

EVENT STUDIES IN SECURITIES LITIGATION: LOW POWER, CONFOUNDING EFFECTS, AND BIAS

ALON BRAV*
J.B. HEATON**

ABSTRACT

An event study is a statistical method for determining whether some event—such as the announcement of earnings or the announcement of a proposed merger—is associated with a statistically significant change in the price of a company’s stock. The main inputs to an event study are historical stock returns for the companies under study, benchmark returns like the return to the broader stock market, and standard statistical tests like t-tests that are used to test for statistical significance. In securities litigation and regulation, event studies are used primarily to detect the impact of disclosures of alleged fraud on the price of a single traded security.

But are event studies in securities litigation reliable? What is interesting about the use of event studies in securities litigation is that the methodology litigants use in court differs from the methodology that economists apply in their research. With few exceptions, securities litigation event studies are single-firm event studies, while almost all academic research event studies are multi-firm event studies. Multi-firm event studies are generally accepted in financial economics research, and peer-reviewed journals contain them by the hundreds. By contrast, single-firm event studies—the mainstay of modern securities fraud litigation—are almost nonexistent in peer-reviewed journals.

* Robert L. Dickens Professor, Duke University Fuqua School of Business and National Bureau of Economic Research. brav@duke.edu.

** Partner, Bartlit Beck Herman Palenchar & Scott LLP. jb.heaton@bartlit-beck.com. The views expressed here are Heaton’s own, and do not express the views of Bartlit Beck Herman Palenchar & Scott LLP, its attorneys, or its clients.

For very helpful comments and suggestions, we thank Josh Ackerman, Reid Bolton, Peter Clayburgh, Brad Cornell, Chris Culp, Kevin Dages, Tiago Duarte-Silva, Marc Gross, Chris Hagale, Jeff Hall, Mike Keable, Ashley Keller, Chris Landgraff, Dan McElroy, Katherine Minarik, Mark Mitchell, Steve Nachtwey, Martha Pacold, David Ross, Cindy Sobel, David Tabak, Sanjay Unni, David Wensel, and participants at the 21st Annual Institute for Law and Economic Policy Conference: New Directions for Corporate and Securities Litigation (co-sponsored by the *Washington University Law Review* and the Institute for Law and Economic Policy). We are especially grateful to Jim Cox for inviting us to think hard about single-firm event studies. All errors are our own.

Importing a methodology that economists developed for use with multiple firms into a single-firm context creates three substantial difficulties. First, single-firm event studies suffer from a severe signal-to-noise problem in that they lack statistical power to detect price impacts unless the price impacts are quite large. Inattention to statistical power lowers the deterrent effect of the securities laws by giving a “free pass” to some economically meaningful price impacts and may encourage more small- and mid-scale fraud than is socially optimal given the costs of litigation. Second, single-firm event studies do not average away confounding effects. While this problem is well known, some courts have unrealistic expectations of litigants’ ability to quantitatively decompose observed price impacts into those caused by alleged fraud and those unrelated to alleged fraud. Third, low statistical power and confounding effects combine to generate sizeable upward bias in detected price impacts and therefore in damages. To improve the accuracy of adjudication in securities litigation, we suggest that litigants report the statistical power of their event studies, that courts allow litigants flexibility to deal with the problem of confounding effects, and that courts and litigants consider the possibility of upward bias in the detection of price impacts and the estimation of damages.

TABLE OF CONTENTS

INTRODUCTION.....	585
I. DIFFICULTY #1: LOW STATISTICAL POWER.....	589
A. <i>Abnormal Returns and Type I Error</i>	589
B. <i>Abnormal Returns and Type II Error</i>	593
C. <i>Statistical Significance and Likelihood Ratios</i>	597
D. <i>Low Power and Statistical Insignificance in Case Law</i>	599
E. <i>An Aside on Power and the MFES</i>	603
II. DIFFICULTY #2: CONFOUNDING EFFECTS	605
A. <i>The Problem</i>	605
B. <i>Confounding Effects in Case Law</i>	606
III. DIFFICULTY #3: BIAS	608
CONCLUSION.....	612

INTRODUCTION

An event study is a statistical method for determining whether some event—such as the announcement of earnings or the announcement of a proposed merger—is associated with a statistically significant change in the price of a company’s stock.¹ The main inputs to an event study are historical stock returns for the companies under study, benchmark returns like the return to the broader market, and standard statistical tests like *t*-tests that are used to test for statistical significance. In securities litigation and regulation, event studies are used primarily to detect the impact of disclosures of alleged fraud on the price of a traded security.

After the Supreme Court endorsed the fraud-on-the-market doctrine in *Basic Inc. v. Levinson*² in 1988, event studies became so entrenched in securities litigation that they are viewed as necessary in every case.³ Based on the efficient markets hypothesis that “the market price of shares traded on well-developed markets reflects all publicly available information, and, hence, any material misrepresentations,”⁴ securities litigants use the event study to help answer two crucial questions. First, was there a price impact

1. Our description here assumes little familiarity with the basics of the event study methodology. A recent book covering event study methodology is DORON KLIGER & GREGORY GUREVICH, *EVENT STUDIES FOR FINANCIAL RESEARCH* (2014). Other good sources for those with an intermediate background on the subject are JOHN Y. CAMPBELL, ANDREW W. LO, & A. CRAIG MACKINLAY, *THE ECONOMETRICS OF FINANCIAL MARKETS* 149–80 (1997) (chapter on event-study analysis); A. Craig MacKinlay, *Event Studies in Economics and Finance*, 35 J. ECON. LIT. 13, 13–14 (1997); and S.P. Kothari & Jerold B. Warner, *Econometrics of Event Studies*, in *HANDBOOK OF CORPORATE FINANCE 3* (B. Espen Eckbo ed., 2007). A good introduction to the Single-Firm Event Study (“SFES”) context is Mark L. Mitchell & Jeffry M. Netter, *The Role of Financial Economics in Securities Fraud Cases: Applications at the Securities and Exchange Commission*, 49 BUS. LAW. 545 (1994); see also M. Laurentius Marais & Katherine Schipper, *Applications of Event Study Methods in Litigation Services*, in *LITIGATION SERVICES HANDBOOK: THE ROLE OF THE ACCOUNTANT AS EXPERT* 45.1 (Roman L. Weil ed., 2d ed. 1995); Charles J. Corrado, *Event Studies: A Methodology Review*, 51 ACCT. & FIN. 207 (2011); Frank Torchio, *Proper Event Study Analysis in Securities Litigation*, 35 J. CORP. L. 159 (2009).

2. 485 U.S. 224 (1988).

3. See, e.g., *Bricklayers & Trowel Trades Int’l Pension Fund v. Credit Suisse Secs. LLC*, 752 F.3d 82, 86 (1st Cir. 2014) (“The usual—it is fair to say ‘preferred’—method of proving loss causation in a securities fraud case is through an event study, in which an expert determines the extent to which the changes in the price of a security result from events such as disclosure of negative information about a company, and the extent to which those changes result from other factors.”); *In re Oracle Sec. Litig.*, 829 F. Supp. 1176, 1181 (N.D. Cal. 1993) (citation omitted) (“Use of an event study or similar analysis is necessary . . . to [more accurately] isolate the influences of information specific to Oracle which defendants allegedly have distorted. As a result of his failure to employ such a study, the results reached by [plaintiffs’ expert] cannot be evaluated by standard measures of statistical significance. Hence, the reliability of the magnitude and direction of his value estimates are incapable of verification.”).

4. *Basic*, 485 U.S. at 246.

at the time of an alleged misrepresentation or corrective disclosure? Second, if there was a price impact, how much of it was caused by the alleged misrepresentation or corrective disclosure as opposed to other, unrelated factors?⁵ In proposing answers to these questions, litigants have not been shy in asserting the event study's impressive academic pedigree.⁶ But the methodology that litigants use in court differs from the methodology used in academic research. In particular, securities litigation event studies are almost always single-firm event studies ("SFESs") that examine the price moves of the security of the single firm involved in the litigation,⁷ while almost all academic research event studies are multi-firm event studies ("MFESs") that examine large samples of securities from multiple firms.⁸ Importing a methodology that economists developed for use with multiple firms into a single-firm context creates three substantial difficulties: low statistical power, confounding effects, and bias.

First, an SFES often has low statistical "power" to detect an economically meaningful price impact, which typically must be at least approximately twice as large as the standard deviation of daily (abnormal) returns for the examined firm. But requiring conventional levels of statistical significance when power is low effectively gives a "free pass" to economically meaningful securities fraud because the SFES simply cannot detect price impacts below a high threshold. Courts, ignoring low power,

5. See, e.g., Marais & Schipper, *supra* note 1, at 45.1 (phrasing the two questions: as "Did the announcement cause a price reaction? What was the price reaction to the announcement alone?").

6. See, e.g., *In re Cendant Corp. Sec. Litig.*, 109 F. Supp. 2d 235, 253–54 (D.N.J. 2000) ("Here, according to [plaintiffs,] the methodology—'event study methodology'—used to calculate shareholder damages during the class period 'has been used by financial economists since 1969 as a tool to measure the effect on market prices from all types of new information relevant to a company's equity valuation.'").

7. See, e.g., *In re Northern Telecom Ltd. Sec. Litig.*, 116 F. Supp. 2d 446, 456 (S.D.N.Y. 2000) ("Defendants' expert . . . conducted an 'event study' analysis to determine whether the six statements at issue had any impact on the trading price of Nortel's stock.").

