Event study methods in corporate governance studies

Manapol Ekkayokkaya
Manapol@cbs.chula.ac.th

Krishna Paudyal
Krishna.paudyal@strath.ac.uk

Poonyawat Sreesing
Poonyawatsrs@au.edu

We thank Anant Chiarawongse and Sira Suchintabandid for valuable comments on the technical aspect of this chapter. All remaining errors are ours.

Manapol Ekkayokkaya is from Chulalongkorn University, Thailand. Krishna Paudyal is from the University of Strathclyde, United Kingdom. Poonyawat Sreesing is from Assumption University, Thailand. **Correspondence**: Manapol Ekkayokkaya, Chulalongkorn Business School, Chulalongkorn University, Bangkok 10330, Thailand. Tel: +66 2 2185671; Fax: +66 2 2185676; Email: Manapol@cbs.chula.ac.th .

This version: November 2021

Event study methods in corporate governance studies

1. Introduction

What is an event study? Fundamentally, it is a study that analyzes how security prices adjust to any specific kind of new information. The event study methodology was first developed by Fama, Fisher, Jensen, and Roll (FFJR) (1969) to examine how share price reacts to the announcement of a stock-split decision, i.e., how the announcement affects stock price of the firm conducting the split. Since then, there has been a great deal of extensions and refinements made to the methodology developed by FFJR (1969). One of the most (if not the most) widespread applications of FFJR (1969), with further developments in testing methods, has been made in the area of corporate finance. That is, to test the wealth effect of a corporate decision or event. In other words, the FFJR (1969) methodology is commonly employed in addressing the question of how a managerial decision (or, corporate event) affects shareholders' wealth. The methodology has been applied to testing the price changes not only around the event announcement, but also over a much longer period following the time of the event (e.g., 3 or 5 years after the event).

Application of the FFJR (1969) methodology has also gained increasing popularity among studies on corporate governance (CG) issues. These include not only studies on firm-specific CG events, but also studies on CG-related regulations. The FFJR (1969) methodology is a fruitful approach since the market reaction, or share price reaction, to the arrival of new information about a CG event can provide a meaningful indication of the expected effect of the event as assessed by market participants in aggregate. Indeed, CG events by nature can often influence firms' activities in the long run, i.e., have long-term implications on firms. Thus, examination of the effect of CG events over a long horizon may conceivably be viewed by many as desirable.

Our objective in this chapter is to point out to potential issues inherent in application of the conventional event study methodology to CG studies. In doing so, we shall refrain from making reference to any particular studies whose reported findings appear to be attributable to their potentially debatable application of the event study methodology.

Another use of this chapter can also be viewed as non-technical notes for those seeking to conduct an event study, but are not yet familiar with *what it is all about*. To our knowledge, the existing work that addresses methodological problems inherent in event studies or insightfully reviews the published methodologies is highly technical and suitable mostly to advanced readers. Given the nature of the topic itself, this is understandable and the way it has to be. Unfortunately, it leaves novices with a big leap to make by themselves. In relation to the materials currently available in the literature, this chapter can be viewed, hopefully, as a primer that covers the groundwork needed for embarking on the more technical and advanced reading in the literature. Naturally, discussion of highly technical issues in fine detail is beyond the scope of this chapter.

This chapter essentially has two main parts. The first part is on the techniques for testing the price changes (more precisely, abnormal return) in the short run (i.e., days) around the time of the event. The second part describes the methods for analyzing price movements during a long period (e.g., 5 years) following the event. The first part of this chapter is largely self-contained so that the second part is optional to those who need to examine only short-term abnormal return. However, it is advisable that those with the research objective of examining long-term abnormal return first familiarize themselves with the tests of short-term abnormal return in order to install an understanding of the rationale underlying an event study. Such an understanding is essential because the tests of long-term abnormal return are fundamentally an outgrowth of the methodology fundamentally designed for a short window.

The first part of this chapter consists of sections 2 and 3. Sections 4 and 5 constitute the second part. In section 2, the process of measuring short-term abnormal return is described. This includes identification of an event date, definition of an event window, and the common procedures in estimating the price reaction to an event announcement. Section 3 provides notes on hypothesis testing, which is necessary for assessing statistical reliability of the observed abnormal returns. The common statistical issues are also discussed in this section. The common approaches to measuring abnormal return in the long run after the event are presented in section 4. In section 5, the methodological issues surrounding tests of long-term abnormal return and a perspective on these issues are discussed. This chapter ends with take-home messages offered in section 6.

2. Measuring announcement-period abnormal return

The process of measuring abnormal return begins by correctly identifying an event date, which is followed by defining the length of the event window. As discussed below, the event window should be as short as possible. Once the appropriate window length is defined, abnormal return can be meaningfully estimated against the benchmark or expected return with reference to the *correctly* identified event date.

2.1 Identifying the event date

Great care must be taken when identifying the event date. In most event studies, the event date is referred to as the *announcement* date. This is the date on which firms conducting an event under investigation publicly announce for the *first time* their decision to conduct the event, e.g., an acquisition, equity offering, CEO appointments, and capital expenditure. Hence, an event date is the day on which the market learns information about the event for the first time. That is, an event date must be the date of a *surprise*. There are other dates related to the announcement date, for example, the date on which a firm files with the SEC its intention to make an equity offering. The filing date cannot be considered as an event date. This is simply because issuers usually do not publicly announce when they make the filing although the filing date is kept as a public record. There is no arrival of new information in the market on the filing date.

The utmost importance of accurately identifying the date of a surprise also applies to CG studies. The official public announcement date for several CG-related events are not always the date of a surprise as such. One example is the date on which an institutional investor publicly announces for the first time its decision to place a firm on its watch list for a substandard governance practice. If such a decision by some protocol is made following publicly disclosed communications between the investor and the firm, the observed or recorded public announcement date of the decision should not be treated as an event date. Much of the information about the decision is already transmitted into the public domain through the preceding public disclosure of the discourse.

In an attempt to gauge the valuation effects of changes in firms' governance mechanisms and/or practice, a number of studies examine how firms' share prices react to the introduction of, or change in, CG-related regulations. While this is a fruitful empirical strategy as regulatory changes are exogenous to firms, accurate identification of the event date remains challenging. In most cases, the official enactment date or the date on which the regulation comes into effect should not be taken as the date of a surprise. The enactment or effective date of regulatory changes is usually known, or at least well anticipated, in advance. Thus, any regulatory effect is likely to be impounded into share prices before the date identified as the event date. To ensure an accurate regulatory event date, researchers can consider the path taken by Larcker et al. (2011). These authors collect *both* the date on which each regulation they examine was officially introduced and the date on which it first appeared in the news media, and see whether the two dates are the same. The date on which a regulatory change is reported in a newspaper in and of itself is not always an accurate event date. For studies examining firms experiencing a change in their CG rating, similarly, the date on which an updated rating table is uploaded on the rating agency's website, or the website of some other organization, is not necessarily the date of a surprise. When extracting an event date from a website, it is crucial to ensure at least that the ratings are updated in a systematic pattern, e.g., always updated on a certain trading day of each month.

The obvious consequence of mis-identifying the true event date is a measurement error that predictably pushes the statistical result, i.e., the average market reaction or abnormal return, towards being insignificant. Assuming that investors significantly react to an observed change in CG (as would be implicitly assumed by researchers in formulating their hypothesis), studies adopting a wrong event date will systematically miss the significant market reaction. To this extent, one inference typically ensues: the effect of the CG change is not important, or the market fails to react to the CG change.

One useful way to analyze the valuation effects of CG attributes of firms is to employ as a laboratory some corporate decisions with an event date that can be accurately identified. Among several decisions, one popular candidate is corporate acquisitions. For instance, one may examine how the market reaction to the acquisition announcement differs across acquirers with various CG attributes (see, e.g., Masulis et al., 2007; Wang and Xie, 2009; Cai and Sevilir, 2012; Schmidt, 2015; Masulis and Zhang, 2019). Within the framework of the conventional event study setup, one may also analyze the relation between event firms' CG attributes (including CEO attributes) and deal characteristics (see, e.g., Grinstein and Hribar, 2004; Harford et al., 2012; Jenter and Lewellen, 2015). It is also possible and fruitful to employ these approaches to announcements of securities issuances (e.g., Di Giuli and Laux, 2021).

2.2 The event window

Suppose the true event date has been accurately identified. An abnormal price movement or return due to, as a proxy for the wealth effect of, the event should *ideally* be measured on this date. An event study is all about precision in measuring a price response, or how the market reacts, to a surprise. Traditionally, announcement dates are collected from a newspaper, e.g., the Wall Street Journal. The section in a newspaper dedicated to reporting public announcements by firms usually lists the announcements made on the stock exchange on the *previous* day. Hence, the actual event date is the day preceding the press date. In reality, not all investors actually go to the stock exchange every day to observe announcements made by firms.

However, it is conceivable that investors on average read newspapers to see if there is any corporate announcement that may be relevant to their portfolio. As a result, studies employing data collected from a newspaper or press document, which are usually the earlier studies, measure abnormal return over a 2-day period or window covering the press date and the previous day. Such a period is commonly referred to as the announcement period, and in this case, would be specified as the event window (−1, 0) where day 0 is the press date. This is why abnormal return is measured over a multiple-day window even though an event is announced only once on a particular date. The event window may be widened to include days before day −1 in order to capture the effect of information leakage, e.g., (−10, 0). Considering the importance of CG mechanisms to firms' stakeholders, it would be difficult to rule out the possibility of information leakage of CG-related events. That is, the nature of the CG event being analyzed needs to be taken into account when selecting the window length. Due to strong belief in market efficiency, as we should also note, an event window covering days after day 0 has traditionally been uncommon.

Announcement dates are also often collected a periodic magazine, e.g., Acquisitions Monthly. The announcement date reported in such a source would be the date on which the announcement is made on the stock exchange. In this case, a reasonable event window may be (−1, +1) or (−5, +5) in order to allow for human errors in entering data into the source. Such symmetric event windows have become common among recent studies. Since the late 1990s, an electronic database such as the Securities Data Company (SDC) database and Bureau van Dike (BvD), have become increasingly available to authors. An electronic database, too, is subject to human errors alike. For example, Masulis et al. (2007) point out that the vast majority of acquisition announcement dates recorded by the SDC are correct: the recorded announcement dates that are inaccurate are usually off by no more than two trading days. Owing to this note, subsequent studies of corporate acquisitions using the SDC database tend to adopt the (−2,+2) window, which is reasonable.[1] Of course, one may decide to adopt the (−10,+10) window to reduce the chances of missing the true announcement date.

[1] For the BvD, we know of no authors that have made a similar note. But then again, to the best of our knowledge, the BvD has been nowhere near as popular as the SDC for announcements of corporate acquisitions or securities offerings.

