also Chiang et al. 2013). Moreover, Campbell et al. (2001) recognize that the increase in the idiosyncratic volatility might potentially affect the inferences of the event study analysis since abnormal event-related returns are highly determined by the volatility of individual stock returns relative to the market. All-in-all, by using this M&As sub-sample analysis we endeavour to clarify whether the specification and power performance of the test statistics we investigate are stable under different market conditions.

Table 6 presents the results for the sub-sample of M&As that exhibit the lowest mean returns in the estimation window. In the case where there is no event-induced increase in variance the most powerful test from Panel C is the MV-TSMM, yet its rejection rates are not as expected (Panel A). Although STAR is the second most powerful test, its rejection rates (Panel A) are close to the expected ones. Rejection rates of GARCH are similar to the rejection rates of the other tests but GARCH is the test with the lowest statistical power.

[21] All conclusions regarding the mean return remain unaltered if we instead pick the 20% deals with the highest mean returns in the estimation window.

BETA-1 is similar to the RANK test regarding rejection rates and power. BMP and V-TSMM have similar rejection rates but V-TSMM clearly outperforms. Panel B (Panel D) provides rejection and power rates in the presence of event-induced increase in variance. We notice that the power of all tests is severely reduced in the presence of event-induced increase in variance. Results show that rejection rates for the STAR and V-TSMM tests are similar to the ones reported in Panel A and these tests are also the ones with the highest level of power. BETA-1 and GARCH are the least powerful tests under the event-induced increase in variance scheme. BMP has similar rejection rates as in Panel A but its power is significantly lower in Panel D than the power reported in Panel C. Rejection rates of MV-TSMM are similar to V-TSMM but the power of MV-TSMM is lower than the power of V-TSMM. The fact that we observe lower power of MV-TSMM compared to V-TSMM is in line with empirical findings as in Aktas et al. (2007a) which suggest that the market model parameters are the same under both regimes. However, this is no longer true since contaminating events affect both mean and variance specification. Therefore the power of the MV-TSMM model could happen to be higher. Another important observation in the presence of event-induced increase in variance is that rejection rates of RANK differ significantly from rejection rates of the other tests. This observation provides additional empirical evidence that the RANK test performs poorly under real stock return data. Overall, empirical observations in Table 6 give support to the STAR event study method which exhibits again reasonable performance under extreme stock returns occurring in the estimation window.

[Insert Table 6 about here]

Table 7 presents results regarding the case of investigating the 20% of M&As that preserve the highest return volatility in the estimation window. In general, this type of analysis again reveals that the regime-switching family of models dominates all other tests in terms of specification and power analysis. The power of the STAR model is yet much better than the power of V-TSMM and MV-TSMM; at 1% significance level, we observe the most intense differences in power rates between the STAR and the other two models. Additionally, rejection rates of STAR, V-TSMM and MV-TSMM are similar in Panels A and B, a fact that empirically demonstrates that these models perform rather well when event-induced increase in variance is present.

[Insert Table 7 about here]

4. Discussion and conclusions

There is a variety of tests that are robust to event-induced increase in variance caused by the cross-sectional variation in the effects of an event. According to Harrington and Shrider (2007), all events induce variance and therefore models that are robust to cross sectional variation must be used.

Using simulated data, we observe that when there is no event-induced increase in variance, the RANK test is an attractive approach while using real M&A data the power of the RANK test deteriorates significantly. Furthermore, the RANK test in the presence of event-induced increase in variance seems to be poorly specified since in all results rejection rates of this test statistics are extremely high compared to rejection rates of all other tests. This result is also supported by previous studies, see for example Aktas et. al. (2007a). The BMP model is a good choice to work with in the case of contaminated events under the simulated data environment. In the real data setting, however, BMP performs reasonably well but not as good as the regime-switching family and in particular when compared to the STAR event study test statistic.

The least powerful tests are BETA-1 and GARCH. When using simulated data, these two tests have similar rejection rates and power regardless of the presence of contaminating events. On the other hand, evidence from using the real data suggests that in the presence of event-induced increase in variance the rejection rates of these tests as well as their power are severely reduced making these tests the less preferable.