8. See, e.g., Kothari & Warner, *supra* note 1, at 8 ("An event study typically tries to examine return behavior for a sample of firms experiencing a common type of event (e.g., a stock split)."). SFESs are rare in the peer-reviewed literature. See, e.g., Corrado, *supra* note 1, at 209 ("The skeletal econometric structure of an event study is well illustrated by the case of a single security-event date combination. While such studies rarely find their way into a research journal, they are common in certain legal proceedings."). One example of a single-firm event study in the peer-reviewed finance literature is Richard S. Ruback, *The Cities Service Takeover: A Case Study*, 38 J. FIN. 319 (1983). For applications of the single-firm event study methodology in law reviews and legal journals, see Jonathan Klick & Robert H. Sitkoff, *Agency Costs, Charitable Trusts, and Corporate Control: Evidence from Hershey's Kiss-Off*, 108 COLUM. L. REV. 749 (2008) (studying the announced sale and then cancellation of the sale of a controlling interest in Hershey Company), and Mark I. Weinstein, *Don't Buy Shares Without It: Limited Liability Comes to American Express*, 37 J. LEGAL STUD. 189 (2008) (studying the effect of transition from unlimited to limited liability on American Express stock).

then conclude that some economically large price impacts are immaterial. Courts err because of their mistaken premise that statistical insignificance indicates the probable absence of a price impact. Overreliance on statistical significance without consideration of statistical power “leads to a decision-making regime in which the probability of an incorrect exoneration far exceeds the probability of an incorrect condemnation.”⁹ While it is possible that this regime reflects a rational policy judgment, we see no evidence such a judgment has been made deliberately.

Second, when an SFES does detect a price impact, it reflects confounding effects that are unrelated to the alleged fraud. Unfortunately, there is no fully reliable, mathematically precise way to decompose an observed event return in an SFES into component parts: the part related to alleged fraud and the part not related to alleged fraud. Financial economists have long understood that our ability to fully explain observed price moves is quite limited;¹⁰ much price movement occurs for reasons unrelated to news, including as a result of the liquidity trades of investors in the market seeking to raise funds for other purposes and the (at least short-term) impact of “noise traders” who trade for irrational reasons.

Third, low statistical power and confounding effects combine to generate sizeable upward bias in detected price impacts and damages (i.e., overstating the magnitude of a price impact and damages). This upward bias problem means that we cannot leave confounding effects unaddressed in the hope that they are as likely to be on one side of the true price effect as on the other. For example, suppose the true price impact is -2.0%, but the requirement of statistical significance is such that price impacts less severe than -2.94% will be rejected as statistically insignificant. In that case, a price impact will be detected only when there are confounding effects that push the observed price impact past -2.94%. As we show later, the expected detected price impact in such situations is -3.9%, substantially higher than -2.0%, the true price impact.

These problems help explain why the SFES methodology is applied so infrequently in peer-reviewed research. But the same problems have not limited the use of the SFES in securities litigation. Securities litigants use SFESs to show that securities did or did not trade in an efficient market, to establish that alleged misrepresentations did or did not impact the stock price for purposes of materiality and reliance, and to determine the

9. Michael I. Meyerson & William Meyerson, *Significant Statistics: The Unwitting Policy Making of Mathematically Ignorant Judges*, 37 PEPP. L. REV. 771, 840 (2010).

10. See, e.g., Richard Roll, *R²*, 43 J. FIN. 541, 541 (1988) (“Even with hindsight, the ability to explain stock price changes is modest.”).

existence or absence of loss causation and amount of damages. We are not the first to point out that SFESs have low statistical power and are subject to problems with confounding effects (though we are, to our knowledge, the first to point out the bias problem in this context).¹¹ But especially since courts are increasingly required to address price impact evidence at the class certification stage by using event studies,¹² it is time to review the limitations of the single-firm event study in securities litigation—particularly those limitations that arise from low power, confounding effects, and bias—in order to provide courts and litigants with a firmer basis for considering evidence based on single-firm event studies.

In Part I, we explain why the SFES as typically applied in securities litigation has low statistical power, in the sense that it cannot detect price impacts reliably unless they are large. In Part II, we explain the problem of confounding effects. In Part III, we explain how low statistical power and confounding effects combine to generate bias in detected price impacts. We conclude with proposals for improving the accuracy of adjudication involving SFESs. These include requiring litigants to report the power of their analyses, allowing litigants flexibility to address the problem of confounding effects, and encouraging courts and litigants to consider the possibility of upward bias in the detection of price impacts and the estimation of damages.¹³

11. See Sanjai Bhagat & Roberta Romano, *Event Studies and the Law: Part I: Technique and Corporate Litigation*, 4 AM. LAW & ECON. REV. 141, 148–50 (2002). Others who have criticized the role of the event study in securities litigation on various grounds include Jill E. Fisch, *The Trouble With Basic: Price Distortion After Halliburton*, 90 WASH. U. L. REV. 895, 918–22 (2013); Michael J. Kaufman & John M. Wunderlich, *Regressing: The Troubling Dispositive Role of Event Studies in Securities Fraud Litigation*, 15 STAN. J.L. BUS. & FIN. 183, 230–33 (2009); William O. Fisher, *Does the Efficient Market Theory Help Us Do Justice in a Time of Madness?*, 54 EMORY L.J. 843, 880 (2005); Bradford Cornell & R. Gregory Morgan, *Using Finance Theory to Measure Damages in Fraud on the Market Cases*, 37 UCLA L. REV. 883, 906–11 (1990).

12. This is a result of the Supreme Court's decision in *Halliburton Co. v. Erica P. John Fund, Inc.*, 134 S.Ct. 2398, 2417 (2014) (holding defendants may defeat the fraud-on-the-market presumption of reliance at the class-certification stage through evidence that the misrepresentation did not in fact affect the stock price).

13. Another important problem of the SFES that we do not address here is non-normality in returns. That problem has been addressed to great extent elsewhere and is remedied by appropriate changes in statistical inference methods. See Jonah B. Gelbach et al., *Valid Inference in Single-Firm, Single-Event Studies*, 15 AM. L. & ECON. REV. 495 (2013).

I. DIFFICULTY #1: LOW STATISTICAL POWER

A. *Abnormal Returns and Type I Error*

The main question addressed in an event study is whether a particular event was associated with a change in the price of a firm's securities. In securities litigation,

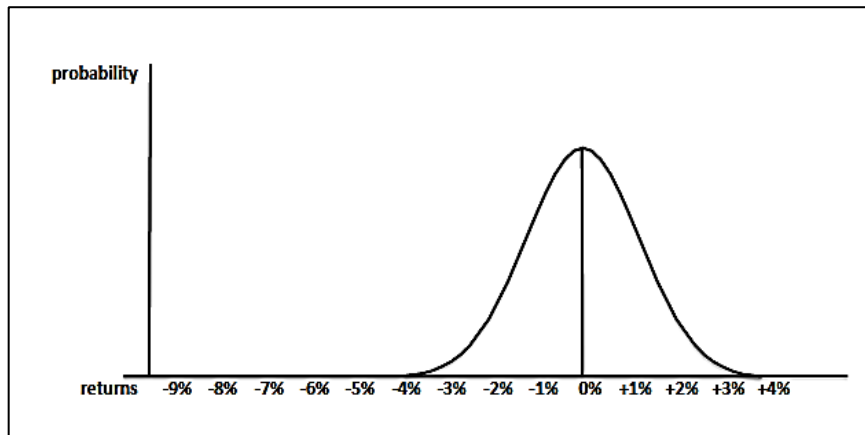
[t]he classic example of a loss-inducing event is a corrective disclosure by the company itself. A corrective disclosure is traditionally an admission by the company that one or more of its previous statements were false or misleading followed by a corrected, truthful and complete version of those statements. The event need not take this form, however. The event could be a credit ratings downgrade, or the collapse of the company.¹⁴

Financial economists typically analyze price changes by analyzing returns (i.e., change in price, plus dividends or other distributions, divided by price, which is more easily comparable across firms than price). An event changes the value of the firm's securities if it changes the probability distribution of security returns.¹⁵ Consider Figure 1, which shows an example probability distribution for the daily return of a hypothetical single stock, and which is reflective of the daily standard deviation and return of a typical large cap stock.

14. *In re Vivendi Universal, S.A. Sec. Litig.*, 605 F. Supp. 2d 586, 598 (S.D.N.Y. 2009) (footnote omitted) (citations omitted).

15. *See, e.g.*, MacKinlay, *supra* note 1, at 14 (describing the question in an event study as whether "the event has no impact on the distributions of returns").

FIGURE 1: EXAMPLE PROBABILITY DISTRIBUTION FOR A SINGLE FIRM'S DAILY ABNORMAL STOCK RETURNS



The mean of this probability distribution is at zero because we are interested in the part of a security's returns that are not explained by long-run average returns, returns in the broader market, or returns of the firm's industry or the like. When the market as a whole goes down, many securities tend to go down as well, since the market is just the collection of all securities. When all the other stocks in a firm's industry are moving in one direction on a given day (perhaps because of news in the market about the prospects for that industry), the returns of a firm in that industry will tend to move in that direction as well. In an event study, we want to remove the part of the returns of the examined firm that are explained by co-movement with market, industry, or other broader moves, so we can better isolate the firm-specific event we are studying. In essence, we are trying to remove any return that would have occurred anyway absent the event. The failure to make adjustments for the effect of market and industry moves nearly always dooms an analysis of securities prices in litigation.¹⁶ Typically, parties remove co-movement using market indexes

16. See, e.g., *Computer Aid, Inc. v. Hewlett-Packard Co.*, 56 F. Supp. 2d 526, 540 (E.D. Pa. 1999) (rejecting expert's event study analysis as "dubious" where it did not account for apparent industry-wide negative returns); *Carpe v. Aquila, Inc.*, No. 02-0388-CV-W-FJG, 2005 WL 1138833, at *4 (W.D. Mo. Mar. 23, 2005) (excluding plaintiffs' expert opinion because he did not perform a proper event study to exclude effects on price of market and industry movements and noting that "[f]ailure to conduct an event study comparing the stock's price to the market as a whole or a selected index of similar businesses is enough to cause an expert's opinion to be excluded").

and industry indexes, but these methods can become quite sophisticated in practice.

