Industry Indices in Event Studies

Joseph M. Marks

Bentley University, AAC 273
175 Forest Street
Waltham, MA 02452-4705
jmarks@bentley.edu

Jim Musumeci*

Bentley University, 107 Morrison
175 Forest Street
Waltham, MA 02452-4705
jmusumeci@bentley.edu

Aimee Hoffmann Smith

Bentley University, AAC 221
175 Forest Street
Waltham, MA 02452-4705
asmith@bentley.edu

Current Draft: January, 2018

Industry Indices in Event Studies
Abstract

Event studies compare a sample of stock returns relative to their benchmark returns at the time of an event and test whether deviations from the benchmarks are significantly different from zero. There are two desirable characteristics of these benchmarks: (1) that they be unbiased, i.e., absent an event, the average deviation from the benchmark is zero, and (2) that the prediction error has as small a variance as possible.

King (1966) found that a firm's industry explains about 10% of its variance of returns. Despite this, event-study benchmarks typically ignore this industry effect. We consider several common factor models and examine the results when an industry factor is used to replace or supplement a market factor. We find that inclusion of an industry factor increases event study test power by about 10%.

EFM classification: 760, 310, 350, 380

Industry Indices in Event Studies

In general, there are two desirable requirements for a benchmark to be used in an event study. First, it should be unbiased; biasedness will necessarily produce Type I error that increases in sample size. Second, it should have as low a variance of prediction error as possible. Ceteris paribus, a lower prediction-error variance will improve test power.

Consider a perhaps not-too-distant future in which worldwide stock markets trade continuously around the clock, and global stock and national stock market indices are readily available. If you were examining an event's effect on a sample of U.S. stocks only, you could use a global market index to find benchmark returns. However, while international events affect the entire world's economy, certainly some countries will be affected more than others. For example, while political upheaval in, say, China, would necessarily have implications for worldwide markets, we would expect it to affect Asian stocks substantially more than U.S. stocks. Similarly, while any disruption of NAFTA would have global implications, we would expect it to have a larger impact on U.S., Canadian, and Mexican stocks than on South African stocks. Thus, if an entire sample of stocks is from the same country, we would expect a global index to overweight events that have minor implications for that country's stocks and underweight events that have major implications. For this reason a U.S. market index seems a better choice than a global index for a sample composed of U.S. companies. In general, a market index that includes the stocks of interest but does not include too much extraneous noise seems a wiser choice than an index that is too broad.[1]

King (1966) found that the market accounts for about half the total variance of a U.S. stock, and that the stock's industry explains about an additional 10%. This suggests use of an industry factor is likely to produce residuals with a smaller variance and thus lead to more powerful tests. Consider, for example, an increase in the value of the dollar. Ceteris paribus, we would expect this to have a positive effect on industries that are net importers and a negative effect on industries that are net exporters. A national index would necessarily reflect only the *average* effect on all firms and would be less informative than an industry index that would capture the net exporter vs. net importer effect.

If all the stocks in an event-study sample were from the same industry, then the same industry index could be used for each stock and there would be no problem with inconsistency. However, it might seem that some comparability problems could occur if we used different indices for different stocks in the same sample. For example, if two stocks in different industries had the same total variance of returns, but one industry contained fairly homogenous firms while the other did not, we would expect the estimation-period residuals

---

[1] However, an index cannot be too narrow, either. For example, an equally weighted index that includes *only* the stocks in the sample would necessarily produce an average residual of zero on the event day or any other day.

and event-day abnormal return for the stock in the homogenous industry to have a smaller variance than those for the stock in the heterogeneous industry. Because the main two event-study methods, Patell (1976) and Boehmer et al. (1991), normalize the event-period abnormal returns by the standard deviation of the estimation-period residuals, we expect the difference in the variances of the raw (unnormalized) abnormal returns to be inconsequial because the point of normalization is to create standardized abnormal returns that have approximately equal variances.

In the following section, we discuss use of equally weighted vs. value-weighted indices in event studies. Next we describe how we use (1) common methods, (2) an industry index replacing the market index, and (3) an industry index supplementing a market index to form benchmarks. We then proceed to describe how we simulate events and abnormal returns. Finally, we compare the specification and power of tests using the various benchmarks. Of particular interest will be comparisons of commonly used methods with benchmarks that either replace or supplement a market index with an industry index. The final section concludes.

