POWER AND STATISTICAL SIGNIFICANCE IN SECURITIES FRAUD LITIGATION

JILL E. FISCH* JONAH B. GELBACH**

Event studies, a half-century-old approach to measuring the effect of events on stock prices, are now ubiquitous in securities fraud litigation. In determining whether the event study demonstrates a price effect, expert witnesses typically base their conclusion on whether the results are statistically significant at the 95% confidence level, a threshold that is drawn from the academic literature. As a positive matter, this represents a disconnect with legal standards of proof. As a normative matter, it may reduce enforcement of fraud claims because litigation event studies typically involve quite low statistical power even for large-scale frauds.

This Article, written for legal academics, judges, and policy makers, makes three contributions. First, it contributes to a nascent literature demonstrating that the standard event-study methodology can be problematic in securities litigation. In particular, the Article documents the tradeoff between power and confidence level and the ensuing impact on the likelihood that valid claims of fraud will erroneously be rejected. In so doing, the Article highlights that the choice of confidence level is a policy judgment about the appropriate balance between the costs of litigation and the costs of securities fraud. Second, the Article argues that the Securities and Exchange Commission (SEC) has both the legal power and the institutional competence to develop litigation standards that balance these costs.

Third, the Article provides a novel and feasible framework through which the SEC can implement such litigation standards. The framework relies on an assessment of the defendant firm's market capitalization and abnormal returns distribution to determine the maximum confidence level (minimum significance level) that is consistent with the minimum required power of detecting a fraud of the benchmark magnitude. The SEC is uniquely positioned to make this determination based on the information it possesses about the level of fraud in the capital markets and the role of private litigation in deterring fraud.

* Saul A. Fox Distinguished Professor of Business Law, University of Pennsylvania Law School.
** Professor of Law, University of California at Berkeley Law School.

Introduction

Event studies, a statistical tool borrowed from the financial economics literature, are "standard operating procedure in federal securities litigation."[1] The purpose of an event study is to measure the extent to which stock prices react to the release of new information into the market. In securities fraud cases, event studies are used in several ways, including analyzing the efficiency of the market in which the securities trade,[2] measuring the price impact of the fraudulent disclosures,[3] determining whether there is a causal relationship between the fraud and the plaintiffs' economic losses,[4] and computing the amount of damages.[5] Although courts vary in the extent to which they require the use of an event study and the degree to which they accept

[1] *In re* Vivendi, S.A. Sec. Litig., 838 F.3d 223, 253–54 (2d Cir. 2016) (quoting United States v. Gushlak, 728 F.3d 184, 201 (2d Cir. 2013)).

[2] *See, e.g.*, Di Donato v. Insys Therapeutics, Inc., 333 F.R.D. 427, 439–40 (D. Ariz. 2019) (using event studies to evaluate market efficiency).

[3] *See, e.g.*, Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 263 (N.D. Tex. 2015) (using event studies to determine price impact).

[4] *See* Bricklayers & Trowel Trades Int'l Pension Fund v. Credit Suisse Sec. (USA) 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.").

[5] *See, e.g.*, *In re* Oracle Sec. Litig., 829 F. Supp. 1176, 1181 (N.D. Cal. 1993) (discussing the role of event studies in calculating damages).

other evidence with respect to these issues, a properly conducted event study is often a critical factor.[6]

The structure of an event study involves using a process known as hypothesis testing to distinguish between normal fluctuations in stock price and a so-called abnormal return associated with the release of material information about the company. An event study seeks to determine whether to reject the so-called null hypothesis—that the price movements in question fall within some measure of normal limits. It concludes that a stock price movement is likely to have been caused by a disclosure if the size of the stock price reaction is sufficiently outside the normal or expected range of stock price fluctuations, a price movement that we have termed, in other work, "highly unusual."[7]

In the academic literature from which the event study methodology is drawn, the basis for inferring a causal relationship is most commonly a confidence level of 95%. A finding that the stock price movement exceeded the magnitude of 95% of movements that occur only by chance is termed statistically significant. This requirement of a 95% confidence level has been imported into the law.[8] Courts have repeatedly held that a "properly conducted" event study demonstrating a statistically significant price effect is necessary to establish or rebut key elements of a securities fraud claim.[9]

---

[6] *See, e.g.*, Goldkrantz v. Griffin, 97 CIV. 9075 (DLC), 1999 U.S. Dist. LEXIS 4445 at *11–13, 24 (S.D.N.Y. Apr. 5, 1999) (granting defendants summary judgment on negative loss causation affirmative defense where expert examined statistically significant residual returns identified by "'standard' 95% confidence interval," used index of peer stock firms, and assumed market efficiency; and plaintiff's expert "conducted no independent statistical analysis" of defendant's stock).

[7] Jill Fisch, Jonah Gelbach & Jonathan Klick, *The Logic and Limits of Event Studies in Securities Fraud Litigation,* 96 Texas L. Rev. 553 (2018). *See also* Michael J. Kaufman & John Wunderlich, *Regressing: The Troubling Dispositive Role of Event Studies in Securities Fraud Litigation*, 15 Stan. J. L. Bus. & Fin. 183, 193–94 (2009) (The "excess price movement is tested for statistical significance to see whether the result is unusual or unlikely to be explained by the normal random variations of the stock price") (quoting David Tabak, *Making Assessments About Materiality Less Subjective Through the Use of Content Analysis* (NERA Econ. Consulting Working Paper, 2007), https://www.nera.com/content/dam/nera/publications/archive1/PUB_Tabak_Content_Analysis_SEC1646-FINAL.pdf). A different methodological approach from conventional hypothesis testing is Bayesian hypothesis testing. To our knowledge this approach has not been used in securities litigation; *see* Jonah B. Gelbach & Jenny R. Hawkins, *A Bayesian Approach to Event Studies for Securities Litigation*, 176 J. Institutional & Theoretical Econ. 86, 86 (2020).

[8] *See, e.g., Halliburton Co.,* 309 F.R.D. at 262 ("To show that a corrective disclosure had a negative impact on a company's share price, courts generally require a party's expert to testify based on an event study that meets the 95% confidence standard, which means 'one can reject with 95% confidence the null hypothesis that the corrective disclosure had no impact on price.'").

[9] *See, e.g.*, Carpenters Pension Tr. Fund of St. Louis v. Barclays PLC, 310 F.R.D. 69, 95 (S.D.N.Y. 2015) (describing the role of "a properly conducted event study" in establishing price impact); United States v. Hatfield, 06-CR-0550 (JS)(AKT), 2014 U.S. Dist. LEXIS 174947, at *18, 26 (E.D.N.Y. Dec. 18, 2014) (observing that "a properly conducted event study" was necessary to determine the amount of investors' losses in criminal securities fraud case).

Similarly, courts have rejected efforts to establish those elements that fail to establish a causal relationship at the 95% confidence level.[10]

While courts have embraced the event study methodology, they have paid limited attention to the question of whether the social science standard of statistical significance and the requirement of the 95% confidence level are appropriate standards for legal sufficiency. In this Article, we argue that they are not. We focus, in particular, on an issue that the courts have mostly overlooked, the relationship between confidence level and power. As we explain, the confidence level controls the probability of rejecting non-meritorious claims—cases in which the disclosure in question did not, in fact, cause a stock price reaction. Power addresses a complementary concern—erroneous rejection of meritorious claims. As we explain, the structure of an event study presents a tradeoff between confidence level and power. Specifically, the requirement that event studies establish a causal relationship at the 95% confidence level leads to low power in many situations. As a result, when courts require event studies that meet the standard of a 95% confidence level, there is a high likelihood that they will reject a substantial number of cases of true fraud.[11]

The tradeoff that we highlight between confidence level and power demonstrates that the choice of what confidence level to use in securities fraud litigation is not a matter of objective scientific truth. Instead, the choice reflects an implicit normative judgment about the appropriate level of difficulty required to establish a securities fraud claim. As we elaborate, the concept of statistical significance involves a fundamental policy choice between reducing the risk of imposing liability for a disclosure that did not affect stock prices and increasing the likelihood that defendants will be held accountable for fraudulent disclosures. In their use of the event study methodology and reliance on the social science literature, the courts have not confronted this policy choice directly.[12]

With the policy choice inherent in the use of statistical significance revealed, it becomes clear that securities litigation should not borrow unthinkingly from empirical practice in the social sciences but instead determine the appropriate threshold of statistical significance based on the tradeoff between these competing concerns. We therefore consider which part of our system of securities regulation is best suited to make this policy choice from the perspective of comparative institutional competence. Although Article III courts currently handle this decision implicitly by the standards they im-

---

[10] *See infra* notes 32–35 and accompanying text.

[11] *See, e.g.*, Taylor Dove, Davidson Heath & J.B. Heaton, *Bias-Corrected Estimation of Price Impact in Securities Litigation*, 21 Am. L. Econ. Rev. 184, 188 (2019) ("[A] consequence of the low statistical power of single-firm event studies is that many smaller (but still economically important) securities frauds will go undetected because they do not reach statistical significance.").

[12] *See* Jonah B. Gelbach, *Estimation Evidence*, 168 U. Pa. L. Rev. 549, 552 (2020) (discussing the disconnect between use of academic standards of statistical significance and stated litigation standards).

pose on the admissibility and persuasiveness of expert testimony, we argue that they are not well-situated to do so. Federal judges are poorly positioned to weigh the policy considerations reflected by the tradeoff between confidence level and power and to consider the impact that shifting the extent of the tradeoff will have on the deterrence of fraud and the promotion of market integrity. Although the policy analysis we describe could be conducted by Congress, we argue that Congress lacks both the expertise and the political will to make this determination and that monitoring the market's response is also likely to require a level of flexibility and adjustment that one-off legislation is poorly suited to provide.

As a result, we propose that the SEC decide the appropriate level of statistical significance to be used in securities litigation event studies. We argue that balancing litigation costs against fraud costs is precisely the type of determination that expert administrative agencies were designed to make. The SEC is well-suited to make this determination based both on the technical expertise of its staff of economists as well as its familiarity with the role of private enforcement in serving the objectives of the federal securities laws. The SEC also has unique access to data allowing it to evaluate the costs involved and the flexibility to adjust its rule in response to changes in market conditions or the behavior of market participants. We therefore call upon the SEC to engage in formal rulemaking to evaluate the applicable policy considerations and to set the level of statistical significance necessary for an acceptable event study in securities fraud litigation.

We conclude by offering a feasible framework that the SEC can use in determining the appropriate tradeoff between power and confidence level. This approach is based on ensuring that a minimum level of power is obtained for a benchmark fraud magnitude, based on the SEC's judgment about the level of enforcement necessary to provide sufficient deterrence of fraud. Given knowledge of the defendant firm's market capitalization and abnormal returns distribution, it is straightforward to determine the maximum confidence level (minimum significance level) that is consistent with the minimum required power of detecting a fraud of the benchmark magnitude.

The rest of this Article proceeds as follows. In Part I, we provide a brief overview of the current role of event studies in securities fraud litigation. In Part II, we explain in more detail the idea of statistical significance and its conceptual twin, the confidence level. We also introduce and describe the tradeoff between the confidence level used in an event study and the power of that study, a tradeoff that highlights the implicit policy judgment reflected in the use of the 95% confidence level by the courts.[13] Part III explains that, because of the policy considerations reflected in the choice of the 95% confidence level, its use by the courts reflects a normative judgment rather than merely some technical inquiry. We argue that the courts and Congress are poorly positioned to make this normative judgment. Instead, we propose that

[13] Equivalently, the tradeoff between Type I and Type II error rates.

this determination be made by the SEC through formal notice-and-comment rulemaking. Finally, Part IV suggests a framework for the SEC to choose the confidence level in a way that consciously takes into account the tradeoff between confidence level and power.

I. Event Studies and Securities Litigation: The State of Play

A. *Event Studies in the Courts*

In *Basic, Inc. v. Levinson*, the U.S. Supreme Court accepted the fraud on the market presumption that stock prices, in an efficient market, reflect material public disclosures.[14] As a result, *Basic* enabled private securities fraud claims to be brought as class actions as long as the plaintiffs could establish that the fraudulent disclosures affected stock price. In *Dura Pharmaceuticals, Inc. v. Broudo*, the Supreme Court cautioned, however, that it was legally insufficient for plaintiffs to prove that they had purchased stock at a price artificially inflated by fraud; plaintiffs were also required to establish a causal relationship between the fraud and their subsequent economic harm,[15] a requirement known as "loss causation."[16]

Litigants now introduce event studies to address the requirements of both *Basic* and *Dura*.[17] Event studies, which seek to measure the extent to which stock prices respond to disclosures, are used to support plaintiffs' claims, pursuant to *Basic*, that the securities in question traded in an efficient market. Similarly, event studies are used in court to prove or refute the claim that a particular fraudulent disclosure affected market price—a requirement that the Supreme Court subsequently termed "price impact."[18] In *Halliburton II*, the Supreme Court explicitly endorsed the ability of defendants to defeat class certification by introducing event studies demonstrating the absence of price impact.[19] Finally, to address the concern identified in *Dura*, litigants introduce event studies to address loss causation, most commonly by evaluating the effect on the market price of a disclosure revealing or

---

[14] Basic Inc. v. Levinson, 485 U.S. 224, 245–47 (1988).

[15] Dura Pharms., Inc. v. Broudo, 544 U.S. 336, 345–46 (2005).

[16] *See* Securities Exchange Act of 1934, 15 U.S.C. § 78u-4(b)(4) (2018) (requiring plaintiffs to establish loss causation).

[17] *See* Fisch et al., *supra* note 7, at 560–62 (describing role of event studies in addressing the requirements of *Basic* and *Dura*).

[18] *See* Halliburton Co. v. Erica P. John Fund, Inc., 573 U.S. 258, 277 (2014) [hereinafter *Halliburton II*] (defining price impact as proof "that a defendant's misrepresentation actually affected the stock").

[19] *Id.* at 280, 283. As this Article went to press, the Supreme Court agreed to hear a case concerning the legal standard applicable to such an effort by a defendant. *See generally* Goldman Sachs Grp., Inc. v. Ark. Teacher Ret. Sys., No. 20-222, 2020 U.S. LEXIS 5993 (Dec. 11, 2020). For a useful discussion of *Halliburton II*'s technical requirements, see generally Merritt B. Fox, *Halliburton II: It All Depends on What Defendants Need to Show to Establish No Impact on Price*, 70 Bus. Law. 437 (2015).

correcting a prior fraudulent statement.[20] Although securities fraud cases rarely go to trial and, as a result, judicial efforts to calculate damages are virtually non-existent, litigants also proffer event studies with respect to damages on motions for summary judgment[21] as well as at the motion for class certification in response to Rule 23's requirement that damages can be calculated on a class-wide basis.[22] Moreover, plaintiffs' decisions whether to litigate fraud claims are made in the shadow of prevailing judicial standards; thus, they are unlikely even to file a complaint unless they can support their claims with an event study likely to pass muster under prevailing standards.

Courts have varied widely in their evaluation and use of the event study methodology.[23] Within this variation, however, courts generally recognize the need to determine whether an event study constitutes reliable evidence with respect to the question for which it is introduced, that is, whether the event study supports (or refutes) the claimed relationship between a disclosure and a stock price movement.[24] Central to that task is the concept of statistical significance.[25]

In the social science literature from which the event study methodology is drawn, the results of hypothesis testing depend on the likelihood that an outcome as extreme as the one observed would occur by chance, given that no event having a causal effect on stock price had actually occurred. Mapping that into the event study context, the question is how probable it would be to observe as large a price drop if the stock price were actually not affected by an alleged corrective disclosure.[26] Commentators have argued that

---

[20] *See, e.g.*, *In re* Novatel Wireless Sec. Litig., No. 08-1689 2013 U.S. Dist. LEXIS 154599, at *15–16 (S.D. Cal. Oct. 25, 2013) ("Event studies are crucial to demonstrate loss causation, and indeed some courts describe them as 'almost obligatory.'" (quoting *In re* Vivendi Universal, S.A. Sec. Lit., 605 F. Supp. 2d 586, 599 (S.D.N.Y. 2009))).

