Mark L. Mitchell
Erik Stafford
Harvard University

Managerial Decisions and Long-Term Stock Price Performance*

I.   Introduction

How reliable are estimates of long-term abnormal returns subsequent to major corporate events? The view expressed in recent long-term event studies is that we can precisely identify systematic mispricings of large samples of equity securities for up to 5 years following major corporate decisions. Taken at face value, these findings strongly reject the notion of stock market efficiency. However, this is at odds with the conventional view that the stock market quickly and completely incorporates public information into the stock price. We try to reconcile these two views by reassessing the reliability of recent long-term abnormal return estimates and providing new estimates that are robust to common statistical problems. In particular, we investigate the impact on inferences of several potential, but often overlooked, problems with common methodologies using three large, well-studied samples of major managerial decisions, namely, mergers, seasoned equity offerings (SEOs), and share repurchases, all of which have been the focus of

A rapidly growing literature claims to reject the efficient market hypothesis by producing large estimates of long-term abnormal returns following major corporate events. The preferred methodology in this literature is to calculate average multiyear buy-and-hold abnormal returns and conduct inferences via a bootstrapping procedure. We show that this methodology is severely flawed because it assumes independence of multiyear abnormal returns for event firms, producing test statistics that are up to four times too large. After accounting for the positive cross-correlations of event-firm abnormal returns, we find virtually no evidence of reliable abnormal performance for our samples.

numerous long-term event studies, and each of which has the benefit of a long sample period, 1958–93.

There are several important components to measuring long-term abnormal stock price performance, including an estimator of abnormal performance and a means for determining the distribution of the estimator. Beginning with Ritter (1991), the most popular estimator of long-term abnormal performance is the mean buy-and-hold abnormal return, $\overline{\text{BHAR}}$. Concerns arising from the skewness of individual-firm long-horizon abnormal returns hampered statistical inference in many initial studies, which either avoided formal statistical inference or relied on assumptions that were later questioned, such as normality of the estimator. To address the skewness problem, Ikenberry, Lakonishok, and Vermaelen (1995) introduce a bootstrapping procedure for statistical inference that simulates an empirical null distribution of the estimator, relaxing the assumption of normality.

Several long-term event-study methodology papers have questioned each aspect of measuring long-term abnormal performance. Barber and Lyon (1997) argue that the $\overline{\text{BHAR}}$ is the appropriate estimator because it ''precisely measures investor experience.'' However, Barber and Lyon (1997) and Kothari and Warner (1997) provide simulation evidence showing that common estimation procedures can produce biased $\overline{\text{BHAR}}$ estimates. In particular, biases arise from new listings, rebalancing of benchmark portfolios, and skewness of multiyear abnormal returns. Proposed corrections include carefully constructing benchmark portfolios to eliminate known biases and conducting inferences via a bootstrapping procedure, as applied by Ikenberry et al. (1995). In addition, large sample sizes mitigate many of these biases. The common conclusion of these methodology papers is that ''measuring long-term abnormal performance is treacherous.''

Fama (1998) argues against the BHAR methodology because the systematic errors that arise with imperfect expected return proxies—the bad model problem—are compounded with long-horizon returns. In addition, any methodology that ignores cross-sectional dependence of event-firm abnormal returns that are overlapping in calendar time is likely to produce overstated test statistics. Therefore, Fama strongly advocates a monthly calendar-time portfolio approach for measuring long-term abnormal performance. First, monthly returns are less susceptible to the bad model problem. Second, by forming monthly calendar-time portfolios, all cross-correlations of event-firm abnormal returns are automatically accounted for in the portfolio variance. Finally, the distribution of this estimator is better approximated by the normal distribution, allowing for classical statistical inference.

Despite the apparent attractiveness of the calendar-time portfolio approach, Lyon, Barber, and Tsai (1999) and Loughran and Ritter (in

press) prefer the BHAR methodology. Lyon et al. again argue that the BHAR is the appropriate estimator because it ''accurately represents investor experience'' and that statistical inference should be performed via either a bootstrapped skewness-adjusted *t*-statistic or the bootstrapping procedure of Ikenberry et al. (1995). Loughran and Ritter primarily argue that the calendar-time portfolio approach has low power to detect abnormal performance because it averages over months of ''hot''and ''cold'' event activity. For example, the calendar-time portfolio approach may fail to measure significant abnormal returns if abnormal performance primarily exists in months of heavy event activity.

Following the prescriptions of the methodology papers described above, there is a second wave of long-term event studies that also find large estimates of abnormal performance. The typical approach is to focus on BHARs, using various benchmarks that are carefully constructed to avoid known biases and assessing statistical significance of the $\overline{BHAR}$ via a bootstrapping procedure. The authors conclude that, since they find evidence of long-term abnormal performance even after taking into account the potential problems highlighted by the methodology papers, their results are especially robust. There is some sense that the recent estimates of abnormal performance are perhaps conservative, but since they are still very statistically significant, market efficiency is strongly rejected.

The general findings from the long-term abnormal performance studies are described by figure 1. Typically, the estimated mean buy-and-hold abnormal return, $\overline{BHAR}$, falls far into the tail of the null distribution of the $\overline{BHAR}$, and often it falls well beyond the maximum, or below the minimum, of the bootstrapped null distribution. The methodology papers emphasize that the benchmarks must be carefully constructed to avoid known biases, which can move the $\overline{BHAR}$ in either direction. However, this has little impact on inferences, in practice, for large samples. For our samples, different methods of constructing benchmark portfolios change estimated mean BHARs roughly 20% in either direction. In other words, if the mean BHAR is −10% (as is the case in fig. 1), modifying the benchmark construction produces estimates ranging from −8% to −12%. Although this may appear significant, this is not a meaningful difference since the minimum of the bootstrapped null distribution is only around −4%, producing a *p*-value of 0.000, regardless of which estimate is used. The common inference from results similar to these is that average BHARs following major corporate events are very far from zero and, thus, market efficiency is rejected.

We are suspicious of this interpretation. It is difficult to reconcile extremely precise measurement of abnormal performance when ex-
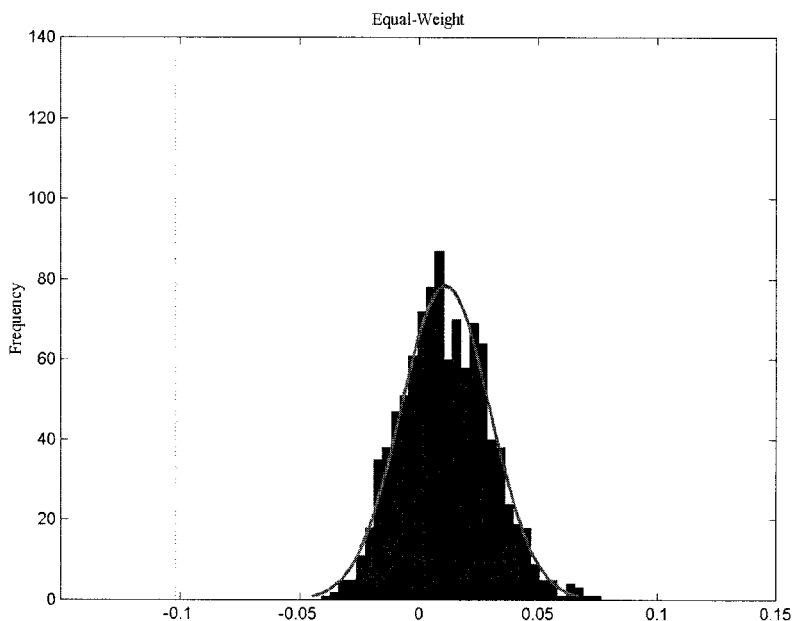
Fig. 1.—Empirical distribution for mean BHAR for seasoned equity issue sample. The histogram plots the empirical distribution of equal-weight average 3-year BHARs for 1,000 bootstrap samples. Each bootstrap sample is created by assigning the completion date of each original sample firm to a randomly selected firm with the same size-BE/ME portfolio assignment at the time of the event. This procedure yields a pseudosample that has the same size-BE/ME distribution, the same number of observations, and the same calendar-time frequency as the original event sample. We then calculate the mean BHAR for this pseudosample in the same way as was done for the original sample. This results in one mean BHAR under the null of the model. We repeat these steps until we have 1,000 mean BHARs and, thus, an empirical distribution of the mean BHAR under the null. A p-value is calculated as the fraction of the mean BHARs from the pseudosamples that are larger in magnitude (but with the same sign) than the original mean BHAR.

pected performance is difficult to determine a priori. Our view of figure 1 is that the popular bootstrapped distribution of 3-year $\overline{\text{BHAR}}$s may be too tight given that we cannot price equity securities very precisely, especially at long horizons, and, thus, it is not clear whether the $\overline{\text{BHAR}}$ is far from zero or not.

Our primary concern with the advocated bootstrapping procedure is that it assumes event-firm abnormal returns are independent. Major corporate actions are not random events, and thus event samples are unlikely to consist of independent 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. We are more comfortable assuming that the mean BHAR is normally distributed with large sample sizes than we are assuming that all multiyear event-firm abnormal returns are independent. The dependence problem increases with sample size, whereas the normality assumption becomes more plausible with large samples.

To gain perspective on the magnitude of the dependence problem, we assume normality of the mean BHAR and impose a simple structure on the covariance matrix to estimate average cross-correlations of 3-year BHARs. The estimated correlations are used to calculate correlation-adjusted standard errors of the $\overline{\text{BHAR}}$ for each of the three samples. We find that the normality assumption for the mean BHAR is reasonable for our three large event samples. However, accounting for dependence has a huge effect on inferences. It is common for *t*-statistics to fall from over 6.0 to less than 1.5 after accounting for cross-correlations. Equivalently, we find that a 3-year $\overline{\text{BHAR}}$ of 15% is not statistically different from zero. In fact, after accounting for the positive cross-correlations of individual-firm BHARs, we find no reliable evidence of long-term abnormal performance for any of the three event samples when using the BHAR approach. It is important that this result is due to accounting for the dependence of individual-event-firm abnormal returns, not due to the construction of the benchmark portfolios. Although our estimated returns are similar in magnitude to those reported in previous studies, our test statistics that allow for dependence are dramatically smaller.

Our results directly contradict the prescriptions of most methodology papers that advocate the BHAR approach in conjunction with bootstrapping. However, the calendar-time portfolio approach advocated by Fama (1998) is robust to the most serious statistical problems. It is interesting to note that the inferences from the calendar-time portfolio approach are quite similar to those from our modified BHAR analysis, after accounting for the positive cross-sectional dependence of event-firm abnormal returns. Moreover, we directly address the concerns raised by Loughran and Ritter (in press), and we find no evidence in their support. In fact, we find that the calendar-time portfolio procedure has more power to identify reliable evidence of abnormal performance in our samples than the BHAR approach, after accounting for dependence. We, like Fama (1998), strongly advocate a calendar-time portfolio approach.

Contemporaneous research by Brav (in press) also emphasizes the problems of long-term abnormal performance methodologies that assume independence. In particular, Brav develops an elaborate Bayesian predictive methodology for measuring long-term abnormal returns that

relaxes the assumption of independence in certain circumstances. How-ever, his approach does not provide a complete correction to the depen-dence problem.

Finally, we show that much of what is typically attributed to an ''event'' is merely a manifestation of known mispricings of the model of expected returns. Following Fama and French (1992, 1993), virtually all recent studies documenting long-term abnormal returns use some form of risk adjustment that assumes the cross-section of expected re-turns can be completely described by size and book-to-market equity. However, this appears to be a misreading of the evidence presented in Fama and French (1993). Although expected returns are systematically related to size and book-to-market attributes, Fama and French point out that three of 25 portfolios formed based on these characteristics are associated with abnormal return estimates that are significantly dif-ferent from zero. In other words, the model of market equilibrium used to identify mispricing cannot completely price the cross-section of ex-pected returns on the dimensions that it is designed to explain. After controlling for the sample composition of our three samples, based on size and book-to-market attributes, reliable evidence of abnormal per-formance is substantially reduced and is restricted to a few subsamples of small stocks. This is consistent with the evidence provided by Brav and Gompers (1997) for initial public offerings (IPOs) and suggests that many of the ''event anomalies'' previously documented by other researchers are actually manifestations of known pricing deficiencies of the model of expected returns. This further highlights the inconsis-tency of being able to reliably identify mispriced assets when asset pricing is imprecise.

II.   Data Description

A.   *Sample Selection*

The datasets for this article consist of three large samples of major managerial decisions, namely, mergers, SEOs, and share repurchases, completed during 1958–93.[1] Each of these event samples has been the focus of numerous recent studies of long-term stock price performance, although data from the 1960s is for the most part unexplored. The use of three large well-explored samples facilitates comparisons of abnormal

---

1. The event samples are from the CRSP-EVENTS database currently under develop-ment at the University of Chicago's Center for Research in Security Prices (CRSP). The CRSP-EVENTS database contains detailed information on mergers and tender offers, pri-mary seasoned equity issues, and stock repurchases from 1958 to the present, where CRSP price data are available at the time of the event announcement. Data sources include corpo-rate annual reports, *Investment Dealer's Digest, Mergers and Acquisitions*, U.S. Securities and Exchange Commission filings, Standard and Poor's Compustat database, *Wall Street Journal* (and Dow Jones News Retrieval Service), and various miscellaneous sources.

performance estimates across both our samples and those of other studies, as well as across various methodologies. In addition, the use of data from the 1960s may be useful in determining how sensitive results are to short sample periods.