However, a longer window may come with a material cost. The longer the window, the noisier the estimate of abnormal return becomes.

Unless there is a specific reason or economic issue to investigate, an event window should be as short as possible. For the reasons discussed above, nevertheless, a realistic event window is longer than one day.

2.3 Estimating benchmark return and abnormal return

There are several alternative ways to measure abnormal return to a firm that conducts a corporate event, or is involved in an event. That is, one can assume a number of alternative return-generating processes in estimating benchmark or expected return. The most commonly adopted expected return model is the OLS-based *statistical* model in equation (1), also known as the market model:

(1) $r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it},$

Where r_{it} is return to event firm *i* observed on day *t*, and r_{mt} return on a market-wide index (as a proxy for market return) observed on day *t*. As a proxy for the market-wide index, the value-weighted index and equally weighted index are equally common. ε_{it} is the regression error term, which is assumed to be normally identically and serially uncorrelated with a mean of zero and homoscedastic variance of $\sigma^2_{\varepsilon_i}$ (this assumption is commonly referred to as the i.i.d. assumption). Both r_{it} and r_{mt} are calculated as continuously compounded return, which is compatible with the classical linear regression model framework. Using continuously compounded return is also intuitively appealing in the sense that an economy works continuously, and hence, reinvestment of output made accordingly. Return calculation should include not only the price change, but also dividends.[2]

By running equation (1) using data from the period *preceding* the event window, abnormal return can be estimated as a *forecast* error or regression error calculated out of sample:

(2) $$ar_{i\tau} = \hat{\varepsilon}_{i\tau} = r_{i\tau} - \left(\hat{\alpha}_i + \hat{\beta}_i r_{m\tau}\right),$$

where $r_{i\tau}$ and $r_{m\tau}$ are, respectively, return on firm *i* and the market return observed on day τ during the event window. Conceptually, the term $\left(\hat{\alpha}_i + \hat{\beta}_i r_{m\tau}\right)$ represents the return firm *i* would be expected to achieve if there was no event, i.e., the benchmark or expected return. Therefore, $ar_{i\tau}$ is a measure of abnormal return to (i.e., abnormal share price movement of) firm *i* on event day τ.

For example, let us define the announcement date (i.e., event date) for a firm's inclusion in the good CG index as day 0, and the announcement period as a 5-day window (−2, +2).[3] The terms $\hat{\alpha}_i$ and $\hat{\beta}_i$ can be estimated by running equation (1) using r_{it} and r_{mt} observed on day −255 through day −6, i.e., days during the period before the event. This period can be referred to as the estimation window (−255, −6). Abnormal return to firm *i* on each day τ during the 5-day event window can then be calculated using equation (2). It should be noted that, by definition, the estimation window is assumed to be a quiet period or period of *normal return*.

Unfortunately, there is no theoretical guidance on how long the estimation window should be. In practice, the window should be long enough so that there are

---

[2] This is because total return to shareholders or equity investors can be divided into the capital-gain and income components. Although failing to include dividends in the return calculation may not materially affect the results based on a short window, it is obviously important to include dividends if one is to analyze long-term abnormal returns.
[3] The inclusion in such an index would typically be hypothesized to convey new information to the market that the true quality of the firm's CG mechanism is good, or better than what investors had expected based on the information set publicly available prior to the announcement of the inclusion.

sufficient observations to get a stable estimate for $\hat{\alpha}_i$ as well as $\hat{\beta}_i$. Also, the longer the estimation window the statistically more precise $\hat{\alpha}_i$ and $\hat{\beta}_i$ become (see also section 3.2). However, use of a very long estimation window may come at a cost. If the nature of the firm's business operations has recently changed, $\hat{\alpha}_i$ and $\hat{\beta}_i$ estimated using a long pre-event window will represent not only the firm's current business profile, but also outdated information about its risk-return characteristics. To the extent that corporate events reflect firms' response to fundamental change in their business environment or product market, it is possible that the risk-return characteristic of typical event firms is materially different from what it was, say five or six years, earlier. A long estimation window may also cover an earlier event. In studies of new CEO appointments, for example, a long estimation window may well contain the announcement of a previous CEO announcement. As discussed below, this can render the estimation of the model parameters problematic.

Considering the efficiency gains and potential costs of a long estimation window, the length of one or two years of daily data seems a reasonable balance for several types of events. Since the choice of the estimation window length is admittedly arbitrary, one useful way to assess the impact of the window length on quality of the results is to employ alternative window lengths and see if the results based on different lengths lead to the same or different conclusions.

Due to the ease of estimation and practicality, the market-adjusted model has recently become increasingly popular. In this model, abnormal return to firm *i* is measured as:

(3) $ar_{i\tau} = r_{i\tau} - r_{m\tau}$.

Here, the market return is therefore the assumed benchmark return for all event firms in the sample. While equation (2) allows the beta risk to vary across event firms, the market-adjusted model in equation (3) assumes that the beta risk is unity across firms. Despite this challenging assumption, the simulation results of Brown and Warner (1980) show that, for short event windows, adjusting for the systematic (beta) risk does not improve the quality of abnormal return estimates.

Though appearing to be a lazy man's approach, this low-cost model has additional merit. Many firms repeatedly conduct an event within a given period of time (e.g., 1 or 2 years). For instance, several recent merger studies report that many firms in their samples make multiple acquisitions within a period of three years (e.g., for U.S., Fuller et al., 2002; Ekkayokkaya and Paudyal, 2015; for U.K., Ekkayokkaya et al., 2009b). In such studies, there is no an estimation window that is free from the event under analysis for a large portion of the sample firms. Studies of CG events are also prone to a similar problem. Given its nature, a CG event may not occur in isolation. The inclusion in the good CG index, for example, is obviously an outcome of economically significant wealth-maximizing decisions observed in the recent past.[4] As a result, it is clearly possible that the estimation window in a CG-related event study covers important events with a predictable valuation effect. That is, the estimation window may well be a period of predictable abnormal return, which in turn would bias $\hat{\alpha}_i$.

Because abnormal return in equation (3) is estimated directly from the event window, the market-adjusted model does not suffer from this problem. In large part due to its practicality, the model has been widely adopted as a return benchmark in recent event studies.

[4] A CG event can also be systematically followed by subsequent material events. To the extent that the good CG index inclusion reduces information asymmetry about firms, for example firms may well raise capital following their index inclusion. As discussed in section 4, subsequent events have an important implication on tests of post-event long-term abnormal return.

3. Hypothesis testing

Once abnormal return is measured for each event firm in the sample, statistical or hypothesis testing is needed to objectively determine whether the event under investigation produces any economically meaningful and statistically reliable value impact on the sample firms. The first step in hypothesis testing is to calculate the average, or median, abnormal return for the sample, i.e., to form an event portfolio and calculate the abnormal return on the portfolio. The next step is to assess the statistical significance of the point estimate (i.e., average or median abnormal return). In doing so, assessment of the economic significance is often overlooked even though we are interested in corporate events primarily because such events are expected (for good theoretical reasons) to produce *material* wealth effects on shareholders. Section 3.1 describes the event-portfolio formation and discusses related issues. In section 3.2, the commonly applicable test statistics are discussed. A perspective on statistical significance in coexistence with economic significance is then offered in section 3.3.

3.1 Aggregation of abnormal returns and forming an event portfolio

For reasons discussed in section 2.2, the wealth effect of corporate events is typically measured over an event window, e.g., a 5-day (−2, +2) window. Thus, $ar_{i\tau}$ needs to be summed across T days during the event window to yield *cumulative abnormal return* (CAR):

(4) $CAR_i = \sum_{\tau=1}^{\mathrm{T}} ar_{i\tau}$,

In order to draw a meaningful inference about the general value impact of the event, we need to aggregate CAR_i's across firms in the sample. That is, we need to form an event portfolio or calculate an average percentage CAR for the sample of *n* event firms:

(5) $\overline{CAR} = \frac{1}{n}\sum_{i=1}^{n} CAR_i$.

Obviously, equation (5) assumes that an equal amount is invested across firms in the sample. In other words, all of the sample firms are given equal importance or weight regardless of their size. An equal weighting scheme, i.e., forming an equally weighted portfolio, is reasonable if the research objective is to measure the *typical* value impact of an event, e.g., how the market assesses the value impact of the good GC index inclusion for a typical firm. When the objective is to measure the *aggregate* wealth effect of an event, on the other hand, calculating an average abnormal *dollar* return ($\overline{CAR}_D$) is appropriate:[5]

(6) $\overline{CAR}_D = \frac{1}{n}\sum_{i=1}^{n}[V_i CAR_i]$,

where V_i is the market capitalization (or, market value) of common equity of firm *i* observed on the day *prior to* the event window, e.g., day −3. Since abnormal return is estimated in event time, it is important, especially when the sample period is long, that V_i is standardized at each point in time using an appropriate deflator. One commonly used deflator is the price level of a value-weighted market index.[6] Importantly, standardization helps ensure that V_i is comparable in real term across time. Estimating abnormal dollar return is equivalent to estimating CAR for a value-weighted portfolio. Percentage CAR for a value-weighted portfolio ($\overline{CAR}_{VW}$) can be calculated as:

(7) $\overline{CAR}_{VW} = \sum_{i=1}^{n} w_i CAR_i$,

where $w_i = \frac{V_i}{\sum_{i=1}^{n} V_i}$. It should be noted that while it is possible to calculate a median abnormal dollar return with equal weighting, it is not possible to do so meaningfully with value weighting.

[5] For an insightful discussion on the economics of abnormal dollar return as a measure of wealth effects, see Malatesta (1983).

[6] See, for example, Mitchell and Stafford (2000), Boehme and Sorescu (2002), Ekkayokkaya et al. (2009a, 2009b).

3.2 Statistical properties and test statistics

The objective of testing whether an event in general has any statistically reliable value impact on the sample firms implies the null hypothesis that the average (or, median) CAR is zero. Therefore, one is to test whether the average CAR significantly differs from zero, and as a result, needs to estimate the variance of the estimated CAR. The statistical assessment of the CAR based on the market model in equation (2) is first described, and then followed by the description of the assessment of the CAR based on the market-adjusted model in equation (3). Given the assumption of the beta of unity across firms underlying the use of the market-adjusted model, an alternative approach to test the null of hypothesis of zero abnormal return is also described.

