Journal of Financial Economics 55 (2000) 361–389

JOURNAL OF Financial ECONOMICS

www.elsevier.com/locate/econbase

Uniformly least powerful tests of market efficiency☆

Tim Loughran[a], Jay R. Ritter[b,*]

[a]P.O. Box 399, University of Notre Dame, Notre Dame IN 46556-0399, USA

[b]Department of Finance, University of Florida, Gainesville FL 32611-7168, USA

Abstract

Defenders of market efficiency argue that anomalies involving long-term abnormal returns are not robust to alternative methodologies. We argue that because various methodologies use different weighting schemes, the magnitude of abnormal returns *should* differ, and in a predictable manner. Three problems are identified that cause low power in value-weighted three-factor time series regressions when abnormal returns following managerial actions are being estimated. We illustrate the sensitivities in the context of the new issues puzzle as well as with simulations. More generally, multifactor models as currently used do not, and cannot, test market efficiency.

JEL classification: G12; G14

Keywords: Market efficiency; Anomalies; New issues puzzle; Risk factors

☆We thank Eugene Fama for providing the Fama-French monthly time series of factor realizations. Brad Barber, Alon Brav, John Campbell (a referee), Eugene Fama (a referee), Kenneth French, Todd Houge, David Ikenberry, Inmoo Lee, Mark Mitchell, Forrest Nelson, Bill Schwert, Richard Thaler, Ivo Welch, and seminar participants at the Universities of Central Florida, Florida, Missouri, and Notre Dame, the Chinese University of Hong Kong, Copenhagen Business School, the Norwegian School of Management, Tulane, Virginia Tech, and the NBER provided helpful comments. We thank Hsuan-Chi Chen for research assistance. http://bear.cba.ufl.edu/ritter/index.html.

* Corresponding author. +1-352-846-2837; fax: +1-352-392-0301.

E-mail addresses: loughran.9@nd.edu (T. Loughran), ritter@dale.cba.ufl.edu (J.R. Ritter)

0304-405X/00/$ - see front matter
PII: S0304-405X(99)00054-9

1. Introduction

In the last two decades, many long-term return anomalies have been uncovered in event studies. Because the magnitude, and sometimes even the sign, of the abnormal returns are sensitive to alternative measurement methodologies, the existence and importance of these abnormal returns have been questioned. These anomalies suffer from the joint-hypothesis problem: to test whether there is an inefficiency, one must know what 'normal' returns should be, and whether the actual returns deviate from this benchmark. Because theoretically motivated asset pricing models have little corroborating empirical support, there is no consensus on how to measure long-term abnormal returns. This joint-hypothesis problem, combined with the low signal-to-noise ratio of stock returns (actual returns have a low correlation with expected returns), has led many to question the importance of these anomalies.

Most prominently, Fama (1998) argues that the anomalies literature suffers from data-mining. He emphasizes that the magnitude of abnormal returns is rarely robust to alternative methodologies. Consequently, Fama argues that the existence of any reliable patterns is unproven, and the paradigm of market informational efficiency should be maintained.

In this paper, we argue that if there are significant misvaluations in the stock market, abnormal returns should *not* be robust to alternative methodologies. In particular, some methods have little power to pick up material misvaluations that display predictable patterns. (The power of a test is the probability of rejecting the null hypothesis when the hypothesis is false.) Other methods have much more power to identify these misvaluations, as manifested in subsequent abnormal returns.

To test whether there are abnormal returns, a benchmark model is needed. We argue that the researcher must choose between normative models such as the capital asset pricing model or positive models such as controlling for size and book-to-market. Tests of market efficiency *require* that a normative (equilibrium) model be used as a benchmark. If a positive (empirically based) model is used, one is not testing market efficiency; instead, one is merely testing whether any patterns that exist are being captured by other known patterns.

A common metric for measuring abnormal returns in recent years is the Fama and French (1993) three-factor model, used in time-series regressions with monthly portfolio returns. We argue that, for three reasons, this metric will have low power to identify abnormal returns for events that occur as a result of behavioral timing, especially when value-weighted portfolio returns are used. Two of these reasons are related to the patterns that we would expect to see if firms are taking advantage of misvaluations, and relate to weighting schemes.

(1) *Equally weighting each firm versus equally weighting each time period*: If there are time-varying misvaluations that firms capitalize on by taking some action (a supply response), there will be more events involving larger

misvaluations in some periods than in others. Examples of this would include junk bond issuance, equity-financed acquisitions, equity issues, and share repurchases.[1] In general, tests that weight firms equally should have more power than tests that weight each time period equally.

(2) *Equally weighting versus value-weighting returns*: If percentage misvaluations are greater among small firms than among big firms, then tests that weight firms equally should find greater abnormal returns than tests that weight firms by market capitalization. Consistent with this statement, just about every known stock market pattern is stronger for small firms than for big firms. Value-weighted portfolios can also have some periods in which a single firm is a large proportion of the portfolio, resulting in a high variance of returns because this firm's unique risk is not diversified away. The resulting low power will manifest itself in large standard errors and low t-statistics.

If significant misvaluations are as common among big firms as among small firms, then the expected point estimates of abnormal returns on value-weighted portfolios would be the same as for equally weighted portfolio returns. There are good reasons, however, to expect that significant misvaluations are more common, and are larger, among small firms than among big firms. The logic is simple. Assume that cognitive biases result in misvaluations that, in the absence of arbitrageurs, are equal in magnitude for both small and big firms. The vast majority of small-cap stocks have wider percentage bid–ask spreads than do large-cap stocks, making trading costs higher. Furthermore, the ability to buy or sell large quantities of shares without affecting the price is less for small-cap stocks, as shown by, among others, Hasbrouck (1991, Table 1). Consequently, the ability to capitalize on the same percentage misvaluation for a small stock will be less than for a big stock. Percentage misvaluations, in equilibrium, will be larger for small stocks. Otherwise, arbitrageurs could make more money, net of costs, by finding misvaluations among big stocks. This is the logic in Shleifer and Vishny (1990,1997). Thus, for any given misvaluations that occur, there will be a stronger force pushing the price towards fundamental value (and thus limiting the magnitude of any misvaluation) for big stocks.[2]

[1] For example, in Loughran and Ritter (1995, Tables I and II), the five-year wealth relative for initial public offerings is 0.70 when each firm is weighted equally, but 0.79 when each of the 21 cohort years are weighted equally. For seasoned equity offerings, the five-year wealth relative is 0.69 weighting each firm equally, but 0.76 when each of the 21 cohort years is weighted equally. A wealth relative of 1.00 represents no abnormal performance.

[2] We are not asserting that it is always correct to use equal weights, rather than value weights. If one is trying to measure the abnormal returns on a value-weighted portfolio with an equal amount of money invested each time period, then it is appropriate to value-weight returns and equally weight time periods. But if one is trying to measure the abnormal returns on the average firm undergoing some event, then each firm should be weighted equally. Alternatively stated, a traditional event study approach in which all observations are weighted equally will produce point estimates that are relevant from the point of view of a manager, investor, or researcher attempting to predict the abnormal returns associated with a random event. More generally, as Fama (1998) notes, the weighting scheme should be determined by the economic hypothesis of interest.

The third reason is a straightforward statistical issue:

(3) *Benchmark contamination*: A test is biased towards high explanatory power and no abnormal returns if it uses a benchmark that is contaminated with many of the firms that are the subject of the test. In the limit, the minimum power test is when the benchmark is composed of the same firms, with the same weights, as the sample being tested. If this is true, there will always be zero abnormal performance. The maximum power test uses a benchmark that is constructed to have none of the stocks in the sample as part of the benchmark.

When abnormal returns are being calculated using buy-and-hold returns, it is common practice to calculate the benchmark buy-and-hold returns by matching on characteristics such as size and book-to-market after excluding event firms. We are recommending that the same purging of the benchmarks be done when multifactor models are being used. (Campbell et al. (1997, Chapter 6) and Fama (1996), among others, discuss multifactor models.) If one views size and book-to-market as priced equilibrium risk factors in a multifactor asset pricing model, as does Fama (1998), then for mean-variance efficiency considerations, it is not appropriate to purge these factors of sample firms. Because we are willing to accept the market factor as an equilibrium priced risk factor, we do not purge the market portfolio of sample firms.

If one does not accept a given multifactor model as an equilibrium model, there is a legitimate question about the validity of its use as a benchmark. Our justification is that the Fama-French three-factor model is widely used for calculating abnormal returns, and when testing an anomaly, it is a legitimate empirical question to ask whether a pattern is distinct from other cross-sectional patterns such as size and book-to-market, or merely a manifestation of those patterns.