[21] *See, e.g.*, *In re* Imperial Credit Indus. Sec. Litig., 252 F. Supp. 2d 1005, 1016 (C.D. Cal. 2003) (granting summary judgment for defendants where plaintiffs failed to introduce an event study to determine damages); *Novatel*, 2013 U.S. Dist. LEXIS 154599, at *16 (observing that "[t]he absence of an event study for damages, in particular, will result in summary judgment in favor of defendants").

[22] *See* Comcast Corp. v. Behrend, 569 U.S. 27, 34 (2013) (requiring that plaintiffs present a methodology showing that damages can be measured on a classwide basis to obtain certification under Fed. R. Civ. P. 23).

[23] *In re* BankAtlantic Bancorp, Sec. Litig., No. 07-61542, 2011 U.S. Dist. LEXIS 48057, at *70 (S.D. Fla. Apr. 25, 2011) ("The greater weight of authority as reflected in many of the circuit and district court opinions that have followed *Dura* and are cited herein, is that a securities-fraud plaintiff can satisfy his burden of proving loss causation only by producing the testimony of an expert who has completed a reliable multiple-regression analysis, event study, and financial analysis in order to quantify the extent to which the claimed losses are the result of the alleged fraud.").

[24] *See, e.g.*, Ohio Pub. Emps. Ret. Sys. v. Fed. Home Loan Mortg. Corp., No. 408-0160, 2018 U.S. Dist. LEXIS 137229, at *30 (N.D. Ohio Aug. 4, 2018) ("Statistical significance, however, is essential to give meaning to statistical evidence.").

[25] *Id.*

[26] Not all courts correctly describe conventional hypothesis testing. For example, it is erroneous to say, as one court recently did, that "a statistically significant result . . . at a five percent level of significance . . . means that there is no more than a five percent chance that the observed relationship is purely random." Carpenters Pension Tr. Fund of St. Louis v. Barclays

the same requirement should apply when event studies are used in the legal context,[27] and most courts have agreed. As one court put it, "The case law establishes that 5% is the standard—though not exclusive—decision rule employed by courts . . . for identifying evidence of market efficiency or price impact in 10b-5 cases."[28]

We have written elsewhere about the challenges in adapting event study methodology for its use in securities fraud litigation, and we identified several important methodological considerations and offered mechanisms for addressing them.[29] In that prior work, we flagged but did not explore an additional concern, the widespread use of the 95% confidence level as the basis for accepting or rejecting the conclusions of an event study as probative on the question of whether a disclosure affected stock price.[30]

Courts use the 95% confidence level in two ways. First, they may determine that an event study that does not establish a relationship at the 95% confidence level does not meet the requirements for admissibility under *Daubert v. Merrell Dow Pharmaceuticals, Inc*.[31] Alternatively, a court may conclude that the inability of an event study to meet the 95% threshold renders it legally insufficient as proof of the relationship between information and stock price for which it is offered.[32] Thus, for example, the court in *Halliburton* concluded that an event study showing a price movement that was statistically significant at a 90% confidence level failed to demonstrate price impact because it was less than the 95% confidence level required.[33] Similarly, the court in *In re American International Group, Inc. Securities Litigation,* rejected plaintiff's expert's findings of price impact where the reported price declines were significant only at the 10% level.[34] And in *In re*

PLC, 310 F.R.D. 69, 81 (S.D.N.Y. 2015). Conventional hypothesis testing does not actually address the probability that an estimated statistical result is the product of chance. Rather, as the text above the line explains, it addresses the probability of observing results at least as extreme as the observed relationship given that only chance is at work. For more on this point, see Gelbach, *supra* note 12.

[27] *See, e.g.*, Jonathan R. Macey, Geoffrey P. Miller, Mark L. Mitchell & Jeffry M. Netter, *Lessons from Financial Economics: Materiality, Reliance, and Extending the Reach of* Basic v. Levinson, 77 VA. L. REV. 1017, 1041 (1991) ("We suggest choosing a significance level such that the probability of a Type 1 error is less than 5%; this is a standard level used by researchers in finance and economics.").

[28] *In re* Chi. Bridge & Iron Co. N.V. Sec. Litig., No. 17 Civ. 01580, 2019 U.S. Dist. LEXIS 180895, at *38 (S.D.N.Y. Mar. 23, 2020).

[29] Fisch et al., *supra* note 7.

[30] *Id.* at 619.

[31] Daubert v. Merrell Dow Pharm., 509 U.S. 579 (1993). *See, e.g.*, Bricklayers & Trowel Trades Int'l Pension Fund v. Credit Suisse Sec. (USA), 752 F.3d 82 (1st Cir. 2014) (rejecting event study for failure to meet the *Daubert* standards).

[32] On both this point and the one involving *Daubert*, see generally Gelbach, *supra* note 12.

[33] Erica P. John Fund, Inc. v. Halliburton Co., 309 F.R.D. 251, 270 (N.D. Tex. 2015) (finding no statistically significant price impact where plaintiffs' expert found a statistically significant impact "only at a 90% confidence level, which is less than the 95% confidence level both experts require in their regression analyses and which the Court finds is necessary").

[34] *In re* Am. Int'l Group, Inc. Sec. Litig., 265 F.R.D. 157, 187 (S.D.N.Y. 2010) (holding that expert findings of statistical significance at the 90% confidence level were not a sufficient

Intuitive Surgical Securities Litigation, the court found an absence of price impact where plaintiff's expert provided event study results reporting price impact at a 90%, but not a 95%, confidence level.[35]

To be sure, not all courts view the 95% standard as critical. As the court observed in *In re Chicago Bridge & Iron Company Securities Litigation*, "it makes little sense to treat a finding just above the 5% level as nearly irrebuttable evidence that the disclosure impacted the stock price, while not considering a finding below the 5% level at all."[36] Similarly, the *United States v. Hatfield* court recognized that, although "the 95% confidence interval is the threshold typically used by academic economists in their work" it was questionable "whether its use is appropriate in a forfeiture hearing, where the Government's burden is by a preponderance of the evidence."[37] The court in *Pirnik v. Fiat Chrysler Automobiles, N.V.* reasoned that "a price impact statistically significant at a confidence level of only 92.12%, which is below the conventional statistical measure of a 95% confidence level, . . . [was] obviously less comfort than a result that is statistically significant at a confidence level of 95%, but it does not prove the absence of price impact."[38]

The reluctance to view the failure to meet the 95% confidence level as dispositive appears most frequently in analyses of market efficiency where some courts have concluded that, while event studies may be useful in establishing market efficiency, they are not required.[39] As a result, a court may consider an event study that establishes the responsiveness of market price to material disclosures at something less than a 95% confidence level as evidence in support of market efficiency. Other courts, however, have held that proof of market efficiency requires that the plaintiff both provide an event study and satisfy the other *Cammer* factors.[40] Thus, in *Ohio Public*, the court concluded that OPERS failed to establish market efficiency based on the four "structural *Cammer* factors" where "[c]orrecting just some of the

basis upon which to find price impact), *vacated on other grounds*, 689 F.3d 229 (2d Cir. 2012).

[35] *In re* Intuitive Surgical Sec. Litig., No. 13-cv-01920, 2016 U.S. Dist. LEXIS 178148, at *45 (N.D. Cal. Dec. 22, 2016).

[36] *In re* Chi. Bridge & Iron Co. Sec. Litig., No. 17 Civ. 01580, 2019 U.S. Dist. LEXIS 180895, at *38–41 (S.D.N.Y. Oct. 18, 2019).

[37] United States v. Hatfield, 795 F. Supp. 2d 219, 234 (E.D.N.Y. 2011).

[38] Pirnik v. Fiat Chrysler Autos., 327 F.R.D. 38, 46–47 (S.D.N.Y. 2018).

[39] *See, e.g.*, City of Cape Coral Mun. Firefighters' Ret. Plan v. Emergent Biosolutions, Inc., 322 F. Supp. 3d 676, 689 (D. Md. 2018) (concluding that plaintiffs established market efficiency even if the court were to ignore the expert's empirical test because "[c]ourts have routinely found a market efficient based on a showing that the first four Cammer factors only are met"). Courts most recently evaluate market efficiency on the basis of the five "Cammer" factors articulated in Cammer v. Bloom, 711 F. Supp. 1264 (D.N.J. 1989). *See also* Carpenters Pension Tr. Fund of St. Louis v. Barclays, 310 F.R.D. 69, 84 (S.D.N.Y. 2015) ("[C]ourts have found market efficiency in the absence of an event study or where the event study was not definitive.").

[40] *See, e.g.*, Ohio Pub. Emps. Ret. Sys. v. Fed. Home Loan Mortg. Corp., No. 08CV0160, 2018 U.S. Dist. LEXIS 137229, at *47–48 (N.D. Ohio Aug. 14, 2018) (holding that plaintiff failed to establish market efficiency based exclusively on deficiencies in the event study conducted by plaintiffs' expert).

flaws in the [expert's empirical analysis] leads to statistically insignificant results."[41]

In requiring a 95% confidence level, few courts have given any consideration to the potential tradeoff between confidence level and power. To our knowledge, only two published opinions in the securities litigation area even address this issue. In *Chicago Bridge*, the court recognized, in evaluating different event study methodologies, the potential tradeoff involved, noting that "by reducing the Type I error rate, multiple comparisons tend to increase the probability of Type II errors."[42] In *In re Barclays Bank Securities Litigation*, plaintiffs challenged the defendant's expert for failing to disclose the Type II error rate generated by his methodology.[43] The court rejected this challenge, stating both that "Type II error rates appear to be irrelevant here" and that the failure to disclose such a rate did not render the methodology unreliable.[44]

Finally, in demanding event studies that meet the 95% confidence level, some courts have failed to distinguish between the two distinct tasks for which event studies are introduced—demonstrating a relationship between information and stock price movement and establishing the absence of such a relationship.[45] As a result, they have concluded that the inability of plaintiffs or their experts to produce an event study that shows a price movement at the 95% confidence level demonstrates the absence of price efficiency or the lack of price impact.[46] But, as more recent decisions increasingly recognize, a failure to reject the null hypothesis is not the same thing as proving that there was no price impact.[47] Indeed, it is hornbook statistics that a failure to reject the null hypothesis does not imply that one accepts the null.

---

[41] *Id.*

[42] *In re* Chi. Bridge & Iron Co. Sec. Litig., No. 17 Civ. 01580, 2019 U.S. Dist. LEXIS 180895, at *45–46 (S.D.N.Y. Oct. 18, 2019).

[43] *In re* Barclays Bank Sec. Litig., No. 09 Civ. 1989, 2017 U.S. Dist. LEXIS 148695, at *81–82 (S.D.N.Y. Sept. 13, 2017).

[44] *Id.*

[45] *See, e.g.*, *In re* Am. Int'l Group, Inc. Sec. Litig., 265 F.R.D. 157, 186–87 (S.D.N.Y 2010) (finding that defendants successfully rebutted the presumption of price impact by showing that price movements failed to meet the 95% threshold necessary for statistical significance).

[46] *Id.*

[47] We identified this issue in our prior work. Fisch et al., *supra* note 7, at 575, 613. Recent decisions increasingly recognize the difference. *See, e.g.*, Li v. Aeterna Zentaris, Inc., 324 F.R.D. 331, 345 (D.N.J. 2018) (holding that the plaintiffs' expert's report "[did] not demonstrate the absence of a price impact," even though he failed to find price impact with 95% confidence); Di Donato v. Insys Therapeutics, Inc., 333 F.R.D. 427, 444 (D. Ariz. 2019) ("The *lack* of statistically significant proof that a statement affected the stock price is not statistically significant proof of the opposite, that is, that it did *not* actually affect the stock price."); *see also* Wilson v. LSB Indus., No. 15 Civ. 7614 (RA), 2018 U.S. Dist. LEXIS 138832, at *37–38 (S.D.N.Y. Aug. 13, 2018) (asserting that courts routinely reject the argument that a non-statistically significant stock price decline proves an absence of price impact).

B. Events Studies in the Academic Literature

As noted above, the event study methodology is drawn from financial economics. Its origins are frequently attributed to a 1969 paper by Eugene Fama, Lawrence Fisher, Michael Jensen, and Richard Roll.[48] Since that time, countless papers have refined the methodology for use in academic research.[49] The use of event studies has expanded from finance research to legal academia. Event studies are used to test for a variety of correlations, to make inferences about causation, and to shed light on the impact of regulatory and policy decisions.[50] Reliance on the 95% confidence level as a threshold for reporting results has an impressive pedigree,[51] although journal articles frequently report results at other confidence levels with appropriate caveats about the significance to be attached to those results.[52]

Courts have generally accepted this literature in determining that judicial reliance on the 95% confidence level is appropriate, but academic commentators have been more transparent in recognizing the policy implications that follow from that choice.[53] In an early article evaluating the use of event study methodology in securities fraud litigation, Macey et al. recognized the tradeoff between power and confidence level and argued that courts should nonetheless apply the 95% confidence level because "this is a standard level used by researchers in finance and economics."[54] Similarly, Fox, Fox, and Gilson identified the tradeoff between Type I and Type II error[55] and

---

[48] Eugene F. Fama, Lawrence Fisher, Michael C. Jensen & Richard Roll, *The Adjustment of Stock Prices to New Information*, 10 INT'L ECON. REV. 1 (1969).

[49] *See, e.g.*, John J. Binder, *The Event Study Methodology Since 1969*, 11 REV. QUANTITATIVE FIN. & ACCT. 111, 120–22 (1998) (describing developments in the academic methodology).

[50] *See, e.g.*, Sanjai Bhagat & Roberta Romano, *Event Studies and the Law: Part I: Technique and Corporate Litigation*, 4 AM. L. & ECON. REV. 141, 142 (2002) ("The event study methodology is well accepted and extensively used in finance" and "[i]ts use in policy analysis in recent years has become more widespread . . . .").

[51] R.A. Fisher adopted the 95% threshold in his influential textbook. *See* R.A. FISHER, STATISTICAL METHODS FOR RESEARCH WORKERS 45, 85 (F.A.E. Crew & D. Ward Cutler, 5th ed. 1934).

[52] *See In re* Chi. Bridge & Iron Co. Sec. Litig., No. 17 Civ. 01580, 2019 U.S. Dist. LEXIS 180895, at *39 (S.D.N.Y. Oct. 18, 2019) (observing that "while the author of a finance article may report findings at the 10% level, in addition to those at the 5% level, that is different than creating a decision rule for identifying evidence of market efficiency or price impact in 10b-5 cases").

[53] Indeed, some academic commentators have gone further and argued that the courts place too much weight on the results of event studies in securities fraud litigation. *See* 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, 260 (2009) (arguing that the "event study requirement poses considerable Seventh Amendment concerns and is inconsistent with the federal securities laws").

[54] Macey et al., *supra* note 27, at 1041.