Comparing the real returns results under different extreme market conditions we observe differences in the power of the test statistics especially in the presence of event-induced increase in variance. Yet, extreme market conditions do not affect the models' ranking performance. Again the best approach is one of the regime-switching models with STAR event study model to show the overall best performance.

In a nutshell, we find empirical evidence to support that the best approach under contamination and event-induced increase in variance is the test statistic computed from the STAR event study model followed by V-TSMM and MV-TSMM. Overall, it is found that the STAR event study test statistic outperforms, in almost all cases, any other rival method. Therefore, our analysis empirically supports that the STAR model should be employed in the application of short-horizon event studies since it appears to be the most credible method in detecting the true size of abnormal returns.

Acknowledgments

The authors would like to thank Milto Hadjikyriakou for providing excellent research assistance. The authors also acknowledge comments and constructive suggestions from participants at the 5th CSDA International Conference on Computational and Financial Econometrics (December 2011, UK).

References

Ahern KR (2009) Sample selection and event study estimation. Journal of Banking and Finance 16: 466-482.

Aktas N, De Bodt E, Cousin JG (2007a) Event studies with a contaminated estimation period. Journal of Corporate Finance 13: 129-145.

Aktas N, De Bodt E, Cousin JG (2007b) Assessing the power and the size of the event study method through the decades. In Finance International Meeting AFFI-EUROFIDAI, Paris, December.

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

Aktas N, De Bodt E, Roll R (2009). Learning, hubris and corporate serial acquisitions. Journal of Corporate Finance 15: 543-561.

Arora RK, Das H, Jai PK (2009) Stock returns and volatility: Evidence from selected emerging markets. Review of Pacific Basin Financial Markets and Policies 12: 567-592.

Ball, CA, Torous, WN (1988) Investigating security-price performance in the presence of event-date uncertainty. Journal of Financial Economics 22: 123-153.

Bhagat S, Dong M, Hirshleifer D, Noah R (2005) Do tender offers create value? New methods and evidence. Journal of Financial Economics 76: 3-60.

Boehmer E, Masumeci J, Poulsen AB (1991) Event-study methodology under conditions of event-induced variance. Journal of Financial Economics 30: 253-272.

Brown SJ, Warner JB (1980) Measuring security price performance. Journal of Financial Economics 8: 205-258.

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

Cam MA, Ramiah V (2014) The influence of systematic risk factors and econometric adjustments in catastrophic event studies. Review of Quantitative Finance and Accounting 42: 171-189.

Campbell CJ, Cowan AR, Salotti V (2010) Multi-country event-study methods. Journal of Banking and Finance 34: 3078-3090.

Campbell JY, Lettau M., Malkiel BG, Xu Y (2001) Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. The Journal of Finance 56: 1-43.

Chang WC, Weiss DE (1991) An examination of the time series properties of beta in the market model. Journal of the American Statistical Association 86: 883-890.

Chiang T, Tan L, Li J, Nelling E (2013) Dynamic herding behavior in Pacific-Basin markets: Evidence and implications. Multinational Finance Journal 17: 165-200.

Corrado CJ (1989) A nonparametric test for abnormal security-price performance in event studies. Journal of Financial Economics 23: 385-395.

Corrado CJ, Zivney TL (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.

Cyree KB, DeGennaro RP (2002) A generalized method for detecting abnormal returns and changes in systematic risk. Review of Quantitative Finance and Accounting 19: 399-416.

Davidson R, MacKinnon JG (1998) Graphical methods for investigating the size and power of hypothesis tests. The Manchester School 66: 1-26.

Délèze F, Hussain S (2014) Information arrival, jumps and cojumps in European financial markets: Evidence using tick by tick data. Multinational Finance Journal 18: 169-213.

Deschamps PJ, (2008) Comparing smooth transition and Markov switching autoregressive models of US unemployment. Journal of Applied Econometrics 23: 435-462.

Dombrow J, Rodriguez M, Sirmans CF (2000) A complete nonparametric event study approach. Review of Quantitative Finance and Accounting 14: 361-380.

Fama E, Fisher L, Jensen M, Roll R (1969) The adjustment of stock prices to new information. International Economic Review 10: 1-21.

