Journal of Empirical Finance 16 (2009) 101–111

Contents lists available at ScienceDirect

Journal of Empirical Finance

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

Long-run performance evaluation: Correlation and heteroskedasticity-consistent tests ☆

Narasimhan Jegadeesh [a,*], Jason Karceski [b,1]

[a] Goizueta Business School Emory University 1300 Clifton Road Atlanta, GA 30322 and NBER, United States
[b] Warrington College of Business University of Florida P.O. Box 117168 Gainesville, FL 32611-7168, United States

ARTICLE INFO

Article history:
Received 5 June 2007
Accepted 11 June 2008
Available online 19 June 2008

JEL classification:
C1
G1

Keywords:
Long horizon performance
Small sample distribution
Specification tests

ABSTRACT

Although there is an extensive literature that evaluates long-run stock returns, the statistical tests that are commonly used are misspecified when event firms share common characteristics. For example, industry clustering or overlapping returns in the sample contribute to test misspecification. We propose a new test of long-run performance that allows for heteroskedasticity and autocorrelation. Our tests are well-specified in random samples and in samples with industry clustering and with overlapping returns.

1. Introduction

One of the most significant challenges to the efficient market hypothesis comes from studies of long-run performance following important corporate events. Several studies document evidence that stocks underperform their benchmarks in the long-run following new equity issues, stock mergers, and dividend omissions. Other studies have found that stocks outperform their benchmarks following stock splits and dividend initiations.[2] Although the magnitudes of underperformance or outperformance that these papers document are non-trivial, Barber and Lyon (1997), Kothari and Warner (1997) and others question the reliability of the methods that these papers use for statistical inference. Specifically, they show that in simulations, empirical rejection levels significantly exceed corresponding theoretical levels of significance.

Lyon et al. (1999) (henceforth LBT) evaluate various tests of long-run abnormal returns. In a traditional event study framework, they recommend that researchers use either a bootstrapped skewness-adjusted *t*-statistic, or the distribution of mean abnormal returns of pseudoportfolios generated by simulation for statistical inference. Many recent papers on long-run stock performance follow their recommendations. Their methods perform well in random samples of firms. However, as LBT caution, "misspecification in nonrandom samples is *pervasive*" (emphasis added) with these methods.[3]

☆ We thank participants at the International Conference on Modeling, Optimization and Risk Measurement in Finance for helpful comments.
* Corresponding author. Tel.: +1 404 727 4821.
E-mail addresses: Narasimhan_Jegadeesh@bus.emory.edu (N. Jegadeesh), jason.karceski@cba.ufl.edu (J. Karceski).
[1] Tel.: +1 352 846 1059.
[2] See Ritter (1991), Loughran and Ritter (1995), Spiess and Affleck-Graves (1995), Jegadeesh (2000), Purnanandam and Swaminathan (2004), and Krishnan and Laux (2005) for evidence on new issues, Loughran and Vijh (1997), Michaely et al. (1995), and Boehme and Sorescu (2002) for dividend initiations and omissions, and Ikenberry et al. (1995) and Chan et al. (2004) for repurchases.
[3] Mitchell and Stafford (2000) also observe that this methodology is "severely flawed."

0927-5398/$ – see front matter
doi:10.1016/j.jempfin.2008.06.002

An important reason for the misspecification is that the LBT approach implicitly 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. In nonrandom samples where the returns for event firms are positively correlated, the variability of the test statistics is larger than in a random sample. Therefore, if the empiricist follows the LBT approach and calibrates the distribution of the test statistics in random samples and uses the empirical cutoff points for nonrandom samples, the tests reject the null hypothesis of no abnormal performance too often.

Such excess rejection is of major concern for event studies because event firms tend to share similar characteristics, and their returns are correlated. For example, new equity issues are concentrated among technology firms in certain years, and in oil and gas or other industries in other years. Such industry concentration induces correlation among the returns of event firms and hence the assumption of nonrandom samples is violated. Therefore, the tests that are currently used could lead to mistaken inferences in practical applications.

This paper proposes test statistics that are well-specified in nonrandom samples. We recommend a *t*-statistic that is computed using a generalized version of the Hansen and Hodrick (1980) standard error. Our generalization allows for heteroskedasticity, and autocorrelation, and for weights to differ across observations. These generalizations accommodate samples that may contain different number of firms in different months and samples with different types of stocks at different times within the sample period.

We evaluate the performance of the tests we recommend in samples that are concentrated in particular industries, and also in samples where event firms enter the sample on multiple occasions within the holding period. In such situations, LBT note pervasive misspecifications when their test statistics are used. In contrast, we find that the distributions of the test statistics that we propose in nonrandom samples are much closer to the corresponding distribution in random samples than the conventional *t*-statistics. For instance, in industry-clustered samples, the size of the conventional *t*-statistic at the 5% level of significance is 15.5% for a three-year holding period with matched-firm benchmarks, and the corresponding size for the serial correlation consistent *t*-statistic is 6.84%. The serial correlation consistent tests are reasonably well-specified in both random samples and nonrandom samples.

Our approach is robust because the tests allow for correlations across observations. Therefore, the standard error reflects the properties of the sample and will be larger if a sample were concentrated in any particular industry, or if the sample were tilted towards any unobservable common factors. As a result, the distributions of the test statistics in nonrandom samples are similar to that in random samples.

One drawback with the approach that we propose is that it is less powerful than the conventional *t*-statistic in nonrandom samples. For instance, with 5% per year abnormal returns and matched-firm benchmarks, the rejection rate for a three-year holding period is 44.8% with the conventional *t*-statistic but is only 38.5% with the serial correlation consistent *t*-statistics. There are larger power differences between tests that use the conventional *t*-statistic and those that use the serial correlation consistent *t*-statistics with portfolio benchmarks.

The rest of the paper is organized as follows. Section 1 presents the methodology that we propose and Section 2 describes our simulation procedure. Section 3 presents simulation results in random samples and examines robustness in nonrandom samples. Section 4 evaluates the power of the tests, and Section 5 concludes the paper.

2. Data and test statistics

Many of the papers in the long horizon literature use the methodology advocated by LBT.[4] To facilitate direct comparison with the LBT methodology, we conduct our experiments using the same data and sample period as LBT. Our sample comprises NYSE, Amex, and Nasdaq stocks, and our sample period is July 1973 through December 1994. We exclude ADRs, closed-end funds, and REITs. We obtain monthly returns and market capitalization data from CRSP, and book value data are from Compustat (data item 60).