[55] Edward G. Fox, Merritt B. Fox & Ronald J. Gilson, *Economic Crisis and the Integration of Law and Finance: The Impact of Volatility Spikes*, 116 COLUM. L. REV. 325, 354 (citing example in which use of 95% confidence level results in a Type II error rate of 83%).

presented evidence that crisis-related spikes in idiosyncratic risk can magnify this tradeoff.[56]

In one of the most extensive discussions of the issue, Alon Brav and J.B. Heaton documented the relationship between confidence level and power in securities fraud event studies and demonstrated the potential of this relationship to skew the results against finding evidence of price effect in cases in which the price movement in question is economically significant.[57] To address this concern, they proposed that experts report the statistical power of their event studies, thereby enabling a court to incorporate the risk of Type II errors into its analysis of the event study's reliability.[58] Brav and Heaton did not explain, however, how courts should conduct this analysis.

In a more recent article, Taylor Dove, Davidson Heath, and J.B. Heaton explore the relationship of low statistical power to confounding effects.[59] They show that, because judicial reliance on statistical significance truncates the sample of cases in which event studies demonstrate price impact, confounding effects have the potential to bias the size of the price impact that is demonstrated. This has the result of increasing price impact and, potentially, damages.

We continue, in this Article, where Brav and Heaton left off in their 2015 article.[60] We agree that the tradeoff between confidence level and power affects the role in litigation of event studies, both in terms of the size of the observed effect and the weight that a court will give to it, but we argue that courts are poorly positioned, for several reasons, to evaluate expert data on confidence level and power and to determine how to navigate the tradeoff. We further identify limitations on Congress's ability to address the issue. As a result, for the reasons we detail in Part III below, we believe that task should be performed by the SEC pursuant to formal notice-and-comment rulemaking.

II. Confidence Levels and Power: The Tradeoff

The core idea of the efficient markets hypothesis is that publicly salient events that reduce investors' perceptions of the value of a firm will lead some holders of the firm's stock who think it is overpriced to sell. This dynamic will continue until the market price is in line with the "marginal" investor's perception. A similar mechanism works in the opposite direction. Events that increase perceived value will cause investors to buy anew or

---

[56] *Id.* at 357–58.

[57] *See generally* Alon Brav & J.B. Heaton, *Event Studies in Securities Litigation: Low Power, Confounding Effects, and Bias*, 93 Wash. U. L. Rev. 583 (2015).

[58] *Id.* at 612–13.

[59] *See generally* Dove et al., *supra* note 11.

[60] Brav & Heaton, *supra* note 57, at 614 ("How that tradeoff should occur is beyond the scope of our Article, but addressing that important problem presents a challenge for future work.").

increase their holdings of the firm's stock until its price has risen enough to equilibrate the market. Event studies use statistical techniques to determine whether stock price changes are sufficiently large to suggest that an event of interest must have changed investors' perceptions on the event date.

We begin in section II.A with a brief discussion of the basic components of an event study. In section II.B, we then describe two types of potential errors. A Type I error occurs when the event study concludes that an event had a nonzero effect on stock price when in fact it had no effect. A Type II error involves a result indicating that an event had no effect when in fact it caused a price move.

In section II.C, we discuss conventional hypothesis testing for statistical significance, which entails a careful explanation of confidence levels and the nearly equivalent concept of the significance level of statistical significance tests. These concepts relate to the probability of avoiding Type I errors, for example, concluding that an alleged corrective disclosure reduced stock price when it really did not.

In section II.D, we turn to the concept of power. In our context, power involves the probability of determining that an alleged corrective disclosure reduced firm price when the event really did affect price. In other words, power relates to the probability of avoiding a Type II error.

In section II.E, we consider the implications of requiring a confidence level of 95%, which is what many experts and courts do. We show that with this confidence level, power can be very low for many firms. This means that in securities litigation involving these firms, there will be a high frequency of false negatives—finding there was no effect of an alleged corrective disclosure that really did reduce firm share price.

In section II.F, we investigate how power varies as we relax the confidence level. We show that there is a tradeoff between the two types of errors: greater confidence levels generally come at the price of reduced power. This tradeoff is practically relevant, because conventional hypothesis testing often requires confidence levels high enough to cause power to be very low. For many firms involved in securities litigation, small reductions in these high confidence levels will lead to comparatively large increases in power.

A. Event Study Basics

To conduct an event study, one first identifies the event of interest. In securities fraud litigation, that event is typically the public disclosure of material information. One then defines a pre-event period of study. For the pre-event period, one estimates a regression model that relates the daily stock return to one or more variables representing the behavior of peer stocks, for example, those of firms operating in similar industries. That requires data on daily returns for the stock of interest and for an index of peer stocks. The daily return for the stock of interest is the percentage change in the stock on a given date,[61] with the daily return for an index of peer stocks computed in a way that accounts for differences in overall market valuation of the peer firms.[62] With the data in hand, one then estimates the linear regression relationship between the daily returns of the stock of interest and the peer index.

The resulting coefficient estimates are used with the event-date peer index return to estimate the expected return, that is, the daily return that would have been expected, on average, in the absence of the event of interest; call this estimate $\widehat{R}_e$.[63] We then compute the estimated abnormal return for the event date as the difference between the firm's actual daily return and the expected one. In other words, the estimated abnormal return is found via the equation:

$$\widehat{AR} = R_e - \widehat{R}_e$$

The final step in an event study is to determine whether $\widehat{AR}$ is sufficiently far from zero that one can confidently say that it would be too unusual to observe such a value if there were in fact no effect, a determination that is typically termed statistical significance. We take up that topic in section II.C; first, though, we discuss the conceptually prior concept of Type I and Type II errors.

B. *Type I and Type II Errors in Event Studies*

To explain Type I and Type II errors, it helps to have a concrete type of event in mind. We will focus on a common type of event in securities litigation: plaintiffs' allegation that a firm's alleged corrective disclosure caused the firm's stock price to fall. In this context, the event is the firm's disclosure.

---

[61] For example, if a stock's price rises from \$100 to \$101 on date *t*, then its return for date *t* is 1%—100% times the ratio of a \$1 increase in the stock price to the original value of \$100; had the stock's price instead fallen to \$99, its daily return would have been -1%. Sometimes an author conducting an event study will use a different dependent variable, the log of $(1 + r)$, where *r* is the ratio of the change in the stock price from $t - 1$ to *t* to the stock price at $t - 1$ (in other words, *r* is the daily return measured in decimal rather than percentage terms). Results rarely depend importantly on which approach is used, so for simplicity we will stick to the percentage-change approach.

[62] The usual way to compute the index return is to base daily returns on a value-weighted index on each date. To calculate the value-weighted index amount on date *t*, one simply adds up the market capitalization of all firms in the index. Then one computes the percentage change in this overall market capitalization from date $t - 1$ to date *t*. For example, if the firms making up the index have overall market capitalization of \$100 billion on date $t - 1$ and \$101 billion on date *t*, then the return on the value-weighted index for date *t* would be 1%.

For an algorithmic approach to choosing the industry index, see Andrew Baker & Jonah B. Gelbach, *Machine Learning and Predicted Returns for Event Studies in Securities Litigation*, 5 J.L., Fin., Acct. 231 (2020).

[63] Let R_e be the event-date return for the firm of interest, let R_{peer} be the daily return for the peer index, let *a* be the estimated intercept in the regression, and let *b* be the estimated coefficient on the industry peer index. Then the model-based estimated event-date return is $\widehat{R}_e = a + \left(b \times R_{peer}\right)$.

A legally relevant "effect" means the disclosure caused firm price to fall, and "zero effect" means the disclosure did not affect firm price.[64]

There are two potential types of correct decisions about the effects of the alleged corrective disclosure. First, in those situations in which the disclosure really did cause firm price to fall, it is correct to decide that it did. Second, in those situations in which the disclosure did *not* cause firm price to fall, it is correct to decide that it did not. Respectively, we will term these "correct positives" and "correct negatives."

There are also two potential types of errors. A false positive error involves deciding that an event had an effect when in fact it did not. Statisticians use the term Type I error to refer to false positive errors. So, when an alleged corrective disclosure had no effect on firm price, but experts or courts wrongly decide that it reduced firm price, they have committed a Type I error.

The second type of error is the false negative—deciding an event had no effect when in fact it had a nonzero effect. False negatives, also known as Type II errors, occur in our context when an alleged corrective disclosure actually reduced stock price, but experts or courts decide it did not.

TABLE 1: TYPE I AND TYPE II ERRORS IN THE ALLEGED CORRECTIVE DISCLOSURE CONTEXT

	Expert or Court Decision About Alleged Corrective Disclosure	
Truth About Disclosure	It Had a Nonzero Effect	It Had Zero Effect
It Had Zero Effect	*False positive: Type I Error*	Correct negative
It Had a Nonzero Effect	Correct positive	*False negative: Type II Error*

Table 1 collects these four possible outcomes of a decision-making process—those involving correct positives and false positives (left column), and correct negatives and false negatives (right column).

C. Understanding Hypothesis Testing and the Confidence Level

This section explains how conventional statistical significance testing works. Statistical significance testing is an instance of a broader class of methods known as hypothesis testing. A hypothesis is simply a state of the world. Experts using conventional hypothesis testing define two hypothe-

[64] Analogous reasoning applies if the disclosure instead caused firm price to rise.

ses—the "null hypothesis" and the "alternative hypothesis"—and use statistical methods to choose between them.

In our example of considering whether an alleged corrective disclosure reduced a firm's stock price, the null hypothesis is the hypothesis that the disclosure had *zero* effect on firm price.[65] The alternative hypothesis is that the disclosure *reduced* firm price.

These hypotheses bear some additional explanation. If we want to know whether a firm's price *changed* on a particular date, we can just compare its price on that date to its price on an earlier date. The challenge is that stock prices move all the time, for all kinds of reasons. As we discussed in section II.A, an event study tries to do more than determine whether a stock price changed at all; it tries to determine whether the price fell so much on an alleged corrective disclosure date so that we should believe something particularly important happened that day. An event study is meant to determine whether the observed drop in stock price is within the normal range for the stock in question, for a day when nothing unusual happened.[66] All else equal, large price movements are more unusual and, therefore, more likely to have been caused by the alleged corrective disclosure. But what does "large" mean?

Statistical significance tests answer this question, in the securities litigation context, by defining a threshold level of price drop such that any price drop smaller than the threshold is considered to be within the usual range of chance variations in stock price, as measured by the abnormal return.[67] This usual range is defined to include those values of the abnormal return that would be observed on most dates when nothing unusual happened—in other words, all dates except those on which the most extreme circumstances occurred. The threshold for a price movement that is extreme enough is known as the "significance level" and is often referred to using the Greek letter α. Many experts use a significance level of $\alpha = 5\%$, so that the usual range of abnormal returns includes all but the 5% most extreme negative values.[68]

---

[65] The defendant's case generally would be made not only when the alleged corrective disclosure has zero effect, but also—*a fortiori*—if the alleged corrective disclosure caused firm price to *rise*. We work with the narrow version of the null hypothesis for expositional reasons.

[66] This means an event study tells us something about how likely the observed data (*D*) would be if the null hypothesis (*NH*) were true. Courts sometimes misunderstand this and think event studies (and null hypothesis testing in general) tells us about the probability that the null hypothesis is true, given the observed data. In other words, conventional event studies tell us something about the probability of observing *D*, given *NH*, whereas the legal question is usually thought of as the probability that *NH* is true, given *D*.

These are not generally the same thing; *see, e.g.*, Gelbach, *supra* note 12. For an approach to event studies that uses Bayes's theorem to learn about the probability of the null hypothesis given the data, see Gelbach & Hawkins, *supra* note 7.

[67] *See* section II.A for the definition of abnormal return; for our purposes this is simply the measure of stock price change.

[68] Experts do not necessarily use the exact wording we use here; we are using simplifying language to explain the ideas at play.

To illustrate, suppose the firm being sued has daily abnormal returns that follow the normal distribution with mean 0.[69] The normal distribution is convenient because its volatility is entirely characterized by its standard deviation, which is typically denoted with the Greek letter σ.[70]

To provide some perspective on the range of stocks' standard deviations, we obtained securities price return data for calendar year 2015 from the Center for Research in Security Prices (CRSP) daily stock returns file.[71] We estimated a simple market model for each security, using the CRSP value-weighted index of all securities as the market index.[72] We then calculated daily abnormal returns for each security, and finally we calculated the standard deviation of these daily abnormal returns. Table 2 reports salient percentiles of the distribution of firm-level daily-return standard deviations. The median value of σ across firms was 1.6%. More than half of the securities we considered had a daily abnormal return standard deviation above $\sigma =$ 1.6%. Fewer than 1 in 10 securities in our data had a σ value as low as 0.5%, and the same share had a σ value greater than 4.5%.[73] A quarter had $\sigma <$ 0.9%, and a quarter had $\sigma >$ 2.8%.

---

[69] Normality of daily abnormal returns may be rejected for many firms' stocks; *see* Jonah B. Gelbach, Eric Helland & Jonathan Klick, *Valid Inference in Single-Firm, Single-Event Studies*, 15 Am. Law Econ. Rev. 495 (2013) and sources cited therein for discussion. We assume normality in this part of the paper only for expositional purposes. We can use ideas related to the SQ test proposed in Gelbach, Helland, & Klick to account for any non-normality in a firm's stock returns. Because doing so requires added notation and complexity, we confine the analysis to Part IV.C.

[70] The standard deviation of a random variable's distribution is the square root of the expected value, or mean, of the squared deviation from the random variable's mean. If X is the variable, and μ is its mean, then the standard deviation is $\sigma = E[(X - \mu)^2]$. For random variables with a normal distribution, the entire distribution is fully characterized by the mean and standard deviation.

[71] Wharton Research Data Services, https://wrds-web.wharton.upenn.edu/wrds/ds/crsp/stock_a/dsf.cfm?navId=128. We included data on all equity securities included in the CRSP database, except that we only used data for those securities for which data were available on 200 or more days during 2015. There were 6,820 such securities.

[72] We offer the results in Table 2 in an illustrative spirit only, but, in practice, the firm-level abnormal return is better estimated with the inclusion of an industry-based peer index regressor. For discussion on this topic, see Baker & Gelbach, *supra* note 62.

[73] In other words, the 10th and 90th percentiles were 0.5% and 4.5%, respectively.

TABLE 2: PERCENTILES OF THE FIRM-LEVEL DAILY STANDARD DEVIATION OF ABNORMAL RETURNS

Percentile	Firm-Level Standard Deviation (σ)
10	0.5%
25	0.9%
50	1.6%
75	2.8%
90	4.5%

Throughout the balance of this Article, we will use a standard deviation level of σ = 1% as our benchmark for a relatively low-volatility firm.[74] We will use a standard deviation of σ = 3% as the corresponding benchmark for a relatively high-volatility firm.[75] We use σ = 0.5% and σ = 4.5% as measures of the standard deviation for firms with especially low and especially high volatility, as these are the 10th and 90th percentiles, as reported in Table 2.

In our subsequent discussion, it will be convenient to use the term "stock price drop" in place of "abnormal return."[76] Although we employ these terms interchangeably throughout the following discussion, it should be understood that the term stock price, as used here, is computed after adjusting for other variables included in the stock return model.

Given that abnormal returns are normally distributed, it can be shown that the threshold for a test with 95% confidence level always equals roughly 1.64 times the standard deviation.[77] This means that for our low-volatility firm, with σ = 1%, if nothing special happened on a date of interest, the appropriate threshold for a 5% significance level—and thus a 95% confidence level—is 1.64%. That is, the low-volatility firm will exhibit a stock price drop of at least 1.64% roughly 5% of the time, when there is no event-driven price impact on the date in question. Experts using conventional hypothesis testing would therefore consider a negative abnormal return of magnitude, say, 1.7%, large enough to reject the null hypothesis of no price effect for the low-volatility firm, but not one of magnitude, say, 1.2%.