I.   Equally Weighted vs. Value-Weighted Indices

For many years the market model using the CRSP equally weighted index was the main event-study benchmark, and indeed it is still the default in Eventus. Although the CRSP value-weighted index is occasionally used, it can lead to an interesting paradox. Consider an economy with 9 small firms, each having 5% of the total market cap, and one large firm constituting the remaining 55%. Without loss of generality, suppose also that each of the 10 companies has a $\hat{\beta}$ of 1 and an $\hat{\alpha}$ of 0. Suppose also that on some event date the large firm went up by 0.9%, while the nine small firms went down by 1.1% each. The value-weighted index didn't change because .45(-.011) + .55(.009) = 0, while an equally weighted index would produce a market return of $\frac{9(-.011)+1(.009)}{10}$ = -0.9%.

An event study testing whether the abnormal return of some "sample" of stocks (which, unbeknownst to you, happened to be all the stocks in this market) will produce an odd result. If you weight the residuals equally but use the value-weighted index, you find that nine stocks underperformed the value-weighted index by 1.1% and one outperformed it by .9%, for an average abnormal return of $\frac{9(-.011)+1(.009)}{10}$ = -.9%. The conclusion is that your sample, i.e., the market, had a negative abnormal performance, i.e., it performed worse than itself. This problem goes away if you use the equally weighted index, because then the nine stocks underperformed that market return by 0.2% each, while the single large firm outperformed it by 1.8%, so average AR = $\frac{9(-.002)+1(.018)}{10}$ = 0. Similarly, the value-weighted index avoids the paradox if you value-weight the residuals (but no one does that). Brown and Warner (1980)

also make this observation on pp. 241-242, and they go on to find use of the value-weighted index produces less powerful tests than when an equally weighted index is used.

This paradox notwithstanding, there is a powerful intuitive motivation for using a value-weighted index. Specifically, if your sample consists primarily of large stocks, then those stocks are likely to have a higher correlation with a value-weighted index than with an equally weighted one. Accordingly, a single-factor model using the value-weighted index may have lower prediction error and so be preferable. Another possibility is to use a benchmark with only large stocks (to keep the high correlation), but to use an equally weighted index of these stocks to avoid the paradox discussed above. To the best of our knowledge, no one has tried this.

More recently, some have used the Fama-French Three-Factor Model or the Carhart Four-Factor Model to find benchmark returns, and these models use a value-weighted market index. They also produce the interesting result observed by Cremers et al. (2012) that common market benchmarks have alphas that are consistently non-zero. The reason for this unusual result is related to the paradox just discussed (except that in our example the residuals are equally weighted and so an equally weighted index avoids the problem, while Cremers et al. examine value-weighted benchmarks, and find their paradox is mitigated when, consistent with their benchmarks examined and with the Fama-French market factor, they value-weight the HML factor as well).

Because we are comparing use of industry indices with commonly used methods, some of which include an equally weighted market index (e.g., the single-factor CRSP equally weighted index that is the default in Eventus), and some of which use a value-weighted market index (e.g., the Fama-French and Carhart models), we examine both equally weighted and value-weighted industry indices. These indices were found by calculating the daily return on an appropriate portfolio of all stocks in the same industry as per the Fama and French (1997) 48 industry classifications.

II. Various Plausible Benchmarks

In the following sections, we compare simulated event-study results using the following 16 possible benchmarks.
  I.   A single-factor model using an equally weighted industry index
  II. A single-factor model using the CRSP equally weighted index (the default in Eventus)
  III. A two-factor model using the CRSP and industry equally weighted indices
  IV. A single-factor model using a value-weighted industry index
  V.   A single-factor model using the CRSP value-weighted index
  VI. A two-factor model using the CRSP and industry value-weighted indices
  VII. The Fama-French Three-Factor Model

VIII.   A Three-Factor Model with an equally weighted industry index as well as the Fama-French SMB and HML factors (i.e., the Fama-French Three-Factor Model with an equally weighted industry index replacing the market index)

  IX.   A Three-Factor Model with a value-weighted industry index as well as the Fama-French SMB and HML factors (i.e., the Fama-French Three-Factor Model with a value-weighted industry index replacing the market index)

   X.   The Fama-French Three-Factor Model with a fourth factor equal to the industry equally weighted index

  XI.   The Fama-French Three-Factor Model with a fourth factor equal to the industry value-weighted index

 XII.   The Carhart Four-Factor Model

XIII.   A Four-Factor Model with an equally weighted industry index as well as the Fama-French SMB and HML factors and the Carhart momentum factor (i.e., the Carhart Four-Factor Model with an equally weighted industry index replacing the market index)

XIV.   A Four-Factor Model with a value-weighted industry index as well as the Fama-French SMB and HML factors and the Carhart momentum factor (i.e., the Carhart Four-Factor Model with a value-weighted industry index replacing the market index)

 XV.   The Carhart Four-Factor Model with a fifth factor equal to the industry equally weighted index

XVI.   The Carhart Four-Factor Model with a fifth factor equal to the industry value-weighted index

Consistent with common practice, in the first six (single- and two-factor) specifications, the market and industry indices are raw values, i.e., not in risk-premium form. In the remaining ten specifications, all market and industry indices are in risk-premium form, i.e., $R_{market} - R_F$ or $R_{industry\ index} - R_F$.