The sample consists of 4,911 underwritten primary and combination SEOs, 2,421 open market and tender-offer share repurchases (excluding odd-lot repurchases), and 2,193 acquisitions of the Center for Research in Security Prices (CRSP) firms. We exclude multiple events by the same firm within any 3-year period. In other words, after the first event, we ignore additional events until after the 3-year event window. Our objective is to focus on the long-run price performance of the broad samples rather than to delve into the cross-sectional particulars, so we limit the cross-sectional characteristics of individual transactions to those that previous research has shown to be important. For example, we classify merger transactions based on form of payment. For repurchase events we distinguish between open market repurchases and self-tenders, which is similar to what has been done in prior research.

B. *Announcement and Completion Dates*

In the long-run abnormal performance tests, we begin the multiyear event window at the end of the completion month rather than at the announcement date. For example, the announcement date for SEOs is the registration date, whereas the completion date is the offering date. The average time between the registration and completion date is one month. For mergers, the interval between announcement and completion is typically several months. Since stock repurchase programs rarely provide a definitive completion date and because these programs can take place over a period of a year or two, we treat the announcement date as the completion date. In the empirical tests examining the pre-event long-term abnormal performance, the event window ends in the month before the event announcement.

C. *Returns, Size, and Book-to-Market Equity*

The return data come from the CRSP monthly NYSE, AMEX, and NASDAQ stock files.[2] Firm size refers to the market value of common equity at the beginning of the month. In part of the empirical analysis that follows, we assign sample firms to portfolios based on size and book-to-market equity (BE/ME). We follow Fama and French (1997)

---

2. Shumway (1997) reports that performance-related delistings often have missing final period returns in CRSP. Using other sources, Shumway estimates final period returns for these firms to be -30%. We find that substituting missing performance-related delisting returns with -30% does not alter our findings, so we report results based on the unaltered data.

and define book equity as total shareholders' equity, minus preferred stock, plus deferred taxes (when available), plus investment tax credit (when available), plus postretirement benefit liabilities (when available). Preferred stock is defined as redemption, liquidation, or carrying value (in this order), depending on availability. If total shareholders' equity is missing, we substitute total assets minus total liabilities. Book-to-market equity is the ratio of fiscal year-end book equity divided by market capitalization of common stock at calendar year end. We use the most recent fiscal year-end book equity, as long as it is no later than the calendar year-end market equity. Consequently, if the fiscal year end occurs in January through May, we use book equity from the prior fiscal year.

The event samples begin in 1958 and extend through 1993. Since the book equity data is limited on Compustat during the early part of the sample period, we fill in much of the missing data by hand-collecting book equity from Moody's Manuals (Chan, Jegadeesh, and Lakonishok 1995; Kothari, Shanken, and Sloan 1995). For example, in the fiscal year 1962, we supplement the 974 firms with Compustat book equity data with 769 firms from the Moody's manuals. In all, we replace about 7,000 missing book equity observations.

D.   Distribution of Event Firms by Size and Book-to-Market

In some of the empirical tests to follow, we compare the stock price performance of the event firms to 25 portfolios formed on size and book-to-market quintiles using NYSE breakpoints (see Fama and French 1992, 1993). Table 1 displays the distribution of the event firms according to the size and BE/ME classifications.

All three event samples have size distributions that are tilted toward large firms relative to the population of CRSP firms. Close to 60% of all CRSP firms fall into the bottom size quintile based on NYSE breakpoints. Acquirers are more likely to have relatively low BE/ME ratios and large equity values. The SEO issuers are also more likely to have low BE/ME ratios, and they tend to be small firms relative to the NYSE breakpoints. Share repurchasers are also relatively small firms, and it is interesting that they are more likely to be low BE/ME firms, which suggests that these firms are not typically ''value'' stocks, as some behavioral stories suggest.

The distribution of sample firms does not change very much between the pre- and postevent years. This suggests that there is little systematic change in the firm-level size and book-to-market characteristics following the event. However, this hides the fact that most firms change portfolio assignments following the event. In particular, only 25% of our sample firms are in their original size and book-to-market portfolio 3 years after the event.

TABLE 1   **Distribution of Event Sample Firms by Size and Book-to-Market Equity (July 1961–December 1993)**

| | Book-to-Market | | | | | |
	Low	2	3	4	High	Total
Mergers:						
Preevent:						
Small	3.3	2.1	2.4	2.5	4.0	14.3
2	3.0	2.8	2.9	2.7	2.8	14.2
3	4.2	2.6	3.3	3.5	1.9	15.5
4	5.1	5.3	5.3	4.2	2.6	22.7
Large	11.4	8.5	6.6	5.0	1.8	33.3
Total	27.1	21.3	20.6	17.9	13.1	100.0
Postevent:						
Small	3.0	2.1	2.5	2.0	4.0	13.6
2	2.5	3.2	2.1	2.3	2.4	12.5
3	3.5	3.5	2.9	3.4	1.9	15.2
4	5.3	5.9	6.1	4.4	2.2	23.9
Large	12.0	8.2	7.0	5.1	2.6	34.8
Total	26.3	22.9	20.7	17.1	13.0	100.0
Seasoned equity offerings:						
Preevent:						
Small	16.1	6.0	4.4	4.6	5.5	36.6
2	8.2	3.0	2.9	2.6	2.5	19.3
3	4.3	2.9	2.9	3.2	2.0	15.3
4	3.0	2.3	3.9	3.3	1.7	14.2
Large	2.4	2.6	3.3	4.2	2.1	14.6
Total	34.1	16.8	17.5	17.9	13.7	100.0
Postevent:						
Small	16.5	5.8	3.8	2.8	2.7	31.6
2	10.9	3.9	2.6	2.2	1.6	21.1
3	7.1	2.7	2.7	2.8	1.4	16.6
4	4.3	2.7	3.7	3.2	1.5	15.4
Large	3.0	2.7	3.6	3.7	2.2	15.2
Total	41.7	17.9	16.3	14.7	9.4	100.0
Share repurchases:						
Preevent:						
Small	4.9	6.6	5.4	5.6	10.0	32.6
2	3.7	4.8	3.8	3.6	2.5	18.3
3	4.2	3.1	3.1	2.5	1.9	14.8
4	4.0	3.7	4.1	2.2	1.8	15.8
Large	6.5	4.2	3.7	3.0	1.0	18.5
Total	23.4	22.4	20.1	16.9	17.2	100.0
Postevent:						
Small	3.7	5.5	7.3	7.8	9.7	34.0
2	2.9	4.1	4.1	3.7	3.1	18.0
3	3.7	3.7	3.0	2.6	1.6	14.6
4	3.5	3.9	3.8	2.9	2.0	16.2
Large	5.8	4.2	3.2	2.7	1.3	17.2
Total	19.6	21.3	21.5	19.8	17.8	100.0

TABLE 1   (*Continued*)

	Book-to-Market					
	Low	2	3	4	High	Total
CRSP universe:						
Small	14.5	8.8	8.2	9.7	16.8	57.9
2	4.0	2.9	2.9	2.6	2.4	14.8
3	3.0	2.3	2.2	1.9	1.4	10.7
4	2.6	2.0	1.8	1.5	0.9	8.7
Large	2.7	1.8	1.5	1.2	0.6	7.9
Total	26.9	17.7	16.5	16.8	22.1	100.0

NOTE.—Size and book-to-market equity groupings are based on independent rankings of sample firms relative to NYSE quintiles in both the year prior $(t - 1)$ and subsequent $(t + 1)$ to the completion of the event.

III.   Buy-and-Hold Abnormal Returns (BHARs)

Buy-and-hold abnormal returns have become the standard method of measuring long-term abnormal returns (see Barber and Lyon 1997; Lyon et al. 1999). Buy-and-hold abnormal returns measure 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.

Barber and Lyon (1997) and Lyon et al. (1999) argue that BHARs are important because they ''precisely measure investor experience.'' While it is true that BHARs capture the investor's experience from buying and holding securities for 3–5 years, this is not a particularly compelling reason to restrict attention to this methodology if reliably assessing long-term stock price performance is the objective. First, this is only one type of investor experience—the buy-and-hold experience. There are other reasonable trading strategies that capture other investors' experiences, for example, periodic portfolio rebalancing. Second, because of compounding, the buy-and-hold abnormal performance measure is increasing in holding period, given abnormal performance during any portion of the return series. For instance, if abnormal performance exists for only the first 6 months following an event and if one calculates 3- and 5-year BHARs, both can be significant, and the 5-year BHAR will be larger in magnitude than the 3-year BHAR. This is important to consider since the length of the holding period is arbitrary and various holding period intervals are often analyzed to determine how long the abnormal performance continues after the event. Finally, and most important, we show in the next section that there are serious statistical problems with BHARs that cannot be easily corrected. Since our objective is to reliably measure abnormal returns, it is imperative that the methodology allow for reliable statistical inferences.

A. Calculating BHARs

We calculate 3-year BHARs for each firm in the three event samples using 25 value-weight, nonrebalanced, size-BE/ME portfolios as expected return benchmarks:[3]

$$\text{BHAR}_i = \prod_{t=1}^{T} (1 + R_{i,t}) - \prod_{t=1}^{T} (1 + R_{\text{benchmark},t}), \qquad (1)$$

where the mean buy-and-hold abnormal return is the weighted average of the individual BHARs:

$$\overline{\text{BHAR}} = \sum_{i=1}^{N} w_i \cdot \text{BHAR}_i. \qquad (2)$$

Both equal-weight (EW) and value-weight (VW) averages are computed, where the value weights are based on market capitalizations at event completion, divided by the implicit value-weight stock market deflator. In other words, we standardize market values of the sample firms by the level of the CRSP VW market index at each point in time before determining the weights. This is to avoid the obvious problems with unstandardized value weights, which would weight recent observations much more heavily than early observations. The size-BE/ME portfolio benchmarks are designed to control for the empirical relation between expected returns and these two firm characteristics (see Fama and French [1992, 1993] for discussion and evidence).

   The benchmark portfolios exclude event firms, but otherwise they include all CRSP firms that can be assigned to a size-BE/ME portfolio. At the end of June of each year t, all stocks are allocated to one of five size groups, based on market capitalization rankings relative to NYSE quintiles. In an independent sort, all stocks are also allocated to one of five BE/ME groups, based on where their BE/ME ranks relative to NYSE quintiles. The returns for the 25 portfolios are calculated for the year, defined July of year t through June of year $t + 1$ as the value-weight average of the individual-firm monthly returns in each of the size-BE/ME quintile intersections. To allow for changing firm characteristics, the size-BE/ME benchmarks are allowed to change at the end of June of each year when new portfolio assignments are available. Moreover, the size-BE/ME portfolios are composed of firms that have reported prior BE data, which partially mitigates any bias caused by including recent IPOs in the benchmarks.

---

   3. Employing long-term windows of different lengths does not alter the substantive results of this article.

In calculating the BHARs for the individual firms, we impose two conditions to ensure that all BHARs represent true 3-year buy-and-hold returns. First, because of delistings, not all of the sample firms have a full 3 years of valid return data following the completion of the event. Therefore, we fill in missing sample firm returns with the benchmark portfolio return. Second, in forming the benchmark portfolios, we do not rebalance, so that each BHAR is a true buy-and-hold return. This means that we compute the 3-year returns for each of the 25 size-BE/ ME portfolios each calendar month.

B.   Statistical Inference via Bootstrapping

Since the BHAR is the difference of a sample firm's 3-year return and the 3-year return on a benchmark portfolio, the distribution of individual-firm BHARs is strongly positively skewed (Barber and Lyon 1997) and generally does not have a zero mean. Therefore, statistical inference for the mean BHAR is often based on an empirical distribution simulated under the null of the model as applied by Brock, Lakonishok, and LeBaron (1992) and Ikenberry et al. (1995). Within this framework, the implied model of expected 3-year returns is simply the average 3-year return of firms that have similar size and BE/ME.

Following the methodology of Brock et al. (1992) and Ikenberry et al. (1995), for each sample firm, we assign the completion date to a randomly selected firm with the same size-BE/ME portfolio assignment at the time of the event. This procedure yields a pseudosample that has the same size-BE/ME distribution, the same number of observations, and the same calendar-time frequency as the original event sample. We then calculate the $\overline{\text{BHAR}}$ for this pseudosample in the same way as for the original sample. This results in one $\overline{\text{BHAR}}$ under the null of the model. We repeat these steps to generate 1,000 $\overline{\text{BHAR}}$s and thus an empirical distribution of the $\overline{\text{BHAR}}$ under the null. A p-value is calculated as the fraction of the $\overline{\text{BHAR}}$s from the pseudosamples that are larger in magnitude (but with the same sign) than the original $\overline{\text{BHAR}}$.