In this paper, we focus on the class of anomalies related to managerial actions involving cash flows (such as equity issues, stock-financed acquisitions, and share repurchases) rather than routine events (quarterly earnings announcements), actions not involving cash flows (stock splits), or samples constructed from rankings (extreme winners and losers). Behavioral timing is a potential explanation for the events that are managerial choice variables, but not for the routine events or rankings. In other words, if managers take advantage of windows of opportunity (temporary misvaluations) to benefit existing shareholders, there should be a supply response, as modeled by Stein (1996). If misvaluations are more extreme for smaller firms, and are correlated among firms with similar characteristics, we expect that events driven by behavioral timing motivations should display time and industry clustering, and should be more common among smaller firms. Note that behavioral timing is not the *cause* of misvaluations, it is the *response* to them.

We present simulation results that are designed to mimic some of the samples employed in current empirical research. Our samples are constructed so that the average abnormal returns that are calculated differ depending upon whether equal- or value-weighting is used, and whether averages are calculated with the weights on individual firms or on calendar periods. In these simulations, we create time variation in the volume of sample observations ('events'), with high-volume periods having greater misvaluations, and with small firms being more misvalued than big firms. We find that abnormal returns calculated using buy-and-hold returns and benchmark portfolios chosen on the basis of size, with each firm weighted equally, capture close to 90% of the true abnormal returns. Using buy-and-hold returns and benchmark portfolios chosen on the basis of both size and book-to-market captures about 80% of the true abnormal returns. But when the Fama-French three-factor model is used with value-weighted portfolios, only about half of the true abnormal returns are captured. This occurs because, by construction, we have more severe misvaluations among small firms and in high-volume periods.

We also examine the underperformance of firms conducting initial public offerings and seasoned equity offerings, a pattern known as the new issues puzzle. We present the results of Fama-French three-factor regressions after purging the factors of new issues. As predicted, we find substantially greater underperformance using decontaminated factors than when using the typical implementation of the Fama-French three-factor model. This underperformance is most severe in high-volume periods. We address the concern, raised by Brav et al. (1999) and others, that the negative intercepts in three-factor regressions are due to the misspecification of the three-factor model for predicting returns on small growth firms. We show that issuing firms underperform nonissuers across all size and book-to-market categories.

The remainder of this paper is organized as follows. Section 2 describes the construction of the factors used in the Fama-French three-factor model and shows how these factors are frequently intensive in the firms being analyzed. Section 3 presents simulation results that show the distribution of abnormal performance measured using various methodologies. Section 4 addresses the new issues puzzle, and shows that the underperformance of new issues is sensitive to how the Fama-French three-factor model is implemented in a predictable manner. Finally, Section 5 summarizes the paper.

2. The construction of the three factors and factor contamination

The Fama and French (1993) three-factor model is a parsimonious model for stock returns taking the form

$$r_{pt} - r_{ft} = a + b(r_{mt} - r_{ft}) + s\text{SMB}_t + h\text{HML}_t + e_{pt}, \tag{1}$$

where $(r_{mt} - r_{ft})$ is the market factor, constructed by subtracting the T-bill return from the value-weighted market return. SMB is the size factor, constructed by taking the return on a portfolio of small stocks minus the return on a portfolio of big stocks. HML is the book-to-market factor, formed by taking the return on a portfolio of value stocks (with high book-to-market ratios) and subtracting the return on a portfolio of growth stocks (with low book-to-market ratios).

An understanding of how the factors are formed is important. Fama and French (1993, pp. 8–9) form six portfolios based upon two size groupings and three book-to-market groupings, and describe the construction of the factors as follows:

> In June of each year t from 1963 to 1991, all NYSE stocks on CRSP are ranked on size (price times shares). The median NYSE size is then used to split NYSE, Amex, and (after 1972) NASDAQ stocks into two groups, small and big (S and B). Most Amex and NASDAQ stocks are smaller than the NYSE median, so the small group contains a disproportionate number of stocks (3616 out of 4797 in 1991). Despite its large number of stocks, the small group contains far less than half (about 8% in 1991) of the combined value of the two size groups.
>
> We also break NYSE, Amex, and NASDAQ stocks into three book-to-market equity groups based on the breakpoints for the bottom 30% (Low), middle 40% (Medium), and top 30% (High) of the ranked values of BE/ME for NYSE stocks. We define book common equity, BE, as the COMPUSTAT book value of stockholder's equity, plus balance-sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use the redemption, liquidation, or par value (in that order) to estimate the value of preferred stock. Book-to-market equity, BE/ME, is then book common equity for the fiscal year ending in calendar year $t-1$, divided by market equity at the end of December of $t-1$. We do not use negative-BE firms, which are rare before 1980, when calculating the breakpoints for BE/ME or when forming the size-BE/ME portfolios. Also, only firms with ordinary common equity (as classified by CRSP) are included in the tests. This means that ADRs, REITs, and units of beneficial interest are excluded.
>
> We construct six portfolios (S/L, S/M, S/H, B/L, B/M, B/H) from the intersections of the two ME and the three BE/ME groups. For example, the S/L portfolio contains the stocks in the small-ME group that are also in the low-BE/ME group, and the B/H portfolio contains the big-ME stocks that also have high BE/MEs. Monthly value-weighted returns on the six portfolios are calculated from July of year t to June of $t+1$, and the portfolios are reformed in June of $t+1$. We calculate returns beginning in July of year t to be sure that book equity for year $t-1$ is known.
>
> To be included in the tests, a firm must have CRSP stock prices for December of year $t-1$ and June of t and COMPUSTAT book common equity for year $t-1$. Moreover, to avoid the survival bias inherent in the way

COMPUSTAT adds firms to its tapes (Banz and Breen, 1986), we do not include firms until they have appeared on COMPUSTAT for two years. (COMPUSTAT says it rarely includes more than two years of historical data when it adds firms).

Size – Our portfolio SMB (small minus big), meant to mimic the risk factor in returns related to size, is the difference, each month, between the simple average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the simple average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). Thus, SMB is the difference between the returns on small- and big-stock portfolios with about the same weighted-average book-to-market equity. This difference should be largely free of the influence of BE/ME, focusing instead on the different return behaviors of small and big stocks.

BE/ME – The portfolio HML (high minus low), meant to mimic the risk factor in returns related to book-to-market equity, is defined similarly. HML is the difference, each month, between the simple average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). The two components of HML are returns on high- and low-BE/ME portfolios with about the same weighted-average size. Thus the difference between the two returns should be largely free of the size factor in returns, focusing instead on the different return behaviors of high- and low-BE/ME firms. As testimony to the success of this simple procedure, the correlation between the 1963–1991 monthly mimicking returns for the size and book-to-market factors is only -0.08.

In Panel A of Table 1, we report the average percentage of market capitalization represented by each of the six portfolios constructed using the Fama-French methodology. Inspection of this panel shows that, on average, large growth stocks represent 44% of market cap, whereas small value stocks represent only 3% of market cap. Thus, the equal weighting of each of the six portfolios greatly underweights large growth stocks and overweights small stocks relative to their market capitalization. It should be noted that since each of the six component portfolios is value-weighted, the three small portfolios are actually more like mid-cap portfolios than the small-cap portfolios that are normally used in academic studies using small stocks.

Our sample period covers 1973–1996. We start a decade later than Fama and French for several reasons. Because much of our empirical work deals with the new issues puzzle, the lack of Compustat data for most small firms before 1973 constrains the empirical work. Furthermore, Nasdaq did not start until February 1971, and CRSP does not include Nasdaq returns before December 1972. Rather than have some tables that use data beginning in 1963, and other tables that use data from 1973, we have chosen to have all of our empirical work cover the same sample period. Our returns end in December 1996 or June 1997,

Table 1
Average annual percent of market value, number of firms, proportion of sample that are IPOs/SEOs, and value-weighted annual returns, July 1973 to June 1997

At the end of June in year t, NYSE, Amex, and Nasdaq operating companies on both the CRSP and Compustat tapes with Compustat book values for years $t-1$ and $t-2$ are allocated into market capitalization groups (as of June of year t) whose cutoff is determined by the median NYSE firm. Firms are then allocated into book-to-market groups on the basis of the 30%, 40%, and 30% values of NYSE firms to form six portfolios. These six portfolios are used to form the SMB (small minus big) and HML (value minus growth) factors. Panel A reports the average (over the 24 cohorts) percent of the total sample market value in each portfolio. Panel B reports the average number of firms in each portfolio. Panel C lists the average value-weighted percentage of firms in the six portfolios that issued equity (either initial or seasoned offerings) during the previous three years in a general cash offer. For Panel D, all annual returns are value-weighted within each size and book-to-market group. The average annual returns equally weight each of the 24 years.

	Book-to-market groups		
	LOW (Growth)	2	HIGH (Value)
Panel A: Average annual percent of total sample market value			
SMALL	4.3%	4.1%	2.9%
BIG	43.8%	31.9%	13.0%
Panel B: Average number of firms			
SMALL	1062.5	1006.1	1185.9
BIG	326.8	297.9	140.6
Panel C: Average proportion that has issued equity (IPOs and SEOs) within last three years			
SMALL	33.8%	15.9%	10.3%
BIG	11.0%	12.3%	12.2%
Panel D: Average valued-weighted annual returns			
SMALL	15.1%	20.1%	23.0%
BIG	13.6%	16.0%	18.3%

depending upon the table, reflecting six years of additional data since Fama and French did their original work.