3.2.1 The market model

When $ar_{i\tau}$ is measured using the market model and the equal weighting scheme adopted as in equation (5), the variance of $\overline{CAR}$ ($\bar{\sigma}^2$) can be calculated as:

(8) $$\bar{\sigma}^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2,$$

where σ_i^2 is the variance of each CAR_i (again, over an event window of T days) and estimated as:

(9) $$\sigma_i^2 = \sum_{\tau=1}^{\mathrm{T}}\left[\hat{\sigma}_{\varepsilon_i}^2 + \hat{\sigma}_{\varepsilon_i}^2\left(\frac{1}{d} + \frac{(r_{m\tau}-\bar{r}_m)^2}{\sum_{t=1}^{d}(r_{mt}-\bar{r}_m)^2}\right)\right].$$

The bracketed term the variance of an individual $ar_{i\tau}$, where $\bar{r}_m$ is the simple average of market returns observed during the estimation window of *d* days. The term $\hat{\sigma}_{\varepsilon_i}^2$ is the variance of the market-model regression in equation (1) for firm *i* and estimated using the return observations in the *d*-day estimation window:

(10) $$\hat{\sigma}_{\varepsilon_i}^2 = \frac{1}{d-k}\sum_{t=1}^{d}\left(r_{it} - \hat{\alpha}_i + \hat{\beta}_i r_{mt}\right)^2.$$

Since there are two parameters in the market model, $k = 2$. As can be seen from both equations (9) and (1), the larger the value of d, i.e., the longer is the estimation window, the smaller $\hat{\sigma}_{\varepsilon_i}^2$ and σ_i^2 become, ceteris paribus. Thus, one statistical benefit of using a long estimation window is an increase in the precision of $ar_{i\tau}$. As discussed in section 2.3, nevertheless, the potential problems associated with a very long estimation period should not be ignored.

For a large sample of event firms, the equally weighted $\overline{CAR}$ from equation (5) under the null hypothesis of zero abnormal return is distributed as: $\overline{CAR} \sim N(0, \bar{\sigma}^2)$. A test of the null can be conducted using the following test statistic:

(11) $t = \overline{CAR}/\bar{\sigma}$.

In practice, it is safe to assume that this test statistic, t, follows Student's t distribution because it approximates the normal distribution for a large number of observations, e.g., 120 or more. Here, it is useful to note that a two-tailed test is generally preferred to a one-tailed test since it is easier to reject the null using the later. For this reason, results reported based on a one-tailed test without strong priors on the direction of the event's impact are often viewed as weak.

Often, a theoretical hypothesis predicts that $\overline{CAR}$ for one group of firms is larger or smaller than that for the other group, e.g., between firms that adopt a pay-for-performance compensation scheme and those that do not. A test statistic for assessing the difference between the CAR for group 1 ($\overline{CAR}_1$) and the CAR for group 2 ($\overline{CAR}_2$) can be calculated as:

(12) $t_{diff} = (\overline{CAR}_1 - \overline{CAR}_2)/\sqrt{\bar{\sigma}_1^2 + \bar{\sigma}_2^2}$,

where $\bar{\sigma}_1^2$ and $\bar{\sigma}_2^2$ are, respectively, the variance of $\overline{CAR}_1$ and $\overline{CAR}_2$. As with equation (11), the test statistic t_{diff} can be assumed to follow Student's t distribution.

As mentioned in section 3.1, measuring dollar CAR ($\overline{CAR}_D$) in equation (6) or value-weighted percentage CAR ($\overline{CAR}_{VW}$) in equation (7) can be desirable from some theoretical point of view. The variance of $\overline{CAR}_D$ ($\bar{\sigma}_D^2$) can be calculated as:

(13) $\bar{\sigma}_D^2 = \frac{1}{n^2}\sum_{i=1}^{n}(\sigma_i^2 V_i^2)$.

Similar to the case of $\overline{CAR}$, the null hypothesis of zero abnormal dollar return can be conducted using the test statistic:

(14) $t_D = \overline{CAR}_D / \bar{\sigma}_D$.

Using the same structure, the variance of $\overline{CAR}_{VW}$ ($\bar{\sigma}_{VW}^2$) can be calculated as:

(15) $\bar{\sigma}_{VW}^2 = \sum_{i=1}^{n}(\sigma_i^2 w_i^2)$,

and the null of hypothesis of zero value-weighted percentage abnormal return can be conducted using the test statistic:

(16) $t_{VW} = \overline{CAR}_{VW} / \bar{\sigma}_{VW}$.

Both t_D and t_{VW} can also be assumed to follow Student's t distribution.

To test the difference in abnormal return between two groups of firms, the test statistics in equations (17) and (18) can be employed for the difference in $\overline{CAR}_D$ and $\overline{CAR}_{VW}$, respectively:

(17) $t_{D,diff} = \left(\overline{CAR}_{D,1} - \overline{CAR}_{D,2}\right) / \sqrt{\bar{\sigma}_{D,1}^2 + \bar{\sigma}_{D,2}^2}$, and

(18) $t_{VW,diff} = \left(\overline{CAR}_{VW,1} - \overline{CAR}_{VW,2}\right) / \sqrt{\bar{\sigma}_{VW,1}^2 + \bar{\sigma}_{VW,2}^2}$,

where subscripts 1 and 2 denote groups 1 and 2, respectively. These test statistics can also be assumed to follow Student's t distribution.

When the sample is relatively small, it is useful to assess the median CAR in addition to $\overline{CAR}$ from equation (5). The average of a small sample can be sensitive to the presence of outliers, if any. The statistical significance of a median CAR can be assessed using the Wilcoxon signed-rank test. The non-parametric equivalent for equations (12), (17) and (18) is the Mann-Whitney U test, also known as Wilcoxon rank-sum test. Since non-parametric tests are essentially tests of location, the only required input for these tests is CAR_i from equation (4). Options for these nonparametric tests are regularly available in most statistical/econometrics software packages. As mentioned in section 3.1, a median value-weighted CAR is not economically meaningful. For technical details and estimation procedure of these tests, readers are referred to Brown and Warner (1985) and Hollander and Wolfe (1999). These non-parametric tests are also applicable to the estimation of CAR based on the market-adjusted model in equation (3).

3.2.2 The market-adjusted model

With an equal weighting scheme, the only way to estimate the variance of $\overline{CAR}$ calculated using $ar_{i\tau}$ from equation (3) ($\sigma^2_{\overline{CAR}}$) is to estimate it as a cross-sectional sample variance:

(19) $\sigma^2_{\overline{CAR}} = \frac{\sum_{i=1}^{n}(CAR_i - \overline{CAR})^2}{n-1}$, and

the null hypothesis of zero percentage abnormal return can be conducted using the test statistic commonly known as the simple t-test:

(20) $t_{simple} = \overline{CAR}/\sigma_{\overline{CAR}} \cdot \sqrt{n}$.

To test the difference between two groups of firms, it is appropriate to use the independent-samples t-test which has the same structure as the test statistic in equation (12). Both the simple t-test and independent-samples t-test are readily available in virtually all statistical software packages. These test statistics and the ones below in this subsection follow Student's t distribution.

If one is to test the null hypothesis of zero abnormal dollar return, the following test statistic can be used:

$$(21) \quad t_{D,simple} = \overline{CAR}_D \Big/ \left(\sigma_{\overline{CAR}} \sqrt{\textstyle\sum_{i=1}^{n} V_i^2} \right) (n) \,.$$

It should be noted here that it would be incorrect to test the null of $\overline{CAR}_D = 0$ by directly applying the simple t-test. This is because, unlike CAR_i, V_i is predetermined rather than being a random variable. Similar reasoning holds for the null of $\overline{CAR}_{VW} = 0$. The test statistic below can be used to test the null of zero value-weighted percentage abnormal return:

$$(22) \quad t_{VW,simple} = \overline{CAR}_{VW} \Big/ \left(\sigma_{\overline{CAR}} \sqrt{\textstyle\sum_{i=1}^{n} w_i^2} \right) .$$

To test the difference in $\overline{CAR}_D$ and $\overline{CAR}_{VW}$ between two groups of firms, a test statistic of the same structure as those in equations (17) and (18), respectively, can be employed. It is also worth noting that the test statistics described in this subsection can be viewed as implicitly assuming that the variance of individual CAR_i's is constant and equal to $\sigma^2_{\overline{CAR}}$.[7]

3.2.3 The cross-sectional regression approach

As mentioned in section 2.3, although the market-adjusted model does not require use of the estimation window, it assumes that the beta risk is unity across firms. It is possible to address, in part, this assumption and still avoid using the estimation window by allowing the benchmark return model to reflect the average beta risk (as well as other risk factors) of the sample firms. In a nutshell, this approach employs a

[7] Therefore, these test statistics are different from their counterparts based on the market model in equation (2), which by structure take account of individual variances of CAR_i observations which may be heteroscadestic.

cross-section of data and measures average abnormal return in the regression framework of Jensen's alpha. Here, the statistical significance of abnormal return can be obtained as part of the regular regression routine. Assuming the CAPM as the return generating process, average abnormal return for the sample firms can be estimated by running the following regression:[8]

(23) $\left(r_{i\mathrm{T}} - r_{f\mathrm{T}}\right) = \alpha_{\mathrm{T}} + \beta_{\mathrm{T}}\left(r_{m\mathrm{T}} - r_{f\mathrm{T}}\right) + \varepsilon_{i\mathrm{T}}$,

where $r_{i\mathrm{T}}$ is return to firm *i* observed over the T-day event window and calculated as: $r_{i\mathrm{T}} = \sum_{\tau=1}^{\mathrm{T}} r_{i\tau}$. If return is calculated as continuously compounded return, $r_{i\mathrm{T}}$ represents buy-and-hold return to firm *i* over a T-day period. $r_{f\mathrm{T}}$ and $r_{m\mathrm{T}}$ are the corresponding risk-free return and market return: both of which are calculated in the same fashion as $r_{i\mathrm{T}}$. Technically speaking, the market return can be calculated from either an equally weighted or value-weighted market index. Since the theoretically optimal market portfolio is a value-weighted portfolio, use of a value-weighted market index is advocated here.

In equation (23), the estimated intercept ($\hat{\alpha}_{\mathrm{T}}$) is a measure of equally weighted average abnormal return to the sample firms over the T-day event window (Jensen's alpha). The standard error for testing the significance of $\hat{\alpha}_{\mathrm{T}}$ is readily provided by the regression procedure. The potential impact of non-constant variances of event firm returns on the statistical significance of all model parameters (including $\hat{\alpha}_{\mathrm{T}}$) can then be conveniently accounted for by employing the White heteroscedasticity-consistent standard error, which is nowadays a standard built-in option in practically all of the

[8] This cross-sectional regression approach to testing the null of zero abnormal return has been adopted by Draper and Paudyal (2006) and Ekkayokkaya et al. (2009b). In the light of the recent evidence that the size effect, book-to-market effect as well as return persistence explain a cross-section of stock returns above and beyond the beta (systematic) risk (e.g., Fama and French, 1996), one can directly extend equation (23) to include these additional priced risk factors.

available econometric software packages. The average systematic risk of the sample firms is captured by the estimated slope coefficient ($\hat{\beta}_{\mathrm{T}}$).