While it is possible to simply look at the results of all 16 possible benchmarks and choose the one that is well specified and most powerful, we believe the better way to analyze the results of the following sections is to consider a model that is currently in use, and see if that model produces better tests when the market index is replaced by or supplemented with an industry index. For example, a researcher who plans to use the default in Eventus (benchmark II, the CRSP equally weighted index) will find a comparison with the equally weighted industry index (benchmark I) or the CRSP and industry equally weighted indices (benchmark III) most useful.

We note that while adding an independent variable will necessarily produce a higher R^2, or equivalently a lower variance of the estimation-period residuals, event-studies generally make out-of-sample predictions. The adjustment for out-of-sample prediction variance is given for a single-factor model in Patell (1976) to be

$$C = 1 + \frac{1}{T} + \frac{(R_{M,E} - \bar{R}_M)^2}{\sum_t (R_{M,t} - \bar{R}_M)^2} \qquad [1]$$

for bivariate models. In the multivariate case, the analogous adjustment is given in Kmenta (1971, p. 375) as

$$C = 1 + \frac{1}{T} + (\underline{X}_0 - \overline{\underline{X}})'(\underline{X}'\underline{X})^{-1}(\underline{X}_0 - \overline{\underline{X}}) \qquad [2]$$

In this latter case, multicollinearity will increase the variance of the prediction error. Thus while adding a factor (e.g., changing from benchmark II to benchmark III) will necessarily produce a lower in-sample variance, it may well produce inferior results because the independent variables are correlated.

III. Generation of Simulated Abnormal Returns and Results

The entire population of stocks listed in the daily CRSP database were our initial candidates for simulated "events." For every observation in CRSP we estimated parameters for the 16 specifications from Section II using an estimation period of the 120 days preceding the "event," provided the stock traded for at least $5 on the event date and that there were no missing observations during the 120 days preceding the "event." This generated 37,587,725 parameter estimates and "event"-day abnormal returns. For all specifications the average event-day abnormal return was essentially zero. A summary of the results is reported in Table I. In general, adding more factors increases R^2 and decreases in-sample residual standard deviation, but also produces larger forecast-error adjustments (C) because of multicollinearity.

We then proceeded to generate 10,000 random pseudo-portfolios, each consisting of 5000 hypothetical events. For each portfolio, we constructed nested subsets with sizes of 2500, 1000, 500, 250, and 100 events. The sampling was without replacement within each portfolio so that the same firm-day event cannot appear twice within a portfolio. Additionally, the sampling was performed without imposing controls on firm or temporal distribution (i.e., the same firm or the same date may appear multiple times within a given portfolio). For each event date for each stock we then simulated an abnormal return with a mean of $\bar{\Delta}$ equal to either 0% or 0.125%, and a variance equal to θ (0 or 1) times the estimation-period variance. Thus, for example, $\bar{\Delta}$ = 0% and θ = 0 simulates no abnormal return at all (and provides evidence regarding test specification), while $\bar{\Delta}$ = 0.125% and θ = 1 simulates an event that causes share price to increase by an average of an eighth of a percent, with a variance that is equal to the stock's residual variance during the estimation period.

To ensure that we did not obtain an aberrant simulation, we repeated the pseudo-portfolio generation process two additional times. The results were consistent across all three simulations. Results when $\bar{\Delta}$ = 0.25% (not reported) were also fairly similar.

A: $\bar{\Delta}$ = 0%

Tables II and III report the results for $\bar{\Delta}$ = 0%, θ = 0 and $\bar{\Delta}$ = 0%, θ = 1. Thus, they provide evidence regarding whether the various benchmarks produce tests that are well-specified, i.e., reject a null hypothesis of no abnormal performance with a frequency equal to the purported value of α. We actually tested α = 1% and α = 5%, but for brevity we report only the results for the latter.

Consistent with Marks and Musumeci (2017), who tested only specification II (CRSP equally weighted index), we find the Patell test is misspecified and rejects H$_0$: SAR = 0 too frequently across all 16 models, even absent any event-induced variance (Table II). We find no evidence that the BMP test is misspecified for any of the 16 models.