We follow the LBT methodology to construct 70 size and book-to-market reference portfolios. Specifically, we first assign stocks to size deciles based on the market value of equity at the end of June each year using CRSP end-of-month prices and total shares outstanding. We then use NYSE size decile breakpoints to assign firms to each of the size categories. We place Amex and Nasdaq stocks in the appropriate NYSE size decile based on their end-of-June market capitalizations. Because many small firms listed on Nasdaq fall in the small firm decile, we further partition it into quintiles using size breakpoints based on all NYSE, Amex, and Nasdaq companies in this decile. This procedure yields a total of 14 size categories. We further subdivide each size category into book-to-market quintiles based on the book-to-market ratios (BM) of the stocks as of the end of the previous calendar year. We use the most recent book value of equity as of December in a particular year and the market value of equity at the end of December to compute BM. We evenly divide each of the 14 size-based portfolios into five portfolios based on the previous December's BM. This procedure gives us a total of 70 size/BM reference portfolios. We reclassify firms into various size and BM categories at the end of June every year.

Our sample comprises *N* "event" firms, which we randomly choose for each simulation. We wish to test whether the event firms exhibit abnormal return performance from the event date through an *H*-month holding period. We define *H*-month abnormal return for stock *i* that enters the sample in event month *t* as:

$$\mathrm{AR}_i(t,H) = \prod_{j=t}^{t+H-1}\left(1+R_{ij}\right) - \prod_{j=t}^{t+H-1}\left(1+R_{bj}\right),$$

[4] For example, see Mitchell and Stafford (2000), Eberhart and Siddique (2002), Boehme and Sorescu (2002), Gompers and Lerner (2003), and Byun and Rozeff (2003).

where $R_{i,t}$ is the return for stock i in month t, and $R_{b,t}$ is the return in month t for the firm's benchmark b. We consider three holding periods: 1 year, 3 years, and 5 years. Therefore, in our simulation experiments, H equals 12, 36, or 60.

We use two types of benchmarks for each test statistic. The first benchmark is the size/BM portfolio that includes the event firm on the event date. The second benchmark is a size/BM-matched individual control firm. We identify all firms with an equity market capitalization between 70% and 130% of the event firm at the most recent end of June. From this set of firms, we select the one with BM closest to the event firm as of the previous December. If a matching firm is delisted sometime during the H-month period, the next closest firm by BM on the event date replaces the delisted firm. If the event firm is delisted prior to the conclusion of the H-month period, we compute the abnormal return only for the time period when the event firm has valid stock returns.

2.1. Conventional t-statistic

To test the null hypothesis that the average abnormal return for these N event firms equals zero, the conventional t-statistic is

$$t = \frac{\overline{\mathrm{AR}}_{\mathrm{sample}}(H)}{\text{standard error}},$$

where $\overline{\mathrm{AR}}_{\mathrm{sample}}(H) = \frac{1}{N}\sum_{i=1}^{N} AR_i(t,H)$ and

$$\text{standard error} = \sqrt{\frac{\frac{1}{N-1}\sum_{i=1}^{N}\left[\mathrm{AR}_i(t,H) - \overline{AR}_{\mathrm{sample}}(H)\right]^2}{N}}. \tag{1}$$

LBT examine the performance of this conventional t-test in small samples. They also examine the small sample performances of a skewness-adjusted t-statistic and a p-value test based on the distribution of abnormal returns in random samples. All these three statistics perform similarly in specification tests. Therefore, to conserve space, we report the results of the specification tests only for the conventional t-test.

2.2. Serial correlation and heteroskedasticity-consistent test statistics

Let N_t equal the number of stocks in the sample in month t, and let N be the total number of stocks in the sample. Therefore,

$$N = \sum_{t=1}^{T} N_t,$$

where T is the number of months in the sample period.

Define the average abnormal holding period return across all event firms that enter the sample on calendar month t (we refer to this group of firms as "month t cohort") as

$$\overline{\mathrm{AR}}(t,H) = \begin{cases} \frac{1}{N}\sum_{i=1}^{N_t} \mathrm{AR}_i(t,H) \text{ if } N_t > 0 \\ 0, \text{ otherwise}, \end{cases}$$

Let $\overline{\mathrm{AR}}(H)$ be a $T\times 1$ column vector where the tth element equals $\overline{\mathrm{AR}}(t,H)$. $\overline{\mathrm{AR}}(t,H)$ is the average long-run abnormal return of each monthly cohort. For example, if there were three stocks in month $t=10$, then $\overline{\mathrm{AR}}(t=10, H=36)$ would be the average 36-month abnormal return of these three event firms.

Define w as a $T\times 1$ column vector of weights where the tth element is the ratio of the number of events that occur in month t divided by N. Specifically,

$$w(t) = \frac{N_t}{N}.$$

The sample average abnormal return is equal to the monthly weight vector w times the average abnormal return of each monthly cohort:

$$\overline{\mathrm{AR}}_{\mathrm{sample}}(H) = w'\overline{\mathrm{AR}}(H).$$

The variance of $\overline{\mathrm{AR}}_{\mathrm{sample}}(H)$ is given by

$$\text{variance}\left[\overline{\mathrm{AR}}_{\mathrm{sample}}(H)\right] = w'Vw, \tag{2}$$

where V is the $T\times T$ variance–covariance matrix of AR(H). Because holding periods overlap for monthly cohorts that are less than H months apart, we allow the first H-1 serial covariances of cohort returns to be non-zero, and we set all higher order serial covariances to zero.

We consider two estimators for the variance–covariance matrix V. The first estimator is the Hansen and Hodrick (1980) estimator. This estimator allows for correlation across monthly cohort returns, but assumes homoskedasticity. We denote this estimator as SC_v, and the ijth element of SC_v is

$$\text{SC\_}v_{i,j}=\begin{cases}\sigma^2=\dfrac{1}{T_N}\sum\limits_{\substack{t=1\\N_t>0}}^{T}\overline{\text{AR}}(t,H)^2, & \text{if } i=j\\ \rho_j=\dfrac{1}{T_{N,j}}\sum\limits_{\substack{t=1\\N_t>0\\N_{t+j}>0}}^{T}\left[\overline{\text{AR}}(t,H)\times\overline{\text{AR}}(t+j,H)\right], & \text{if } 1\le|i-j|\le H-1 \text{ and } T_{N,j}\ge 5\\ 0, & \text{otherwise,}\end{cases} \tag{2}$$

where T_N is the number of months that have at least one event (i.e., $N_t>0$), $T_{N,j}$ is the number of months where both month t and month $t+j$ have at least one event (i.e., $N_t>0$ and $N_{t+j}>0$). σ^2 is the variance of H-period abnormal returns of monthly cohorts. ρ_j is the estimator of covariance between the abnormal returns of cohorts that are separated by j months.[5] To guard against large estimation errors in covariances, we require at least five cases where month t and month t+j both have at least one event. If $T_{N,j}<5$, we set the covariance to equal zero.