As shown in Panel A of Table 1, the SMB 'arbitrage' portfolio in a typical year has about 11% of total market cap in the small-stock portfolio, and about 89% in the big-stock portfolio. Recall that Fama and French's HML factor returns are constructed by forming six value-weighted portfolios [independent sorts of size (two categories) and book-to-market (three categories)]. The middle book-to-market portfolios are then discarded. The H portfolio is then constructed by *equally weighting* the small value portfolio (3% of market cap) and the big value portfolio (13% of market cap). The L portfolio is constructed by *equally weighting* the small growth portfolio (4% of market cap) and the big growth portfolio (44% of market cap). As shown in Panel C of Table 1, on a value-weighted basis an average of 34% of the stocks in the small growth portfolio have issued equity during the prior three years, versus 11% of the big growth stocks. In contrast to this large difference between small and big growth stocks, 10% of the small value stocks and 12% of the large value stocks have recently issued equity.

The B (big firm) portfolio is constructed as an equally weighted average of three value-weighted portfolios (B/L, B/M, and B/H). Inspection of Table 2 shows that in almost every year from 1973 to 1996, the S portfolio is more intensive in recent new issues than the B portfolio. Similarly, in almost every year, the H portfolio is less intensive in recent new issues than the L portfolio. The relatively high percentage of the H (value stocks) portfolio that is composed of recent new issues in some years is partly attributable to the fact that many issuing firms have migrated into the value portfolio within three years after the equity issue, even though they were growth firms at the time of issuing. On average the S portfolio has 8.2 percentage points more of its weight in recent issuers than the B portfolio (20.0% vs. 11.8%). On average the H portfolio has 11.1 fewer percentage points of its weight in recent issuers than the L portfolio (11.3% vs. 22.4%).

Because the S part of the SMB factor is usually intensive in recent issuers, this will bias a three-factor regression to have an intercept near zero and a high loading (slope coefficient) on SMB for a portfolio of issuing firms. (To be precise, the slope coefficients are subject to an omitted variable bias, where the hypothesized omitted variable is a 'new issues' factor (the returns on nonissuers minus issuers) whose factor realizations are correlated with SMB and HML. Of course, we do not consider this new issues factor to be an equilibrium risk factor.) Because the L part of the HML factor is usually intensive in recent issuers, a related bias will exist. A similar logic applies for the Loughran and Vijh (1997) stock-financed acquisition anomaly: At certain points many of the large growth firms in the HML arbitrage portfolio could have been recent stock-financed acquirers. In Section 4 below, we show how this contamination of the HML and SMB portfolios biases the coefficients when we address the new issues puzzle.

3. The power of the three-factor model to find misvaluations

We have argued that the Fama-French three-factor model, as commonly applied, has low power to find abnormal performance for events that are subject to behavioral timing considerations. In this section, we show the quantitative magnitude of the misspecification of the abnormal return distribution using

Table 2
Percentage of new issue firms in Fama and French's three-factor CRSP and Compustat universe, 1973–1996

On June 30 of each year t from 1973–1996, portfolios are formed that include CRSP-listed firms that have Compustat accounting data for years $t-1$ and $t-2$. For example, a company that went public in May 1987 and was listed on CRSP on June 30, 1987 would be included in the June 30, 1988 portfolios but not the June 30, 1987 portfolios if Compustat included its financials for the fiscal years ending during calendar years 1986 and later, but not for 1985. The last column lists the equally weighted percentage of the firms in the Fama-French factor portfolios that issued equity (either initial or seasoned offerings) during the previous three years in a general cash offer and are included in the Securities Data Co. new issues database. The three-year period covers July 1 of year $t-4$ to June 30 of year t. The proportion of firms in the S (small), B (big), H (value), and L (growth) groups are valued-weighted numbers, based upon the weighting schemes used by Fama and French (1993) and described in Section 2 of this paper.

Portfolio formation year (1)	(Small) % of S (2)	(Big) % of B (3)	Diff (2)–(3)	(High B/M) % of H (4)	(Low B/M) % of L (5)	Diff (5)–(4)	% of firms issuing equity during prior three years
1973	17.8	6.3	11.5	5.4	20.8	− 15.4	16.8
1974	17.6	4.9	12.7	8.0	17.9	− 9.9	16.0
1975	7.8	4.0	3.8	3.1	10.2	− 7.1	7.3
1976	4.5	9.1	− 4.6	0.0	10.2	− 10.2	3.5
1977	4.7	10.5	− 5.8	1.3	9.9	− 8.6	4.0
1978	5.5	10.5	− 5.0	2.6	7.7	− 5.1	4.4
1979	5.9	7.8	− 1.9	2.3	8.7	− 6.4	4.7
1980	7.9	10.6	− 2.7	12.7	12.3	0.4	5.9
1981	14.3	14.2	0.1	13.8	23.7	− 9.9	11.7
1982	16.8	13.1	3.7	14.1	22.8	− 8.7	15.9
1983	24.8	18.5	6.3	15.9	32.0	− 16.1	23.2
1984	23.9	13.0	10.9	9.2	30.8	− 21.6	23.7
1985	31.5	16.4	15.1	15.7	32.9	− 17.2	27.9
1986	26.0	13.1	12.9	13.8	27.3	− 13.5	24.5
1987	23.3	15.7	7.6	16.2	24.2	− 8.0	22.2
1988	26.5	14.2	12.3	21.1	22.9	− 1.8	25.1
1989	19.5	8.3	11.2	11.3	17.6	− 6.3	20.0
1990	16.7	6.2	10.5	7.3	16.4	− 9.1	13.1
1991	19.4	7.9	11.5	8.8	20.5	− 11.7	13.9
1992	26.7	16.3	10.4	17.2	27.9	− 10.7	21.1
1993	33.9	16.1	17.8	17.1	34.9	− 17.8	27.2
1994	37.1	20.2	16.9	23.8	37.9	− 14.1	31.9
1995	34.2	14.9	19.3	15.8	33.4	− 17.6	28.1
1996	33.0	11.9	21.1	13.5	34.0	− 20.5	27.8
Mean	20.0	11.8	8.2	11.3	22.4	− 11.1	17.5

simulations. For our size-adjusted and size- and book-to-market-adjusted returns, we compute abnormal returns two ways. The first procedure, similar to that in Barber and Lyon (1997, Table 6), assigns firms to portfolios and then adds in subsequent abnormal performance. The second procedure assigns firms to portfolios *after* their size and book-to-market ratios have been altered by the misvaluations. The reason for our second procedure is that we want to see how much of the abnormal returns caused by the misvaluations is incorrectly attributed to the benchmark returns. For example, if a firm with a fundamental value of $200 million and a book value of $200 million is overvalued by 100%, its market cap will be $400 million and its book-to-market ratio will be 0.50. If this misvaluation gets corrected over three years, a well-specified test should find that it will have a monthly abnormal return of minus 191 basis points, or about 23% per year.

Almost all managerial decisions that have been the subject of long-run performance studies (equity-financed acquisitions, equity issues, and share repurchases) display time-clustering that is correlated with prior market movements. The behavioral timing, or windows of opportunity, paradigm views these patterns as consequences of managers attempting to take advantage of time-varying misvaluations. Thus, for our simulations using undervalued firms, we will assume that there are more events following bear markets, with the undervaluations being more severe in the high-volume periods. Furthermore, we assume that the misvaluations are less severe among large firms (those with a market cap that would place them in the largest quintile of NYSE stocks). Symmetrically, for our simulations of overvaluations, we will assume that there are more events following bull markets, and that there are more severe misvaluations in the high-volume periods, with the overvaluations being less pronounced among large firms.

3.1. Simulation methodology

For our simulations, we use the same sample selection criteria as in Fama and French (1993), with the exception that we exclude stocks with a price of less than $3.00 at the time of portfolio formation. We do not impose the $3.00 minimum price screen on the benchmarks. The $3.00 screen is imposed at the time of the 'event'. If a stock subsequently falls below $3.00, it remains in the sample. The purpose of the $3.00 screen is to avoid an upward bias due to bid–ask spread 'bounce' in regressions using equally weighted monthly returns. For a firm to be included in our sample on June 30 of year t, it must be listed on the Center for Research in Securities Prices (CRSP) database on this June 30, and it must have Compustat book values for fiscal years $t-1$ and $t-2$, with the year $t-1$ book value being positive. The specifics of our undervaluation simulations are as follows. Each month, a random firm among the universe of qualifying CRSP- and Compustat-listed firms is chosen, and 5% is subtracted from its market value. The undervaluation is assumed to be corrected over a 36-month period by adding 14 basis points per month to the stock return during the 36 months following the event month.