C.   Results from the BHAR Analysis Assuming Independence

Tables 2–4 display the BHAR results for the three event samples over the period July 1961 through December 1993. We report both EW and VW results and compare them with previous research when possible.

Mergers.   As displayed in table 2, the 3-year EW $\overline{\text{BHAR}}$ for acquirers is −1%, and it has a p-value of .164. The wealth relative, measured as the average gross return of the event firms divided by the average gross return of the benchmark firms, is 0.994, implying that investing in these acquirers generated total wealth 0.6% less after

TABLE 2   **Three-Year Mean Buy-and-Hold Abnormal Returns (BHARs) for Acquirers (July 1961–December 1993)**

	Sample	Benchmark	Wealth Relative	BHAR	*p*-Value	*n*
Equal weight:						
Postevent	.514	.524	.994	−.010	.164	2,068
Preevent	.881	.597	1.178	.285	.000	1,796
Financed with stock	.393	.477	.943	−.084	.000	1,029
Financed without stock	.634	.570	1.041	.064	.047	1,039
Growth firms	.390	.372	1.013	.018	.481	526
Value firms	.718	.691	1.016	.027	.471	257
Value weight:						
Postevent	.381	.419	.973	−.038	.027	2,068
Preevent	.468	.420	1.034	.048	.320	1,796
Financed with stock	.297	.350	.961	−.053	.009	1,029
Financed without stock	.483	.503	.986	−.021	.291	1,039
Growth firms	.302	.333	.977	−.031	.186	526
Value firms	.556	.697	.917	−.142	.149	257

NOTE.—BHARs are calculated as the difference between the equal- and value-weight average 3-year return for the event firms and the benchmark portfolios. The 3-year returns begin the month following completion of the event. The benchmark portfolios are 25 value-weight nonrebalanced portfolios formed on size and book-to-market equity (BE/ME), based on New York Stock Exchange (NYSE) breakpoints. Statistical inference is based on an empirical distribution created by simulating 1,000 pseudosamples with characteristics similar to those of the event-sample firms and then calculating the mean BHAR for each pseudosample. The *p*-value is the fraction of the mean BHARs from the pseudosamples larger in magnitudes (but with the same sign) than the original mean BHAR. The wealth relative is the average 3-year gross return of the sample firms divided by the average 3-year gross return of the benchmark firms. Growth firms are identified as firms with BE/ME ratios in the lowest quintile of all NYSE firms. Value firms are identified as firms with BE/ME ratios in the highest quintile of all NYSE firms.

TABLE 3   **Three-Year Mean Buy-and-Hold Abnormal Returns (BHARs) for Seasoned Equity Issuers (July 1961–December 1993)**

	Sample	Benchmark	Wealth Relative	BHAR	*p*-Value	*n*
Equal weight:						
Postevent	.348	.450	.930	−.102	.000	4,439
Preevent	1.519	.673	1.505	.845	.000	2,982
Excluding utilities	.343	.432	.938	−.089	.000	3,842
Growth firms	.259	.237	.017	.022	.714	1,410
Value firms	.427	.668	.856	−.240	.000	538
Value weight:						
Postevent	.411	.453	.971	−.042	.165	4,439
Preevent	.424	.445	.985	−.022	.516	2,982
Excluding utilities	.435	.441	.996	−.006	.450	3,842
Growth firms	.451	.324	1.096	.127	.052	1,410
Value firms	.448	.694	.855	−.246	.000	538

NOTE.—For definitions of the variables and a description of the statistical inference, see table 2.

TABLE 4   **Three-Year Mean Buy-and-Hold Abnormal Returns (BHARs) for Equity Repurchasers (July 1961–December 1993)**

	Sample	Benchmark	Wealth Relative	BHAR	*p*-Value	*n*
Equal weight:						
Postevent	.787	.642	1.088	.145	.000	2,292
Preevent	.558	.529	1.019	.029	.346	1,919
Open market	.776	.620	1.096	.156	.000	1,942
Tender offer	.854	.767	1.049	.087	.228	350
Growth firms	.590	.491	1.067	.099	.021	503
Value firms	1.096	.852	1.132	.244	.030	369
Value weight:						
Postevent	.541	.475	1.044	.066	.227	2,292
Preevent	.458	.428	1.021	.030	.459	1,919
Open market	.559	.480	1.053	.079	.112	1,942
Tender offer	.442	.449	.995	−.007	.482	350
Growth firms	.382	.353	1.021	.028	.460	503
Value firms	.824	.641	1.111	.183	.134	369

NOTE.—For definitions of the variables and a description of the statistical inference, see table 2.

3 years than a strategy that invested in similar size-BE/ME firms. The VW $\overline{\text{BHAR}}$ is −0.038 with a *p*-value of .027 and a wealth relative of 0.973.[4]

The closest comparison to our results is with Loughran and Vijh (1997), who study 947 acquisitions during 1970–89. They report an EW 5-year $\overline{\text{BHAR}}$ of −6.5%, with a *t*-statistic of −0.96. The most extreme abnormal returns in other papers are usually documented for special groupings of event firms, based on BE/ME rankings or form of payment. Popular groupings based on BE/ME are commonly denoted as ''growth'' (or glamour), which refers to firms with low BE/ME, and ''value,'' which refers to firms with high BE/ME. The growth and value groupings are important for several recent behavioral theories of stock market overreaction and underreaction following major corporate decisions. These theories have been interpreted as predicting negative long-term abnormal returns for growth firms and positive abnormal returns for value firms completing major corporate actions. For example, Rau and Vermaelen (1998) document large differentials in performance between glamour and value acquirers. Specifically, they report bias-adjusted 3-year cumulative abnormal returns (CARs) for glamour acquirers of −17.3% (*t*-statistic = −14.45) and value acquirers of 7.6% (*t*-statistic = 14.23). In direct contrast, when we analyze the stock price

4. The preevent buy-and-hold abnormal returns (BHARs) document superb performance before the event for small acquirers relative to the benchmark portfolios, with an equal-weight (EW) average 3-year raw return of 88%, an abnormal return of 28.5%, and a wealth relative of 1.18. However, this abnormal performance appears to be stronger for small acquirers as the variable weight (VW) average preevent BHAR is insignificant.

performance of growth and value acquirers, we find no evidence of reliable differential performance (1% on an EW basis, representing differential performance several times smaller than the amount previously documented).

Similar to previous research, we find that acquirers that use stock to finance the merger perform worse than those that abstain from equity financing. The EW average unadjusted 3-year return for acquirers that finance mergers with at least some stock is roughly two-thirds that of acquirers that abstain from stock financing (39% vs. 63%). The EW $\overline{\text{BHAR}}$ for stock acquirers is -8.4% (p-value $= .000$), while the $\overline{\text{BHAR}}$ for nonstock acquirers is 6.4% (p-value $= .047$). The VW results are similar, although the differences are much smaller in magnitude. Loughran and Vijh (1997) document a similar pattern in abnormal returns related to financing, albeit their differences are larger by a magnitude of three. They report a 5-year EW $\overline{\text{BHAR}}$ of -24.2% (t-statistic $= -2.92$) for stock mergers and 18.5% (t-statistic $= 1.27$) for cash mergers.

Seasoned equity offerings (SEOs). Table 3 reports the BHAR results for SEOs. The EW average 3-year return for the SEO sample is 35%, while the average size-BE/ME 3-year return is 45%, producing a $\overline{\text{BHAR}}$ of -10% with a p-value of .000. The EW $\overline{\text{BHAR}}$ of -10% is 2.5 times smaller than the minimum of the empirical distribution, which is only -4.1%. In other words, not a single one of the 1,000 pseudosample $\overline{\text{BHAR}}$s even comes close to the -10% EW $\overline{\text{BHAR}}$, which is typical of many of the $\overline{\text{BHAR}}$s reported. The VW $\overline{\text{BHAR}}$ is -4.2%, with a p-value of .165.[5] To facilitate comparison with other studies, we also report results after excluding utilities. These results are virtually identical to the full-sample results.

Overall, our results are similar to those obtained by other research. Spiess and Affleck-Graves (1995) and Brav, Geczy, and Gompers (in press) both report EW 3-year raw returns of 34% and 32%, respectively, whereas the average benchmark returns are on the order of 57% and 44%, respectively. Brav et al. also document that the abnormal performance is largely confined to small low-BE/ME firms and that it is substantially reduced with value-weighting.

Share repurchases. The EW BHAR for repurchases is 14.5%, with p-value $= .000$ (table 4). Again, not a single one of the 1,000 pseudosample $\overline{\text{BHAR}}$ even comes close to the 14.5% EW $\overline{\text{BHAR}}$. These results are virtually identical to those reported by Ikenberry et al. (1995) for repurchase programs announced during 1980–90. The abnormal

5. During the 3-year preevent period, small issuers experience enormous average returns. The EW average unadjusted preevent buy-and-hold return is 144%, corresponding to an abnormal 3-year return of 75%. These extremely large preevent returns appear strongest for small stocks as the VW BHAR is slightly negative and insignificant.

performance is considerably reduced with value weighting; VW $\overline{\text{BHAR}}$ equals 6.6% (*p*-value = .227).[6]

The EW $\overline{\text{BHAR}}$ is larger for firms repurchasing their shares on the open market (15.6%, *p*-value = .000) rather than through tender offers (8.7%, *p*-value = .228). This is interesting in light of a tender offer being a more dramatic event in the life of a firm. In addition, the abnormal returns are twice as large for value firms as for growth firms (24.4% vs. 9.9%), similar to Ikenberry et al. (1995). But the VW $\overline{\text{BHAR}}$s are not reliably different from zero for the subsamples.

IV. Assessing the Statistical Reliability of BHARs

As described in Section II, it is common to simulate an empirical distribution in order to perform statistical inference of the $\overline{\text{BHAR}}$s since the individual BHARs have poor statistical properties, producing biased test statistics in random samples (see Barber and Lyon 1997; Kothari and Warner 1997; Lyon et al. 1999). Most researchers view this bootstrapping procedure as a robust solution to the known statistical problems associated with the BHAR methodology. In this section, we question the robustness of the bootstrapping procedure with respect to statistical inference for event samples.

The empirical distribution is simulated under the null hypothesis assuming (1) the 25 size-BE/ME benchmark portfolios completely describe expected returns, and (2) the randomly selected firms used to construct the empirical distribution have the same covariance structure as the sample firms. Fama (1998, p. 291) details the problems associated with the first assumption as the ''bad model'' problem, arguing that ''all models for expected returns are incomplete descriptions of the systematic patterns in average returns during any sample period.'' In other words, if the model for expected returns does not fully explain stock returns, measured abnormal performance is likely to exist with respect to event samples exhibiting common characteristics. The best means of checking the robustness of our results with respect to this assumption is to repeat the analysis with a different model of expected returns and a different methodology.

We focus on the second assumption, which is crucial to the statistical reliability of BHARs and is unique to the manner in which the empirical distribution is constructed. The bootstrapping procedure makes two implicit assumptions: (1) the residual variances of sample firms are no different from randomly selected firms, and (2) the observations are independent. The first assumption may pose a problem if the sample firms' returns are more or less volatile than those of the firms that are used to create the pseudosamples. Although on average, the $\overline{\text{BHAR}}$

---

6. One misconception about firms that repurchase their shares is that they perform poorly before the event. The preevent abnormal performance is insignificant on both an EW and VW basis.

will be correct, the empirical distribution may be too "tight," leading to an overstatement of significance (see Brav, in press). The second assumption may be problematic if the events themselves are driven by some underlying factor not captured by size and BE/ME. Andrade and Stafford (1999) show that mergers (from the acquirer's perspective) tend to cluster in calendar time by industry. Similarly, Mitchell and Mulherin (1996) identify fundamental industry shocks that lead to increased takeover activity at the industry level, and Comment and Schwert (1995) do the same at the aggregate level. Ritter (1991) states that IPOs cluster by industry at given points in time. It is also likely that SEOs and share repurchases cluster by industry. If the event-clustering leads to positively correlated individual BHARs, statistical significance will be overstated by any methodology that assumes independence.

We have some reason to be concerned that the p-values reported in tables 2–4 are overstated because we find strong statistical significance for economically small estimates. For example, the VW $\overline{\text{BHAR}}$ for acquirers is -3.8%, with a p-value of .027 and a wealth relative of 0.973. In other words, the average 3-year investment in acquiring firms generated 2.7% less wealth than an otherwise similar investment in nonacquirers on a value-weight basis. In economic terms, this does not seem significant, but the test statistic suggests that this represents statistically significant long-term mispricing. In addition, we find $\overline{\text{BHARs}}$ 2.5 times larger (in magnitude) than the extreme of the empirical distribution. For example, the EW $\overline{\text{BHAR}}$ for SEOs is -10.2%, whereas the minimum of the empirical distribution is only -4.1%. Assuming normality of the empirical distribution of the mean BHAR, this corresponds to a p-value of less than .000000001.