Fuller K, Netter J, Stegemoller M (2002) What do returns to acquiring firms tell us? Evidence from firms that make many acquisitions. The Journal of Finance 57: 1763-1793.

Guo JM, Petmezas D (2012) What are the Causes and Effects of M&As? The UK Evidence. Multinational Finance Journal 16: 21-47.

Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57: 357-384.

Hamilton JD (1994) Time series analysis. Princeton: Princeton university press.

Hansen B (2011) Threshold Autoregression in Economics. Statistics and Interface 4: 123-127.

Harford J. (2005) What drives merger waves? Journal of Financial Economics 77: 529-560.

Harrington SE, Shrider DG (2007) All events induce variance: Analyzing abnormal returns when effects vary across firms. Journal of Financial and Quantitative Analysis 42: 229-256.

Hays PA, Upton DE (1986) A shifting regimes approach to the stationarity of the market model parameters of individual securities. Journal of Financial and Quantitative Analysis 21: 307-321.

Karafiath I (1988) Using dummy variables in the event methodology. Financial Review 23: 351-357.

Kenourgios D, Dimitriou D, Christopoulos A (2013) Asset markets contagion during the global financial crisis. Multinational Finance Journal 17: 49-76.

Klein A, Rosenfeld J (1987) The influence of market conditions on event-study residuals. Journal of Financial and Quantitative Analysis 22: 345-351.

Kolari JW, Pynnönen S (2010) Event study testing with cross-sectional correlation of abnormal returns. Review of Financial Studies 23: 3996-4025.

Kolari JW, Pynnönen S (2011) Nonparametric rank tests for event studies. Journal of Empirical Finance 18: 953-971.

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

Laopodis NT (2010) Dynamic linkages between monetary policy and the stock market. Review of Quantitative Finance and Accounting 35: 271-293.

Moeller SB, Schlingemann FP, Stulz RM (2004) Firm size and the gains from acquisitions. Journal of Financial Economics 73: 201-228.

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: 165-178.

Sharpe WF (1963) A simplified model for portfolio analysis. Management Science 9: 277-293.

Terasvirta T (1994) Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models, Journal of the American Statistical Association 89: 208–218.

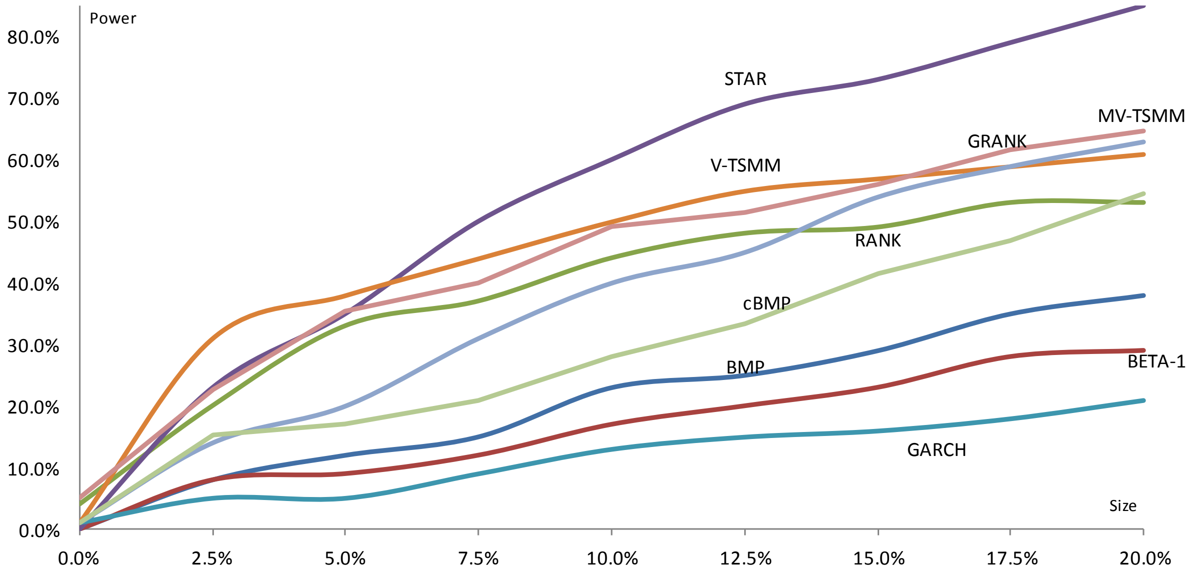
Terasvirta T, Anderson HM (1992) Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. Journal of Applied Econometrics 7: 119-36.