[74] This is the 27th percentile value of firm-level *s* in our data.

[75] This is the 77th percentile value of firm-level *s* in our data.

[76] This will allow us to avoid repeatedly emphasizing that the estimated abnormal return must be negative to reject the null hypothesis.

[77] This number results from solving the equation $P(Z \leq z_{CL}) = CL$ for the parameter z_{CL}, where Z is a random variable with a normal distribution having mean 0 and standard deviation 1. The object z_{CL} is known as "the *CL*-quantile," and for the standard normal distribution with $CL = 0.95$ (that is, 95% expressed in decimal form), we have $z_{CL} = \Phi^{-1}(0.95)$ which equals 1.64 to two decimal places.

For our high-volatility firm, the one with $\sigma = 3\%$, the threshold is 1.64 times $\sigma = 3\%$, which results in 4.92%.[78] Thus for the high-volatility firm, an estimated stock price drop of 5% would be enough to reject the null hypothesis, but one of 4.8% would not. Intuitively, for a more volatile stock, a wider range of price drops is within the range of observation on typical days. For a firm with greater standard deviation, σ, a more substantial price drop must be observed before an expert can reject the null hypothesis at a given confidence level.

This discussion indicates there are two key variables relevant to the threshold for significance testing. The first is the volatility of the firm's daily stock price, as measured by the standard deviation, which is often referred to using the Greek letter σ. The second variable is the significance level α.

In the foregoing discussion, we used the terms significance level and confidence level. The significance level is the likelihood that we will reject the null hypothesis *when it is true*—meaning that we will incorrectly find a price effect.[79] The lower the value of the significance level α, the larger the price drop on the event date will have to be to reject the null hypothesis.

An equivalent way to describe the concepts underlying the significance level is the confidence level, which henceforth we denote CL. To understand confidence levels, observe that the following are equivalent statements:

(i) an abnormal return is not among the 5% most extreme negative values for typical dates, and

(ii) an abnormal return is among the 95% values that are *not* the most extreme negative values.

For example, we have seen that for a low-volatility firm, with $\sigma = 1\%$, a price drop of greater than 1.64% is among the 5% most extreme observations on a typical date. This means that a price drop of *less* than 1.64% (say, a drop of 1.2%, or a price increase of 0.4%) would be among the other 95% of observations, the range that we considered, within this context, typical price fluctuations. A 95% confidence level means that, if the event didn't cause our price to drop, the observed price change will lie *outside* the 5% most extreme price drops with probability 95%.

The confidence level and the significance level play mirror-image roles in separating the sets of observed price changes into those that do not make us reject the null hypothesis and those that do. Thus, it is equivalent to say that an expert uses a significance level of $\alpha = 5\%$ and that the expert uses a confidence level of CL = 95%. Similarly, an expert using a significance level of $\alpha = 10\%$ has confidence level of CL = 90%, and so on.[80] All else

[78] The exact number is closer to 4.93%; the difference is inconsequential and due to rounding.

[79] That is, the threshold z_{CL} is chosen to ensure this result.

[80] In precise terms, $CL = 100\% - \alpha$.

equal, a higher confidence level requires a larger event-date price drop to determine there was a statistically significant drop in firm price on that date, and a lower confidence level requires a smaller price drop.

In the foregoing discussion, a price drop of 1.64% reflects the threshold value above which a price drop is considered most extreme. We use the term Threshold(σ,CL) to emphasize that the threshold's value depends on both volatility, σ, and the confidence level, CL.[81] As we touched on above, the threshold value equals the product of the standard deviation σ and an appropriate percentile of the normal distribution with mean 0 and standard deviation equal to 1.[82]

Table 3 provides four examples of the threshold values used to implement conventional statistical significance testing. The rows of the table list two confidence levels commonly used—95% and 90%. The columns provide values of the standard deviation for a low-volatility firm (σ = 1%) and a high-volatility firm (σ = 3%).

TABLE 3: EXAMPLES OF THRESHOLD(σ,CL) FOR REJECTING NULL HYPOTHESIS

	Volatility Level	
Confidence Level:	Low: σ = 1%	High: σ = 3%
CL = 95%	1.64%	4.92%
CL = 90%	1.28%	3.84%

Each cell reports the minimum value of the drop in firm price, measured by the event-date abnormal return, necessary to reject the null hypothesis that the alleged corrective disclosure had no impact on the event date.

The table shows that when the confidence level is 95%, an expert would reject the null hypothesis if the low-volatility firm had a price drop of more than 1.64% but will reject the null hypothesis for a high-volatility firm whose price dropped more than 4.92%. Experts requiring the lower confi-

[81] For a firm whose abnormal returns are not normally distributed, the threshold will depend on the shape of the abnormal returns distribution in a way that cannot generally be summarized with just the standard deviation, σ. For our discussion of this more general case, see Part IV.C.

[82] The equation is *Threshold(σ,CL)* = σ × z_{CL}, where z_{CL} is the *CL*-percentile of the standard normal distribution. This threshold reflects the fact that under the null hypothesis that nothing unusual happened on the event date, the event-date abnormal return has a normal distribution with mean 0% and standard deviation σ. It is a fact about the normal distribution that if *X* has a normal distribution with standard deviation σ, its percentiles equal σ times the corresponding percentiles of the standard normal distribution. Thus, the *CL*-percentile of the distribution of *X* equals σz_{CL} *for any choice of CL*. It is straightforward to show that when the null hypothesis is true, a statistician doing null hypothesis testing using the threshold level *Threshold(σ,CL)* will *fail* to reject the null hypothesis *CL* percent of the time; equivalently, the statistician will reject the null hypothesis *100% - CL* = α of the time.

dence level of 90% would reject the null hypothesis whenever there was an event-date price drop of more than 1.28% for the low-volatility firm and 3.84% for the high-volatility firm. These figures show that the threshold for rejecting the null hypothesis is sensitive to both the volatility of the firm's share price and the confidence level the expert uses. Table 4 summarizes the relationship between the null hypothesis, the significance level, and the confidence level.

TABLE 4: CONCEPTS RELATED TO CONVENTIONAL STATISTICAL SIGNIFICANCE TESTING

Concept	Explanation
Null Hypothesis	The hypothesis that the event had no effect on the firm's stock price.
Significance Level: α	Chance of a false positive—a Type I—error when the null hypothesis is true.
Confidence Level: $CL = 1 - \alpha$	Chance of a correct negative result when the null hypothesis is true.

The foregoing discussion summarizes at a high level the key methodology used in conventional statistical significance testing.[83] This discussion paints only half the picture. We turn now to power, which is the other half of the story.

D. Understanding Power

We saw in the previous section that the threshold used for statistical significance testing is based on assuming the null hypothesis is true. That means the test's performance is geared to characteristics of decision-making when the null hypothesis is actually true—namely the probability of making either the false positive (Type I) error or the correct negative decision. But it also means that conventional hypothesis testing is not based on the probability of correct or erroneous decisions given that the null hypothesis is *false*. As a result, it is natural to wonder how conventional statistical significance testing performs when the alternative hypothesis, rather than the null hypothesis, is true.

[83] We have deliberately abstracted from the details of how the standard deviation and the event-date abnormal return are estimated, as well as from various other nuts-and-bolts issues such as non-normality of abnormal returns. These issues are discussed in detail in our other work; *see* Fisch et al., *supra* note 7.

The probability of a false negative (Type II) error is often denoted with the Greek letter β. Power is the probability of correctly rejecting an actually false null hypothesis and is often denoted with the Greek letter π. Because the probability of a false negative and the probability of a correct positive must add up to 1, we have $\pi = 100\% - \beta$.[84]

The greater the power of a test, the higher the chance that the test will detect a real effect. Just as higher confidence is a good thing, so, too, is greater power. Table 5 summarizes the relationship between types of errors, types of correct decisions, significance level, confidence levels, and power.[85]

TABLE 5: CONCEPTS RELATED TO THE TRUTH VALUE OF NULL HYPOTHESIS AND HYPOTHESIS TEST OUTCOMES

	Statistician's Decision		
State of World:	Reject Null	Don't Reject Null	Relationship
Null is True	Type I error (false positive) *Test threshold is set so that this equals the significance level, α*	Correct negative *Test threshold is set so that probability this happens is confidence level, CL*	*CL = 100% – α*
Null is False	Correct positive *Probability it happens is power, π*	Type II error (false negative) *Probability it happens is β*	*β = 100% – π*

Power varies with the magnitude of the actual share price movement. To analyze a test's power, we must therefore specify this magnitude, which we will represent with the Greek letter γ. If the alleged corrective disclosure caused share price to fall by one percent, then $\gamma = 1\%$.

Suppose for the moment that we are somehow able to know that a corrective disclosure actually caused share price to fall by $\gamma = 3\%$. For a low-volatility firm ($\sigma = 1\%$), this is a large effect—three times its daily standard deviation. As in the previous section, the threshold, using the 95% confidence level, to reject the null hypothesis is a price drop of 1.64%. The probability of rejecting the null hypothesis is thus the probability that the event-date estimated abnormal return will reflect a price drop of more than

[84] This follows because, when the alternative hypothesis is true, the only two possible decisions are (i) to reject the null hypothesis, which is a correct positive, or (ii) to fail to reject the null hypothesis, which is a false negative.

[85] This table differs from Table 1 only in that we have added nomenclature.

1.64%, *given that the event actually did cause a price drop of γ = 3%*. In our example, the chances of this happening are 91%.[86]

Table 6 provides the values, for this example, of each of the conceptual variables in Table 5. The top row shows that the test has a confidence level of CL = 95% (corresponding to a Type I error rate of α = 5%) when the null hypothesis is true. This fact is true by construction: that is, we chose the threshold price drop of 1.64% to make sure it would hold, as discussed in section II.C. The bottom row of the table shows what we determined in the previous paragraph: When the event in question actually caused price to drop γ = 3%, a test with a confidence level of CL = 95% correctly rejects the null hypothesis about π = 91% of the time; the other β = 9% of the time, this test commits the Type II error of failing to reject a false null hypothesis. This example shows that the 95% confidence level coexists with high power for our low-volatility firm (σ = 1%) when there is a large price effect (γ = 3%).[87]

Table 6: Power and Error Rate for Low-Volatility Firm with 3% Price Drop at 5% Significance Level

	Reject Null	Don't Reject Null
Null is True	Type I error rate is α = 5% (significance level)	Probability of correct decision is CL = 95% (confidence level)
Null is False	Power—probability of correct decision—is π = 91%	Type II error rate is β = 9%

Now consider the high-volatility firm, the one with standard deviation σ = 3%. For this firm, price drops of γ = 3% happen with reasonably high frequency. Thus, it will be relatively unlikely for the abnormal return to show a large enough price drop that an expert using statistical significance testing will reject the null hypothesis. The expert will reject the null hypothesis under these circumstances only about π = 26% of the time.[88] Thus, when the firm has standard deviation σ = 3% and the expert uses the confi-

86 The probability in question may be found by letting *X* be a normally distributed random variable with mean -3 and standard deviation 1. The probability in question is the probability that *X* will take on a value less than -1.64:

$$Pr(X<-1.64) = Pr(X-(-3) < -1.64-(-3)) = Pr(Z<1.36),$$

where *Z* is a standard normal random variable. This probability is approximately 0.91, or 91%.

87 Equivalently, both types of error are relatively unlikely for such a firm when the alleged corrective disclosure really caused a 3% share price drop.

88 The probability that a normally distributed random variable *X* having mean -3 and standard deviation $\sigma = 3$ will take on a value less than $-1.64 \cdot \sigma = 4.92$ is:

dence level CL = 95%, if the true effect is γ = 3%, the expert will commit a Type II error β = 74% of the time.

TABLE 7: POWER AND ERROR RATE FOR HIGH-VOLATILITY FIRM WITH 3% PRICE DROP AT 5% SIGNIFICANCE LEVEL

	Reject Null	Don't Reject Null
Null is True	Type I error rate is $\alpha = 5\%$ (significance level)	Probability of correct decision is CL = 95% (confidence level)
Null is False	Power—probability of correct decision—is $\pi = 26\%$	Type II error rate is $\beta = 74\%$

One way to think of these results is that (1) the statistical cost of insisting on a 95% confidence level varies with the volatility of the firm's daily abnormal return, and (2) this cost can be substantial, especially for high-volatility firms. An expert using a test with confidence level 95% for a high-volatility firm (σ = 3%) would find no actionable price drop in roughly three of four cases in which there *was* a price drop of γ = 3%. For a firm with market capitalization of $4 billion, a fraud of that magnitude represents $120 million (3% of $4 billion). That is a substantial amount.

There is a general formula that relates power (π), the confidence level (CL), volatility (σ), and the true effect magnitude (γ). It can be shown that higher power is associated with a lower confidence level, with a lower level of volatility, and with a higher true event effect.[89]

We now turn to a systematic discussion of the tradeoff between confidence level and power highlighted by the foregoing discussion.

$$Pr\left(X < -4.92\right) = Pr\left(\frac{X-(-3)}{3} < \frac{-4.92-(-3)}{3}\right) = Pr\left(Z < -0.64\right)$$

where Z is a standard normal random variable. This probability is approximately 0.26, or 26%.

[89] When abnormal returns have a normal probability distribution, this formula is: $\pi = \Phi\left(z_{1-CL} + \frac{\gamma}{\sigma}\right)$, where Φ is the cumulative distribution function of the standard normal distribution, z_p is the p^{th} percentile of this distribution (for example, with $p = 0.5$, $z_p = 0$, which is another way of saying the median of the standard normal distribution is 0). Because both the cumulative distribution function and the percentile function are strictly increasing for a continuous random variable, the partial derivatives of power with respect to CL, γ, and σ are negative, positive, and negative, which implies that power is negatively associated with the confidence level and volatility, but positively associated with event effect magnitude. Finally, we note that the formula above may be re-written in terms of Type I and Type II error rates by substituting $\beta = 1 - \pi$ and $\alpha = 1 - CL$. Thus, when the Type I error rate is α, the Type II error rate is given by the equation $\beta = 1 - \Phi\left(z_\alpha + \frac{\gamma}{\sigma}\right)$. The same reasoning above indicates that the two types of error rates are negatively associated—which is another way of saying there is a tradeoff between them—and that the Type II error rate is positively associated with volatility (σ) and negatively associated with true event effect magnitude (γ).

E. The Power Implications of Requiring a 95% Confidence Level

To understand the power implications of the 95% confidence level in more general terms, it will help to express abnormal returns as multiples of the standard deviation. This approach helps put stocks whose abnormal returns are naturally more variable on a similar playing field. Recall that a stock with an abnormal return standard deviation of $\sigma = 3\%$ is more volatile than one with a standard deviation of $\sigma = 1\%$, so that it is much more likely that the high-volatility firm will see a sizable price drop even in the absence of an important event.[90] Failing to account for that fact would confuse issues as we consider the tradeoff between confidence level and power.

For this reason, it can be useful to express daily abnormal returns in multiples of the firm's σ. For a firm with standard deviation $\sigma = 3\%$, a daily abnormal return of $\gamma = 3\%$ amounts to a one-σ price drop. Our Table 7 example involved an event that caused a one-σ price drop, showing that when daily abnormal returns are normally distributed, we will reject the null hypothesis only $\pi = 26\%$ of the time when the event actually caused a one-σ price drop. This is equally the case any time a firm's standard deviation equals the price drop, that is, whenever $\sigma = \gamma$. Thus, power is $\pi = 26\%$ not only for our Table 7 example with $\sigma = \gamma = 3\%$, but also for the combination of a low-volatility firm with a small true price drop, that is, $\sigma = \gamma = 1\%$, as well as for a very high-volatility firm with a very high price drop, that is, $\sigma = \gamma = 5\%$.