Each month we compound the CRSP equally weighted market return during the prior 12 months. If this annual market return is worse than − 10%, five additional firms are chosen randomly from among the universe of eligible firms excluding the large firms in the highest quintile (using NYSE breakpoints) of market cap. These five firms are assumed to be undervalued by 50%, with the undervaluation corrected over a 36-month period by adding 194 basis points per month to the stock return during the 36 months following the event month. Note that $(1.0194)^{36} = 2.00$, so a stock that is undervalued by 50% will double its value relative to the market over the following three years.

For our overvaluation simulations, we add 5% to the market value, and subtract 14 basis points per month from the monthly returns, for one random firm per month. In months when the market return has exceeded 30% during the prior 12 months, five additional firms are chosen randomly from among the universe of eligible firms excluding the highest quintile of market cap, and 100% is added to their market value. The overvaluation is assumed to be corrected during the 36 months following the event month by subtracting 191 basis points per month from the stock return.

These assumptions result in a situation in which our sample 'event' firms display clustering, with the high-volume periods (six, rather than one, observations per month) tilted towards smaller firms and greater misvaluations. Thus, as is the case with most (all?) of the long-run anomalies studies, there is a correlation of misvaluations with both size and with volume. This correlation structure is intended to mimic the behavioral patterns in which firms take voluntary actions to capitalize on time-varying misvaluations that are larger in amplitude for smaller firms.

We have created a sample of 'events' in which on average small firms are misvalued by more than large firms, and there are greater misvaluations in high-volume periods than low-volume periods. So we have stacked the deck against the Fama-French three-factor model by using value-weighted portfolios and equally weighting each time period. Another way of looking at it, however, is that if there are greater misvaluations among small firms, and if there are windows of opportunity that firms take advantage of, the three-factor model, as typically used, is stacked against finding the misvaluations. That is, it has low power. And we should note that there is another reason for why the three-factor model is stacked against finding abnormal returns: the benchmark is partly composed of the sample. In our simulations, this last reason is not captured. This is because by adding or subtracting a constant return per month, there is no covariation with the factor returns, and the slope coefficients will thus be unaffected. None of the abnormal return will be improperly attributed to factor returns because, by construction, our abnormal returns are orthogonal to the factor realizations.[3]

When we compute abnormal returns using three-factor model time-series regressions, the event portfolio is reconstituted at the end of every month using all firms that have had an event during the previous 36 months. When we compute abnormal returns using reference portfolios matched by either size deciles or a 5×5 sort of size and book-to-market, we use two different algorithms for choosing reference portfolios. We assign our sample firms to the appropriate size decile or 5×5 portfolio either before or after the misvaluation is added or subtracted from market value. The purpose is to observe how much of the true abnormal return is incorrectly attributed to differences in the benchmark returns. When we assign firms to portfolios *after* the misvaluation affects the size and book-to-market ratio, by construction smaller firms are more likely to be undervalued and firms with low book-to-market ratios are more likely to be overvalued.

3.2. Simulation results

In Table 3, we report the results of our simulations on undervalued (Panel A) and overvalued (Panel B) firms. Using each methodology, we report the mean, minimum, median, and maximum values from 1000 simulations. When we weight each firm equally, we calculate the average annual abnormal percentage return as

$$\left\{\left[\prod_{j=1}^{3}\left\{1+\frac{1}{n_k}\sum_{i=1}^{n_k}\left[\prod_{t=k}^{k+11}(1+r_{it})-\prod_{t=k}^{k+11}(1+r_{bt})\right]\right\}\right]^{1/3}-1\right\}\times 100\% \quad \text{for } j=1,2,3 \tag{2}$$

where $k=12(j-1)+1$, and n_1, n_{13}, and n_{25} are the number of firms present in, respectively, event months 1, 13, and 25; r_{it} is the return on firm i in event month t; and r_{bt} is the return on the benchmark in event month t. If a firm is delisted during an event year, the CRSP equally weighted market return is substituted for the remainder of the year to calculate an annual return. In all cases, the annual benchmark return that is subtracted from the annual raw return on a sample firm is constructed by compounding the monthly benchmark portfolio return. Eq. (2) computes the average annual abnormal return in each of

[3] Eugene Fama has pointed out to us that there may actually be an induced correlation in our simulations. Because after bear markets our undervaluation simulations add high positive returns to our (small value) sample firms, if there are high subsequent return realizations on the three factors, our abnormal returns will not be orthogonal to the factor realizations. Brav (1998) presents simulations to calculate confidence intervals when clustering is present.

Table 3

Mean, minimum, median, and maximum abnormal performance calculations, equally weighting each observation or each calendar month, for 1000 samples of undervalued (Panel A) or overvalued (Panel B) firms

The sample is drawn, with replacement, for the 12 months after June 30 of year *t* from the population of CRSP-listed Amex, Nasdaq, and NYSE firms with Compustat-listed book value of equity for years *t*-1 and *t*-2, and with a price on June 30 of at least $3.00. Book value is defined as Compustat book value of equity plus deferred taxes and investment tax credits, minus the book value of preferred stock (using, in order, redemption, liquidation, and par values). Firms with negative book values in year *t*-1 are deleted. Returns are measured over January 1973 to December 1996, a period of 288 months.

In Panel A, the sample is constructed by taking one random firm each month and subtracting 5% from its market value. In addition, for each month in which the equally weighted (EW) market return is below − 10% during the prior 12 months, five random firms with a market cap below that of the top size quintile are chosen, and 50% is subtracted from the market value of each of these firms. There are 51 months that meet this criteria, so 255 firms are added to the sample. For all of these undervalued sample firms, it is assumed that their misvaluations are corrected over the 36 months following their event month. Thus, positive returns of 14 basis points per month are added to the 288 firms between January 1973 and December 1996 that are undervalued by 5%, and 194 basis points per month are added to the 255 firms that are undervalued by 50%. Because the rows 10 and 11 procedures use portfolios of firms that have undergone an event during the prior three years, we record 'events' beginning in July 1970. There are a total of 58 months starting in July 1970 meeting the bear market criteria, so 290 severely undervalued firms are added to the sample of 318 slightly undervalued firms from July 1970 to December 1996.

In Panel B, the sample is constructed in a manner similar to Panel A, with the difference that five firms are added in each of the 68 months (starting in January 1973) following an annual EW market return above + 30%, with 100% added to the market value of these companies. Returns of 191 basis points per month are subtracted from these firms during the following 36 months. For the one random firm per month that is chosen (288 firms), 5% is added to its market value, and returns of 14 basis points per month are subtracted during the next 36 months. When each observation is weighted equally, there are 628 overvalued firms from January 1973 to December 1996. When each time period is weighted equally, there are 677 sample overvalued firms beginning in July 1970.

In rows 1–9, with each firm weighted equally, average annual raw returns are the geometric average of the arithmetic average annual returns in calendar months 1–12, 13–24, and 25–36 relative to the event month. Benchmark-adjusted annual returns are calculated by subtracting the compounded monthly benchmark return from the raw annual return on a sample firm. For sample firms that are delisted early, an annual return is calculated by splicing in the CRSP EW market return for the remainder of the months in the event year. Since the return measurement period is truncated on December 31, 1996, for observations from late in the sample period, we treat the truncated returns starting in 1996 as if they are an annual return. Market-adjusted returns are calculated using the CRSP EW and VW market returns (including Amex, Nasdaq, and NYSE stocks). Size-adjusted returns are calculated using EW monthly returns on size decile portfolios, based upon NYSE market value rankings. Size- and book-to-market-adjusted annual returns are calculated by subtracting the compounded monthly returns on the 5×5 portfolio to which a sample firm is assigned. Size and book-to-market portfolios are rebalanced every month, but each sample firm is matched with the same reference portfolio to which it is originally matched. Size quintiles are based upon NYSE cutoffs, and book-to-market quintile cutoffs are calculated separately for each size quintile using all eligible Amex, Nasdaq, and NYSE stocks in each year.

In rows 10 and 11, with each time period weighted equally, Fama-French three-factor model intercepts are annualized by multiplying the average monthly intercept by 12. Intercepts are from the time-series regression $r_{pt} - r_{ft} = a + b(r_{mt} - r_{ft}) + s\mathrm{SMB}_t + h\mathrm{HML}_t + e_{pt}$. The sample firm portfolio is composed of all firms that had an event during the 36 months prior to a given month, with the portfolio rebalanced every month.

The mean, minimum, median, and maximum are the values from the 1000 sample portfolio means. The numbers in parentheses are the standard deviations of the 1000 sample portfolio means.