The results are even more dramatic when the event creates an increase in variance as in Table III ($\bar{\Delta}$ = 0%, θ = 1). Here, the Patell test rejects a true null between three and four times as often as it should. This is consistent with previous research [Boehmer et al (1991), Harrington and Shrider (2007), and Marks and Musumeci (2017)], except we extend the analysis beyond only specification II and find the problem occurs for any of the 16 specifications. As was the case when θ = 0, we find no evidence that the BMP test is misspecified when θ = 1.

Because the Patell test is misspecified both in the absence or presence of event-induced variance, we consider only the BMP in our examination of test power.

B: $\bar{\Delta}$ = 0.125%

Table IV reports the results for $\bar{\Delta}$ = 0.125%, θ = 0 and $\bar{\Delta}$ = 0.125%, θ = 1. Not surprisingly, we did not find a dramatic difference in the absolute difference in rejection rates. For example, absent any event-induced variance (Panel A) and when N =500 events, the two least powerful specifications were V (CRSP value-weighted index, rejection rate = 29.48%) and VII (Fama-French Three-Factor Model, rejection rate = 29.79%). On the other extreme, the two most powerful specifications were III (CRSP equally weighted index plus an equally weighted industry index, rejection rate = 33.81%) and XI (Fama-French Three-Factor Model plus a value-weighted industry index, rejection rate = 33.41%).

While the absolute difference between the highest (33.81%) and lowest (29.48%) rejection rates does not appear large, it represents a proportional increase of $\frac{33.81\%}{29.48\%} - 1$ = 14.69% and is statistically significant at the 5% level.

The results were fairly similar in the presence of event-induced variance (Panel B). Here, the average rejection rates were unsurprisingly lower than those of Panel A (more noise invariably produces less powerful tests, and event-induced variance is essentially a type of noise), specifically, the average rejection rate across all 16 specifications for N = 500 was 32.18% when $\theta = 0$, but only 19.44% when $\theta = 1$. The worst two performers were again specification V (CRSP value-weighted index, rejection rate = 17.67%) and VII (Fama-French Three-Factor Model, rejection rate = 18.08%). The two best performers were XVI (Carhart Four-Factor Model with a value-weighted industry index, rejection rate = 20.45%) and III (CRSP equally weighted index plus an equally weighted industry index, rejection rate = 20.25%). Once again, the absolute difference in rejection rates appears small, but the proportional increase from least to most powerful is $\frac{20.45\%}{17.67\%} - 1$ = 15.73% and is statistically significant at the 5% level.

However, the purpose of this paper is to determine whether an industry index improves performance of commonly used models when it is used instead of or in addition to a market index. Thus, we find it natural not to compare all the models with each other, but to compare them within similar groups.

Our first comparison of this type is what happens when we take a commonly used model and replace or supplement the market index with an industry index. For example, the default in Eventus is a single-factor model using the CRSP equally weighted index (specification II). The two most natural comparisons involving an industry index are specification I (an equally weighted industry index is used *instead of* the market index) and specification III (an equally weighted industry index is used *in addition to* the market index). For convenience, these results are summarized in Table V. Whether $\theta = 0$ or $\theta = 1$, with the sole exception of N = 1000, specification I slightly outperforms the more common specification II. However, regardless of sample size or event-induced variance, specification III consistently outperforms both of them. This was not a foregone conclusion, as while it is true that an additional independent variable will necessarily increase in-sample R^2, it will also increase out-of-sample forecast error because of the rather large correlation between industry and market indices. Nevertheless, the simulations suggest the benefit of adding an industry index to the CRSP equally-weighted market index outweighs the costs created by multicollinearity. Very similar results occur in Table VI when we use value weighted indices (specifications IV, V, and VI).

We next consider natural peers of the Fama-French Three-Factor Model (specification VII), specifically what happens if the market index is replaced with an equally weighted industry index (specification VIII) or a value-weighted industry index (specification IX), or if it is supplemented by these industry indices (specifications X and XI). The comparisons are summarized in Table VII. Once again there is a familiar theme: use of an industry index improves test power. Here it makes little difference whether this industry index is value weighted or equally weighted, or whether it is used to replace the market index in the Fama-French Three-Factor model or used to supplement it. For example, in the presence of event-

induced variance when N = 500, the FF Model's rejection rate is 18.08%. This rejection rate is improved by anywhere from a little over 7% (when the industry index replaces the market index) to a little over 10% (when the industry indices supplement the market index).