Further, there is nothing special about the one-σ aspect of these examples. If the ratio of σ to γ is the same, then the power of the test will be the same.[91] Consider again the low-volatility firm, with $\sigma = 1\%$, but now consider an event that causes a $\gamma = 3\%$ price drop. That is exactly the example we considered in Table 6, where we saw that power was $\pi = 91\%$. Thus, any time the true price effect caused by an event is three times the stock return's standard deviation—so that $\gamma = 3\sigma$—power will be $\pi = 91\%$ when the confidence level is CL = 95%.

We can determine test power for other values of price effect and volatility besides those for which $\gamma = \sigma$ or $\gamma = 3\sigma$. Table 8 does this, showing how power, π, varies with the ratio of event effect size, γ, to volatility, σ, given that we continue to require a CL = 95% confidence level. The examples just discussed, from the two preceding tables, correspond to the rows of Table 8

---

[90] For the same reason, the firm with $\sigma = 3\%$ is also more likely to see larger price *increases* than is the firm with $\sigma = 1\%$, even in the absence of events that would tend to cause price to rise.

[91] This invariance can be shown to be a property of the normal distribution, whose shape is entirely governed by the standard deviation parameter. That means changes in the *scale* of the distribution—how dispersed daily abnormal returns will typically be—are fully captured by changes in the standard deviation parameter. Accordingly, as long as we are willing to maintain the assumption of normality, we lose nothing by focusing only on the event effect's magnitude in standard deviation multiples.

for which the value in the first column is 1.00 times σ (Table 7) and 3.00 times σ (Table 6).

TABLE 8: HOW POWER VARIES WITH ACTUAL EVENT EFFECT SIZE (EFFECT SIZE IS MEASURED IN UNITS OF STANDARD DEVIATION)

True Event Effect Size, γ, Measured in Multiples of σ	Power: π (Probability of Rejecting Null When CL = 95%)
0.00*	5%*
0.25	8%
0.50	13%
0.75	19%
1.00**	26%**
1.25	35%
1.50	44%
1.64	50%
1.75	54%
2.00	64%
2.25	73%
2.50	80%
2.75	87%
3.00***	91%***
3.25	95%
3.50	97%

* The null hypothesis is true when the event effect is 0.
** Table 6 example.
*** Table 7 example.

Additionally, Table 8 shows that the test's power is very low when the effect size is less than 1 unit of standard deviation. Power meets or exceeds 50% whenever the true effect size γ is at least 1.64 multiples of σ, and it is 80% or greater whenever the true effect size γ exceeds 2.50 multiples of σ. The overall lesson that results from Table 8 is simple: the greater is the true effect size, γ, *relative to abnormal return volatility*, σ, the greater will be the test power for a given confidence level. Thus, holding the confidence level constant, greater effect size and greater power go hand in hand.

Our power discussion so far has effectively held constant the volatility of firms' daily abnormal returns. To understand how volatility itself matters, it is necessary to ask what magnitude of true effect is necessary to have a particular degree of power. We illustrate the answer to this question by comparing the required event effect magnitude for low- and high-volatility firms. Once again, we use a firm with a standard deviation of daily abnormal returns of $\sigma = 1\%$ as our low-volatility example, and one with a standard deviation of $\sigma = 3\%$ as a high-volatility firm.

The first column of Table 9 shows various levels of power that one might demand. The table's second column provides the required true event effect magnitude necessary to achieve the power listed in the first column, for our low-volatility firm (the one that has $\sigma = 1\%$).[92] The third column shows the corresponding required event effect magnitude necessary for our high-volatility firm (the one with $\sigma = 3\%$). To achieve an even chance—50%—of rejecting a false null hypothesis requires a $\gamma = 1.64\%$ price drop with the low-volatility firm. By contrast, the test achieves power of 50% with the high-volatility firm only if the true event effect corresponds to a price drop of $\gamma = 4.92\%$. To obtain power of 80%, a level often sought in applied statistics,[93] we need the true event effect to be $\gamma = 2.49\%$ for the low-volatility firm and $\gamma = 7.46\%$ for the high-volatility firm.[94]

---

[92] By design, the first two columns of Table 9 are just the two columns of Table 8, but in reverse order.

[93] *See, e.g.*, Jacob Cohen, Statistical Power Analysis for the Behavioral Sciences (2d ed. 1988).

[94] Careful readers will note that the numbers in the third column all equal three times those in the first column. That occurs because the impact of increasing volatility by a factor of 3—moving from a firm with $\sigma = 1\%$ to one with $\sigma = 3\%$—is exactly offset by tripling the event effect size, γ, and vice-versa. With normally distributed price effects, this relationship holds generally, that is, not just with a factor of 3. Thus, as we scale up the volatility, we also scale up the event effect size, γ, required to achieve any given desired power, π.

TABLE 9: REQUIRED EFFECT SIZE TO ACHIEVE DESIRED POWER WHEN CL = 95%

	Event Effect Magnitude Needed to Achieve Desired Power, π	
Power, π (Probability of Rejecting the Null)	Low-Volatility Firm: σ = 1%	High-Volatility Firm: σ = 3%
8%	0.24%	0.72%
13%	0.52%	1.56%
19%	0.77%	2.30%
26%	1.00%	3.00%
35%	1.26%	3.78%
44%	1.49%	4.48%
50%	1.64%	4.93%
54%	1.75%	5.24%
64%	2.00%	6.01%
73%	2.26%	6.77%
80%	2.49%	7.46%
87%	2.77%	8.31%
91%	2.99%	8.96%
95%	3.29%	9.87%

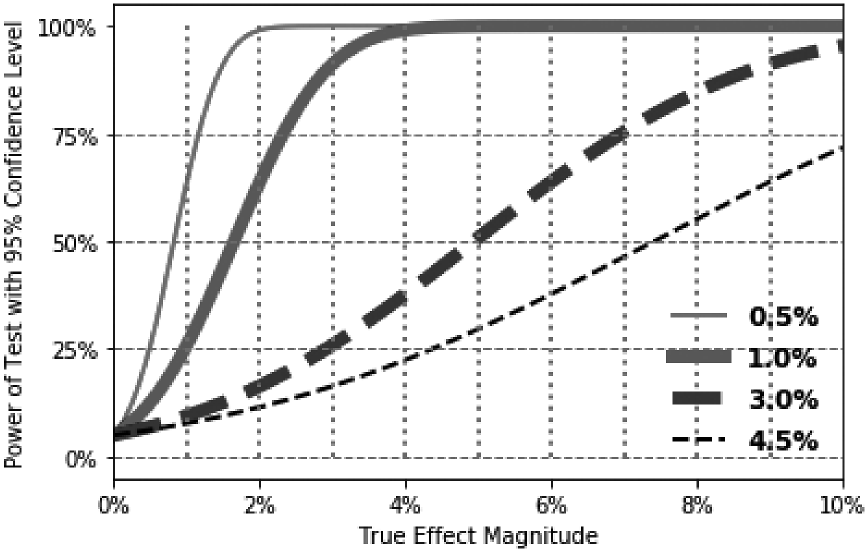
* The null hypothesis is true when the event effect is 0.

It is clear, then, for firms with greater price volatility, a given level of power may require a considerably higher true event effect size. For a high-volatility firm with market capitalization of $4 billion, power of 80% exists only for frauds that inflated firm value by over $300 million. To get power on par with the confidence level of 95% for a high-volatility firm would require that the corrective disclosure caused a price drop of almost 10%, indicating that the fraud inflated firm value by nearly $400 million.

Figure 1 plots the required effect size and power for firms with the volatility levels we have discussed, that is, σ = 1% (thick, solid line) and σ = 3% (thick, dashed line). It also adds the extreme low volatility and extreme high volatility levels of σ = 0.5% and σ = 4.5%.[95]

[95] Recall from the discussion of Table 2 above that these are the 10th and 90th percentiles of the firm-level distribution of σ values.

Figure 1: How the Power-Effect Size Relationship Varies with Standard Deviation
(Test with 95% Confidence Level)



The figure's horizontal axis lists various potential values of the true effect size of a corrective disclosure, ranging from 0% to 10% of firm value. The vertical axis shows the power of a test with a confidence level of CL = 95%. Thus, for the σ = 1% and 3% cases, the figure is essentially a graphical version of the information in Table 9.

Now consider the top curve, which is for a firm with very low volatility, σ = 0.5% (the tenth percentile of the firm-level distribution of σ in our CRSP data for 2015). For such a firm, test power, π, tops 50% even for true event effects as small as γ = 1% (that effect size is not enough even to obtain power of π = 10% for our σ = 3% firm). For a firm with σ = 0.5% and an event effect size of only γ = 2%, power reaches nearly π = 100%; thus, a correct determination that there was an event effect is virtually certain for such a firm even with an event effect that is quite small in absolute terms.[96]

At the other end of the extreme is the bottom curve, which is for a firm with standard deviation of abnormal returns equal to σ = 4.5% (the 90th percentile of the firm-level distribution of σ in our CRSP data for 2015). A test with confidence level CL = 95% is very unlikely to reject the null hypothesis for this firm even when a corrective disclosure has quite a substantial effect on stock price. Even a γ = 5% effect of the corrective disclosure

[96] Note that an event effect of 1% is a two-σ effect for a firm with volatility *σ = 0.5%*; this explains the results just described.

on such a firm's value is only a touch more than a one-σ effect, so it brings only about a $\pi = 30\%$ chance of correctly rejecting the null hypothesis of no event effect.

The takeaways from Figure 1 are straightforward. If a firm has low volatility of daily abnormal returns, then only relatively small effects are needed to have reasonably high power, that is, to identify an effect that really occurred. But for highly volatile firms, power is quite low even for substantively sizable event effects. This is true because it just is not very surprising to see a 5% daily stock price move for a firm with $\sigma = 4.5\%$, even on a day when nothing especially unusual happened for that firm. Thus, Figure 1 illustrates the fact that power depends importantly on both the true effect of the event of interest and the volatility of a firm's daily abnormal returns.

To put all this in perspective, it is helpful to recall from our discussion above that the meat of the firm-level distribution of σ values in our 2015 data lies between $\sigma = 1\%$ and $\sigma = 3\%$. This means that for many firms, power will be relatively low for event effect sizes between $\gamma = 2\%$ and $\gamma = 6\%$. Additional perspective follows from observing that Halliburton had a standard deviation of about 1.7% during the period considered by experts who wrote reports in the *Halliburton* litigation,[97] putting Halliburton close to the middle of our examples of low- and high-volatility firms.

F. The Tradeoff Between the Confidence Level and Power

As noted above, the social science literature commonly uses a 95% confidence level. Our analysis demonstrates the power implications of insisting on that particular confidence level. We have seen that for a large majority of firms, requiring this very demanding confidence level results in a test that has quite low power, except when event effects are very large. In other words, except for high-value frauds, it is very unlikely that an expert witness will reject the null hypothesis and find that the event had a sufficient effect on the stock price even when the fraud is real. In this section we explain how this problem can be addressed by reducing the required confidence level.

Suppose we substitute a 75% confidence level for the required 95% confidence level. How would that change affect the required magnitude of the stock price drop? Consider first our low-volatility firm, the one with $\sigma = 1\%$. For that firm, rejecting the null hypothesis at the 75% confidence level requires a stock price drop of roughly 0.67%, or larger.[98] This is considerably less than the price drop of 1.64% we require to reject the null hypothesis when we use a confidence level of 95%. Table 10 repeats the format of Table

---

[97] *See* Fisch et al., *supra* note 7.

[98] With the threshold for rejection set at -0.67%, the probability of rejecting the null hypothesis when it is true is $Pr(Z<-0.67) = \Phi(-0.67) = 0.25$, where Z is once again a standard normal random variable. Since the probability of rejecting the true null hypothesis, or size, is 0.25, the confidence level is $1-0.25 = 0.75$, or 75%.

8, but with an additional column added to show the power associated with various actual event effect magnitudes when we set the confidence level to CL = 75%.

TABLE 10: HOW POWER VARIES WITH THE CONFIDENCE LEVEL FOR A LOW-VOLATILITY FIRM
(σ = 1%, SIGNIFICANCE THRESHOLD = -1.64%)

True Effect, γ (In Percentage Points)	Confidence Level	
	95%	75%
0.25	8%	34%
0.5	13%	43%
0.75	19%	53%
1	26%	63%
1.25	35%	72%
1.5	44%	80%
1.65	50%	84%
1.75	54%	86%
2	64%	91%
2.25	73%	94%
2.5	80%	97%
2.75	87%	98%
3	91%	99%
3.25	95%	99%
3.5	97%	100%
3.75	98%	100%
4	99%	100%
4.25	100%	100%
4.5	100%	100%
4.75	100%	100%

Comparing the second and third columns of Table 10 shows that for relatively small event effect magnitudes, power is much greater when we reduce the confidence level from CL = 95% to CL = 75%. For example, the table shows that even for a corrective disclosure that reduces firm value by only 0.25%, power is 34%—substantially above the 8% power obtained when the confidence level is CL = 95%. To obtain power above 50%—so that we have a better-than-even chance of detecting a real event effect—it is enough for the event effect to cause a price drop of only $\gamma = 0.75\%$. And for an event that causes price to drop by $\gamma = 1.64\%$—the estimated daily abnormal return required to reject the null hypothesis of no event effect when we insist on confidence 95%—power is 84%. With a confidence level of CL = 95%, power that great would require that the event actually caused stock price to drop at least $\gamma = 2.5\%$.

The same pattern holds when we consider a high-volatility firm, one with $\sigma = 3\%$. Table 11 reports the power level for confidence levels CL = 95% and CL = 75% for such a firm. Reducing the required confidence level from 95% to 75% increases power from $\pi = 6\%$ to $\pi = 28\%$ when the event causes a drop in stock price of $\gamma = 0.25\%$. With a confidence level of CL = 75%, power rises to $\pi = 50\%$, representing an even chance of correctly rejecting the null hypothesis, for an event that causes a drop in stock price of $\gamma = 2\%$; that is more than three times the power of $\pi = 16\%$ achieved when we use a confidence level of 95%. In fact, with a confidence level of 95%, even an event that causes firm price to fall by $\gamma = 4.75\%$ is not enough to yield an even chance of correctly detecting that there was an effect size: power is just $\pi = 48\%$ in this case.