	Panel A: Undervalued firms Annual percentage returns				*Panel B: Overvalued firms* Annual percentage returns			
	Mean	Minimum	Median	Maximum	Mean	Minimum	Median	Maximum
Weighting each observation equally:								
(1) Raw returns	37.1% (1.5%)	32.7%	37.1%	42.1%	6.2% (1.1%)	3.3%	6.2%	9.4%
(2) Raw returns prior to adding misvaluations	21.0% (1.3%)	17.3%	21.0%	25.3%	20.8% (1.3%)	17.5%	20.8%	24.6%
(3) **True abnormal returns [(1) − (2)]**	**16.1%**				**−14.6%**			
(4) Equally weighted market-adjusted returns	6.6% (1.5%)	2.1%	6.6%	11.5%	− 22.1% (1.1%)	− 24.9%	− 22.1%	− 18.8%
(5) Value-weighted market-adjusted returns	22.0% (1.5%)	17.5%	22.0%	27.0%	− 8.2% (1.1%)	− 11.1%	− 8.3%	− 5.0%
(6) Size-adjusted returns (assigning sample firms to size decile portfolios BEFORE misvaluations are added)	14.8% (1.5%)	10.5%	14.9%	19.7%	− 14.9% (1.1%)	− 18.2%	− 15.0%	− 11.6%
(7) Size-adjusted returns (assigning sample firms to size decile portfolios AFTER misvaluations are added)	14.2% (1.5%)	9.8%	14.2%	18.9%	− 14.0% (1.1%)	− 17.4%	− 14.1%	− 10.8%
(8) Size- and book-to-market-adjusted returns (assigning sample firms to 5 × 5 portfolios BEFORE misvaluations are added)	14.0% (1.4%)	9.8%	14.0%	18.8%	− 14.3% (1.1%)	− 17.7%	− 14.3%	− 10.9%
(9) Size- and book-to-market-adjusted returns (assigning sample firms to 5 × 5 portfolios AFTER misvaluations are added)	11.8% (1.4%)	7.5%	11.8%	16.6%	− 11.8% (1.1%)	15.4%	− 11.9%	− 8.5%
Weighting each time-period equally:								
(10) Annualized EW 3-factor model intercepts	11.3% (1.2%)	7.0%	11.2%	14.6%	− 9.5% (1.2%)	− 13.3%	− 9.6%	− 5.4%
(11) Annualized VW 3-factor model intercepts	7.1% (2.0%)	− 0.4%	7.1%	13.1%	− 7.0% (1.8%)	− 12.6%	− 7.1%	− 0.7%

the three years after an event, and then takes the geometric mean of these three average returns.

In row 1 of Panel A (undervalued firms), we report the average annual raw return in the three years after the event. The average annual return of 37.1% is quite high, reflecting the artificially induced abnormal returns. Taking the difference between rows 1 and 2 shows that the average annual abnormal return is 16.1%. Thus, any well-specified methodology that weights each observation equally should calculate average abnormal returns of close to 16.1%.

In rows 4 and 5, we report the average annual market-adjusted return, using, respectively, the CRSP equally weighted (EW) and value-weighted (VW) market return. As can be seen, the market-adjusted returns are quite different using the EW and VW indices, reflecting the higher returns on small stocks during the first half of the sample period.

We report average annual size-adjusted returns in row 6 of Table 3, with firms assigned to portfolios *before* misvaluations are added. The mean size-adjusted return of 14.8% in Panel A is close to the artificially induced average abnormal return of 16.1% in our sample, suggesting that size-adjusted returns have some desirable properties.[4] In row 7, we report average annual size-adjusted returns with abnormal returns added *after* assigning firms to different size portfolios. Here, the mean abnormal return is 14.2%. The smaller abnormal return (14.2% versus 14.8%) that is measured when misvaluations are reflected in market cap changes suggests that part of the abnormal return is being improperly attributed to size effects in row 7.

In row 8, we report the average annual abnormal return calculated using size and book-to-market reference portfolios, with firms assigned to portfolios *before* misvaluations have been added. Here, the average abnormal return is 14.0%. In row 9, we report the average abnormal return using size- and book-to-market reference portfolios, with the assignment of firms *after* the misvaluations are added. The average abnormal return of 11.8% reflects the fact that many of our undervalued firms are matched with benchmark portfolios with high returns, underestimating the true abnormal performance.

Rows 10 and 11 of Table 3 report the Fama-French intercepts (annualized by multiplying the monthly intercepts by 12) for, respectively, equally weighted and value-weighted portfolios. The three-factor model estimates of abnormal performance for undervalued firms are 11.3% (EW) and 7.1% (VW).

These three-factor regressions estimate even less abnormal performance than when we weight each observation equally and adjust for size (rows 6 and 7 in

---

[4] Because our sample firms have been chosen randomly from among firms in specified size quintiles, before we add artificial misvaluations, there is no tilt towards extreme value firms or extreme growth firms. Certain real-world event studies, such as Loughran and Ritter (1995), do have a sample tilt, in which case using size-adjusted returns may not be as well-specified as in our simulations.

Table 3) or size and book-to-market (rows 8 and 9 in Table 3). The reason for this difference is that the Fama-French procedure misses the correlation of misvaluations and volume by equally weighting each time period. Also as expected, value-weighted three-factor regressions can substantially under-estimate abnormal performance when firms (especially small firms) take advantage of windows of opportunity. Furthermore, the value-weighted three-factor intercepts have the largest standard errors of any of the procedures, making it harder to find statistically significant abnormal returns. The higher standard errors are at least partly attributable to some periods when one or two large firms dominate a portfolio, so that little of the unique risk of the firms is diversified away in the portfolio.

In Panel B of Table 3, we report the results of our simulations on overvalued firms. The qualitative results are similar to those reported for undervalued firms in Panel A.

When using buy-and-hold returns, why do our simulations find that size-adjusted returns are better specified than the alternatives? The Barber and Lyon (1997), Kothari and Warner (1997), and Lyon et al. (1999) simulations do not find this. These articles present simulations in which firms are assigned to portfolios based on book-to-market and size rankings, and then a certain level of abnormal performance is specified.[5] The empirical rejection probabilities are then calculated, in order to examine how well-specified various procedures are. This methodology gives a false sense of confidence in the ability of portfolios formed on the basis of size and book-to-market to pick up misvaluations. In general, a firm that is overvalued to start with will have a lower book-to-market and higher size to begin with. The size effect will generally be less important than the book-to-market effect, because of the wider dispersion of size at a point in time than book-to-market ratios. *Ceteris paribus*, overvalued firms will tend to be assigned to low book-to-market portfolios, and undervalued firms will be in high book-to-market portfolios. These benchmark portfolios will have low returns for the overvalued firms and high returns for the undervalued firms, so there will be a bias towards finding zero abnormal returns on misvalued firms. When using buy-and-hold returns with benchmark portfolios in our Table 3, the quantitative effect of this bias is about 16–17% (− 11.8% instead of − 14.3% for our overvalued simulations, and 11.8% instead of 14.0% for our undervalued simulations) of the abnormal returns that are measured before reclassifying firms to reflect their altered market values.

[5] In Barber and Lyon (1997), ten size deciles (based upon NYSE rankings) and five book-to-market quintiles are used. In Lyon et al. (1999), the bottom size decile is split into five quintiles (based upon Amex-Nasdaq-NYSE rankings), resulting in 14 size groupings, and five book-to-market quintiles are used. The motivation for the finer partitioning of small stocks is that bid–ask bounce results in substantial effects for the tiniest firms, many of which are Amex (especially in the pre-1973 portfolios) stocks with bid prices below $3 and wide percentage bid–ask spreads.

4. The new issues puzzle

We have raised a number of issues with regard to the power of three-factor time-series regressions to identify abnormal returns. In this section, we evaluate the importance of these criticisms in the context of the new issues puzzle: the tendency for firms issuing stock to subsequently have very low returns. We could use another example to illustrate the arguments, such as the long-run return anomalies that have been documented with regard to share repurchases (Ikenberry et al., 1995) or stock-financed acquisitions (Loughran and Vijh, 1997).

Ritter (1991), Loughran and Ritter (1995), and Spiess and Affleck-Graves (1995) all report that firms issuing equity have very low returns in the three to five years after issuing. Controlling for size and either book-to-market or industry, and sometimes for momentum as well, Spiess and Affleck-Graves (1995), Mitchell and Stafford (1999), Jegadeesh (1999), and Brav et al. (1999) all report buy-and-hold abnormal returns on firms issuing seasoned equity offerings (SEOs). While the sample selection procedures vary somewhat, all report that firms issuing SEOs underperform by about 4% per year in the three to five years after issuing. Loughran and Ritter report both average annual returns, weighting each firm equally, and Fama-French three-factor regressions, weighting each time period equally. Loughran and Ritter find that in their three-factor regressions the magnitude of abnormal performance is less than when each firm is weighted equally, and attribute the difference to the tendency for there to be worse performance following high-volume periods. Brav and Gompers (1997), Brav, Geczy, and Gompers, and Mitchell and Stafford all report value-weighted three-factor regression results as well, generally finding slightly less negative intercepts than those reported by Loughran and Ritter.

4.1. The effects of contamination

We have argued that the contamination of the factors with new issues biases the three-factor model towards finding zero abnormal returns. When recent issuers have poor returns, the standard SMB (small minus big) factor has a low return and the HML (value minus growth) factor has a high return. So, to some degree, low returns on equity issuers are being used to explain low returns on equity issuers. Obviously, the intercept is biased towards finding no abnormal returns. And this bias will be worse in those periods when there has been a lot of recent equity issuance, which is exactly when Loughran and Ritter (1995, Tables I, II, and VIII) find underperformance to be most severe.