What is left over after subtracting the return predicted by co-movement with other stocks is called the “abnormal return.” On any given day, the abnormal return may fluctuate within a large range. For example, the standard deviation of the abnormal daily return in Figure 1 is 1.5%, so we would expect 95% of abnormal returns to occur between -2.94% and +2.94%. This is based on the normal distribution we’ve drawn here, normality still being the assumption of most work in practice.¹⁷

Now consider the occurrence of an event of interest like a corrective disclosure. The standard approach is to calculate a test statistic based on the observed abnormal return on the event day divided by the standard deviation of abnormal returns.¹⁸ If the test statistic exceeds a “critical” value,¹⁹ the return is said to be “statistically significant,” in the sense that a return of that magnitude has a relatively small probability of occurring if the event did not have a price impact.²⁰ Consider Figure 2. The lower and upper critical values are the abnormal returns (1.96 standard deviations from zero) that generate values of the test statistic beyond which the abnormal return would be considered statistically significant.²¹ In the usual

17. We focus on the normal case because standard practice still rests heavily on the normality assumption, despite strong evidence that daily abnormal returns are non-normal. See Stephen J. Brown & Jerold B. Warner, *Using Daily Stock Returns: The Case of Event Studies*, 14 J. FIN. ECON. 3, 4 (1985) (“The daily stock return for an individual security exhibits substantial departures from normality that are not observed with monthly data.”); Gelbach et al., *supra* note 13, at 511 (“[F]ew firms’ distributions are consistent with normality.”).

18. See, e.g., Corrado, *supra* note 8, at 211 (setting forth test statistic as the abnormal return divided by the square root of the variance of the abnormal return).

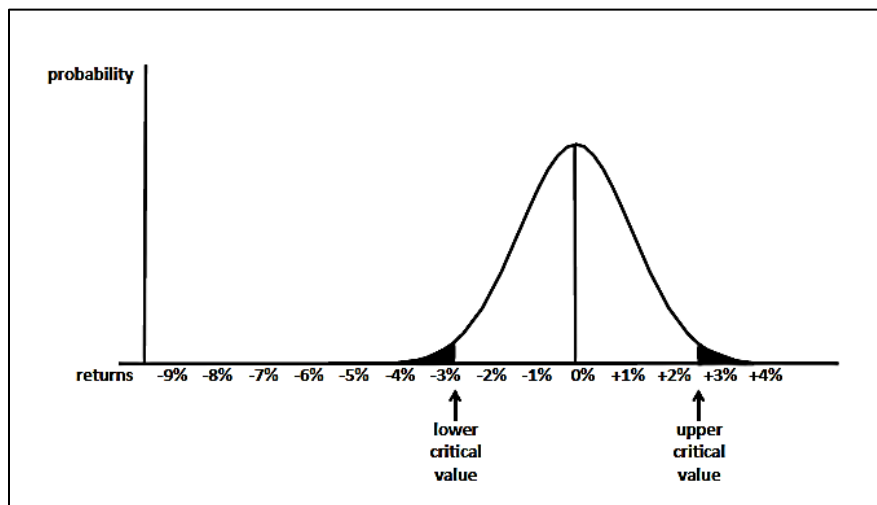
19. *Id.* (“Assuming returns are normally distributed, the test statistic . . . is distributed as Student-*t* with *n*-2 degrees of freedom (df).”).

20. *Id.* (“A test statistic larger than the upper-tail critical value . . . provides statistical evidence that the merger announcement had a significant positive impact on the stock price on the event date. Similarly, a test statistic less than the lower-tail critical value . . . would provide evidence that the announcement had a significant negative impact.”).

21. Note that while we have represented a two-tailed test here, one-tailed tests may be more appropriate in testing for the alternative of a price impact that is less than zero (the usual case for a corrective disclosure) or greater than zero (the usual case at the time of a misrepresentation that allegedly inflates the security price). See, e.g., Brown & Warner, *supra* note 17, at 12–13 (using a one-tailed 5% test for power analysis). There is usually no basis for objecting to an expert’s analysis on the grounds that the expert used a one-tailed test. Cf. *In re Novatel Wireless Sec. Litig.*, 910 F. Supp. 2d 1209, 1214 (S.D. Cal. 2012) (“Defendants contend that [plaintiffs’ expert] should have utilized a two-tailed test rather than a one-tailed test when considering the statistical significance of the stock price decline on July 20, 2007.”), *vacated on other grounds*, Civil No. 08cv1689 AJB (RBB), 2013 WL 494361 (S.D. Cal. Feb. 7, 2013). Of course, using a two-tailed test at the same significance level makes it less likely to detect statistical significance against an alternative known to be negative, and this is no doubt what defendants hoped to accomplish when they argued against the one-tailed test. The court rejected the arguments. *Id.* at 1216 (“Whereas [defendants’ expert] finds the use of the one-tailed

case, we will say that an event day return is statistically significant if it is greater in magnitude than one of the critical values (i.e., if it falls in the darkened tails, which theoretically extend infinitely in both directions with increasingly small probability).

FIGURE 2: STATISTICAL SIGNIFICANCE



The idea behind the selection of critical values is that, all else equal, we do not want to conclude that a price impact occurred from the event when what in fact happened is that we observed a routine return from the same abnormal return distribution centered at zero. In other words, we want to avoid improperly rejecting the possibility that the negative return reflected normal fluctuations and not the event. To address this problem, we require that the observed return be sufficiently extreme that we will make that mistake only a small percentage of the time. That percentage is called the significance level, or “size,” of the test. Type I error is the probability of concluding that the event caused a price impact when it did not.²² Ideally, we want a low probability of Type I error, and it is common to fix size at 5%, though other choices like 1%, 2.5%, or 10% are also used in research

test to be unreliable and [plaintiffs’ expert] has provided a reasonable explanation for using the one-tailed test, the Court finds that it is not an issue of admissibility, but rather of probative value to be addressed at trial.”)

22. See generally GEORGE CASELLA & ROGER L. BERGER, STATISTICAL INFERENCE 358–65 (1990).

practice.²³ It is crucial to understand that statistical significance is simply describing a set of returns that would be unusual to observe if there was *no* price impact. Lack of statistical significance does *not* tell us that it is more probable than not that there was no price impact.²⁴

B. Abnormal Returns and Type II Error

Power is the ability to detect a true effect (e.g., a price impact due to fraud) when it exists. “Analysis of the power of statistical tests is an important part of planning any scientific research study”²⁵ As described above, a Type I error is the error of concluding that there was an effect when there was none. The second kind of mistake is the converse: failing to find an effect that exists. That is, we observe an abnormal return that is not statistically significant and conclude that there was no price impact, when in fact there was one. This is known as a Type II error. The “power” of a test is one minus the probability of a Type II error.²⁶

Suppose that the true price impact of a corrective disclosure was -2.0%, so that the abnormal distribution on the event date had the same standard deviation as the no-event distribution but with a -2.0% mean rather than a zero mean. Figure 3 illustrates why the SFES will miss detecting that price impact with high probability.

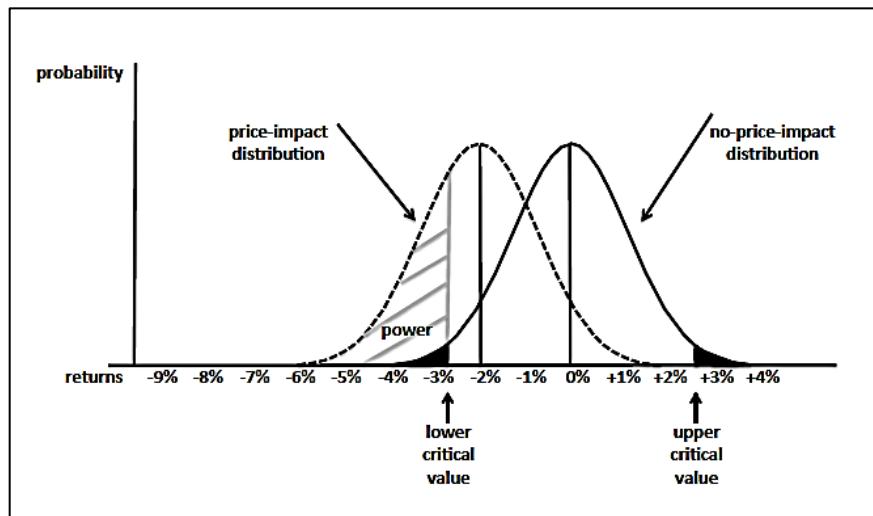
23. See, e.g., *id.* at 361 (“Experimenters commonly specify the level of the test they wish to use, with typical choices being [size] = .01, .05, and .10.”).

24. Steven Goodman, *A Dirty Dozen: Twelve P-Value Misconceptions*, 45 SEMINARS IN HEMATOLOGY 135, 136 (2008) (“A nonsignificant [effect] merely means that a null effect [here, no price impact] is statistically consistent with the observed results, together with the range of effects included in the confidence interval. It does not make the null effect [i.e., the hypothesis of no price impact] the most likely. The effect best supported by the data from a given experiment is always the observed effect, regardless of its significance.”).

25. Larry V. Hedges & Therese D. Pigott, *The Power of Statistical Tests in Meta-Analysis*, 6 PSYCHOL. METHODS 203, 216 (2001).

26. PAUL D. ELLIS, *THE ESSENTIAL GUIDE TO EFFECT SIZES: STATISTICAL POWER, META-ANALYSIS, AND THE INTERPRETATION OF RESEARCH RESULTS* 52 (2010) (“Statistical power describes the probability that a test will correctly identify a genuine effect. Technically, the power of a test is defined as the probability that it will reject a false null hypothesis. Thus, power is inversely related to . . . the probability of making a Type II error.”); JACOB COHEN, *STATISTICAL POWER ANALYSIS FOR THE BEHAVIORAL SCIENCES* 5 (2d. ed. 1988) (describing Type II error as (1-power)).

FIGURE 3: POWER



We can refer to the mean zero distribution as the “no-price-impact distribution” and the mean -2.0% distribution as the “price-impact distribution.” Having chosen the size of our test (which determines the areas of the darkened tails in the no-price-impact distribution), we can see that when the price-impact distribution generates the abnormal return (i.e., there is a price impact), we will make a Type II error anytime the return from the price-impact distribution is to the right of the lower critical value return but to the left of the upper critical value return. Our requirement of statistical significance leads us to the wrong inference in those cases where the price-impact distribution generated the return in that range. As we have drawn the distributions here, we will make a Type II error 73.4% of the time that there is a price impact, because 73.4% of the price-impact distribution is to the right of the critical value generating return and to the left of the upper critical value. Obviously, the further away the price-impact distribution is from the no-price-impact distribution, the more power we have. So the SFES can detect very large price impacts reliably, but it cannot detect smaller price impacts reliably. Unless the price impact is large relative to the standard deviation of returns (i.e., more than about two standard deviations away from zero), an SFES can easily miss it.