The autocorrelation-consistent t-statistic that we propose is

$$\text{SC\_}t=\frac{\overline{\text{AR}}_{\text{sample}}(H)}{\sqrt{w'\,\text{SC\_}vw}}.$$

The second estimator for V that we use allows for heteroskedasticity as well as serial correlation, and we denote it as HSC_v. Allowing for heteroskedasticity could be important for at least two reasons. First, stock return volatility varies over time. Second, the number of events each month and the characteristics of event firms, such as their industry and size, typically vary across sample months. These factors may cause the volatility of monthly cohort abnormal returns to change over time.

To take heteroskedasticity into account, we specify the ijth element of HSC_v as

$$\text{HSC\_}v_{i,j}=\begin{cases}\overline{\text{AR}}(i,H)^2, & \text{if } i=j,\\ \text{AR}(i,H)^{*}\text{AR}(j,H), & \text{if } 1\le|i-j|\le H-1\\ 0, & \text{otherwise.}\end{cases} \tag{3}$$

This estimator generalizes White's heteroskedasticity-consistent estimator and allows for serial covariances to be non-zero whenever the holding periods of monthly cohorts overlap. The heteroskedasticity and autocorrelation-consistent t-statistic that we propose is

$$\text{HSC\_}t=\frac{\overline{\text{AR}}_{\text{sample}}(H)}{\sqrt{w'\,\text{HSC\_}vw}}.$$

Both SC_t and HSC_t statistics follow a t-distribution in large samples. However, we do not know their distributions in small samples. Therefore, later sections use bootstrap experiments to examine the small sample properties of these test statistics.

The appendix in LBT considers a test statistic that attempts to explicitly account for cross-sectional correlations using a different approach, but LBT find that approach does not mitigate the misspecification in the test statistics that they recommend. The LBT approach explicitly estimates the covariance between each pair of event firm returns. For a sample size of N firms, this approach requires estimating $\frac{N(N+1)}{2}$ variance and covariance terms. Quite likely, the approach LBT consider fails to address the misspecification problem because it entails the estimation of a large number of parameters.

In contrast, our approach uses estimates of covariances across monthly cohorts with overlapping holding periods, rather than estimates of covariances across individual firms. As a result, the number of parameters that we require is far smaller than the approach that LBT evaluate. In addition, the number of parameters that we need to estimate does not increase with either the sample size or the length of the sample period. For example, the SC_t statistic requires estimating only $H-1$ serial covariances and one variance, regardless of the sample size.

The HSC_t statistic, however, requires the estimation of a larger number of variance and covariance terms because it allows for heteroskedasticity and for covariances to vary through time. However, as in White (1980), we only need to consistently estimate the product in Eq. (2) rather than each of the individual terms in the variance–covariance matrix. Consequently, the larger number of different terms in the HSC_v estimator does not necessarily add to its estimation error.[6]

[5] We could conceivably impose the null hypothesis that the mean equals zero for the covariance estimator, rather use the sample mean in the estimator. However, we found that the power of the test was adversely affected when we imposed the null hypothesis of zero mean. Intuitively, because the covariances are typically positive in our setting, imposing a zero mean tends to increase the sample standard error estimates, which in turn leads to decreased power of the tests.

[6] In fact, we cannot rely on the consistency of the estimate of any of the individual variance or covariance terms in Eq. (3) because we use only one observation to compute each term. See White (1980) for a further discussion of this point in the context of his heteroskedasticity-consistent standard error estimator.

Table 1
Summary of statistical tests

Method description	Benchmark	Test statistic
t-statistic	Size/BM portfolio	$t = \frac{\overline{AR}_{sample}(H)}{\text{standard error}}$
t-statistic	Size/BM-matched control firm	$t = \frac{\overline{AR}_{sample}(H)}{\text{standard error}}$
Serial correlation consistent SC_*t* statistic	Size/BM portfolio	$SC\_t = \frac{\overline{AR}_{sample}(H)}{\sqrt{w' SC\_vw}}$
Serial correlation consistent SC_*t* statistic	Size/BM-matched control firm	$SC\_t = \frac{\overline{AR}_{sample}(H)}{\sqrt{w' SC\_vw}}$
Heteroskedasticity and serial correlation consistent HSC_*t* statistic	Size/BM portfolio	$HSC\_t = \frac{\overline{AR}_{sample}(H)}{\sqrt{w' HSC\_vw}}$
Heteroskedasticity and serial correlation consistent HSC_*t* statistic	Size/BM-matched control firm	$HSC\_t = \frac{\overline{AR}_{sample}(H)}{\sqrt{w' HSC\_vw}}$

This table summarizes the statistical methods that we evaluate. Standard error is the traditional standard error computed across all event firm abnormal returns and is given by Eq. (1) in the main text. SC_*v* is the serial correlation consistent variance–covariance matrix given by Eq. (2) in the main text, and HSC_*v* is the heteroskedasticity serial correlation consistent variance covariance matrix given by Eq. (3) in the main text.

Although the procedure that we illustrate here assigns weights for each monthly cohort proportional to the number of stocks in the cohort, our approach can be adapted to any weighting scheme that the empiricist chooses to use. The proportional weighting scheme that we use is common in the literature (for example, see Loughran and Ritter, 2000). However, Fama (1998) recommends a calendar-time approach where each month is given equal weight, regardless of the number of observations in that month. If the empiricist chooses to use this approach, then the weight for each monthly cohort should be set equal to the inverse of the number of months in the sample period.

3. Small sample distribution of the test statistics

Table 1 summarizes the statistical methods that we examine. The first two methods use the conventional *t*-statistic, first with size/BM-matched portfolio benchmarks and then with size/BM individual firm benchmarks. These methods that LBT recommend are commonly used in the literature, and we present them here for direct comparison. The next four methods use the autocorrelation-consistent test statistics that we propose.