van Dijk D, Franses PH (1999) Modelling Multiple Regimes in the Business Cycle. Macroeconomic Dynamics 3: 311-340.

van Dijk D, Teravirta T, Franses PH (2002) Smooth Transition Autoregressive Models – a Survey of Recent Developments. Econometric Reviews 21: 1-47.

Figures

Panel A. Without contaminating events



Panel B. With contaminating events

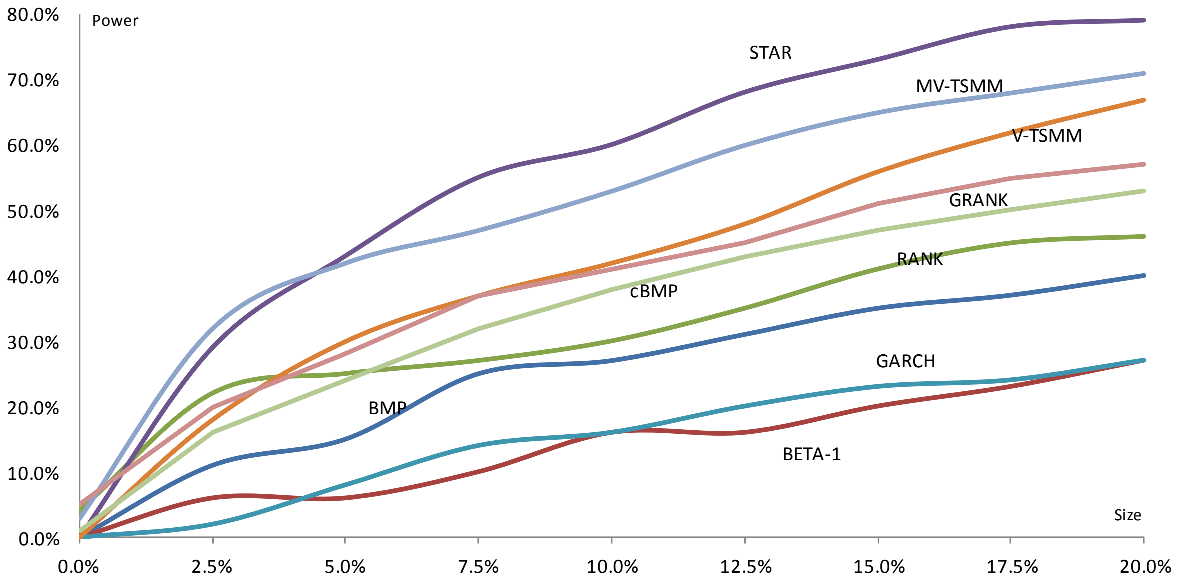


Figure 1. This figure provides size-power curves with event-induced return of 1% and an event-induced increase in return variance with simulated data. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003), BETA-1 is the cross-sectional test using the constrained version of the market model, cBMP is the adjusted BMP test as in Kolari and Pynnönen (2010) and GRANK is the generalized RANK test as in Kolari and Pynnönen (2011). STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a).

