Vaasan yliopisto
UNIVERSITY OF VAASA

ANUPAM DUTTA

Essays on Testing Long-Run Abnormal Stock Returns

ACTA WASAENSIA 324

MATHEMATICS AND STATISTICS 11

Reviewers

Professor James W. Kolari
JP Morgan Chase Professor of Finance
Director, Commercial Banking Program
Department of Finance
Texas A&M University
College Station, Texas 77843-4218
USA

Professor Johan Knif
Department of Finance and Statistics
Hanken School of Economics
P.O. Box 287
FI-65101 Vaasa
Finland

Julkaisija Vaasan yliopisto	Julkaisupäivämäärä Toukokuu 2015	
Tekijä(t) Anupam Dutta	**Julkaisun tyyppi** Collection of Essays **Julkaisusarjan nimi, osan numero** Acta Wasaensia, 324	
Yhteystiedot Vaasan Yliopisto Teknillinen tiedekunta Matemaattisten tieteiden yksikkö PL 700 64101 Vaasa	**ISBN** 978-952-476-613-5 (print) 978-952-476-614-2 (online) **ISSN** 0355-2667 (Acta Wasaensia 324, print) 2323-9123 (Acta Wasaensia 324, online) 1235-7928 (Acta Wasaensia. Mathematics 11, print) 2342-9607 (Acta Wasaensia. Mathematics 11, online)	
	Sivumäärä 99	**Kieli** Englanti
Julkaisun nimike Essays on Testing Long-Run Abnormal Stock Returns		

Tiivistelmä

Osakemarkkinoiden pitkän aikavälin (1–5 vuotta) event-tutkimuksen menetelmällisessä kehityksessä on tapahtunut vuosien kuluessa merkittävää edistystä. Kuitenkaan yleisesti hyväksytystä lähestymistavasta arvioida osakkeen hinnan mahdollista epänormaalia käyttäytymistä merkittävän yritystapahtuman (event) jälkeen ei ole yksimielisyyttä. Toistaiseksi kaksi käytetyintä tapaa ovat osta-ja-pidä sijoitusstrategiaan perustuva (buy-and-hold abnormal return, BHAR) lähestymistapa ja kalenteriaikaan perustuvan portfolion (calendar time protfolio, CTP) lähestymistapa. Molempia näitä kritisoidaan puutteidensa vuoksi kirjallisuudessa. Niinpä tutkimus on edelleen vilkasta tällä alalla.

Väitöskirjassa esitetään menetelmä, joka parantaa kalenteriakaan perustuvassa portfoliomenetelmässä olevia puutteita. Ehdotettu lähestymistapa, joka tunnetaan myös nimellä standardoitu kalenteriaika (standardized calendar time approach, SCTA), nojautuu kahteen osatekijään: event-yrityksen standardoitu tuotto ja painotettu kalenterikuukauden portfolio. Standardointi vähentää volatiilisten osakkeiden painoarvoa ja kuukausittainen painotus korostaa ajanjaksoja joissa on useita tapahtumia (events).

Ehdotetun lähestymistavan robustisuuden tutkimiseksi tuloksia verrataan traditionaalisiin BHAR ja CTP menetelmien tuloksiin. Tutkimuksessa hyödynnetään Yhdysvaltojen, Englannin ja merkittävimpien Aasian markkinoiden osaketuottoja. Näiden markkinoiden todellisiin tuottoihin perustuvat simulointikokeet osoittavat, että tutkimuksen SCTA menetelmä on tilastollisessa mielessä paremmin spesifiotu ja voimakkaampi kuin edellä mainitut BHAR ja CTP menetelmät. Samalla tutkimus osoittaa myös, että ehdotettu menetelmä toimii hyvin Yhdysvaltojen markkina-aineiston lisäksi myös Aasian ja Englannin markkinoihin perustuvilla aineistoilla.

Asiasanat

Pitkän aikavälin event-tutkimus, standardoitu epänormaali tuotto, yritystapahtumat, testin spesifiointi, testin voimakkuus

Publisher	**Date of publication**	
Vaasan yliopisto	May 2015	
Author(s) Anupam Dutta	**Type of publication** Collection of Essays	
	Name and number of series Acta Wasaensia, 324	
Contact information University of Vaasa Department of Mathematics and Statistics P.O. Box 700 FI-65101 Vaasa, Finland	**ISBN** 978-952-476-613-5 (print) 978-952-476-614-2 (online)	
	ISSN 0355-2667 (Acta Wasaensia 324, print) 2323-9123 (Acta Wasaensia 324, online) 1235-7928 (Acta Wasaensia. Mathematics 11, print) 2342-9607 (Acta Wasaensia. Mathematics 11, online)	
	Number of pages 99	**Language** English

Title of publication Essays on Testing Long-Run Abnormal Stock Returns

Abstract

Although long-run event studies have seen many advances over the years, the proper methodology for measuring the stock price performance of firms for periods of one to five years following certain corporate events is much debated in the literature. While a large number of recent studies consider applying the buy-and-hold abnormal return (BHAR) approach and the calendar time portfolio (CTP) method for investigating long-term anomalies, each of the methods is a subject to criticisms. A fundamental choice for many recent studies, therefore, concerns the measure of long-run stock price performance.

The purpose of the present dissertation is to propose a refined calendar time approach to moderate such pitfalls to some extent. The proposed calendar time portfolio approach, also known as standardized calendar time approach (SCTA), consists of two major components: standardization of event firms' abnormal returns and weighting the monthly portfolios. While standardizing diminishes the effect of event firms having volatile future returns, weighting allows monthly portfolios containing more event firms to receive more weight.

In order to investigate the robustness of the proposed approach, the results from BHAR methodology and other traditional CTP methods are also reported. The study utilizes the U.S., the U.K. and the leading Asia-Pacific security market data. Simulations show that SCTA documents better specification and power than the conventional long-run event study methodologies. However, the findings further conclude that in addition to the U.S. stock market, the event study methodologies considered perform well in Asia-Pacific and the U.K. security markets as well.

Keywords

Long-run Event Studies, Standardized abnormal returns, Corporate Events, Test specification, Power of test.

Acknowledgement

Foremost, I wish to express my deepest gratitude to my advisor, Dr. Seppo Pynnönen, for his outstanding assistance, caring and endurance. The most important part of his guidance is that he is always there to help me throughout the course of this PhD project. In our first meeting, which took place in August, 2012, Professor Pynnönen gave me the research Idea and this was really essential for the on time completion of my doctoral project. He had also gone through all the papers carefully before I submitted the dissertation for pre-examination. His thoughtful comments and suggestions are cordially acknowledged.

I would like to thank Dr. Seppo Hassi and Dr. Tommi Sottinen for providing me with an excellent atmosphere for doing research. As the head of the department, both of them always gave me the support and inspiration to further my research career. My research would not have been possible without their generous helps. I am using this opportunity to thank Dr. Bernd Pape who constantly keeps trying to help me and give his best suggestions. It would have been a lonely period without his kind presence. I would like to thank all other people in the department of Mathematics and Statistics for their support and guidance during the last 3 years. Special thanks go to Jaakko Tyynelä who always provides me with the required data to carry out my research project.

I express my warm thanks to the pre-examiners of this dissertation, Professor James Kolari from Texas A&M University and Professor Johan Knif from Hanken School of Economics. Their insightful and comprehensive comments certainly helped me a lot to improve my thesis. Their supervisions and recommendations are highly appreciated.

I would also like to thank my parents and my elder brother who are continuously supporting me and encouraging me with their best wishes. Today I truly feel happy that I have become successful to fulfill my mother’s sweetest dream. Finally, my heartfelt thanks go to my wife, Aditi Acharyya, who is always there for cheering me up and supporting me in my good as well as bad times.

Vaasa, May 2015

Anupam Dutta

Contents

1 INTRODUCTION ... 1
1.1 Event study methodology ... 2
1. 2 Literature Review ... 6
1.3 Standardized Calendar Time Approach (SCTA) ... 9

2 SUMMARY OF THE ESSAYS ... 11
2.1 Parametric and Nonparametric Event Study Tests: A Review ... 11
2.2 Improved Calendar Time Approach for Measuring Long-Run Anomalies ... 11
2.3 Does Calendar Time Portfolio Approach Really Lack Power? ... 12
2.4 Investigating Long-Run Stock Returns after Corporate Events: the UK Evidence ... 12
2.5 Conducting Long-Run Event Studies in Asia-Pacific Security Markets ... 13

REFERENCES ... 14

This PhD dissertation consists of the following five essays and the introductory chapter:

1. Dutta, A. (2014). Parametric and Nonparametric Event Study Tests: A Review, *International Business Research*, 7 (12): pp. 136-142.

2. Dutta, A. (2015). Improved Calendar Time Approach for Measuring Long-Run Anomalies. Under review in *Cogent Economics and Finance*.

3. Dutta, A. (2014). Does Calendar Time Portfolio Approach Really Lack Power?. *International Journal of Business and Management*, 9 (9): pp. 260-266.

4. Dutta, A. (2014). Investigating Long-Run Stock Returns after Corporate Events: the UK Evidence. *Corporate Ownership and Control*, 12 (1): pp. 298-307.

5. Dutta, A. and Pynnönen, S. (2015). Conducting Long-Run Event Studies in Asia-Pacific Security Markets. Under review in *Australian Economic Papers*.

1 INTRODUCTION

Since the seminal paper of Fama, Fisher, Jensen and Roll (1969) in the late 1960s, the event study methodology has become an important tool of testing market efficiency. Such methodology is employed for the purpose of analyzing the stock market responses to certain corporate events such as mergers and acquisitions, IPOs, stock split etc. That is, event studies are empirical procedures for investigating the effect of an event on stock returns. However, typical events are of two types: Firm-specific events and Economy-wide events. Firm-specific events usually indicate a change in the company policy. Examples of such events include earnings, investment, mergers and acquisitions, issues of new debt or equity, stock splits, etc. announcements. Economy-wide events, on the other hand, are used to assess the impact of a particular event on relevant securities. This type of events includes inflation, interest rate, consumer confidence, trade deficient, etc. announcements.

In event studies, the data to be analyzed can be daily, weekly, monthly, or annually. While the earlier studies in financial economics such as Brown and Warner (1980, 1985), Corrado (1989), Campbell and Wasley (1993), Kolari and Pynnönen (2011) etc. focus on the characteristics of abnormal returns measured on a particular day or, at the most cumulated over several months, a large number of recent studies investigate the stock price performance of firms for periods of one to five years following significant corporate events. The extensive literature of long-horizon event studies includes Barber and Lyon (1997), Kothari and Warner (1997), Fama (1998), Lyon, Barber, and Tsai (1999), Mitchell and Stafford (2000), Boehme and Sorescu (2002) and so on.

Although long-run event studies have a long history, serious limitations still exist. Kothari and Warner (1997), for example, document that while short-horizon methods are quite reliable, inferences from long-horizon tests require extreme caution. Lyon et al. (1999) also conclude that the analysis of long-run abnormal performance is treacherous. Short-run event studies, on the other hand, are relatively stable and free of limitations. For instance, Fama (1991) report that short-horizon tests represent the cleanest evidence we have on efficiency, but the interpretation of long-horizon results is problematic. Further filtering of the existing long-run methodologies (e.g., the buy-and-hold abnormal return methodology and the calendar time portfolio approach) is thus required.

The objective of this dissertation is to propose a refined calendar time approach to mitigate such limitations to some extent. Our proposed calendar time portfolio (CTP) approach has two major components: standardization of event firms' abnormal returns and weighting the monthly portfolios. While standardizing diminishes the impact of event firms having volatile future returns, weighting allows monthly portfolios containing more event firms to receive more weight. Simulations reveal that these two innovations document better specification and power than the conventional long-run event study methodologies.

The rest of this Introductory chapter is structured as follows. The next section outlines the event study methodology. The existing literature of event studies is then reviewed. Our proposed approach has been discussed in the last section.

1.1 Event study methodology

This section discusses the methodological issues of event study. Campbell, Lo and MacKinlay (1997) outline the following steps for a typical event study.

Event Definition and Event Window

The initial task of conducting an event study is to define the event of interest (e.g., the announcement of quarterly earnings for a firm) and identify the period over which the prices of the relevant financial instruments will be examined. This period is called the event window. The choice of event window is somewhat arbitrary and there does not appear to be any sound empirical basis for choosing a particular time period around an event. It is a matter for judgment for the researcher.

Selection Criteria

The next step is to determine the selection criteria for the firms to be included in the study. Suggested approaches are to look at firms only on major exchanges with frequent trading. Also, there may be a need to exclude firms with more than one event over the periods of the event window. This is necessary if one cannot determine which event is driving the returns of the stock.

Normal and Abnormal Returns

In order to assess the event's impact, a measure of abnormal returns is required. The normal return is the return that would be expected if the event did not take place. For each firm *i*, the abnormal return at at time *i* is calculated as follows:

$$AR_{it} = R_{it} - E(R_{it})$$

where AR_{it}, R_{it} and $E(R_{it})$ indicate abnormal, actual and normal returns respectively. In the following sections, we discuss different methodologies to determine the normal or expected return.

Measuring Normal Performance

Models and methods used for measuring normal performance are as follows:

(i) Constant Mean-Return Model

The mean return model assumes that the mean of the stock's return over the event window is expected to be the same as the mean over the estimation period. The abnormal return using this model is

$$AR_{it} = R_{it} - \mu_i,$$

where R_{it} denotes the return of stock i at time t and μ_i is the mean return of stock i over the period.

Although this is the simplest model for measuring normal returns, it becomes problematic if the firms in the sample of event firms cluster in time. Another problem related to the mean-return model is that it does not respond well when the market trends up or down. In such cases, the estimates will also trend up or down, but those conditions may not exist during the event window. This model, however, also does not respond well when certain industries experience uncertainty and significant variation in returns.

(ii) Market Model

The market model represents a potential improvement over the constant-mean-return model. By removing the portion of the return that is related to variation in the market's return, the variance of the abnormal return is reduced. The abnormal return using the market model is

$$AR_{it} = R_{it} - \alpha_i - \beta_i R_{mt},$$

where R_{mt} is the market return at time t and α and β are the market model parameters.

Problems occur with the market model when the event dates for the firms in the sample occur around the same period (clustering problem). Otherwise, this method is as efficient as more advanced methods are.

(iii) Capital Asset Pricing Model (CAPM)

The abnormal return using the Capital Asset Pricing Model (CAPM) is

$$AR_{it} = R_{it} - R_{ft} - \beta_i(R_{mt} - R_{ft}),$$

where R_{ft} is the risk free rate and β_i is the slope parameter of the CAPM model.

(iv) Fama-French Three-Factor Model

Using the three-factor model, proposed by Fama and French (1993), the abnormal return is

$$AR_{it} = R_{it} - R_{ft} - \beta_{i1}(R_{mt} - R_{ft}) - \beta_{i2}SMB_t - \beta_{i3}HML_t,$$

where R_{ft} is the risk-free rate, $R_{mt} - R_{ft}$ is the excess return of the market, *SMB* is the difference between the return on the portfolio of small stocks and big stocks, *HML* is the difference between the return on the portfolio of high and low book-to-market stocks, and the β's are the slope parameters.

(v) Carhart Four-Factor Model

Carhart (1997) extends the Fama-French three-factor model to include the momentum factor. The abnormal return using this model is

$$AR_{it} = R_{it} - R_{ft} - \beta_{i1}(R_{mt} - R_{ft}) - \beta_{i2}SMB_t - \beta_{i3}HML_t - \beta_{i4}UMD_t,$$

where *UMD* is the difference between returns of winners and losers. However, this model is not able to explain the anomalies of small firms.

(vi) Reference Portfolio Method

Lyon et al. (1999) report that the calendar-time portfolio methods based on reference-portfolio abnormal returns generally dominate those based on asset pricing models (e.g., Fama-French three-factor model) for two reasons. First, the three-factor model implicitly assumes linearity in the constructed market, size, and book-to-market factors. But Lyon et al. find that this assumption is unlikely to be the case for the size and book-to-market factors. Second, while the Fama- French three-factor model assumes there is no interaction between the three factors, Lyon et al. document that this assumption is also likely violated because the relation between book-to-market ratio and returns is most pronounced for small firms. Later, Loughran and Ritter (2000) also argue that the three-factor model is not an equilibrium model since it only detects anomalies in financial markets and fails to test market efficiency. Barber and Lyon (1997) and Lyon et al. (1999), therefore, employ characteristics-based reference portfolios to measure the abnormal performance. These studies construct reference portfolios on the basis of market value and book-to-market ratio. However, although the use of reference portfolios alleviates the problem of new listing and re-balancing biases, the skewness bias still remains.

(vii) Control Firm Approach

In this approach, sample firms are matched to a control firm on the basis of specified firm characteristics such as market value, book-to-market ratio etc. Barber and Lyon (1997) and Lyon et al. (1999) prefer control firm approach to reference portfolio approach as the former mitigates the new listing, re-balancing and skewness biases. The new listing bias is eliminated as both the sample and control firms are listed in the identified month. The re-balancing bias is also eliminated since both sample and control firm returns are computed without re-balancing. Finally, employing the control firm approach alleviates the skewness problem since the sample and control firms are equally likely to experience large positive returns. However, Lyon et al. (1999) report that standard tests based on the control firm approach are not as powerful as those based on the reference portfolio approach.

Testing Procedure

Two commonly used approaches for testing the null hypothesis of no abnormal performance are parametric tests and nonparametric tests. While parametric tests require a specific distributional assumption, nonparametric tests refer to as distribution-free tests. Although a number of event studies rely on parametric test statistics, Brown and Warner (1985) report that stock prices are not normally distributed. Consequently, when this assumption of normality is violated, parametric tests are not well-specified. Non-parametric tests, on the other hand, are well-specified and more powerful at detecting a false null hypothesis of no abnormal returns. The most successful among these tests are the nonparametric sign and rank tests advanced in Corrado (1989), Zivney and Thompson (1989), and Corrado and Zivney (1992). Well-known studies of this type are Cowan (1992), Campbell and Wasley (1993), and Corrado and Truong (2008). Each of these studies reports that sign and rank tests provide better specification and power than parametric tests. Kolari and Pynnönen (2011) recently develop a generalized rank test which is robust and documents superior empirical power relative to popular parametric tests. Detailed discussions on these event study tests can be found in the 1st article of the current dissertation.

Empirical Results

The presentation of the empirical results follows the formulation of the econometrical design. In addition to presenting the basic empirical results, the presentation of diagnostics can be fruitful. Occasionally, especially in studies with a limited number of event observations, the empirical results can be heavily influenced by one or two firms. Knowledge of this is crucial for gauging the importance of the results.

Interpretation and Conclusions

Ideally the empirical results will lead to insights about the mechanisms by which the event affects security prices. Additional analysis may be included to distinguish between competing explanations.

1. 2 Literature Review

While short-run event study methods are relatively straightforward and reliable (Fama, 1991) the proper methodology for measuring long-run abnormal stock returns is still much debated in the literature. Financial economists are always in search of the appropriate measure of long-run abnormal stock returns and the appropriate statistical methodology for testing the significance of any measured abnormal performance. Kothari and Warner (2007), for instance, argue that the question of which model of expected returns is correct remains an unresolved issue. Fama (1998) also concludes that not a single model for expected returns can present a complete description of the systematic patterns in average returns.

However, beginning with Ritter (1991), the most popular estimator of long-run abnormal performance is the mean buy-and-hold abnormal return (BHAR). Mitchell and Stafford (2000) define BHARs as the average multiyear return from a strategy of investing in all firms that complete an event and selling at the end of a prespecified holding period versus a comparable strategy using otherwise similar nonevent firms. An appealing feature of using BHAR is that buy-and-hold returns better resemble investors actual investment experience than periodic (monthly) re-balancing entailed in other approaches to measuring risk-adjusted performance.

Fama (1998), however, argues against the BHAR methodology because of the statistical problems associated with the use of the BHAR and the associated test statistics. In addition, any methodology ignoring the cross-sectional dependence of event-firm abnormal returns that do overlap in calendar time is likely to produce overstated test statistics. Eckbo et al. (2000) also argue against the application of buy-and-hold abnormal return method. They document that the BHAR methodology is not a feasible portfolio strategy because the total number of stocks is not known in advance. Later, Jegadeesh and Karceski (2009) criticize the BHAR approach arguing that it assumes the cross-sectional independence of abnormal returns, while such assumption is violated in nonrandom samples, where the event firm returns are positively correlated.

Barber and Lyon (1997) and Lyon et al. (1999) identify new listing, re-balancing, and skewness biases with inference in long-run event studies using the BHAR. They use simulations to investigate the impact of these biases on inference when BHAR is exercised to measure the abnormal performance and standard tests are applied. However, in case of using a reference portfolio to capture expected return, the new listing and rebalancing biases can be addressed in a relatively simple way by careful construction of the reference portfolio [see Lyon et al. (1999)].

Unfortunately, the use of a reference portfolio to capture the expected return gives rise to the skewness bias. This bias arises due to the fact that the long-run return of a portfolio is compared with the long-run return of an individual asset. The long-run return of an individual security is highly skewed; whereas the long-run return for a reference portfolio (due to diversification) is not. Consequently, the BHAR, the difference between these returns, is also skewed. Barber and Lyon (1997) report that since BHAR is positively skewed, its use causes the standard tests to have the wrong size and causes the power of the test to be asymmetric; rejection rates are far higher when induced abnormal returns are negative than when they are positive.

To avoid the skewness bias, a control firm rather than a reference portfolio can be used as the long-run return benchmark. BHAR is then measured as the difference between the long-run holding-period returns of the event firm's equity and that of a control firm. Although the distribution of each asset's holding-period return is highly skewed, the distribution of their difference is not. As a result, standard statistical tests based on the control firm approach have the right size in random samples.

However, standard tests based on the control firm approach are not as powerful as those based on the reference portfolio approach. Lyon et al. (1999), for instance, argue that the use of a control firm is a noisier way to control for expected returns than is the use of a reference portfolio and this added noise reduces the power of the test. The variance of the difference between the returns on two individual assets is generally much higher than the variance of the difference between the return of an asset and that of a portfolio, even when the control firm is chosen carefully. Powerful tests thus require very large samples when control firm approach is applied.