What is the likely importance of this problem in practice? In one review, power in the SFES context ranges from 5% to 17%, depending on assumptions of effect size and standard deviation.²⁷ But it may be easier to understand the power problem in dollar terms. In Table I, we use data from Wharton Research Data Services (“WRDS”) to sort U.S. firms with ordinary common shares into deciles by size from smallest to largest. We use a four-factor model to estimate abnormal returns for 2014 for each of 4298 firms that had data available for all 252 trading days.²⁸

TABLE I: SFES MINIMUM DETECTABLE PRICE IMPACT
USING 1.96X STANDARD DEVIATION TEST STATISTIC

Size Deciles	Avg. Mkt. Cap. (\$ millions)	Avg. Std. Dev. (% return)	Min. Detectable Impact. (\$ millions)
smallest 1	29	4.1	2.3
2	81	2.7	4.3
3	167	2.6	8.5
4	297	2.2	12.8
5	517	2.2	22.3
6	882	2.0	34.6
7	1,543	1.8	54.4
8	2,888	1.6	90.6
9	6,238	1.4	171.2
largest 10	41,597	1.1	896.8

The first column gives the ten size deciles, with the 10% of firms that are smallest by beginning-of-year 2014 equity market capitalization being in the 1st decile and so on through the 10% of firms that are largest in the 10th decile. The second column is the average beginning of year equity market capitalization for firms in that decile. The third column is the average standard deviation of 2014 abnormal returns for each firm in that decile. The fourth column is the minimum price impact on the equity value of the average-sized firm in that decile that would be detectable with a

27. See MacKinlay, *supra* note 1, at 29 tbl.2; see also Kothari & Warner, *supra* note 1, at 17–18 (providing charts appearing to show power starting at sample size of one).

28. Specifically, for each firm with available daily return data for the entire year we estimate a regression of the firm’s daily return in excess of the daily risk free rate on four factors, obtained from Ken French’s website at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. These factors are the market factor in excess of the risk free rate, portfolio factors meant to capture size, book-to-market effects in stock returns, and a portfolio factor meant to capture stock momentum effects. See KLIGER & GUREVICH, *supra* note 1, at 31 (four-factor model using Fama-French and Carhart factors “may allow for more precise measurement of [normal returns] and, consequently, more accurate estimation of market reaction to the studied event”). For each firm, we then compute the residual standard deviation obtained from these least squares regressions.

1.96 average standard deviation cut-off for that decile. Consider decile 8 firms—those in the top 30% of firms by size but not in the top 20%. These firms have an average beginning-of-year 2014 equity market capitalization of \$2.888 billion. The standard deviation of their abnormal returns in 2014 was 1.6%. If the test statistic requires that a detected abnormal return on event day be 1.96 standard deviations from zero to be statistically significant, then the event day price impact must be at least \$90.6 million for the SFES to detect it, assuming there are no confounding effects pushing it toward significance. For firms in decile 10—the biggest 10% of firms—the price impact must be nearly \$900 million to be detectable by an SFES. Of course, if there were confounding effects going in the opposite direction (e.g., positive news announced with a corrective disclosure), then the price impact would have to be even bigger for the SFES to detect it.

This is a strange state of affairs. Suppose, for example, that an average decile 10 firm announced that a now-fired management team had fraudulently overstated its cash by \$300 million, but the company could not recover against the management or the company's auditors who, we may assume, found the overstatement in the course of their audit. Suppose further that the market recognized the \$300 million loss and the value of the firm fell a further \$200 million on the belief that further problems might be forthcoming. All else equal, the value of the firm would decline \$500 million. There would be little doubt that the fraud caused at least \$300 million of that fall. But even the total price move (\$300 million plus \$200 million) would be statistically insignificant in an SFES because it falls below the approximately \$900 million detectability threshold for decile 10 firms. It is hard to believe this is the right conclusion. And indeed, we would expect a court to look beyond event study evidence to the more direct measure of impact on value that would be available in this example.

In Table II, we present power for combinations of standard deviations and price impacts. We assume a two-tailed test statistic as in Figures 2 and 3, and a test statistic that is 1.96 standard deviations from zero on both sides (i.e., a standard two-tailed significance level of 5%).²⁹

29. In terms of Figure 3, power is the area to the left of the lower critical value under an alternative distribution centered at the price impact with the same standard deviation, plus the (tiny) area to the right of the upper critical value under the same alternative distribution. See, e.g., MacKinlay, *supra* note 1, at 28 (presenting the power calculation formula).

TABLE II: POWER UNDER ALTERNATIVE ASSUMPTIONS FOR
STANDARD DEVIATION AND MEAN PRICE IMPACT
(5% SIGNIFICANCE LEVEL, NORMAL DISTRIBUTION, SAMPLE SIZE = 1)

Std. Dev. (%)	Mean Price Impact (%)						
	-1.0	-2.0	-3.0	-4.0	-5.0	-7.5	-10.0
4.0	5.7%	7.9%	11.7%	17.0%	24.0%	46.6%	70.5%
3.0	6.3%	10.2%	17.0%	26.6%	38.5%	70.5%	91.5%
2.0	7.9%	17.0%	32.3%	51.6%	70.5%	96.3%	99.9%
1.5	10.2%	26.6%	51.6%	76.0%	91.5%	99.9%	100.0%
1.0	17.0%	51.6%	85.1%	97.9%	99.9%	100.0%	100.0%

Table II demonstrates how power is a function of effect size and standard deviation, holding significance level constant, given a normal distribution assumption for abnormal returns. The probability of Type II error is one minus power. So, for example, when standard deviation is 1.5% and the true price impact is -2.0%, power is only about 26.6%, thus making the probability of Type II error about 73.4%, as in Figure 3.

C. Statistical Significance and Likelihood Ratios

Consider Figure 3 again. One reason to be concerned that the price impact of -2.0% will show up as statistically insignificant is that -2.0% is much more probable under the price-impact distribution than under the no-price-impact distribution. That is, the ratio of the height of the price-impact distribution to the height of the no-price-impact distribution, both evaluated at -2.0%, shows that it is more likely than not that there was a price impact. This ratio—a likelihood ratio—is a standard calculation in statistics and may be particularly useful for the kinds of problems that arise in litigation.³⁰

Table III presents likelihood ratio calculations for the price impact and standard deviation combinations presented in Table II. The number in each cell is the ratio of the probability of the price-impact distribution to the probability of the no-price-impact distribution at the posited price impact, where we assume normality of both distributions and where the mean of the price-impact distribution is the posited price impact and the mean of the no-price-impact distribution is zero. Asterisks denote that the observed

30. For an example of the intriguing general analysis of likelihood ratios in a legal context, see Louis Kaplow, *Likelihood Ratio Tests and Legal Decision Rules*, 16 AM. L. & ECON. REV. 1 (2014). For a more general look at likelihood ratios in a Bayesian context, see Robert E. Kass & Adrian E. Raftery, *Bayes Factors*, 90 J. AM. STAT. ASS'N 773 (1995).

return is at least 1.96 standard deviations from 0% (i.e., that the observed return is “statistically significant” at the two-tailed 5% level at that level of standard deviation).

TABLE III: LIKELIHOOD RATIOS AND STATISTICAL SIGNIFICANCE (*) FOR PRICE IMPACT VERSUS NO PRICE IMPACT UNDER ALTERNATIVE ASSUMPTIONS FOR STANDARD DEVIATION AND PRICE IMPACT (NORMAL DISTRIBUTION, SAMPLE SIZE = 1)

Std. Dev. (%)	Observed and Mean Price Impact (%)						
	-1.0	-2.0	-3.0	-4.0	-5.0	-7.5	-10.0
4.0	1.03	1.13	1.32	1.65	2.18	5.80	22.76*
3.0	1.06	1.25	1.65	2.43	4.01	22.76*	258.67*
2.0	1.13	1.65	3.08	7.39*	22.76*	1.13E+03*	2.68E+05*
1.5	1.25	2.43	7.39*	35.01*	258.67*	2.68E+05*	4.48E+09*
1.0	1.65	7.39*	90.02*	2.98E+03*	2.68E+05*	1.64E+12*	5.18E+21*

Table III demonstrates that an observed price impact is always more probable under a distribution centered at that price impact than under the no-price-impact distribution. This is the “maximum likelihood” estimate of the mean price impact that maximizes the likelihood at that point. It illustrates quite a different picture than significance levels provide. For instance, consider our example of a price impact of -2.0% when the standard deviation is 1.5%. Figure 3 demonstrates why the -2.0% return will fail a statistical significance test, which is shown by the absence of an asterisk for that entry. Figure 3 and Table II show, however, that the test statistic that failed to show statistical significance had only about 26.6% power to find the price impact centered at -2.0%. Table III illustrates a concern with this result: -2.0% is 2.43 times more likely under a price-impact distribution centered at -2.0% than under the no-price-impact distribution centered at 0%. Notwithstanding statistical insignificance, it is far more likely than not that there was a price impact centered at -2.0% than that there was no price impact.

