Journal of Empirical Finance 47 (2018) 1–24

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

A robust and powerful test of abnormal stock returns in long-horizon event studies

Anupam Dutta [a], Johan Knif [b], James W. Kolari [c], Seppo Pynnonen [d,*]

[a] Department of Accounting and Finance, University of Vaasa, P.O. Box 700, FI-65101, Vaasa, Finland
[b] Department of Finance and Statistics, Hanken School of Economics, FI-65101 Vaasa, Finland
[c] Texas A&M University, Department of Finance, College Station, TX 77843-4218, United States
[d] Department of Mathematics and Statistics, University of Vaasa, P.O. Box 700, FI-65101, Vaasa, Finland

ARTICLE INFO

JEL classification:
C4
G1

Keywords:
Abnormal returns
Long-run event study
Standardized returns
IPOs
SEOs

ABSTRACT

This paper proposes a novel standardized test for abnormal returns in long-horizon event studies that takes into account cross-sectional correlation, autocorrelation, and heteroskedasticity of stock returns. Extensive simulation analyses demonstrate improved size and power of testing relative to existing long-run test methodologies. Application to initial public offerings and seasoned equity offerings further demonstrates robustness to extreme return outliers inherent in these long-run studies.

1. Introduction

A growing body of empirical literature finds significant long-run stock price reactions to a wide variety of firm-specific news announcements.[1] Behavioralists argue that long-run anomalies can occur due to short-term over- and underreactions by investors to information and subsequent long-term return reversals.[2] By contrast, Fama (1998) contends that market efficiency implies that long-term abnormal returns should disappear after risk adjustment.[3] The present study is agnostic on the behavioral/efficiency debate; instead, by improving the major weaknesses of size, power, and robustness of long-run statistical tests (Kothari and Warner, 2007, Table 2), we seek to contribute empirical methods to better understand and interpret the empirical evidence.

* Corresponding author.
E-mail addresses: anupam.dutta@uva.fi (A. Dutta), JKnif@hanken.fi (J. Knif), j-kolari@tamu.edu (J.W. Kolari), sjp@uwasa.fi (S. Pynnonen).

[1] For example, see studies on long-term return reversals (DeBondt and Thaler, 1985), initial public offerings (Ibbottson, 1975; Ritter, 1991; Loughran and Ritter, 1995; Billett et al., 2011; Liu et al., 2014, and others), mergers and acquisitions (Becher et al., 2015), spinoffs (Miles and Rosenfeld, 1983; Cusatis et al., 1993, and others), earnings announcements (Ball and Brown, 1968; Bernard and Thomas, 1990, and others), and private equity sponsors (Harford and Kolasinski, 2014). Many other corporate actions have been investigated in the context of long-run abnormal returns, including seasoned equity offerings, dividend omissions, new exchange listings, share repurchases, proxy fights, stock splits, etc. Also, long-run event studies have been published on a wide variety of business topics, particularly in accounting, economics, and law.

[2] See Barberis et al. (1998), Daniel et al. (1998), Shiller (2000), and others.

[3] Studies by Eckbo et al. (2000) and Bessembinder and Zhang (2013) find that imperfect control-firm matching leads to inappropriate risk adjustment of long-run abnormal returns. After taking into account various macroeconomic, market, or idiosyncratic risks, these studies find no significant long-run abnormal returns with respect to selected firm-specific news announcements.

https://doi.org/10.1016/j.jempfin.2018.02.004
Received 13 January 2016; Received in revised form 4 October 2017; Accepted 25 February 2018
Available online 5 March 2018
0927-5398/

It is well known that calendar time (CT) abnormal returns suffer from misspecification and/or bad model problems.[4] Lacking reliable mean models, many researchers[5] employ buy-and-hold abnormal returns (BHARs) popularized by Lyon et al. (1999) (henceforth LBT) that utilizes carefully chosen matched control firm portfolios or control firm stocks. However, results in Eckbo et al. (2000) and (Bessembinder and Zhang, 2013) suggest that BHARs also suffer from statistical and modeling problems (e.g., skewness, cross-sectional correlation, and matching quality degrades rapidly over time) that weaken the size, power, and robustness of tests.