A. *Properties of the Empirical Distribution of the Mean BHAR*

Figure 1 plots the simulated empirical null distribution of the EW $\overline{\text{BHAR}}$ for the SEO sample, as described above and in Section I. The plots reveal that the distributions are quite symmetric and reasonably well approximated by the normal density superimposed on the graphs. Because of the large sample size, the mean should be close to normal regardless of the underlying distribution of the individual firm BHARs.[7] However, the Jarque-Bera test statistic rejects normality of both the EW and VW empirical distributions. Although the empirical distributions are not normal, it is interesting to see how poor an assumption normality is. Table 5 reports various critical values based on the empirical distributions for the three event samples, and it compares them to the critical values assuming normality. The critical values assuming

---

7. The Central Limit Theorem guarantees that a standardized sum of random variables will converge to a normal (0, 1) random variable, even if the individual random variables are correlated (see Chung [1974] for a discussion).

TABLE 5   **Critical Values for Mean Buy-and-Hold Abnormal Returns Assuming Independence**

	Equal Weight				Value Weight			
	.5	2.5	97.5	99.5	.5	2.5	97.5	99.5
Mergers:								
Empirical distribution (%)	−4.3	−2.9	5.9	7.0	−6.1	−4.1	9.1	10.6
Assuming normality (%)	−5.6	−3.3	5.5	7.8	−7.3	−4.1	8.5	11.7
Seasoned equity offerings:								
Empirical distribution (%)	−3.6	−2.4	5.9	7.5	−10.5	−8.2	10.2	21.6
Assuming normality (%)	−4.8	−2.7	5.8	7.9	−14.6	−9.7	10.0	14.9
Share repurchases:								
Empirical distribution (%)	−4.5	−3.3	7.0	8.9	−6.5	−4.6	15.3	21.3
Assuming normality (%)	−6.2	−3.5	7.0	9.6	−11.4	−6.4	13.7	18.7

NOTE.—Critical values are from the empirical distributions of the equal- and value-weight mean buy-and-hold abnormal returns (BHARs) from the merger, seasoned equity offering, and share repurchase event samples. The empirical distribution is created by simulating 1,000 pseudosamples with size and book-to-market characteristics similar to those of the event sample firms and then calculating the mean BHAR for each pseudosample. The critical values assuming normality of the empirical distribution are calculated by adding (subtracting) 2 SDs to the mean for the 2.5th (97.5th) percentile and by adding (subtracting) 3 SDs to the mean for the 0.5th (99.5th) percentile, where the mean and standard deviation are calculated from the empirical distribution.

normality are calculated based on the mean and standard deviation of the empirical distribution. For the most part, assuming normality seems reasonable, and inferences would be unaffected at either the 1% or 5% levels for both EW and VW for all three samples.

We also create a bootstrapped distribution of the $\overline{\text{BHAR}}$ (not reported) for the three samples by resampling from the event-firm BHARs themselves (see Efron and Tibshirani 1993). This again assumes that the events are independent, but the bootstrapping makes no assumptions about the event-firm residual variances relative to randomly selected firms, as resampling is done using the original BHAR data. This allows us to isolate the effect of differential residual variance on inferences via the empirical distribution. Comparison of the empirical and the bootstrapped distributions reveals no noticeable difference in dispersion, which suggests that increased residual variance of event firms is not a serious problem with these three samples.[8]

B.   *Cross-Sectional Dependence of BHARs*

Our primary statistical concern is that major corporate actions are not random events and thus may not represent independent observations. The very nature of an event sample is that all of these firms have chosen to participate in an event, while other firms have chosen not to participate. As indicated above, major corporate events cluster through time by industry. This may lead to cross-correlation of abnormal returns,

8. In general, this may be a major problem. Brav (in press), e.g., documents that IPO firms have significantly larger residual variances than otherwise similar non-IPO firms.

which could flaw inferences from methodologies that assume indepen-
dence. There is an extensive accounting literature documenting cross-
sectional dependence of individual-firm residuals (see Collins and Dent
1984; Sefcik and Thompson 1986; Bernard 1987). These studies find
that contemporaneous market model residuals for individual firms are
significantly correlated, on the order of 18% within individual indus-
tries. Since major corporate events cluster in certain industries at any
given point in time, correlated residuals will pose a significant problem
for the BHAR methodology, which assumes independence of all obser-
vations, including those that are overlapping in calendar time.[9]

To gain perspective on the magnitude of this problem, we calculate
average pairwise correlations of monthly and annual BHARs for each
of our three event samples where there is perfect calendar-time overlap.
In other words, all possible unique correlations are calculated using 5
years of either monthly or annual abnormal return data for firms that
complete events in the same month. Below we show the grand average
of these average pairwise correlations of BHARs with complete calen-
dar-time overlap:

Frequency	Mergers	Seasoned Equity Offerings	Repurchases
Monthly	.0020	.0177	.0085
Annual	.0175	.0258	.0175

The average correlations increase substantially with the interval for all
three event samples, which is consistent with previous research. For
example, Bernard (1987) finds that average intraindustry correlations
in individual-firm market model residuals increase with holding period,
nearly doubling from 0.18 to 0.30 when the interval is increased from
monthly to annual.

Although the average correlations appear small, they can have a sig-
nificant impact on inferences with large samples. This can be seen by
inspecting the formulas for the sample standard deviation, equation (3),
and the ratio of the standard deviation assuming independence to the
standard deviation accounting for dependence, equation (4):[10]

$$\sigma_{\text{BHAR}} = \sqrt{\frac{1}{N} \cdot \overline{\sigma}_i^2 + \frac{(N-1)}{N} \cdot \overline{\rho}_{i,j} \overline{\sigma}_i \overline{\sigma}_j}; \qquad (3)$$

$$\frac{\sigma_{\text{BHAR}} \text{ (independence)}}{\sigma_{\text{BHAR}} \text{ (dependence)}} \approx \frac{1}{\sqrt{1 + (N-1)\overline{\rho}_{i,j}}}. \qquad (4)$$

9. Note that overlapping observations on the same firm are excluded from the samples.
The overlapping observations that are important for this analysis are those of similar firms,
such as those in the same industry.

10. The approximation of the ratio of σ(independence) to σ(dependence) assumes equal
variances of the individual BHARs.

where, N = number of sample events, σ_i^2 = average variance of individual BHARs, $\rho_{i,j}\sigma_i\sigma_j$ = average covariance of individual BHARs, and $\rho_{i,j}$ = average correlation of individual BHARs. In large samples with positive cross-correlations, the covariance term comes to dominate the individual variances. As such, ignoring cross-correlations will lead to overstated test statistics.

To determine the severity of overstated test statistics for our event samples, we calculate ''corrected'' t-statistics that account for dependence in BHARs under various assumptions about the average correlation of 3-year BHARs and the covariance structure. We report results only for the EW test statistics because EW results are the largest and tend to be the focus of most previous research. We assume that the average correlation for overlapping observations is linear in the number of months of calendar-time overlap, ranging from 0.0 for nonoverlapping observations to the estimated average correlation of 3-year BHARs of firms with complete overlap. Table A1 in the appendix displays the assumed covariance structure for the SEO sample. It is difficult to estimate directly average correlations of 3-year BHARs because of limited data. Therefore, we report a range of estimates and show the impact on t-statistics over this range. Table 6 displays the results.

First, we should point out that we are assuming that the empirical distribution is normal, which, although not technically true, appears to be a reasonable approximation. This assumption allows us to calculate t-statistics for the $\overline{\text{BHAR}}$s using the mean and standard deviation from the empirical distribution. We are also able to calculate t-statistics using standard deviations that account for cross-correlations. Second, because the average correlation appears to be increasing in holding period, it is unlikely that the average correlation of 3-year BHARs is less than the average correlation from the annual BHARs. Therefore, the t-statistics that assume that the average correlation of 3-year BHARs is equal to our estimate of average correlations from annual BHARs are still likely to be overstated.

The corrected t-statistics reveal that there is no statistical evidence of abnormal returns for any of the three event samples. The massive t-statistics of -6.05 (SEOs) and 4.86 (repurchases) that assume independence fall to -1.49 and 1.91, respectively, after accounting for the positive cross-correlations of individual BHARs. Since the average correlation of 3-year BHARs is almost surely larger than that of annual BHARs, the t-statistics that assume average correlations of 0.02 for mergers and repurchases and 0.03 for SEOs are probably more reflective of the true level of significance.

C. The Bottom Line on BHARs

The literature on long-term stock price performance heavily emphasizes results from the BHAR methodology despite well-known poten-

TABLE 6   **Corrected *t*-Statistics for Average 3-Year Buy-and-Hold Abnormal Returns (BHARs)**

	Mergers	Seasoned Equity Offerings	Repurchases
Descriptive statistics:			
Average BHAR	−.010	−.102	.145
Mean of empirical distribution	.0111	.0111	.0172
Standard deviation of empirical distribution	.0222	.0187	.0263
Average correlation of annual BHARs with complete overlap	.0175	.0258	.0175
t-statistics:			
Assuming independence	−.95	−6.05	4.86
Using average correlation of annual BHARs	−.42	−1.49	1.91
Using average correlation of annual BHARs + 25%	−.38	−1.34	1.73
Assuming average correlation = .01	−.52	−2.28	2.39
Assuming average correlation = .02	−.40	−1.67	1.80
Assuming average correlation = .03	−.34	−1.38	1.51

NOTE.—The corrected *t*-statistics are adjusted to account for cross-correlation of individual BHARs using the correlation structure described in the appendix. Average BHARs are calculated as the difference between the equal-weight average 3-year return for the sample of event firms and the benchmark portfolios. The 3-year returns begin the month following completion of the event. The benchmark portfolios are 25 value-weight nonrebalanced portfolios formed on size and book-to-market equity based on New York Stock Exchange breakpoints. The empirical distribution is created by simulating 1,000 pseudosamples with characteristics similar to those of the event-sample firms and then calculating the mean BHAR for each pseudosample. The *t*-statistic that assumes independence is calculated as the average BHAR minus the mean of the empirical distribution, all divided by the standard deviation of the empirical distribution. The corrected *t*-statistics are adjusted using the following approximation:

$$\frac{\sigma_{\text{BHAR}}(\text{independence})}{\sigma_{\text{BHAR}}(\text{dependence})} \approx \frac{1}{\sqrt{1 + (N-1)\bar{\rho}_{i,j}}}.$$

tial problems. Barber and Lyon (1997), Kothari and Warner (1997), and Lyon et al. (1999) provide simulation evidence showing that $\overline{\text{BHAR}}$ estimates can be biased because of poor statistical properties of individual-firm BHARs. Many of these biases are mitigated with large sample sizes and careful construction of benchmark portfolios. However, the problems associated with standard error estimates for BHARs on nonrandom samples cannot easily be corrected, and they are generally increasing in sample size. This point is often missed in methodology papers and dismissed in long-term event studies, which frequently claim that bootstrapping solves all dependence problems. However, that claim is not valid. Event samples are clearly different from random samples. Event firms have chosen to participate in a major corporate action, while nonevent firms have chosen to abstain from the action. An empirical distribution created by randomly selecting firms with similar size-BE/ME characteristics does not replicate the covariance structure underlying the original event sample. In fact, the typical bootstrapping approach does not even capture the cross-sectional correlation

structure related to industry effects that has been documented by Bernard (1987), Brav (in press), and others. Moreover, Bernard shows that the average interindustry cross-sectional correlation of individual abnormal returns is also positive, which suggests that dependence corrections concentrating only on industry effects will not account for all cross-correlations.

Our results suggest that the popular BHAR methodology, in its traditional form, should not be used for statistical inference. Some type of correction for positive cross-correlations of individual event firm BHARs should be made. Finally, it is worth noting that, for our three major events, there is no statistical evidence of long-term abnormal returns after accounting for positive cross-correlations of individual event firm BHARs.

V. Calendar-Time Portfolio Approach

An alternative approach to measuring long-term stock price performance is to track the performance of an event portfolio in calendar time relative to either an explicit asset-pricing model or some other benchmark. The calendar-time portfolio approach was first used by Jaffe (1974) and Mandelker (1974) and is strongly advocated by Fama (1998). The event portfolio is formed each period to include all companies that have completed the event within the prior n periods. By forming event portfolios, the cross-sectional correlations of the individual event firm returns are automatically accounted for in the portfolio variance at each point in calendar time. In light of our strong evidence that the individual event firm abnormal returns are cross-sectionally correlated, calendar-time portfolios represent an important improvement over the traditional BHAR methodology, which assumes independence of individual-firm abnormal returns.