3.1. Bootstrap experiment

This section examines the small sample distribution of these tests based on bootstrap experiments. The experiment first randomly draws, with replacement, 200 event months over the July 1973 to December 1994 sample period. For each of these 200 event months, we randomly draw one stock from the population of all stocks that are active in the CRSP database for that month. We then compute one-, three- and five-year abnormal returns using the size and BM-matched portfolio and the control firm benchmarks. When we use the control firm benchmarks, for each sample of event firms, we generate 1000 sets of control firms. We then compute the test statistics we describe in Table 1 for each holding period for each simulation run. We repeat this experiment 1000 times to generate the distribution of the test statistics.

3.2. Critical values

Table 2 reports the two-tailed critical values for each test statistic. In large samples the test statistics converge to a standard normal distribution, and the critical values for two-tailed tests at the 1%, 5%, and 10% level are 2.33, 1.96, and 1.65, respectively. The results in Table 2 will help us gauge the divergence between the critical values in small samples and their large sample limits.

One of the patterns that is evident in Table 2 is that the critical values are closer to being symmetric with control firm benchmarks that with portfolio benchmarks. For example, the left and right hand side critical values for the conventional *t*-statistic for a three-year holding period are −1.93 and 2.09 with the control firm benchmark. The corresponding cutoffs with control portfolio benchmark are −2.46 and 1.80. This pattern reflects the fact that the abnormal return distributions are symmetric with the control firm benchmark since we randomly draw both the event firms and the control firms. However, the distributions are positively skewed with the portfolio benchmark because individual stock returns exhibit more positive skewness than portfolio returns.

There is a clear ordering of the deviation of the critical value test statistics from the tabulated values across the three test statistics. The critical values of the conventional *t*-statistics are closer to the tabulated values than the corresponding the critical values for the SC_*t*, which in turn are closer than the critical values for the HSC_*t* statistics. For example, critical values at the 5% level for the three-year abnormal returns with control firm benchmarks are −1.93 and 2.09 for the conventional *t*-statistics, −2.19 and 2.35 for the SC_*t*-statistic and −2.93 and 3.25 for the HSC_*t*-statistics.

These results indicate that the deviation from the large sample distribution is related to the restrictiveness of the assumptions about the time-series of return series. Specifically, the conventional *t*-statistic assumes homoskedasticity and no serial correlation while the SC_*t* statistic allows for serial correlation. The HSC_*t* statistic allows for heteroskedasticity while the SC_*t* statistic does not. Consequently, the number of parameters that are estimated to compute the conventional *t*-statistic is less than that for the

Table 2
Critical values in random samples

Statistic	Benchmark	Two-tailed significance level					
		1%		5%		10%	
		Empirical cumulative density function (%)					
		0.5	99.5	2.5	97.5	5.0	95.0
Panel A: One-year ARs							
t-statistic	Size/BM portfolio	−3.02	1.99	−2.37	1.64	−1.98	1.49
t-statistic	Size/BM control firm	−2.69	2.46	−2.07	1.87	−1.74	1.63
SC_*t* statistic	Size/BM portfolio	−3.36	2.35	−2.20	1.70	−1.81	1.40
SC_*t* statistic	Size/BM control firm	−2.60	2.67	−1.84	1.90	−1.59	1.60
HSC_*t* statistic	Size/BM portfolio	−5.12	3.12	−2.80	2.14	−2.31	1.64
HSC_*t* statistic	Size/BM control firm	−4.22	3.67	−2.62	2.54	−2.06	1.95
Panel B: Three-year ARs							
t-statistic	Size/BM portfolio	−3.06	2.27	−2.46	1.80	−2.01	1.60
t-statistic	Size/BM control firm	−2.43	2.66	−1.93	2.09	−1.62	1.71
SC_*t* statistic	Size/BM portfolio	−5.26	3.15	−2.93	2.12	−2.32	1.85
SC_*t* statistic	Size/BM control firm	−3.93	3.70	−2.19	2.35	−1.72	1.80
HSC_*t* statistic	Size/BM portfolio	−9.44	5.80	−3.37	3.12	−2.42	2.16
HSC_*t* statistic	Size/BM control firm	−7.88	6.47	−2.93	3.25	−2.15	2.48
Panel C: Five-year ARs							
t-statistic	Size/BM portfolio	−3.31	2.05	−2.70	1.60	−2.31	1.42
t-statistic	Size/BM control firm	−2.41	2.42	−1.89	1.89	−1.64	1.57
SC_*t* statistic	Size/BM portfolio	−5.55	4.23	−3.61	2.09	−2.75	1.78
SC_*t* statistic	Size/BM control firm	−5.12	4.45	−2.76	2.50	−2.06	2.03
HSC_*t* statistic	Size/BM portfolio	−8.39	4.68	−3.74	2.66	−2.79	1.96
HSC_*t* statistic	Size/BM control firm	−8.51	10.12	−3.11	3.60	−2.29	2.19

This table presents two-tailed critical values for 1000 random samples of 200 firms at various levels of significance. Panels A, B and C present the critical values for tests with one-, three- and five-year holding period abnormal returns. The critical values are determined based on the empirical distribution of the test statistics in 1000 replications.

SC_*t* statistic, which in turn is less than that for the HSC_*t* statistic. Heuristically, the statistics that entail the estimation of more parameters deviate more from the large sample limits within a finite sample of a given length.

The critical values for all statistics are larger than the corresponding large sample limits in most instances. As a result, the use of large sample critical values would overstate the rejection rates under the null hypothesis for all of the tests. Therefore, it is important to use critical values based on the empirical distribution of the test statistics for statistical inference.

4. Test specification

This section examines the specification of various test methodologies. We first consider the case where event firms and event times are randomly drawn. Later, we allow for two types of nonrandomness in the sample. The first constrains the event firms to be clustered within industries. The other nonrandom case includes the same firms in the sample on multiple occasions within a given holding period. In our random sample and in our industry-clustered sample, we do not allow multiple occurrences of the same firm within the holding period under consideration. Therefore, neither of these samples permit overlapping returns for the same event stocks.

4.1. Random sample

Table 3 presents the rejection rates in 1000 simulations with a random sample of 200 stocks. This table reports rejection rates based on both theoretical critical values and also based on the empirical critical values from Table 2. The results here indicate that virtually all of these tests are misspecified when theoretical critical values are used. For example, with size and BM portfolio benchmarks and a three-year holding period, the rejection rates at the 5% level of significance based on tabulated critical values is 4.40% and 0.80% for the conventional *t*-statistic, 6.96% and 3.03% for the SC_*t* statistic and 11.19% and 6.12% for the HSC_*t* statistic. These misspecifications reflect the fact that the empirical critical values deviate from the corresponding theoretical values.