A test of there being a statistically significant difference in abnormal return between two groups of firms can be carried out by adding to equation (23) an indicator variable for the grouping and an interaction term between the indicator variable and the risk factor. This is to run the following regression model:

$$(24) \quad \left(r_{i\mathrm{T}} - r_{f\mathrm{T}}\right) = \alpha_{\mathrm{T}} + \beta_{\mathrm{T}}\left(r_{m\mathrm{T}} - r_{f\mathrm{T}}\right) + \gamma_{\mathrm{T}}(G) + \delta_{\mathrm{T}}\left(\left(r_{m\mathrm{T}} - r_{f\mathrm{T}}\right) \cdot G\right) + \varepsilon_{i\mathrm{T}} \,,$$

where G is the indicator variable taking the value of 1 when the term $\left(r_{i\mathrm{T}} - r_{f\mathrm{T}}\right)$ is observed for firm group 1 (or, group 2) and zero otherwise. The estimated coefficient $\hat{\gamma}_{\mathrm{T}}$ measures the difference in abnormal return (i.e., Jensen's alpha) between the two groups of firms. The interaction term is incorporated to account for the difference in the slope coefficient, i.e., average systematic risk, between the firm groups. Omitting the interaction term when such a difference is important would be to load on G not only the difference in abnormal return but also the difference in risk, making $\hat{\gamma}_{\mathrm{T}}$ of little use in practice.[9] To the extent there are theoretical priors to warrant a statistical test of the difference in abnormal return between firm groups, it is likely that the groups also have materially different risk-return characteristics. For example, the risk-return combination of investments is likely to vary between firms that adopt a pay-for-performance compensation scheme and those that do not.

Equation (23) can be adjusted to test the null hypothesis of zero abnormal dollar return over a T-day event window. To do so, following Eckbo and Thorburn (2000), the terms $\left(r_{i\mathrm{T}} - r_{f\mathrm{T}}\right)$ and $\left(r_{m\mathrm{T}} - r_{f\mathrm{T}}\right)$ are pre-multiplied by V_i. The regression model can then be re-run. The resulting intercept provides a measure of an equally weighted

[9] The statistical importance of the difference in the slope coefficient can be assessed from the significance of the estimated interaction term coefficient ($\hat{\delta}_{\mathrm{T}}$).

average abnormal dollar return. The structure of equation (24) can be directly applied should one have theoretical priors to expect a difference in abnormal dollar return between firm groups. It is noted here that it is not feasible to estimate value-weighted average percentage abnormal return using the framework of equations (23) and (24).

3.3 Interpretation of results – statistical and economic significance

For most of the times, statistically significant results are also economically significant or meaningful. This is because a point estimate generally has to be economically sizeable for it to be statistically significant. Often enough, however, a statistically significant abnormal return is very small in magnitude. A small, and yet significant, point estimate can easily arise from a large sample.

As an illustration, consider an equally weighted $\overline{CAR}$ of 0.3% for a sample of 250 observations with the standard deviation of 2.99. This $\overline{CAR}$ would not be statistically significant, based on the simple t-test, at the conventional levels as the test statistic would be 1.59 with p-value of 0.114. Note here that sample size of 250 observations is not small, and indeed, is more than enough for Student's t distribution to approximate the Normal distribution. Now, suppose with an electronic database, the available sample size becomes substantially larger (which is empirically desirable), say 950. With this sample size, *other things constant*, the test statistic for this 0.3% $\overline{CAR}$ would mechanically become 3.09 with p-value of 0.002, thereby making the same $\overline{CAR}$ significant at the 0.01 level. This pattern tends to be observed more often among recent studies as the sample size employed typically gets larger and larger, thanks to the burgeoning availability of electronic databases. Though naïve, this simple illustration offers another important message. In a large sample, weak statistical significance, e.g., at the 0.10 level, is unlikely to be viewed as sufficiently reliable evidence to reject the null of zero abnormal return. This message is compatible with a number of event studies in the mainstream journals interpreting their results as statistically significant if the *p*-value is 0.05 or less.

A large number of authors condition their interpretation of results primarily on statistical significance. While focusing on statistical significance yields objectivity and is the right thing to do, it is also equally important to pay attention to the economic significance of the point estimate(s). When we are empirically interested in analyzing a given corporate event, we are so because we have theoretical reasons to expect the event to produce an economically meaningful value impact, e.g., to destroy shareholders' wealth in a material fashion. In the above illustration, an appropriate inference that can be drawn assuming the sample size of 950 observations is that the event can be expected to systematically affect wealth, but its economic effect is only small. To this extent, focusing solely on statistical significance may well give an incomplete picture of the wealth effect of the event.

It is difficult to make a judgment call on what is and is not economically significant. Unfortunately, there exists no hard and fast rule (i.e., theoretical guidance) on how to gauge economic significance. One way to do so is to compare the point estimate, e.g., 0.3% CAR over a 5-day event window, with the results reported by the comparable existing studies. Given this average CAR, another useful way is to compare it with the degree of price movement other strands of the finance literature considers as a large price movement. For example, a 1.5%, 2.0% or 2.5% increase or decrease in the CRSP market index, either equally weighted or value-weighted, is defined in studies of institutions' trading behavior as a large price movement (e.g., Dennis and Strickland, 2002; Lipson and Puckett, 2010). Clearly, a statistically significant $\overline{CAR}$ of 0.3% is far from being economically significant: it would be so if it were, say, 1.1%.

3.4 Statistical problems

There are two common problems associated with the inference about the statistical significance of abnormal return estimates. These problems are related to the i.i.d. assumption; namely, heteroscedasticity or varying specific variances, and cross-sectional correlation among abnormal returns. In the following subsections, these problems and their empirical importance are discussed.

3.4.1 Problem of heteroscedastic variances

For heteroscedasticity, the conventional practice is to standardize each individual CAR using its own standard deviation, i.e., $[CAR_i/\sigma_i]$. In this case, the applicable test statistic under the null of hypothesis of zero abnormal return by definition becomes standard normal (also commonly known as the Z-statistic):

(25) $$Z = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{CAR_i}{\sigma_i}.$$

This standardization accounts for heteroscedasticity in the estimated abnormal return and can make the test statistic more powerful in detecting abnormal return (Brown and Warner, 1985). Since equation (25) requires specific variances of individual CARs, it is applicable only when the market model (equation (2)) is used to estimate CAR_i. Alternatively, one can also account for heteroscedasticity in the variance component of the test statistic. By structure, $\bar{\sigma}^2$ in equation (8) incorporates a specific variance of CAR_i (σ_i^2). Therefore, the structure of the test statistic in equation (11) and its variants in section 3.2.1 accounts for varying specific variances. The key difference between equations (25) and (11) is that the former assumes that the distribution of the standardized CAR is unit normal whereas the latter makes no such an assumption.

As discussed earlier, the test statistics in section 3.2.2 make use of a cross-sectional sample variance. Because the market-adjusted $ar_{i\tau}$ from equation (3) is calculated entirely within the event window, the variance of the resulting average CAR (whether percent or dollar) is by structure the cross-sectional sample variance. Therefore, these test statistics are not subject to the heteroscedastic variances inherent in the case of the abnormal return estimation based on the market model from equation (2). Since the cross-sectional regression approach in section 3.2.3 also estimates abnormal return entirely within the event window, it, too, is not subject to such heteroscedastic variances. As mentioned in the section, although this regression approach is still subject to the non-constant variances of event firm returns, the remedy is readily available from the regression procedure.

3.4.2 Problem of cross-sectional correlation

A potentially more concerning statistical problem is the cross-sectional correlation among abnormal returns: also known as the cross-dependence problem. This problem arises when individual abnormal returns (i.e., CARs) are contemporaneously, or cross-sectionally, correlated across firms. As has been widely observed, abnormal returns to firms that conduct the same event tend to move together *in event time* in a noticeable manner. That is, the cross dependence is positive in nature.

While all of the test statistics discussed above assume random sampling, corporate events, as correctly noted by Mitchell and Stafford (2000), are *not* random events. While some firms choose to participate in a certain event (e.g., make an SEO), some others choose not to do so. It is also well documented that M&As do not take place randomly across industries or time. Rather, these activities occur in waves or clusters, exhibit industry clustering, and are also systematically associated with macro-economic conditions (e.g., Mitchell and Mulherin, 1996; Harford, 2005). Other corporate events, such as security issues and share repurchases also occur in waves (see Rau and Stouraitis, 2011). A sample of firms involved in a CG event is no exception. Firms that get included in a good CG index, for example, are likely to have chosen to make wealth-maximizing decisions prior to their index inclusion and have been experiencing good performance, either stock price or operating, or both. As another example, firms adopting a pay-for-performance compensation scheme are likely to have chosen to do so for a similar set of reasons. Accordingly, the stock prices of firms in these examples would move in a similar direction during the event window, i.e., would be positively cross-sectionally correlated. Intuitively, firms that choose to conduct an event do so because they expect to benefit from the event. Thus, event firms are bound to be of certain characteristics, and the market is likely to react to event firms *systematically* differently from how it would react to random non-event firms.

One key consequence of cross dependence is systematic underestimation of the true variance of the estimated average abnormal return, i.e., the mean CAR calculated in equations (5), (6) and (7). This is because while the cross dependence is by and large positive in nature, the conventional variance estimation (e.g., that in section 3.2) assumes zero correlation among abnormal returns. In other words, CAR_i's are assumed to be cross-sectionally independent. Such systematic underestimation of true variance in turn leads to an overstatement of a conventional test statistic, with the upshot being too many rejections of the null of zero abnormal return when it is true.

Cross dependence is most severe when there are overlapping return-calculation periods. That is when a given sample firm is observed: (i) twice or more in the sample; and (ii) within the time span shorter than the event window. Consider a simple case of a single firm conducting the event twice. During the overlapping period, the firm's abnormal return will be counted, i.e., observed, twice. This is where the problem arises. The average of two identical abnormal returns will be the same as each estimate. This is not the case for variance. Assuming cross-sectional independence, the estimated variance of the average (say, equally weighted) would be only half that of each constituent abnormal return. However, the true variance of this average abnormal return is, as it should be, identical to the variance of the constituent abnormal returns, which are essentially a single abnormal return counted twice in the calculation of the average CAR.

One way to account for cross dependence is to make the Crude Dependent Adjustment (CDA) suggested by Brown and Warner (1980). With the CDA, the standard deviation of the average abnormal return is calculated as the standard deviation of the event-time series of abnormal returns observed during the pre-event period (see Brown and Warner, 1980, equations A.4 through A.6). Since the standard deviation is estimated from the average, or portfolio, abnormal returns observed in event time, any cross dependence among individual abnormal returns is captured in the calculation. Another way to deal with cross dependence is to form a portfolio of the sample firms in calendar time. Forming the event firm portfolio in calendar time completely eliminates the cross-dependence problem. Since cross dependence is unlikely to materially affect the statistical inference of abnormal returns estimated from a short event window, the details of the calendar-time portfolio approach are discussed in section 5 where the tests of long-term abnormal return are discussed.