A. Calculating Calendar-Time Abnormal Returns (CTARs)

For each month from July 1961 to December 1993, we form EW and VW portfolios of all sample firms that participated in the event within the previous 3 years.[11] Portfolios are rebalanced monthly to drop all companies that reach the end of their 3-year period and add all companies that have just executed a transaction. The portfolio excess returns are regressed on the three Fama and French (1993) factors, as in equation (5):

$$R_{p,t} - R_{f,t} = a_p + b_p(R_{m,t} - R_{f,t}) + s_p \text{SMB}_t + h_p \text{HML}_t + e_{p,t}. \quad (5)$$

The three factors are zero-investment portfolios representing the ex-

---

11. We exclude multiple observations on the same firm that occur within 3 years of the initial observation.

cess return of the market; the difference between a portfolio of ''small'' stocks and ''big'' stocks, SMB; and the difference between a portfolio of ''high'' BE/ME stocks and ''low'' BE/ME stocks, HML. Within this framework, the intercept, a_p, measures the average monthly abnormal return on the portfolio of event firms, which is zero under the null of no abnormal performance, given the model. If the Fama and French model provides a complete description of expected returns, then the intercept measures mispricing. However, if the model provides only an imperfect description of expected returns, then the intercept represents the combined effects of mispricing and model misspecification. This is what Fama (1970) refers to as the ''joint-test problem''—tests of market efficiency are necessarily joint tests of market efficiency and the assumed model of expected returns.

Table 7 reports the intercepts from regressions of the 25 EW and VW size-BE/ME portfolios on the Fama and French 3-factor model. These are the original assets used in Fama and French (1993) to test the model. As pointed out by Fama and French, the three-factor model is unable to completely describe the cross-section of expected returns even on the dimensions on which it is based. This is illustrated by the several significant intercepts in table 7. This suggests that the null hypothesis—intercept equals zero—may be problematic for samples tilted toward characteristics that the model cannot price in the first place. This can be seen most easily with IPOs. The IPO firms are overwhelmingly small, low-BE/ME firms. When the abnormal returns of IPO firms are estimated with the Fama and French three-factor model, the estimates are on the order of −12% to −15% for an EW portfolio over a 3-year period, or about −0.35% to −0.42% per month. However, Brav and Gompers (1997) argue that the underperformance of IPOs is not an IPO effect per se. They find that similar size and book-to-market firms that have not issued equity perform as poorly as IPOs. Note that the intercept for all small, low-BE/ME firms reported in table 7 is −0.37, which is essentially identical to that found for IPOs.

In order to gain perspective on whether the known pricing deficiencies of the Fama and French three-factor model affect the three samples studied in this article, we decompose the intercepts into two components: (1) the expected abnormal performance, given the sample composition (based on size-BE/ME portfolio assignment and calendar-time frequency); and (2) the amount of abnormal performance attributable to other sources, including the event. In particular, we estimate the expected intercept, conditional on the sample composition, as the mean intercept from 1,000 calendar-time portfolio regressions of random samples of otherwise similar nonevent firms. This is directly comparable to the empirical distribution used in the BHAR analysis. However, we are using this methodology to determine the mean of the null distribution, not to measure the dispersion. Each of the 1,000 random sam-

TABLE 7   **Intercepts from Excess Stock Return Regressions of 25 Size and Book-to-Market Equity Portfolios on the Fama and French Three-Factor Model (July 1963–December 1993)**

	Low	2	3	4	High
Intercepts:					
Equal-weight portfolios:					
Small	−.37	.02	.06	.23	.26
2	−.21	−.01	.11	.11	−.04
3	−.14	.07	−.01	.14	−.03
4	.09	−.14	.02	.01	.03
Large	.10	.00	−.02	−.07	−.09
Value-weight portfolios:					
Small	−.49	−.09	−.05	.06	.01
2	−.09	.03	.13	.14	.03
3	−.06	.10	.02	.12	−.04
4	.14	−.15	.02	.01	.01
Large	.15	−.01	−.02	−.04	−.24
t-statistics:					
Equal-weight portfolios:					
Small	−2.65	.21	.67	2.75	2.52
2	−2.54	−.09	1.61	1.79	−.60
3	−1.85	1.00	−.13	2.18	−.41
4	1.15	−1.69	.26	.19	.33
Large	1.67	−.02	−.29	−1.05	−.87
Value-weight portfolios:					
Small	−4.80	−1.22	−.76	1.09	.18
2	−1.02	.43	1.94	2.29	.46
3	−.88	1.34	.33	1.82	−.45
4	1.94	−1.93	.29	.16	.12
Large	2.37	−.08	−.27	−.61	−2.13

NOTE.—Dependent variables are 25 size and book-to-market equity portfolio returns, *Rp,* in excess of the 1-month Treasury-bill rate, *Rf,* observed at the beginning of the month. The 25 size and book-to-market equity portfolios are formed on New York Stock Exchange size and book-to-market equity quintiles. The three factors in the Fama and French model are zero-investment portfolios representing the excess return of the market, $Rm - Rf$; the difference between a portfolio of small stocks and big stocks, SMB; and the difference between a portfolio of high book-to-market stocks and low book-to-market stocks, HML. See Fama and French (1993) for details on the construction of the factors. Their empirical model is $Rp_t - Rf_t = a + b(Rm_t - Rf_t) + sSMB_t + hHML_t + e_t$.

ples has the same calendar-time frequency, and at each point in time, the portfolio of randomly selected firms has the same size-BE/ME composition as the corresponding event portfolio. A new *t*-statistic is calculated using the expected intercept $\hat{a}_0$ as the null and the original intercept and standard error estimates. We refer to the difference between the estimated intercept and the expected intercept as the ''adjusted intercept'':

$$t = \frac{\hat{a} - \hat{a}_0}{\hat{s}}. \tag{6}$$

B. Calendar-Time Portfolio Regression Results

Tables 8–10 display the EW and VW calendar-time portfolio regression results for the three samples over the period July 1961 through

TABLE 8 — Calendar-Time Fama and French Three-Factor Model Portfolio Regressions of Acquirers (July 1961–December 1993)

	Equal Weight			Value Weight		
	a (t-statistic)	Adjusted a (t-statistic)	Adjusted R^2 [N]	a (t-statistic)	Adjusted a (t-statistic)	Adjusted R^2 [N]
Full sample (postevent)	−.20 (−3.70)	−.14 (−2.61)	.97 [390]	−.03 (−.48)	−.04 (−.68)	.95 [390]
Full sample (preevent)	.49 (9.49)	.52 (10.09)	.97 [378]	.19 (3.05)	.15 (2.48)	.94 [378]
Financed with stock	−.33 (−4.64)	−.25 (−3.59)	.95 [390]	−.14 (−2.00)	−.12 (−1.81)	.93 [390]
Financed without stock	−.09 (−1.14)	−.04 (−.54)	.94 [388]	.14 (1.68)	.10 (1.19)	.89 [388]
Growth firms	−.37 (−3.64)	−.18 (−1.76)	.90 [373]	−.13 (−1.23)	−.20 (−1.95)	.87 [373]
Value firms	.00 (.01)	−.08 (−.51)	.85 [355]	.06 (.32)	.03 (.14)	.73 [355]
Hot market	.00 (.00)	.04 (.08)	.97 [390]	−.12 (−.89)	−.13 (−.93)	.95 [390]
Cold market	.11 (.85)	.14 (1.12)	.97 [390]	.19 (1.42)	.19 (1.40)	.95 [390]

NOTE.—Dependent variables are event portfolio returns, Rp, in excess of the 1-month Treasury-bill rate, Rf, observed at the beginning of the month. Each month, we form equal- and value-weight portfolios of all sample firms that have completed the event within the previous 3 years. The event portfolio is rebalanced monthly to drop all companies that reach the end of their 3-year period and add all companies that have just executed a transaction. The three factors are zero-investment portfolios representing the excess return of the market, $Rm − Rf$; the difference between a portfolio of small stocks and big stocks, SMB; and the difference between a portfolio of high book-to-market stocks and low book-to-market stocks, HML. See Fama and French (1993) for details on the construction of the factors. The intercept, a, measures the average monthly abnormal return, given the model. The adjusted intercept measures the difference between the intercept estimated using the event portfolio and the average intercept estimated from 1,000 random samples of otherwise similar (based on size and book-to-market) nonevent firms. A minimum of 10 firms in the event portfolio is required. The t-statistic is in parentheses, and N, the number of monthly observations, is in square brackets. The equation is $Rp_t − Rf_t = a + b(Rm_t − Rf_t) + sSMB_t + hHML_t + e_t$.

TABLE 9   Calendar-Time Fama and French Three-Factor Model Portfolio Regressions of Seasoned Equity Issuers (July 1961–December 1993)

	Equal Weight			Value Weight		
	a (t-statistic)	Adjusted a (t-statistic)	Adjusted R^2 [N]	a (t-statistic)	Adjusted a (t-statistic)	Adjusted R^2 [N]
Full sample (postevent)	−.33 (−5.19)	−.22 (−3.51)	.96 [390]	−.03 (−.44)	.00 (−.02)	.90 [390]
Full sample (preevent)	1.28 (17.39)	1.30 (17.62)	.95 [378]	.19 (2.55)	.17 (2.30)	.90 [378]
Excluding utilities	−.37 (−5.58)	−.26 (−3.86)	.97 [390]	.06 (.77)	.09 (1.08)	.91 [390]
Growth firms	−.32 (−3.45)	−.07 (−.76)	.94 [390]	.19 (1.62)	.16 (1.39)	.88 [390]
Value firms	−.31 (−2.24)	−.34 (−2.52)	.86 [314]	−.24 (−1.36)	−.15 (−.87)	.66 [314]
Hot market	−.12 (−.81)	.02 (.15)	.96 [390]	.07 (.40)	.07 (.41)	.90 [390]
Cold market	.07 (.45)	−.05 (−.31)	.96 [390]	.32 (1.79)	.42 (2.34)	.90 [390]

NOTE.—For definitions of variables and descriptions of the estimation, see table 8. The t-statistic is in parentheses, and N, the number of monthly observations, is in square brackets.

TABLE 10 Calendar-Time Fama and French Three-Factor Model Portfolio Regressions of Equity Repurchasers (July 1961–December 1993)

	Equal Weight			Value Weight		
	a (t-statistic)	Adjusted a (t-statistic)	Adjusted R^2 [N]	a (t-statistic)	Adjusted a (t-statistic)	Adjusted R^2 [N]
Full sample (postevent)	.08 (1.24)	.08 (1.17)	.96 [374]	−.01 (−.11)	−.04 (−.43)	.90 [374]
Full sample (preevent)	−.07 (−.91)	−.05 (−.72)	.95 [378]	−.05 (−.57)	−.08 (−.94)	.90 [378]
Open market	.14 (1.82)	.18 (2.39)	.95 [334]	.04 (.42)	.01 (.06)	.90 [334]
Tender offer	.14 (1.14)	.01 (.07)	.87 [323]	−.02 (−.12)	−.09 (−.50)	.73 [323]
Growth firms	−.20 (−1.35)	.03 (.20)	.87 [286]	−.08 (−.50)	−.13 (−.80)	.79 [286]
Value firms	.65 (4.99)	.48 (3.72)	.86 [306]	.29 (1.65)	.17 (.96)	.75 [306]
Hot market	.14 (.92)	.04 (.26)	.96 [374]	.22 (1.16)	.17 (.88)	.90 [374]
Cold market	−.13 (−.86)	−.20 (−1.31)	.96 [374]	.27 (1.37)	.06 (.31)	.90 [374]

NOTE.—For definitions of variables and descriptions of the estimation, see table 8. The *t*-statistic is in parentheses, and *N*, the number of monthly observations, is in square brackets.

December 1993. The number of monthly observations varies slightly
for different samples and subsamples as we require a minimum of 10
firms in each monthly event portfolio.

The EW 3-year acquirer portfolio exhibits statistically significant av-
erage abnormal returns: −0.20% per month or −7.2% after 3 years
(−0.20% × 36 months), with a t-statistic of −3.70 (table 8). When
the intercept is adjusted to control for the size and BE/ME characteris-
tics of the sample, the abnormal performance is lower. The adjusted
intercept is −0.14%, which corresponds to a 3-year abnormal return
of −5% (t-statistic = −2.61). The intercept from the VW regression
is not significant at −0.03 (t-statistic = −0.48), translating into a 3-
year average abnormal return of only −1.1%. The adjusted intercept
for the VW acquirer portfolio is virtually identical. Since the abnormal
returns are only significant when the event firms receive equal weight
in the portfolio, it appears that small acquirers are more prone to under-
performance in the postevent period. This finding is similar to that pre-
viously documented by Brav and Gompers (1997) and Brav et al. (in
press) with equity issuers.

With respect to portfolios composed of firms in the lowest BE/ME
quintile (growth or glamour firms), EW acquirer portfolios have sig-
nificantly negative intercepts of −0.37 (t-statistic = −3.64), corre-
sponding to 3-year average abnormal returns of −13.3%. However,
this does not appear to be entirely a ''merger effect'' as the adjusted
intercept is only −0.18% (−6.5% over 3 years), with a t-statistic of
−1.76. Acquirers with high-BE/ME (value firms) have an EW portfo-
lio intercept of 0.00 and an adjusted intercept of −0.08, suggesting that
these firms are fairly priced during the postevent period. Again, when
the firms are value weighted, the intercepts are statistically insignificant
for both the growth and value portfolios.[12]

Like other researchers, we find that acquirer underperformance is
limited to those firms that use at least some stock to finance the acquisi-
tion. The EW-stock-financed acquirer portfolio intercept is −0.33 (t-
statistic = −4.64) as compared with −0.09 (t-statistic = −1.14) for
the EW-no-stock portfolio. The differential performance based on the
type of financing survives value weighting. The t-statistic for the differ-
ence between the stock-financed and the non-stock-financed adjusted
intercepts is 2.06, although neither adjusted intercept is significant on
its own.