The tests are all generally well-specified when using the empirical critical values from Table 2. For example, with size and BM portfolio benchmarks and a three-year holding period, the rejection rates at the 5% level are 1.70% and 1.70% for the conventional *t*-statistic, 3.13% and 2.52% for the SC_*t* statistic, and 4.22% and 1.80% for the HSC_*t* statistic.

4.2. Industry clustering

It should not be surprising that the tests are well-specified with random samples when we use critical values from Table 2. These critical values were determined based on the distributions of the corresponding test statistics in random samples. So the distributions of the test statistics in the specification tests should be similar to those used for calibration.

Table 3
Size of alternative test statistics in random samples

Part 1: 200 firms		Two-tailed theoretical significance level					
Statistic	Benchmark	1%		5%		10%	
		Theoretical cumulative density function (%)					
		0.5	99.5	2.5	97.5	5.0	95.0
Panel A: One-year ARs							
t-statistic (T)	Size/BM portfolio	1.60	0.00	5.30	0.60	8.70	1.90
t-statistic (T)	Size/BM control firm	0.30	0.40	2.10	2.00	4.40	5.30
t-statistic (E)	Size/BM portfolio	0.40	0.50	2.80	1.90	5.20	2.90
t-statistic (E)	Size/BM control firm	0.20	0.50	1.40	2.60	3.50	5.50
SC_*t* statistic (T)	Size/BM portfolio	0.90	0.10	4.00	1.00	6.40	2.10
SC_*t* statistic (T)	Size/BM control firm	0.40	0.70	1.30	1.80	3.80	3.40
SC_*t* statistic (E)	Size/BM portfolio	0.10	0.30	2.60	1.90	5.40	4.40
SC_*t* statistic (E)	Size/BM control firm	0.40	0.60	2.20	2.00	4.40	3.60
HSC_*t* statistic (T)	Size/BM portfolio	3.91	0.90	7.31	3.61	11.12	6.31
HSC_*t* statistic (T)	Size/BM control firm	1.90	1.40	4.00	4.20	7.10	6.90
HSC_*t* statistic (E)	Size/BM portfolio	0.10	0.60	2.61	2.61	5.31	6.31
HSC_*t* statistic (E)	Size/BM control firm	0.70	0.50	1.90	1.60	3.50	4.30
Panel B: Three-year ARs							
t-statistic (T)	Size/BM portfolio	1.50	0.00	4.40	0.80	7.30	3.10
t-statistic (T)	Size/BM control firm	0.20	0.50	2.70	2.90	5.10	5.20
t-statistic (E)	Size/BM portfolio	0.60	0.30	1.70	1.70	4.10	3.50
t-statistic (E)	Size/BM control firm	0.60	0.50	3.00	1.90	5.30	4.60
SC_*t* statistic (T)	Size/BM portfolio	4.04	1.51	6.96	3.03	10.09	5.55
SC_*t* statistic (T)	Size/BM control firm	2.43	1.82	5.67	4.36	8.41	6.79
SC_*t* statistic (E)	Size/BM portfolio	0.40	0.91	3.13	2.52	4.64	3.73
SC_*t* statistic (E)	Size/BM control firm	0.41	0.51	4.15	2.33	7.50	5.47
HSC_*t* statistic (T)	Size/BM portfolio	7.60	3.17	11.19	6.12	13.52	8.34
HSC_*t* statistic (T)	Size/BM control firm	4.55	3.10	8.54	5.88	11.64	8.76
HSC_*t* statistic (E)	Size/BM portfolio	0.95	0.53	4.22	1.80	8.34	4.65
HSC_*t* statistic (E)	Size/BM control firm	0.44	0.00	3.33	1.88	6.98	3.10
Panel C: Five-year ARs							
t-statistic (T)	Size/BM portfolio	3.20	0.00	7.20	0.40	11.60	1.50
t-statistic (T)	Size/BM control firm	0.30	0.20	3.10	1.20	6.30	2.80
t-statistic (E)	Size/BM portfolio	0.80	0.30	2.10	1.70	4.70	3.90
t-statistic (E)	Size/BM control firm	0.80	0.20	3.90	1.40	6.40	3.60
SC_*t* statistic (T)	Size/BM portfolio	6.71	1.55	10.94	3.61	13.73	5.68
SC_*t* statistic (T)	Size/BM control firm	3.13	3.34	7.63	3.54	12.02	7.84
SC_*t* statistic (E)	Size/BM portfolio	1.03	0.31	2.99	3.10	5.26	4.44
SC_*t* statistic (E)	Size/BM control firm	0.31	0.73	2.82	3.45	6.69	5.43
HSC_*t* statistic (T)	Size/BM portfolio	9.74	4.01	13.97	6.99	17.64	9.97
HSC_*t* statistic (T)	Size/BM control firm	5.87	3.32	10.97	6.63	15.31	9.31
HSC_*t* statistic (E)	Size/BM portfolio	0.69	1.37	3.78	3.67	8.48	6.99
HSC_*t* statistic (E)	Size/BM control firm	0.26	0.26	2.93	1.91	7.78	5.23

This table presents the percentages of 1000 random samples of 200 firms that reject the null hypothesis of no abnormal returns over one-year (Panel A), three-year (Panel B), and five-year (Panel C) holding periods. Table 1 presents the description of the test statistics. "(T)" indicates that critical values are based on the tabulated distribution of the test statistics, and "(E)" indicates that critical values are based on the empirical distribution of the test statistics based on 1000 replications.

The specification of various tests becomes an issue only when we consider more general scenarios where event firms come from a nonrandom sample but the empirical distribution is calibrated using a random sample. Such scenarios are more realistic because event firms tend to cluster around common factors that are not always observable. For instance, event firms may be concentrated in stocks that are sensitive to interest rates, but the empiricist may not be aware of this concentration. However, because it is hard to specify all characteristics that particular event firms share, the empirical tests in the literature calibrate critical values based on random samples and use them to evaluate the performance of event firms.

This section evaluates the performance of various tests in samples that are concentrated in particular industries. Industry concentration is a special case of samples with stocks that have common characteristics, and the specification tests with such samples will help us evaluate the robustness of various test methodologies. To facilitate comparison, we follow the LBT methodology to form industry-clustered samples.

Each simulation begins with a randomly selected two-digit SIC code. All firms in the sample are required to have this two-digit SIC code. Once a firm enters the sample, we do not allow it to reenter until the end of the holding period. If we run out of firms in the randomly selected industry, we select another two-digit SIC code to complete the sample. If we do not find 200 firms with these two SIC codes, we discard all event firms and start the simulation over. On average, 90% of our simulations include a single industry, and the remaining 10% include two industries.