Likelihood ratios are not, however, a panacea for the low power plaguing SFESs, because choosing a price impact based on likelihood ratios generates a higher probability of Type I error. For example, consider a -1.0% observed and mean price impact when the standard deviation is 1.5%. According to Table II, power is only 10.2%, so the probability of a Type II error is 89.8%. The likelihood ratio evaluated at -1.0% is 1.25, which slightly favors price impact over no price impact. But the probability of Type I error at -1.0% is now just over 50% (which is the

cumulative probability under the no price impact distribution up to -1.0% with standard deviation of 1.5% and assumed normality, times two (to reflect both tails)). So while the likelihood ratio recognizes that price impact may be more probable than no price impact even when there is statistical insignificance, a decision rule that replaced statistical significance with a likelihood ratio test based on picking the most likely state—price impact or no price impact—comes at the cost of higher probability of a Type I error (i.e., a greater chance of a false positive when testing for price impact). Nevertheless, the likelihood ratio—properly defined to balance Type I and Type II error—may have a better connection to evidentiary burdens than simple reliance on Type I error as it is unconnected to such burdens and insensitive to Type II error.³¹

D. Low Power and Statistical Insignificance in Case Law

Many cases address statistical significance of event study results. Consider *Goldkrantz v. Griffin*,³² where plaintiff shareholders brought a securities class action against defendants for an alleged misrepresentation. When the defendant company cured the alleged misrepresentation by corrective disclosure in a 10-K filing, the stock price fell by 2.64%.³³ Defendants' expert submitted an event study where the critical return for statistical significance was -4.41%, so the price impact of -2.64% was rejected as statistically insignificant. The Plaintiffs submitted a report of an expert criticizing the SFES as "highly suspect," but the court was unconvinced:

Plaintiff's argument that adopting a different confidence interval [i.e., different lower and upper critical values] could make the 2.64% stock price change in March statistically significant is of course true. Plaintiff has, however, provided no explanation for why a 95% confidence interval [i.e., a size of 5%] is inappropriate, other than its failure to pick up the 2.64% price change. In particular,

31. For an example of one court questioning the connection between conventional levels of statistical significance and evidentiary burden, see *United States v. Hatfield*, 795 F. Supp. 2d 219, 234 (E.D.N.Y. 2011) (citations omitted) ("The Court recognizes that the 95% confidence interval is the threshold typically used by academic economists in their work. But the Court strongly questions whether its use is appropriate in a forfeiture hearing, where the Government's burden is by a preponderance of the evidence. Indeed, when the Government cross-examined [Defendant's expert], it raised similar concerns. Preponderance of the evidence does not anywhere near require 95% certainty, and [the Government's expert's] study should have made accommodations for this lower evidentiary burden.").

32. No. 97 Civ. 9075(DLC), 1999 WL 191540 (S.D.N.Y. Apr. 6, 1999).

33. *Id.* at *2.

plaintiff has made no case for application of a different confidence level that would pick up 2.64% as statistically significant, without also picking up other statistically insignificant residual returns. As a result, defendants have met their burden of showing that any decline in stock price between March 20, 1997 and December 9, 1997, is not attributable to the revelation of the alleged misrepresentation. Plaintiff's conclusory critique of defendants' statistical analysis is insufficient to create a material question of fact.³⁴

We should be concerned with this result. The critical value of -4.41% meant that any price impact less than that amount was rendered effectively non-actionable. But a price impact of -2.64% certainly seems *economically* significant. Assuming that -4.41% was 1.96 standard deviations from zero, the standard deviation was 2.25%. In that case, the power to detect a price impact of, say, -4.0% was only 42.8%, so that the probability of a Type I error was only 5% while the probability of a Type II error was 57.2%. The power to detect a price impact of -2.64%—the observed impact—was only 21.7%, with the same probability of Type I error of 5% but a probability of Type II error of 78.3%. The likelihood ratio in favor of a -2.64% price impact was 1.99. Thus, the court's decision exonerated the defendant by relying on a decision rule that was heavily stacked against the plaintiffs.

Even where defense experts have asserted that "a price decline would have to be greater than 6.3% in order to be statistically significant,"³⁵ courts have not stopped to consider whether it is sensible to limit the application of the securities fraud laws to such large price moves. While courts have examined statistical power in cases from other areas of law where proper statistical practice is important,³⁶ serious consideration of statistical power is absent from opinions addressing event studies.³⁷

34. *Id.* at *5.

35. *In re Motorola Sec. Litig.*, 505 F. Supp. 2d 501, 511 n.14 (N.D. Ill. 2007). According to data from the Center in Research on Security Prices ("CRSP") available on the Wharton Research Data Services ("WRDS"), Motorola had a market capitalization in excess of \$45 billion in January 2001, the month at issue in the case. Therefore, the defense expert's claim was that fraud with a price impact of less than \$2.8 billion was statistically insignificant.

36. *See, e.g.*, *NRDC v. EPA.*, 658 F.3d 200, 218 (2d Cir. 2011) (addressing statistical power in risk assessment of pesticide; "[m]oreover, while the Gledhill study may have had sufficient statistical power to detect a cholinesterase inhibition greater than 0, EPA did not explain whether the 9-person study (six dosed subjects, 3 placebo subjects) had sufficient power to determine with any level of precision the magnitude of the cholinesterase inhibition.").

37. *See, e.g.*, *Sanner v. Bd. of Trade*, No. 89 C 8467, 2001 WL 1155277, at *5-6 (N.D. Ill. Sept. 28, 2001) (court declines to find expert's methodology unreliable in light of challenges that the study had insufficient power).

Instead, these opinions reject price impacts as statistically insignificant even though it is likely that the event study methodology had no power to detect highly probable and economically meaningful price impacts.³⁸

Thus far, problems with power have so far shown up most noticeably in the use of SFESs to determine whether a market is “efficient” for purposes of invoking the fraud-on-the-market presumption of reliance. Judicial findings of market efficiency for fraud-on-the-market purposes have always been somewhat at odds with academic research. There is no generally accepted method in financial economics for determining whether a security trades in an efficient market, and the so-called *Cammer* factors that are important in the courts have no general acceptance—indeed, are not even used—in academic research for that purpose.³⁹ Nevertheless, under existing law, courts must decide when information transmission in markets is sufficient to allow the use of the fraud-on-the-market doctrine, and markets need not be efficient—as economists understand the term—for that doctrine to make sense.⁴⁰

38. See, e.g., *In re Credit Suisse First Bos. Corp. Analyst Sec. Litig.*, 250 F.R.D. 137, 143 (S.D.N.Y. 2008) (describing an event study that showed no statistically significant price increases from disclosures of three analyst reports, where abnormal returns were -0.59%, 3.6%, and 2%).

39. See *Cammer v. Bloom*, 711 F. Supp. 1264, 1286–87 (D.N.J. 1989) (identifying five factors as ones to be considered in ascertaining whether a given security trades in an efficient market: (1) the average weekly trading volume of the security, (2) the number of securities analysts following and reporting on the security, (3) the extent to which market makers traded in the security, (4) the issuer’s ability to file an SEC registration Form S-3, and (5) the demonstration of a cause and effect relationship between unexpected, material disclosures and changes in prices of the security). We made this point in an earlier paper as well. See Alon Brav & J.B. Heaton, *Market Indeterminacy*, 28 J. CORP. L. 517, 535 (2003) (“[T]here is no reliable way to distinguish ‘efficient’ and ‘inefficient’ markets in fraud-on-the-market cases.”). More than ten years later, we are still aware of no peer-reviewed study in the finance literature that uses the *Cammer* factors to test whether a security traded in an efficient market or not. See Mukesh Bajaj et al., *Assessing Market Efficiency for Reliance on the Fraud-On-The-Market Doctrine After Wal-Mart and Amgen*, in *THE LAW AND ECONOMICS OF CLASS ACTIONS* 161, 183–90 (James Langenfeld ed., 2014) (criticizing the *Cammer* factors); Brad M. Barber et al., *The Fraud-on-the-Market Theory and the Indicators of Common Stocks’ Efficiency*, 19 J. CORP. L. 285, 290 (1994) (demonstrating that several of the *Cammer* factors do not explain efficiency).

40. See Fisch, *supra* note 11, at 898 (“Although market efficiency is neither a necessary nor a sufficient condition to establish that misinformation has distorted prices, most courts have concluded that the threshold inquiry in *Basic* is satisfied by proof that the misrepresentations were publicly made and ‘that the stock traded in an efficient market.’”); Jonathan R. Macey et al., *Lessons from Financial Economics: Materiality, Reliance, and Extending the Reach of Basic v. Levinson*, 77 VA. L. REV. 1017, 1018 (1991) (“We suggest that the focus of the Supreme Court’s holding in *Basic* is misplaced: what determines whether investors were justified in relying on the integrity of the market price is not the efficiency of the relevant market but rather whether a misstatement distorted the price of the affected security.”). For a discussion of how market efficiency could be better understood in the context of the fraud-on-the-market theory, see Lucian A. Bebchuk & Allen Ferrell, *Rethinking Basic*, 69 BUS. LAW. 671 (2014).

In practice, this has meant, in part, that plaintiffs must show the existence of statistically significant price reactions to firm-specific news.⁴¹ From the start, this is an odd test, and one that seems to assume that prices are not efficient unless price reactions are large enough to be statistically significant. This reflects a misunderstanding of efficiency. Prices can react efficiently to information even though the price reactions themselves are not so unusual in size as to approach statistical significance.⁴² The question for efficiency is whether the price reaction accurately reflected the impact of the news on the value of the firm—whether that impact was .1, .5, 1, 1.96, 3, or more standard deviations of value change. Thus, while a statistically significant reaction to a firm-specific news event is evidence that information was reflected in the price (absent confounding effects), the converse is not true—the failure of the price to react so extremely as to be two standard deviations from average does not establish that the market is inefficient; it may mean only that the correctly-sized value impact that occurred was less than 1.96 standard deviations from the mean. While some courts have been sensitive to this distinction by rejecting bad arguments to the contrary,⁴³ other courts have remained inattentive to this fact, which has generated inaccurate findings in some securities cases. A troubling recent example is *In re American International Group, Inc.*,⁴⁴ a case where the court determined that defendants had rebutted the fraud-on-the-market presumption on certain dates because price moves were statistically significant only at the 10% level (size) but not the 5% level.⁴⁵

41. Defendants occasionally argue that market efficiency requires that any statistically significant price move be linked to firm specific news. This is wrong. Financial economists have long known that price changes are only partly explainable by identifiable news events, and courts have rejected the argument. *See, e.g., In re DVI Inc. Sec. Litig.*, 249 F.R.D. 196, 211–12 (E.D. Pa. 2008) (rejecting defendant’s criticism that event study methodology showing market efficiency was flawed because 40% of statistically significant price moves could not be matched with identifiable news events).

42. *See, e.g., supra* note 35.