To deal with the power and specification issues, Lyon et al. (1999) discuss two modes to modify the reference portfolio approach for fixing the associated size problem. The first of these two ways refers to the use of p-values generated from the empirical distribution of long-run abnormal returns, while the other suggests the use of skewness-adjusted *t*-statistics. Such methods, combined with careful construction of reference portfolios to remove the rebalancing and new listing biases, solve the size problem in random samples. However, Lyon, Barber, and Tsai observe that these corrections do not produce well-specified tests in many of the non-random samples considered in their study. In non-random samples the use of a standard reference portfolio approach often fails to match the expected return of the event firm with the expected return of the reference portfolio resulting in a misspecified test. Furthermore, when the return on a diversified portfolio is employed to capture expected returns, there is no offset of any contemporaneous correlation of idiosyncratic returns that may exist across firms. This problem is likely to be heightened when the events get highly clustered in time. Fama (1998) strongly recommends the use of CTP methodology on the grounds that monthly returns are less susceptible to the bad model problem as they are less skewed and by forming monthly calendar time portfolios, all cross-correlations of event-firm abnormal returns are automatically accounted for in the portfolio variance. Fama also documents that the distribution of this estimator is better approximated by the normal distribution, allowing for classical statistical inference. Mitchell and Stafford (2000), like Fama (1998), also prefer the CTP approach to BHAR methodology as the latter assumes independence of multi-year event firm abnormal returns.

While many recent studies strongly advocate the CTP approach, it has a number of potential pitfalls. Loughran and Ritter (2000), for example, criticize the use of calendar time approach arguing that it gives equal weight to each month, regardless of whether the month has heavy or light event activities. They conclude that the calendar time portfolio regressions have low power to identify the abnormal performance because it averages over months of ‘hot’ and ‘cold’ event activity. Lyon et al. (1999), however, claim that the CTP approach is misspecified in non-random samples, while the BHAR approach is relatively robust.

The bottom line is that despite these positive developments in long-run event study methodology, the power and specification issues still remain unsolved and further refinement of the existing methods is required for solving these issues. Kothari and Warner (2007), for instance, conclude that whether calendar time, BHAR methods or some combination can best address long-horizon issues remains an open question.

In this study, we propose to refine the traditional calendar time portfolio approach in order to deal with the ongoing debates discussed in prior literature. To serve this purpose, a variant of calendar time method is proposed where we first standardize the abnormal returns for each of the event firms in the sample and then construct the monthly portfolios. However, we also propose to weight the monthly portfolios such that periods of heavy event activity receive more weight than periods of low event activity. In addition to the U.S. stock market, we also analyze the data from the UK stock market and a number major Asia-Pacific security markets. Simulations show that our proposed approach is robust in each of the security markets considered.

1.3 Standardized Calendar Time Approach (SCTA)

While analyzing the stock returns after certain corporate events, a number of firms in the sample often produce volatile returns. Because of this volatility, the distributions of stock returns tend to have fat tails. But one possible solution to this problem is standardizing the abnormal returns by their volatility measures. This helps improve the testing power. Previous empirical studies, for example, Patell (1976) and Kolari and Pynnönen (2010) also argue that short-run tests using standardized returns document better power than those based on unstandardized returns. However, employing standardized abnormal returns is well-documented in long-run event studies as well. For example, Jaffe (1974) and Mandelker (1974) use standardized portfolio returns for assessing the long-run abnormal performance. Later, Mitchell and Stafford (2000) use standardized abnormal returns to alleviate the heteroskedasticity problem that often occurs in CTP approach due to the varying portfolio construction.

In our proposed standardized calendar time methodology, we use standardized abnormal returns to compute the calendar time abnormal returns (CTARs). Under the standardization approach, if an event firm has very volatile future returns, its impact on the overall portfolio return series is diminished. This improves the power of the test. The whole procedure of standardization is done in two steps. The first step involves the calculation of standardized abnormal returns for each of the sample firms. In doing so, the abnormal returns for firm *i* are computed as $\varepsilon_{it} = r_{it} - E(r_{it}); t = 1, \ldots, H$, where r_{it} denotes the log return on event firm *i* in the calendar month t and $E(r_{it})$ is the expected return which is proxied by size/book-to-market reference portfolios and size/book-to-market matched control firm and H is the holding period which equals 12, 36 or 60 months. The next task is to estimate the event-portfolio residual variances using the H-month residuals computed as monthly differences of i-th event firm returns and control firm returns. Dividing ε_{it} by the estimate of its standard deviation yields the corresponding standardized abnormal return, say, z_{it}, for event firm i in month t. Now let N_t refer to the number of event firms in the calendar month t. We then calculate the calendar time abnormal return for portfolio t as:

$$CTAR_t = \sum_{i=1}^{N_t} x_{it} z_{it}, \qquad (1)$$

where the weight x_{it} equals $\frac{1}{N_t}$ when the abnormal returns are equally-weighted and $\frac{MV_{it}}{\sum MV_{it}}$ when the abnormal returns are value-weighted by size.

We propose to weight each of the monthly CTARs by $1/\sqrt{\left(\sum_{i=1}^{N_t} x_{it}^2\right)}$. For instance, when the abnormal returns are equally weighted i.e., when $x_{it} = \frac{1}{N_t}$, then $1/\sqrt{\left(\sum_{i=1}^{N_t} x_{it}^2\right)} = \sqrt{N_t}$. This weighting scheme is lucrative as it gives more loadings to periods of heavy event activity than the periods of low event activity. Now the grand mean monthly abnormal return, denoted by $\overline{CTAR}$, is calculated as:

$$\overline{CTAR} = \frac{1}{T} \sum_{t=1}^{T} CTAR_t \qquad (2)$$

While finding $\overline{CTAR}$, it might be the case that a number of portfolios do not contain any event firm. In such situations, those months are dropped from the analysis. To test the null hypothesis of no abnormal performance, the t-statistic of $\overline{CTAR}$ is computed by using the intertemporal standard deviation of the monthly CTARs defined in equation (1).

2 SUMMARY OF THE ESSAYS

2.1 Parametric and Nonparametric Event Study Tests: A Review

The objective of this paper is to review the existing methodologies for measuring short-run abnormal performance of firms following certain corporate events. In doing so, the study outlines standard parametric tests, Generalized Sign Test, Wilcoxon Signed-Rank Test and Corrado's Rank Test (1989) in details. Recent developments in non-parametric event study tests are also discussed. For example, Kolari and Pynnönen (2011) recently introduce a generalized rank test based on generalized standardized abnormal returns which is used to test both single abnormal returns and cumulative abnormal returns. Reviewing the prior literature reveals that the nonparametric sign and rank tests are better specified than parametric procedures. However, in case of detecting the short-run anomalies, we document that nonparametric approaches have superior power relative to standard parametric tests.

2.2 Improved Calendar Time Approach for Measuring Long-Run Anomalies

The proper methodology for analyzing the long-term return anomalies has been much debated in the literature. Although a large number of event studies have employed the BHAR methodology and the calendar time portfolio approach for investigating the long-run abnormal performance, each method has potential pitfalls. For example, Mitchell and Stafford (2000) argue against using the BHAR methodology as it assumes event-firm abnormal returns to be independent. They, like Fama (1998), strongly advocate the use of CTP approach. Loughran and Ritter (2000), however, report that the calendar time portfolio approach weights each month equally so that months that reflect heavy event activity are treated the same as months with low activity. Thus the CTP approach may fail to detect significant abnormal returns if abnormal performance primarily exists in months of heavy event activity. Lyon et al. (1999), however, claim that the CTP approach is misspecified in nonrandom samples, while the BHAR approach is relatively robust. This paper proposes a modified calendar time portfolio approach which has two major components: standardization of event firms' abnormal returns and weighting the monthly portfolios. While standardizing diminishes the impact of event firms having volatile future returns, weighting allows monthly portfolios containing more event firms to receive more weight. The empirical analysis shows that these two innovations improve the size and power properties of statistical tests used in long-run event studies.

2.3 Does Calendar Time Portfolio Approach Really Lack Power?

Although long-run event study methodologies have seen many advances over the years, very few studies focus on the power issue. In order to extend the limited literature, the present study aims to compare the power of alternative methodologies. To be more specific, this paper investigates whether the calendar time methodology lacks power in detecting the long-run abnormal performance. In addition, the study uses a modified calendar time approach by forming the monthly portfolios in a variant way. To assess the robustness of this refined method, the results from buy-and-hold abnormal return approach and the mean monthly calendar time abnormal return methodology are also documented. Simulations show that the modified calendar time approach improves the power in random samples and in samples with calendar clustering.

2.4 Investigating Long-Run Stock Returns after Corporate Events: the UK Evidence

This paper investigates the robustness of existing long-run event study methodologies using the UK security market data. In doing so, the study employs the buy-and-hold abnormal return approach and the calendar time portfolio method to identify the long-term abnormal performance following corporate events. Although many recent studies consider the application of these two widely used approaches, each of the methods is a subject to criticisms. This paper uses the standardized calendar time approach (SCTA) of Dutta (2014a) which presents a number of significant improvements over the traditional calendar time methodology. The simulated results reveal that all the traditional methodologies perform well in the UK stock market. Our findings further report that SCTA documents better specification and power than the conventional approaches.

2.5 Conducting Long-Run Event Studies in Asia-Pacific Security Markets

The main purpose of this study is to check the robustness of existing long-run event study methodologies in the leading Asia-Pacific stock markets. To serve this purpose, the study employs the buy-and-hold abnormal return approach and the mean monthly calendar time abnormal return method to measure the return anomalies. However, we also consider the application of standardized calendar time approach (SCTA) of Dutta (2014b) as an alternative methodology. To measure the abnormal performance of the sample firms, both control firm approach and reference portfolio approach have been adopted. Our empirical analysis indicates that the traditional methods are found to be effective in leading Asia-Pacific security markets. We further document that test statistics based on SCTA are generally well-specified in all types of nonrandom samples considered. The BHAR approach, on the other hand, yields reasonably well-specified test statistics only when the control firm approach is employed. In addition, simulations show that the mean monthly calendar time abnormal return methodology performs well when the abnormal returns are calculated using the control firm approach. We, therefore, advocate the use of control firm method for measuring the long-run abnormal performance of event firms. However, in case of detecting the abnormal performance, SCTA documents higher power than other empirical procedures used in this study.

REFERENCES

Barber, B. and J. Lyon (1997). Detecting Long-run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics, Journal of Financial Economics, 43, 341-372.

Boehme, D. R. and S. M. Sorescu (2002). The Long-Run Performance Following Dividend Initiations and Resumptions: Underreaction orProduct of Chance?, Journal of Finance, 57, 871-900.

Brown, S. and J. Warner (1980). Measuring Security Price Performance, Journal of Financial Economics, 8, 205-258.

Brown, S. and J. Warner (1985). Using Daily Stock Returns: The Case of Event Studies, Journal of Financial Economics, 14, 3-31.

Campbell, C. and C. Wasley (1993). Measuring Security Price Performance Using Daily NASDAQ Returns, Journal of Financial Economics, 33, 73-92.

Carhart, M. (1997). On Persistence in Mutual Fund Performance, Journal of Finance, 52, 57-82.

Corrado, C. J. (1989). A Nonparametric Test for Abnormal Security-price Performance in Event Studies, Journal of Financial Economics, 23, 385-395.

Corrado, C. J. and C. Truong (2008). Conducting Event Studies with Asia-Pacific Security Market Data, Pacific-Basin Finance Journal, 16, 493-521.

Corrado, C. J. and T. L. Zivney (1992). The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns, Journal of Financial and Quantitative Analysis, 27, 465-478.

Cowan, A. R. (1992). Nonparametric Event Study Tests, Review of Quantitative Finance and Accounting, 2, 343-358.

Dutta, A. (2014a). Does Calendar Time Portfolio Approach Really Lack Power?, International Journal of Business and Management, 9, 260-266.

Dutta, A. (2014b). Investigating Long-Run Stock Returns after Corporate Events: the UK Evidence, Corporate Ownership and Control, 12, 298-307.

Eckbo, B. E., R. W. Masulis and Ø. Norli (2000). Seasoned Public Offerings: Resolution of the ‘New Issues Puzzle', Journal of Financial Economics, 56, 251-291.

Fama, E. (1991). Efficient Capital Markets: II, Journal of Finance, 46, 1575-1617.

Fama, E. (1998). Market Efficiency, Long-term Returns, and Behavioral Finance, Journal of Financial Economics, 49, 283-306.

Fama, E., L. Fisher, M. Jensen and R. Roll (1969). The Adjustment of Stock Prices to New Information, International Economic Review, 10, 1-21.

Fama, F. and K. French, (1993). Common Risk Factors in Returns on Stocks and Bonds, Journal of Financial Economics, 33, 3-56.

Jaffe, J. F. (1974). Special Information and Insider Trading, Journal of Business, 47, 410-428.

Jegadeesh, N. and J. Karceski (2009). Long-run Performance Evaluation: Correlation and Heteroskedasticity-Consistent Tests, Journal of Empirical Finance, 16, 101-111.

Knif, J., J. W. Kolari and S. Pynnönen (2013). A Robust and Powerful Test of Abnormal Stock Returns in Long-horizon Event Studies, Working Papers Series: Available at SSRN: http://ssrn.com/abstract=2292356

Kolari, J. W. and S. Pynnönen (2010). Event Study Testing with Cross-sectional Correlation of Abnormal Returns, Review of Financial Studies, 23, 3996-4025.

Kolari, J. W. and S. Pynnönen (2011). Nonparametric Rank Tests for Event Studies. Journal of Empirical Finance, 18, 953-971.

Kothari, S. P. and J. B. Warner (1997). Measuring Long-horizon Security Price Performance, Journal of Financial Economics, 43, 301-339.

Kothari, S. P. and J. B. Warner (2007). Econometrics of Event Studies, in: B. Eckbo Espen, ed., Handbook of Corporate Finance: Empirical Corporate Finance (Handbooks in Finance Series, Elsevier, North-Holland), 3-36.

Loughran, T. and J. Ritter (2000). Market Efficiency, Uniformly least powerful tests of market efficiency, Journal of Financial Economics, 55, 361-389.

Lyon, J., B. Barber and C. Tsai (1999). Improved Methods of Tests of Long-horizon Abnormal Stock Returns, Journal of Finance, 54, 165-201.

MacKinlay, A. C. (1997). Event Studies in Economics and Finance, Journal of Economic Literature, 35, 13-39.

Mandelker, G. (1974). Risk and Return: The Case of Merging Firms, Journal of Financial Economics, 1, 303-336.

Mitchell, M. and E. Stafford (2000). Managerial Decisions and Long-term Stock Price Performance, Journal of Business, 73, 287-329.

Patell, J. (1976). Corporate Forecasts of Earnings per Share and Stock Price Behavior: Empirical Tests, Journal of Accounting Research, 14, 246-276.

Ritter, J. (1991). The Long-Term Performance of Initial Public Offerings, Journal of Finance, 46, 3-27.

Zivney, T. L. and D. J. Thompson II (1989). The Specification and Power of the Sign Test in Measuring Security Price Performance: Comments and Analysis, The Financial Review, 24, 581-588.

International Business Research; Vol. 7, No. 12; 2014
ISSN 1913-9004 E-ISSN 1913-9012
Published by Canadian Center of Science and Education

Parametric and Nonparametric Event Study Tests: A Review

Anupam Dutta[1]

[1] Department of Mathematics & Statistics, University of Vaasa, Vaasa, Finland

Correspondence: Anupam Dutta, Department of Mathematics & Statistics, University of Vaasa, Vaasa, Finland. E-mail: adutta@uwasa.fi

Received: July 24, 2014 Accepted: October 24, 2014 Online Published: November 25, 2014

doi:10.5539/ibr.v7n12p136 URL: http://dx.doi.org/10.5539/ibr.v7n12p136

Abstract

This paper presents a modest attempt to review the existing methodologies for measuring short-run abnormal performance of firms following certain corporate events. In doing so, the study discusses different parametric as well as nonparametric testing procedures available in the literature. Reviewing the prior literature reveals that the nonparametric sign and rank tests are better specified than parametric procedures. However, in case of detecting the short-run anomalies, we document that nonparametric tests have higher power relative to standard parametric approaches.

Keywords: event study, short-run anomalies, sign test, rank test

1. Introduction

An event study is an empirical procedure that measures the effect of new information on the price of an asset, i.e. an event study is concerned with the impact of an event on the market prices of a company's publicly traded securities. In particular, researchers are concerned with the hypothesis that an event will have impact on the value of a firm or firms, and that this impact will be reflected on the stock and other security prices, manifesting itself in abnormal security returns. For instance, an event study might be conducted for the purpose of determining the impact of corporate earnings announcements on the stock price of the company.

The event study methodology is widely used in finance, accounting and economics. Many types of events are studied with event studies. Such events may include takeover announcements, environmental regulation enactments, patent filing announcements, competitor bankruptcy announcements, CEO resignation announcements, etc. Event studies are employed to measure market efficiency and to determine the impact of a given event on security prices. Such methodology refers to the set of econometric techniques used to measure and interpret the effects of an event on the value of a firm.

It is a difficult task to determine how many event studies have been published so far. Kothari and Warner (2007), for example, report that the number of published papers that deal with the event study methodology easily exceeds 500 and continues to grow. Although there have been many advances in this methodology over the years, the core elements of a typical event study can be found in two landmark papers by Ball and Brown (1968) and Fama, Fisher, Jensen, and Roll (1969) (henceforth FFJR).

The prime objective of this paper is to highlight the important parametric and nonparametric tests used in short-run event study methodology. To serve this purpose, we first review the existing literature of short-run event studies and then try to compare standard parametric tests with different nonparametric approaches available in the literature. Reviewing a large number of elementary studies suggests that the nonparametric sign and rank tests provide better specification and power than standard parametric approaches in detecting the abnormal performance.

The rest of the paper is organized as follows: Section 2 reviews the existing literature and reports the significant developments in the event study methodology. Parametric as well as nonparametric event study tests are discussed in Section 3. Section 4 outlines some recent developments in nonparametric approaches. Section 5 concludes.

2. Literature Review

Although the core elements of a typical event study are extensively summarized by Ball and Brown (1968) and FFJR, these papers are not the first that portray event studies. MacKinlay (1997) reports an early event study by Dolley (1933) which examines the stock price reaction to stock splits by studying nominal price changes at the time of the split. Using a sample of 95 splits from 1921 to 1931, Dolley finds that the price increased in 57 of the cases and the price declined in only 26 instances.

In the late 1960s, Ball and Brown (1968) and FFJR introduced the methodology that is essentially equivalent to that which is in use today. Ball and Brown (1968) conclude that annual accounting income data contains information that is related to stock prices. They found that income forecast errors, which are measured by the difference between announced and expected accounting earnings, have a positive impact on the abnormal performance index around the annual report announcement date.

FFJR also note that stock prices appear to adjust to new information. Stock splits generally occur following periods when stock prices significantly increase relative to the market. They found that, after a split announcement, stock prices seem to quickly reflect all available information and do not generate any abnormal returns. The results demonstrate the efficiency of the capital market.

Since these pioneering studies, numerous modifications have been developed in order to investigate the impact of a number of potential problems of concern in the literature which include non-normality of returns and excess returns, bias in OLS estimates of market model parameters in the presence of non-synchronous trading and estimation of the variance to be used in hypothesis tests concerning the mean excess return. Brown and Warner (1980, 1985) deal with the practical importance of these complications. In the 1980 paper, they consider implementation issues for data sampled at a monthly interval, while the 1985 paper deals with issues for daily data.

However, the issue of event-induced volatility has been a source of concern in the literature for some time. Brown and Warner (1980, 1985) report that increases in variance may result in misspecification of the traditional test statistics and that the power of tests can be improved by appropriately modeling the volatility process. Other studies such as Aktas et al. (2007), Harrington and Shrider (2007) and Higgins and Peterson (1998) also document that all events induce an increase in cross-sectional variance that must be estimated and adjustments embodied in all tests used to assess the statistical significance of event date abnormal returns. Boehmer, Musumeci, and Poulsen (BMP) (1991) argue that the event-period returns should be standardized by the estimation-period standard deviation, and the cross-sectional mean of the standardized returns needs to be divided by their cross-sectional standard deviation to yield the test statistic. BMP approach implicitly assumes that the event-induced variance is the same for all securities in the sample. Corrado (1989) introduces the nonparametric rank test to deal with the issue of event-induced variance. Simulations show higher power of the rank test relative to the traditional tests. Simulations in BMP approach also confirm the same.

In traditional event study methodology, however, it is assumed that the abnormal returns are cross-sectionally uncorrelated. This assumption is valid when the event day is not common to the firms. Brown and Warner (1980, 1985) show that even when the event day is common for the firms which are not from the same industry, use of the market model to derive the abnormal return reduces the inter-correlations virtually to zero. But, if the firms are from the same industry, extracting market factor may not reduce the cross-sectional residual correlation. Consequently, using the traditional standardized return test statistics, even moderate cross-sectional correlation in an event study causes substantial over-rejection of the null hypothesis of no abnormal performance. Kolari and Pynnönen (2010) propose simple corrections to the popular Patell (1976) and Boehmer, Musumeci, and Poulsen (1991) statistics to account for the correlation. They show that, when there is no event-induced volatility increase, each of these corrected test statistics is approximately equally powerful and rejects the null hypothesis of no abnormal performance at the correct nominal rate when it is true.

3. Event Study Tests

A number of event studies rely on parametric test statistics. But, one disadvantage of using parametric test statistics is that they do require essential assumptions about the probability distribution of returns. Brown and Warner (1985) report that stock prices are not normally distributed. Consequently, when this assumption of normality is violated, parametric tests yield misspecified test statistics.

Non-parametric tests, on the other hand, are well-specified and more powerful at detecting a false null hypothesis of no abnormal returns. The most successful among these tests were the nonparametric sign and rank tests advanced in Corrado (1989), Zivney and Thompson (1989), and Corrado and Zivney (1992). Well-known studies of this type are Cowan (1992), Campbell and Wasley (1993, 1996), and Corrado and Truong (2008). Each of these studies documents that sign and rank tests provide better specification and power than parametric tests.

In this section, we review different types of nonparametric event study tests available in the literature. In doing so, we first discuss standard parametric procedures for testing the null of no abnormal performance. Reviewing nonparametric tests will follow.