Table 4
Size of alternative test statistics in industry-clustered samples

Part 1: 200 firms		Two-tailed theoretical significance level					
Statistic	Benchmark	1%		5%		10%	
		Theoretical cumulative density function (%)					
		0.5	99.5	2.5	97.5	5.0	95.0
Panel A: One-year ARs							
t-statistic	Size/BM portfolio	2.30	2.70	7.60	8.10	10.60	11.40
t-statistic	Size/BM control firm	1.40	1.50	5.80	4.80	8.80	6.80
SC_*t* statistic	Size/BM portfolio	0.60	0.90	4.00	4.70	7.90	7.90
SC_*t* statistic	Size/BM control firm	1.00	1.10	3.90	3.30	7.50	6.00
HSC_*t* statistic	Size/BM portfolio	0.40	0.50	3.20	4.00	6.40	7.30
HSC_*t* statistic	Size/BM control firm	0.60	0.30	3.12	2.71	6.03	4.92
Panel B: Three-year ARs							
t-statistic	Size/BM portfolio	4.90	6.40	12.20	11.20	17.40	15.00
t-statistic	Size/BM control firm	4.20	1.70	9.40	6.10	13.30	9.00
SC_*t* statistic	Size/BM portfolio	1.40	0.90	4.91	4.81	8.32	8.53
SC_*t* statistic	Size/BM control firm	0.91	0.50	4.43	2.41	7.75	4.73
HSC_*t* statistic	Size/BM portfolio	0.31	1.36	3.77	4.29	5.55	7.96
HSC_*t* statistic	Size/BM control firm	0.33	0.33	3.32	2.10	7.09	4.98
Panel C: Five-year ARs							
t-statistic	Size/BM portfolio	7.70	7.80	16.80	13.90	21.90	17.10
t-statistic	Size/BM control firm	4.40	2.10	13.00	5.70	18.00	9.10
SC_*t* statistic	Size/BM portfolio	0.82	1.02	5.00	4.18	10.41	6.94
SC_*t* statistic	Size/BM control firm	0.92	0.21	5.13	1.85	9.75	4.41
HSC_*t* statistic	Size/BM portfolio	0.44	0.33	4.55	2.77	8.43	5.54
HSC_*t* statistic	Size/BM control firm	0.13	0.50	4.40	1.26	7.66	3.02

This table presents the percentages of 1000 industry-clustered samples of 200 firms that reject the null hypothesis of no abnormal returns over one-year (Panel A), three-year (Panel B), and five-year (Panel C) holding periods. The statistics and benchmarks are described in detail in the main text. All firms are selected from two randomly chosen industries (based on two-digit SIC coddles). The critical values are reported in Table 2.

Table 4 presents the results for industry-clustered samples. In this table and all subsequent tables, we report only tests that use empirical critical values since the tabulated values do not accurately capture the small sample distribution even in the case of random samples.

The rejection rates with the conventional *t*-tests are substantially larger than the theoretical levels. For instance, for a three-year holding period and control firm benchmarks, the total rejection rate at the 5% level of significance is 9.40% and 6.10% with the conventional *t*-statistic.[7] The level of misspecification generally increases as the holding period increases. For instance, the total rejection rates at the 5% significance level with the control firm benchmarks are 10.60%, 15.50% and 18.70% for one-, three- and five-year holding periods. In unreported tests we also found similar results for the skewness-adjusted *t*-statistic and the *p*-value tests proposed by LBT.

The level of misspecification increases substantially with control portfolio benchmarks. For instance, for a five-year holding period, the rejection rates are 30.70% with the control portfolio benchmark compared with 18.70% for the control firm benchmark. These results have important implications for the choice of benchmarks in empirical tests. LBT report that tests with control portfolio benchmarks are more powerful than those with control firm benchmarks in random samples. However, the level of misspecification would be larger with the control firm benchmark in nonrandom samples.

The rejections rates with the SC_*t*-statistics are far less than the rejection rates with the conventional *t*-statistics. For example, at the 5% level of significance with control firm benchmarks, the total rejection rates with the SC_*t* statistic are 7.20%, 6.84%, and 6.98% with one-, three- and five-year holding periods.

These rejection rates, however, are greater than the theoretical rejection rates. Although the rejection rates are larger than the theoretical size, they do not increase with the holding period. Here again, the rejection rates are larger with the control portfolio benchmark than with the control firm benchmark. However, the differences in the rejection rates with these two benchmarks are only about 2%.

In marked contrast with the other tests, the HSC_*t* tests with the control firm benchmarks are well-specified for all holding periods. For instance, the rejection rates at the 5% significance level with the control firm benchmark are 5.83%, 5.42% and 5.66%. The rejection rates with control portfolio benchmarks are larger than the theoretical level by about 2% to 3%.

Intuitively, the reason why the empirical critical values for the conventional *t*-statistics are misspecified in industry-clustered samples is because they do not account for the fact that standard deviations are larger in these samples than in random samples. Therefore, the critical values for *t*-statistic that are calibrated using random samples are not appropriate when the sample is concentrated within industries.

[7] These rejection rates are somewhat larger than those reported by LBT. One possible reason for this difference is that our industry clustering may be more concentrated than LBT's. LBT do not report how often their 200-firm sample is drawn from one industry versus two industries.

The SC_*t* tests and the HSC_*t* tests, on the other hand, are relatively robust because they compute standard errors using sample return data. Therefore, the standard error estimates reflect the properties of the sample, and they would be larger if a sample is concentrated in any particular industry, or more generally, if it is tilted towards any unobservable common factors. As a result, the distributions of these two tests are closer to their random sample counterparts.

The HSC_*t* statistics account for heteroskedasticity while the other statistics do not. Evidently, ignoring heteroskedasticity inherent in the data contributes to test misspecification. Therefore, among all these three tests, the HSC_*t* test yields the most reliable inference.

4.3. Overlapping returns

Another instance where conventional *t*-tests are misspecified is when the same event firms enter the sample on multiple occasions during the holding period. Here again, because holding periods overlap for repeat firms, standard errors in a random sample understate the true standard errors. This subsection examines the performance of the autocorrelation-consistent test statistics we propose when there are repeat observations in the sample.