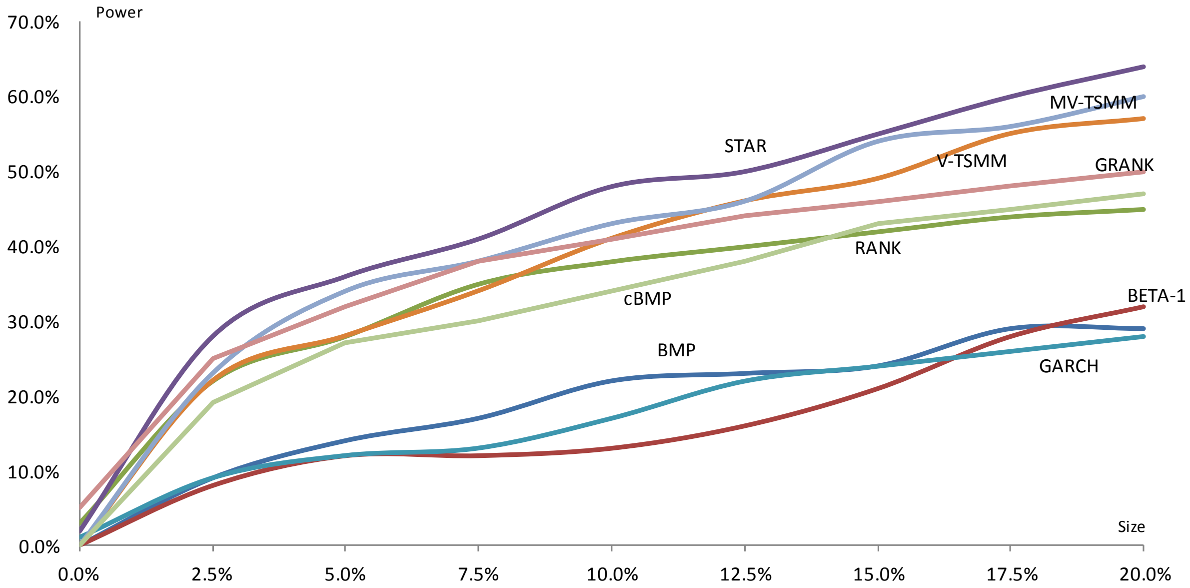


Figure 2. This figure provides size-power curves with event-induced return of 1% and an event-induced increase in return variance with M&As stock returns data. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003), BETA-1 is the cross-sectional test using the constrained version of the market model, cBMP is the adjusted BMP test as in Kolari and Pynnönen (2010) and GRANK is the generalized RANK test as in Kolari and Pynnönen (2011). STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a).

Tables

Table 1
Rejection Rates of Test Statistics: no event-induced returns – no event-induced variance

	Significance Level		
	1%	5%	10%
Panel A: Without contaminating events			
BMP	0.70%	5.30%	10.40%
BETA-1	0.60%	5.50%	9.60%
RANK	1.50%	8.00%	14.10%
STAR	0.10%	2.50%	8.60%
GARCH	2.00%	5.90%	11.20%
V-TSMM	0.70%	4.80%	10.30%
MV-TSMM	0.70%	5.30%	10.90%
Panel B: With contaminating events			
BMP	0.90%	5.00%	10.40%
BETA-1	0.80%	5.40%	9.70%
RANK	1.40%	7.60%	13.50%
STAR	0.20%	2.50%	8.90%
GARCH	1.60%	6.50%	13.10%
V-TSMM	0.60%	5.20%	12.60%
MV-TSMM	0.90%	4.60%	9.60%

The table presents rejection rates for cross-sectional test statistics when an event creates no abnormal returns and no increase in variance. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a). Panel A provides the analysis when the estimation window is not contaminated and Panel B when it is contaminated.

Table 2
Power Analysis: event-induced returns – no event-induced variance

	Significance Level		
	1%	5%	10%
Panel A: Without contaminating events			
BMP	42.30%	67.80%	77.50%
BETA-1	10.80%	26.70%	38.80%
RANK	53.60%	75.40%	82.30%
STAR	20.70%	52.30%	68.60%
GARCH	8.90%	23.00%	32.90%
V-TSMM	18.40%	34.90%	45.10%
MV-TSMM	31.80%	51.70%	62.20%
Panel B: With contaminating events			
BMP	42.10%	66.90%	77.10%
BETA-1	10.10%	27.10%	38.70%
RANK	51.70%	75.00%	81.40%
STAR	22.80%	52.90%	69.70%
GARCH	6.40%	16.50%	26.20%
V-TSMM	20.80%	36.00%	45.90%
MV-TSMM	47.70%	71.70%	81.50%

The table presents power rates for cross-sectional test statistics when an event creates an increase in returns of 1% and no increase in returns variance. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a). Panel A provides the analysis when the estimation window is not contaminated and Panel B when it is contaminated.