3.1 T-Test-Mean Excess Returns

Let ε_{it} denote the abnormal return of security *i* on day *t*, i.e., $\varepsilon_{it} = R_{it} - E(R_{it})$, where R_{it} represents the return of security *i* on day *t* and $E(R_{it})$ indicates the expected return generated by a particular benchmark model. Also let *t=0* be the event date. Now for each day *t*, the cross-sectional average excess return of *N* securities is calculated as:

$$\overline{\varepsilon_t} = \frac{1}{N}\sum_{i=1}^{N} \varepsilon_{it}$$

The day 0 test statistic is then given by:

$$J_1 = \frac{\overline{\varepsilon_0}}{S(\bar{\varepsilon})}$$

with $S(\bar{\varepsilon})$ being an estimate of standard deviation of the average abnormal returns.

3.2 T-Test-Mean Standardized Excess Returns

In this case, each ε_{it} is divided by its estimated standard deviation to produce a standardized excess return computed as:

$$\varepsilon'_{it} = \frac{\varepsilon_{it}}{S(\varepsilon_i)}$$

Then the day 0 test statistic is defined as:

$$J_2 = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \varepsilon'_{it}$$

3.3 Cross-Sectional Dependence (Crude Adjustment)

Brown and Warner (1980) suggest a crude dependence adjustment for cross sectional dependence. The variance of the average abnormal return of the event day is estimated using the time series variance of the average of the abnormal returns.

In this case, the day 0 test statistic is given by:

$$J_3 = \frac{\overline{\varepsilon_0}}{\sqrt{\widetilde{\sigma^2}}}$$

Here, $\widetilde{\sigma^2} = \frac{1}{T-1}\sum_{t=1}^{T}(\overline{\varepsilon_t} - \bar{\varepsilon})^2$ and $\bar{\varepsilon} = \frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\varepsilon_{it}$

3.4 Generalized Sign Test

The sign test, often used in event studies, refers to a simple binomial test of whether the frequency of positive abnormal residuals equals 50 percent. Brown and Warner (1980) point out that correct specification of the sign test requires equal numbers of positive and negative abnormal returns, absent a reaction to an event. Cowan, Nayar and Singh (1990) and Sanger and Peterson (1990), use a refined version of this sign test by allowing the null hypothesis to be different from 0.5 and this modified approach is called generalized sign test.

In order to implement this test, we first need to determine the proportion of securities in the sample having non-negative abnormal returns under the null hypothesis of no abnormal performance. The value for the null is estimated as the average fraction of stocks with non-negative abnormal returns in the estimation period. If abnormal returns are independent across stocks, under the null hypothesis the number of non-negative values of abnormal returns has a binomial distribution with parameter p.

The statistic for the generalized sign test is defined as:

$$z = \frac{|p_0 - p|}{\sqrt{p(1-p)N}}$$

where p_0 denotes the observed fraction of positive returns computed across stocks in one particular event week, or the average fraction of firms with non-negative abnormal returns for events occurring over multiple weeks. This statistic is approximately distributed as normal distribution with zero mean and variance 1.

The advantage of the generalized sign test is that it takes into account the evidence of skewness in security returns. However, the power and specification of the generalized sign test have not been documented.

3.5 Wilcoxon Signed-Rank Test

Employing Wilcoxon signed-rank test is handy, since it considers that both the sign and the magnitude of abnormal returns are significant. The test statistic in this case is given by:

$$W = \sum_{i=1}^{N} r_i^+$$

Where r_i^+ is the positive rank of the absolute value of abnormal returns. This test assumes that none of the absolute values are equal, and that each is a nonzero value. Under the null hypothesis of equally likely positive or negative abnormal returns and when N is large, W asymptotically follows a normal distribution with the following mean and variance:

$$E(W) = \frac{N(N+1)}{4}$$

$$V(W) = \frac{N(N+1)(2N+1)}{12}$$

3.6 Corrado's Rank Test

Corrado (1989) observes that another nonparametric test, known as the rank test, is more powerful than the standard parametric tests. Like the generalized sign test, the rank test does not require symmetry of the cross-sectional abnormal return distribution.

In order to implement this test, it is first necessary to transform each firm's abnormal returns into their respective ranks. To do so, let K_{it} denote the rank of the abnormal return ε_{it} in security *i*'s time series of T excess returns, i.e., $K_{it} = rank(\varepsilon_{it})$; t=1, 2,…, T. Here $\varepsilon_{it} \geq \varepsilon_{ij}$ means $K_{it} \geq K_{ij}$ and $T \geq K_{it} \geq 1$. The average rank is then calculated as $\bar{K} = \frac{T+1}{2}$ and the day 0 test statistic is given by

$$R = \frac{\frac{1}{N}\sum_{i=1}^{N}(K_{i0} - \bar{K})}{S(K)}$$

where the standard deviation S(K) is computed as:

$$S(K) = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left[\sum_{i=1}^{N}(K_{it} - \bar{K})/N\right]^2}$$

This statistic is distributed asymptotically as unit normal. Cowan and Sergeant (1996) document that if the return variance is unlikely to increase, then Corrado's rank test is better specified and more powerful than parametric tests. With the increase in variance, however, this test is misspecified.

Table 1 presents parametric and nonparametric event study tests reviewed in this paper. Each of these approaches is employed to investigate the short-run abnormal performance of firms following major corporate events.

Table 1. Summary of alternative methodologies

Methods	Test Statistics
t-test-mean excess returns	$J_1 = \frac{\overline{\varepsilon_0}}{S(\bar{\varepsilon})}$
t-test-mean standardized excess returns	$J_2 = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varepsilon'_{it}$
Cross-Sectional Dependence (Crude Adjustment)	$J_3 = \frac{\overline{\varepsilon_0}}{\sqrt{\overline{\sigma^2}}}$
Generalized Sign Test	$z = \frac{\lvert p_0 - p \rvert}{\sqrt{p(1-p)N}}$
Wilcoxon Signed-Rank Test	$W = \sum_{i=1}^{N} r_i^+$
Corrado's Rank Test	$R = \frac{\frac{1}{N}\sum_{i=1}^{N}(K_{i0} - \bar{K})}{S(K)}$

Note. This table summarizes the test statistics of different empirical procedures discussed in this study. The first three methodologies refer to parametric event study approaches and the rest indicate nonparametric procedures.

4. Recent Developments in Non-parametric Event Study Tests

The rank tests, introduced by Corrado (1989) and Corrado and Zivney (1992), are applied for testing single day event abnormal returns. Corrado (1989), however, reports that implementing rank test for CARs requires defining multiple-day returns that match the number of days in the CARs. This can be done by dividing the estimation period and event period into intervals matching the number of days in the CAR. Unfortunately, this procedure is not very effective, because the number of observations quickly becomes impracticably small as the CAR-period lengthens and the resultant loss of observations weakens the abnormal return model estimation. Cowan (1992) and Campbell and Wasley (1993) conduct Corrado's rank test for testing cumulative abnormal returns by simply accumulating daily ranks of abnormal returns within the CAR-period. Like the multi-day approach, cumulated ranks approach also has potential shortcomings. Cowan (1992) and Kolari and Pynnönen (2010) report that such procedure quickly loses power in detecting abnormal returns, especially in longer event windows. Because this approach involves transferring the returns to rank numbers and hence the returns no longer capture the magnitudes of returns, only their relative ranks. Thus, if one large return is randomly assigned to one day within the event window independently for each stock, there is only one potentially outstanding rank for each stock that is randomly scattered across the window. This is likely to average largely out in the cumulative rank sum resulting in poor power properties of the test.

In order to overcome these puzzles, Kolari and Pynnönen (2011) introduce a generalized rank test based on generalized standardized abnormal returns which can be applied for testing both single abnormal returns and cumulative abnormal returns. The proposed test is robust to abnormal return serial correlation, event-induced volatility and cross-sectional correlation of abnormal returns due to event day clustering. Further details can be found in Kolari and Pynnönen (2011).

5. Conclusions

Event studies are conducted for the purpose of investigating the effect of an event on stock returns. Typical events include firm-specific events and Economy-wide events. Firm-specific events usually indicate a change in the company policy. Examples of such events involve earnings, investment, mergers and acquisitions, issues of new debt or equity, stock splits, etc. announcements. Economy-wide events, on the other hand, are employed in large sample event studies which investigate the impact of a particular event on relevant securities. This type of events includes inflation, interest rate, consumer confidence, trade deficient, etc. announcements.

This paper presents a modest attempt to portray the short-run event study methodology beginning with FFJR in the late 1960s. The main objective of this article is to outline the existing parametric and nonparametric tests used in short-run event studies. To serve this purpose, standard parametric tests, Generalized Sign Test, Wilcoxon Signed-Rank Test and Corrado's Rank Test are discussed. Recent developments in non-parametric event study tests are also reviewed. For example, Kolari and Pynnönen (2011) recently introduce a generalized rank test based on generalized standardized abnormal returns which is used to test both single abnormal returns and cumulative abnormal returns. Reviewing the prior studies concludes that nonparametric sign and rank tests are well specified and have more power than the standard parametric approaches in detecting the short-run anomalies.

Acknowledgement

The author thanks Professor Seppo Pynnönen for his valuable comments and suggestions. His assistance and guidance are also acknowledged.

References

Aktas, N., de Bodt, E., & Cousin, J. G. (2007). Event Studies with a Contaminated Estimation Period. *Journal of Corporate Finance, 13*(1), 129–145. http://dx.doi.org/10.1016/j.jcorpfin.2006.09.001

Ball, R., & Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research, 6*, 159–177.

Boehmer, E., Musumeci, J., & Poulsen, A. B. (1991). Event-study Methodology under Conditions of Event-induced Variance. *Journal of Financial Economics, 30*(2), 253–272. http://dx.doi.org/10.1016/0304-405X(91)90032-F

Brown, S., & Warner, J. (1980). Measuring Security Price Performance. *Journal of Financial Economics, 8*, 205–258. http://dx.doi.org/10.1016/0304-405X(80)90002-1

Brown, S., & Warner, J. (1985). Using Daily Stock Returns: The Case of Event Studies. *Journal of Financial Economics, 14*, 3–31. http://dx.doi.org/10.1016/0304-405X(85)90042-X

Campbell, C., & Wasley, C. (1993). Measuring Security Price Performance Using Daily NASDAQ Returns. *Journal of Financial Economics, 33*, 73–92. http://dx.doi.org/10.1016/0304-405X(93)90025-7

Corrado, C. J. (1989). A Nonparametric Test for Abnormal Security-Price Performance in Event Studies. *Journal of Financial Economics, 23*, 385–395. http://dx.doi.org/10.1016/0304-405X(89)90064-0

Corrado, C. J., & Truong, C. (2008). Conducting Event Studies with Asia-Pacific Security Market Data. *Pacific-Basin Finance Journal, 16*, 493–521. http://dx.doi.org/10.1016/j.pacfin.2007.10.005

Corrado, C. J., & Zivney, T. L. (1992). The Specification and Power of the Sign Test in Event Study Hypothesis Tests Using Daily Stock Returns. *Journal of Financial and Quantitative Analysis, 27*(3), 465–478. http://dx.doi.org/10.2307/2331331

Cowan, A. R. (1992). Nonparametric Event Study Tests. *Review of Quantitative Finance and Accounting, 2*, 343–358. http://dx.doi.org/10.1007/BF00939016

Cowan, A. R., & Sergeant, A. M. A. (1996). Trading Frequency and Event Study Test Specification. *Journal of Banking and Finance, 20*(10), 1731–1757. http://dx.doi.org/10.1016/S0378-4266(96)00021-0

Cowan, A. R., Nayar, N., & Singh, A. K. (1990). Stock Returns before and after Calls of Convertible Bonds. *Journal of Financial and Quantitative Analysis, 25*, 549–554. http://dx.doi.org/10.2307/2331016

Dolley, J. C. (1933). Characteristics and Procedure of Common Stock Split-ups. *Harvard Business Review, 11*, 316–326.

Dyckman, T., Philbrick, D., & Stephan, J. (1984). A Comparison of Event Study Methodologies Using Daily Stock Returns: A Simulation Approach. *Journal of Accounting Research, 22*, 1–30.

Fama, E., Fisher, L., Jensen, M., & Roll, R. (1969). The Adjustment of Stock Prices to New Information. *International Economic Review, 10*, 1–21.

Harrington, S. E., & Shrider, D. G. (2007). All Events Induce Variance: Analyzing Abnormal Returns When Effects Vary Across Firms. *Journal of Financial and Quantitative Analysis, 42*(1), 229–256. http://dx.doi.org/10.1017/S002210900000226X

Higgins, E. J., & Peterson, D. R. (1998). The Power of One and Two Sample t-statistics Given Event-induced Variance Increases and Nonnormal Stock Returns: A Comparative Study. *Quarterly Journal of Business and Economics, 37*(1), 27–50.

Kolari, J. W., & Pynnönen, S. (2010). Event Study Testing with Cross-sectional Correlation of Abnormal Returns. *Review of Financial Studies, 23*, 3996–4025. http://dx.doi.org/10.1093/rfs/hhq072

Kolari, J. W., & Pynnönen, S. (2011). Nonparametric Rank Tests for Event Studies. *Journal of Empirical Finance,* *18*, 953–971. http://dx.doi.org/10.1016/j.jempfin.2011.08.003

Kothari, S. P., & Warner, J. B. (2007). Econometrics of Event Studies. In B. Eckbo Espen (Ed.), *Handbook of Corporate Finance: Empirical Corporate Finance* (pp. 3–36). Handbooks in Finance Series, Elsevier, North-Holland.

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature, 35*(1), 13–39.

Patell, J. (1976). Corporate Forecasts of Earnings per Share and Stock Price Behavior: Empirical Tests. *Journal of Accounting Research, 14*, 246–276.

Sanger, G. C., & Peterson, J. D. (1990). An Empirical Analysis of Common Stock Delistings. *Journal of Financial and Quantitative Analysis, 25*(2), 261–272. http://dx.doi.org/10.2307/2330828

Zivney, T. L., & Thompson, D. J. (1989). The Specification and Power of the Sign Test in Measuring Security Price Performance: Comments and Analysis. *The Financial Review, 24*, 581–588. http://dx.doi.org/10.1111/j.1540-6288.1989.tb00362.x

Improved Calendar Time Approach for Measuring Long-Run Anomalies

Anupam Dutta
Department of Mathematics and Statistics
University of Vaasa

Abstract

Although a large number of recent studies employ the buy-and-hold abnormal return methodology and the calendar time portfolio approach to investigate the long-run anomalies, each of the methods is a subject to criticisms. This paper introduces a variant of calendar time portfolio approach where we first standardize the abnormal returns of the event firms forming the monthly portfolios and then weight each of these portfolios such that periods of heavy event activity receive higher loading than periods of light event activity. In our proposed method, we, however, use characteristics-based reference portfolios to measure the abnormal returns. Simulations show that the refined calendar time approach documents better specification and power than the conventional methodologies.

Keywords: Long-run anomalies, Standardized abnormal returns, Test specification, Power of test.

JEL Classification: C1, G1.

1. Introduction

Two commonly used methodologies for investigating the long-term stock price performance following major corporate events are the buy-and-hold abnormal return (BHAR) approach and the calendar time portfolio (CTP) method. The BHAR is defined as the difference between the long-run holding period return of a sample firm and that of some benchmark asset. Mitchell and Stafford (2000) explain BHAR returns as the average multiyear return from a strategy of investing in all firms that complete an event and selling at the end of a prespecified holding period versus a comparable strategy using otherwise similar nonevent firms. The calendar time method, on the other hand, is based on the mean abnormal time series returns to monthly portfolios of event firms.

Following the work of Ritter (1991), BHAR becomes one of the most popular estimators in the literature of long-horizon event studies. A large number of papers have applied the BHAR approach in Measuring long-horizon security price performance. Important examples include Barber and Lyon (1997) and Lyon, Barber, and Tsai (1999) (henceforth LBT). These studies document that an appealing feature of using BHAR is that buy-and-hold returns better resemble investors' actual investment experience than periodic (monthly) rebalancing entailed in other approaches to measuring risk-adjusted performance.

Fama (1998), however, argues against the BHAR methodology because of the statistical problems associated with the use of BHAR and the relevant test statistics. He reports that BHAR does not address the issue of potential cross-sectional correlation of event-firm abnormal returns. Mitchell and Stafford (2000) also question the application of BHAR approach suggesting that the assumption of independence of observations is violated and hence the cross-sectional correlations significantly bias the test statistics that are computed from the BHARs. In addition, Eckbo et al. (2000) document that the BHAR methodology is not a feasible portfolio strategy because the total number of stocks is not known in advance. As an alternative, the calendar time portfolio approach, developed by Jaffe (1974) and Mandelker (1974), is widely used to resolve the issue of cross-sectional dependence of abnormal returns (Fama, 1998 and Mitchell and Stafford, 2000). Fama (1998) strongly recommends the use of CTP methodology on the grounds that monthly returns are less susceptible to the bad model problem as they are less skewed and by forming monthly calendar time portfolios, all cross-correlations of event-firm abnormal returns are automatically accounted for in the portfolio variance. Fama also documents that the distribution of this estimator is better approximated by the normal distribution, allowing for classical statistical inference. Mitchell and Stafford (2000), like Fama (1998), also prefer the CTP approach to BHAR methodology as the latter assumes independence of multi-year event firm abnormal returns.

While many recent studies strongly advocate the CTP approach, it has a number of potential pitfalls. For example, if there is a differential abnormal performance in periods of heavy event activity versus periods of light event activity, the regression approach will average over these, and it may be less likely to identify the abnormal performance. Loughran and Ritter (2000) argue that corporate executives time the events to exploit mispricing, but the CTP approach, by forming calendar-time portfolios, under-weights managers' timing decisions and over-weights other observations. Since the CTP approach weights each period equally, it has lower power to detect abnormal performance if managers time corporate events to coincide with misvaluations. LBT, however, claim that the CTP approach is misspecified in nonrandom samples, while the BHAR approach is relatively robust.

The purpose of this paper is to refine the calendar time portfolio approach in order to deal with the ongoing debates discussed in previous studies. To do so, a variant of calendar time method is proposed where we first standardize the abnormal returns for each of the event firms in the sample and then construct the monthly portfolios. That is, we first obtain the event firms' monthly abnormal returns (adjusted by characteristics-based reference portfolio returns), then standardize them by their standard deviations over the long-run holding period and finally form the monthly portfolios from the event firms. We also propose to weight each monthly portfolio return which takes into account the number of event firms in the month. Weighting the portfolios in this way gives more weight to periods of heavy event activity than periods of low event activity. To assess the robustness of the proposed method, we also present the results from Fama-French three-factor (henceforth FF3F) model, and the BHAR approach.

Standardizing the abnormal returns is a common practice in the short-run event studies. Patell (1976) report that short-run tests using standardized returns have superior power compared to those based on unstandardized returns. Besides, standardizing event returns also reduces the heteroskedastic error term. The use of standardized abnormal returns is well-documented in long-run event studies as well. Fama (1998), for example, recommends to the use of standardized abnormal returns to improve the power of the calendar time methodology. Following Fama (1998), Mitchell and Stafford (2000) also standardize each month's abnormal return by an estimate of its standard deviation.

However, LBT report that the calendar-time portfolio methods based on reference-portfolio abnormal returns generally dominate those based on asset pricing models (e.g., Fama-French three-factor model) for two reasons. First, the three-factor model implicitly assumes linearity in the constructed market, size, and book-to-market factors. But LBT find that this assumption is unlikely to be the case for the size and book-to-market factors. Second, while the Fama- French three-factor model assumes there is no interaction between the three factors, LBT document that this assumption is also likely violated because the relation between book-to-market ratio and returns is most pronounced for small firms. Later, Loughran and Ritter (2000) also argue that the three-factor model is not an equilibrium model since it only detects anomalies in financial markets and fails to test market efficiency. In our proposed method, we, therefore, employ the reference portfolio approach based on size and book-to-market ratio to measure the abnormal returns. LBT and Mitchell and Stafford (2000) also use the reference portfolio method for computing long-run abnormal stock returns.

The standardized calendar time approach (henceforth SCTA) introduced in this paper contributes to the existing literature of long-run event studies in several ways. First, it resolves the concerns raised by Loughran and Ritter after achieving higher power than the traditional methodologies in detecting the anomalies. Second, it considers - for the first time- the standardized abnormal returns of the event firms constituting the monthly portfolios. Third, it allows months of hot event activity to receive more weight than months of cold event activity. Finally, our proposed approach is generally well-specified when nonrandom samples are taken into account. Hence SCTA presents a number of potential improvements over the conventional calendar time portfolio approach.

The remainder of the paper is structured as follows. Section 2 outlines the data and methodology. Our simulation procedure is explained in Section 3. Section 4 presents the specification of the tests. Section 5 reports power of the tests, and Section 6 concludes the paper.

2. Data and Methodology

The data used in this paper consist of NYSE, Amex, and Nasdaq stocks, and our sample period ranges from July 1978 to December 2007. We obtain monthly returns, market value (MV) or size and book-to-market (BM) value data from Datastream.

We construct 25 size-BM portfolios as expected return benchmarks. In doing so, at the end of June of year t, we allocate all the stocks to one of five size groups, based on size rankings relative to NYSE quintiles. In an independent sort, all stocks are also allocated to one of five BM groups, based on their BM ranks relative to NYSE quintiles.

2.1. Standardized Calendar Time Approach (SCTA)

The conventional way of calculating the mean monthly calendar time abnormal return (CTAR) is the following

$$\text{mean monthly CTAR} = \frac{1}{T}\sum_{t=1}^{T} CTAR_t, \tag{1}$$

where

$$CTAR_t = R_{pt} - E(R_{pt}), \tag{2}$$

Within this framework, R_{pt} is the monthly return on the portfolio of event firms, $E(R_{pt})$ is the expected return on the event portfolio which is proxied by the raw return on a reference portfolio and T is the total number of months in the sample period.

However, a number of firms in the sample often produce volatile returns. Small firms, for instance, usually exhibit such pattern and because of this volatility, the distributions of long-run returns tend to have fat tails. One possible solution to this problem is standardizing the abnormal returns by their volatility measures. In this paper, we use standardized abnormal returns to compute the CTARs. The whole procedure is done in two steps. We first calculate the standardized abnormal returns for each of the sample firms. In doing so, the abnormal return for firm i is computed as $\varepsilon_{it} = R_{it} - E(R_{it}); t = 1, \ldots, H$, where R_{it} denotes the return on event firm i in the calendar month t, $E(R_{it})$ is the expected return on the event portfolio which is proxied by 25 size-BM reference portfolios and H is the holding period which equals 12, 36, or 60 months. The next task is to estimate the event-portfolio residual variances using the H-month residuals computed as monthly differences of i-th event firm returns and size-BM portfolio returns. Dividing ε_{it} by the estimate of its standard deviation yields the corresponding standardized abnormal return, say z_{it}, for event firm i in month t. Now let N_t refer to the number of event firms in the calendar month t. We then calculate the calendar time abnormal return for portfolio t as

$$CTAR_t = \sum_{i=1}^{N_t} x_{it} z_{it}, \tag{3}$$

where the weight x_{it} equals $\frac{1}{N_t}$ when the abnormal returns are equally-weighted and $\frac{MV_{it}}{\sum MV_{it}}$ when the abnormal returns are value-weighted by size.