43. In one reported case, counsel ignored this rudimentary fact, but the judge rejected their argument. *See In re Nature’s Sunshine Product’s Inc.*, 251 F.R.D. 656, 664 (D. Utah 2008) (rejecting defendants’ argument that market for stock was not efficient because “of the 93 event days chosen by [plaintiffs’ expert], only 23 of those days (or less than 25%) result in a statistically significant change to Nature’s stock price,” and according to defendants, “in an efficient market material news should result in a statistically significant change in Nature’s stock price”).

44. 265 F.R.D. 157, 187 (S.D.N.Y. 2010), *vacated and remanded*, 689 F.3d 229 (2d Cir. 2012).

45. The court determined that the lack of a statistically significant price move at the 5% level on March 30 and 31, 2005, demonstrated that the fraud had not impacted the stock price. *Id.* at 185–87. The same court’s determinations regarding AIG bonds have been severely criticized as well. *See* Michael Hartzmark et al., *Fraud on the Market: Analysis of the Efficiency of the Corporate Bond Market*, 2011 COLUM. BUS. L. REV. 654, 654–55 (criticizing the AIG court, which “found insufficient empirical evidence to hold that the \$1.71 billion in AIG bonds, issued by the world’s largest insurance company, traded in open, developed, and efficient markets.”).

Similarly, when courts say things such as “[i]n an efficient market, stock prices should show statistically significant abnormal returns on days in which unexpected, material information is released into the market,”⁴⁶ they must recognize that they have implicitly *defined* materiality to mean a 1.96 standard deviation price move, so that, referring back to Table 1, a fraud of \$700 million is not “material” for a \$42 billion dollar company—a proposition with which many would disagree.

The Supreme Court has held that information is material if it “would have been viewed by the reasonable investor as having significantly altered the ‘total mix’ of information made available.”⁴⁷ Materiality “depends on the significance the reasonable investor would place on the withheld or misrepresented information,”⁴⁸ not on “statistical” significance.⁴⁹ In the model of informationally-responsive financial markets that underlies the fraud-on-the-market doctrine, materiality means “that there has been an effect on the market price.”⁵⁰ To be sure, the Supreme Court has recognized that there could be such a thing as “too low a standard of materiality,”⁵¹ but there is no reason to believe that the answer to that problem is to define the materiality threshold as 1.96 standard deviations from zero.

E. An Aside on Power and the MFES

It may be helpful to understand why low statistical power does not plague the MFES like it does the SFES. Suppose we have not one but sixteen firms that experienced the event we studied above. We can assume that each firm is otherwise the same and can be represented on an individual-firm basis to have the probability distributions given in Figure 3. When we combine the abnormal returns from the sixteen firms into a

46. *In re Fed. Home Loan Mortg. Corp. (Freddie Mac) Sec. Litig.*, 281 F.R.D. 174, 178 (S.D.N.Y. 2012).

47. *Basic Inc. v. Levinson*, 485 U.S. 224, 231–32 (1988) (quoting *TSC Industries, Inc. v. Northway, Inc.*, 426 U.S. 438, 449 (1976)).

48. *Id.* at 240.

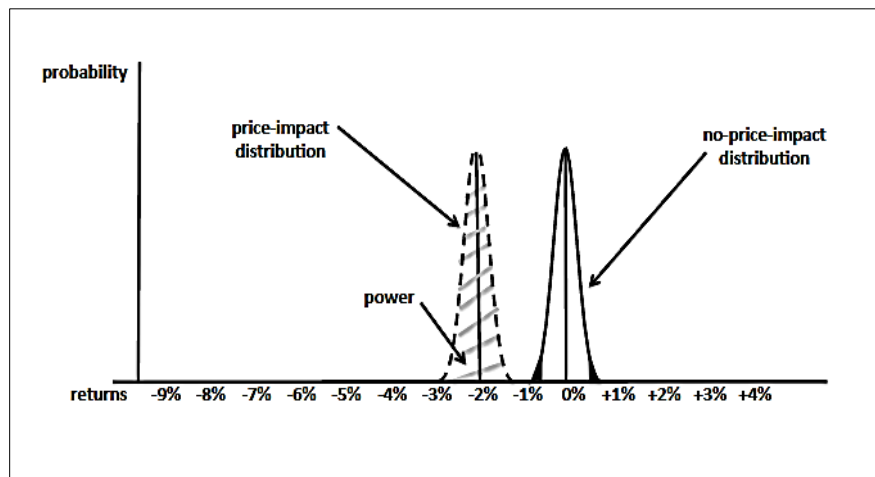
49. See David Tabak & Frederick Dunbar, *Materiality and Magnitude: Event Studies in the Courtroom* 16–17 (Nat’l Econ. Research Assocs., Working Paper No. 34, 1999), available at <http://www.nera.com/content/dam/nera/publications/archive1/3841.pdf> (“It is not clear what level of statistical significance corresponds to a legal definition of materiality. As Mitchell and Netter [*supra* note 1] point out, the 95% confidence level is commonly used, while the 90% and 99% levels are also options.”).

50. Daniel R. Fischel, *Use of Modern Finance Theory in Securities Fraud Cases Involving Actively Traded Securities*, 38 BUS. LAW. 1, 11 (1982).

51. *Basic*, 485 U.S. at 231.

sample average, the result is a very different set of distributions for the average, as shown in Figure 4.

FIGURE 4: POWER WITH 16 FIRMS



The distributions for the sample average abnormal return are now much tighter, because the standard deviation of a sample mean's distribution (as opposed to the standard deviation of a distribution generating a single abnormal return) falls as the number of observations reflected in the sample mean increases. Concentrating on the average abnormal return of 16 sample firms raises the "signal-to-noise ratio of the measured security price reaction to the studied event."⁵² This greater precision (i.e., the lower standard deviation) shows up in excellent statistical power. With the same size test as before (5% chance of Type I error), our statistical power is now in excess of 99.9%. We are virtually certain to detect the existence of the price impact if it exists.

The problem is that we are unable to apply the multi-firm approach in securities litigation. While increasing the number of firms would increase power to detect the effect of events, the event at issue in securities litigation typically is unique to the firm in the litigation. It is unlikely that a court would be satisfied with an analysis that attempted to incorporate ripple effects of the event in the prices of related firms, or with an analysis that attempted to identify 'similar' instances of misrepresentation by other

52. KLIGER & GUREVICH, *supra* note 1, at 32.

defendants in other litigation, because the incorporated events would not be probative enough of the price impact of the event at issue on the defendant firm in the securities litigation at hand.

II. DIFFICULTY #2: CONFOUNDING EFFECTS

A. *The Problem*

Causes of price impacts unrelated to the event under study are “confounding effects.” Sometimes confounding effects are apparent, such as when, in an event study of dividend omissions, a firm simultaneously announces bad earnings, which makes it more difficult to determine how much of the observed price impact resulted from the dividend omission and how much from the negative earnings announcement.⁵³ But most stock movements on a given day are impossible to decompose into their constituent parts by individual causal factor, and the abnormal return of an event study is no exception.⁵⁴

Clearly, security returns react to causes other than the announcement under consideration. To assess the effect of the announcement itself . . . it would be useful to identify those causes and to remove their effects from the data. It is unlikely, however, that *all* relevant factors can be identified or, even if they could, that their cumulative effect on each observed return could be measured.⁵⁵

In an MFES, this is of less concern because—as Figure 4 illustrates—averaging numerous abnormal returns across firms tends to average away the effects of unrelated price movements (or, more precisely, average them to their mean of zero), which leaves a better estimate of the mean price impact of the event under study. The matter is much more difficult in an SFES because there is no averaging away of confounding effects. When we observe a single abnormal return of, say, -4.0%, we could have observed the actual price impact without contamination by confounding effects. But we also could have observed a combination of an event price impact of, say, -2.5% and an unrelated price impact of -1.5%. Or we could

53. See Roni Michaely et al., *Price Reactions to Dividend Initiations and Omissions: Overreaction or Drift?*, 50 J. FIN. 573, 585 (1995) (investigating the influence of concurrent earnings announcements on event study results of dividend omissions and noting that “[c]learly, there is some incremental (negative) information content in the earnings release”).

54. See generally Roll, *supra* note 10.

55. Marais & Schipper, *supra* note 1, at 45.6.

have observed a combination of an event price impact of -5.0% and an unrelated price impact of +1.0% in the other direction. In general, the abnormal return we observe is comprised of a component related to the event plus a component related to non-event movements, but we have no mathematically precise way to separate them from each other.⁵⁶

B. Confounding Effects in Case Law

Courts are well aware that confounding effects are “[a] recurring problem in event studies” in securities litigation.⁵⁷ For example, a key issue in securities fraud cases is loss causation—demonstrating that the defendant’s misrepresentation caused the losses that plaintiff seeks to recover.⁵⁸ The event study plays a prominent role in addressing loss causation. As one court summarized:

Sorting out which declines were caused by such extraneous factors and which were caused by [disclosure of the fraud] is generally the province of an expert. It is an expert that produces the almost obligatory ‘event study’ that begins by isolating stock declines associated with market-wide and industry-wide downturns from those specific to the company itself. Once plaintiff’s expert has isolated days where the stock declines were statistically significant relative to these downturns, he must consider firm-specific events that might have caused those declines.⁵⁹

Failing to distinguish the fraud versus non-fraud causes of a price decline is generally fatal, especially where there are obvious confounding events occurring on the same day.⁶⁰ As one group of commentators has observed, many cases involve defense expert witnesses

56. In addition, the abnormal return will also contain a component related to the fact that there is measurement error in the estimation of the parameters of the model generating the expected return. *See* MacKinlay, *supra* note 1, at 21 (observing that the variance of abnormal returns contains components due to the sampling error in measuring predicted returns).

57. *Bricklayers & Trowel Trades Int’l Pension Fund v. Credit Suisse Secs. LLC*, 752 F.3d 82, 89 (1st Cir. 2014).

58. *See generally* *Dura Pharms., Inc. v. Broudo*, 544 U.S. 336 (2005).

59. *In re Vivendi Universal, S.A. Sec. Litig.*, 605 F. Supp. 2d 586, 599 (S.D.N.Y. 2009) (citations omitted).