In Table 4, we show the quantitative magnitude of this bias using monthly returns on a portfolio of new issues from January 1973 to December 1996.

Table 4
Time-series regressions of EW and VW monthly percentage returns on a portfolio of new issues with market, size, and book-to-market as explanatory variables

The sample period is January 1973 to December 1996 (288 months). Operating firms that publicly issued equity (either initial or seasoned offerings (SEOs)) for cash are considered new issues for the 36 months following the month of the offering. No Compustat screen is implemented on the data set. New issues beginning with offerings from 1970 are from the Loughran and Ritter (1995) dataset and Securities Data Co. No firms with SIC code of 491–494 (utilities) are allowed to be SEOs. The maximum number of firms in the new issues portfolio occurs in November 1996 (2464 firms). The minimum number of firms occurs in March of 1976 (144). The first row of each panel uses equally weighted (EW) returns, and the second row uses value-weighted (VW) returns. Purged HML (value minus growth) and SMB (small minus big) factors have been constructed after purging from the universe of firms all stocks that publicly issued equity for cash during the prior five years. All t-statistics are calculated using White's (1980) method.

$$r_{pt} - r_{ft} = a + b(r_{mt} - r_{ft}) + s\mathrm{SMB}_t + h\mathrm{HML}_t + e_{pt}.$$

Item	a	b	s	h	Adjusted R^2
Panel A: The Fama–French (1993) method					
(1) Equally weighted	− 0.40	1.08	1.22	− 0.08	0.94
	(− 4.10)	(32.93)	(25.79)	(− 1.44)	
(2) Value-weighted	− 0.30	1.00	0.25	− 0.26	0.91
	(− 3.00)	(34.96)	(5.87	(− 5.39)	
Panel B: HML is purged of new issues					
(1) Equally weighted	− 0.44	1.10	1.22	− 0.03	0.94
	(− 4.44)	(34.97)	(27.12)	(− 0.43)	
(2) Value-weighted	− 0.32	1.03	0.27	− 0.28	0.91
	(− 3.18)	(37.42)	(6.04)	(− 5.72)	
Panel C: SMB is purged of new issues					
(1) Equally weighted	− 0.49	1.12	1.18	− 0.15	0.92
	(− 4.22)	(28.61)	(19.81)	(− 2.16)	
(2) Value-weighted	− 0.33	1.01	0.29	− 0.28	0.92
	(− 3.29)	(34.94)	(6.56)	(− 5.92)	
Panel D: HML and SMB are purged of new issues					
(1) Equally weighted	− 0.56	1.16	1.18	− 0.04	0.92
	(− 4.70)	(31.72)	(19.93)	(− 0.47)	
(2) Value-weighted	− 0.35	1.04	0.29	− 0.29	0.91
	(− 3.52)	(37.47)	(6.40)	(− 5.89)	

Inspection of row 1 of Panel A shows that the equally weighted (EW) portfolio of new issues underperforms by 40 basis points per month. On a value-weighted (VW) basis, row 2 shows underperformance of 30 basis points per month. All t-statistics have been adjusted for heteroskedasticity using White's (1980) method.[6] In Panel B, we run similar regressions, but only after changing the construction of the HML factor by deleting all firms that have publicly issued equity for cash during the prior five years. Although we define our dependent variable portfolio to be firms issuing equity during the prior 36 months, we decontaminate the factor portfolios by removing firms that have issued during the last 60 months. The reason for using this longer period for the screen is that the new issues puzzle literature shows continued underperformance beyond 36 months.

The Panel B HML factor is value firms minus growth firms, with issuing firms removed. Consistent with the contamination hypothesis, the intercepts reported in Panel B are more negative than in the Panel A regressions. The effects for both the EW and VW regressions, however, are quite small.

In Panel C of Table 4, we report three-factor regression results after changing the construction of the SMB factor to remove issuing firms from the factors. The small growth firm portfolio is much more intensive in new issues than is the large growth firm portfolio, so previously the SMB factor was partially measuring issuers minus nonissuers. As shown by Jegadeesh (1992), low correlations at the individual firm level can have substantial effects on statistical inference when portfolio returns are used. Once again, consistent with the contamination hypothesis, in Table 4 the Panel C intercepts are more negative than the Panel A intercepts.

Lastly, Panel D reports the three-factor regression results in which we have eliminated the contamination of both the HML and SMB factors. In constructing the decontaminated factors, we remove from the universe of eligible firms (as described by Fama and French and in our Section 2) all stocks that have been CRSP-listed for less than 60 months or have publicly issued equity for cash during the prior 60 months. Independent sorts on size and book-to-market are then conducted every June 30 using NYSE breakpoints, and value-weighted style portfolio returns are constructed. Note that we do not remove firms that issue equity in the future, for this is unknowable at the time of portfolio construction. As expected, the intercepts are even more negative, with the new issues portfolio underperforming by 56 basis points per month on an equally weighted basis and 35 basis points per month on a value-weighted basis. These intercepts are 17–40% larger than those reported in Panel A.

In Table 5, we test the supply response hypothesis by segmenting the 1973–1996 period into the calendar months in which new issues are a low percentage of the market and the months in which they are a high percentage of the market. Each month, we calculate the percentage of CRSP-listed domestic operating companies that have issued equity during the prior 36 months. The

[6] An alternative approach to adjusting for heteroskedastity is contained in Jaffe (1974) and Mandelker (1974), where monthly parameter estimates are divided by their cross-sectional standard deviations to get a standardized variable.

Table 5
Time-series regressions of EW and VW monthly percentage returns on a portfolio of new issues with market, size, and book-to-market as explanatory variables

The sample period is January 1973 to December 1996 (288 months). Operating firms that issued equity (either initial or seasoned offerings) for cash are considered new issues for the 36 months following the month of the offering. The first row of each panel uses equally weighted (EW) returns, and the second row uses value-weighted (VW) returns. Purged (decontaminated) HML and SMB factors have been constructed after purging from the universe of firms all stocks that publicly issued equity for cash during the prior five years. The 144 months when the ratio of recent new issues to the total number of domestic operating firms on CRSP is highest are defined as high-volume markets. The other 144 months are designated as low-volume markets. All t-statistics are calculated using White's (1980) method.

$$r_{pt} - r_{ft} = a + b(r_{mt} - r_{ft}) + s\mathrm{SMB}_t + h\mathrm{HML}_t + e_{pt}.$$

Item	a	b	s	h	Adjusted R^2
Panel A: Low-volume markets, regular factors					
(1) Equally weighted	– 0.29	1.07	1.27	– 0.13	0.94
	(– 1.99)	(24.82)	(19.17)	(– 1.62)	
(2) Value-weighted	– 0.34	0.95	0.16	– 0.21	0.88
	(– 2.07)	(24.89)	(2.34	(– 3.35)	
Panel B: Low-volume markets, purged factors					
(1) Equally weighted	– 0.34	1.16	1.18	– 0.04	0.92
	(– 1.87)	(23.06)	(13.48)	(– 0.40)	
(2) Value-weighted	– 0.34	9.97	0.21	– 0.30	0.89
	(– 2.18)	(27.55)	(3.20)	(– 4.97)	
Panel C: High-volume markets, regular factors					
(1) Equally weighted	– 0.55	1.10	1.18	– 0.05	0.95
	(– 4.19)	(20.17)	(19.81)	(– 0.67)	
(2) Value-weighted	– 0.27	1.12	0.38	– 0.21	0.96
	(– 2.85)	(31.03)	(6.73)	(– 3.48)	
Panel D: High-volume markets, purged factors					
(1) Equally weighted	– 0.78	1.16	1.15	– 0.02	0.92
	(– 5.12)	(21.13)	(17.85)	(– 0.22)	
(2) Value-weighted	– 0.39	1.16	0.42	– 0.16	0.95
	(– 3.96)	(30.52)	(5.96)	(– 2.68)	

'low-volume' months are those below the median ratio, and the 'high-volume' months are those above the median. The supply response hypothesis predicts that there should be more severe underperformance in the high-volume periods.

Table 5 reports three-factor regression results for the low-volume periods with the standard Fama-French factors (Panel A) and the purged factors (Panel B).

Whether EW or VW regressions are used, the intercepts are − 34 basis points per month when purged factors are used. In Panels C and D of Table 5, we report the three-factor regression results for the high-volume periods. On both an EW and a VW basis, the intercepts are economically and statistically significant, with the point estimates larger in absolute value than for the low-volume periods. This is exactly what the supply response would predict. In the high-volume periods, the purged factors produce a noticeable change in the intercepts relative to the standard Fama-French factors. This is as predicted by the factor contamination hypothesis, for it is in the high-volume periods that the factors are most contaminated with the issuing firms.