We also propose to weight each of the monthly CTARs by $\frac{1}{\sqrt{(\sum_{i=1}^{N_t} x_{it}^2)}}$. For instance, when the abnormal returns are equally weighted i.e., when $x_{it} = \frac{1}{N_t}$, then $\frac{1}{\sqrt{(\sum_{i=1}^{N_t} x_{it}^2)}} = \sqrt{N}_t$. This weighting scheme is lucrative as it gives more loadings to periods of heavy event activity than the periods of low event activity. Now the grand mean monthly abnormal return ($\overline{CTAR}$) is calculated as

$$\overline{CTAR} = \frac{1}{T} \sum_{t=1}^{T} CTAR_t \tag{4}$$

While finding $\overline{CTAR}$, it might be the case that a number of portfolios do not contain any event firm. In such situations, those months are dropped from the analysis. To test the null hypothesis that there is no abnormal performance, the t-statistic of $\overline{CTAR}$ is computed by using the intertemporal standard deviation of the monthly CTARs defined in equation (3).

2.2. Buy-and-Hold Abnormal Return (BHAR)

Once a reference portfolio is identified, computing BHARs is straightforward. A H-month BHAR for event firm i is defined as:

$$BHAR_{iH} = \prod_{t=1}^{H} (1 + R_{it}) - \prod_{t=1}^{H} (1 + R_{Bt}), \tag{5}$$

where R_{it} denotes the return on event firm i at time t and R_B indicates the return on 25 size-BM reference portfolios.

To test the null hypothesis that the mean buy-and-hold return equals zero, the conventional t-statistic is given by

$$t_{BHAR} = \frac{\overline{BHAR_H}}{\sigma(BHAR_H)/\sqrt{n}} \ , \tag{6}$$

where $\overline{BHAR_H}$ implies the sample mean and $\sigma(BHAR_H)$ refers to the cross-sectional sample standard deviation of abnormal returns for the sample containing n firms.

However, the earlier studies (e.g. Mitchell and Stafford (2000), Boehme and Sorescu (2002), Jegadeesh and Karceski (2009)) report that the BHAR approach does not control well for the cross-sectional correlation among individual firms in nonrandom samples and thus yields misspecified t-statistics. Moreover, the test statistics based on BHARs also suffers from this misspecification problem due to the severe skewness of the distribution of BHARs. Though bootstrapping corrects for the skewness problem to some extent, it ignores the cross-sectional dependence of abnormal returns.

2.3. Fama-French Three-Factor Model

For each calendar month t, we form portfolios consisting of all sample firms that have participated in the event within the last H months, where H equals 12, 36, or 60 in our study. For each calendar month, the portfolios are rebalanced, i.e., the firms that reach the end of their H-month period drop out and new firms that have just executed a transaction are added. We then calculate the portfolio mean monthly abnormal return α_p by regressing its excess return on the three Fama-French factors:

$$R_{pt} - R_{ft} = \alpha_p + \beta_p(R_{mt} - R_{ft}) + s_p SMB_t + h_p HML_t + e_{pt}, \tag{7}$$

where R_{pt} is the equal or value-weighted return on portfolio t, R_{ft} is the risk-free rate, $R_{mt} - R_{ft}$ is the excess return of the market, SMB is the difference between the return on the portfolio of small stocks and big stocks, HML is the difference between the return on the portfolio of high and low book-to-market stocks, α_p measures the mean monthly abnormal return of the calendar time portfolio which is zero under the null hypothesis of no abnormal performance and β_p, s_p and h_p are sensitivities of the event portfolio to the three factors.

However, since the number of firms changes over the sample period, this may cause the error term to be heteroskedastic and hence the ordinary least squares estimate becomes inefficient. Fama (1998), therefore, suggests to apply the weighted least squares technique instead of ordinary least squares to control for heteroskedasticity. In this study, we estimate regression (7) using weighted least squares (WLS) procedures. Monthly returns in the WLS model are weighted by $\sqrt{N_t}$, where N_t stands for the number of event firms in month t.

3. Simulation Method

To test the specification of the t-statistics, we randomly select 1,000 samples of 200 event months without replacement. For each of these 200 event months, we randomly draw one stock from the population of all stocks that are active in the database for that month. For a well-specified test statistic, 1000α tests reject the null hypothesis. A test is conservative if fewer than 1000α null hypotheses are rejected and is anticonservative if more than 1000α null hypotheses are rejected. Based on this procedure, we test the specification of the t-statistic at 5% theoretical levels of significance. A well-specified null hypothesis rejects the null at the theoretical rejection level in favor of the alternative hypothesis of negative (positive) abnormal returns in $1000\alpha/2$ samples.

4. Test Specification

This section reports the specification of various methodologies used in our study. We first discuss the results in random samples. Later, we consider different types of nonrandom samples based on firm size, book-to-market ratio, pre-event return performance, and overlapping returns.

4.1. Random Samples

Table 1 indicates the rejection rates in 1000 simulations with a random sample of 200 firms. Findings reveal that all the *t*-statistics based on buy and hold abnormal returns are negatively biased. For example, when the horizons are 5 years, the rejection rates at the 5% level of significance are 4.8% and 0%. These results are consistent with those documented by LBT where the tests have higher rejection rates in the lower tail.

As anticipated, all the calendar time portfolio methods considered in our analysis are well-specified in random samples regardless of whether equally-weighted or value-weighted portfolios are employed. For example, for a 3-year holding period and with equally-weighted portfolios, the rejection rates at 5% level of significance are 2.4% and 2.8% and 2% and 0.8% for the *t*-statistics produced by standardized calendar time approach and Fama-French three-factor model respectively.

[Table 1]

However, our proposed calendar time approach involves two components: standardization and weighting. In order to verify whether it is the standardization or the weighting approach that improves the specification of tests, the table below presents a set of simulation results for the following cases: a) Standardization only approach and b) Weighting only approach. Table 1A shows that the standardization only approach produces well-specified test statistics when the length of the investment periods is either three or five years. But, for a one-year horizon, there is some evidence of misspecification. For the weighting only approach, we document misspecifications when the holding periods are one year and five years. Similar comments also apply when the portfolios are value-weighted. Hence, we conclude that the standardized calendar time approach, which involves both standardization and weighting components, improves the size of tests as shown in Table 1.

[Table 1A]

4.2. Nonrandom Samples

4.2.1. Firm Size

In order to investigate the effect of size-based sampling biases on the employed methods, we randomly choose 1000 samples separately from the largest size decile and smallest size decile. These results are presented in Tables 2A and 2B. Our analysis indicates that among the three methods, SCTA is better specified for each type of samples based on size. For instance, for a five-year horizon and with small firms and value-weighted portfolios, the rejection rates at 5% level of significance are 0.4% and 2.8% for SCTA, and 0.8% and 4.4% for FF3F model. However, the t-statistics based on BHAR approach are negatively skewed for samples containing small firms and positively skewed for samples consisting of large firms.

[Table 2A]

[Table 2B]

4.2.2. Book-to-Market (BM) Ratio

Firms are deciled into ten groups based on rankings of BM ratio at the end of June each year. We choose the groups with the highest BM ratio and the lowest BM ratio for robustness check. For each group, we select a random sample of 200 firms. The procedure is repeated 1000 times and Tables 3A and 3B report the rejection rates. Inspection of these tables suggests that the standardized CTP approach yields reasonably well-specified test statistics in each case. For example, for a 3-year holding period and with equally-weighted portfolios and firms with low book-to-market ratio, the rejection rates at 5% level of significance are 1.6% and 2.0% for SCTA, and 0.2% and 5.6% for FF3F model. The BHAR method, on the other hand, produces either negatively or positively skewed test statistics depending on low or high book-to-market value respectively.

[Table 3A]

[Table 3B]

4.2.3. Pre-event Return Performance

To assess the specification of the tests under study, we consider drawing firms on the basis of pre-event return performance. Following LBT, we compute the preceding six-month buy-and-hold return on all firms in each month from July 1978 through December 2012. We then decile this six-month return and separately select 1,000 samples of 200 firms from the high-return decile and the low-return decile. The findings of our analysis are shown in Tables 4A and 4B.

Scrutinizing these two tables suggests that most of the tests produce either positively or negatively biased test statistics and these results resemble those of LBT. Jegadeesh and Titman (1993), however, also report similar types of findings. LBT, however, suggest to match sample firms to firms of similar pre-event return performance to avoid this type of misspecification. Following LBT, we construct $5 \times 5 \times 5$ reference portfolios based on size, BM and pre-event return performance. Simulated results, shown in Tables 4C and 4D, report that the level of specification of the methods considered improves. We employ the four-factor model, proposed by Carhart (1997), to include the momentum factor computed as the difference between returns of winners and losers. We, therefore, exclude the three-factor model while constructing Tables 4C and 4D. Unfortunately, such portfolios are not equally effective (not reported in the table) when analyzing random samples as well as other nonrandom samples.

[Table 4A]

[Table 4B]

[Table 4C]

[Table 4D]

4.2.4. Overlapping Returns

With a view to inspecting how the methods used in this paper behave in the presence of cross-sectional correlation of abnormal returns, we consider nonrandom samples based on overlapping returns. Selecting these samples consists of two steps. The first stage involves a random selection of 100 firms from the population. In the second stage, for each of these 100 firms, we randomly choose a second event month that is within $H - 1$ periods of the original event month (either before or after), where H equals 12, 36 or 60. Hence we have 200 firms with 200 event months where the same firm appears in the sample twice and this generates the issue of overlapping returns. We repeat this procedure 1000 times and the results are presented in Table 5.

Findings indicate that the BHAR approach yields misspecified test statistics and these results are consistent with those reported in previous studies (e.g., LBT and

Mitchell and Stafford (2000)). This misspecification is due to the fact that BHAR assumes that the observations are cross-sectionally uncorrelated. This assumption is tenable in random samples of event firms, but it would be violated in nonrandom samples, where the returns for event firms are positively correlated (Jegadeesh and Karceski, 2009).

Results further confirm that each of the calendar time methods performs well when return calculations do overlap. In most of the cases, the equally-weighted scheme produces higher rejection compared with the value-weighted scheme. For example, for a 3-year holding period and with equally-weighted portfolios, the rejection rates at 5% level of significance are 2.9% and 3.6% for the t-statistics produced by SCTA. The corresponding rejection rates are 1.8% and 3.1% when the portfolios are value-weighted. It indicates that the value-weighted scheme should be taken into account if misspecifications occur due to overlapping returns. Last but not the least, the CTP approach controls well for the problem of cross-sectional dependence and is thus recommended while dealing with cross-sectionally correlated returns.

[Table 5]

5. Power

In this section, we compare the power of all the three methods employed in our study. Note that we only choose random samples since the t-tests are not, in general, well-specified in nonrandom samples. To examine the power of test, we introduce a constant level of abnormal return ranging from -20% to 20% at an interval of 5% to event firms. We also consider equally-weighted portfoilos to make a direct comparison with BHAR approach. Table 6 indicates the percentages of 1000 samples of 200 firms that reject the null hypothesis of zero abnormal returns over a three-year holding period. Figure 1 also plots power of the tests.

It is evident from Table 6 and Figure 1 that our proposed standardized calendar time approach produces the most powerful t-statistic, followed by the BHAR method, and the test statistic based on Fama-French three-factor model is the least powerful. For instance, with 15% per year abnormal returns, the rejection rate is 92% for SCTA, 80% BHAR method and 58% for FF3F model. We, therefore, conclude that the standardized calendar time approach produces more power to detect the abnormal performance than the BHAR method after accounting for cross-sectional correlation of the event firms.

[Table 6]

[Figure 1]

6. Conclusion

The proper methodology for analyzing the long-term return anomalies has been much debated in the literature. Kothari and Warner (2007), for instance, report that the question of which model of expected returns is correct remains an unresolved issue. Fama (1998) also concludes that not a single model for expected returns can fully describe the systematic patterns in normal returns and hence the anomalies arise because of the misspecification of models and the statistical tests applied. A fundamental choice for many recent studies, therefore, concerns the measure of long-run stock price performance.

Although numerous event studies have employed the BHAR methodology and the calendar time portfolio approach for investigating the long-run abnormal performance, each method has serious limitations. For example, Mitchell and Stafford (2000) argue against using the BHAR methodology as it assumes event-firm abnormal returns to be independent. They argue that major corporate actions are not random events, and thus event samples are likely to consist of dependent observations. In particular, major corporate events cluster through time by industry. This leads to positive cross-correlation of abnormal returns, making test statistics that assume independence severely overstated. They, like Fama (1998), strongly recommend the use of CTP approach. Loughran and Ritter (2000), however, report that the calendar time portfolio approach weights each month equally so that months that reflect heavy event activity are treated the same as months with low activity. Thus CTP approach may fail to detect significant abnormal returns if abnormal performance primarily exists in months of heavy event activity.

In this paper, we propose a modified calendar time portfolio approach which has two major components: standardization of event firms' abnormal returns and weighting the monthly portfolios. While standardizing diminishes the impact of event firms having volatile future returns, weighting allows monthly portfolios containing more event firms to receive more weight. The empirical analysis shows that these two innovations improve the size and power properties of statistical tests used in long-run event studies.

References

Barber, B., and J. Lyon 1997. Detecting Long-run Abnormal Stock Returns: The Empirical Power and Specification of Test Statistics. *Journal of Financial Economics* 43: 341-372.

Boehme, D. R., and S. M. Sorescu 2002. The Long-Run Performance Following Dividend Initiations and Resumptions: Underreaction orProduct of Chance?. *Journal of Finance* 57: 871-900.

Carhart, M. 1997. On Persistence in Mutual Fund Performance. *Journal of Finance* 52: 57-82.

Eckbo, B. E., R. W. Masulis, and Ø. Norli 2000. Seasoned Public Offerings: Resolution of the 'New Issues Puzzle'. *Journal of Financial Economics* 56: 251-291.

Fama, E. 1998. Market Effciency, Long-term Returns, and Behavioral Finance. *Journal of Financial Economics* 49: 283-306.

Jaffe, J. F. 1974. Special Information and Insider Trading. *Journal of Business* 47: 410-428.

Jegadeesh, N., and S. Titman 1993. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance* 48: 65-91.

Jegadeesh, N., and J. Karceski 2009. Long-run Performance Evaluation: Correlation and Heteroskedasticity-Consistent Tests. *Journal of Empirical Finance* 16: 101-111.

Kothari, S. P., and J. B. Warner 2007. Econometrics of Event Studies. in: B. Eckbo Espen, ed., Handbook of Corporate Finance: Empirical Corporate Finance (Handbooks in Finance Series, Elsevier, North-Holland), 3-36.

Loughran, T., and J. Ritter 2000. Market Efficiency, Uniformly least powerful tests of market efficiency. *Journal of Financial Economics* 55: 361-389.

Lyon, J., B. Barber, and C. Tsai 1999. Improved Methods of Tests of Long-horizon Abnormal Stock Returns. *Journal of Finance* 54: 165-201.

Mandelker, G. 1974. Risk and Return: The Case of Merging Firms. *Journal of Financial Economics* 1: 303-336.

Mitchell, M., and E. Stafford 2000. Managerial Decisions and Long-term Stock Price Performance. *Journal of Business* 73: 287-329.

Patell, J. 1976. Corporate Forecasts of Earnings per Share and Stock Price Behavior: Empirical Tests. *Journal of Accounting Research* 14: 246-276.

Ritter, J. 1991. The Long-Term Performance of Initial Public Offerings. *Journal of Finance* 46: 3-27.

Table 1: Specification of Tests in Random Samples

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	1.2	2.8	2.4	2.8	1.2	2.4
Buy-and-Hold Abnormal Return Method	6.2*	0.8	4.4*	0.4	4.8*	0.0
Fama-French Three-Factor Model	4.4*	0.8	2.0	0.8	3.2	1.6
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	1.6	2.4	1.2	3.2	1.2	3.4
Fama-French Three-Factor Model	3.6	1.2	2.0	2.0	0.8	2.8

This table presents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level. We, however, employ 25 size-BM reference portfolios to measure the BHARs and CTARs.

Table 1A: Specification of Tests in Random Samples

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardization Only Approach	1.3	8.3*	2.0	3.2	3.4	2.8
Weighting Only Approach	5.6*	0.4	2.0	3.0	1.1	4.2*
Panel B: Value-Weighted Portfolios						
Standardization Only Approach	0.2	4.0*	1.6	2.7	4.1*	2.3
Weighting Only Approach	3.8*	2.6	1.2	3.1	0.3	5.8*

This table reports the results from Standardization Only Approach and Weighting Only Approach. The Standardization Only Approach does not weight the monthly portfolios, while the Weighting Only Approach does not standardize the event firm anomalies. The numbers shown in this table present the percentages 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table 2A: Specification of Tests in Samples with Large Firms

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	0.4	4.8*	0.8	2.0	1.6	2.0
Buy-and-Hold Abnormal Return Method	0.0	16.8*	0.0	18.8*	0.0	22.0*
Fama-French Three-Factor Model	0.8	4.0*	1.6	2.8	2.8	2.0
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	0.0	4.0*	0.8	2.4	1.6	3.4
Fama-French Three-Factor Model	0.8	4.8*	0.8	2.8	2.8	2.4

This table presents the percentages of 1000 samples of 200 large firms that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table 2B: Specification of Tests in Samples with Small Firms

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	4.0*	0.8	2.4	1.6	1.6	3.6
Buy-and-Hold Abnormal Return Method	24.6*	0.0	9.2*	0.0	9.2*	0.0
Fama-French Three-Factor Model	6.2*	0.8	4.4*	1.6	2.4	2.4
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	4.6*	2.4	3.2	2.4	0.4	2.8
Fama-French Three-Factor Model	4.0*	1.8	2.8	2.4	0.8	4.4*

This table presents the percentages of 1000 samples of 200 small firms that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table 3A: Specification of Tests in Samples of firms with High BM Value

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	2.0	3.2	1.6	2.0	0.0	6.0*
Buy-and-Hold Abnormal Return Method	0.0	32.2*	0.0	29.2*	0.1	17.8*
Fama-French Three-Factor Model	0.4	7.2*	0.2	5.6*	0.8	3.6
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	1.2	3.2	1.2	2.4	0.0	4.0*
Fama-French Three-Factor Model	0.0	10.1*	0.4	5.4*	0.8	5.6*

This table presents the percentages of 1000 samples of 200 firms with high BM value that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table 3B: Specification of Tests in Samples of firms with Low BM Value

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	2.8	2.4	1.6	0.4	2.0	0.4
Buy-and-Hold Abnormal Return Method	14.0*	0.0	10.4*	0.2	8.2*	0.0
Fama-French Three-Factor Model	14.0*	0.0	8.8*	0.4	6.8*	0.0
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	1.6	2.4	2.4	0.8	1.2	0.4
Fama-French Three-Factor Model	11.0*	0.0	6.0*	0.8	4.0*	0.0

This table presents the percentages of 1000 samples of 200 firms with low BM value that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table 4A: Specification of Tests in Samples with Good Pre-event Returns: Case I

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	0.4	9.2*	2.0	4.0*	1.2	4.8*
Buy-and-Hold Abnormal Return Method	0.0	6.8*	14.0*	0.2	21.4*	1.6
Fama-French Three-Factor Model	0.4	8.0*	3.2	2.8	3.2	1.2
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	0.0	12.0*	1.2	3.6	0.4	4.8*
Fama-French Three-Factor Model	0.4	10.0*	2.0	5.2*	2.4	4.6*

This table presents the percentages of 1000 samples of 200 firms with good pre-event returns that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level. We, however, employ 25 size-BM reference portfolios to measure the BHARs and CTARs.

Table 4B: Specification of Tests in Samples with Poor Pre-event Returns: Case I

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	14.0*	0.4	0.4	2.8	1.2	1.6
Buy-and-Hold Abnormal Return Method	21.2*	0.0	8.4*	0.1	4.6*	0.3
Fama-French Three-Factor Model	16.8*	0.0	2.6	4.2*	2.1	2.0
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	6.8*	0.4	0.4	4.0*	0.8	4.0*
Fama-French Three-Factor Model	9.6*	0.4	0.6	2.4	0.4	2.8

This table presents the percentages of 1000 samples of 200 firms with poor pre-event returns that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level. We, however, employ 25 size-BM reference portfolios to measure the BHARs and CTARs.

Table 4C: Specification of Tests in Samples with Good Pre-event Returns: Case II

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	0.3	4.6*	3.1	3.6	2.4	0.7
Buy-and-Hold Abnormal Return Method	0.2	4.2*	3.8*	2.3	3.0	1.1
Carhart Four-Factor Model	0.0	5.6*	2.3	2.4	0.9	3.1
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	2.8	4.0*	1.8	3.4	3.6	2.1
Carhart Four-Factor Model	0.4	3.8*	2.6	1.9	2.9	3.6

This table presents the percentages of 1000 samples of 200 firms with good pre-event returns that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level. However, the difference between this table and Table 4A is that here we employ $5 \times 5 \times 5$ size-BM-past returns reference portfolios to measure the BHARs and CTARs. We also estimate the four-factor model of Carhart (1997) to include the momentum factor computed as the difference between returns of winners and losers. We, therefore, exclude the three-factor model while constructing Table 4C.

Table 4D: Specification of Tests in Samples with Poor Pre-event Returns: Case II

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	3.7*	1.6	2.8	2.0	3.5	1.3
Buy-and-Hold Abnormal Return Method	4.8*	1.2	3.6	2.7	2.6	2.2
Carhart Four-Factor Model	4.4*	0.3	3.7*	1.9	2.3	1.6
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	4.1*	2.1	2.6	0.3	3.4	2.0
Carhart Four-Factor Model	3.8*	0.2	1.1	2.3	3.6	2.7

This table presents the percentages of 1000 samples of 200 firms with poor pre-event returns that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level. However, the difference between this table and Table 4B is that here we employ $5 \times 5 \times 5$ size-BM-past returns reference portfolios to measure the BHARs and CTARs. We also estimate the four-factor model of Carhart (1997) to include the momentum factor computed as the difference between returns of winners and losers. We, therefore, exclude the three-factor model while constructing Table 4D.