3.4.3 Empirical importance of statistical problems

Both of the heteroscedasticity and cross dependence problems are important statistical aspects of event studies. Regardless of their observed empirical importance, it is essential that one be aware of these aspects to be able to make judgment on the statistical reliability of abnormal return estimates. Empirically, the findings of the simulation studies by Brown and Warner (1980, 1985) indicate that, for a short event window, neither of the problems appears to affect in an important way the statistical significance or the magnitude of abnormal return estimates. Indeed, the findings of a number of empirical event studies also show that the choice of return benchmark (e.g., market model vs. market-adjusted model) does not materially affect the detection of abnormal return.

For instance, in examining abnormal returns around takeover announcements, Draper and Paudyal (1999) employ the return benchmarks in Brown and Warner (1985) with and without the CDA. Their results show that the abnormal return estimates are insensitive to model specification in terms of statistical significance as well as the magnitude of the estimates. Since Draper and Paudyal (1999) employ test statistics that involve and do not involve the standardization discussed in section 3.4.1, the insensitivity of their results serves as an empirical indication that the presence of varying specific variances of abnormal returns is unlikely to affect the statistical inferences of abnormal return estimates. Indeed, the insensitivity of their results also provides non-simulation evidence that adjusting for cross dependence does not affect the quality of abnormal return estimates. In their subsequent large-sample study of takeovers, Draper and Paudyal (2006) employ the cross-sectional regression approach similar to that in equation (23). They assume three return generating processes: the CAPM; Fama-French three-factor model; and the three-factor model with a variable representing the average past return for the sample firms. Especially for the 3-day window surrounding the announcement date, their abnormal return estimates are notably comparable across the employed benchmarks in both magnitude and statistical significance. This pattern can be viewed as a non-simulation indication that accounting for other risk factors in addition to the market risk factor affects neither the magnitude nor statistical significance of abnormal return estimates.

Employing the market-adjusted model as in equation (3) and a large sample of takeovers by European firms, Faccio et al. (2006) examine whether the announcement-period abnormal returns to acquirers differs between takeovers of listed targets and takeovers of unlisted targets. As part of their empirical analysis, they categorically address the potential problem of cross dependence using the calendar-time portfolio approach. Their results show that the difference in abnormal return between the two types of takeovers is comparable in both magnitude and statistical significance whether the average CAR is calculated in event time or calendar time. This pattern is also persistently observed across their various sample partitions. Thus, the findings of Faccio et al. (2006) serve as an additional non-simulation testimony to the insensitivity of short-window abnormal return estimates to cross dependence, if any. Given that cross dependence is severe when there is return overlap, such insensitivity is intuitive. Firms generally do not repeat the same event within a very short period of time, e.g., within 5 or 10 days, thereby making return overlap unlikely.

In sum, it is well documented in the event study literature that neither the choice of return benchmarks nor the statistical problems of heteroscedastic variances and cross dependence of abnormal returns is a serious concern in conducting tests of abnormal return in a short event window. As discussed below, however, this is not the case for tests of abnormal return in a long event window, e.g., a three-year post-event window.

4. Measuring long-term abnormal return

In addition to testing the announcement-period or short-term abnormal return, event studies often include, as material part, examination of abnormal return over a long period of time following the event. For an analysis of post-event abnormal return, this period is essentially the event window. As with the short-horizon analysis, there is no theoretical guidance on how long the event window in the post-event analysis should be. Almost all of the existing studies employ 1-, 2- and 3-year post-event windows with several also focusing on the 5-year window. Obviously, the key difference between tests of short-term and long-term abnormal return is simply the length of the event window.

An attempt to measure an event-induced value impact over a very long period of time exposes tests of long-term abnormal return to contamination. This is so regardless of the statistical approaches, and CG-related event studies are no exception. For instance, a successfully implemented change in firms' CG mechanism can pave the way for subsequent corporate activities. That is, it is clearly possible that a CG event is systematically followed by other events that are of substance to firms. To this extent, the abnormal return detected during the post-event window, if any, is attributable not only to the CG event being studied but also to the subsequent activity(ies).

For the long window analysis, moreover, there are several issues and complications – both conceptual and statistical – that appear to remain debatable. In section 4.1, a brief background of the literature on tests of post-event long-term abnormal return is introduced. In the remaining subsections, the commonly adopted approaches to measuring long-term abnormal return are described.

4.1 Background

Much of the empirical event study literature has, and still does, primarily focused on tests of announcement-period abnormal return (i.e., short-term abnormal return). Such a focus on short-term abnormal return is primarily due to “the strong belief in market efficiency” (Agrawal and Jaffe, 2000, p. 8; see also Masulis et al., 2007, footnote 9), which would dictate zero abnormal return in the period following the event outcome. As an illustration, Jensen and Ruback (1983, p. 20) remark that negative post-event abnormal returns are “unsettling because they are inconsistent with market efficiency and suggest that changes in stock prices overestimate the future efficiency gains from mergers”. Especially during the 1990s, the literature saw a burgeoning interest in tests of post-event long-term abnormal return. Not surprisingly, the typical motivation or discussion of results given in a number of long-term abnormal return studies alludes to market inefficiency, or investors suffering from a behavioral bias(es) one way or another.[10]

Several intuitively and statistically appealing empirical approaches have been put forward in the literature. In the main, these approaches can be categorized into the *event-time* and *calendar-time* approaches, and for both approaches, the benchmark return can be estimated using a k-factor asset pricing model or return to a characteristic-based control firm/portfolio. Yet, each of these approaches is plagued with methodological problems one way or another. Unsurprisingly, there appears to be no consensus on the fool-proof way to measure long-term abnormal return. Based on the existing evidence, it would indeed be safe to generalize that long-term abnormal return estimates are sensitive to model specification. Such sensitivity is, at least in part,

[10] For example, Loughran and Ritter (1995), Loughran and Vijh (1997), Rau and Vermaelen (1998), Boehme and Sorescu (2002). For theoretical arguments predicting patterns of long-term abnormal return, see Shleifer and Vishny (2003).

attributable to the length of the window itself, and yet, the objective is to measure abnormal return during a long window. Since the conventionally adopted windows in a long-horizon test are extremely long, considerable noise is inevitable. To the contrary, the event study methodology is fundamentally designed to capture an abnormal price reaction to the new information, or surprise, that arrives at the market at *a very specific point in time*.[11] In other words, the methodology is not designed to isolate noise from the valuation effect of new information.

Because there are a number of methodological issues surrounding tests of long-term abnormal return, it is worthwhile to briefly mention here the central message of this part of the chapter, in the hope of maintaining tractability of the discussion. Different approaches address different aspects abnormal return measurement. It is important for researchers to contemplate both the economic and statistical properties of abnormal return estimates. Given the known sensitivity of the estimates of long-term abnormal return to model specification, one useful empirical strategy is to employ at least one event-time specification and one calendar-time specification, and then compare the results.

4.2 Firm-specific Fama-French three-factor model

One intuitive way to measure long-term abnormal return earned by an average event firm in the sample is to use a *k*-factor asset pricing model as a return benchmark as in Barber and Lyon (1997a). Here, abnormal return to event firm *i* during the *T*-month window following the month of the event outcome can be estimated in the regression framework of Jensen's alpha. One important implicit assumption underlying

[11] As discussed in section 2.1, it is crucial to identify the true event date, or the precise time at which the market *first* learns about the event.

the choice of window length is that the value impact of an event under the alternative hypothesis lasts for the length of the window. Due to the findings of Fama and French (1992, 1993), subsequent studies of corporate events commonly control for the size and book-to-market (BM) effects when formulating an expected-return benchmark.[12]

(26) $R_{it} - R_{ft} = \alpha_i + \beta_i(R_{mt} - R_{ft}) + s_i SMB_t + h_i HML_t + \varepsilon_{it}$.[13]

R_{it} is return to firm *i* observed in month *t* during the T-month event window. R_{ft} and R_{mt} are the corresponding risk-free return and market return. SMB_t and HML_t are the return spreads observed in month *t* between small and big firms, and between firms with high and low book-to-market ratios, respectively. In other words, these spreads are the size and book-to-market risk factors as in Fama and French (1993).

The estimated intercept ($\hat{\alpha}_i$) is a measure of abnormal return firm *i* earns *per month* during the T-month event window.[14] Here, one could get a sense of per-annum or holding-period return by multiplying $\hat{\alpha}_i$ by 12 or the number of months in the window, respectively. The average, equally weighted or value-weighted, abnormal return can then be calculated by averaging $\hat{\alpha}_i$'s across the sample in the fashion similar to equation (5) or (7), respectively. Similarly, the average monthly dollar abnormal return can be calculated following equation (6). Because each $\hat{\alpha}_i$ is estimated relative to the event month, this approach measures abnormal return in *event time*.

Employing the cross-sectional sample variance as in Barber and Lyon (1997a), the null hypothesis of zero monthly abnormal return can be tested using the test

---

[12] Fama and French (1992) find that firm size and the BM ratio are important and the most robust factors in explaining the cross-section of expected stock returns for the U.S non-financial firms. A subsequent study by Barber and Lyon (1997b) documents that these relations also hold for financial firms in the U.S. market. In the U.K. market, the size and BM factors have also been found as important risk factors (see, e.g., Davies et al., 1999; Gregory et al., 2013).

[13] The subscripts used from this point onwards are not related to and not to be confused with those used in the discussion of tests of short-term abnormal return. Subscripts '*t*' and '*T*' are used in this part simply to maintain compatibility with the convention.

[14] If abnormal return is to be measured during a 3-year window, for example, the number of monthly returns to be included in the regression model is 36.

statistic similar to equation (20) or (22) for equally weighted or value-weighted average percentage abnormal return, respectively. The structure of equation (21) is applicable to the null of zero monthly abnormal dollar return. To account for varying specific variances, the test statistics in equations (27), (28) and (29) can be used to test the null of zero equally weighted and value-weighted percentage monthly abnormal return, and monthly abnormal dollar return, respectively:

(27) $$t_{EW} = \left(\frac{\frac{1}{n}\sum_{i=1}^{n} \hat{\alpha}_i}{\sqrt{\sum_{i=1}^{n} se_i^2}} \right) \cdot n \,,$$

(28) $$t_{VW} = \left(\frac{\sum_{i=1}^{n} w_i \hat{\alpha}_i}{\sqrt{\sum_{i=1}^{n} w_i^2 se_i^2}} \right),$$

(29) $$t_{dollar} = \left(\frac{\frac{1}{n}\sum_{i=1}^{n} V_i \hat{\alpha}_i}{\sqrt{\sum_{i=1}^{n} V_i^2 se_i^2}} \right) \cdot n \,,$$

where se_i is the regression standard error for $\hat{\alpha}_i$. The terms *n*, V_i, and w_i are defined as in section 3. These test statistics follow Student's *t* distribution. Alternatively, one may choose to assume that the distribution of standardized $\hat{\alpha}_i$, i.e., $[\hat{\alpha}_i / se_i]$, is unit normal in testing the relevant null hypothesis of zero abnormal return. To test the difference in average abnormal return between two groups of firms, the structure of the test statistics in equation (12), (17) or (18) can be adopted.