To construct the sample with overlapping returns, we randomly select $\frac{N}{2}$ firms from the population, each with its own event month τ. Then for each of these firms, we randomly choose a second event month between $\tau-(H-1)$ and $\tau+(H-1)$, where $H=12$, 36, or 60 depending on the holding period of the experiment. The second event month guarantees that the same firm is in the sample twice, with at least one month of overlapping returns. We repeat this procedure in each of our 1000 simulations, and Table 5 reports the rejection rates.

The general pattern of misspecification for this overlapping sample is similar to that of the industry-clustered sample, although the level of misspecification here is smaller. For instance, the total rejection rate at the 5% level of significance in Table 5 for the conventional *t*-test is 12.10% with control firm benchmarks and a three-year holding period, compared with 15.60% in Table 4. However, the rejection rates for the *t*-test do not increase with the holding period.

The level of misspecification is fairly small with the SC_*t* statistic and the HSC_*t* statistic. For instance, at the 5% level of significance, the total rejection rates are less than 6% in virtually all instances with both statistics. Therefore, both these statistics are reasonably well-specified for the overlapping sample.

Overall, these results indicate that return correlation across sample firms due to common factors such as industry concentration would lead to more severe misspecification of the conventional *t*-test than that due to overlapping samples. Intuitively, the misspecification is not as severe in the overlapping sample because each return observation in the sample is correlated with only

Table 5
Size of alternative test statistics in samples with overlapping returns

Statistic	Benchmark	Two-tailed theoretical significance level					
		1%		5%		10%	
		Theoretical cumulative density function (%)					
		0.5	99.5	2.5	97.5	5.0	95.0
Panel A: One-year ARs							
t-statistic	Size/BM portfolio	1.40	3.00	4.80	8.20	7.40	12.30
t-statistic	Size/BM control firm	1.00	1.50	4.70	4.90	7.80	8.60
SC_*t* statistic	Size/BM portfolio	0.50	1.00	3.00	4.40	5.90	8.10
SC_*t* statistic	Size/BM control firm	0.30	0.80	2.10	3.20	5.10	5.70
HSC_*t* statistic	Size/BM portfolio	0.10	0.20	2.50	3.40	4.20	6.90
HSC_*t* statistic	Size/BM control firm	0.10	0.50	1.80	2.40	3.30	4.60
Panel B: Three-year ARs							
t-statistic	Size/BM portfolio	1.50	2.90	5.30	8.30	8.60	11.10
t-statistic	Size/BM control firm	1.50	2.00	6.70	5.40	10.10	9.20
SC_*t* statistic	Size/BM portfolio	0.40	0.40	2.51	3.71	5.82	6.93
SC_*t* statistic	Size/BM control firm	1.01	0.71	3.83	4.24	7.06	6.56
HSC_*t* statistic	Size/BM portfolio	0.31	0.63	2.09	3.45	4.60	6.17
HSC_*t* statistic	Size/BM control firm	0.23	0.23	2.82	2.25	5.74	4.73
Panel C: Five-year ARs							
t-statistic	Size/BM portfolio	1.50	2.90	5.00	6.70	8.60	10.10
t-statistic	Size/BM control firm	1.30	1.20	4.90	4.20	8.60	7.60
SC_*t* statistic	Size/BM portfolio	0.51	0.82	2.36	3.49	4.93	6.47
SC_*t* statistic	Size/BM control firm	0.21	0.41	2.99	2.47	6.69	5.87
HSC_*t* statistic	Size/BM portfolio	0.80	0.80	2.86	3.78	5.15	5.84
HSC_*t* statistic	Size/BM control firm	0.13	0.38	2.38	3.63	5.75	5.25

This table presents the percentages of 1000 samples of 200 firms with overlapping returns that reject the null hypothesis of no abnormal returns over one-year (Panel A), three-year (Panel B), and five-year (Panel C) holding periods. The statistics and benchmarks are described in detail in the main text. Each sample is constructed in two steps. First, 100 event months in are randomly selected. For each of these event months, a second event month is selected within 12 months of the original event month (either before or after) to guarantee the presence of overlapping returns. The critical values are reported in Table 2.

Table 6
Power of alternative tests in random samples

Statistic	Benchmark	Induced level of annual abnormal return						
		−15%	−10%	−5%	0	5%	10%	15%
Panel A: One-year ARs								
t-statistic	Size/BM portfolio	84.6	59.2	22.4	5.0	31.9	84.8	99.6
t-statistic	Size/BM control firm	74.0	44.9	17.5	5.0	11.9	39.1	69.9
SC_*t* statistic	Size/BM portfolio	80.4	55.9	22.0	5.0	23.3	74.1	98.0
SC_*t* statistic	Size/BM control firm	57.3	28.9	8.3	5.0	5.8	24.4	51.2
HSC_*t* statistic	Size/BM portfolio	76.3	51.9	19.6	5.0	21.8	69.9	96.9
HSC_*t* statistic	Size/BM control firm	47.0	23.4	6.8	5.0	16.3	38.3	52.3
Panel B: Three-year ARs								
t-statistic	Size/BM portfolio	90.1	69.2	28.2	5.0	52.1	99.2	100.0
t-statistic	Size/BM control firm	89.1	64.0	24.4	5.0	20.4	61.1	89.3
SC_*t* statistic	Size/BM portfolio	80.0	57.2	21.5	5.0	35.0	88.0	98.2
SC_*t* statistic	Size/BM control firm	79.2	53.5	19.6	5.0	18.9	50.7	78.2
HSC_*t* statistic	Size/BM portfolio	69.0	47.3	19.1	5.0	19.2	56.9	85.7
HSC_*t* statistic	Size/BM control firm	61.8	35.4	13.0	5.0	13.1	30.0	51.8
Panel C: Five-year ARs								
t-statistic	Size/BM portfolio	84.2	64.1	28.6	5.0	48.3	98.4	100.0
t-statistic	Size/BM control firm	86.1	59.5	22.8	5.0	16.2	52.5	78.7
SC_*t* statistic	Size/BM portfolio	73.0	49.6	20.3	5.0	35.4	85.4	97.7
SC_*t* statistic	Size/BM control firm	77.9	54.2	22.8	5.0	18.9	47.1	72.7
HSC_*t* statistic	Size/BM portfolio	50.6	34.5	15.7	5.0	22.9	52.2	81.1
HSC_*t* statistic	Size/BM control firm	50.8	31.1	14.0	5.0	10.5	24.5	37.3

This table presents the percentages of 1000 samples of 200 firms that reject the null hypothesis of zero abnormal returns over one-year (Panel A), three-year (Panel B), and five-year (Panel C) holding periods. We choose the sample firms randomly and then add the levels of annual abnormal return indicated in the column heading. The statistics and benchmarks are described in detail in the main text. The critical values are reported in Table 2.

one other observation. However, in the case of industry concentration, return observations are correlated with returns for all sample firms within the same industry.