Table 5: Specification of Tests in Samples with Overlapping Returns

Method Description	1 Year		3 Years		5 Years	
	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios						
Standardized Calendar Time Approach	3.2	2.0	2.9	3.6	2.4	3.2
Buy-and-Hold Abnormal Return Method	7.2*	0.4	9.4*	3.6	10.2*	4.8*
Fama-French Three-Factor Model	1.2	1.6	2.3	2.4	2.8	2.1
Panel B: Value-Weighted Portfolios						
Standardized Calendar Time Approach	3.2	1.6	1.8	3.1	1.6	2.0
Fama-French Three-Factor Model	2.8	0.8	1.6	3.4	2.2	3.3

This table presents the percentages of 1000 samples of 200 firms with over-lapping returns that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods at the theoretical significance level of 5 percent in favor of the alternative hypothesis of a significantly negative intercept (i.e., calculated p value is less than 2.5 percent at the 5 percent significance level) or a significantly positive intercept (calculated p value is greater than 97.5 percent at the 5 percent significance level). The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level. Selecting these samples is a two-step procedure. The first step involves a random selection of 100 firms from the population. In the second step, for each of these 100 firms, we randomly choose a second event month that is within $H-1$ periods of the original event month (either before or after), where H equals 12, 36 or 60. Hence we have 200 firms with 200 event months where the same firm appears in the sample twice and this generates the issue of overlapping returns. We repeat this procedure 1000 times.

Table 6: Power of alternative Methods in Random Samples

Method Description	Induced Level of Abnormal Return (%) over 3 Years								
	−20	−15	−10	−5	0	5	10	15	20
Standardized Calendar Time Approach	1.00	0.88	0.56	0.11	0.05	0.18	0.71	0.92	1.00
Buy-and-Hold Abnormal Return Method	0.94	0.71	0.46	0.11	0.05	0.17	0.59	0.80	0.97
Fama-French Three-Factor Model	0.78	0.55	0.34	0.08	0.04	0.10	0.39	0.58	0.85

This table presents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over a three-year holding period. We add the levels of annual abnormal return indicated in the column heading. However, only equally weighted portfolios are considered to make a direct comparison with the BHAR approach.

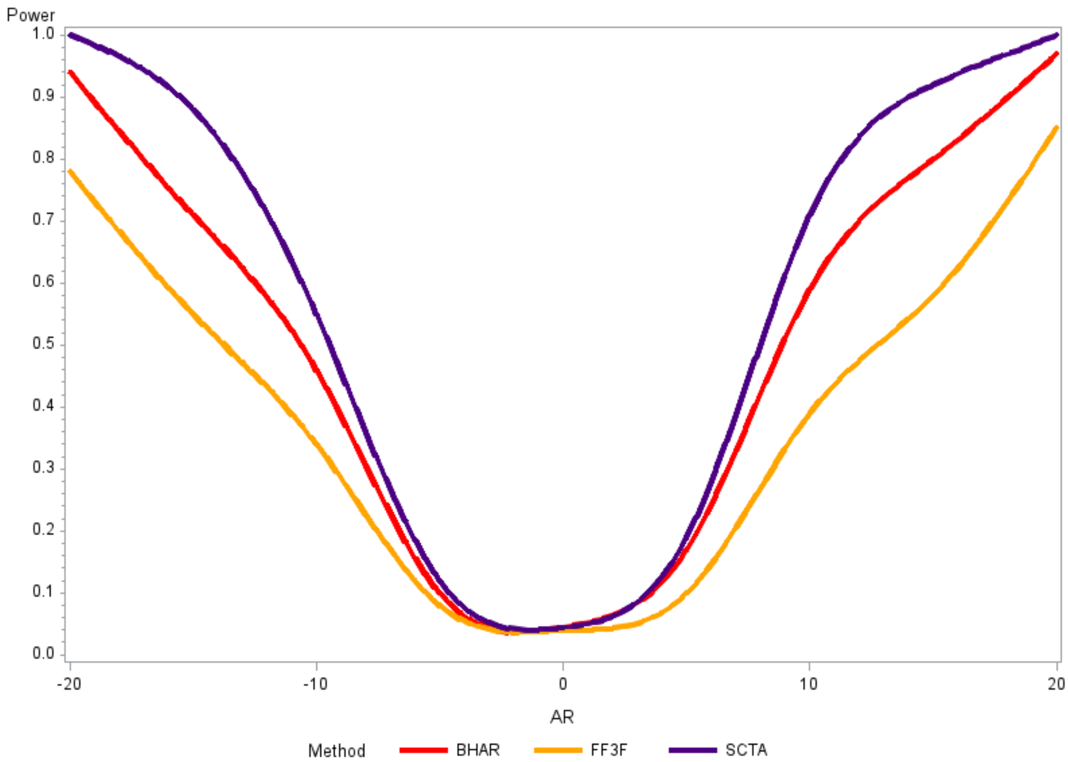

Figure 1: Simulated Power of Different Methods in Random Samples. This figure represents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over three-year holding period. We consider equally weighted portfolios to make a direct comparison with BHAR approach. The horizontal axis indicates the induced level of abnormal returns (%), while the rejection rates are shown in the vertical axis.

International Journal of Business and Management; Vol. 9, No. 9; 2014
ISSN 1833-3850 E-ISSN 1833-8119
Published by Canadian Center of Science and Education

Does Calendar Time Portfolio Approach Really Lack Power?

Anupam Dutta[1]

[1] Department of Mathematics & Statistics, University of Vaasa, Vaasa, Finland

Correspondence: Anupam Dutta, Department of Mathematics & Statistics, University of Vaasa, Vaasa, Finland. E-mail: adutta@uwasa.fi

Received: May 28, 2014 Accepted: July 14, 2014 Online Published: August 22, 2014

doi:10.5539/ijbm.v9n9p260 URL: http://dx.doi.org/10.5539/ijbm.v9n9p260

Abstract

This paper investigates whether the calendar time methodology lacks power in detecting the long-run abnormal performance of the firms after major corporate events. In addition, the study proposes a variant of calendar time approach by standardizing the abnormal returns of the event firms forming the monthly portfolios. To assess the robustness of the modified method, the results from buy-and-hold abnormal return approach and the mean monthly calendar time abnormal return method are also reported. The empirical analysis documents that the proposed approach improves the power in random samples and in samples with small firms and with calendar clustering.

Keywords: event study, long-run anomalies, standardized abnormal returns, specification issue, power issue

1. Introduction

A large number of recent studies examine the price behavior of equity for periods of one to five years following significant corporate events (e.g., IPOs, SEOs, repurchases, or bond rating changes). Although there have been many advances in long-run event study methodology over the years, the elementary papers in this area include Ritter (1991), Barber and Lyon (1997), Kothari and Warner (1997), Fama (1998), Lyon, Barber, and Tsai (1999), Mitchell and Stafford (2000), Loughran and Ritter (2000), and Jegadeesh and Karceski (2009).

While investigating the long-term abnormal returns of the event firms, two important issues are taken into account. These are the power issue and the specification issue. Prior literature suggests that these two issues still remain unsolved and further filtering of the existing methods is required for solving such issues. Kothari and Warner (2007), for instance, conclude that whether calendar time portfolio (CTP) method or buy-and-hold abnormal return (BHAR) approach can best address these long-horizon issues remains an open question.

However, although each of these two issues is important in inspecting the long-run abnormal performance, only a few studies focus on the power issue. Ang and Zhang (2004), for instance, is the solo study in the literature that extensively reports the power of several empirical procedures at three different investment horizons. In order to extend this limited literature, the present study makes a modest attempt to compare the power of alternative methodologies. In doing so, we analyze the power of buy-and-hold abnormal return approach and the mean monthly calendar time abnormal return (CTAR) methodology for random samples and samples with small firms and with calendar clustering.

Loughran and Ritter (2000), however, criticize the calendar time methodology claiming that it has low power. Ang and Zhang (2004) also document that the power of calendar time portfolio approach decreases as the holding period increases. Mitchell and Stafford (2000), on the other hand, find no evidence that supports the concern raised by Loughran and Ritter (2000). This disagreement gives us the motivation to investigate whether the CTP approach really lacks power in detecting the long-run anomalies.

The empirical findings reveal that the mean monthly calendar time abnormal return method has low power than the buy-and-hold abnormal return approach. Although each of these methodologies is well-specified in random samples and in samples with small firms and with calendar clustering, the CTAR approach lacks power in detecting the long-term anomalies. The study, therefor, proposes to refine the mean monthly calendar time abnormal return methodology by considering the standardized abnormal returns of the event firms forming the monthly portfolios. In our modified approach, the monthly portfolios are also weighted in such a way that periods of heavy event activity receives more loadings than the periods of low event activity. Our analysis further shows that the refined calendar time approach produces well-specified test statistics andimproves the power in all the sampling schemes under consideration.However, one potential limitation with our proposed approach is that it is, like other existing long-run event study methodologies, is not well specified in all types of nonrandom samples.

The remainder of the paper is organized as follows. Section 2 reviews the existing literature. Section 3 outlines the data and methodology. Results are discussed in Section 4, and Section 5 concludes the paper.

2. Literature Review

Previous research documents that the buy-and-hold abnormal return methodology and the calendar time portfolio approach are commonly employed for examining the long-term abnormal stock returns. Barber and Lyon (1997) and Lyon et al. (1999) claim that the BHARs most accurately capture investor experience. Fama (1998), however, argues against the BHAR methodology as it experiences the bad model problems. Fama and later Mitchell and Stafford (2000)strongly recommend the use of calendar time methodology to deal with the bad model problems.

Barber and Lyon (1997) and Lyon et al. (1999) also identify new listing, re-balancing, and skewness biases with inference in long-run event studies using the BHAR. They use simulations to investigate the impact of these biases on inference when BHAR is exercised to measure the abnormal performance and standard tests are applied. However, in case of using a reference portfolio to capture normal or expected return, the new listing and rebalancing biases can be addressed in a relatively simple way by careful construction of the reference portfolio. Unfortunately, considering reference portfolio also gives rise to the skewness bias.

To avoid thisskewness bias, a control firm rather than a reference portfolio can be used as the long-run return benchmark. The BHAR is then measured as the difference between the long-run holding-period returns of the event firm's equity and that of a control firm. Although the distribution of each asset's holding-period return is highly skewed, the distribution of their difference is not. As a result, standard statistical tests based on the control firm approach have the right size in random samples.

Barber and Lyon (1997), however, report that standard tests based on the control firm approach are not as powerful as those based on the reference portfolio approach. This is because of the fact that the use of a control firm is a noisier way to control for expected returns than is the use of a reference portfolio and this added noise reduces the power of the test. The variance of the difference between the returns on two individual assets is generally much higher than the variance of the difference between the return of an asset and that of a portfolio, even when the control firm is chosen carefully. Powerful tests thus require very large samples when control firm approach is applied.

To resolve these problems that the BHAR methodology encounters, the calendar time portfolio approach is considered as a possible alternative. Unfortunately, Loughran and Ritter (2000) argue that the CTP approach lacks power while identifying the abnormal performance.Ang and Zhang (2004) also report the same in their simulation study. In this paper, we, therefore, make an attempt to modify the conventional calendar time approach such that its power improves. In the following section, we discuss our proposed methodology. Reviewing the conventional approaches will follow.

3. Data and Methodology

The data employed in this paper comprise NYSE, Amex, and Nasdaq stocks, and our sample period ranges from July 1980 to December 2012. We obtain monthly stock prices, market value (MV) or size and book-to-market (BM) value data from DataStream.

In this study, we consider a size-BM-matched control firm to calculate the abnormal returns. Identifying such a control firm is a 2-step procedure. First, we identify all the firms with a market value of equity between 70% and 130% of the sample firm at the most recent end of June. Then from this set of firms, we choose the firm with BM closest to that of the sample firm as of the previous December. We do not use reference portfolios as test statistics based on buy-and-hold abnormal return calculated employing a reference portfolio approach are generally misspecified.

3.1 Mean Monthly Calendar Time Abnormal Return (CTAR)

The calculation of mean monthly calendar time abnormal return (CTAR) is the following:

$$mean\ monthly\ CTAR = \frac{1}{T}\sum_{t=1}^{T} CTAR_t, \tag{1}$$

where

$$CTAR_t = R_{pt} - E(R_{pt}) \tag{2}$$

Within this framework, R_{pt} is the monthly return on the portfolio of event firms, $E(R_{pt})$ is the expected return on the event portfolio which is proxied by the raw return on a control firm and T is the total number of months in the sample period. To test the null hypothesis that there is no abnormal performance, the *t*-statistic of the mean monthly CTAR is calculated using the intertemporal standard deviation of the monthly CTARs defined in equation (2).

3.2 Standardized Calendar Time Approach (SCTA)

In our proposed approach, the monthly portfolios are constructed in an alternative way. We consider standardized abnormal returns to compute the monthly CTARs. Using standardized returns is advantageous in the sense that many firms, especially small firms, often produce volatile returns and consequently, the distributions of long-run returns tend to have fat tails. One possible solution to this problem is standardizing the abnormal returns by their volatility measures. However, to reduce the skewness problem, we consider using log returns.

Now the formation of the monthly portfolios involves two steps. We first calculate the standardized abnormal returns for each of the sample firms. In doing so, the abnormal returns for firm *i* are computed as $\varepsilon_{it} = R_{it} - E(R_{it}); t = 1, \dots, H$, where R_{it} denotes the return on event firm *i* in the calendar month *t* and $E(R_{it})$ is the expected return which is proxied by the raw return on a control firm and *H* is the holding period which equals 12, 24 or 36 months. The next task is to estimate the event-portfolio residual variances using the H-month residuals computed as monthly differences of i-th event firm returns and control firm returns. Dividing ε_{it} by the estimate of its standard deviation yields the corresponding standardized abnormal return, say, z_{it}, for event firm*i* in month*t*. Now let $N_t$refer to the number of event firms in the calendar month *t*.We then calculate the calendar time abnormal return for portfolio *t* as:

$$CTAR_t = \frac{1}{N_t}\sum_{t=1}^{N_t} z_{it} \quad (3)$$

We also propose to weight each of the monthly CTARs by$\sqrt{N_t}$. This weighting scheme is lucrative as it gives more loadings to periods of heavy event activity than the periods of low event activity. However, Loughran and Ritter (2000) argue that when a small number of firms include a large proportion of a value-weighted portfolio, unsystematic risk is not diversified away. In this paper, we, therefore, consider only equally-weighted portfolios to estimate the abnormal returns. The grand mean monthly abnormal return, denoted by$\overline{CTAR}$, is then calculated as:

$$\overline{CTAR} = \frac{1}{T}\sum_{1}^{T} CTAR_t \quad (4)$$

While finding$\overline{CTAR}$, it might be the case that a number of portfolios do not contain any event firm. In such situations, those months are dropped from the analysis. To test the null hypothesis of no abnormal performance, the *t*-statistic of $\overline{CTAR}$is computed by using the intertemporal standard deviation of the monthly CTARs defined in equation (3).

3.3 Buy-and-Hold Abnormal Return (BHAR)

An*H*-month BHAR for event firm *i*is defined as:

$$BHAR_{iH} = \prod_{t=1}^{H}(1 + R_{it}) - \prod_{t=1}^{H}(1 + R_{Bt}), \quad (5)$$

whereR_{it} denotes the return on event firm *i* at time *t* and R_{Bt} indicates the return on a control firm.

To test the null hypothesis that the mean buy-and-hold return equals zero, the conventional*t*-statistic is given by:

$$t_{BHAR} = \frac{\overline{BHAR_H}}{\sigma(BHAR_H)/\sqrt{n}} \quad (6)$$

where $\overline{BHAR_H}$ implies the sample mean and $\sigma(BHAR_H)$ refers to the cross-sectional sample standard deviation of abnormal returns for the sample containing *n* firms.

Table 1 summarizes the test statistics of different methods used in our study to investigate the long-term abnormal performance. We use size-BM matched control firm to measure the anomalies. The standard error shown in the numerator is the traditional standard error computed using the CTARs defined in equation (2) and equation (3).

Table 1. Summary of alternative methodologies

Method Description	Test Statistics
Standardized Calendar Time Approach (SCTA)	$t = \frac{\overline{CTAR}}{Standard\ Error}$
Buy-and-Hold Abnormal Return (BHAR)	$t_{BHAR} = \frac{\overline{BHAR_H}}{\sigma(BHAR_H)/\sqrt{n}}$
Mean Monthly Calendar Time Abnormal Return (CTAR)	$t = \frac{Mean\ Monthly\ \overline{CTAR}}{Standard\ Error}$

Note. This table summarizes the test statistics of different empirical procedures employed in this study. The standard error is the conventional standard error computed using the CTARs defined in equation (2) and equation (3).

3.4 Simulation Method

To assess the specification of the employed methodologies, we randomly select 1000 samples of 200 event months without replacement. For each of these 200 event months, we randomly draw one stock from the population of all stocks that are active in the database for that month. For a well-specified test statistic, 1000α tests reject the null hypothesis. A test is conservative if fewer than 1000α null hypotheses are rejected and is anticonservative if more than 1000α null hypotheses are rejected. Based on this procedure, we test the specification of the *t*-statistic at 5% theoretical levels of significance. A well-specified null hypothesis rejects the null at the theoretical rejection level in favor of the alternative hypothesis of negative (positive) abnormal returns in $1000\alpha/2$ samples.

4. Empirical Results

This section reports the specification and power of various methodologies used in our study. We first discuss the specification issue. Later, we focus on the power issue. In addition to random samples, two types of nonrandom samples based on small firms and calendar clustering are also considered in our analysis.

4.1 Specification of Tests

Table 2 indicates the rejection rates in 1000 simulations with a sample of 200 firms. Panels A, B and C present the size of alternative tests for random samples and samples with small firms and with calendar clustering respectively. The simulated results reveal that all the empirical methods considered in our analysis are well-specified in random samples as well as in nonrandom samples. For example, Panel A suggests that when the holding period is one year, the rejection rates at the 5% level of significance are 2.4% and 2.8% for SCTA, 2.8% and 0.4% for BHAR, and 2.7% and 1.3% for CTAR. Panel B, on the other hand, indicates that with a five-year investment horizon, the rejection rates at the 5% level of significance are 3.5% and 1.6% for SCTA, 3.6% and 0.7% for BHAR, and 2.4% and 1.2% for CTAR. Panel C reveals that with a three-year holding period the rejection rates at the 5% level of significance are 3.6% and 2.2% for SCTA, 2.7% and 2.1% for BHAR, and 2.1% and 2.8% for CTAR.

Table 2. Specification of tests in random samples and samples with small firms and with calendar clustering

	Holding Period					
	12 Months		36 Months		60 Months	
	Theoretical Cumulative Density Function (%)					
Method	2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Random Samples						
SCTA	2.4	2.8	1.6	2.8	2.0	3.6
BHAR	2.8	0.4	3.2	1.2	2.6	2.3
CTAR	2.7	1.3	2.4	2.4	1.1	1.6
Panel B: Samples with Small Firms						
SCTA	3.2	0.9	2.6	2.1	3.5	1.6
BHAR	2.4	3.6	3.2	2.8	3.6	0.7
CTAR	1.1	0.3	2.9	1.2	2.4	1.2
Panel C: Samples with Calendar Clustering						
SCTA	1.9	2.7	3.6	2.2	2.0	3.1
BHAR	3.4	0.8	2.7	2.1	3.6	1.1
CTAR	3.1	0.9	2.1	2.8	3.2	2.0

Note: This table presents the percentages of 1000 samples of 200 firms that reject the null hypothesis of no abnormal returns over one-year, three-year, and five-year holding periods. Panel A shows the results for random samples, while Panel B and Panel C indicate the findings for samples with small firms and with calendar clustering.

4.2 Power

Tables 3–5 show the power of alternative methodologies in random samples and in samples with small firms and with calendar clustering respectively. Since the methods we employ in this paper are not, in general, well-specified in samples based on book-to-market-ratio or large firms, we do not report the power of different approaches using these nonrandom samples. To examine the power of the employed methods, we introduce a constant level of abnormal return ranging from -20% to 20% at an interval of 5% to event firms. Tables 3–5 indicate the percentages of 1000 samples of 200 firms that reject the null hypothesis of zero abnormal returns over one, three and five-year holding periods.

It is evident from Table 3 that in case of detecting the anomalies, the calendar time methodologies have more power than the BHAR approach does. For example, with +10% (-10%) per year abnormal returns and with a three-year holding period, the rejection rate is 25% (16%) for SCTA, 23% (16%) for BHAR, and 17% (13%) for CTAR. Table 4, however, indicates that with +15% (-15%) per year abnormal returns and with a five-year investment horizon, the rejection rate is 30% (26%) for SCTA, 29% (25%) for BHAR, and 21% (17%) for CTAR. Finally, Table 5 reveals that with +20% (-20%) per year abnormal returns and with a three-year holding period, the rejection rate is 49% (41%) for SCTA, 43% (39%) for BHAR, and 41% (36%) for CTAR.

Table 3. Power of Alternative methodologies in random samples

	Induced Level of Abnormal Returns (%)								
Methods	-20	-15	-10	-5	0	5	10	15	20
Panel A: One Year Holding Period									
SCTA	0.98	0.81	0.51	0.19	0.05	0.27	0.62	0.87	1.0
BHAR	0.96	0.79	0.46	0.17	0.03	0.21	0.55	0.83	0.99
CTAR	0.81	0.67	0.43	0.12	0.04	0.18	0.52	0.74	0.91
Panel B: Three Years Holding Period									
SCTA	0.54	0.35	0.16	0.08	0.04	0.12	0.25	0.44	0.63
BHAR	0.51	0.34	0.16	0.08	0.04	0.10	0.23	0.39	0.69
CTAR	0.47	0.31	0.13	0.07	0.05	0.07	0.17	0.37	0.63
Panel C: Five Years Holding Period									
SCTA	0.36	0.19	0.08	0.06	0.05	0.07	0.12	0.26	0.41
BHAR	0.33	0.17	0.09	0.06	0.05	0.08	0.12	0.24	0.40
CTAR	0.32	0.14	0.06	0.04	0.02	0.06	0.09	0.21	0.36

Note: This table documents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal performance over one-year (Panel A), three-year (Panel B), and five-year (Panel C) holding periods. We add the levels of annual abnormal return indicated in the column heading.