Finally, we consider similar comparisons for the Carhart Four-Factor Model, the results of which are summarized in Table VIII. The results here are a bit different, specifically, there is virtually no difference when an industry index replaces or supplements the market index in the presence of event-induced variance when sample size is only N = 100, but there is a slight improvement under other conditions. The largest improvement in test power (12.4%) occurs for specification XVI (industry value-weighted index supplements the Carhart model's four factors) when N = 1000.

IV. Conclusions

King (1966) found that an industry index explains about 10% of a stock's variance of returns, but subsequent event-study methods have relied on market indices instead of industry indices. We find that consideration of industry indices generally improves test power by around 10% when it is used to supplement a market index, and improves test power (albeit by a slightly smaller amount) when it is used to replace the market index. Given that more powerful tests are better, we recommend that industry indices be more widely used in event studies.

References

Boehmer, E., J. Musumeci, and A. Poulsen, 1991, "Event Study Methodology under Conditions of Event-Induced Variance," *Journal of Financial Economics*, 30: 253-272.

Brown, S.J., and J.B. Warner, 1980, "Measuring Security Price Performance," *Journal of Financial Economics*, 8(3): 205-258.

Brown, S.J., and J.B. Warner, 1985, "Using Daily Stock Returns: The Case of Event Studies," *Journal of Financial Economics*, 14(1): 3-31.

Carhart, M., 1997, On Persistence in Mutual Fund Performance, *The Journal of Finance*, 52(1): 57-82.

Cowan, A.R., 2007, Eventus®, Eventus User's Guide: Software Version 8.0, Standard Edition 2.1, Cowan Research L.C.

Cremers, M., A. Petajisto, and E. Zitzewitz, 2012, Should Benchmark Indices Have Alpha? Revisiting Performance Evaluation, *Critical Finance Review* 2: 1-48.

Fama, E., and K. French, 1992, The Cross-Section of Expected Stock Returns, *The Journal of Finance*, 47(2): 427-465.

Fama, E., and K. French, 1997, Industry Costs of Equity, *Journal of Financial Economics*, 43(2): 153-193.

Harrington, S., and D. Shrider, 2007, "All Events Induce Variance: Analyzing Abnormal Returns When Effects Vary across Firms," *Journal of Financial and Quantitative Analysis*, 42(1): 229-256.

King, B., 1966, Market and Industry Factors in Stock Price Behavior, *Journal of Business* 39(1): 139-190.

Kmenta, J., 1971, *Elements of Econometrics*, (MacMillan, New York, NY).

Marks, J., and J. Musumeci, 2017, Misspecification in Event Studies, *Journal of Corporate Finance*, 45: 333-341.

Patell, J., 1976, "Corporate Forecasts of Earnings Per Share and Stock Price Behavior: Empirical Test," *Journal of Accounting Research*, 14(2): 246-276.

Table I: Summary Statistics for Regressions

Specification:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
Alpha	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
CRSP EW		1.070	0.270													
CRSP VW					0.858	0.417	0.934			0.296	0.547	0.934			0.302	0.555
SMB							0.677	-0.021	0.465	0.181	0.601	0.676	-0.019	0.448	0.186	0.604
HML							0.195	-0.094	-0.014	0.035	0.161	0.184	-0.090	-0.014	0.036	0.155
MOM												-0.035	0.040	0.023	0.006	-0.030
Industry EW	0.948		0.778					0.909		0.723			0.895		0.717	
Industry VW				0.703		0.438			0.668		0.384			0.646		0.378
R^2	0.1572	0.1366	0.1743	0.1567	0.1338	0.1785	0.1713	0.1855	0.1896	0.2028	0.2092	0.1828	0.1963	0.2005	0.2125	0.2186
Residual σ	0.0241	0.0245	0.0240	0.0242	0.0245	0.0240	0.0242	0.0239	0.0239	0.0237	0.0237	0.0241	0.0238	0.0238	0.0237	0.0236
C (forecast error adjustment)	1.0089	1.0091	1.0137	1.0087	1.0088	1.0134	1.0186	1.0185	1.0184	1.0234	1.0234	1.0242	1.0240	1.0239	1.0291	1.0291

Parameters based on a 120-day estimation period preceding each "event" candidate in the CRSP database. There were 37,587,725 observations satisfying the $5 price filter. Not surprisingly, specifications with more factors generally produced higher values for R^2 and smaller in-sample standard deviations of the residual, but larger values of the forecast error adjustment, C, because of correlation between the factors.