Table 3
Rejection Rates of Test Statistics: no event-induced returns – but event-induced variance

	Significance Level		
	1%	5%	10%
Panel A: Without contaminating events			
BMP	1.60%	6.00%	11.70%
BETA-1	0.70%	6.70%	11.60%
RANK	11.40%	22.90%	30.00%
STAR	0.20%	3.50%	9.90%
GARCH	1.70%	5.00%	9.50%
V-TSMM	1.30%	6.80%	9.60%
MV-TSMM	1.20%	5.90%	9.50%
Panel B: With contaminating events			
BMP	0.80%	5.60%	10.50%
BETA-1	0.80%	5.10%	10.60%
RANK	9.50%	20.40%	29.60%
STAR	0.20%	4.20%	10.10%
GARCH	1.90%	6.80%	11.30%
V-TSMM	1.30%	6.30%	9.00%
MV-TSMM	1.20%	6.30%	9.50%

The table presents rejection rates for cross-sectional test statistics when an event creates no abnormal returns but an increase in returns variance. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a). Panel A provides the analysis when the estimation window is not contaminated and Panel B when it is contaminated.

Table 4
Power Analysis: event-induced returns – event-induced variance

	Significance Level		
	1%	5%	10%
Panel A: Without contaminating events			
BMP	4.00%	11.40%	18.70%
BETA-1	1.60%	6.20%	13.30%
RANK	12.50%	24.80%	33.60%
STAR	10.80%	37.60%	56.20%
GARCH	2.10%	7.10%	12.70%
V-TSMM	16.90%	32.20%	42.10%
MV-TSMM	12.80%	33.20%	52.10%
Panel B: With contaminating events			
BMP	4.80%	12.30%	22.60%
BETA-1	2.20%	8.00%	15.00%
RANK	12.30%	26.70%	33.90%
STAR	14.20%	41.60%	59.00%
GARCH	2.00%	6.70%	11.80%
V-TSMM	15.70%	40.30%	55.10%
MV-TSMM	16.50%	36.90%	52.70%

The table presents power rates for cross-sectional test statistics when an event creates an increase in returns of 1% and an increase in returns variance. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a). Panel A provides the analysis when the estimation window is not contaminated and Panel B when it is contaminated.

Table 5
Rejection Rates and Power Analysis of Test Statistics: Pooled sample of M&A deals

	Significance Level		
	1%	5%	10%
Panel A: Rejection Rates of Test Statistics: No event-induced variance			
BMP	4.80%	14.40%	20.80%
BETA-1	4.60%	12.60%	22.40%
RANK	4.00%	13.10%	19.90%
STAR	1.30%	6.50%	12.10%
GARCH	6.70%	16.40%	20.60%
V-TSMM	1.60%	5.90%	11.70%
MV-TSMM	1.30%	7.00%	12.20%
Panel B: Rejection Rates of Test Statistics: Event-induced variance			
BMP	2.10%	7.30%	13.10%
BETA-1	1.70%	5.90%	12.40%
RANK	13.80%	26.00%	35.20%
STAR	1.10%	4.90%	9.50%
GARCH	3.40%	10.50%	13.90%
V-TSMM	0.80%	4.90%	10.60%
MV-TSMM	0.90%	5.00%	10.30%
Panel C: Power Analysis: No event-induced variance			
BMP	45.00%	53.50%	65.90%
BETA-1	47.20%	52.90%	61.70%
RANK	52.80%	56.90%	71.90%
STAR	75.00%	82.40%	89.90%
GARCH	20.80%	25.20%	33.50%
V-TSMM	65.20%	75.90%	86.90%
MV-TSMM	71.50%	82.50%	89.40%
Panel D: Power Analysis: Event-induced variance			
BMP	14.80%	31.90%	41.60%
BETA-1	7.90%	20.90%	30.90%
RANK	20.60%	35.50%	43.80%
STAR	25.90%	39.80%	50.10%
GARCH	14.50%	18.70%	24.10%
V-TSMM	18.30%	37.20%	49.70%
MV-TSMM	18.30%	37.90%	48.60%

Rejection and power rates without or with an event-induced increase in return variance for the pooled sample of M&As. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a).