Table 4. Power of alternative methodologies in samples with small firms

Induced Level of Abnormal Returns (%)									
Methods	-20	-15	-10	-5	0	5	10	15	20
Panel A: One Year Holding Period									
SCTA	0.96	0.77	0.56	0.19	0.04	0.21	0.63	0.69	0.88
BHAR	0.91	0.74	0.54	0.18	0.06	0.15	0.59	0.66	0.82
CTAR	0.63	0.45	0.31	0.12	0.02	0.06	0.26	0.38	0.58
Panel B: Three Years Holding Period									
SCTA	0.61	0.40	0.22	0.11	0.06	0.14	0.29	0.49	0.71
BHAR	0.57	0.38	0.19	0.10	0.05	0.13	0.31	0.47	0.68
CTAR	0.42	0.29	0.12	0.07	0.05	0.08	0.21	0.34	0.50
Panel C: Five Years Holding Period									
SCTA	0.52	0.26	0.09	0.07	0.06	0.08	0.11	0.30	0.56
BHAR	0.48	0.25	0.10	0.08	0.07	0.07	0.11	0.29	0.54
CTAR	0.29	0.17	0.07	0.05	0.04	0.08	0.09	0.21	0.38

Note. This table shows the percentages of 1000 samples of 200 small firms that reject the null hypothesis of no abnormal performance over one-year (Panel A), three-year (Panel B), and five-year (Panel C) holding periods. We add the levels of annual abnormal return indicated in the column heading.

Table 5. Power of alternative methodologies in samples with calendar clustering

Induced Level of Abnormal Returns (%)									
Methods	-20	-15	-10	-5	0	5	10	15	20
Panel A: One Year Holding Period									
SCTA	0.73	0.52	0.31	0.12	0.05	0.17	0.39	0.61	0.82
BHAR	0.68	0.47	0.29	0.11	0.04	0.14	0.38	0.58	0.82
CTAR	0.61	0.41	0.25	0.08	0.04	0.11	0.31	0.49	0.68
Panel B: Three Years Holding Period									
SCTA	0.67	0.43	0.24	0.09	0.06	0.11	0.31	0.51	0.76
BHAR	0.59	0.36	0.21	0.06	0.04	0.10	0.32	0.49	0.72
CTAR	0.54	0.32	0.18	0.06	0.05	0.08	0.19	0.37	0.52
Panel C: Five Years Holding Period									
SCTA	0.41	0.28	0.10	0.06	0.05	0.08	0.14	0.32	0.49
BHAR	0.39	0.26	0.10	0.06	0.05	0.07	0.11	0.30	0.43
CTAR	0.36	0.24	0.08	0.06	0.04	0.07	0.11	0.31	0.41

Note. This table documents the percentages of 1000 samples of 200 firms with calendar clustering that reject the null hypothesis of no abnormal performance over one-year (Panel A), three-year (Panel B), and five-year (Panel C) holding periods. We add the levels of annual abnormal return indicated in the column heading.

These findings suggest three important implications. First, with the increase in the investment period, the power starts decreasing for all the methods employed in this study. Ang and Zhang (2004), however, also report the same. Second, the mean monthly calendar time abnormal return has low power to detect the long-term anomalies. Third, our proposed standardized calendar time approach improves the power in all the sampling schemes under study.

5. Conclusion

Although long-run event studies have an extensive history, the power issue generally receives less attention than the specification issue. To conceal this research gap, the present study aims to compare the power of alternative methodologies under different sampling schemes. In doing so, this paper extends the prior literature in two aspects. First, the study inspects whether the calendar time methodology really lacks power in detecting the long-term abnormal performance of the firms following major corporate events. To serve this purpose, we compare the results of buy-and-hold abnormal return approach and the mean monthly calendar time abnormal return methodology. The empirical analysis indicates that although each of these methodologies is well-specified in random samples as well as in nonrandom samples, the mean monthly calendar time abnormal return method lacks power in detecting the anomalies. Second, we propose to refine the mean monthly calendar time approach by forming the monthly portfolios in a variant way and further analysis shows that our proposed calendar time approach, with only a few exceptions, improves the power in random samples and in samples with small firms and with calendar clustering.

References

Ang, S., & Zhang, S. (2004). An evaluation of testing procedures for long horizon event studies. *Review of Quantitative Finance and Accounting*, *23*(3), 251–274. http://dx.doi.org/10.1023/B:REQU.0000042344.27369.0d

Barber, B., & Lyon, J. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, *43*(3), 341–372. http://dx.doi.org/10.1016/S0304-405X(96)00890-2

Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, *49*(3), 283–306. http://dx.doi.org/10.1016/S0304-405X(98)00026-9

Jegadeesh, N., & Karceski, J. (2009). Long-run performance evaluation: correlation and hetero skedasticity-consistent tests. *Journal of Empirical Finance*, 16(1), 101–111. http://dx.doi.org/10.1016/j.jempfin.2008.06.002

Kothari, S. P., & Warner, J. B. (2007). Econometrics of event studies. In Eckbo, E. (Ed.), *Handbook of Corporate Finance: Empirical Corporate Finance* (vol. 1, pp. 3–36). Elsevier/North-Holland, Amsterdam.

Kothari, S., & Warner, J. B. (1997). Measuring long-horizon security price performance. *Journal of Financial Economics, 43*(3), 301–339. http://dx.doi.org/10.1016/S0304-405X(96)00899-9

Loughran, T., & Ritter, J. (2000). Market efficiency, uniformly least powerful tests of market efficiency. *Journal of Financial Economics*, *55*(3), 361–389. http://dx.doi.org/10.1016/S0304-405X(99)00054-9

Lyon, J., Barber, B., & Tsai, C. (1999). Improved methods of tests of long-horizon abnormal stock returns. *Journal of Finance*, *54*(1), 165–201. http://dx.doi.org/10.1111/0022-1082.00101

Mitchell, M., & Stafford, E. (2000). Managerial decisions and long-term stock price performance.*Journal of Business*, *73*(3), 287–329. http://dx.doi.org/10.1086/209645

Ritter, J. (1991). The long-term performance of initial public offerings. *Journal of Finance*, *46*(1), 3–27. http://dx.doi.org/10.1111/j.1540-6261.1991.tb03743.x

Corporate Ownership & Control / Volume 12, Issue 1, 2014, Continued - 2

INVESTIGATING LONG-RUN STOCK RETURNS AFTER CORPORATE EVENTS: THE UK EVIDENCE

Anupam Dutta*

Abstract

The objective of this paper is to assess the robustness of the existing long-run event study methodologies in the UK stock market. In doing so, the study employs the buy-and-hold abnormal return approach and the calendar time portfolio method to identify the long-term abnormal performance following corporate events. Although many recent studies consider the application of these two widely used approaches, each of the methods is a subject to criticisms. This paper uses the standardized calendar time approach (SCTA) which presents a number of important improvements over the traditional calendar time methodology. The empirical analysis reveals that all the traditional methodologies perform well in the UK security market. Our findings further report that SCTA documents better specification and power than the conventional approaches.

Keywords: Long-run Anomalies, Standardized Abnormal Returns, Specification Issue, Power Issue

*Department of Mathematics and Statistics, University of Vaasa, Wolffintie 34, Vaasa-65200, Finland
Tel: +358458690502
Email: adutta@uwasa.fi

1 Introduction

Many recent studies investigate the long-term performance of firms after certain corporate events such as IPOs, SEOs, or repurchases. The key articles in this area include Ritter (1991), Barber and Lyon (1997), Kothari and Warner (1997), Fama (1998), Lyon, Barber, and Tsai (1999), Mitchell and Stafford (2000), Loughran and Ritter (2000), Boehme and Sorescu (2002), and Jegadeesh and Karceski (2009). While long-horizon event studies have enjoyed many advances over the years, all the elementary papers focus on United States security markets. Although a number of studies investigate the long-term performance of the UK IPOs (Levis, M., 1993, Espenlaub, Gregory and Tonks, 2000 etc.), not a single simulation study concentrating on this security market is found in the literature. Therefore, the objective of this paper is to conceal such gaps by conducting a simulation study with the UK stock market data. Since the choice of proper methodology plays a key role in investigating the long-run performance, we, like other fundamental studies, employ Buy-and-Hold Abnormal Return (BHAR) approach and Calendar Time Portfolio (CTP) method to measure the long-run anomalies.

However, previous studies document that each of these widely used methods has a number of potential pitfalls. Fama (1998), for example, reports that the BHAR method ignores the issue of potential cross-sectional correlation of event-firm abnormal returns and hence produces misspecified test statistics. Loughran and Ritter (2000), on the other hand, claim that CTP approach has low power to identify the abnormal performance because it gives equal weight to each month, regardless of whether the month has heavy or light event activities. Following the work of Dutta (2014), this paper considers applying the Standardized Calendar Time Approach (SCTA) where we first standardize the abnormal returns for each of the event firms forming the monthly portfolios and then each portfolio is weighted such that periods of heavy event activity receive more weight than periods of low event activity. However, employing standardized abnormal returns is well-documented in the literature. For example, Jaffe (1974) and Mandelker (1974) employ standardized portfolio returns for investigating the long-run abnormal performance. Fama (1998) also suggests to standardize the abnormal returns to resolve the issues raised by Loughran and Ritter (2000). Later, Mitchell and Stafford (2000) use standardized abnormal returns to alleviate the heteroscedasticity problem that often occurs in CTP approach due to the varying portfolio construction.

The empirical analysis reveals that all these methodologies are robust in the UK stock market as well. Our findings also report that the standardized calendar time approach produces reasonably well-specified test statistics in all types of nonrandom samples. The results further show that SCTA documents better power than the existing approaches to identify the long-term abnormal performance. One striking output of our analysis is that the test statistics based on the Fama-French three-factor model are not well specified even in random samples. However, the simulated result shows that the adjusted three-factor model controls well for size and book-to-market ratio biases. The BHAR approach, on the other hand, yields reasonably well-specified test statistics when the control firm approach is employed. While using a reference portfolio as a benchmark, the BHAR methodology does not produce well-specified test statistics. These results are consistent with those reported by Lyon, Barber, and Tsai (1999).

This paper extends the prior literature in three aspects. First, it uses simulated results to assess the performance of the existing long-run event study methodologies using the UK security market data. Second, the study employs a variant of calendar time methodology which yields well-specified test statistics in nonrandom samples. Third, this refined calendar time approach improves the power while inspecting the long-term abnormal performance. However, one major limitation of our proposed approach is that it does not yield well-specified test statistics (not reported in the table) for samples based on pre-event return performance. Lyon, Barber, and Tsai (1999) also report that the buy-and-hold abnormal return approach as well as the traditional calendar time method have this limitation. Therefore, further filtering of the existing methodologies is needed to resolve this problem. Another important drawback of our study is that it does not present the results based on industry-clustered samples due to the non-availability of industry codes.

The remainder of the paper is structured as follows. Section 2 outlines the data and methodology. Section 3 explains the simulation procedure. Section 4 discusses the specification of the tests. Section 5 reports power of the tests and Section 6 concludes the paper.

2 Data and methodology

We obtain stock prices, market value (MV) or size and book-to-market (BM) value data of the UK stock market from DataStream. The sample period ranges from July 1983 to December 2013.

In this paper, we construct 25 size-BM portfolios as expected return benchmarks. In doing so, at the end of June of year t, firms are quantiled into five groups on the basis of their market values. Firms are further quantiled into five groups based on their book-to-market ratios. However, we also consider a size-BM-matched control firm to calculate the abnormal returns. Identifying this control firm is a 2-step procedure. First, we identify all the firms with a market value of equity between 70% and 130% of the sample firm at the most recent end of June. Then from this set of firms, we choose the firm with BM closest to that of the sample firm as of the previous December.

2.1 Standardized calendar time approach (SCTA)

The calculation of mean monthly calendar time abnormal return (CTAR) is the following:

$$MMCTAR = \frac{1}{T}\sum_{t=1}^{T} CTAR_t \quad (1)$$

Where

$$CTAR_t = R_{pt} - E(R_{pt}) \quad (2)$$

Within this framework, R_{pt} is the monthly return on the portfolio of event firms, $E(R_{pt})$ is the expected return on the event portfolio which is proxied by the raw return on either a reference portfolio or a control firm and T is the total number of months in the sample period.

Following the work of Dutta (2014), this paper uses standardized abnormal returns to compute the monthly CTARs. Dutta argues that since a number of firms in the sample might produce volatile returns, it would cause the distributions of long-run returns to have fat tails. Consequently, test statistics will be seriously misspecified. But standardizing the abnormal returns by their volatility measures is a possible solution to this problem. Although Dutta uses simple return, we consider log return to minimize the skewness problem. Bessembinder and Zhang (2013) and Knif, Kolari and Pynnönen (2013) also document that employing log returns produces better specified test statistics.

The construction of the monthly portfolios in the standardized calendar time approach consists of two steps. We first calculate the standardized abnormal returns for each of the sample firms. In doing so, the abnormal returns for firm i are computed as $\varepsilon_{it} = r_{it} - E(r_{it}); t = 1, \ldots, H$, where r_{it} denotes the log return on event firm i in the calendar month t and $E(r_{it})$ is the expected return which is proxied by the raw return either 25 size-BM reference portfolios or a size-BM matched control firm and H is the holding period which equals 12, 24 or 36 months. The next task is to estimate the event-portfolio residual variances using the H-month residuals computed as monthly differences of i-th event firm returns and control firm returns. Dividing ε_{it} by the estimate of its standard deviation yields the corresponding standardized abnormal return, say, z_{it}, for event firm i in month t. Now let N_t refer to the number of event firms in the calendar month t. We then calculate the calendar time abnormal return for portfolio t as:

$$CTAR_t = \sum_{t=1}^{N_t} x_{it} z_{it} \quad (3)$$

The weight x_{it} equals $\frac{1}{N_t}$ when the abnormal returns are equally-weighted and $\frac{MV_{it}}{\sum MV_{it}}$ when the abnormal returns are value-weighted by size.

We, like Dutta (2014), also assign weights to each of the monthly CTARs by $1/\sqrt{(\sum_{t=1}^{T} x_{it}^2)}$. For instance, when the abnormal returns are equally weighted i.e., when $x_{it} = \frac{1}{N_t}$, then $1/\sqrt{(\sum_{t=1}^{T} x_{it}^2)} = \sqrt{N_t}$. This weighting scheme is lucrative as it gives more loadings to periods of heavy event activity than the periods of low event activity. Now the grand mean monthly abnormal return, denoted by $\overline{CTAR}$, is calculated as:

$$\overline{CTAR} = \frac{1}{T} \sum_{1}^{T} CTAR_t \quad (4)$$

While finding $\overline{CTAR}$, it might be the case that a number of portfolios do not contain any event firm. In such situations, those months are dropped from the analysis. To test the null hypothesis of no abnormal performance, the *t*-statistic of $\overline{CTAR}$ is computed by using the intertemporal standard deviation of the monthly CTARs defined in equation (3).

2.2 Buy-and-hold abnormal return (BHAR)

An *H*-month BHAR for event firm *i* is defined as:

$$BHAR_{iH} = \prod_{t=1}^{H}(1 + R_{it}) - \prod_{t=1}^{H}(1 + R_{Bt}) \quad (5)$$

R_{it} denotes the return on event firm i at time t and R_{Bt} indicates the return on a control firm.

To test the null hypothesis that the mean buy-and-hold return equals zero, the conventional *t*-statistic is given by:

$$t_{BHAR} = \frac{\overline{BHAR_H}}{\sigma(BHAR_H)/\sqrt{n}} \quad (6)$$

$\overline{BHAR_H}$ implies the sample mean and $\sigma(BHAR_H)$ refers to the cross-sectional sample standard deviation of abnormal returns for the sample containing *n* firms.

However, the earlier studies such as Mitchell and Stafford (2000), Boehme and Sorescu (2002), Jegadeesh and Karceski (2009) report that the BHAR approach does not control well for the cross-sectional correlation among individual firms in nonrandom samples and thus yields misspecified *t*-statistics. Moreover, the test statistics based on BHARs also have this misspecification problem, since the distribution of BHARs is highly skewed. Though bootstrapping corrects for the skewness problem to some extent, it ignores the cross-sectional dependence of abnormal returns.

2.3 Fama-French three-factor model

For each calendar month *t*, we form portfolios consisting of all sample firms that have participated in the event within the last *H* months, where *H* equals 12, 36, or 60 in our study. For each calendar month, the portfolios are rebalanced, i.e., the firms that reach the end of their *H* -month period drop out and new firms that have just executed a transaction are added. We then calculate the portfolio mean monthly abnormal return α_p by regressing its excess return on the three Fama-French factors:

$$R_{pt} - R_{ft} = \alpha_p + \beta_p(R_{mt} - R_{ft}) + s_p SMB_t + h_p HML_t + e_{pt} \quad (7)$$

R_{pt} is the equal or value-weighted return on portfolio *t*, R_{ft} is the risk-free rate, $(R_{mt} - R_{ft})$ is the excess return of the market, *SMB* is the difference between the return on the portfolio of small stocks and big stocks, *HML* is the difference between the return on the portfolio of high and low book-to-market stocks, α_p measures the mean monthly abnormal return of the calendar time portfolio which is zero under the null hypothesis of no abnormal performance and β_p, s_p and $h_p$are sensitivities of the event portfolio to the three factors.

However, since the number of firms changes over the sample period, this may cause the error term to be heteroskedastic and hence the ordinary least squares estimate becomes inefficient. Fama (1998), therefore, suggests to apply the weighted least squares technique instead of ordinary least squares to control for heteroskedasticity. In this study, we estimate regression (7) using weightedleast squares (WLS) procedures. Monthly returns in the WLS model are weighted by $\sqrt{N_t}$, where N_t stands for the number of event firms in month *t*.

2.4 Adjusted Fama-French three-factor model

Fama and French (1993) document that the traditional three-factor model is not able to completely explain the cross section of stock returns. However, Mitchell and Stafford (2000) and later Boehme and Sorescu (2002) refine this three-factor model to deal with the bad model problem. In this paper, we also try to modify the conventional Fama-French three-factor model to moderate the size and book-to-market ratio biases. Our adjusted three-factor model assumes the following form:

$$(R_{event} - R_{control})_{pt} = \alpha_p + \beta_p(R_{mt} - R_{ft}) + s_p SMB_t + h_p HML_t + e_{pt} \quad (8)$$

$(R_{event} - R_{control})_{pt}$ is the equal- or value-weighted monthly portfolio return between the simple returns of each event firm and its size-BM matched control firm. Moreover, for portfolio t, $(R_{event} - R_{control})_{pt}$ contains those firms whose event period includes the month t. In this adjusted model, α_p is a measure of long-term abnormal performance which is zero under the null hypothesis that no abnormal performance exists. Now, to test this null hypothesis, the t-statistic is given as:

$$t = \frac{\hat{\alpha}_p}{s(\hat{\alpha}_p)}$$

$\hat{\alpha}_p$ is an estimator of α_p, and $s(\hat{\alpha}_p)$ is the corresponding standard error of $\hat{\alpha}_p$.

3 Simulation method

Following the work of Lyon, Barber, and Tsai (1999), we randomly select 1000 samples of 200 event months without replacement to assess the specification of the employed methodologies. For each of these 200 event months, we randomly draw one stock from the population of all stocks that are active in the database for that month. For a well-specified test statistic, 1000α tests reject the null hypothesis. A test is conservative if fewer than 1000α null hypotheses are rejected and is anticonservative if more than 1000α null hypotheses are rejected. Based on this procedure, we test the specification of the t-statistic at 5% theoretical levels of significance. A well-specified null hypothesis rejects the null at the theoretical rejection level in favor of the alternative hypothesis of negative (positive) abnormal returns in $1000\alpha/2$ samples.

4 Test specification

In this section, we report the specification of various methodologies under consideration. We first discuss the results in random samples. Later, we consider different types of nonrandom samples based on firm size, book-to-market ratio and overlapping returns.

4.1 Random samples

Table 1 shows the rejection rates in 1000 simulations with a random sample of 200 firms. These results indicate that the BHAR method based on size-BM control firms yields well-specified test statistics in each of the three investment horizons. However, the test statistics are severely missepcified when the BHARs are calculated using the reference portfolios. For example, for a three-year holding period the rejection rates at the 5% level of significance are 3.6% and 0.9% for control firm approach and 4.2% and 0% for reference portfolio method.

The numbers presented in Table 1 further reveal that among the calendar time portfolio methods considered in this paper, our proposed approach is better specified than the rest in each case. One striking finding is that test statistics based on the Fama-French three-factor (henceforth FF3F) model are misspecified regardless of whether equally-weighted or value-weighted portfolios are employed. These findings conform to those reported by Yan (2012) in his empirical research. Moreover, while using the adjusted three-factor model, the size improves. For example, for a five-year holding period and with value-weighted portfolios, the rejection rates at 5% level of significance are 5.2% and 2.0% for FF3F and 2.8% and 1.6% for the adjusted version.

4.2 Nonrandom samples

4.2.1 Firm size

To assess the effect of size-based sampling biases on the employed methods, we randomly choose 1000 samples separately from the largest size decile and smallest size decile. Tables 2A and 2B display these results. The empirical analysis shows that SCTA produces well-specified test statistics for size-based samples. For instance, for a three-year horizon and with large firms and value-weighted portfolios, the rejection rates for SCTA at 5% level of significance are 1.8% and 3.0% when reference portfolios are used and 2.6% and 2.8% when the control firm approach is considered. The BHAR methodology, however, yields either negatively or positively skewed test statistics when the reference portfolio approach is taken into consideration. The size improves if the BHAR is estimated on the basis of control firms. The rejection rates for FF3F, on the other hand, are much higher than the theoretical levels. But, the level of misspecification decreases when the adjusted Fama-French three-factor model is used as an alternative.