Table II: Rejection rates at α = 5% when $\bar{\Delta}$ = 0% and θ = 0 (no abnormal return and no event-induced variance)

Panel A: Patell test

Specification:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
N = 100	6.31%	6.48%	6.30%	6.26%	6.32%	6.56%	6.41%	6.01%	6.52%	6.80%	6.37%	6.03%	6.12%	6.09%	6.11%	6.45%
N = 250	6.34%	6.76%	6.33%	5.94%	6.52%	6.47%	6.45%	6.42%	6.42%	6.21%	6.35%	6.46%	6.52%	6.58%	6.30%	6.77%
N = 500	6.78%	6.89%	6.44%	6.72%	6.68%	6.52%	6.81%	6.88%	6.67%	6.49%	6.27%	6.33%	6.82%	6.18%	5.99%	6.43%
N = 1000	6.80%	7.31%	6.45%	6.64%	6.86%	6.11%	6.69%	7.22%	5.82%	6.55%	6.54%	6.68%	7.29%	6.58%	6.68%	6.62%
N = 2500	7.28%	7.06%	7.51%	6.49%	6.70%	6.52%	7.02%	7.41%	6.92%	6.80%	6.80%	6.94%	7.37%	6.60%	6.97%	6.71%
N = 5000	7.91%	8.68%	8.25%	6.78%	7.24%	6.45%	7.68%	7.40%	6.90%	7.80%	7.05%	7.08%	7.02%	7.07%	7.48%	6.97%

Panel B: BMP test

Specification:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
N = 100	4.95%	5.08%	5.08%	5.18%	4.97%	5.08%	5.09%	4.99%	5.16%	4.92%	4.74%	4.61%	5.04%	4.77%	5.02%	5.13%
N = 250	5.07%	5.37%	5.03%	4.43%	5.14%	5.01%	5.25%	4.80%	5.01%	4.77%	4.90%	4.68%	4.93%	5.23%	4.79%	5.25%
N = 500	5.34%	5.01%	5.19%	5.14%	5.01%	4.96%	4.97%	5.37%	5.10%	4.92%	4.84%	4.55%	5.39%	4.54%	4.54%	4.97%
N = 1000	5.01%	5.41%	5.05%	5.07%	4.95%	4.37%	5.01%	5.58%	4.56%	5.02%	4.99%	4.89%	5.45%	5.01%	5.12%	4.94%
N = 2500	5.37%	5.18%	5.55%	4.91%	4.84%	4.95%	5.28%	5.46%	5.18%	5.06%	5.29%	5.37%	5.39%	4.83%	5.15%	4.94%
N = 5000	5.90%	6.20%	6.22%	5.12%	5.28%	4.65%	5.50%	5.30%	5.30%	5.90%	5.32%	5.42%	5.16%	5.29%	5.69%	5.06%

The theoretical rejection rates should be 5.00%. Consistent with Marks and Musumeci (2017), we find the Patell test is misspecified, but no evidence that the BMP test is.

Table III: Rejection rates at α = 5% when $\bar{\Delta}$ = 0% and θ = 1 (no abnormal return and event-induced variance equal to estimation-period residual variance)

Panel A: Patell test

Specification:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
N = 100	17.02%	17.20%	17.25%	17.64%	17.57%	17.51%	17.61%	16.99%	17.31%	17.16%	16.94%	16.84%	16.81%	16.57%	17.08%	16.76%
N = 250	17.40%	17.44%	17.24%	16.94%	17.24%	17.44%	17.17%	17.89%	17.41%	16.86%	16.82%	16.98%	16.18%	17.46%	17.11%	17.08%
N = 500	17.82%	17.79%	17.74%	18.11%	18.22%	17.46%	17.24%	17.90%	17.61%	17.14%	16.92%	17.25%	17.64%	17.11%	17.43%	16.80%
N = 1000	18.23%	17.65%	17.65%	17.50%	18.28%	16.51%	17.50%	17.68%	17.04%	17.36%	16.99%	17.00%	17.36%	17.41%	17.51%	16.98%
N = 2500	18.36%	18.26%	18.13%	17.24%	17.54%	17.02%	18.24%	17.74%	17.34%	18.13%	18.28%	17.65%	17.96%	17.39%	17.34%	17.59%
N = 5000	18.74%	19.59%	18.45%	18.06%	17.81%	17.63%	18.62%	17.89%	17.92%	18.65%	17.98%	17.13%	17.80%	17.96%	17.36%	18.00%