To ensure that the regression estimates are reasonably stable, firms are typically required to have a minimum of 12 or 24 valid returns. Such requirement clearly gives rise to the survivorship bias as only firms that have survived the required minimum post-event are included in the analysis. One way to avoid this survivorship bias is to adopt the regression framework of equation (23). To enter equation (23), firms need to have only 1 month of valid return following the month of the event outcome.

4.3 Event-time characteristic-based return benchmark

Due to the bad model problem (for a more detailed discussion, see section 5.1), a number of studies measure buy-and-hold abnormal return (BHAR) against a characteristic-based return benchmark over a period of *T* months following the month of the event outcome:

(30) $BHAR_{iT} = \prod_{t=1}^{T}[1 + R_{it}] - \prod_{t=1}^{T}[1 + \mathrm{E}(R_{it})]$.

As mentioned earlier, the commonly adopted window length (T) is 12, 24, 36 or 60 months. R_{it} is return to firm *i* observed in month *t* during the *T*-month event window. $\mathrm{E}(R_{it})$ is the return firm *i* is expected to earn during month *t*, i.e., its benchmark return. For estimation purposes, a number of studies calculate return as a simple return. It is also possible to use continuously compounded return as an input to equation (30) – in which case, the right-hand side of the equation becomes: $(\sum_{t=1}^{T} R_{it}^{L}) - (\sum_{t=1}^{T} \mathrm{E}(R_{it}^{L}))$, where superscript *L* denotes continuous compounding. [15]

$\mathrm{E}(R_{it})$ can be estimated as return to the characteristic-based *control firm* or *control portfolio*. Due to the Fama-French findings as mentioned above, most studies select a control firm or firms to constitute a control portfolio that are comparable to event firm *i* in terms of size (market capitalization of common equity) and book-to-market ratio. As with the window length, there is no theoretical guidance on how to do the size and book-to-market matching. Here, the rule of the day is common sense, and one can follow the widely adopted sequential sorting method in Barber and Lyon

[15] Several authors advocate the use of simple return as it is not subject to a downward bias due to Jensen's inequality. More simply, for a given increase (decrease) in price, continuously compounded return is always less positive (more negative) than simple return. The magnitude of this effect can be of some concern if a sample contains a big number of large price changes, e.g., changes larger than 15 percentage points per calculation interval. For small price changes, however, simple and continuous return calculations should yield results that lead to the same conclusion.

(1997a). Loughran and Vijh (1997) employ a slightly different sorting method as well as the Barber-Lyon sequential sorting, and report that their results are qualitatively similar between the two methods.

Unfortunately, there is a bit more to the process of identifying control firms. By definition, a control firm must not conduct the event under analysis at any time during (i) the event window and (ii) the pre-event period of the same length. This requirement also applies to firms to enter a control portfolio. Due to the assumption that the value impact of an event during a T-month window under the alternative hypothesis lasts for T months, it is *not* enough that control firms are a non-event firm only during the event window. It is this requirement that can call the reality of this approach into question in terms of both costs of tracing the activities by the candidate control firms and finding firms that qualify as control firms.

As in the case of tests of short-term abnormal return, $BHAR_{iT}$'s can be averaged with equal or value-weighting, or as an equally weighted average abnormal dollar return. Because $BHAR_{iT}$ is calculated within the event window, the feasible test statistics necessarily rely on the cross-sectional sample variance. The test statistics in section 3.2.2 are therefore applicable to testing an average BHAR. To address the impact of outliers on the abnormal return estimates, the non-parametric tests described in section 3.2.1 can be employed. These non-parametric tests are also applicable to the approach discussed in section 4.5.

4.4 Calendar-time Fama-French three-factor model

In response to the cross-dependence problem, many studies estimate the Fama-French three-factor model in calendar time. That is, a portfolio of event firms is formed in calendar time, and the time series of portfolio returns is regressed on the corresponding time series of the three risk factors. For each calendar month, return is calculated for a portfolio of firms that conduct the event within the previous T months ($T = 12$, 24, 36 or 60 months). The portfolio is *rebalanced* (or, reformed) monthly to drop all firms that reach the end of their period of T months and to add all firms that have just conducted the event.[16] This portfolio formation yields a time series of monthly portfolio returns (R_{pt}). Also importantly, it gives only *one* return for any given calendar month, and as a result, there is *no* cross-correlation at any given point in time. For each portfolio, the Fama-French three-factor model can be estimated in the following regression framework:

$$(31) \quad R_{pt} - R_{ft} = \alpha_p + \beta_p\left(R_{mt} - R_{ft}\right) + s_p SMB_t + h_p HML_t + \varepsilon_{pt}\ .$$

In this *time-series* regression model, R_{pt} is the return to the event-firm portfolio observed in month t, and can be equally weighted or value-weighted. All other variables are defined similarly to those in equation (26). Accordingly, $\hat{\alpha}_p$ is a measure of average monthly abnormal return to the event-firm portfolio during the T-month window following the event. The heteroscedasticity-autocorrelation-consistent (HAC) standard error readily provided by the regression procedure can be used to assess the statistical significance of $\hat{\alpha}_p$.

In correcting for heteroscedasticity, several authors employ the weighted least square (WLS) estimator, instead of the OLS estimator. Commonly, the weights are set proportional to $\sqrt{j}$, where j is the number of firms in each monthly portfolio (see Mitchell and Stafford, 2000). As Mitchell and Stafford show, this weighting assumes that individual-firm residuals are uncorrelated, and hence, "completely defeats the purpose of forming calendar-time portfolios, which is to account for the fact that individual-firm

[16] In this procedure, firms that become delisted before the end of the window are automatically dropped out of the portfolio at the beginning of the month of delisting.

residuals are cross-sectionally correlated" (p. 317). Because heteroscedasticity is a problem of inference common in regression analysis, not just the calendar-time Fama-French three-factor model, it will be reasonable to resort to the readily provided HAC standard error.

To examine whether there is a statistically significant difference in the average monthly abnormal return between two groups of firms, one can apply the dummy variable technique in equation (24). Since the data used in equation (31) is time-series data, however, the observations must be stacked in a chronological order (i.e., time order) for the estimated autocorrelation-consistent standard error to be meaningful.

4.5 Calendar-time characteristic-based return benchmark

Alternative to the calendar-time Fama-French three-factor model is to apply the characteristic-based return benchmark in calendar time. This approach is also known as the calendar-time rolling portfolio approach, first employed by Jaffe (1974) and Mandelker (1974). As its main appeal, it is free from the assumption of parameter stability over time which underlies a *k*-factor asset pricing model. Here, the formation of an event-firm portfolio is similar to that in section 4.4, i.e., assumes monthly rebalancing. For each calendar month *t* and firm *i* that conducts the event within the previous *T* months, abnormal return (AR_{it}) can first be measured against a characteristic-based return benchmark where:

(32) $AR_{it} = R_{it} - \mathrm{E}(R_{it})$.

As with its event-time counterpart in section 4.3, $\mathrm{E}(R_{it})$ can be return on the *control firm* or *control portfolio*. Following Lyon et al. (1999), each month *t*, abnormal return on the event-firm portfolio (MAR_t) can then be calculated as:

(33) $MAR_t = \sum_{i=1}^{j} w_{it} AR_{it}$,

where *j* is the number of event firms with valid return in month *t*. With an equal weighting scheme, $w_{it} = \frac{1}{j}$, and $w_{it} = \frac{V_{it-1}}{\sum_{i=1}^{j} V_{it-1}}$ for a value-weighting scheme. The average monthly portfolio abnormal return for the sample is:

$$(34) \quad \overline{MAR} = \frac{1}{m}\sum_{t=1}^{m} MAR_t \text{ ,}$$

where *m* is the number of months in the time series of MAR_t. Also note that equation (34) applies whether MAR_t is equally weighted or value-weighted. This is because the weighting takes place when a portfolio is formed in each month, i.e., in equation (33). Naturally, equation (34) weights monthly returns equally across the sample period.

The null of zero $\overline{MAR}$ can be conducted using the simple *t*-test as follows:

$$(35) \quad t_{\overline{MAR}} = \left(\frac{\overline{MAR}}{\hat{\sigma}_{\overline{MAR}}}\right) \cdot \sqrt{m} \text{ ,}$$

where $\hat{\sigma}_{\overline{MAR}}$ is the intertemporal sample standard deviation of $\overline{MAR}$, calculated within the sample using the time series of MAR_t. At variance with equation (35), Fama (1998) advocates the method of standardizing MAR_t by its time-series standard deviation, which is normally calculated using a series of between 50 and 60 lagged values of MAR_t (e.g., Jaffe, 1974; Mandelker, 1974; Spiess and Affleck-Graves, 1999). While doing so helps address the problem of heteroscedasticity due to the varying number of firms that enter the monthly portfolios (i.e., *j* varying across months *t*), many observations from the early part of the sample period will be lost. This standardization therefore introduces an inadvertent selection bias in the sample, which will be of great concern when the sample period is relatively short, e.g., 10 years.

5. Issues surrounding tests of long-term abnormal return

The dispute over the best model of expected stock return is far from settled. A test of long-term abnormal return is essentially a joint test of market efficiency and the effect of the event. If one assumes that the adopted return generating process is correctly specified, the observation of statistically significant long-term abnormal return can be interpreted as evidence against market efficiency. With the assumption of market efficiency, alternatively, significant long-term abnormal return implies that the expected return model is mis-specified, or a bad model. Apparent from section 4 is that several attempts have been made to address various problems inherent in tests of long-term abnormal return. Naturally, different approaches give different pictures of abnormal return. It is important to understand what the adopted model(s) does and does not do, and be aware of the potential problem that comes with the chosen test method(s).

This section reviews the known methodological problems surrounding tests of long-term abnormal return. The remedies offered in the existing literature are also discussed in this section. The measurement issues are discussed in section 5.1, and statistical issues in section 5.2. Section 5.3 attempts to draw a perspective on the methodological issues as one's research objective may call for an analysis of long-term abnormal return.