Table 1. Specification of tests in random samples

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	2.4	1.2	2.0	2.8	2.8	2.8
Standardized Calendar Time Approach	Size-BM Control Firm	2.0	0.4	2.8	1.2	1.6	2.4
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	9.2*	0.0	4.2*	0.0	3.9*	0.0
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	2.0	1.4	3.6	0.9	3.0	2.3
Fama-French Three-Factor Model	Not Applicable	4.4*	0.0	3.8*	0.0	4.0*	0.0
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	3.9*	1.7	3.6	2.8	1.2	2.3
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	2.0	2.8	2.8	3.2	2.4	2.0
Standardized Calendar Time Approach	Size-BM Control Firm	2.8	1.6	3.6	2.0	2.8	3.1
Fama-French Three-Factor Model	Not Applicable	0.8	3.0	6.4*	2.6	5.2*	2.0
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	5.6*	1.2	3.2	0.7	2.8	1.6

Note: This table presents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table 2A. Specification of tests in samples with small firms

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	0.8	7.2*	1.8	3.0	2.1	3.8*
Standardized Calendar Time Approach	Size-BM Control Firm	1.4	6.1*	2.6	2.8	0.9	3.1
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	2.8	36.4*	0.0	25.6*	0.0	15.6*
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	4.8*	1.2	2.0	1.3	3.6	2.1
Fama-French Three-Factor Model	Not Applicable	0.0	42.2*	0.3	34.8*	1.1	22.0*
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	0.4	3.8*	3.6	1.6	0.8	6.4*
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	0.0	6.8*	2.8	3.4	3.1	4.7*
Standardized Calendar Time Approach	Size-BM Control Firm	1.6	3.4	2.1	2.4	2.3	3.0
Fama-French Three-Factor Model	Not Applicable	0.0	34.8*	0.0	16.4*	0.2	9.2*
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	2.1	4.8*	1.9	2.8	3.2	3.9*

Note: This table presents the percentages of 1000 samples of 200 large firms that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Corporate Ownership & Control / Volume 12, Issue 1, 2014, Continued - 2

Table 2B. Specification of tests in samples with small firms

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	5.4*	0.4	2.4	2.6	3.6	3.6
Standardized Calendar Time Approach	Size-BM Control Firm	6.8*	1.2	3.2	0.8	1.6	2.8
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	21.2*	0.0	15.6*	0.0	11.8*	0.3
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	3.4	1.8	2.4	2.0	4.0*	3.2
Fama-French Three-Factor Model	Not Applicable	8.8*	0.2	8.0*	3.1	12.4*	3.5
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	4.0*	0.8	6.0*	1.4	5.6*	2.3
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	6.0*	0.0	4.1*	3.1	2.7	3.5
Standardized Calendar Time Approach	Size-BM Control Firm	4.4*	0.2	2.8	1.3	1.6	2.0
Fama-French Three-Factor Model	Not Applicable	9.2*	0.6	7.6*	1.8	5.2*	2.0
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	4.2*	1.6	3.8*	0.8	2.4	1.2

Note*:* This table presents the percentages of 1000 samples of 200 small firms that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

4.2.2 Book-to-market (BM) ratio

To investigate the specification of the tests under study, we consider drawing firms on the basis of BM values. To do so, firms are deciled into ten groups based on rankings of BM ratio at the end of June each year. We choose the groups with the highest BM ratio and the lowest BM ratio for robustness check. For each group, we select a random sample of 200 firms. We repeat the procedure 1000 times and present the result in Tables 3A and 3B. Inspecting these tables suggests that our proposed calendar time methodology yields reasonably well-specified test statistics for each type of samples based on book-to-market ratios. Our analysis also documents that the BHAR approach produces better specified *t*-statistics when the control firm approach is employed, but the rejection level increases while using reference portfolios. In addition, the conventional Fama-French three-factor model as well as its modified version produce misspecified test statistics.

4.2.3 Overlapping returns

We consider nonrandom samples based on overlapping returns to inspect the behaviour of the employed methods in the presence of cross-sectional correlation of abnormal returns. Selection of these samples involves two steps. The first stage involves a random selection of 100 firms from the population. In the second stage, for each of these 100 firms, we randomly choose a second event month that is within H - 1 periods of the original event month (either before or after), where H equals 12, 36 or 60. Hence we have 200 firms with 200 event months where the same firm appears in the sample twice and this generates the issue of overlapping returns. This procedure is repeated 1000 times and Table 5 presents the results.

The empirical procedure reveals that the BHAR approach produces misspecified test statistics and these results are consistent with those documented in previous studies (e.g. Lyon, Barber, and Tsai (1999) and Mitchell and Stafford (2000)). Such misspecifications are observed, because the BHAR approach assumes that the observations are cross-sectionally uncorrelated. Jegadeesh and Karceski (2009), however, report that this assumption is tenable in random samples of event firms, but it would be violated in nonrandom samples, where the returns for event firms are positively correlated. Our analysis further shows that all the calendar time portfolio methods yield well-specified test statistics with few exceptions occurring in the one year horizon. This result is expected, since Fama (1998) and Mitchell and Stafford (2000) document that by forming monthly calendar time portfolios, all cross-correlations of event-firm abnormal returns are automatically accounted for in the portfolio variance and hence calendar time methodology performs better than BHAR approach in the presence of cross-sectional correlation of event firm anomalies.

Table 3A. Specification of tests in samples of firms with high BM value

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	0.2	4.8*	2.4	2.8	1.6	3.4
Standardized Calendar Time Approach	Size-BM Control Firm	0.8	6.0*	2.8	2.3	3.2	1.8
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	0.0	8.4*	0.0	3.9*	0.7	3.4
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	1.6	2.9	1.8	4.0*	2.8	3.6
Fama-French Three-Factor Model	Not Applicable	1.2	27.4*	0.0	14.2*	0.7	8.2*
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	1.6	3.8*	2.6	1.4	1.1	5.2*
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	0.3	7.2*	3.2	0.8	1.1	2.4
Standardized Calendar Time Approach	Size-BM Control Firm	0.0	3.9*	3.4	1.9	3.6	2.0
Fama-French Three-Factor Model	Not Applicable	0.0	14.8*	0.0	7.8*	0.8	5.6*
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	0.8	4.4*	2.9	0.6	3.2	3.8*

Note: This table presents the percentages of 1000 samples of 200 firms with high BM value that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table 3B. Specification of tests in samples of firms with low BM value

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	4.4*	0.6	1.3	2.1	1.2	3.0
Standardized Calendar Time Approach	Size-BM Control Firm	1.3	3.6	1.2	1.6	2.1	3.4
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	9.5*	0.0	3.2	0.8	4.4*	0.0
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	1.8	2.4	3.6	1.9	3.6	2.1
Fama-French Three-Factor Model	Not Applicable	17.8*	0.0	12.0*	0.4	3.6	1.3
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	3.6	0.8	5.4*	1.2	4.2*	0.7
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	1.2	2.8	2.8	3.2	2.4	2.0
Standardized Calendar Time Approach	Size-BM Control Firm	1.1	2.6	3.6	2.0	2.8	3.1
Fama-French Three-Factor Model	Not Applicable	4.8*	0.3	3.9*	0.4	3.1	0.8
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	3.4	1.6	3.7*	2.7	4.0*	2.8

Note: This table presents the percentages of 1000 samples of 200 firms with low BM value that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Corporate Ownership & Control / Volume 12, Issue 1, 2014, Continued - 2

Table 4. Specification of tests in samples of firms with overlapping returns

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	4.2*	2.3	3.2	0.9	2.8	2.6
Standardized Calendar Time Approach	Size-BM Control Firm	2.8	1.6	2.0	2.8	2.8	1.3
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	5.8*	1.6	3.6	4.8*	0.9	3.9*
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	4.0*	0.2	2.0	4.1*	1.7	4.6*
Fama-French Three-Factor Model	Not Applicable	1.4	7.1*	2.6	5.2*	3.7*	0.8
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	1.6	6.1*	1.1	2.0	1.6	2.8
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	1.2	3.9*	2.3	2.6	1.2	2.4
Standardized Calendar Time Approach	Size-BM Control Firm	1.6	3.4	2.1	2.4	2.3	3.0
Fama-French Three-Factor Model	Not Applicable	1.2	5.2*	1.6	3.8*	2.7	1.1
Adjusted Fama-French Three-Factor Model	Size-BM Control Firm	0.2	4.4*	3.6	2.0	2.1	2.3

Note: This table presents the percentages of 1000 samples of 200 firms with overlapping returns that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

5 Power

This section documents the power of alternative methodologies in random samples. Note that we exclude nonrandom samples from our analysis, since the *t*-tests based on such samples are generally misspecified. To examine the power of test, we introduce a constant level of abnormal return ranging from -20% to 20% at an interval of 5% to event firms. However, we employ only equally-weighted portfolios to make a direct comparison with the BHAR approach. In addition, we consider the estimates based on control firm approach as BHAR estimators based on 25 size-BM reference portfolios are severely skewed in random samples. We also exclude the traditional Fama-French three-factor model from our power analysis as the test statistics based on this model are not well-specified in random samples. Table 6 indicates the percentages of 1000 random samples of 200 firms that reject the null hypothesis of zero abnormal returns over a three-year holding period. Figure 1 also plots power of the tests.

It is evident from Table 5 and Figure 1 that our proposed standardized calendar time approach produces the most powerful *t*-statistic, followed by the BHAR method. The adjusted Fama-French three-factor model, on the other hand, has low power to identify the long-run anomalies. For instance, with 10% (-10%) per year abnormal returns, the rejection rate is 95% (91%) for SCTA, 74%(67%) for the BHAR method and 53% (44%) for the modified three-factor model. We, therefore, conclude that in case of detecting the abnormal performance, the standardized calendar time approach achieves higher power than the BHAR methodology.

6 Conclusions

This paper investigates the performance of the existing long-run event study methodologies with the UK security market data. Doing so employs the buy-and-hold abnormal return approach and the calendar time portfolio method to measure the return anomalies. While numerous recent studies examine the long-term stock price performance by exercising these two popular approaches, none of the methods is free of criticisms. This paper makes the use of a refined calendar time methodology, proposed by Dutta (2014), to resolve the ongoing debates regarding this approach. The empirical analysis indicates that the standardized calendar time approach of Dutta yields reasonably well-specified test statistics in all types of nonrandom samples. The results further show that in case of detecting the abnormal performance, this standardized calendar time methodology has higher power than other empirical procedures used in this study. One of the major findings of this study is that the Fama-French three-factor model produces misspecified test statistics even in random samples. Our simulation also reveals that the adjusted three-factor model performs well after controlling for size and book-to-market ratio biases. In addition, the buy-and-hold abnormal return approach yields reasonably well-specified test statistics when control firm approach is employed. But, the BHAR methodology possesses lower power than the modified calendar time approach. However, we document that all the employed approaches perform well in the UK stock market.

Table 5. Power of alternative methods in random samples

Methods	Induced Level of Abnormal Return (%) over 3 Years								
	-20	-15	-10	-5	0	5	10	15	20
Standardized Calendar Time Approach	1.00	0.91	0.62	0.18	0.04	0.23	0.69	0.95	1.00
Buy-and-Hold Abnormal Return Method	0.89	0.67	0.34	0.11	0.03	0.15	0.43	0.74	0.91
Adjusted Fama-French Three-Factor Model	0.68	0.44	0.21	0.09	0.04	0.12	0.29	0.53	0.78

Note: This table presents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over three-year holding period. We add the levels of annual abnormal return indicated in the column heading. In order to make a direct comparison with BHAR approach, only equally-weighted portfolios are considered in our analysis. In addition, we exclude the reference portfolio approach while calculating the power of tests, since the BHAR estimates based on 25 size-BM reference portfolios are generally biased in random samples.

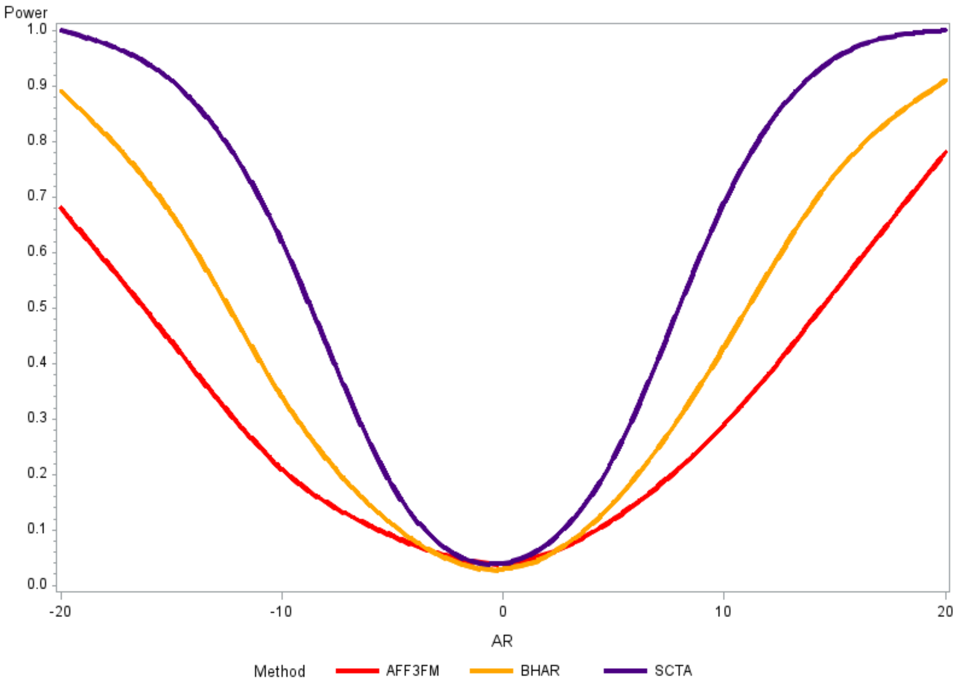

Figure 1. Simulated power of different methods

This figure represents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over three-year holding period. We consider equally weighted portfolios to make a direct comparison with BHAR approach. The horizontal axis indicates the induced level of annual abnormal returns (%), while the rejection rates are shown in the vertical axis. In addition, AFF3FM indicates the adjusted Fama-French three-factor model.

References

1. Barber, B. & Lyon, J. 1997. Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, 43 (3): 341-372.
2. Bessembinder, H. & Zhang, F. 2013. Firm characteristics and long-run stock returns after corporate events. *Journal of Financial Economics*, 109 (1): 83-102.
3. Boehme, D. R. & Sorescu, S. M. 2002. The long-run performance following dividend initiations and

resumptions: Underreaction or Product of Chance?. *The Journal of Finance*, 57 (2): 871-900.
4. Dutta, A. 2014. Does Calendar Time Portfolio Approach Really Lack Power?. *International Journal of Business and Management*, 9 (9): 260-266.
5. Espenlaub, S., Gregory, A. & Tonks, I. 2000). Re-assessing the long-term under-performance of UK initial public offerings. *European Financial Management*, 6: 319-42.
6. Fama, E, & French, K. 1993. Common risk factors in returns on stocks and bonds. *Journal of Financial Economics* 33: 3-56.
7. Fama, E. 1998. Market effciency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49 (3): 283-306.
8. Huang, Y. 2000. Long-term Abnormal Stock Performance: UK evidence. Ph.D. Thesis. University of Exeter: U.K.
9. Jaffe, J. F. (1974). Special information and insider trading. *Journal of Business*, 47: 410-428.
10. Jegadeesh, N. & Karceski, J. 2009. Long-run performance evaluation: correlationand heteroskedasticity-consistent tests. *Journal of Empirical Finance*, 16 (1): 101-111.
11. Kothari, S. & Warner, J. B. 1997. Measuring long-horizon security price performance. *Journal of Financial Economics,* 43 (3), 301-339.
12. Knif, J., Kolari, J. W. & Pynnönen, S. 2013. A Robust and powerful test of abnormal stock returns in long-horizon event studies. Working Papers Series: Available at SSRN: http://ssrn.com/abstract=2292356 or http://dx.doi.org/10.2139/ssrn.2292356.
13. Levis, M. 1993. The long-run performance of initial public offerings: the UK experience 1980-1988. *Financial Management*, 22: 28-41.
14. Loughran, T. & Ritter, J. 2000. Market efficiency, uniformly least powerful tests of market efficiency. *Journal of Financial Economics*, 55 (3): 361-389.
15. Lyon, J., Barber, B. & Tsai, C. 1999. Improved methods of tests of long-horizon abnormal stock returns. *Journal of Finance*, 54 (1): 165-201.
16. Mandelker, G. 1974. Risk and return: The case of merging firms. *Journal of Financial Economics*, 1: 303-336.
17. Mitchell, M. & Stafford, E. 2000. Managerial decisions and long-term stock price performance. *Journal of Business*, 73 (3): 287-329.
18. Ritter, J. 1991. The long-term performance of initial public offerings. *Journal of Finance*, 46 (1): 3-27.

Conducting Long-Run Event Studies in Asia-Pacific Security Markets

Anupam Dutta

Department of Mathematics and Statistics,

University of Vaasa

&

Seppo Pynnönen

Department of Mathematics and Statistics,

University of Vaasa

Abstract

This paper investigates the robustness of existing long-run event study methodologies using the Asia-Pacific security market data. In doing so, the study employs the buy-and-hold abnormal return approach and the calendar time portfolio method to measure the return anomalies. Since each of these two widely used approaches has a number of potential pitfalls, we also use standardized calendar time approach proposed by Dutta (2014a). The empirical analysis shows that all the employed methods are effective in the selected Asia-Pacific security markets. Simulations also show that the modified calendar time approach of Dutta documents better specification and power than the conventional methodologies.

Keywords: Long-run anomalies, Standardized abnormal returns, Test specification, Power of test.

JEL Classification: C1, G1.

I. Introduction

Although long-run event studies have an extensive history during the last two decades, not a single simulation study focusing on Asia-Pacific security markets is found in the literature. Corrado and Truong (2008), however, perform a short-run event study employing the stock market data from a number of major Asia-Pacific countries. But, investigating the robustness of existing long-run event study methodologies using such non-U.S. security market data is yet to be documented. The objective of the present study is to conceal this research gap by conducting a long-run event study in leading Asia-Pacific security markets such as Australia, China, India, Japan, Singapore and South Korea.

Long-run event study methodologies have seen many advances over the years. The extensive literature includes Barber and Lyon (1997), Kothari and Warner (1997), Fama (1998), Lyon, Barber and Tsay (1999), Mitchell and Stafford (2000), Boehme and Sorescu (2002) and Jegadeesh and Karceski (2009) Bessembinder and Zhang (2013) and Knif, Kolari and Pynnönen (2013). Reviewing this existing literature suggests that the Buy-and-Hold Abnormal Return (BHAR) Approach and the Calendar Time Portfolio (CTP) method are commonly used in inspecting the long-horizon abnormal performance following major corporate events. Following the prior literature, we also employ these two empirical procedures in our simulation study.

While numerous recent studies consider the applications of these two widely used methodologies, each of them is a subject to criticism. Fama (1998), for example, documents that the BHAR approach produces misspecified test statistics, since it does not address the issue of cross-sectional correlation of event-firm anomalies. Mitchell and Stafford (2000) also report that the BHAR methodology should not be used in its traditional form. Loughran and Ritter (2000), on the other hand, argue that CTP approach loses its power as it weights each month equally.

Following the work of Dutta (2014a), this paper considers applying the Standardized Calendar Time Approach (henceforth SCTA) where we first standardize the abnormal returns for each of the event firms forming the monthly portfolios and then each portfolio is weighted such that periods of heavy event activity receive more weight than periods of low event activity. However, employing standardized abnormal returns is well-documented in the literature. For example, Jaffe (1974) and Mandelker (1974) employ standardized portfolio returns for investigating the long-run abnormal performance. Fama (1998) also suggests to standardize the abnormal returns to resolve the issues raised by Loughran and Ritter (2000). Later, Mitchell and Stafford (2000) use standardized abnormal returns to alleviate the heteroscedasticity problem that often occurs in CTP approach due to the varying portfolio construction.

The empirical analysis reveals that the Standardized Calendar Time Approach, in general, produces well-specified test statistics in all types of nonrandom samples. The BHAR approach, on the other hand, yields reasonably well-specified test statistics when the control firm approach is employed. While using a reference portfolio as the benchmark, the test statistics based on the BHAR methodology are not well-specified. Lyon, Barber, and Tsai (1999) also document similar output. To assess the robustness of SCTA, the results of mean monthly calendar time abnormal returns (MMCTAR) method are also reported. Simulations show that MMCTAR performs well when the abnormal returns are calculated using the control firm approach. Our findings further document that SCTA achieves higher power than other conventional methodologies. We, however, document that all the methodologies considered perform well in the selected Asia-Pacific security markets. We also recommend the use of control firm approach for measuring the long-run abnormal performance of event firms.

This paper contributes to the prior literature in two ways. First, it investigates - for the first time - the robustness of conventional long-run event study approaches in Asia-Pacific security markets. Second, it uses a refined version of the traditional calendar time methodology which documents better specification and power than the common approaches. The remainder of the paper is structured as follows. The data and methodology are discussed in section 2. Section 3 illustrates the simulation procedure. Section 4 discusses the test specification. Section 5 documents power of the tests, and Section 6 concludes the paper.

II. Data and Methodology

We obtain stock prices, market value (MV) and book-to-market (BM) value data from DataStream. The sample period ranges from July 1973 to December 2008. In this paper, we construct 25 size-BM portfolios as expected return benchmarks. In doing so, at the end of June of year *t*, firms are quantiled into five groups on the basis of their market values. Firms are further quantiled into five groups based on their book-to-market ratios. However, we also consider a size-BM-matched control firm to calculate the abnormal returns. Identifying such a control firm is a 2-step procedure. First, we identify all the firms with a market value of equity between 70% and 130% of the sample firm at the most recent end of June. Then from this set of firms, we choose the firm with BM closest to that of the sample firm as of the previous December. However, since the currencies of the financial markets differ from one to another, we construct the benchmarks for each market separately and then merge all the data sets.

a) Mean Monthly Calendar Time Abnormal Returns (MMCTAR)

The calculation of mean monthly calendar time abnormal return (MMCTAR) is the following:

$$MMCTAR = \frac{1}{T}\sum_{t=1}^{T} CTAR_t, \quad (1)$$

where

$$CTAR_t = R_{pt} - E(R_{pt}) \quad (2)$$

Within this framework, R_{pt} is the monthly return on the portfolio of event firms, $E(R_{pt})$ is the expected return on the event portfolio which is proxied by the raw return on either a reference portfolio or a control firm and *T* is the total number of months in the sample period. To test the null hypothesis of no abnormal returns, the *t*-statistic of MMCTAR is obtained by using the intertemporal standard deviation of the monthly CTARs defined in equation (2).

b) Standardized Calendar Time Approach (SCTA)

Forming the monthly portfolios in the standardized calendar time approach involves two steps (Dutta, 2014a). We first calculate the standardized abnormal returns for each of the sample firms. In doing so, the abnormal returns for firm *i* are computed as $\varepsilon_{it} = r_{it} - E(r_{it}); t = 1, \ldots, H$, where r_{it} denotes the log return on event firm *i* in the calendar month *t* and $E(r_{it})$ is the expected return which is proxied by the raw return either 25 size-BM reference portfolios or a size-BM matched control firm and *H* is the holding period which equals 12, 36 or 60 months. The next task is to estimate the event-portfolio residual variances using the *H*-month residuals computed as monthly differences of *i*-th event firm returns and control firm returns. Dividing ε_{it} by the estimate of its standard deviation yields the corresponding standardized abnormal return, say, z_{it}, for event firm *i* in month *t*. Now let N_t refer to the number of event firms in the calendar month *t*. We then calculate the calendar time abnormal return for portfolio *t* as:

$$CTAR_t = \sum_{i=1}^{N_t} x_{it} z_{it}, \quad (3)$$

where the weight x_{it} equals $\frac{1}{N_t}$ when the abnormal returns are equally-weighted and $\frac{MV_{it}}{\sum MV_{it}}$ when the abnormal returns are value-weighted by size.