Panel B: BMP test

Specification:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
N = 100	5.06%	4.98%	5.28%	5.29%	5.27%	5.55%	5.22%	5.09%	5.31%	5.37%	5.12%	4.94%	4.95%	5.26%	5.53%	5.15%
N = 250	4.80%	4.99%	4.77%	4.87%	5.04%	5.45%	5.01%	5.21%	5.04%	4.74%	5.38%	4.84%	4.67%	5.06%	5.06%	5.45%
N = 500	4.88%	5.18%	4.92%	5.16%	5.15%	4.78%	4.91%	5.54%	5.07%	4.86%	4.71%	4.94%	5.48%	5.00%	4.94%	4.68%
N = 1000	5.19%	5.11%	4.94%	5.23%	4.97%	4.60%	5.05%	5.33%	4.82%	5.23%	4.49%	4.91%	5.16%	5.15%	5.02%	4.68%
N = 2500	5.53%	5.52%	5.54%	5.04%	4.98%	5.22%	5.25%	5.30%	5.16%	5.20%	5.40%	5.33%	5.31%	5.30%	5.13%	5.14%
N = 5000	5.37%	5.83%	5.51%	5.26%	5.11%	4.89%	5.31%	5.03%	5.47%	5.71%	5.02%	5.01%	5.20%	5.01%	5.33%	4.98%

The theoretical rejection rates should be 5.00%. Consistent with the previous literature [Boehmer et al (1991), Harrington and Shrider (2007), and Marks and Musumeci (2017)], we find the Patell test is substantially misspecified, but no evidence that the BMP test is.

Table IV: BMP Rejection rates at α = 5% when $\bar{\Delta}$ = 0.125%

Panel A: θ = 0 (no abnormal return and no event-induced variance)

Specification:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
N = 100	10.06%	10.04%	10.52%	9.68%	9.57%	10.54%	9.45%	10.44%	9.81%	10.13%	10.37%	9.92%	10.57%	10.07%	10.31%	9.97%
N = 250	18.89%	18.61%	19.33%	18.00%	17.33%	19.00%	17.25%	18.56%	18.24%	18.55%	18.41%	18.02%	18.50%	19.18%	19.18%	18.82%
N = 500	33.11%	32.19%	33.81%	31.53%	29.48%	32.91%	29.79%	32.58%	32.00%	32.73%	33.41%	30.62%	32.11%	32.24%	33.18%	33.23%
N = 1000	57.06%	57.17%	58.63%	55.49%	53.03%	58.01%	53.95%	56.65%	56.43%	57.80%	57.56%	54.08%	56.28%	56.81%	57.59%	57.72%
N = 2500	92.06%	92.14%	93.35%	91.25%	89.55%	92.07%	90.01%	92.09%	91.91%	92.72%	92.93%	90.01%	91.81%	91.78%	92.00%	92.55%
N = 5000	99.78%	99.73%	99.75%	99.70%	99.52%	99.80%	99.46%	99.79%	99.76%	99.79%	99.73%	99.50%	99.78%	99.63%	99.72%	99.66%

Panel B: θ = 1 (no abnormal return, event-induced variance equal to estimation-period residual variance)

Specification:	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
N = 100	7.87%	7.75%	8.48%	7.49%	7.79%	7.92%	7.45%	8.05%	7.73%	7.78%	7.80%	7.91%	8.06%	7.61%	7.87%	7.95%
N = 250	11.91%	11.90%	12.08%	11.42%	11.34%	11.94%	11.31%	12.28%	11.25%	12.36%	12.43%	12.15%	12.33%	12.34%	12.67%	12.24%
N = 500	19.43%	19.31%	20.25%	19.21%	17.67%	19.25%	18.08%	19.44%	19.42%	19.97%	19.93%	18.67%	19.72%	19.90%	20.38%	20.45%
N = 1000	33.33%	34.26%	36.03%	32.83%	31.46%	35.04%	32.57%	34.53%	34.41%	35.09%	34.31%	32.42%	33.58%	34.73%	35.00%	36.44%
N = 2500	69.64%	69.82%	71.11%	67.00%	64.55%	69.16%	65.88%	68.79%	68.96%	69.75%	70.35%	66.74%	68.90%	68.66%	69.85%	71.40%
N = 5000	93.19%	93.64%	94.83%	93.18%	91.41%	93.68%	91.76%	93.91%	93.27%	94.26%	94.05%	92.28%	93.82%	93.45%	93.96%	94.43%

Table V: Comparison of rejection rates by specifications I (equally weighted industry index), II (equally weighted market index), and III (both an equally weighted industry index and an equally weighted market index) when simulated abnormal performance ($\bar{\Delta}$) = 0.125%