The conclusion from the acquirer regressions is that, on an EW basis,
acquirers tend to significantly underperform in the 3 years following
the acquisition, but this appears to be limited to those acquirers using
stock financing. On a VW basis, there is virtually no evidence of ac-

12. Acquiring firms have significantly positive average monthly abnormal returns in the
3-year period prior to the merger, on both an equal- and value-weight basis (17.6% and
6.8% over the 3-year period, respectively).

quirer stock price underperformance. This could be the result of either the VW regressions having low power to detect abnormal performance or the larger firms actually not underperforming. We try to distinguish between these two scenarios in the subsequent section, but the high R^2's (generally over 0.90) hint that these regressions indeed have considerable power. Moreover, the VW adjusted-intercept estimates are small in economic terms, ranging from −0.20 to 0.15, which is consistent with these firms being fairly priced on average.

Table 9 displays the calendar-time SEO portfolio regression results. The EW-issuer portfolio has significantly negative abnormal returns in the 3 years following the equity issue, averaging −0.33% per month or −12% over 3 years (*t*-statistic = −5.19). Again, not all of the measured abnormal performance is attributable to the equity-issue event as the adjusted intercept is −0.22, or −7.9% over 3 years (*t*-statistic = −3.51).[13] Although utilities account for a sizable fraction of the equity-issue sample, the measured underperformance is essentially unchanged when we exclude utilities. It appears that the underperformance of the seasoned equity issuers is confined to the EW value portfolio. When the equity-issuer event portfolio is value-weighted, there is virtually no evidence of underperformance either for the full sample or for the various subsamples.

The full-sample repurchase portfolios show no signs of abnormal performance on either an EW or VW basis.[14] The EW postrepurchase average abnormal return is 0.08% per month—less than 3% after 3 years—and the VW portfolio return is less than half this large. On an EW basis, there is strong support for the notion that value repurchasers outperform their expected return benchmark. The EW-portfolio adjusted intercept is 0.48% per month, or 17.3% after 3 years, which is largely consistent with Ikenberry et al. (1995). However, this relation disappears when the repurchase firms are value weighted.

C.   *Robustness of the Calendar-Time Portfolio Regressions*

While the calendar-time portfolio approach solves the dependence problem associated with event-time abnormal performance measures, it has several potential problems that should be addressed. First, the regressions assume that the factor loadings are constant through time, up to 390 months, which is unlikely since the composition of the event portfolio changes each month. These events tend to cluster through time by industry, and different industries have different factor loadings (Fama and French 1997). The portfolio composition is likely heavily

---

13. On an EW basis, issuing firms have significantly positive abnormal returns in the 3 years prior to issuing equity, averaging 1.19% per month or 42.8% over 3 years (*t*-statistic = 14.05), while the VW issue portfolios show no evidence of significant abnormal returns in the preissue period.

14. There is no evidence of abnormal performance prior to the stock repurchase.

weighted in a few industries at each point in time but in different industries at longer intervals. This may lead to biased estimates. Second, the changing portfolio composition may introduce heteroskedasticity as the variance is related to the number of firms in the portfolio. This may cause the ordinary least squares estimator to be inefficient, but it will not lead to biased estimates. A third concern of this procedure is 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 (Loughran and Ritter, in press). If there is differential abnormal performance in periods of high activity versus periods of low activity, the regression approach will average over these, and it may be less likely to uncover abnormal performance. In other words, the full-sample-period regression, which tests for average monthly abnormal returns (given the model), will have low power against the alternative of abnormal performance in ''hot'' markets and no abnormal performance otherwise. A final concern is that the calendar-time portfolio regressions have low power to detect abnormal performance, as argued by Loughran and Ritter (in press).

Our first approach is to address directly the specific concerns of Loughran and Ritter (in press). In addition, we address these concerns simultaneously, by repeating the calendar-time portfolio analysis using the calendar-time abnormal return (CTAR) methodology that was first used by Jaffe (1974) and Mandelker (1974) and that is strongly advocated by Fama (1998).

Heteroskedasticity. Two important statistical issues are whether and how to control for heteroskedasticity. Since the number of firms in the event portfolio changes through time, the portfolio residual variance may also be changing through time. We mitigate the heteroskedasticity problem substantially by requiring at least 10 firms in the event portfolio at each point in time, which accounts for the majority of the diversification effect of the portfolio residual variance. The question is whether more should be done. One common ''correction'' for EW portfolios is weighted least squares, where the weights are proportional to $\sqrt{n}$ (e.g., see Franks, Harris, and Titman 1991). Here, the effects on the residual variance of the number of firms in the event portfolio changing through time can be neutralized by transforming the regression as in equation (7):

$$\sqrt{n} \cdot R_{p,t} = \sqrt{n} \cdot (X_t \cdot \beta_p + \varepsilon_{p,t});$$

$$n \cdot \text{Var}(R_{p,t}) = n \cdot \text{Var}(X_t \cdot \beta_p) + n \cdot \text{Var}(\varepsilon_{p,t}); \qquad (7)$$

$$n \cdot \text{Var}(\varepsilon_{p,t}) = n \cdot \text{Var}\left(\frac{1}{n} \sum_{i=1}^{n} e_{i,t}\right) = n \cdot \frac{n \cdot \sigma_t^2}{n^2} = \sigma_t^2.$$

This transformation assumes that the individual-firm residuals are independent, and it effectively gives equal weight to all observations.

However, this completely defeats the purpose of forming calendar-time portfolios, which is to account for the fact that individual-firm residuals are cross-sectionally correlated.

Our approach to deal with the potential problems of heteroskedasticity is to calculate the finite-sample critical values using a general nonparametric bootstrap procedure detailed in Horowitz (1996). This procedure is quite general and can be used for unknown forms of heteroskedasticity. Specifically, the bootstrapping procedure amounts to sampling 1,000 (y,X) pairs from the original data with replacement. We estimate b^*, s^*, and f^* for each bootstrap sample, where b^* and s^* are the ordinary least squares coefficient and standard error estimates, and t^* or equal, $(b^* - \beta)/s^*$, where β is the original ordinary least squares estimate. The empirical distribution of the t-statistics is used to determine the finite-sample critical values. We reject the null if $|t| > z^*$, where t is the original t-statistic and z^* is the critical value from the empirical distribution of t-statistics.

We find that inferences are unaffected using the bootstrapped critical values rather than the traditional 5%-level critical value of 1.96 for all of the intercepts. As can be seen below, the full-sample bootstrapped critical values are tightly scattered around the theoretical 5% critical value of 1.96. Here we show the bootstrapped critical values for the full-sample calendar-time portfolio regressions:

	Mergers	Seasoned Equity Offerings	Repurchases
Equal weight	2.09	2.13	1.92
Value weight	2.08	1.91	1.92

Hot versus cold markets. We test for differential performance in heavy- and low-event activity periods. We rerun the full-sample portfolio regressions after including two dummy variables, formed on whether there are an unusually large or small number of firms in the event portfolio during that calendar month (see tables 8–10). We define monthly activity as the number of firms in the event portfolio divided by the number of firms listed in the CRSP monthly-return files that month, accounting for the changing composition of the CRSP population through time. Because NASDAQ firms are added to CRSP in the middle of our sample period, we first define event activity for each exchange and then create a CRSP-level activity index as the weighted average of the exchange-level activity indices. The HOT variable is equal to one if event activity lies above the seventieth percentile of all monthly activities and zero otherwise; while the COLD variable is equal to one if event activity lies below the thirtieth percentile of all monthly activities, and zero otherwise. The coefficient on the HOT

dummy variable is insignificant for all of the samples, regardless of whether EW or VW portfolios are analyzed and regardless of whether the coefficients are adjusted or not. The coefficient on the COLD dummy variable is only significant for the VW SEO portfolio. Specifically, the average monthly abnormal returns (adjusted) for the VW SEO portfolio are -0.15% (t-statistic $= -1.26$) in normal-activity periods, 0.07% (t-statistic $= 0.41$) in heavy-activity periods, and 0.42% (t-statistic $= 2.34$) in low-activity periods. These results suggest that the measured abnormal performance for our event samples is not systematically related to the intensity of the event activity. Our evidence is inconsistent with the hypothesis advanced by Loughran and Ritter (in press) that abnormal performance is concentrated in periods when there are a relatively large number of events.

Calendar-time abnormal returns (CTARs). The CTAR is the average abnormal return calculated each calendar month for all sample firms that have completed the event within the prior 3 years:

$$CTAR_t = R_{p,t} - E(R_{p,t}), \qquad (8)$$

where $R_{p,t}$ is the monthly return on the portfolio of event firms, and $E(R_{p,t})$ is the expected return on the event portfolio. The expected return on the event portfolio is proxied by both the Fama and French three-factor model and with the 25 size-BE/ME portfolios. The CTAR approach is similar in spirit to the portfolio regression method in that event-portfolio abnormal returns are calculated in calendar time, such that the portfolio variance accounts for the cross-sectional correlation in the individual-event-firm abnormal returns.

When the 25 size-BE/ME portfolios serve as the expected return proxy, the benchmark can change through time to reflect changes in the firm's characteristics. Moreover, even when expected returns are proxied by the Fama and French three-factor model, the changing parameter problem is mitigated. Proxying expected returns with the Fama and French three-factor model amounts to estimating individual-firm factor loadings over a 5-year postevent estimation period (requiring at least 36 months of valid returns), and then averaging these to form the monthly portfolio factor loadings. Although parameter estimates are assumed constant for each sample firm over the 5-year estimation period, the event portfolio parameters are allowed to change each month as new firms are added and seasoned firms are dropped.

The monthly CTARs are standardized by estimates of the portfolio standard deviation, which serves two purposes (see tables 11–13 for details). First, by standardizing the monthly CTARs, we control for heteroskedasticity. Second, standardizing effectively gives more weight to periods of heavy event activity than periods of low event activity because the portfolio residual variance is decreasing in portfolio size, all else equal. Each standardized monthly CTAR should have

TABLE 11   Mean Calendar-Time Portfolio Abnormal Returns (CTARs) for Acquirers (July 1961–December 1993)

	Equal Weight		Value Weight	
	Fama and French Three-Factor	25 Size-BE/ME	Fama and French Three-Factor	25 Size-BE/ME
Full sample (postevent)	−.14	−.04	−.07	−.03
	(−2.70)	(.78)	(−1.75)	(−.58)
	[366]	[389]	[366]	[389]
Full sample (preevent)	.43	.50	.17	.14
	(8.77)	(9.59)	(3.33)	(2.95)
	[364]	[389]	[364]	[389]
Financed with stock	−.23	−.16	−.13	−.15
	(−3.59)	(−1.22)	(−2.51)	(−2.27)
	[366]	[388]	[366]	[388]
Financed without stock	−.07	.07	.04	.09
	(−.80)	(1.86)	(.34)	(.98)
	[366]	[371]	[366]	[371]
Growth firms	−.19	−.16	−.05	−.14
	(−2.05)	(−1.06)	(−.57)	(−1.23)
	[338]	[338]	[338]	[338]
Value firms	.01	.01	−.08	−.17
	(−.24)	(.48)	(−.76)	(−1.40)
	[346]	[346]	[346]	[346]

Note.—The CTARs are calculated each month as the difference between the event-portfolio return and the expected return on the portfolio, standardized by the portfolio residual standard deviation. Each month, we form equal- and value-weight event portfolios containing all sample firms that have completed the event within the previous 3 years. The event portfolio is rebalanced monthly to drop all companies that reach the end of their 3-year period and add all companies that have just executed a transaction. The portfolio expected returns are proxied by both 25 value-weight portfolios formed on size and book-to-market equity based on NYSE breakpoints (25 size-BE/ME), and the Fama and French three-factor model, which amounts to estimating individual-firm factor loadings over a 5-year postevent estimation period (requiring at least 36 months of valid returns), and then averaging these to form the monthly portfolio factor loadings. We calculate event-portfolio residual variances using 60 months of residuals. Residuals are calculated from portfolio regressions on the Fama and French three-factor model and as monthly differences of event portfolio returns and size-BE/ME portfolio returns. Mean CTARs and standard errors are calculated from the time-series of monthly CTARs. The *t*-statistic is in parentheses and the number of monthly observations is in square brackets.

a mean of zero and be independent. Therefore, statistical inference is based on the time-series mean of the monthly standardized CTARs and standard error of the mean.[15]

The results from the CTAR analysis are presented in tables 11–13. For the most part, the CTAR results are similar to the portfolio regression results for all three of the event samples, indicating that the regression results are quite robust. The primary difference is that virtually all of the CTARs are smaller in magnitude than the corresponding re-

15. The time series mean monthly CTARs is the measure of abnormal performance, while *t*-statistics are calculated from the time-series of monthly standardized CTARs. In some cases, it is possible for the mean CTAR to be associated with a *t*-statistic of opposite sign, depending on the estimated variances.