TABLE 11: HOW POWER VARIES WITH THE CONFIDENCE LEVEL FOR A HIGH-VOLATILITY FIRM
(σ = 3%, SIGNIFICANCE THRESHOLD = -4.92%)

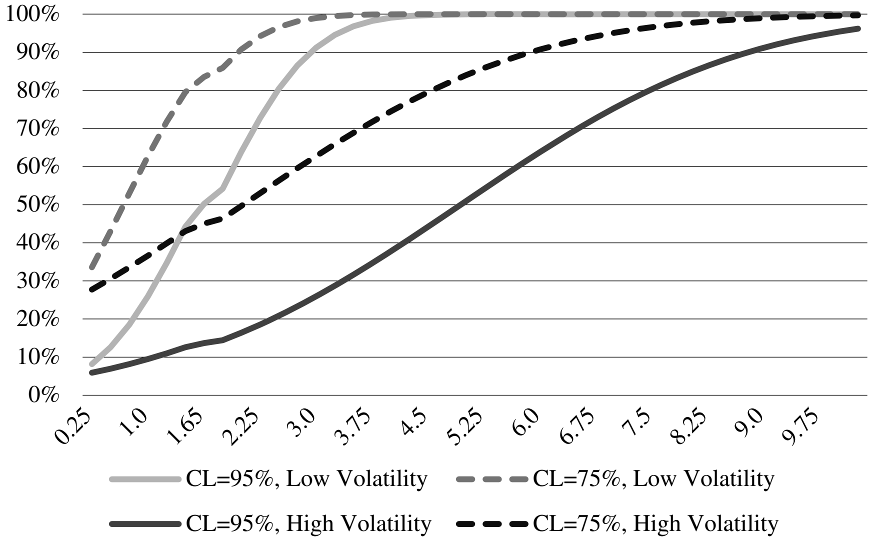
True Effect, γ (In Percentage Points)	Confidence Level	
	95%	75%
0.25	6%	28%
0.5	7%	31%
0.75	8%	34%
1	9%	37%
1.25	11%	40%
1.5	13%	43%
1.65	14%	45%
1.75	14%	46%
2	16%	50%
2.25	19%	53%
2.5	21%	56%
2.75	23%	60%
3	26%	63%
3.25	29%	66%
3.5	32%	69%
3.75	35%	72%
4	38%	75%
4.25	41%	77%
4.5	44%	80%
4.75	48%	82%

We draw together the information in Table 10 and Table 11 in Figure 2, which plots the power level, on the vertical axis, against the true event effect size, on the horizontal axis. The solid lines plot power for the traditional confidence level of CL = 95%, with the dashed lines plotting power for the more relaxed confidence level of CL = 75%. All else equal, power is always greater with (i) less volatility or (ii) a lower confidence level. That explains why the highest-power curve is the one for the low-volatility firm (σ = 1%) when we use the confidence level CL = 75%, and also why the lowest-power curve is the one for the high-volatility firm (σ = 3%) when we use the confidence level CL = 95%.

It is interesting to compare the two middle curves. The dashed one plots power for the high-volatility firm with a confidence level of CL = 75%, and the solid one plots power for the low-volatility firm with a confidence level of CL = 95%. For small event effects—those causing stock price to drop less than about γ = 1.5 percentage points—the high-volatility firm with a relaxed confidence level has greater power than the low-volatility firm with the traditional confidence level. Evidently, the effect of relaxing the confidence level is enough to overcome the power disadvantage caused by greater volatility when the true event effect is relatively small. For larger event effects, this relationship is reversed: power is greater with the low-volatility firm and the traditional, demanding confidence level of 95% than for the high-volatility firm with a relaxed confidence level.

This discussion illustrates the complex interplay between power, volatility, and confidence level. Power can always be increased by reducing the confidence level, and power is always greater for firms with lower volatility. But changing *both* the confidence level and volatility at once can have ambiguous effects on power.

Figure 2: Illustrating How the Power-Effect Size Relationship Varies with Volatility and Confidence Level



Horizontal axis: event effect size

Vertical axis: Power (probability of rejecting the null hypothesis of zero event effect)

Table 12 provides a final vantage point from which to assess the relationship between power and the confidence level. The table relates values of the confidence level, between CL = 95% and CL = 50%, to power for a test of the null hypothesis that the event had no effect.[99] The second and third columns involve an event that actually reduced firm price by $\gamma = 1\%$, and the fourth and fifth columns involve an event that actually reduced firm price by $\gamma = 3\%$.

TABLE 12: POWER RISES AS THE REQUIRED CONFIDENCE LEVEL FALLS (PRESENTED BY EVENT EFFECT MAGNITUDE AND FIRM VOLATILITY)

	$\gamma = 1\%$ Event Effect Firm Volatility Level		$\gamma = 3\%$ Event Effect Firm Volatility Level	
Confidence Level	Low: $\sigma = 1\%$	High: $\sigma = 3\%$	Low: $\sigma = 1\%$	High: $\sigma = 3\%$
95%	26%	9%	91%	26%
90%	39%	17%	96%	39%
85%	49%	24%	98%	49%
80%	56%	31%	98%	56%
75%	63%	37%	99%	63%
70%	68%	42%	99%	68%
65%	73%	48%	100%	73%
60%	77%	53%	100%	77%
55%	81%	58%	100%	81%
50%	84%	63%	100%	84%

All four columns show that as we reduce the required confidence level from CL = 95%, the power of the test rises.[100] Even with a relatively small event effect of $\gamma = 1\%$ and a high-volatility firm ($\sigma = 3\%$), it is possible to achieve power of nearly $\pi = 50\%$—that is, a nearly even chance of rejecting the null hypothesis—by reducing the confidence level from CL = 95% to CL = 65%.

The table also shows that reductions in the confidence level increase power more when the confidence level is initially high. For example, reducing it from 95% to 90% increases power by thirteen percentage points (from $\pi = 26\%$ to $\pi = 39\%$) for a 1% event effect with a low-volatility firm, while reducing the confidence level from 75% to 70% increases power by only

[99] A confidence level of *CL = 50%* entails an even chance of committing a Type I error, that is, rejecting the null hypothesis when it is actually true. *See* Gelbach, *supra* note 12.

[100] It is not erroneous that the last column exactly equals the second column. Power is the same with a 3% event effect and $\sigma = 3\%$ as it is with a 1% event effect and $\sigma = 1\%$. Thus, the combination of the low-volatility firm and the small event effect yields the same power as the combination of the high-volatility firm and the larger event effect.

five percentage points (from $\pi = 63\%$ to $\pi = 68\%$). The former power increase is more consequential in relative terms as well, because it comes from a lower initial level than the latter—just 26% for a confidence level of CL = 95%, compared to 63% for a confidence level of CL = 75%. This observation suggests that the tradeoff between confidence level and power is characterized by decreasing returns to relaxing the confidence level.

III. The Securities and Exchange Commission Should Determine the Appropriate Confidence Level For Event Studies in Securities Fraud Litigation

A. Confidence Level Reflects a Policy Choice

We have shown in Part II that the methodology of event studies involves a tradeoff between confidence level and power. This tradeoff is manifested in the various ways that event studies can be used in securities fraud litigation: market efficiency, price impact, loss causation, and damages. In each case, a higher confidence level increases the probability that a price movement will not be deemed sufficiently extreme to meet the legal requirement at issue, increasing the difficulty of a successful fraud claim. A higher power level increases the probability that a price movement will be accurately characterized as extreme, reducing the difficulty of a successful claim. The true impact of the tradeoff depends on the legal context, including both the extent to which the results of the event study are dispositive of the legal issue as well as whether the plaintiff bears the burden of establishing an extreme price drop or the defendant bears the burden of disproving such a drop.

Importantly, we have also demonstrated that the 95% threshold is not an objective measure of scientific validity—it is simply the threshold at which the likelihood of a false positive is less than 5%. Whether that threshold should be used as the standard for determining whether an event study is admissible or probative is a legal, not a scientific, question. Thus, for example, when a court considers an event study purporting to show whether stock price has reacted to a corrective disclosure, the court's task is to determine whether the plaintiff, through the introduction of the event study, either alone or in conjunction with other evidence, has met its burden of establishing loss causation. The event study, and the significance of its result, are simply evidence, and it is for the factfinder to determine the weight given to that evidence. An event study that demonstrates price impact with a confidence level of 95% is, presumably, more probative than one that demonstrates price impact only at a 90% level, because all else equal, the possibility of a false positive is greater in the latter case. But the standard the evidence must meet in any given context is a legal question, not one for an expert witness. Similarly, the relationship between the probabilities associated with confidence level and power and legal standards of proof such as the preponderance standard typically used in the civil context is uncertain.[101]

To be fair, courts appear to believe that they are applying proven standards of scientific validity to the question of whether an event study purports to demonstrate a causal effect between a disclosure and a stock price movement.[102] The problem is that the question of when a statistical result is compelling enough to count as scientific knowledge is quite a different one from the question of when evidence is strong enough to satisfy the policy considerations that drive standards of evidence in regulation or court.[103] Several commentators have criticized courts for their willingness to treat empirical data as scientific fact instead of recognizing that the decision to treat it as reliable or probative is inherently a legal question.[104]

In addition, in the social sciences, event studies do not generally take the form that they do in securities fraud litigation and are used for different purposes. As we have observed in other work, "there are important differences between the scholarly contexts for which event studies were originally designed and the use of event studies in securities fraud litigation."[105] In importing the event study methodology into securities fraud litigation and, in particular, to address the question of whether a particular disclosure has affected the price of a single firm, courts must make adjustments to the standard methodology as well as make judgments about the legal significance of reported results.[106]

How should this legal judgment be made? We argue that, because of the tradeoff between confidence level and power and the resulting effect on the scope of viable securities fraud cases, the confidence level should be chosen in a way that is sensitive to the likelihood of detecting frauds when they have actually occurred.[107]

---

[101] *See generally* Gelbach, *supra* note 12, for an extensive discussion on this point. For a judicial acknowledgment of the point, see Hatfield, 795 F. Supp. 2d at 234 ("Preponderance of the evidence does not anywhere near require 95% certainty, and Professor Harris' study should have made accommodations for this lower evidentiary burden.").

[102] Whether courts are competent to resolve that scientific question is unclear and delegating to expert witnesses both the standards by which their opinions are to be evaluated and the legal significance to be given them is problematic. *See* Ryan D. Enos, Anthony Fowler & Christopher S. Havasy, *The Negative Effect Fallacy: A Case Study of Incorrect Statistical Reasoning by Federal Courts*, 14 J. Empirical Legal Stud. 618, 647 (2017) (identifying "the challenges associated with interpreting statistical evidence in federal courts" and offering proposals for reform).

[103] For more on this set of issues, see generally Gelbach, *supra* note 12.

[104] *See, e.g.*, Frederick Schauer & Barbara A. Spellman, *Is Expert Evidence Really Different?* 89 Notre Dame L. Rev. 1 (2013); Laurens Walker & John Monahan, *Social Authority: Obtaining, Evaluating, and Establishing Social Science in Law*, 134 U. Pa. L. Rev. 477, 490–91 (1986) (arguing that empirical information should be viewed as analogous to law and receive similar treatment by courts as legal precedent).

[105] Fisch et al., *supra* note 7, at 557.

[106] We identify several such adjustments in Fisch et al., *supra* note 7.

[107] We note that use of event studies is primarily in the context of private litigation as, to date, most event studies have been directed to the elements of reliance and loss causation. Our observations apply to government enforcement actions, although to a more limited degree. *See,*

Consider, for example, the context of loss causation. Plaintiffs have the burden of establishing loss causation,[108] and, as noted in Part I, courts have consistently concluded that the production of a reliable event study is a necessary prerequisite for plaintiffs to meet that burden.[109] Thus, the choice of the appropriate confidence level necessary for an event study is likely to be outcome determinative.[110] Simply put, frauds that impact stock price below a certain magnitude, which varies depending on both stock price and volatility, simply are not actionable despite the fact that the economic impact of these frauds on the market and investors can be substantial.[111]

There is nothing inherently problematic in basing a determination of the appropriate confidence level on its expected effect on the incidence and outcome of securities fraud cases and, in turn, on the effect of overall litigation levels and success rates on investors and the capital markets.[112] An extensive literature debates the merits of securities fraud litigation and questions whether the existing legal standards adequately balance the benefits of deterring fraud and compensating injured investors against the cost of frivolous litigation and the systemic burdens imposed by litigation. Our claim is simply that the choice of confidence level plays an important role in that balance and should therefore be informed by these considerations rather than being treated as some kind of exogenously determined scientific truth.

B. The SEC is Best Positioned to Make the Choice of Confidence Level

Federal courts currently determine the appropriate confidence level when event studies are introduced in securities fraud litigation, either through the application of *Daubert* standards to the introduction of expert testimony or in determining the extent to which a particular event study

e.g., SEC v. Berlacher, 2010 U.S. Dist. LEXIS 95759, at *21 (E.D. Pa. Sept. 13, 2010) (rejecting SEC's insider trading case based on event study by defendant's expert asserting that size of stock price movements in responses to disclosures reflected typical market volatility); United States v. Ferguson, 584 F. Supp. 2d 447, 448 (D. Conn. Oct. 31, 2008) (relying on event studies to calculate size of the loss caused by defendants' conduct for purposes of criminal sentencing).

[108] *See Halliburton II.*

[109] In contrast, some courts have concluded that event studies may be useful in establishing market efficiency but are not required. *See supra* notes 39–41 and accompanying text. Similarly, the question of price impact typically arises at the class certification stage at which it is the defendant's burden to establish lack of price impact.

[110] *See, e.g.*, Robert N. Rapp, *Plausible Cause: Exploring the Limits of Loss Causation in Pleading and Proving Market Fraud Claims Under Securities Exchange Act § 10(b) and SEC Rule 10b-5*, 41 Ohio N.U.L. Rev. 389, 394 (2015) (highlighting "the determinative role of loss causation in pleading and proving [securities fraud]").

[111] For example, a $1 per share fraud has more impact if the stock price is $20 than if it is $100, because in the former case the true event effect is $\gamma = 5\%$, whereas in the latter case it is only $\gamma = 1\%$.

[112] We recognize that the policy considerations that we identify may operate differently depending on the element to which an event study is addressed. Accordingly, our Article does not argue that a single level of statistical significance is required across the different legal contexts.

meets the standard required to prove an element of the case. As indicated above, courts have overwhelmingly required that event studies demonstrate statistical significance at the 95% confidence level commonly used in the social science literature. As we have explained, this requirement has a substantial impact on the likelihood of false negatives. With limited exceptions, however, courts have not considered either the tradeoff between confidence level and power or the extent to which their choice of confidence level imposes too high a burden on establishing fraud.

Because federal securities fraud litigation is based on an implied cause of action under Section 10(b) of the Securities Exchange Act of 1934 ('34 Act), the task of determining the required elements of proof as well as the standards necessary to satisfy those elements has fallen to the federal courts as a matter of common law rulemaking.[113] Courts have used a variety of approaches, from seeking guidance from the statutory text or legislative history to an explicit analysis of policy considerations. But with respect to confidence levels, the courts' approach appears to be a matter of inertia. Because federal judges are not typically trained empiricists,[114] they rely on expert testimony which cites the standard used in academic studies and do not engage with the question of whether that standard should be applied in the litigation context.

Once the policy considerations implicit in the choice of confidence level are exposed, however, it becomes clear that although courts are wrong to accept blindly the 95% confidence level, the judiciary also is not the best-situated branch of government to determine whether or how to modify that standard. First, as noted above, judges are not trained empiricists. Most judges are lawyers, and although modern legal training has been heavily influenced by law and economics, "a standard legal education does not include rigorous training in statistics or the evaluation of scientific evidence."[115] Indeed, extensive literature details the various ways in which courts get statistical analysis wrong.[116]

Moreover, determining an appropriate confidence level requires more than skill in empirical methods. Courts must understand and evaluate the

---

[113] *See, e.g.*, Jill E. Fisch, *Federal Securities Fraud Litigation as a Lawmaking Partnership*, 93 WASH. U.L. REV. 453, 459 (2015) (observing that "the courts have taken primary responsibility for developing the scope of the private right of action and articulating the legal requirements for a successful claim").

[114] *See* Enos et al., *supra* note 102, at 619.