60. *See, e.g.*, *Oscar Private Equity Invs. v. Allegiance Telecom, Inc.*, 487 F.3d 261, 271 (5th Cir. 2007) (vacating class certification order and noting that “[t]he plaintiffs’ expert does detail event studies supporting a finding that Allegiance’s stock reacted to the *entire bundle* of negative information contained in the 4Q01 announcement, but this reaction suggests only market efficiency, not loss causation, for there is no evidence linking the *culpable* disclosure to the stock-price movement”), *abrogated by* *Erica P. John Fund, Inc. v. Halliburton Co.*, 131 S. Ct. 2179 (2011); *In re*

parsing every announcement by a company to attempt to isolate factors that negatively impacted the stock price, but which were not related to the alleged fraud. Not surprisingly, plaintiffs' experts take a different approach, generally defining the fraud as broadly as possible so that all of a company's announcements that harmed the stock price are deemed to be fraud-related. This "battle of the experts" is at the heart of many of today's securities cases.⁶¹

While the event study itself is a first step in loss causation analyses, the expert must then work to decompose price effects outside the framework of the event study methodology.⁶² Because that effort is necessarily partly subjective and less constrained by generally accepted quantitative methods, it has proven unsatisfactory to some courts. These courts have complained that such efforts are based on "unprovable and often unexplained assumptions,"⁶³ that they are no "more than observations and market rumors,"⁶⁴ and that they reflect little more than "a judgment call as to confounding information without any methodological underpinning."⁶⁵ Courts have recognized the possibility of studying intraday returns when confounding events occur sequentially instead of overlapping, and other means of addressing the confounding effects problem have been suggested as well.⁶⁶ At the end of the day, however, most stocks "are routinely

Omnicom Grp., Inc., 541 F. Supp. 2d 546, 554 (S.D.N.Y. 2008) (rejecting expert's event study which did not isolate effects of alleged corrective disclosures from other factors), *aff'd*, 597 F.3d 501 (2d Cir. 2010); Fener v. Belo Corp., 560 F. Supp. 2d 502, 508 (N.D. Tex. 2008) (citation omitted) ("[Expert's] event study tends to establish that the market reacted to the bundle of August 5 news pieces with an August 6 stock price drop of 5.47% [but] . . . fails to target the corrective disclosure at issue."), *aff'd*, 579 F.3d 401 (5th Cir. 2009).

61. Jay W. Eisenhofer et al., *Securities Fraud, Stock Price Valuation, and Loss Causation: Toward a Corporate Finance-Based Theory of Loss Causation*, 59 BUS. LAW. 1419, 1428 (2004).

62. See, e.g., *Vivendi*, 605 F. Supp. 2d at 591–93 (holding that expert's date-by-date subjective attribution of price moves to fraud and non-fraud causes created an issue of fact on loss causation sufficient to withstand summary judgment); *In re Pfizer Inc. Sec. Litig.*, 936 F. Supp. 2d 252, 266 (S.D.N.Y. 2013) (stating that after running an event study to identify statistically significant events, the expert "cited market commentary confirming that it was those events—not market forces or non-fraudulent Pfizer-specific news—that caused the price declines").

63. *United States v. Olis*, Crim. No. H-03-217-01, 2006 WL 2716048, at *8 (S.D. Tex. Sept. 22, 2006).

64. *Hershey v. Pac. Inv. Mgmt. Co. LLC*, 697 F. Supp. 2d 945, 955 (N.D. Ill. 2010).

65. *Bricklayers & Trowel Trades Int'l Pension Fund v. Credit Suisse Secs. LLC*, 752 F.3d 82, 95 (1st Cir. 2014).

66. See *Bricklayers & Trowel Trades Int'l Pension Fund v. Credit Suisse First Bos.*, 853 F. Supp. 2d 181, 190 (D. Mass. 2012) (citing Laura Starks, *Discussion of Market Microstructure: An Examination of the Effects on Intraday Event Studies*, 10 CONTEMP. ACCT. RES. 355, 383–86 (1994)); *In re Novatel Wireless Sec. Litig.*, 910 F. Supp. 2d 1209, 1218–21 (S.D. Cal. 2012) (examining objections to expert's intraday causation analysis in detail and holding that the testimony was admissible); DAVID TABAK, MAKING ASSESSMENTS ABOUT MATERIALITY LESS SUBJECTIVE

affected by events that may not be pinpointed by inspecting the media sources: generally, stock prices continuously change, while only rarely these changes may be linked to specific information releases.”⁶⁷ As a result, courts may have excessively high expectations of the ability of litigants—whether plaintiff or defendant—to decompose an observed return into a component caused by fraud and a component caused by other factors.⁶⁸ There simply is no fully reliable, mathematically precise way to do so. Financial economists have long understood that our ability to fully explain observed price moves is quite limited;⁶⁹ much price movement probably occurs for reasons unrelated to news, including the liquidity trades of investors in the market seeking to raise funds for other purposes and the trades of noise traders who trade for irrational reasons.⁷⁰

III. DIFFICULTY #3: BIAS

When low statistical power meets confounding effects, the result is biased estimates of detected price impacts. When statistical power is low, small price impacts may get detected only *because* of confounding effects. Consider Figure 3 again. The true price impact was -2.0%, but the SFES had no power to detect that impact on its own; -2.0% is between the critical values and will show up as statistically insignificant if it occurs as a “clean” price move. The SFES will detect the price impact as statistically significant only when there are enough confounding effects to push the overall price impact past -2.94% (which is -1.96 multiplied by the standard deviation of 1.5%). When we detect a price impact in such situations, it is as if we are sampling from a truncated distribution, as shown below in

THROUGH THE USE OF CONTENT ANALYSIS (2007), available at http://www.nera.com/extImage/PUB_Tabak_Content_Analysis_SEC1646-FINAL.pdf); Esther Bruegger & Frederick C. Dunbar, *Estimating Financial Fraud Damages With Response Coefficients*, 35 J. CORP. L. 11, 13 (2009).

67. KLIGER & GUREVICH, *supra* note 1, at 32.

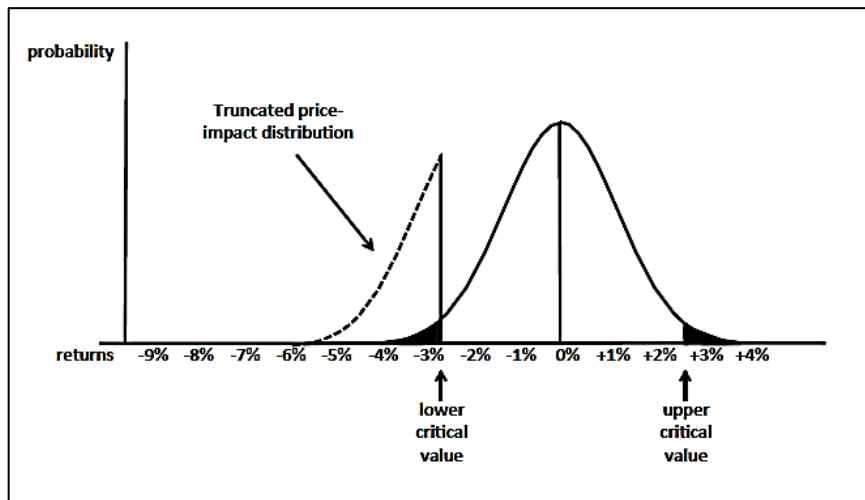
68. *See, e.g., In re Moody's Corp. Sec. Litig.*, No. 07 Civ. 8375(GBD), 2013 WL 4516788, at *12 (S.D.N.Y. Aug. 23, 2013) (noting that Plaintiffs provided insufficient evidence of loss causation where “neither [plaintiffs’ expert’s] report nor any other evidence proffered by Plaintiffs establish that market forces and other factors unrelated to Moody’s alleged mismanagement of its conflicts of interest did not play a significant role in Plaintiffs’ economic loss”).

69. *See, e.g.,* Roll, *supra* note 10, at 541 (“Even with hindsight, the ability to explain stock price changes is modest.”).

70. The idea that prices may move for reasons unrelated to news is well understood in the financial economics literature. *See, e.g.,* Fischer Black, *Noise*, 41 J. FIN. 529 (1986); J. Bradford De Long et al., *Noise Trader Risk in Financial Markets*, 98 J. POL. ECON. 703 (1990); Lawrence R. Glosten & Paul R. Milgrom, *Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders*, 14 J. FIN. ECON. 71 (1985); Albert S. Kyle, *Continuous Auctions and Insider Trading*, 53 ECONOMETRICA 1315 (1985).

Figure 5. Unfortunately, the expected value of that distribution is not the true price impact, which was cut off by our critical value.

FIGURE 5: BIAS



As power improves—in other words, as the alternative distribution gets further away from the null distribution (i.e., the effect size we are looking for increases)—the bias goes to zero, as shown below in Figures 6 and 7.