5. Power

This section compares the power of various tests with the three test statistics. We confine our power analysis to the random sample case since we know that the *t*-tests are misspecified in nonrandom samples.

To evaluate power, we add a constant level of abnormal return, ranging from −15% per year to 15% per year in increments of 5% to the return of each sample firm. Table 6 reports the rejection rates at the 5% significance level for one-, three- and five-year holding periods. The null hypothesis is that the mean sample long-run abnormal return is zero.

Table 6 reports the rejection rates for various tests under the alternative hypotheses. The conventional *t*-test is the most powerful test, followed by the SC_*t* test, and the HSC_*t* test is the least powerful. For instance, with −10% per year abnormal returns and control firm benchmarks, the rejection rate for a three-year holding period is 64.0% with the conventional *t*-statistic, 53.5% with the SC_*t* and 35.4% with the HSC_*t* statistics. There are larger power differences between tests using the conventional *t*-statistic and those using the serial correlation consistent *t*-statistics with portfolio benchmarks.

Therefore, the improved specification of the tests with less restrictive assumptions comes at a cost — loss of power. As we discussed earlier, the number of parameters that we need to estimate increases as we make less restrictive assumptions. The most general HSC_*t* test entails the estimation of the largest number of parameters and the most restrictive *t*-test the smallest. Heuristically, each additional parameter that is estimated from the data adds to the noise, and thereby adversely affect the power of the tests.

6. Conclusions

A number of papers in the literature examine long-run performance following events such as new issues, stock repurchases and stock splits to examine whether the market reacts efficiently to these events. Recent long-run performance studies typically use the tests recommended by Lyon et al. (1999) to test the hypothesis that abnormal returns following the events are not different from zero. However, the LBT tests are misspecified when the event firms are not randomly drawn.[8]

To address the shortcomings of the LBT methodology, this paper proposes two new autocorrelation-consistent test statistics. We find that both these tests are well-specified in samples of firms that share common characteristics. In particular, we find that

[8] Many papers also use a calendar-time portfolio formation approach that buys and holds the event stocks at monthly intervals, and thereby avoid the issue of overlapping returns. However, this approach computes monthly returns and not buy-and-hold returns that are typically of interest. Loughran and Ritter (2000) discuss additional shortcomings of the calendar-time approach.

the tests are well-specified when the sample is concentrated in certain industries and when the sample contains the same observations on multiple occasions.

The serial correlation consistent (SC_*t*) and the heteroskedasticity and serial correlation consistent (HSC_*t*) tests, however, have lower power than the conventional *t*-test in random samples. The empiricist, therefore, is faced with a dilemma when deciding on which test to use for evaluating long horizon performance. Should he choose the HSC_*t* test for its robustness or should he choose the conventional *t*-test for its power in random samples?

Since many common characteristics that are shared by different stocks are not readily observable, it would require a leap of faith to assume that a sample of event firms is randomly drawn. Therefore, we recommend that the empiricist use either the SC_*t* test or the HSC_*t* test for their robustness. Even if the empiricist firmly believes that the sample is randomly drawn, it would be useful to report the SC_*t* and the HSC_*t* statistics so that readers can reach their own conclusion about the strength of the results.

References

Barber, Brad M., Lyon, John D., 1997. Detecting long-run abnormal stock returns: the empirical power and specification of test statistics. Journal of Financial Economics 43, 341–372.

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

Byun, Jinho, Rozeff, Michael S., 2003. Long-run performance after stock splits: 1927 to 1996. Journal of Finance 58, 1063–1085.

Chan, Konan, Ikenberry, David, Lee, Inmoo, 2004. Economic sources of gain in stock repurchases. Journal of Financial and Quantitative Analysis 39, 461–479.

Eberhart, Allan C., Siddique, Akhtar, 2002. The long-term performance of corporate bonds (and stocks) following seasoned equity offers. Review of Financial Studies 15, 1385–1406.

Fama, Eugene F., 1998. Market efficiency, long-term returns, and behavioral finance. Journal of Financial Economics 49, 285–306.

Gompers, Paul A., Lerner, Josh, 2003. The really long-run performance of initial public offerings: the pre-Nasdaq evidence. Journal of Finance 58, 1355–1392.

Hansen, Lars Peter, Hodrick, Robert J., 1980. Forward exchange rates as optimal predictors of future spot rates: an econometric analysis. Journal of Political Economy 88, 829–853.

Ikenberry, David, Lakonishok, Josef, Vermaelen, Theo, 1995. Market underreaction to open market share repurchases. Journal of Financial Economics 39, 181–208.

Jegadeesh, N., 2000. Long-term performance of seasoned equity offerings: benchmark errors and biases in expectations. Financial Management 29, 5–30.

Kothari, S.P., Warner, Jerold B., 1997. Measuring long-horizon security price performance. Journal of Financial Economics 43, 301–339.

Krishnan, C.N.V., Laux, Paul A., 2005. Misreaction. Journal of Financial and Quantitative Analysis 40, 403–434.

Loughran, Tim, Ritter, Jay R., 1995. The new issues puzzle. Journal of Finance 50, 23–51.

Loughran, Tim, Vijh, Anand M., 1997. Do long-term shareholders benefit from capital acquisitions? Journal of Finance 52, 1765–1790.

Loughran, Tim, Ritter, Jay R., 2000. Uniformly least powerful tests of market efficiency. Journal of Financial Economics 55, 361–389.

Lyon, John D., Barber, Brad M., Tsai, Chih-Ling, 1999. Improved methods for tests of long-run abnormal stock returns. Journal of Finance 54, 165–201.

Michaely, Roni, Thaler, Richard H., Womack, Kent L., 1995. Price reactions to dividend initiations and omissions: overreaction or drift? Journal of Finance 50, 573–608.

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

Purnanandam, Amiyatosh K., Swaminathan, Bhaskaran, 2004. Are IPOs really underpriced? Review of Financial Studies 17, 811–848.

Ritter, Jay R., 1991. The long-run performance of initial public offerings. Journal of Finance 46, 3–28.

Spiess, D. Katherine, Affleck-Graves, John, 1995. Underperformance in long-run stock returns following seasoned equity offerings. Journal of Financial Economics 38, 243–267.

White, Halbert, 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.