5.1 Measurement issues

Central to the measurement of abnormal return is that the estimate actually represents the true abnormal return earned by investors and is unbiased. When an asset pricing model is used, e.g., as in sections 4.2 and 4.4, it is implicitly assumed that the model is a correct return generating process. As pointed out by Fama (1998, p. 292), however, "any asset pricing model is just a model and so does not completely describe expected returns". This conclusion was drawn from extensive evidence of significant long-term abnormal return following several corporate events. As Fama and French

(1993) themselves observe, their *three-factor model* is designed to price stocks on the size and book-to-market dimensions, and even so, it still misprices small high-growth (i.e., low book-to-market) firms. Moreover, Fama and French (1996) report that the return persistence or momentum documented in Jegadeesh and Titman (1993) is not explained by their three-factor model. These empirical observations on the performance of the Fama-French three-factor model may well be the reason why several subsequent studies of long-term abnormal return employ the Fama-French three-factor model plus the Carhart (1997) momentum factor (see, e.g., Moeller et al., 2004; Bouwman et al., 2009). Whether or not this four-factor model completely describes the return generating process is not obvious. To this end, it should also be noted that while the CAPM is a theoretically founded asset pricing model, the size and book-to-market as well as momentum factors are originally motivated by empirics.

The use of an asset pricing model has also been criticized for the regular rebalancing of the event firm portfolio. That is, the portfolio is mechanically rebalanced at the end of every return calculation interval. For the methods described in sections 4.2 and 4.4, for instance, the portfolio is rebalanced every month. Such a regular rebalancing strategy is likely to incur substantial transaction costs for investors, and is hence, an unlikely description of typical investors' portfolio strategy (also Loughran and Vijh, 1997). The punchline of this criticism is that using an asset pricing model as a return benchmark leads to a measure of abnormal return that is inconsistent with investors' experience. As advocated by many, especially Barber and Lyon (1997a) and Lyon et al. (1999), long-term abnormal return should be measured as *buy-and-hold* return as it does not assume such regular rebalancing and allows for compounding. Because the event window is typically very long, it is conceptually important to incorporate compounding in measuring long-term abnormal return earned on a portfolio of event firms.

Apparently, the buy-and-hold return calculation is intuitively appealing. However, the compounding itself poses a serious problem of artificial abnormal return. As pointed out by Mitchell and Stafford (2000), as long as abnormal return exists in any portion of the return series, BHAR can artificially grow even in the absence of true abnormal return, and this artificial BHAR grows in the length of the window, i.e., the holding period.[17] In a nutshell, such artificial BHAR growth is caused by the reinvesting of abnormal return in period(s) of zero abnormal return at the rate of normal return. To correct for the artificial BHAR problem, one can adopt the wealth relative measure (practically, a variant of BHAR) as in Loughran and Ritter (1995). For firm i, a wealth relative measure over a period of T months (WR_{iT}) can be calculated as:

(36) $$WR_{iT} = \frac{\prod_{t=1}^{T}[1+R_{it}]}{\prod_{t=1}^{T}[1+\mathrm{E}(R_{it})]}.$$

Because WR_{iT} is a ratio of the end of period wealth in an event-firm to the end of period expected (benchmark) wealth, this measure of long-term abnormal return takes account of compounding but does not grow in the length of the holding period.

In response to the bad model problem and regular rebalancing, several authors sidestep the use of an asset pricing model altogether, and adopt a characteristic-based return benchmark described in section 4.3. Both control-firm and control-portfolio returns are commonly employed as a proxy for expected return. However, this approach is not trouble-free. In relation to the use of a control portfolio, Barber and Lyon (1997a) identify three biases: namely; new listing bias, rebalancing bias, and skewness bias. The *new listing bias* is expected to systematically drive the abnormal

[17] Fama (1998, p. 294) illustrates a simple example of this artificial BHAR problem.

return estimate upwards. This is because the control portfolio includes not only seasoned firms, but also newly listed firms which generally underperform market averages. It is the inclusion of newly listed firms that systematically drives the control portfolio return downwards. Compounding gives rise to the *rebalancing bias* when constituent returns in the control portfolio reverse.[18,19] While the control portfolio effectively gets rebalanced periodically in order to maintain equal weights, there is no rebalancing for the sample firm.[20] With return reversal, periodic rebalancing of the control portfolio translates into the purchase (sale) of stocks that perform well (badly) in the next period, inflating the long-term return earned by the portfolio. If the control portfolio is value-weighted, however, periodic rebalancing does not lead to the rebalancing bias. The *skewness bias* arises from the dissimilarity in distribution between long-term return to a sample firm and long-term return to the control portfolio. Because returns on an individual stock can be very large whereas portfolio returns typically are relatively small, the resulting abnormal return is generally skewed. Since the risk factor(s) in an asset pricing model is essentially a spread in portfolio returns, these biases are also present when using an asset pricing model as a return benchmark. The outcome of these biases is misspecification of test statistics, i.e., rejection of the null hypothesis of zero abnormal return when it is true.

Barber and Lyon (1997a) show that the use of control-firm return eliminates the new listing and rebalancing biases, and largely alleviates the skewness bias. First, the identified control firm must be listed at the beginning of the event period of interest.

[18] Note that this rebalancing bias is not the same problem as the problem of regular rebalancing discussed above, which pertains to the event-firm or sample-firm portfolio.
[19] Barber and Lyon (1997a) note that the observed return reversals in the control portfolio are not necessarily sufficient for profit making since these reversals may well be the outcome of a bid-ask bounce.
[20] Although the control-portfolio return may be measured as buy-and-hold return, rebalancing still mechanically occurs with the use of interval returns (e.g., monthly returns) as an input.

Second, there is no portfolio rebalancing when the benchmark is a control firm. Finally, the random chances of both individual sample firms and their respective control firm experiencing large positive return are equally likely. Nevertheless, one question remains. Is it empirically possible to find a true control firm as such? As explained in section 3.4.2, firms conducting the same event are likely to share similar characteristics. To this extent, it is highly likely that the available pool of candidate control firms (or, firms to enter a control portfolio) is much smaller than what one would expect on the surface. In reality, moreover, there is no such a thing as a single firm (or, even a portfolio) that represents an *exact* match for a given event firm. This may explain why significant long-term abnormal return, as reported in many studies, still survives the characteristic-based return benchmark.

In the interest of completeness, it is worth mentioning the remedies for the new listing and rebalancing biases suggested by Lyon et al. (1999). To avoid these biases, Lyon et al. calculate long-term return to a control portfolio by first compounding over the event window the returns to individual constituent firms in the portfolio and then averaging across firms. In order to alleviate the skewness bias, Lyon et al. advocate the use of bootstrapped skewness-adjusted *t*-statistics and a nonparametric bootstrap approach based on an empirical distribution of returns in a pseudo-portfolio, which contains one randomly selected firm for each sample firm.

The measurement issues also include the controversy over how the event-firm portfolio should be weighted. For tests of long-term abnormal return, the choice of portfolio weighting has an implication further from whether the objective is to measure the typical or aggregate wealth effect of an event. On the one hand, Fama (1998) argues that value-weighting gives an appropriate measurement of abnormal return since it more accurately captures the total wealth effects experienced by investors. On the other hand, Loughran and Ritter (2000) contend that value-weighting leads to low power to detect true abnormal return, particularly when abnormal return is expected to persist among smaller firms, and therefore, advocate the use of equally weighted returns. Employing the Fama-French three-factor model, Brav and Gompers (1997) report that underperformance of IPO firms is much weaker with value-weighting than with equal weighting, and that small high-growth firms earn significantly negative abnormal return whether or not they are issuers. Indeed, Fama (1998) also points out that the significant long-term abnormal return on an equally weighted portfolio, which is found in many studies, shrinks a lot and often disappears when value-weighting is adopted. Given the existing empirical patterns, it appears that the use of equal weighting can be expected to produce significant abnormal return, especially if the sample is tilted towards small high-growth firms.

5.2 Statistical issues

At the heart of an ideal method for measuring abnormal return lie two key statistical properties. First, the method must not reject the null hypothesis of zero abnormal return when it is true. Second, the method must have sufficient power to reject the null when it is false, i.e., to detect abnormal return when it exists. As mentioned earlier, Lyon et al. (1999) propose the use of skewness-adjusted *t*-statistics and empirical distribution of pseudo-portfolio returns via bootstrapping as remedies to measurement problems. Lyon et al. report that these techniques are well-specified and powerful in detecting abnormal return. However, this result holds only in the random sample situation. In the presence of industry clustering and calendar clustering, these methods yield mis-specified test statistics. Intuitively, the non-random nature of corporate events is likely to violate the two implicit assumptions made by the bootstrapping procedure: (i) the residual variances of sample firms are identical to those of the randomly selected firms; and (ii) the observations are independent of each other (Mitchell and Stafford, 2000).

As Mitchell and Stafford (2000) point out because event firms are clearly different from random non-event firms, an empirical distribution created from characteristic-based matching does not replicate the covariance structure underlying the original event sample, and this in case, an overstatement of statistical significance could result. Indeed, Lyon et al. (1999) recognize that the bootstrapping methods do not solve the problem of cross-sectional correlations among abnormal returns. As described in section 3.4.2, it is possible to adjust for cross dependence by using Brown and Warner's (1980) CDA. However, the CDA requires the use of pre-event data, which would exacerbate the new listing bias.

As explained in section 4, forming the portfolio of event firms in calendar time *eliminates* the cross dependence problem. In fact, this is the only way to eliminate cross-sectional correlation among abnormal returns. In section 4, two alternative calendar-time approaches are described. Lyon et al. (1999) report that the calendar-time Fama-French three-factor model yields well-specified statistics in random samples although it is generally mis-specified in situations of extreme sample biases. More importantly, this approach yields well-specified statistics in the case of return overlap: the sample situation in which cross dependence is most severe. This result is observed for both equal and value-weighting. As reported by Lyon et al., the characteristic-based benchmark approach such as one described in section 4.5 is comparably well-specified, but suffers from having lower power than the Fama-French three-factor model. Nevertheless, whether or not even the three-factor model with value-weighting is sufficiently powerful remains debatable. Loughran and Ritter (2000) report that, on a value-weighted basis, the three-factor model captures only half of the true abnormal returns. Taking an alternative perspective, Mitchell and Stafford (2000) observe high R^2 values (generally in excess of 0.90) for their value-weighted Fama-French three-factor regressions, and thereby argue that the model has "considerable power" (p. 315). Mitchell and Stafford also report that the evidence of statistically significant BHAR practically disappears once the average covariance and correlation of individual BHARs are controlled for, and on this basis, argue that a calendar-time approach is more powerful than BHAR (i.e., an event-time approach) after controlling for cross-sectional correlation. On balance, empirical evidence on the relative power of the calendar-time approaches to tests of long-term abnormal return on balance remains inclusive.