FIGURE 6: BIAS

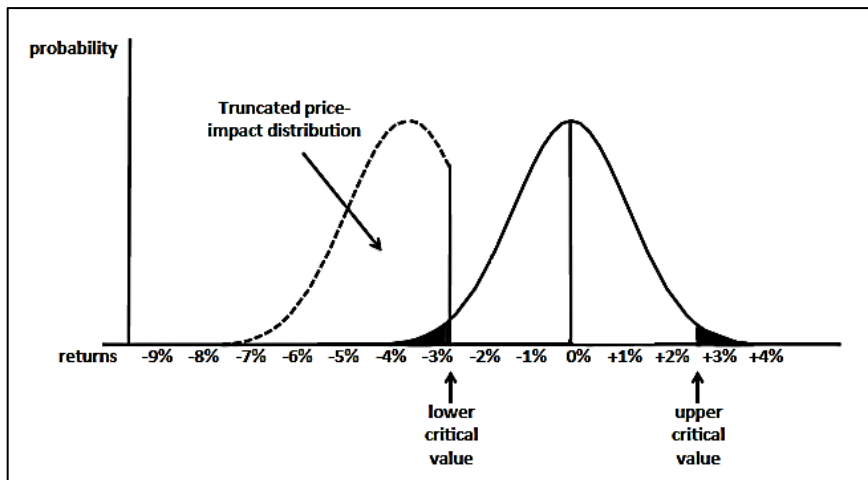
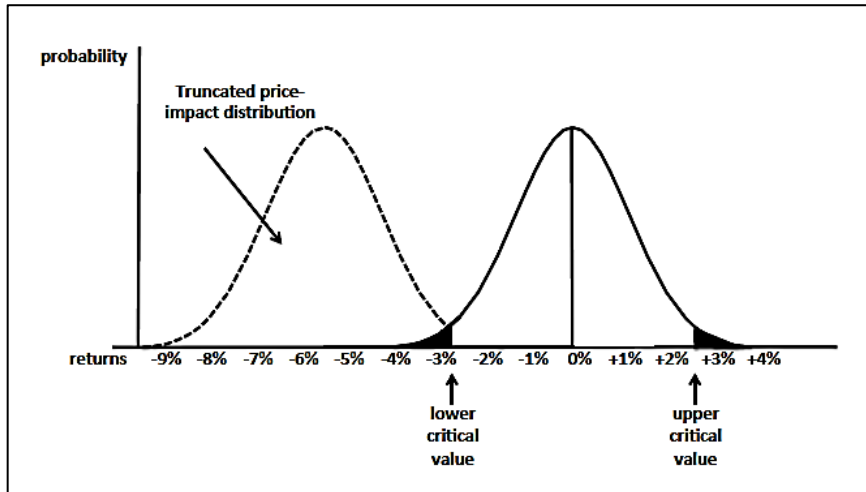


FIGURE 7: BIAS



This bias problem is well known in scientific contexts. For example, in a leading neuroscience journal, authors described how the problem plagues neuroscience studies:

[E]ven when an underpowered study discovers a true effect, it is likely that the estimate of the magnitude of that effect provided by that study will be exaggerated. This effect inflation is often referred to as the ‘winner’s curse’ and is likely to occur whenever claims of discovery are based on thresholds of statistical significance (for example, $p < 0.05$) or other selection filters (for example, a Bayes factor better than a given value or a false-discovery rate below a given value). Effect inflation is worst for small, low-powered studies, which can only detect effects that happen to be large. If, for example, the true effect is medium-sized, only those small studies that, by chance, overestimate the magnitude of the effect will pass the threshold for discovery.⁷¹

71. Katherine S. Button et al., *Power Failure: Why Sample Size Undermines the Reliability of Neuroscience*, 14 NATURE REVIEWS: NEUROSCIENCE 365, 366 (2013) (footnote omitted); see also John P. A. Ioannidis, *Why Most Discovered True Associations Are Inflated*, 19 EPIDEMIOLOGY 640 (2008); Satish Iyengar & Joel B. Greenhouse, *Selection Models and the File Drawer Problem*, 3 STAT. SCI. 109 (1988); Robert Rosenthal, *The “File Drawer Problem” and Tolerance for Null Results*, 86 PSYCH. BULL. 638 (1979).

But the bias problem seems to have gone unnoticed in the SFES context, despite its likely importance given the combination of low power and confounding effects. In Table IV, we present the expected value of truncated distributions.⁷² In each case, the cutoff point is the point that is -1.96 standard deviations from zero under the no-price-impact distribution.

TABLE IV: EXPECTED VALUE OF TRUNCATED NORMAL DISTRIBUTIONS (TRUNCATION FROM ABOVE AT $-1.96 \times \text{STD. DEV.}$ FROM ZERO)

Std. Dev. (%)	Mean Price Impact (%)						
	-1.0	-2.0	-3.0	-4.0	-5.0	-7.5	-10.0
4.0	-9.5%	-9.6%	-9.8%	-10.0%	-10.2%	-10.9%	-12.0%
3.0	-7.1%	-7.3%	-7.5%	-7.7%	-8.0%	-9.0%	-10.5%
2.0	-4.8%	-5.0%	-5.2%	-5.5%	-6.0%	-7.7%	-10.0%
1.5	-3.6%	-3.9%	-4.2%	-4.6%	-5.3%	-7.5%	-10.0%
1.0	-2.5%	-2.8%	-3.3%	-4.1%	-5.0%	-7.5%	-10.0%

Continuing our example, consider again Figure 5, where the standard deviation is 1.5% and the alternative distribution was centered at a price impact of -2.0%. The lower critical value of -2.94% is the upper cutoff of the alternative distribution. If the true price-impact distribution is centered at -2.0%, we will observe statistically significant results only when we get a return from the tail of the price-impact distribution to the left of -2.94%. The expected value (average draw) from that truncated distribution is -3.9%. That is, while low power (recall that power is 26.6% in this scenario) means we are unlikely to get a statistically significant result, the problem is that, when we do, the expected value of the statistically significant return we do see is -3.9%, which is significantly biased from -2.0%.

Note that as power improves in the lower right-hand corner of Table IV, the expected value converges on its true value. So, for example, when the true price impact is -10% and the standard deviation is 1.5%, the expected value is -10.0%. The large negative values in the upper left-hand corner may appear anomalous, but they are not. Consider the case when the standard deviation is 3.0% and the true price impact is -1.0%. In that case, it is unlikely that we will observe a statistically significant result;

72. For the formulas to calculate the expected values, *see, e.g.*, WILLIAM H. GREENE, *ECONOMETRIC ANALYSIS* 836 (7th ed. 2012) (equation (19-3b)); Donald R. Barr & E. Todd Sherrill, *Mean and Variance of Truncated Normal Distributions*, 53 *AM. STATISTICIAN* 357, 359 (1999) (equation (8)).

power is only 6.3% in Table II. But if we do get statistical significance, it is because we have observed a return that is at least -5.88% ($3.0\% \times -1.96$). The average of the values from minus infinity to -5.88% under the truncated probability distribution is indeed -7.1%.

These results demonstrate that we cannot leave confounding effects unaddressed in the hope that they are as likely to be on one side of the true price effect as on the other. To the contrary, when power is low and confounding effects are likely, the bias in price impacts—and therefore in damages—is upward (i.e., in the direction of more severe price impacts) and not downward.⁷³

CONCLUSION

The multi-firm event study was a breakthrough research method in financial economics,⁷⁴ but the single-firm event study differs from the multi-firm event study in a fundamental sense. By examining only one firm event at a time, the single-firm event study runs into trouble with low statistical power, confounding effects, and bias. No economic methodology is perfect, of course, and the questions that securities litigation presents need answers. Still, a review of the case law suggests that courts and litigants have often failed to recognize the special difficulties that the SFES presents, especially the problems that result from low power, and accuracy of securities litigation adjudication has suffered in some cases as a result.

We offer the following suggestions for improving the use of event study evidence in securities litigation. First, Courts should require litigants and their expert witnesses to report the results of a power analysis for all event studies. Financial economists using event studies in their academic work take power seriously. As one prominent reviewer wrote of power in this context, “An important consideration when setting up an event study is the ability to detect the presence of a non-zero abnormal return. The inability to distinguish between the null hypothesis and economically

73. One source of upward confounding effects at the time of a corrective disclosure is the effect of likely litigation. See Janet Cooper Alexander, *The Value of Bad News in Securities Class Actions*, 41 UCLA L. REV. 1421 (1994).

74. As Nobel Laureate Eugene Fama wrote in 1991: “Event studies are now an important part of finance, especially corporate finance. In 1970 there was little evidence on the central issues of corporate finance. Now we are overwhelmed with results, mostly from event studies.” Eugene F. Fama, *Efficient Capital Markets: II*, 46 J. FIN. 1575, 1600 (1991). In 1997, Shleifer and Vishny wrote that event study analysis “has surely become the most common empirical methodology of corporate governance and finance.” Andrei Shleifer & Robert W. Vishny, *A Survey of Corporate Governance*, 52 J. FIN. 737, 746 (1997) (citation omitted).

interesting alternatives would suggest the need for modification of the design.”⁷⁵ That same level of rigor should apply in litigation.⁷⁶ A power analysis will tell the court whether the litigant’s event study was reliable for detecting price impacts of various sizes. Securities litigants should not be allowed to say that a misrepresentation or corrective disclosure caused no price impact based on a test that had little or no power to detect a price impact that the court determines to be material, and one could imagine a court excluding any such evidence under Federal Rule of Evidence 403 as misleading. Likelihood ratios may shed further light on event study results, though it is important to report probabilities of Type I and Type II error in doing so.

Second, to address the problem of confounding effects, courts should allow litigants flexibility to present other evidence to prove that a price impact from misrepresentation or corrective disclosure did or did not occur. Such evidence may include, as it does now, intraday analyses and analyses of other news about the firm that day and quantitative analysis of its potential effect on value, but also evidence from multi-firm event studies of similar events that may shed light on the magnitude of price impacts of that type, estimates of the effects of liquidity and noise traders, and opinions based on other valuation methods such as discounted cash flow to value posited confounding effects.

Third, Courts and litigants should recognize the potential effect of low statistical power and confounding effects on detected price impacts, considering the possibility that statistically significant price impacts are biased estimates of true price impacts. Because of bias, detected price impacts are more likely to overestimate price impact than underestimate it. It is especially important to consider this possibility in evaluating securities damages. Litigants should be permitted to present evidence of the upward bias phenomenon alongside their specific arguments regarding confounding factors, even if their evidence on confounding factors is insufficient to prove a complete lack of liability.

Finally, our analysis suggests a real need to more carefully consider tradeoffs between Type I and Type II error in this litigation setting. Currently, the use of statistical significance without attention to power implements a legal regime where the probability of incorrectly exonerating securities defendants is much higher than the probability of incorrectly

75. MacKinlay, *supra* note 1, at 28.

76. *See generally* Kumho Tire Co. v. Carmichael, 526 U.S. 137, 152 (1999) (stating that an expert should employ in court “the same level of intellectual rigor that characterizes the practice of an expert in the relevant field”).

finding securities defendants liable. As shown in Table I, the unspoken impact of the current situation is that courts are likely to ignore some non-negligible frauds if they are below the minimum detectable price impact in a low power SFES. This lowers the deterrent effect of the securities laws and may encourage more small- and mid-scale fraud on markets than is socially optimal given the costs of litigation. How that tradeoff should occur is beyond the scope of our Article, but addressing that important problem presents a challenge for future work.