TABLE 12   Mean Calendar-Time Portfolio Abnormal Returns (CTARs) for
            Seasoned Equity Issuers (July 1961–December 1993)

	Equal Weight		Value Weight	
	Fama and French Three-Factor	25 Size-BE/ME	Fama and French Three-Factor	25 Size-BE/ME
Full sample (postevent)	−.25	−.21	−.01	−.08
	(−4.62)	(−3.38)	(−1.01)	(−1.36)
	[366]	[390]	[366]	[390]
Full sample (preevent)	1.09	1.33	.19	.19
	(17.34)	(21.84)	(2.62)	(2.62)
	[365]	[389]	[365]	[389]
Excluding utilities	−.28	−.20	.03	−.02
	(−4.65)	(−2.68)	(−.52)	(−.62)
	[366]	[390]	[366]	[390]
Growth firms	−.14	−.02	.04	.00
	(−1.41)	(−.24)	(.07)	(−.13)
	[301]	[301]	[301]	[301]
Value firms	−.13	−.03	−.06	−.06
	(−1.22)	(−.07)	(−.75)	(−.57)
	[257]	[257]	[257]	[257]

NOTE.—For definitions of variables and description of the estimation, see table 11. The *t*-statistic is in parentheses and the number of monthly observations is in square brackets.

TABLE 13   Mean Calendar-Time Portfolio Abnormal Returns (CTARs) for
            Equity Repurchasers (July 1961–December 1993)

	Equal Weight		Value Weight	
	Fama and French Three-Factor	25 Size-BE/ME	Fama and French Three-Factor	25 Size-BE/ME
Full sample (postevent)	.09	.16	−.04	.02
	(2.06)	(3.65)	(−.14)	(.68)
	[366]	[373]	[366]	[373]
Full sample (preevent)	.08	.04	.03	−.13
	(1.94)	(2.21)	(.78)	(−1.35)
	[363]	[379]	[363]	[379]
Open market	.14	.25	.02	.05
	(2.20)	(4.26)	(.44)	(1.01)
	[332]	[334]	[332]	[334]
Tender offer	.03	.11	−.15	.09
	(.21)	(1.06)	(−1.15)	(.67)
	[315]	[320]	[315]	[320]
Growth firms	.13	.26	.01	.01
	(1.09)	(2.19)	(.15)	(.00)
	[249]	[250]	[249]	[250]
Value firms	.24	.12	−.02	−.04
	(1.78)	(1.20)	(.02)	(−.28)
	[244]	[244]	[244]	[244]

NOTE.—For definitions and description of the estimations, see table 11. The *t*-statistic is in parentheses and the number of monthly observations is in square brackets.

gression estimate, suggesting that the regression intercepts are not biased toward zero as some of the potential concerns predict.

The main result from the acquirer sample—significantly negative abnormal returns are limited to stock mergers—is robust to the CTAR methodology, with the exception of the EW portfolio relative to the 25 size-BE/ME portfolios. In fact, when the 25 size-BE/ME portfolios are used as the expected return benchmark, none of the EW postevent acquirer portfolios have significant abnormal returns—the 3-year abnormal return for the full sample is only -1.44% with a *t*-statistic of 0.78. Moreover, there is no significant difference between the value and growth acquirer portfolios for either of the benchmarks, regardless of whether the portfolios are EW or VW.

Again, the CTAR results for issuers largely confirm those from the calendar-time portfolio regressions. The EW portfolio experiences abnormally low returns following the equity issue, on the order of -9% over 3 years, with a *t*-statistic around -4.6 or -3.4, depending on which benchmark is used. There is no hint of differential abnormal performance between the growth and value portfolios, as the *t*-statistics for the difference range from -0.07 to 0.57. Value weighting eliminates reliable underperformance for the full sample.

The EW CTAR results for repurchasers have a greater tendency to be statistically significant than the calendar-time portfolio regressions, but they are of similar magnitude. The most notable difference is that the EW repurchase portfolio average abnormal returns are now significant. In particular, the 3-year EW abnormal returns are 3.2% (*t*-statistic $= 2.06$) and 5.8% (*t*-statistic $= 3.65$) for the Fama and French three-factor and 25 size-BE/ME adjusted CTARs, respectively. Again, there is no difference between the abnormal returns of the value and growth portfolios, and there is no evidence of abnormal performance when the repurchaser portfolios are value-weighted.

Overall, the CTAR results tend to confirm our inferences from the calendar-time portfolio regressions, but they indicate that the point estimates are slightly smaller. To the extent that the CTAR methodology is plagued by fewer statistical flaws, more faith should be placed in these results.

Power of the Calendar-Time Portfolio Approach. In order to directly address the concerns raised in Loughran and Ritter (in press), we assess the specification and power of the calendar-time portfolio regressions and the Fama and French CTARs. Specifically, we calculate abnormal performance for random samples of size and time period similar to those analyzed in this study. We draw 1,000 random samples of 2,000 firms over the period July 1963–December 1993 from the population of CRSP firms with at least one valid return. For each of the 1,000 random samples, we induce 3-year abnormal returns ranging from -20% to 20%. Table 14 and figure 2 and table 15 and figure 3

TABLE 14   **Percentage of Samples Rejecting the Null Based on Calendar-Time Portfolio Regressions**

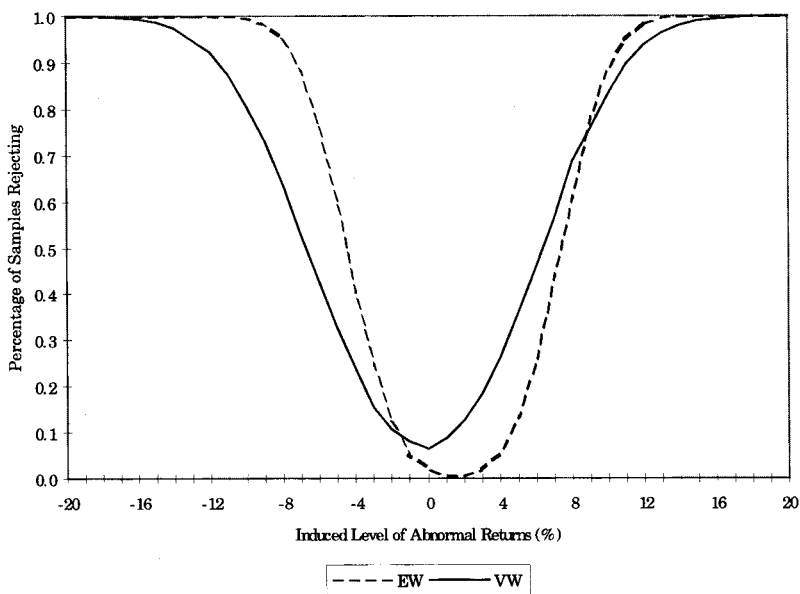| | Induced Level of Abnormal Return (%) over 3 Years | | | | | | | | |
Sample Type	−20	−15	−10	−5	0	5	10	15	20
Equal weight	1.00	1.00	1.00	.60	.02	.13	.89	1.00	1.00
Value weight	1.00	.99	.81	.33	.07	.36	.83	.99	1.00

FIG. 2.—Simulated power of equal-weight and variable-weight calendar-time portfolio regressions test statistics: the percentage of 1,000 random samples of 2,000 firms rejecting the null hypothesis of no abnormal performance at various induced levels of abnormal return.

report the results and plot the power functions. The calendar-time portfolio regressions and CTARs are better specified when the portfolios are value weighted rather than equally weighted. Moreover, both of the calendar-time portfolio methodologies reject over 99% of the time with induced abnormal performance of ±15% over 3 years, regardless of the weighting scheme.

We conclude that the calendar-time portfolio approach has sufficient power to detect abnormal performance over economically important ranges. Moreover, in direct contrast to the claims of Loughran and Ritter (in press), we find that the calendar-time portfolio approach has

TABLE 15   **Percentage of Samples Rejecting the Null Based on Mean Calendar-Time Portfolio Abnormal Returns**

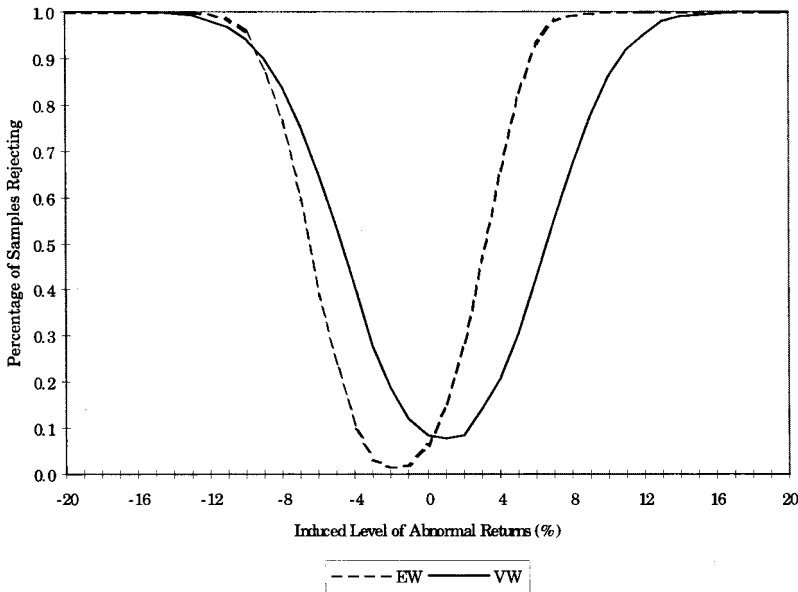| Sample Type	Induced Level of Abnormal Return (%) over 3 Years								
	−20	−15	−10	−5	0	5	10	15	20
Equal weight	1.00	1.00	.96	.24	.06	.83	1.00	1.00	1.00
Value weight	1.00	1.00	.94	.54	.09	.30	.86	1.00	1.00

FIG. 3.—Simulated power function for calendar-time abnormal returns (CTARs). The percentage of 1,000 random samples of 2,000 firms rejecting the null hypothesis of no abnormal performance at various induced levels of abnormal returns. CTARs are calculated relative to the Fama and French three-factor model.

more power to detect abnormal performance than the BHAR approach after accounting for cross-sectional dependence of individual-firm abnormal returns.

VI.   Interpreting the Results

A.   *Violating the Assumption of Independence*

Because of positive cross-sectional correlation of individual firm abnormal returns, which is increasing in holding period, the *p*-values associated with the BHARs are surely overstated. For our three samples,

standard errors that account for cross-sectional correlations of annual abnormal returns are over four times as large for the SEO sample and roughly 2.5 times those that assume independence for the merger and share repurchase samples. Consistent with earlier research by Bernard (1987), we find that the average correlation of individual-firm abnormal returns increases dramatically with the holding period, suggesting that the standard errors assuming independence for 3-year BHARs are even more severely overstated.

This point is often dismissed in long-term event studies, which frequently claim that bootstrapping solves all dependence problems. However, this claim is not valid. An empirical distribution created by randomly selecting firms with similar size-BE/ME characteristics does not replicate the covariance structure underlying the original event sample. In fact, the typical bootstrapping approach does not even capture the cross-sectional correlation structure related to industry effects documented by Bernard (1987), Brav (in press), and others. Moreover, Bernard shows that the average interindustry cross-sectional correlation of individual abnormal returns is also positive. Our estimate of the average correlation of 3-year BHARs for SEOs is only 0.0035, but this is important when there are over 9.8 million unique correlations.

There are essentially three approaches to dealing with cross-sectional correlation of abnormal returns. The first approach is to ignore the problem by assuming that all event announcements are independent and that event firms are directly comparable to randomly selected nonevent firms. The second approach is to recognize that cross-sectional dependence may be a serious problem and estimate the covariance structure. The final approach is to form calendar-time portfolios, which completely avoids the problems associated with cross-sectional dependence. In light of the evidence that the event-firm abnormal returns for our three samples show considerable cross-sectional dependence, ignoring the problem is clearly not appropriate. Moreover, the approaches prescribed by Brav (in press) and Lyon et al. (1999) do not provide a complete correction to the dependence problem. Like Fama (1998), we strongly advocate the use of a calendar-time portfolio approach.

B.   On the Joint-Test Problem and on Attributing Mispricing to "Event" Samples

All tests of long-term stock price performance are necessarily joint tests of market efficiency and the assumed model of expected returns, whether an asset-pricing model, such as the Fama and French three-factor model, or some other expected return benchmark, such as the 25 size-BE/ME portfolios. With this in mind, it is interesting to examine whether the measured abnormal performance is merely a manifestation of known mispricings of the Fama and French three-factor model. In

other words, it is clear that there is some reliable mispricing of securities, given the model. However, it is not clear that all of the mispricing is unique to the event firms.