In this paper we propose a robust and powerful test of abnormal stock returns that improves the statistical properties of long-horizon event study tests and helps to simplify existing practices. Our testing procedure seeks to determine whether an event firm's (long-run) return performance deviates from that of its benchmark (i.e., a matching reference firm or portfolio). Because the event period in long-run studies extends over several years, the event periods of different firms tend to overlap which introduces a serious problem with cross-sectional correlation in long-horizon tests (e.g., see Kothari and Warner, 2007). Additionally, stocks are highly heteroskedastic which affects estimation and testing of return effects. For example, in our sample period CRSP stock (annual) volatilities range from less than 10% up to 200% with a median of about 50%.[6] Heteroskedasticity can be expected to be even more pronounced in event studies as the event will most likely change the event firm's exposure to underlying risk factors and, at least temporally, increase idiosyncratic volatility (e.g., see Eckbo et al., 2007; Kothari and Warner, 2007). We propose a statistical procedure that accounts for the above issues, thereby providing a test that improves robustness and statistical power of inference over existing long-run tests.

Popular short-run event study tests by Patell (1976), Boehmer et al. (1991), Kothari and Warner (2007, Sec. 4.2), and Kolari and Pynnonen (2010, 2011) have found that weighting returns by their statistical precision (i.e., standard deviation) materially boosts power relative to unstandardized tests. Despite the fact that this power advantage is widely accepted in the literature (Fama, 1998, p. 265), application of standardized returns in long-horizon event studies is rare.[7] Here we extend short-run studies by developing a novel long-run test that incorporates standardized returns (hereafter *abnormal standardized returns*, or *ASR*s) to detect abnormal performance. The standardization efficiently captures the heteroskedasticity problem, but the cross-correlation issue needs additional consideration. While Kolari and Pynnonen (2010) have a solution for completely overlapping event periods, it is not applicable to long-run event studies with partially overlapping event periods. The calendar time portfolio approach solves the cross-correlation problem but is prone to low statistical power which is corroborated in the simulation experiments of this paper. Jegadeesh and Karceski (2009) propose a modification of the calendar time approach and utilize heteroskedasticity and autocorrelation robust approaches to overcome cross-sectional correlation and other issues. Again, while robustness of the approach generally improves, the cost is loss of power due to the large number of estimated parameters (Jegadeesh and Karceski, 2009, p. 110). To avoid estimation of many covariance parameters, we adopt existing clustering techniques (e.g., see Cameron et al., 2011) to compute robust standard errors. The resulting test accounts for potential biases arising from returns' cross-sectional correlation, autocorrelation, and volatility clustering without power loss.

Based on monthly U.S. stock returns from July 1973 to December 2009, extensive simulation analyses are conducted to comparatively test the size and power properties of our ASR test versus BHAR and CT tests over long horizons. Interestingly, our empirical findings indicate that in nonrandom samples benchmarking against complicated size and book-to-market matched reference portfolios in BHARs can be replaced by a market index portfolio. Importantly, regardless of the chosen reference portfolio, the proposed ASR test exhibits accurate size properties and superior power relative to conventional *t*-tests in long-horizons. Also, we find that event-date clustering in random samples does not distort the size or power of test statistics. In nonrandom samples, market index, size, and book-to-market matched portfolios do not result in reliable inferences. In this case, rather than matching on firm characteristics, in combination with the ASR test, we find that sample subpopulations are recommended as benchmark portfolios. Market index and subpopulation reference portfolios not only provide more reliable results but simplify the testing procedure. If a subpopulation benchmark cannot be identified, a size/book-to-market matched reference stock is the safest option to control the size of the test, but the cost is diminished power which our ASR test can help to offset. To demonstrate our new test, we investigate long-run abnormal returns of initial public offerings and seasoned equity offerings. These finding reveal that, unlike conventional tests, ASR tests are robust to extreme return outliers inherent in many long-run studies.

The next section discusses return metrics and related test statistics. Section 3 describes the methodology, including simulation design. Section 4 reports the simulation results. Section 5 conducts tests of long-run abnormal returns after initial public offerings (IPOs) and seasoned equity offerings (SEOs). Section 6 concludes.

2. Return metrics and related test statistics

2.1. Continuously compounded returns

Because simple returns are bounded from below but not above, skewness is a major statistical problem. The central limit theorem guarantees normal approximation of test statistics, but convergence to the limiting distribution slows down materially for skewed parent distributions and makes normal approximation inaccurate even for fairly large sample sizes (e.g., Sutton, 1993).

[4] See Fama (1998), Eckbo et al. (2000), Bessembinder and Zhang (2013), and others.

[5] See Mitchell and Stafford (2000), Eberhart and Siddique (2002), Boehme and Sorescu (2002), Gompers and Lerner (2003), Byun and Rozeff (2003), Gao et al. (2013), Chen et al. (2014), and others.

[6] For example in bond market event studies Ederington et al. (2015) observe that accounting for the heteroskedasticity materially improves the power of event effect tests.

[7] To our knowledge, Mitchell and Stafford (2000, Section V.C) is the only empirical application.