In Tables 6 and 7, respectively, we decompose the new issues portfolio into an initial public offering (IPO) portfolio and an SEO portfolio, and report the results of three-factor time-series regressions with and without any decontamination of the factors for both low- and high-volume periods. In Panel A of Table 6, using the standard methodology, IPOs underperform by 24 basis points per month on an EW basis, but by only seven basis points per month on a VW basis, with the latter number being unreliably different from zero. This is consistent with the evidence in Ritter (1991) and Brav and Gompers (1997) that small IPOs underperform by more than large IPOs. In Panel B, the factors are constructed to eliminate the contamination. Now, IPOs underperform by 42 basis points per month (EW) and 26 basis points per month (VW). This is consistent with the evidence that the small growth stock portfolio is frequently intensive in recent IPOs, so using it to measure IPO performance without removing new issue firms biases the Panel A regressions towards a zero intercept.

Panels C and D of Table 6 report the results for the low-volume and high-volume periods. In low-volume periods, IPOs do not appreciably underperform, whereas in high-volume periods, IPOs severely underperfrom on both an EW (− 90 basis points per month) and a VW (− 49 basis points per month) basis. This is consistent with the supply response hypothesis, and with the results and interpretation in Ritter (1991) and Loughran and Ritter (1995).

In Table 7, we report three-factor regression results for SEOs, with Panel A using the traditional Fama-French factors and Panel B using the decontaminated factors. As with IPOs, the decontaminated factor regressions show more underperformance. In Panel C of Table 7, we report the results for low-volume periods, and in Panel D for high-volume periods. As with IPOs, there is more severe performance for SEOs in the high-volume periods, consistent with the supply response hypothesis.

4.2. Is the underperformance of new issues merely a manifestation of a misspecified model?

Brav and Gompers (1997) argue that IPOs have low returns, but that this is just a manifestation of the fact that many IPOs are small growth firms, which

Table 6
Time-series regressions of EW and VW monthly percentage returns of IPO portfolios on market, size, and book-to-market return realizations

The sample period is January 1973 to December 1996 (288 months). Operating firms that issued equity for cash are considered initial public offerings (IPOs) for the 36 months following the month of the offering. The first row of each panel uses EW returns, the second row uses VW returns. Low-volume months are defined as those months in the bottom half of the ratio of issuing firms to total domestic operating firms on CRSP. Purged (decontaminated) HML and SMB factors have been constructed after purging from the universe of firms all stocks that publicly issued equity for cash during the prior five years. All *t*-statistics are calculated using White's (1980) method. The *t*-statistics for the difference in intercepts between Panels C and D are 3.39 for EW portfolios and 1.34 for VW portfolios, using a difference in means with unequal variances test.

$$r_{pt} - r_{ft} = a + b(r_{mt} - r_{ft}) + s\mathrm{SMB}_t + h\mathrm{HML}_t + e_{pt}.$$

Item	a	b	s	h	Adjusted R^2
Panel A: The Fama–French (1993) method, IPOs					
(1) Equally weighted	− 0.24	(1.03	1.39	− 0.10	0.90
	(− 1.73)	(27.94)	(24.13)	(− 1.57)	
(2) Value-weighted	− 0.07	1.08	0.93	− 0.55	0.88
	(− 0.52)	(26.95)	(15.38	(− 8.39)	
Panel B: HML and SMB are purged of new issues, IPOs					
(1) Equally weighted	− 0.42	1.12	1.32	− 0.04	0.87
	(− 2.64)	(26.46)	(18.10)	(− 0.44)	
(2) Value-weighted	− 0.26	1.20	0.90	− 0.46	0.85
	(− 1.62)	(26.74)	(12.80)	(− 5.09)	
Panel C: Low-volume markets, purged factors, IPOs					
(1) EW-low volume	− 0.00	1.11	1.38	− 0.08	0.87
	(− 0.01)	(18.33)	(13.89)	(− 0.66)	
(2) VW-low volume	− 0.06	1.19	0.88	− 0.51	0.82
	(− 0.22)	(18.25)	(8.73)	(− 3.69)	
Panel D: High-volume markets, purged factors, IPOs					
(1) EW-high volume	− 0.90	1.14	1.13	− 0.04	0.87
	(− 5.16)	(19.15)	(12.70)	(− 0.35)	
(2) VW-high volume	− 0.49	1.24	0.95	− 0.35	0.91
	(− 2.92)	(19.32)	(10.73)	(− 3.72)	

have low returns in general. In other words, once one knows that a firm has a small size and a low book-to-market ratio, there is no incremental information associated with whether it is a recent IPO or not. Evidence suggesting that this is potentially a plausible explanation for the low returns on IPOs is contained in Fama and French (1993, Table 9a, Panel (iv)) and Mitchell and Stafford (1999,

Table 7
Time-series regressions of EW and VW monthly percentage returns of SEO portfolios on market, size, and book-to-market return realizations

The sample period is January 1973 to December 1996 (288 months). Publicly traded firms issuing equity in general cash offers during the prior 36 months are considered to be seasoned equity offering (SEO) firms. The first row of each panel uses EW returns, and the second row uses VW returns. Low-volume months are defined as those months in the bottom half of the ratio of issuing firms to total domestic operating firms on CRSP. Purged (decontaminated) HML and SMB factors have been constructed after purging from the universe of firms all stocks that publicly issued equity for cash during the prior five years. All t-statistics are calculated using White's (1980) method. The t-statistics for the difference in intercepts between Panels C and D are 1.00 for EW portfolios and 0.31 for VW portfolios, using a difference in means with unequal variances test.

$$r_{pt} - r_{ft} = a + b(r_{mt} - r_{ft}) + s\mathrm{SMB}_t + h\mathrm{HML}_t + e_{pt}.$$

Item	a	b	s	h	Adjusted R^2
Panel A: The Fama–French (1993) method, SEOs					
(1) Equally weighted	− 0.47	1.15	1.10	− 0.12	0.96
	(− 5.42)	(32.64)	(23.07)	(− 2.04)	
(2) Value-weighted	− 0.32	1.01	0.17	− 0.27	0.90
	(− 3.00)	(34.06)	(3.89	(− 5.54)	
Panel B: HML and SMB are purged of new issues, SEOs					
(1) Equally weighted	− 0.61	1.23	1.07	− 0.08	0.94
	(− 6.08)	(33.81)	(19.88)	(− 0.11)	
(2) Value-weighted	− 0.35	1.03	0.22	− 0.27	0.91
	(− 3.38)	(36.45)	(4.67)	(− 5.54)	
Panel C: Low-volume markets, purged factors, SEOs					
(1) EW-low volume	− 0.49	1.26	0.97	− 0.00	0.95
	(− 3.68)	(26.10)	(13.46)	(− 0.02)	
(2) VW-low volume	− 0.34	0.97	0.13	− 0.30	0.89
	(− 2.08)	(26.38)	(1.90)	(− 4.78)	
Panel D: High-volume markets, purged factors, SEOs					
(1) EW-high volume	− 0.69	1.20	1.14	− 0.11	0.94
	(− 4.63)	(20.51)	(18.34)	(− 1.57)	
(2) VW-high volume	− 0.40	1.14	0.33	− 0.16	0.94
	(− 3.68)	(31.24)	(4.86)	(− 2.74)	

Table 6), where it is reported that small growth firms have negative intercepts in three-factor regressions.

In Table 8, we directly address the possibility that the new issues puzzle is merely due to a misspecified three-factor model. In testing this hypothesis, we are constrained by the fact that, especially in the 1970s, there are many years

Table 8
Intercepts and t-statistics from monthly three-factor regressions using decontaminated factors for six subportfolios (formed on the basis of size and book-to-market) of new issues, nonissuers, and issuers minus nonissuers

The sample period is July 1973 to December 1996 (282 months) except for Panels A and C where the small value portfolio uses 261 months and the large value portfolio uses 222 months. Months with less than five firms in a portfolio are not included for the regressions. Operating firms that issued equity (either initial or seasoned offerings) for cash are considered new issues for the 36 months following the month of the offering. Only firms used in Tables 1 and 2 are included in the universe: all firms must be CRSP-listed and have two prior years of Compustat book value available. Furthermore, nonissuers must be listed on CRSP for 60 months before entering the sample. The left three columns of each panel use equally weighted (EW) returns, and the right three columns use value-weighted (VW) returns. All t-statistics are calculated using White's (1980) method. All panels use decontaminated factors where HML and SMB have been constructed after purging from the universe of firms all stocks that publicly issued equity for cash during the prior five years. In Panel C the dependent variable is $r_{\text{issuers,t}} - r_{\text{nonissuers,t}}$.

$$r_{pt} - r_{ft} = a + b(r_{mt} - r_{ft}) + s\text{SMB}_t + h\text{HML}_t + e_{pt}$$