Fama and French (1993, 1997), Fama (1998), and Davis, Fama, and French (in press) emphasize that the Fama and French three-factor model is unable to describe completely the cross-section of expected returns, even on the dimensions on which it is based. This is illustrated by the several statistically significant intercepts from the generic size-BE/ME portfolios reported in table 7. It is interesting that across all three event samples, for both EW and VW, none of the full-sample postevent average abnormal returns fall outside of the range of the generic size-BE/ME portfolio intercepts. Although the EW acquirer and issuer portfolios have significantly negative intercepts, the estimates are smaller in magnitude than some of the intercepts that are deemed sufficiently small to justify use of the model in the first place.

This suggests that the null hypothesis—intercept equals zero—may be problematic for samples tilted toward characteristics that the model cannot price in the first place. We find that much of the mispricing typically attributed to an ''event'' is actually due to a more general phenomenon with which the event firms happen to be correlated. For example, comparison of the estimated intercepts and the adjusted intercepts for the EW portfolios suggests that one-third of the estimated abnormal performance for mergers and SEOs is due to model misspecification rather than to the ''event.'' The model misspecification is especially severe for the EW event portfolios composed of ''growth'' firms. In particular, only one-half of the estimated abnormal return for growth acquirers and virtually none of the estimated abnormal returns for growth-equity issuers and share repurchasers are unique to the event firms per se. Instead, the Fama and French three-factor model overestimates the expected returns of otherwise similar nonevent growth firms just as poorly. This is especially severe for small growth firms.

C. Economic Significance

Interpreting the economic significance of the estimates is nearly as tricky as assessing the statistical reliability. Consider the SEO sample, which is associated with the most reliable abnormal return estimate. The EW and VW calendar-time portfolio regressions imply 3-year abnormal returns of −7.9% and 0.0%, respectively, after adjusting for the expected abnormal performance, given the sample composition. On an equal-weight basis, where all firms are treated as equally important, this is suggestive of serious mispricing. However, the reliable evidence of abnormal performance is restricted to the smallest stocks. After controlling for the known pricing deficiencies of the Fama and French three-factor model, only the smallest quintile of SEO firms have reliably significant abnormal returns, and this only on an EW basis (results

not reported). Fama and French (1993) report that the firms in the smallest quintile (based on NYSE breakpoints) account for only 2.8% of the value of the CRSP VW stock market, on average. Although this represents a large number of firms, it is not clear how economically important this portion of the market is for assessing overall stock market efficiency.

In general, the VW average abnormal returns are not very large in economic terms. For the full samples, the VW calendar-time portfolio adjusted intercepts correspond to a range of -1.4% to 0.0% over 3 years. Finally, it is interesting to note that for the firms in the largest size quintile—on average representing 73.9% of the value of the CRSP VW stock market—none of the adjusted intercepts from the calendar-time portfolio regressions is reliably different from the expected null for any of the three samples, regardless of whether EWs or VWs are used (results not reported).

VII.   Conclusion

This article reexamines the reliability of recent long-term stock price performance estimates using three large well-explored samples of major corporate events. We find that the popular approach of measuring long-term abnormal performance with mean BHARs in conjunction with bootstrapping is not an adequate methodology because it assumes independence of multiyear event-firm abnormal returns. We show that event-firm abnormal returns are positively cross-correlated when overlapping in calendar time. As such, assuming independence is problematic for any long-term abnormal performance methodology. Moreover, this is likely to be a problem for most event samples, not just the mergers, SEOs, and share repurchases examined in this article, since major corporate actions are not random. As a result, we strongly advocate a methodology that accounts for the dependence of event-firm abnormal returns, such as the calendar-time portfolio approach.

The primary implication of our results is that most of the evidence against market efficiency contained in recent studies measuring significant long-term abnormal returns following major corporate events is largely irrelevant because these studies assume independence. Our estimates of long-term abnormal performance that account for the positive cross-correlations of event-firm abnormal returns produce very little evidence of long-term abnormal performance.

Appendix

Calculating ''Corrected'' *t*-Statistics

To calculate ''corrected'' *t*-statistics that account for cross-correlation of 3-year BHARs, we assume a very simple correlation structure to calculate the average

TABLE A1   Correlation Structure for the Seasoned Equity Issuer Sample

Number of Months of Overlap	Number of Unique Correlations $n(n-1)/2$	Assumed Correlation Structure	Estimated Correlation
36	50,241	ρ	.02576
35	96,914	$35/36 \cdot \rho$	.02504
34	92,094	$34/36 \cdot \rho$	.02433
33	88,297	$33/36 \cdot \rho$	.02361
32	84,531	$32/36 \cdot \rho$	.02290
31	80,937	$31/36 \cdot \rho$	.02218
30	79,158	$30/36 \cdot \rho$	.02147
29	75,881	$29/36 \cdot \rho$	.02075
28	73,894	$28/36 \cdot \rho$	.02004
27	72,001	$27/36 \cdot \rho$	.01932
26	71,554	$26/36 \cdot \rho$	.01860
25	70,956	$25/36 \cdot \rho$	.01789
24	71,347	$24/36 \cdot \rho$	.01717
23	69,456	$23/36 \cdot \rho$	.01646
22	68,503	$22/36 \cdot \rho$	.01574
21	66,132	$21/36 \cdot \rho$	.01503
20	66,093	$20/36 \cdot \rho$	.01431
19	65,627	$19/36 \cdot \rho$	.01360
18	65,674	$18/36 \cdot \rho$	.01288
17	63,926	$17/36 \cdot \rho$	.01216
16	63,106	$16/36 \cdot \rho$	.01145
15	62,344	$15/36 \cdot \rho$	.01073
14	62,531	$14/36 \cdot \rho$	.01002
13	63,139	$13/36 \cdot \rho$	.00930
12	64,265	$12/36 \cdot \rho$	.00859
11	62,811	$11/36 \cdot \rho$	.00787
10	63,017	$10/36 \cdot \rho$	.00716
9	61,357	$9/36 \cdot \rho$	.00644
8	62,370	$8/36 \cdot \rho$	.00572
7	61,621	$7/36 \cdot \rho$	.00501
6	61,945	$6/36 \cdot \rho$	.00429
5	61,800	$5/36 \cdot \rho$	.00358
4	62,704	$4/36 \cdot \rho$	.00286
3	63,196	$3/36 \cdot \rho$	.00215
2	63,684	$2/36 \cdot \rho$	.00143
1	63,357	$1/36 \cdot \rho$	.00072
0	7,373,678	0	.00000

NOTE.—The average of the estimated correlation is 0.00351.

correlation of 3-year BHARs across all of our 4,439 observations in the SEO sample. The correlation matrix requires only an estimate of the average correlation of 3-year BHARs for sample firms with complete (36 months) calendar-time overlap. This is because we assume that the correlation is decreasing linearly as the amount of overlap falls from complete calendar-time overlap of 36 months to no overlap between observations (see table A1). We assume that the average correlation of BHARs with 36 months of calendar-time overlap is $\rho = 0.02576$. Recall that this is our estimate using annual BHARs, which almost surely understates the true correlation of 3-year BHARs with 36 months of overlap. We then

estimate the correlation of 3-year BHARs with 35 months of overlap as $35/36 \cdot \rho = 0.02504$, and so on. The estimate for nonoverlapping observations is zero. This procedure produces an average correlation for the BHARs of 0.00351.

We then use the average correlation of the 3-year BHARs to adjust the t-statistic that assumes independence. Using equation (4), we are able to determine by how much the t-statistic that assumes independence is understated:[16]

$$\frac{\sigma_{\text{BHAR}}(\text{independence})}{\sigma_{\text{BHAR}}(\text{dependence})} \approx \frac{1}{\sqrt{1 + (N-1)\overline{\rho}_{i,j}}} = \frac{1}{\sqrt{1 + (4{,}439 - 1) \cdot 0.00351}}. \quad (4)$$

Although the average correlation is small, the standard deviation that assumes independence is less than one-fourth the magnitude of the standard deviation that accounts for cross-correlation. This translates directly into t-statistics that are four times too large if observations are assumed to be independent (-6.05 vs. -1.49).

References

Andrade, Gregor, and Stafford, Erik. 1999. Investigating the economic role of mergers. Working paper. Boston: Harvard Business School.

Barber, Brad M., and Lyon, John D. 1997. Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics* 43 (March): 341–72.

Bernard, Victor L. 1987. Cross-sectional dependence and problems in inference in market-based accounting research. *Journal of Accounting Research* 25 (Spring): 1–48.

Brav, Alon. In press. Inference in long-horizon event studies: A parametric bootstrap approach with application to initial public offerings. *Journal of Finance*.

Brav, Alon; Geczy, Christopher C.; and Gompers, Paul A. In press. Is the abnormal return following equity issuances anomalous? *Journal of Financial Economics*.

Brav, Alon, and Gompers, Paul A. 1997. Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture- and nonventure-capital-backed companies. *Journal of Finance* 52 (December): 1791–1820.

Brock, William; Lakonishok, Josef; and LeBaron, Blake. 1992. Simple trading rules and the stochastic properties of stock returns. *Journal of Finance* 47 (December): 1731–64.

Chan, Louis K. C.; Jegadeesh, Narasimhan; and Lakonishok, Josef. 1995. Evaluating the performance of value versus glamour stocks: The impact of selection bias. *Journal of Financial Economics* 38 (July): 269–96.

Chung, K. L. 1974. *A Course in Probability Theory*. New York: Academic Press.

Collins, Daniel W., and Dent, Warren T. 1984. A comparison of alternative testing methodologies used in capital market research. *Journal of Accounting Research* 22 (Spring): 48–84.

Comment, Robert, and Schwert, G. William. 1995. Poison or placebo? Evidence on the deterrent and wealth effects of modern antitakeover measures. *Journal of Financial Economics* 39 (September): 3–43.

Davis, James L.; Fama, Eugene F.; and French, Kenneth R. In press. Characteristics, covariances, and average returns: 1929 to 1997. *Journal of Finance*.

Efron, Bradley, and Tibshirani, Robert J. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall.

Fama, Eugene F. 1970. Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25 (May): 383–417.

Fama, Eugene F. 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49 (September): 283–306.

Fama, Eugene F., and French, Kenneth R. 1992. The cross-section of expected stock returns. *Journal of Finance* 46 (June): 427–65.

16. The approximation of the ratio of $\sigma(\text{independence})$ to $\sigma(\text{dependence})$ assumes equal variances of the individual BHARs.

Fama, Eugene F., and French, Kenneth R. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (February): 3–56.

Fama, Eugene F., and French, Kenneth R. 1997. Industry costs of equity. *Journal of Financial Economics* 43 (February): 153–93.

Franks, Julian; Harris, Robert S.; and Titman, Sheridan. 1991. The postmerger share-price performance of acquiring firms. *Journal of Financial Economics* 29 (March): 81–96.

Horowitz, Joel L. 1996. Bootstrap methods in econometrics: Theory and numerical performance. In D. M. Kreps and K. F. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications*. Seventh World Congress. Cambridge: Cambridge University Press.

Ikenberry, David; Lakonishok, Josef; and Vermaelen, Theo. 1995. The underreaction to open market share repurchases. *Journal of Financial Economics* 39 (October): 181–208.

Jaffe, Jeffrey. 1974. Special information and insider trading. *Journal of Business* 47 (July): 411–28.

Kothari, S. P.; Shanken, Jay; and Sloan, Richard G. 1995. Another look at the cross-section of expected stock returns. *Journal of Finance* 50 (March): 185–224.

Kothari, S. P., and Warner, Jerold B. 1997. Measuring long-horizon security performance. *Journal of Financial Economics* 43 (March): 301–39.

Loughran, Tim, and Ritter, Jay R. In press. Uniformly least powerful tests of market efficiency. *Journal of Financial Economics*.

Loughran, Tim, and Vijh, Anand M. 1997. Do long-term shareholders benefit from corporate acquisitions? *Journal of Finance* 52 (December): 1765–90.

Lyon, John D.; Barber, Brad; and Tsai, Chih-Ling. 1999. Improved methods for tests of long-run abnormal stock returns. *Journal of Finance* 54 (February): 165–201.

Mandelker, Gershon. 1974. Risk and return: The case of merging firms. *Journal of Financial Economics* 1 (December): 303–35.

Mitchell, Mark L., and Mulherin, J. Harold. 1996. The impact of industry shocks on takeover and restructuring activity. *Journal of Financial Economics* 41 (June): 193–229.

Rau, P. Raghavendra, and Vermaelen, Theo. 1998. Glamour, value, and the post-acquisition performance of acquiring firms. *Journal of Financial Economics* 49 (August): 223–53.

Ritter, Jay R. 1991. The long-run performance of initial public offerings. *Journal of Finance* 46 (March): 3–27.

Sefcik, Stephen E., and Thompson, Rex. 1986. An approach to statistical inference in cross-sectional models with security abnormal returns as dependent variables. *Journal of Accounting Research* 24 (Autumn): 316–34.

Shumway, Tyler. 1997. The delisting bias in CRSP data. *Journal of Finance* 52 (March): 327–40.

Speiss, D. Katherine, and Affleck-Graves, John. 1995. Underperformance in long-run stock returns following seasoned equity offerings. *Journal of Financial Economics* 38 (July): 243–67.