[115] *Id.*

[116] *See, e.g.*, D. James Greiner, *The Quantitative Empirics of Redistricting Litigation: Knowledge, Threats to Knowledge, and the Need for Less Districting*, 29 YALE L. & POL'Y REV. 527, 536–38 (2011); Wendy K. Tam Cho & Albert H. Yoon, *Strange Bedfellows: Politics, Courts, and Statistics: Statistical Expert Testimony in Voting Rights Cases*, 10 CORNELL J. L. & PUB. POL'Y 237, 239 (2001); Bernard Grofman, *Multivariate Methods and the Analysis of Racially Polarized Voting: Pitfalls in the Use of Social Science by the Courts*, 72 SOC. SCI. Q. 826, 827–28 (1991); Thomas J. Campbell, *Regression Analysis in Title VII Cases: Minimum Standards, Comparable Worth, and Other Issues Where Law and Statistics Meet*, 36 STAN. L. REV. 1299, 1302–05 (1984).

tradeoff in the context of its effect on securities fraud litigation. Here again, federal courts are inexpert. Securities fraud cases filed by private plaintiffs number in the hundreds each year[117] and tend to be concentrated in courts in three districts.[118] The consequence is that most district court judges are likely to go years without encountering a single securities fraud case, because 677 federal judgeships were authorized nationwide as of 2019.[119] In addition, district court judges may not be exposed to their colleagues' cases, and they lack ready tools to share information and coordinate their decisions outside the normal process of publishing their decisions.

Third and most importantly, federal judges can neither monitor the market and determine whether the existing level of enforcement is appropriate nor evaluate the potential impact on market protection from adopting a given confidence level. A court hearing a securities fraud case learns only about a single firm and a single set of disclosures. There is no place in the case for the introduction of the disclosure practices of other firms, the percentage of disclosures that give rise to litigation, or the extent to which litigation successfully deters fraud or compensates injured investors. Indeed, these factors fail Fed. R. Evid. 401's definition of relevance, because none makes a fact of consequence in the instant action more or less probable. Federal courts do not see cases that lawyers do not file. They do not see investor losses that are not pursued through litigation. And they do not evaluate developments in issuer disclosure practices that result from the modification of legal standards[120] or the effect of those disclosure changes on the market.[121]

What about Congress, then? Despite the extensive development through decisional law, at its core, federal securities fraud litigation is based on a statute. Moreover, the policy questions that we raise certainly feel like ones that are appropriate for legislation.

Here, we have three primary concerns. First, it seems implausible that the technical issue that we raise will command congressional attention. Congress moves slowly, and there is little reason to believe that Congress either

---

[117] *See* Press Release, Cornerstone Res., Securities Class Action Filings Reach Record Levels in 2019 (Jan. 20, 2020), https://www.cornerstone.com/Publications/Press-Releases/Securities-Class-Action-Filings-Reach-Record-Levels-in-2019 (reporting that plaintiffs filed more than 400 cases in 2019 for the third year in a row).

[118] These districts are the Central and Northern Districts of California and the Southern District of New York. *See* Lynn A. Baker, Michael A. Perino & Charles Silver, *Is the Price Right? An Empirical Study of Fee-Setting in Securities Class Actions*, 115 COLUM. L. REV. 1371, 1390 (2015) (terming these "high volume" districts).

[119] ADMIN. OFF. OF THE COURTS, THE JUDICIARY FISCAL YEAR 2020 CONGRESSIONAL BUDGET SUMMARY (Feb. 2019), https://www.uscourts.gov/sites/default/files/fy_2020_congressional_budget_summary_0.pdf.

[120] *See, e.g.*, Karen K. Nelson & A. C. Pritchard, *Carrot or Stick? The Shift from Voluntary to Mandatory Disclosure of Risk Factors*, 13 J. EMPIRICAL LEGAL STUD. 266, 267 (2016) (detailing responsiveness of issuer disclosures to litigation risk).

[121] *See, e.g.*, John L. Campbell, Hsinchun Chen, Dan S. Dhaliwal, Hsin-min Lu & Logan B. Steele, *The Information Content of Mandatory Risk Factor Disclosures in Corporate Filings*, 19 REV. ACCT. STUD. 396, 398 (2014) (identifying the effects of risk factor disclosures on market beta and stock price volatility).

has the interest in addressing these issues appropriately or would usefully harness the appropriate expert input. Second, changing the confidence level is likely to change the viability of securities litigation, and it will be difficult to predict the full effects of such changes. As a result, legal changes in this area are best implemented in a flexible manner that can be refined over time. This approach does not characterize legislation; although Congress can occasionally be prompted to turn its attention to this type of policymaking, it rarely revisits a matter even after it has introduced substantial changes. The last major legislation related to securities litigation was enacted more than 20 years ago.[122] Third, the analysis we suggest is precisely the type of determination appropriately made by an administrative agency, which has both the scientific tools and the detailed familiarity with the securities markets required. The SEC's expertise is precisely the reason that Congress delegates lawmaking to administrative agencies through the rulemaking process.

This takes us then to the SEC. To be clear, the case for the SEC to determine confidence level through rulemaking is more than just process of elimination. For one thing, we believe it is well—and deliberately—positioned to regulate the private securities litigation landscape. As Professor Joseph Grundfest wrote a quarter century ago, before the Private Securities Litigation Reform Act's enactment:

> Congress created the Commission as an expert agency with the capacity to address significant problems affecting the nation's securities markets. Congress also created the Commission as an agency that could thoughtfully address problems too politically charged to be easily resolved on Capitol Hill. Congress then delegated to the Commission substantial authority to define the contours of market activity that would create liability for fraud. In light of the rationales for the Commission's existence and the scope of the relevant Congressional delegation, and in light of the Commission's expertise in litigation matters and the contentious nature of the underlying policy claims, the private securities litigation debate is precisely the sort of intricate, labor-intensive task for which delegation to an expert body is appropriate. For the Commission to continue to avoid the private securities fraud litigation debate, which now stands as one of the major policy disputes facing this nation's securities markets—and which is just the sort of controversy that Congress created the Commission to resolve—is for the Commission to evade its responsibility and to betray its *raison d'etre*.[123]

---

[122] Private Securities Litigation Reform Act of 1995, Pub. L. No. 104-67, 109 Stat. 737.

[123] Joseph Grundfest, *Disimplying Private Rights of Action Under the Federal Securities Laws: The Commission's Authority*, 107 HARV. L. REV. 963, 966–67 (1994).

This extended quotation raises two noteworthy issues. First, the SEC exists to deploy judgment on at least two dimensions. One is the familiar dimension of technical prowess—addressing issues that require knowing how the capital markets' plumbing works, understanding academic or technical economics, and so on. The other dimension of judgment involves combining such technical expertise with normative policy considerations. Ultimately, decision-makers at the SEC make policy choices. It is appropriate for them to decide, for example, that they are willing to tolerate more fraud in order to reduce the extent of litigation costs, or vice-versa. Agency policy goals will more effectively achieve the decision-makers' aims if the decision-makers are better informed with respect to technical questions; this point underscores the classical role of agency expertise. But as at other agencies, SEC officials make policy. Deciding on a standard of evidence for event studies used in securities litigation fits both dimensions of agency judgment, because it involves the combination of normative policy judgments with technical judgments.

Professor Grundfest wrote this passage to address the question of whether the SEC could or should act administratively to redirect enforcement authority away from private litigants and toward itself. He pointed out that the Rule 10b-5 cause of action is a judicial creation, "articulated in neither statute, legislative history, nor regulation."[124] The same is true of the common use of the 95% confidence level. This standard is used in litigation because it has been imported from academia by expert witnesses and courts that lacked a clearly articulated alternative. There is no reason this state of affairs should be regarded as unchangeable.

The SEC has long employed economists with technical expertise in statistical reasoning and methods. The SEC's Division of Economic and Risk Analysis (DERA), which was created in 2009, has the express purpose of "integrat[ing] financial economics and rigorous data analytics into the core mission of the SEC."[125] The SEC currently employs approximately sixty-four trained economists who can help translate the technical issues for its Commissioners.[126] The deployment of their expertise in the ways we suggest would vindicate the long-recognized justification for situating policy-making authority in regulatory agencies. Moreover, the SEC is currently structured to address these questions. One of DERA's activities is reviewing market developments for the purpose of "identifying and analyzing issues, trends,

---

[124] *Id.* at 964.

[125] U.S. Sec. & Exch. Comm'n, *About the Division of Economic and Risk Analysis*, https://www.sec.gov/dera/about (last visited Sept. 24, 2020).

[126] Indeed, SEC Commissioners are not exclusively lawyers and may themselves have technical expertise in empirical methods. For example, Michael S. Piwowar, who served as a Commissioner from 2013 to 2018, had a PhD in finance. Biography, Commissioner Michael S. Piwowar, https://www.sec.gov/biography/piwowar-michael-s, Similarly, Robert J. Jackson, Jr., who served as a Commissioner from 2018 to 2020, held an MBA in finance. Biography, Commissioner Robert J. Jackson, Jr., https://www.sec.gov/biography/commissioner-robert-j-jackson.

and innovations in the marketplace."[127] Within DERA, the SEC has set up the Office of Litigation Economics, which provides litigation support and analysis.

Although there is no indication from the SEC's website that the SEC staff currently evaluates the effectiveness of either public or private enforcement of the antifraud provision, the agency has both the tools and the competence to do so. The SEC already collects and reports statistics on its enforcement of Rule 10b-5.[128] Private organizations such as Cornerstone Research collect and report data on both public and private enforcement actions.[129] It is possible to analyze this data according to a variety of metrics to evaluate the criteria by which filing decisions are made, such as the extent of price movement or overall market loss associated with an alleged fraudulent disclosure as well as the market reaction to those cases. In recent work, Stephen Choi analyzes much of this data to offer preliminary measures to assess the SEC's use of its enforcement discretion and the impact of its enforcement decisions over time.[130]

In addition, the SEC has the ability to expand its research beyond publicly-accessible data by collecting and evaluating data on cases that do *not* result in either public or private enforcement actions. The SEC could collect data on corrective disclosures that do not result in litigation to evaluate the price impact associated with such disclosures and to determine whether litigation under existing empirical standards would be viable. The SEC could analyze non-public data from its investigations and settled cases. The SEC could incorporate data from preliminary investigations that do not result in enforcement actions. Moreover, the SEC could compare the scope of its own enforcement activity with private litigation to determine the extent to which public and private enforcement actions are complementary or duplicative in addressing fraudulent behavior.

The SEC already evaluates, on the individual case level, the effect of enforcement on the securities market in the exercise of its prosecutorial discretion. Market impact—including the message that enforcement will send to other market participants—is one of the factors that influences both the

---

[127] *Id.*

[128] *See* Urska Velikonja, *Reporting Agency Performance: Behind the SEC's Enforcement Statistics*, 101 CORNELL L. REV. 901, 902 (2016) (describing SEC's current disclosure of its enforcement activities and calling for better quality data collection and reporting).

[129] *See* Cornerstone Res. and N.Y.U. Pollack Ctr. L. & Bus., *SEC Enforcement Activity: Public Companies and Subsidiaries Fiscal Year 2019 Update* (2019), https://www.cornerstone.com/Publications/Reports/SEC-Enforcement-Activity-FY-2019-Update (reporting data on SEC enforcement against public companies); Cornerstone Res., *Securities Class Action Filings, 2018 Year in Review* (2018), https://www.cornerstone.com/Publications/Reports/Securities-Class-Action-Filings-2018-Year-in-Review (reporting data on private class actions).

[130] Stephen J. Choi, *Measuring the Impact of SEC Enforcement Decisions*, FORDHAM L. REV. (Working Paper 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3601881.

SEC's decision to bring an enforcement action and the type and level of sanctions it seeks.[131]

The SEC's economists also have the empirical skills to model the potential impact of the methodological choices reflected in this Article. Indeed, the SEC has, on occasion, engaged in detailed economic analyses for the purpose of quantifying the impact of regulatory changes on market behavior. Thus, for example, the SEC created a one-year pilot in 2004 to evaluate the effect of revised regulations concerning short selling.[132] The SEC staff evaluated the pilot in a report which was released in 2007.[133] In 2012, in connection with its consideration of reforms to the rules governing money market funds, the SEC staff empirically analyzed both the efficacy of prior regulatory reforms and the potential impact of additional reforms on the market, using a variety of complex empirical procedures, including Monte Carlo simulations.[134]

Finally, the SEC has the flexibility necessary to engage in effective rulemaking with respect to confidence level. As detailed further below, the SEC can evaluate the extent to which the policy-based choice of confidence level is affected by volatility, market capitalization, and industry as well as tailoring its requirements to reflect differences in the legal question to which the event study evidence is addressed. The SEC can also study the effect of its initial rule and make subsequent adjustments without the barriers associated with formal legislation.

C. The SEC has the Authority to Regulate the Use of Event Studies in Litigation

We also believe that, as a matter of administrative law, the SEC would be on solid ground if it adopted rules related to the use of event studies in securities litigation. Perhaps the *least* persuasive case would be that an interpretative rule interpreting Rule 10b-5 with respect to the use of event studies in court would pass muster under *Auer v. Robins*.[135] Although we think this position is likely right under current law, relying on *Auer* deference is not

---

[131] *See, e.g.*, Stephanie Avakian, Director, U.S. Sec. & Exch. Comm'n Div. of Enf't, Measuring the Impact of the SEC's Enforcement Program, Sept. 20, 2018, https://www.sec.gov/news/speech/speech-avakian-092018 (explaining that the SEC's enforcement objective is "to have as broad an impact on the landscape we police as possible; to bring cases that send messages of general and specific deterrence; and to seek and obtain remedies tailored to the conduct at issue and the message we want to send").

[132] *See, e.g.*, David P. McCaffrey, *Review of the Policy Debate over Short Sale Regulation During the Market Crisis*, 73 Alb. L. Rev. 483, 487–88 (2010) (describing the pilot).

[133] Off. of Econ. Analysis, U.S. Sec. & Exch. Comm'n, *Economic Analysis of the Short Sale Price Restrictions Under the Regulation SHO Pilot 56* (2007), https://www.sec.gov/news/studies/2007/regshopilot020607.pdf.

[134] Div. of Risk, Strategy, & Fin. Innovation, U.S. Sec. & Exch. Comm'n, *Response to Questions Posed by Commissioners Aguilar, Paredes, and Gallagher* (Nov. 30, 2012), http://www.sec.gov/news/studies/2012/money-market-funds-memo-2012.pdf.

[135] *Auer v. Robins*, 519 U.S. 452 (1997).

the only route, nor obviously the best choice by itself. The SEC has broad legislative rulemaking powers under the '34 Act.[136] Indeed, Rule 10b-5 itself[137] was promulgated under a provision of the '34 Act that explicitly grants the SEC the power to create "such rules and regulations as the Commission may prescribe as necessary or appropriate in the public interest or for the protection of investors" in connection with the purchase or sale of securities.[138]

Thus, the SEC has authority to use notice-and-comment rulemaking to promulgate a rule setting forth a broad framework for the use of event studies in 10b-5 actions. The SEC could then issue interpretative rules that put meat on the bones of the framework rule in order to answer specific questions such as the confidence level to be used in court. This approach would allow successive administrations to set and change policy regarding the use of event studies in securities litigation as they saw fit, as we believe is appropriate, provided that the usual requirements of administrative lawmaking are satisfied.

One might wonder, though, whether a litigation standard of evidence is the sort of legal question on which the SEC does or should be considered to have lawmaking authority. Perhaps one might think that litigation standards should be left to courts. We have already explained, in Part III.B, why we think common law judging is not the best source of lawmaking for the standard of evidence in event study use for litigation. However, could the Federal Rules of Civil Procedure or the Federal Rules of Evidence be amended to address the question of statistical proof?