Three-factor cell	EW			VW		
	Growth	Middle	Value	Growth	Middle	Value
Panel A: New issues intercepts and t-statistics						
Small firms	− 0.77	− 0.34	− 0.17	− 0.39	− 0.21	− 0.33
	(− 5.59)	(− 2.64)	(− 0.82)	(− 2.91	(− 1.73)	(− 1.93)
Large firms	− 0.37	− 0.30	− 0.35	− 0.20	− 0.22	− 1.41
	(− 3.15)	(− 2.19)	(− 1.64)	(− 1.50)	(− 1.30)	(− 1.95)
Panel B: Nonissuers intercepts and t-statistics						
Small firms	− 0.19	0.10	0.19	− 0.08	0.06	0.10
	(− 1.73)	(1.51)	(2.02)	(− 1.29)	(1.28)	(2.61)
Large firms	0.10	1.05	0.05	0.11	0.04	− 0.03
	(1.76)	(0.74)	(0.83)	(2.30)	(0.64)	(− 0.61)
Panel C: New issue firms minus nonissuers intercepts and t-statistics						
Small firms	− 0.58	− 0.44	− 0.41	− 0.30	− 0.27	− 0.44
	(− 5.22)	(− 3.85)	(− 2.09)	(− 2.56)	(− 2.12)	(− 2.51)
Large firms	− 0.46	− 0.34	− 0.40	− 0.31	− 0.26	− 0.35
	(− 4.13)	(− 2.87)	(− 1.86)	(− 2.28)	(− 1.49)	(− 1.60)

with relatively few new issues, so that too fine a partitioning of the data results in new issues portfolios with zero or one firm in many months. Consequently, we do not separate IPOs from SEOs, which would exacerbate this problem. Furthermore, we use the six Fama-French 'style' subportfolios, rather than the 25 portfolios that would result from a 5×5 sort of size and book-to-market.

Even combining IPOs and SEOs and restricting our analysis to six subportfolios, we still have many months in which the large value portfolio has fewer than five new issues in it.

In Table 8, Panel A, we report the intercepts and associated *t*-statistics from three-factor regressions using decontaminated factors for portfolios of new issues on both an EW and a VW basis. We do not report the slope coefficients (the factor loadings) to avoid overburdening the reader with too many numbers. All 12 intercepts are negative, with five of them having *t*-statistics greater than two in absolute value. In Panel B, we report the intercepts and *t*-statistics for nonissuers. Consistent with the findings of other researchers, using both EW and VW returns, the small growth stock portfolio has a negative intercept, and the small value stock portfolio has a positive intercept. In Panel C, we report regression results in which the dependent variable is the return on the new issues portfolio minus the return on the nonissuers portfolio. Here, all 12 intercepts are negative, with nine of the 12 having *t*-statistics greater than two in absolute value.

For SEOs, this is not too surprising, given that Spiess and Affleck-Graves (1995), Jegadeesh (1999), Brav et al. (1999), Eckbo et al. (1999), and Mitchell and Stafford (1999) all report underperformance for U.S. SEOs using buy-and-hold returns adjusting for size and book-to-market effects. And this is in spite of the fact that both Brav, Geczy, and Gompers and Mitchell and Stafford include utility stocks in their SEO samples. For IPOs, Brav and Gompers report no reliable underperformance for IPOs once size and book-to-market are controlled for.[7]

5. Summary and conclusions

This paper has several purposes. First, we argue that the Fama-French three-factor model using value-weighted returns tends to underestimate abnormal returns when the event being studied is a managerial choice variable. We do this with logical arguments, simulations, and sensitivity tests. In our simulations, misvaluations are most severe among small firms in high-volume periods. Value-weighted three-factor time-series regressions pick up only about half of the abnormal returns that are present when each firm is weighted equally.

[7] In their buy-and-hold returns, Brav and Gompers (1997) and Brav et al. (1999) do not impose a five-year CRSP-listing screen for nonissuers, unlike Loughran and Ritter (1995) and this paper. This seemingly minor departure from the Loughran and Ritter methodology changes the buy-and-hold abnormal returns by about 1% per year. CRSP lists about 200–300 stocks per year that have very low returns, and are not included in the standard IPO databases of firm commitment offerings. Many of these new listings are small bank IPOs.

The second purpose of the article is to argue for a minimalist interpretation of multifactor time-series regression results. These are not tests of market efficiency, but just tests of whether a given pattern in returns is independent of other previously documented patterns in returns. One cannot test market efficiency unless one uses a normative model of what 'normal' returns should look like. Thus, empirically based asset pricing models such as the Fama-French three-factor model (or a four-factor model with a momentum factor) have no power to test market efficiency. This logic applies to tests using arbitrage pricing theory as well, since the factors that are used are always empirically based. Event studies examining long-run returns must distinguish between two types of tests: (1) tests of the joint hypothesis of market efficiency and a normative model, and (2) tests of whether any patterns that are found are distinct from other previously documented empirical patterns.

Third, we argue that the lack of robustness of the magnitude of abnormal returns to alternative methodologies is not evidence in favor of market efficiency. Instead, we argue that because some methodologies have more power than others, if there are true abnormal returns, there *should* be predictable differences in abnormal return estimates across different methodologies.

Time-series regressions do have an advantage over traditional event-study procedures in that the regressions can explicitly control for temporal dependence in returns. However, the common practice of running time-series regressions with each time period weighted equally has consequences above and beyond convenience for calculating statistical significance. If firms are voluntarily taking actions in response to misvaluations, there is almost certain to be time clustering of observations. We would expect the greatest misvaluations (and largest subsequent abnormal returns) to be present in periods when a lot of events take place. But if all periods are equally weighted, this supply response will be missed. Furthermore, because it is plausible that small stocks are more misvalued than big stocks, value-weighting portfolio returns will tend to reduce abnormal returns.

Fourth, using the new issues puzzle, we show that the usual implementation of the three-factor model is biased towards finding zero abnormal returns because of the use of factors that are partly composed of the firms being tested. We propose an alternative method of implementing a multifactor time-series regression. If there is reason to believe that one or more of the style portfolios composing the factors is relatively intensive in the firms being analyzed, a reconstruction of the factors with sample firms omitted can eliminate the contamination. Furthermore, segmenting the sample period into periods when there are relatively few or many events will permit a time-series regression to capture the supply response when the event under study is a managerial choice variable.

A fifth purpose of this article is to reexamine the new issues puzzle. When we run three-factor regressions using factors that have been purged of new issues, we find that new issues reliably underperform on both an equally weighted and a value-weighted basis. For all months combined, our point estimate of underperformance for the value-weighted portfolio is 35 basis points per month, or over 4% per year. This underperformance is more severe in high-volume periods than in low-volume periods. Once size and book-to-market effects are controlled for, IPOs appear to underperform only in high-volume periods. Firms conducting SEOs, however, reliably underperform in both low- and high-volume markets. Consistent with the windows of opportunity framework, the underperformance of SEOs is more severe following high-volume markets.

References

Banz, R., Breen, W., 1986. Sample dependent results using accounting and market data: some evidence. Journal of Finance 41, 779–793.

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

Brav, A., 1998. Inference in long-horizon event studies: A Bayesian approach with application to initial public offerings. Unpublished working paper. Duke University, Durham, NC.

Brav, A., Geczy, C., Gompers, P., 1999. Is the abnormal return following equity issuance anomalous? Journal of Financial Economics, forthcoming.

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

Campbell, J., Lo, A., MacKinlay, C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

Eckbo, B. E., Masulis, R. W., Norli, O., 1999. Seasoned public offerings: resolution of the 'new issues puzzle'. Journal of Financial Economics, forthcoming.

Fama, E., 1996. Multifactor portfolio efficiency and multifactor asset pricing. Journal of Financial and Quantitative Analysis 31, 441–465.

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

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

Hasbrouck, J., 1991. The summary informativeness of stock trades: an econometric analysis. Review of Financial Studies 4, 571–595.

Ikenberry, D., Lakonishok, J., Vermaelen, T., 1995. Market underreaction to open market share repurchases. Journal of Financial Economics 39, 181–208.

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

Jegadeesh, N., 1992. Does market risk really explain the size effect? Journal of Financial and Quantitative Analysis 27, 337–351.

Jegadeesh, N., 1999. Long-run performance of seasoned equity offerings: benchmark errors and biases in expectations. Unpublished working paper. University of Illinois, Urbana, IL.

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

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

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

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

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

Mitchell, M., Stafford, E., 1999. Managerial decisions and long-term stock price performance. Unpublished working paper. Harvard University, Cambridge, MA.

Ritter, J., 1991. The long-run performance of initial public offerings. Journal of Finance 46, 3–27.

Shleifer, A., Vishny, R., 1990. Equilibrium short horizons of investors and firms. American Economic Review 80, 148–153.

Shleifer, A., Vishny, R., 1997. The limits of arbitrage. Journal of Finance 52, 35–55.

Spiess, K., Affleck-Graves, J., 1995. Underperformance in long-run stock returns following seasoned equity offerings. Journal of Financial Economics 38, 243–268.

Stein, J., 1996. Rational capital budgeting in an irrational world. Journal of Business 69, 429–455.

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