5.3 Taking perspective on the methodological problems

In terms of guidance on what to do exactly when conducting tests of long-term abnormal return, the above two sections are far from encouraging. Again, it is crucial to be fully aware of what the adopted test method(s) does and does not do, and what problem is likely to come with it. As Lyon et al. (1999) emphasize in their abstract, "analysis of long-run abnormal returns is treacherous". This is true, and the testimony is provided by Masulis et al. (2007, footnote 9): ". .. . , given the serious methodological concerns that long-run stock return studies raise and the controversial nature of the evidence they produce (see Mitchell and Stafford (2000) and Andrade, Mitchell, and Stafford (2001)[21] for detailed discussions of the evidence for acquisition activity), we choose to focus on short-run stock price reactions instead."

[21] Cited in Masulis et al. 2007.

For many, unfortunately, a formal test of long-term abnormal return is called for by the orientation of their research objective. Given the problems discussed above and a lack of definite solution, robustness checks are of absolute necessity. To obtain any empirical standing, results must be strong enough to show that they stand up to both the measurement and statistical problems.

It is economically important that the abnormal return estimates to be reported are compatible with a real-world strategy for investors. To many, it is more realistic to expect investors to invest in event time and demand interest on interest than to invest in calendar time with monthly rebalancing. In terms of the choice of return benchmark, the question is whether or not one can accept the price of identifying ideal control firms or portfolios. If one is to view, and can justify, that a true control firm/portfolio exists, then an approach that utilizes a characteristic-based return benchmark is a plausible test method. Otherwise, a k-factor asset pricing model provides a more definite return benchmark. In event time, it should not be too time-consuming or costly to implement the approach in equation (26) and apply the setup of equation (23).

It is of equal importance that abnormal return estimates are statistically reliable. For instance, Fama (1998) contends that formal tests of abnormal return should be based on short-interval returns for which normality is a better approximation.[22] It is the cross dependence problem, which is a serious concern for a long-horizon test, that renders the event-time methods statistically inadequate, and based on the findings of Mitchell and Stafford (2000), powerless in detecting abnormal return. In order to establish statistical reliability of abnormal return estimates, it is therefore crucial to

[22] See Fama (1998, p. 294) for a detailed discussion on this debatable point.

employ a calendar-time approach, using either a formal asset pricing model or characteristic-based return benchmark, or both.

In sum, there is no single fool-proof approach to measuring long-term abnormal return. At the very least, one event-time method and one calendar-time method should be employed. If the two sets of results are similar and lead to the same conclusion, then it may be claimed that there is discernible evidence of event-induced long-term abnormal return. Otherwise, the results are just inconclusive. To this end, another caveat is in order. Inconclusive results may be interpreted as "no finding". Because long-horizon tests are basically plagued with methodological problems, no finding also means (to virtually all of our well experienced and meticulous colleagues) that "your methods are problematic" and/or "there is just a lot of noise in your data". Accordingly, it would inevitably follow as a conclusion that the comprehensive analysis we have just conducted tells us nothing about the event: rather than the event, on balance, empirically having no long-term value impact.

6. Take-home messages

Event studies, either short-horizon or long-horizon, all begin with accurate identification of the true event date, or the date of a surprise. In measuring the effect of an event, both the short-term windows and long-term windows are subject to measurement as well as statistical issues. However, the existing insights from simulation and empirical studies consistently indicate that the short-horizon tests appear more robust to the potential statistical problems. To properly conduct long-horizon tests requires a sizable investment. Yet, it is well documented in the literature that these tests remain "treacherous". Understandably, long-horizon tests may appear appealing to the orientation of some research questions. To this extent, it is likely to be useful for one to ponder over the following question. Is an analysis of long-term abnormal return the only meaningful way to answer the research question?

References

Agrawal, A., Jaffe, J.F., 2000. The post-merger performance puzzle. in Cooper, C., Gregory, A., eds.: Advances in Mergers and Acquisitions, Vol. 1. Elsevier Science.

Andrade, G., Mitchell, M., Stafford, E., 2001. New evidence and perspectives on mergers. Journal of Economic Perspectives 15, 103-120.

Barber, B.M., Lyon, J.D., 1997a. Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. Journal of Financial Economics 43, 341-372.

Barber, B.M., Lyon, J.D., 1997b. Firm size, book-to-market ratio, and security returns: a holdout sample of financial firms. Journal of Finance 52, 875-883.

Boehme, R.D., Sorescu, S.M., 2002. The long-run performance following dividend initiations and resumptions: Underreaction or product of chance? Journal of Finance 57, 871-900.

Bouwman, C.H.S., Fuller, K., Nain A.S., 2009. Market valuation and acquisition quality: Empirical evidence. Review of Financial Studies 22, 633-679.

Brav, A., Gompers, P.A., 1997. Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture and nonventure capital-backed companies. Journal of Finance 52, 1791-1821.

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

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

Cai, Y., Sevilir, M., 2012. Board connections and M&A transactions. Journal of Financial Economics 103, 327-349.

Carhart, M.M., 1997. On persistence in mutual fund performance. Journal of Finance 52, 57-82.

Davies, J.R., Unni, S., Draper, P., Paudyal, K., 1999. The Cost of Equity Capital. CIMA Publishing, London.

Dennis, P.J., Strickland, D., 2002. Who blinks in volatile markets, individuals or institutions? Journal of Finance 57, 1923-1949.

Di Giuli, A., Laux, P.A., 2021. The effect of media-linked directors on financing and external governance. Journal of Financial Economics, forthcoming.

Draper, P., Paudyal, K., 1999. Corporate takeovers: Mode of payment, returns and trading activity. Journal of Business Finance and Accounting 26, 521-558.

Draper, P., Paudyal, K., 2006. Acquisitions: Private versus Public. European Financial Management 12, 57-80.

Eckbo, B.E., Thorburn, K.S., 2000, Gains to bidder firms revisited: Domestic and foreign acquisitions in Canada. Journal of Financial and Quantitative Analysis 35, 1-25.

Ekkayokkaya, M., Holmes, P., Paudyal, K., 2009a. The Euro and the changing face of European banking: Evidence from mergers and acquisitions. European Financial Management 15, 451-476.

Ekkayokkaya, M., Holmes, P., Paudyal, K., 2009b. Limited information and the sustainability of unlisted target acquirers' returns. Journal of Business Finance and Accounting 36, 1201-1227.

Ekkayokkaya, M., Paudyal, K., 2015. A trade-off in corporate diversification. Journal of Empirical Finance 34, 275-292.

Faccio, M., McConnell, J.J., Stolin, D., 2006. Returns to acquirers of listed and unlisted targets. Journal of Financial and Quantitative Analysis 41, 197-220.

Fama, E.F., 1998. Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics 49, 283-306.

Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. Journal of Finance 47, 427-465.

Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Fama, E.F., French, K.R., 1996. Multifactor Explanations of Asset Pricing Anomalies. Journal of Finance 51, 55-84.

Fama, E.F., Fisher, L., Jensen, M.C., 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. Journal of Finance 57, 1763-1793.

Gregory, A., Tharyan, R., Christidis, A., 2013. Constructing and testing alternative versions of the Fama-French and Carhart models in the UK. Journal of Business Finance and Accounting 40, 172-214.

Grinstein, Y., Hribar, P., 2004. CEO compensation and incentives: Evidence from M&A bonuses. Journal of Financial Economics 73, 119-143.

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

Harford, J., Humphery-Jenner, M., Powell, R., 2012. The sources of value destruction in acquisitions by entrenched managers. Journal of Financial Economics 106, 247-261.

Hollander, M., Wolfe, D.A., 1999. Nonparametric Statistical Methods. Wiley, New York.

Jaffe, J.F., 1974. Special information and insider trading. Journal of Business 47, 410-428.

Jegadeesh, N., Titman S., 1993. returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65-91.

Jensen, M.C., Ruback, R.S., 1983. The market for corporate control: The scientific evidence. Journal of Financial Economics 11, 5-50.

Jenter, D., Lewellen, K., 2015. CEO preferences and acquisitions. Journal of Finance 70, 2813-2851.

Larcker, D.F., Ormazabal, G., Taylor, D.J., 2011. The market reaction to corporate governance regulation. Journal of Financial Economics 101, 431-448.

Lipson, M., Puckett, A., 2010. Institutional trading during extreme market movements. Working paper, University of Virginia.

Loughran, T., Vijh, A.M., 1997. Do long-term shareholders benefit from corporate acquisitions? Journal of Finance 52, 1765-1790.

Loughran, T., Ritter, J., 1995. The new issues puzzle. Journal of Finance 50, 23-51.

Loughran, T., Ritter, J.R., 2000. Uniformly least powerful tests of market efficiency. Journal of Financial Economics 55, 361-389.

Lyon, J.D., Barber, B.M., Tsai, C.L., 1999. Improved methods for tests of long-run abnormal stock returns. Journal of Finance 54, 165-201.

Malatesta, P.H., 1983. The wealth effect of merger activity and the objective functions of merging firms. Journal of Financial Economics 11, 155-181.

Mandelker, G., 1974. Risk and return: The case of merging firms. Journal of Financial Economics 1, 303-335.

Masulis, R.W., Wang, C., Xie, F., 2007. Corporate governance and acquirer returns. Journal of Finance 62, 1851-1889.

Masulis, R.W., Zhang, E.J., 2019. How valuable are independent directors? Evidence from external distractions. Journal of Financial Economics 132, 226-256.

Mitchell, M.L., Mulherin, J.H.,1996. The impact of industry shocks on take-over and restructuring activity. Journal of Financial Economics 41, 193-229.

Mitchell, M.L., Stafford, E., 2000. Managerial decisions and long-term stock price performance. Journal of Business 73, 287-329.

Moeller, S.B., Schlingemann, F.P., Stulz, R.M., 2004. Firm size and gains from acquisitions. Journal of Financial Economics 73, 201-228.

Rau, P.R., Stouraitis, A., 2011. Patterns in the timing of corporate event waves. Journal of Financial and Quantitative Analysis 46, 209-246.

Rau, P.R., Vermaelen, T., 1998. Glamour, value and the post-acquisition performance of acquiring firms. Journal of Financial Economics 49, 223-253.

Schmidt, B., 2015. Costs and benefits of friendly boards during mergers and acquisitions. Journal of Financial Economics 117, 424-447.

Shleifer, A., Vishny, R.W., 2003. Stock market driven acquisitions. Journal of Financial Economics 70, 295-311.

Spiess, D.K., Affleck-Graves, J., 1999. The long-run performance of stock returns following debt offerings. Journal of Financial Economics 54, 45-73.

Wang, C., Xie, F., 2009. Corporate governance transfer and synergistic gains from mergers and acquisitions. Review of Financial Studies 22, 829-858.