Table 6

Rejection Rates and Power Analysis of Test Statistics: 20% of M&A deals with lowest mean returns

	Significance Level		
	1%	5%	10%
Panel A: Rejection Rates of Test Statistics: No event induced variance			
BMP	1.40%	7.80%	14.60%
BETA-1	2.60%	10.80%	18.30%
RANK	2.40%	11.80%	16.20%
STAR	0.80%	4.90%	9.10%
GARCH	2.40%	10.30%	16.70%
V-TSMM	1.30%	5.80%	15.50%
MV-TSMM	4.80%	13.30%	22.50%
Panel B: Rejection Rates of Test Statistics: Event-induced variance			
BMP	1.70%	6.40%	11.50%
BETA-1	1.40%	5.90%	12.50%
RANK	11.90%	23.60%	31.80%
STAR	0.70%	4.90%	11.30%
GARCH	2.50%	7.90%	12.90%
V-TSMM	1.40%	5.80%	12.20%
MV-TSMM	1.30%	5.70%	10.90%
Panel C: Power Analysis: No event-induced variance			
BMP	49.00%	50.30%	62.70%
BETA-1	46.40%	56.80%	76.90%
RANK	47.30%	58.10%	77.20%
STAR	65.90%	75.40%	86.20%
GARCH	17.30%	36.20%	48.50%
V-TSMM	63.10%	77.00%	84.50%
MV-TSMM	69.80%	81.20%	86.30%
Panel D: Power Analysis: Event-induced variance			
BMP	4.40%	12.30%	19.70%
BETA-1	2.60%	9.70%	15.20%
RANK	12.10%	24.40%	31.90%
STAR	12.10%	24.90%	35.40%
GARCH	2.30%	8.10%	13.40%
V-TSMM	12.40%	25.60%	36.00%
MV-TSMM	7.10%	18.10%	28.50%

Rejection and power rates without or with an event-induced increase in return variance for an M&As sub-sample with the lowest mean returns in the estimation window. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a).

Table 7
Rejection Rates and Power Analysis of Test Statistics: 20% of M&A deals with highest standard deviation

	Significance Level		
	1%	5%	10%
Panel A: Rejection Rates of Test Statistics: No event-induced variance			
BMP	3.90%	11.20%	16.90%
BETA-1	4.40%	10.90%	17.80%
RANK	4.00%	11.40%	17.50%
STAR	0.90%	5.10%	9.80%
GARCH	3.70%	12.10%	18.60%
V-TSMM	1.10%	5.40%	10.20%
MV-TSMM	1.00%	5.20%	10.10%
Panel B: Rejection Rates of Test Statistics: Event-induced variance			
BMP	1.90%	6.00%	12.00%
BETA-1	1.80%	6.30%	12.10%
RANK	11.60%	24.10%	32.70%
STAR	1.10%	5.10%	10.30%
GARCH	2.90%	7.30%	12.40%
V-TSMM	1.20%	5.30%	10.20%
MV-TSMM	1.20%	5.20%	10.10%
Panel C: Power Analysis: No event-induced variance			
BMP	41.80%	49.00%	62.40%
BETA-1	42.50%	49.20%	58.40%
RANK	48.70%	53.60%	68.40%
STAR	71.20%	78.60%	87.40%
GARCH	18.10%	21.20%	30.20%
V-TSMM	61.20%	72.20%	80.80%
MV-TSMM	66.70%	80.80%	87.40%
Panel D: Power Analysis: Event-induced variance			
BMP	2.70%	9.70%	15.50%
BETA-1	2.70%	9.10%	15.30%
RANK	13.10%	25.20%	33.50%
STAR	24.90%	38.50%	50.20%
GARCH	2.20%	8.50%	13.90%
V-TSMM	15.70%	36.60%	46.90%
MV-TSMM	18.80%	38.80%	49.50%

Rejection and power rates without or with an event-induced increase in return variance for an M&As sub-sample with the highest return volatility in the estimation window. BMP is as in Boehmer et al. (1991), RANK is as in Corrado (1989), GARCH is as in Savickas (2003) and BETA-1 is the cross-sectional test using the constrained version of the market model. STAR estimates the market model according to a two-state transition variable, while V-TSMM and MV-TSMM are, respectively, the two-state market model extensions for variance and mean returns and variance following Aktas et al. (2007a).