Following the work of Dutta, each of the monthly CTARs is weighted by $1/\sqrt{\left(\sum_{i=1}^{N_t} x_{it}^2\right)}$. For instance, when the abnormal returns are equally weighted i.e., when $x_{it} = \frac{1}{N_t}$, then $1/\sqrt{\left(\sum_{i=1}^{N_t} x_{it}^2\right)} = \sqrt{N_t}$. This weighting scheme is lucrative as it gives more loadings to periods of heavy event activity than the periods of low event activity. Now the grand mean monthly abnormal return, denoted by $\overline{CTAR}$, is calculated as:

$$\overline{CTAR} = \frac{1}{T}\sum_1^T CTAR_t \quad (4)$$

While finding $\overline{CTAR}$, it might be the case that a number of portfolios do not contain any event firm. In such situations, those months are dropped from the analysis. To test the null hypothesis of no abnormal performance, the *t*-statistic of $\overline{CTAR}$ is computed by using the intertemporal standard deviation of the monthly CTARs defined in equation (3).

c) Buy-and-Hold Abnormal Return (BHAR)

To assess the robustness of our findings, we investigate the result of the BHAR methodology. An *H*-month BHAR for event firm *i* is defined as:

$$BHAR_{iH} = \prod_{t=1}^{H}(1 + R_{it}) - \prod_{t=1}^{H}(1 + R_{Bt}), \quad (5)$$

where R_{it} denotes the return on event firm *i* at time *t* and R_{Bt} indicates the return on a control firm.

To test the null hypothesis that the mean buy-and-hold return equals zero, the conventional *t*-statistic is given by:

$$t_{BHAR} = \frac{\overline{BHAR_H}}{\sigma(BHAR_H)/\sqrt{n}}, \quad (6)$$

where $\overline{BHAR_H}$ implies the sample mean and $\sigma(BHAR_H)$ refers to the cross-sectional sample standard deviation of abnormal returns for the sample containing *n* firms.

However, the earlier studies such as Mitchell and Stafford (2000), Boehme and Sorescu (2002), Jegadeesh and Karceski (2009) report that the BHAR approach does not control well for the cross-sectional correlation among individual firms in nonrandom samples and thus yields misspecified *t*-statistics. Moreover, the test statistics based on BHARs also have this misspecification problem, since the distribution of BHARs is highly skewed. Though bootstrapping corrects for the skewness problem to some extent, it ignores the cross-sectional dependence of abnormal returns.

III. SIMULATTION METHOD

To test the specification of the *t*-statistics, we randomly select 1000 samples of 200 event months without replacement. For each of these 200 event months, we randomly draw one stock from the population of all stocks that are active in the database for that month. For a well-specified test statistic, 1000α tests reject the null hypothesis. A test is conservative if fewer than 1000α null hypotheses are rejected and is anticonservative if more than 1000α null hypotheses are rejected. Based on this procedure, we test the specification of the *t*-statistic at 5% theoretical levels of significance. A well-specified null hypothesis rejects the null at the theoretical rejection level in favor of the alternative hypothesis of negative (positive) abnormal returns in $1000\alpha/2$ samples.

IV. TEST SPECIFICATION

This section addresses the specification of tests under study. Results of both random as well as nonrandom samples are reported. We consider nonrandom samples based on book-to-market ratio and overlapping returns. Since the currencies differ from market to market, specification of tests in samples with small or large firms is not included in our analysis.

a) Random Samples

Table I reports the rejection rates in 1000 simulations with a random sample of 200 firms. The empirical analysis suggests that all the *t*-statistics based on buy and hold abnormal returns calculated using the reference portfolio approach are negatively biased implying that these tests have higher rejection rates in the lower tail. However, when the BHARs are calculated from control firms, such misspecifications are significantly reduced. For example, for a 3-year holding period, the rejection rates at 5% level of significance are 28.4% and 0% for reference portfolio method and 3.3% and 1.8% for control firm approach.

Though there exists some evidence of negative bias when the reference portfolio approach is used at an annual horizon, the standardized calendar time approach generally produces well-specified test statistics in random samples regardless of whether reference portfolio method or control firm approach is adopted. For example, for a 5-year holding period and with equally-weighted portfolios, the rejection rates at 5% level of significance are 1.1% and 3.6% for reference portfolio method and 2.1% and 1.3% for control firm approach. Finally, the *t*-statistics based on mean calendar time abnormal returns are better specified when control firm approach is employed.

Table I: Specification of Tests in Random Samples

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	5.6*	0.0	3.2	1.2	1.1	3.6
Standardized Calendar Time Approach	Size-BM Control Firm	1.2	2.8	2.8	2.4	2.1	1.3
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	20.8*	0.0	28.4*	0.0	38.2*	0.0
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	4.6*	1.9	3.3	1.8	3.1	2.4
Mean Calendar time Abnormal returns	Size-BM Portfolio	5.2*	1.6	2.0	1.2	4.2*	0.9
Mean Calendar time Abnormal returns	Size-BM Control Firm	3.2	0.8	2.2	1.4	2.3	2.1
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	6.0*	0.0	3.6	1.2	2.0	3.2
Standardized Calendar Time Approach	Size-BM Control Firm	1.2	1.2	2.4	2.0	2.6	2.4
Mean Calendar time Abnormal returns	Size-BM Portfolio	2.4	1.2	4.8*	0.9	6.2*	0.8
Mean Calendar time Abnormal returns	Size-BM Control Firm	2.6	0.8	1.8	3.2	2.1	2.3

Note. This table presents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

b) Nonrandom Samples

Book-to-Market (BM) Ratio

For assessing the specification of the methods considered, we sample firms on the basis book-to-market (BM) ratio. To serve this purpose, firms are first deciled into ten groups based on rankings of BM ratio at the end of June each year. We then choose the groups with the highest BM ratio and the lowest BM ratio for robustness check. For each group, we select a random sample of 200 firms. We repeat the procedure 1000 times. The results are presented in Table II and Table III which reveal that test statistics based on standardized calendar time approach are well-specified in most cases. For example, for a 3-year holding period and with value-weighted portfolios and for firms with high book-to-market ratio, the rejection rates at 5% level of significance are 1.2% and 3.0% for SCTA, and 1.2% and 4.4% for MMCTAR method when we consider the control firm approach. The BHAR method, on the other hand, produces either positively or negatively skewed test statistics depending on high or low book-to-market value respectively.

Table II: Specification of Tests in Samples of Firms with High BM Value

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	0.0	6.0*	2.0	2.4	0.8	3.2
Standardized Calendar Time Approach	Size-BM Control Firm	3.8*	0.4	1.2	2.6	3.6	2.1
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	0.0	13.8*	0.0	22.1*	0.0	26.8*
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	0.0	6.4*	2.3	2.8	0.9	3.1
Mean Calendar time Abnormal returns	Size-BM Portfolio	0.0	6.8*	0.0	5.6*	4.0*	2.0
Mean Calendar time Abnormal returns	Size-BM Control Firm	0.2	4.8*	3.6	0.8	2.2	3.0
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	0.4	5.6*	3.6	2.8	0.3	5.1*
Standardized Calendar Time Approach	Size-BM Control Firm	2.0	2.3	1.2	3.0	2.8	1.8
Mean Calendar time Abnormal returns	Size-BM Portfolio	0.8	5.4*	3.7*	1.2	2.0	3.2
Mean Calendar time Abnormal returns	Size-BM Control Firm	0.4	6.0*	1.2	4.4*	2.3	3.6

Note. This table presents the percentages of 1000 samples of 200 firms with high BM value that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Table III: Specification of Tests in Samples of Firms with Low BM Value

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	4.8*	0.0	3.8*	1.1	2.3	2.7
Standardized Calendar Time Approach	Size-BM Control Firm	3.7*	1.6	2.8	0.4	3.6	1.3
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	14.0*	0.2	22.6*	0.0	28.2*	0.0
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	4.2*	0.4	3.8*	0.6	2.4	2.6
Mean Calendar time Abnormal returns	Size-BM Portfolio	6.8*	0.0	4.6*	1.2	3.1	0.8
Mean Calendar time Abnormal returns	Size-BM Control Firm	4.4*	0.0	3.8*	2.6	3.2	0.4
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	4.0*	0.2	3.7*	1.8	1.6	2.0
Standardized Calendar Time Approach	Size-BM Control Firm	2.8	2.1	2.6	0.3	3.2	2.0
Mean Calendar time Abnormal returns	Size-BM Portfolio	4.9*	0.6	3.8*	1.6	2.8	2.6
Mean Calendar time Abnormal returns	Size-BM Control Firm	4.2*	0.4	2.1	2.3	3.4	2.8

Note. This table presents the percentages of 1000 samples of 200 firms with low BM value that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

Overlapping Returns

In order to investigate the performance of the employed methods in the presence of cross-sectional correlation of abnormal returns, we draw nonrandom samples based on overlapping returns. Such samples are selected in two steps. First, we randomly select 100 firms from the population. Then for each of these 100 firms, we randomly draw a second event month that is within *H-1* periods of the original event month (either before or after), where *H* is equal to 12, 36 or 60. This two-step procedure yields 200 firms with 200 event months where the same firm appears in the sample twice and hence generates the issue of overlapping returns. We repeat the same approach 1000 times and Table IV indicates the output.

Our findings report that the BHAR approach yields misspecified test statistics at all three investment periods. Other empirical studies such as Lyon, Barber and Tsay (1999) and Mitchell and Stafford (2000) document similar results. This misspecification occurs since the BHAR approach assumes that the observations are not cross-sectionally correlated. Jegadeesh and Karceski (2009) argue that such assumption is tenable in random samples of event firms, but it would be violated in nonrandom samples, where the event firm returns are positively correlated.

We further document that test statistics based on the calendar time methods are, in general, well-specified when return calculations overlap. The numbers presented in Table IV reveal that the value-weighted scheme produces lower rejection compared with the equally-weighted scheme. For example, for a 1-year holding period and with equally-weighted portfolios and when the reference portfolio method is considered, the rejection rates at 5% level of significance are 5.4% and 0% for the *t*-statistics produced by the modified CTP approach. The corresponding rejection rates are 2.8% and 0.4% when the portfolios are value-weighted. We, therefore, recommend that the value-weighted scheme should be employed if misspecifications occur due to overlapping returns. However, when the abnormal performance is measured on the basis of control firm approach, the performance of the mean monthly calendar time abnormal returns method improves.

Table IV: Specification of Tests in Samples of Firms with Overlapping Returns

Methods	Benchmark	1 Year		3 Years		5 Years	
		2.5	97.5	2.5	97.5	2.5	97.5
Panel A: Equally-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	5.4*	0.0	0.4	2.3	2.4	0.8
Standardized Calendar Time Approach	Size-BM Control Firm	4.0*	1.6	3.2	3.6	1.9	3.1
Buy-and-Hold Abnormal Return Method	Size-BM Portfolio	12.8*	0.6	21.6*	0.0	29.2*	0.0
Buy-and-Hold Abnormal Return Method	Size-BM Control Firm	6.2*	0.2	11.4*	0.0	20.2*	0.0
Mean Calendar time Abnormal returns	Size-BM Portfolio	6.4*	1.4	3.6	1.2	2.1	1.3
Mean Calendar time Abnormal returns	Size-BM Control Firm	5.6*	1.2	3.4	2.4	3.2	1.8
Panel B: Value-Weighted Portfolios							
Standardized Calendar Time Approach	Size-BM Portfolio	2.8	0.4	2.1	2.8	2.9	1.6
Standardized Calendar Time Approach	Size-BM Control Firm	1.6	1.2	3.2	1.4	2.0	3.6
Mean Calendar time Abnormal returns	Size-BM Portfolio	6.0*	0.4	2.3	2.6	2.8	1.2
Mean Calendar time Abnormal returns	Size-BM Control Firm	3.4	0.2	2.0	1.2	1.6	2.1

Note. This table presents the percentages of 1000 samples of 200 firms with overlapping returns that reject the null hypothesis of no annual, three-year and five-year abnormal returns at 5% level of significance. Panel A and Panel B indicate the specification of tests for equally- and value-weighted portfolios respectively. The numbers marked with * suggest that the empirical size is significantly different from the 5% significance level.

V. POWER

In this section, we document the power of alternative methods considered. We, however, consider only random samples, since the *t*-tests are generally misspecified in nonrandom samples. In order to evaluate the power of test, we introduce a constant level of abnormal return ranging from -20% to 20% at an interval of 5% to event firms. We also consider equally-weighted portfolios to make a direct comparison with BHAR approach. In addition, we consider the estimates based on control firm approach as BHAR estimators based on 25 size-BM reference portfolios are severely skewed in random samples. Table V reports the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal performances over a three-year investment period. Figure 1 plots power of the tests, too. Our results are also robust when the holding periods are one year as well as five years.

Inspecting Table V and Figure 1 confirms that test statistics based on the standardized calendar time approach are more powerful than those based on traditional methods. For instance, with 20% per year abnormal returns, the rejection rate is 100% for SCTA, 98% BHAR method and 92% for MMCTAR approach. We, like Dutta (2014b), conclude that the standardized calendar time approach has sufficient power to detect the abnormal performance of event firms. Loughran and Ritter (2000), however, claim that calendar time methodology has lower power than the event time approaches. Angand Zhang (2004) also argue that the calendar time portfolio approach starts lacking power as the holding period increases. In this paper, we find no evidence supporting these statements. In fact, we show that the standardized calendar time approach documents better power than the BHAR methodology.

Table V: Power of Alternative Methods in Random Samples

Methods	Induced Level of Abnormal Return (%) over 3 Years								
	-20	-15	-10	-5	0	5	10	15	20
Standardized Calendar Time Approach	0.98	0.78	0.61	0.19	0.05	0.33	0.78	0.97	1.00
Buy-and-Hold Abnormal Return Method	0.97	0.71	0.58	0.18	0.05	0.27	0.64	0.81	0.98
Mean Calendar time Abnormal returns	0.88	0.68	0.49	0.16	0.04	0.22	0.59	0.78	0.92

Note. This table presents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over a three-year holding period. We add the levels of annual abnormal return indicated in the column heading. In order to make a direct comparison with BHAR approach, only equally-weighted portfolios are considered in our analysis. In addition, we exclude the reference portfolio approach while calculating the power of tests, since the BHAR estimates based on 25 size-BM reference portfolios are generally biased in random samples.

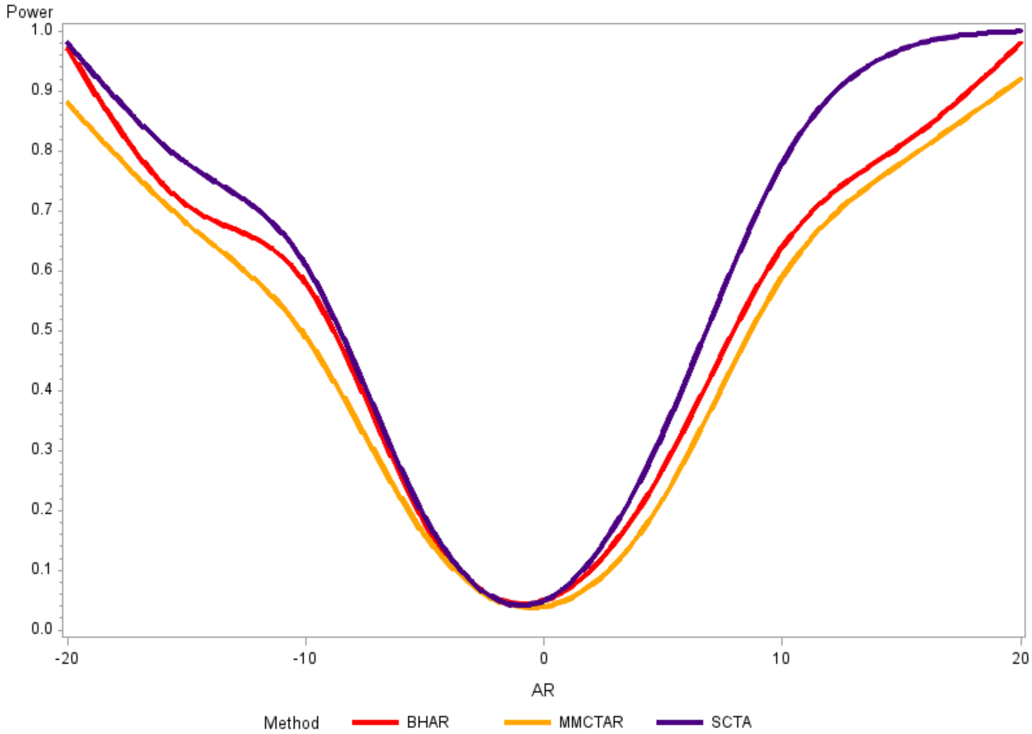

Figure 1: This figure represents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over a three-year holding period. We consider equally weighted portfolios to make a direct comparison with BHAR approach. The horizontal axis indicates the induced level of annual abnormal returns (%), while the rejection rates are shown in the vertical axis.

VI. CONCLUSIONS

This paper investigates the robustness of existing long-run event study methodologies in the leading Asia-Pacific stock markets. In doing so, the present study employs the buy-and-hold abnormal return approach and the calendar time portfolio method to measure the return anomalies. While a large number of studies examine the long-term stock price performance by exercising these two popular approaches, none of the methods is free of criticisms. In order to solve the ongoing debates, we consider the application of standardized calendar time approach (SCTA) of Dutta (2014a). To measure the abnormal performance of the sample firms, both control firm approach and reference portfolio approach have been adopted. The empirical analysis indicates that the traditional methods are found to be effective in leading Asia-Pacific security markets. We report that test statistics based on SCTA are generally well-specified in all types of nonrandom samples considered. The BHAR approach, on the other hand, yields reasonably well-specified test statistics only when the control firm approach is employed. Our simulations show that the mean monthly calendar time abnormal return methodology performs well when the abnormal returns are calculated using the control firm approach. We, therefore, advocate the use of control firm method for measuring the long-run abnormal performance of event firms. However, in case of detecting the abnormal performance, SCTA documents higher power than other empirical procedures used in this study. It is noteworthy that a well-specified test statistic is useless if it does not have power to correctly detect the signal of an abnormal return. Alternatively, instead of ability to detect the alternative hypothesis when it is true, power is the probability that a test correctly rejects the null hypothesis when it is false. Without power, statistical tests are useless in making inferences about a statistical population. Thus we strongly recommend the application of standardized calendar time approach in the analysis of long-term stock returns after corporate events.

REFERENCES

Ang, S. and Zhang, S. 2004, ‘An evaluation of testing procedures for long horizon event studies’, *Review of Quantitative Finance and Accounting*, vol. *23*, pp. 251–274.

Barber, B. and Lyon, J. 1997, ‘Detecting long-run abnormal stock returns: The empirical power and specification of test statistics’, *Journal of Financial Economics*, vol. 43, pp. 341-372.

Bessembinder, H. and Zhang, F. 2013, ‘Firm characteristics and long-run stock returns after corporate events’, *Journal of Financial Economics*, vol. 109, pp. 83-102.

Boehme, D. R. and Sorescu, S. M. 2002, ‘The long-run performance following dividend initiations and resumptions: Underreaction or Product of Chance?’, *The Journal of Finance*, vol. 57, pp. 871-900.

Corrado, C.J. and Truong, C. 2008, ‘Conducting event studies with Asia-Pacific security market data’, *Pacific-Basin Finance Journal*, vol. 16, pp. 493–521.

Dutta, A. 2014a, ‘Improved calendar time approach for measuring long-run anomalies’, Working paper.

Dutta, A. 2014b, ‘Investigating long-run stock returns after corporate events: the UK evidence’, *Corporate Ownership and Control*, vol. 12 (In press).

Fama, E. 1998, ‘Market effciency, long-term returns, and behavioral finance’, *Journal of Financial Economics*, vol. 49, pp. 283-306.

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

Jegadeesh, N. and Karceski, J. 2009, 'Long-run performance evaluation: correlationand heteroskedasticity-consistent tests', *Journal of Empirical Finance*, vol. 16, pp. 101-111.

Kothari, S. and Warner, J. B. 1997, 'Measuring long-horizon security price performance. *Journal of Financial Economics*', vol. 43, pp. 301-339.

Knif, J., Kolari, J. W. and Pynnönen, S. 2013, 'A Robust and powerful test of abnormal stock returns in long-horizon event studies', Working Papers Series: Available at SSRN: http://ssrn.com/abstract=2292356 or http://dx.doi.org/10.2139/ssrn.2292356.

Loughran, T. and Ritter, J. 2000, 'Market efficiency, uniformly least powerful tests of market efficiency', *Journal of Financial Economics*, vol. 55, pp. 361-389.

Lyon, J., Barber, B. and Tsai, C. 1999, 'Improved methods of tests of long-horizon abnormal stock returns', *Journal of Finance*, vol. 54, pp. 165-201.

Mandelker, G. 1974. 'Risk and return: The case of merging firms', *Journal of Financial Economics*, vol. 1, pp. 303-336.

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

Ritter, J. 1991, 'The long-term performance of initial public offerings', *Journal of Finance*, vol. 46, pp. 3-27.