	$\theta = 0$ N = 100	$\theta = 0$ N = 250	$\theta = 0$ N = 500	$\theta = 0$ N = 1000		$\theta = 1$ N = 100	$\theta = 1$ N = 250	$\theta = 1$ N = 500	$\theta = 1$ N = 1000
I (equally weighted industry index)	10.06%	18.89%	33.11%	57.06%		7.87%	11.91%	19.43%	33.33%
II (equally weighted market index)	10.04%	18.61%	32.19%	57.17%		7.75%	11.90%	19.31%	34.26%
III (equally weighted industry index and an equally weighted market index)	10.52%	19.33%	33.81%	58.63%		8.48%	12.08%	20.25%	36.03%

Table VI: Comparison of rejection rates by specifications IV (value-weighted industry index), V (value-weighted market index), and VI (both a value-weighted industry index and a value-weighted market index) when simulated abnormal performance ($\bar{\Delta}$) = 0.125%

	$\theta = 0$ N = 100	$\theta = 0$ N = 250	$\theta = 0$ N = 500	$\theta = 0$ N = 1000		$\theta = 1$ N = 100	$\theta = 1$ N = 250	$\theta = 1$ N = 500	$\theta = 1$ N = 1000
IV (value-weighted industry index)	9.68%	18.00%	31.53%	55.49%		7.49%	11.42%	19.21%	32.83%
V (value-weighted market index)	9.57%	17.33%	29.48%	53.03%		7.79%	11.34%	17.67%	31.46%
VI (value-weighted industry index and a value-weighted market index)	10.54%	19.00%	32.91%	58.01%		7.92%	11.94%	19.25%	35.04%

Table VII: Comparison of rejection rates by specifications VII (Fama-French Three Factor Model), VIII (FF market index replaced with an equally weighted industry index), IX (FF market index replaced with a value-weighted industry index), X (FF 3-factor supplemented with an equally weighted industry index), and XI (FF 3-factor supplemented with a value-weighted industry index) when simulated abnormal performance ($\bar{\Delta}$) = 0.125%

	$\theta = 0$ N = 100	$\theta = 0$ N = 250	$\theta = 0$ N = 500	$\theta = 0$ N = 1000		$\theta = 1$ N = 100	$\theta = 1$ N = 250	$\theta = 1$ N = 500	$\theta = 1$ N = 1000
VII (FF 3-factor)	9.45%	17.25%	29.79%	53.95%		7.45%	11.31%	18.08%	32.57%
VIII (FF 3-factor with market index replaced with an equally weighted industry index)	10.44%	18.56%	32.58%	56.65%		8.05%	12.28%	19.44%	34.53%
IX (FF 3-factor with market index replaced with a value-weighted industry index)	9.81%	18.24%	32.00%	56.43%		7.73%	11.25%	19.42%	34.41%
X (FF 3-factor supplemented with an equally weighted industry index)	10.13%	18.55%	32.73%	57.80%		7.78%	12.36%	19.97%	35.09%
XI (FF 3-factor supplemented with a value-weighted industry index)	10.37%	18.41%	33.41%	57.56%		7.80%	12.43%	19.93%	34.31%

Table VIII: Comparison of rejection rates by specifications XII (Carhart Four-Factor Model), XIII (Carhart market index replaced with an equally weighted industry index), XIV (Carhart market index replaced with a value-weighted industry index), XV (Carhart 4-factor supplemented with an equally weighted industry index), and XVI (Carhart 4-factor supplemented with a value-weighted industry index) when simulated abnormal performance ($\bar{\Delta}$) = 0.125%

	$\theta = 0$ N = 100	$\theta = 0$ N = 250	$\theta = 0$ N = 500	$\theta = 0$ N = 1000		$\theta = 1$ N = 100	$\theta = 1$ N = 250	$\theta = 1$ N = 500	$\theta = 1$ N = 1000
XII (Carhart 4-factor)	9.92%	18.02%	30.62%	54.08%		7.91%	12.15%	18.67%	32.42%
XIII (Carhart 4-factor with market index replaced with an equally weighted industry index)	10.57%	18.50%	32.11%	56.28%		8.06%	12.33%	19.72%	33.58%
XIV (Carhart 4-factor with market index replaced with a value-weighted industry index)	10.07%	19.18%	32.24%	56.81%		7.61%	12.34%	19.90%	34.73%
XV (Carhart 4-factor supplemented with an equally weighted industry index)	10.31%	19.18%	33.18%	57.59%		7.87%	12.67%	20.38%	35.00%
XVI (Carhart 4-factor supplemented with a value-weighted industry index)	9.97%	18.82%	33.23%	57.72%		7.95%	12.24%	20.45%	36.44%