We contend that such a rule would be illegitimate, because it would be outside the domain of the Rules Enabling Act (REA).[139] The REA limits Congressional delegation of rulemaking authority to procedural, practice, and evidence rules that do not "abridge, enlarge or modify any substantive right."[140] The confidence level is a component of the standard of evidence—akin to the choice between the preponderance, clear-and-convincing, or reasonable doubt standards. To the extent that the substance-procedure dichotomy is useful (and the REA makes it a necessary evil), the standard of evidence is therefore substantive.[141] Thus, the confidence level applied to event studies used in securities litigation is not the sort of legal rule that could be addressed through the REA.[142]

---

136 Securities Exchange Act of 1934, 73 P.L. 291, 48 Stat. 881, 891 (1934).

137 Employment of Manipulative and Deceptive Devices, 17 C.F.R. § 240.10b-5 (2012).

138 15 U.S.C. § 78j(b) (2018).

139 28 U.S.C. § 2072 (2018).

140 28 U.S.C. § 2072(b) (2018).

141 *See* Gelbach, *supra* note 12, at 617–26 (discussing Supreme Court decisions under which preponderance is the default rule because "'[a]ny other standard expresses a preference for one side's interests,'" (quoting Herman & MacLean v. Huddleston, 459 U.S. 375, 390 (1983))).

142 There is some question as to whether the operative statutory text should be ignored. Justice Scalia argued that it should in *Shady Grove Orthopedic Assocs., P.A. v. Allstate Ins.*

Our discussion throughout this Article shows that the confidence level plays a critical role in the likelihood that alleged securities frauds will be treated as proved ones. Accordingly, this statistical standard is a core aspect of the Rule 10b-5 action. Recall Professor Grundfest's description of the SEC's role: the agency was created to "thoughtfully address problems too politically charged to be easily resolved" by Congress; it was "delegated . . . substantial authority to define the contours of market activity that would create liability for fraud"; it is well-positioned for an "intricate, labor-intensive task" carried out by an "expert body"; and it has a "responsibility" to weigh in on major policy issues.[143] Thus, given how well-positioned the SEC is to address the issue of the confidence level to be applied to event studies used in securities litigation, and given that the substantive nature of the question places it outside the purview of the judicial rulemaking process, it is hard to see why the SEC's legislative rulemaking authority should somehow fail to apply.[144]

D. Summary

In sum, conventional administrative law arguments establish that the SEC has authority to set general standards of evidence for securities litigation. Such authority includes the power to devise rules for the confidence level to be applied in hypothesis testing used with event studies in Rule 10b-5 litigation. For familiar reasons, neither Congress nor courts in their common law function are best situated to make such policy choices especially well. Moreover, the substantive law nature of the task makes the judicial rulemaking process illegitimate. Thus, the SEC is the right institutional player to address standard of evidence in securities litigation.

IV. A Framework for Choosing the Confidence Level in Light of the Tradeoff with Power

Our discussion in Part II shows that there is a tradeoff between a test's confidence level and its power. Choosing the confidence level amounts to a straightforward tradeoff of the costs associated with our two types of errors.

Co., 559 U.S. 393, 413 (2010) (siding with precedential statutory interpretation that concededly is "hard to square with § 2072(b)'s terms") (Scalia, J., in a portion of the opinion that spoke for three Justices). But the Court has since repudiated that view in two additional cases. *See* Wal-Mart Stores, Inc. v. Dukes, 564 U.S. 338, 367 (2011) (citing § 2072(b) as the reason for the majority's decision in the case) (Scalia, J.); Tyson Foods, Inc. v. Bouaphakeo, 136 S. Ct. 1036, 1046 (2016) (rejecting petitioner's attempt to have evidence deemed improper, simply because the case is a Rule 23 class action, because that would contravene § 2072(b) by abridging a substantive right) (Kennedy, J.).

[143] Grundfest, *supra* note 123, at 966–67.

[144] In addition, Professor Grundfest has argued that the SEC has the lawful power to disimply the 10b-5 action entirely. *Id.* If the SEC has power to eliminate 10b-5 actions altogether, then presumably the SEC has the power to alter the standards by which such actions are proved.

By reducing the confidence level, we increase the frequency of Type I errors, which induces added litigation by plaintiffs and additional pre-litigation defensive behavior by defendants. We will call these costs the "litigation costs" related to a drop in the confidence level. However, because the frequency of Type II errors falls when we reduce the confidence level, doing so entails a drop in the countervailing costs related to Type II errors that occur when we fail to discern frauds that *did* occur. Reducing the confidence level means that actual fraudsters will be less likely to avoid detection, so that "fraud costs" are reduced when we reduce the confidence level.

Choosing a confidence level involves trading off these litigation costs and fraud costs. It is important to recognize that the *optimal* tradeoff between these types of costs is not simply a positive, technical question. Rather, it is a normative question and thus a policy choice: which frauds are costly enough to justify incurring litigation costs?

The discussion in Part II shows that power can be very low, thus, Type II error rates can be very high, when courts impose the 5% Type I error rate. It also shows that power can be increased if one is willing to tolerate a higher Type I error rate, or, equivalently, a lower confidence level. One view of the state of play in Rule 10b-5 litigation is that courts have consciously chosen to use the 95% confidence level because they have determined that it yields the optimal tradeoff between litigation and fraud costs. This is the position taken by Fox, Fox, and Gilson.[145]

We question the claim that the current use of the 95% confidence level reflects a conscious policy choice as opposed to a reflexive reliance on the social science literature. As a result, we offer a framework to make that policy choice: what critical value best balances the litigation costs associated with Type I error rates against the fraud costs associated with Type II error rates?

Section A introduces a formal framework based on conventional neoclassical economic welfare analysis, premised on optimal deterrence analysis. We show that optimal policy depends on (i) positive components related to the impact, the number of frauds and the amount of litigation, the critical value for finding a test significant, and (ii) normative components related to the degrees to which fraud and litigation each impose costs on society. Unfortunately, even the positive components are likely to be difficult if not impossible to estimate convincingly. Accordingly, section B provides a practically feasible alternative approach, one that we believe forges an appropriate compromise between practicality and analytical precision.

A. A Formal Analytical Framework

One approach to our policy choice problem would be to conceive it as one of minimizing total social costs related to fraud costs and litigation

[145] *See* Fox et al., *supra* note 55.

costs. Thus, policy would be chosen to minimize C(fraud,litigation), where C is a function that tells us total social costs for given levels of the variables fraud and litigation.

A decrease in the confidence level CL has the effect of making it easier for plaintiffs to prove fraud. Accordingly, a reduction in CL can be expected to reduce fraud and increase litigation. Other things equal, total social costs fall when fraud falls and rise when litigation rises. The net impact on total social costs of a reduction in CL may be written:

$$\Delta SC = \underbrace{\Delta SC_{\text{Fraud Reduction}} \times \Delta \text{Fraud}}_{\text{Fraud-reduction impact on SC}} + \underbrace{\Delta SC_{\text{Litigation Increase}} \times \Delta \text{Litigation}}_{\text{Litigation-increase impact on SC}}$$

where $\Delta SC_{\text{Fraud Reduction}}$ is the reduction in total social costs due to a reduction in fraud when CL is reduced, ΔFraud is the magnitude of the reduction in due to that reduction in CL, $\Delta SC_{\text{Litigation Increase}}$ is the increase in social costs due to the reduction in CL, and ΔLitigation is the corresponding increase in litigation caused by a reduction in CL.

A reduction in CL will reduce total social costs—ΔSC will be negative—if the fraud-reduction effect outweighs the litigation-increase effect. A reduction in CL will increase total social costs—ΔSC will be positive—if the litigation-increase effect is the larger one. It will make sense to reduce CL further in the former situation, when ΔSC is negative, and it will make sense to instead increase CL in the latter situation, when ΔSC is positive. That means the value of CL is optimal only if the marginal social cost effect of fraud reduction exactly counterbalances the marginal social cost effect of litigation increase.

For example, suppose one fraud reduces social costs by $1 million, and suppose that a contemplated reduction in CL would eliminate four frauds. Then the reduction in CL would have a fraud-reduction impact on total social costs of $4 million. If we suppose that increasing the number of cases litigated by one increases total social costs by $2 million, then the contemplated reduction in CL would make sense as long as it increased litigation by fewer than two cases; for any smaller increase in litigation, the increase in social costs due to increased litigation would amount to less than $4 million. On the other hand, if litigation were to increase by more than two cases, the litigation-increase impact on social costs would outweigh the fraud-reduction impact, in which case it would not make sense to reduce CL (indeed, it would probably make sense to increase it). This example shows that it can make sense to allow a more relaxed evidentiary standard for plaintiffs in securities litigation even if the number of frauds eliminated differs from the number of additional cases litigated. The key is that the social cost impact of each eliminated fraud may differ from the social cost impact of each additional litigated case.

Our discussion shows that a full-fledged optimal policy analysis requires both positive and normative information. The positive information involves the impact of changes in the confidence level on the prevalence of litigation and the impact of changes in the confidence level on the prevalence of securities fraud. These two factors tell us how litigant and executive behavior change with the confidence level. The normative information required involves how social welfare is affected by litigation costs and by the prevalence of securities fraud.

In principle, the positive questions could be answered using statistical analysis. In practice, providing high-quality answers to these questions is likely to be difficult. One issue is that the prevalence of fraud will be hard to measure in those instances in which frauds are not detected. Another is that we lack meaningful empirical variation in the confidence level used in litigation to date. Surely there are other hurdles that will complicate a complete analysis of how the confidence level affects litigation and fraud prevalence.

The normative questions are also difficult to answer. Valuing social welfare effects requires a complete specification of distributional concerns. In the securities litigation context, that requires considering the returns to lawyers and class action plaintiffs, the costs borne by firms, the incidence of those costs borne distinctly by shareholders and workers, the effects on employment, and so on.

We regard the classical social welfare analysis as functionally infeasible for the SEC (or anyone else) to carry out. We therefore offer a practical approach to choosing the confidence level used in securities litigation in Part IV.B.

B. An Informal but Feasible Approach to Setting Policy

We have seen that the current approach leads to a situation in which Type II error rates can be expected to be very high even when fraud of a significant magnitude has occurred. Our feasible approach to policy making is founded on the idea of directly assessing a policy maker's willingness to accept a given rate of failing to detect actual frauds of a given magnitude, in return for keeping as low as possible the rate of concluding that a fraud has occurred when it actually has not. The key variables in our method are the magnitude of fraud, the Type II error rate for that magnitude, and the Type I error rate. As we explain, values of these three variables are together sufficient to determine the confidence level that should be used for any given litigation.

We emphasize that the SEC might have reason to let the values of these variables differ by firm type, so that different cases will have different confidence levels. For example, the SEC's determination may be affected by evidence about the relationship between firms' market capitalization and the incidence of fraud or its social costs. Additionally, firms with higher volatility might face greater risk of frivolous litigation based on large magnitude stock drops. For these reasons, the SEC might reasonably vary the minimum required power depending on firms' market capitalization, volatility, or other characteristics that can be quantified and studied in advance.

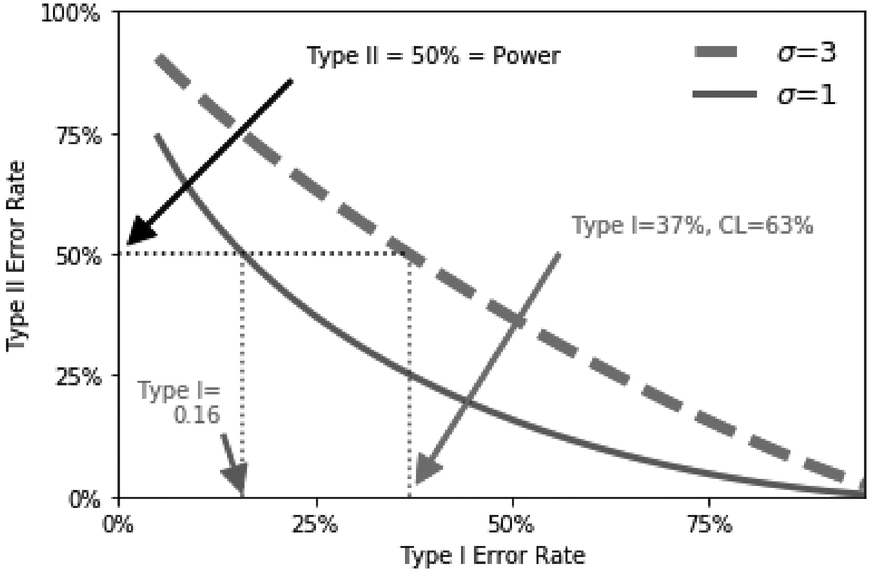
Whatever the mechanism the SEC uses to answer these questions, we suggest it use the following procedure:

1. Choose a policy-salient magnitude of fraud. This amount might be \$50 million; or maybe it is \$1 billion. The requirement is only to pick some value of interest.
2. Choose a minimum required power level, π^{min}, for detecting frauds of the stated magnitude. For example, the SEC might decide that a \$100 million fraud should be detected at least half the time when it happens. In this case, $\pi^{min} = 50\%$; equivalently, this means the maximum allowable Type II error rate is $\beta^{max} = 100\% - \pi^{min} = 50\%$. We emphasize that the SEC's minimum required power level might be influenced by the agency's information about the likely level of undetected capital markets fraud, based on information from its enforcement division, market statistics such as trading spreads, volume of customer complaints, and so forth. Thus, the minimum required power could be different for different types of firms.
3. For the firm being sued, set γ^* equal to the ratio of the magnitude of fraud, from step 1, to the firm's market capitalization. We refer to γ^* as the policy-relevant event effect magnitude, expressed in percentage terms.
4. Calculate the defendant firm's daily abnormal return volatility as measured by its standard deviation, σ.[146]
5. Set α^* equal to the Type I error rate that corresponds to a Type II error rate of β^{max} for a firm with volatility σ when the effect size is γ^*.
6. Do a standard event study-based hypothesis test at significance level α^* (equivalently, confidence level $CL^* = 100\% - \alpha^*$).

To illustrate this procedure, suppose a firm with market capitalization of \$5 billion has been sued. Suppose that a policy maker considers a \$50 million fraud to be significant. Then $\gamma^* = 1\%$. The bottom curve in Figure 3 shows the possible combinations of Type I and Type II error rates for a $\gamma^* = 1\%$ fraud when the firm's standard deviation is $\sigma = 1\%$. We will call this the "error rate combination curve." Any point on the error rate combination curve is feasible for the SEC, in the sense that choosing a given Type I error rate will yield the Type II error rate on the curve when the true effect is $\gamma^* = 1\%$.

[146] In this section, we continue to assume, for the sake of exposition, that the firm's abnormal returns are normally distributed. For the required modification when abnormal returns are not normally distributed, see Part IV.C.

FIGURE 3: THE ERROR RATE COMBINATION CURVE SHOWING THE TRADEOFF BETWEEN TYPE II (β) AND TYPE I (α) ERROR RATES FOR A MAGNITUDE OF FRAUD OF $\gamma = 1\%$



We saw above in Table 8 that for a $\gamma = 1\%$ true effect with a low-volatility firm ($\sigma = 1\%$), a confidence level of CL = 95% corresponds to power of only $\pi = 26\%$. This is equivalent to saying that the Type I error rate of $\alpha = 5\%$ brings a Type II error rate of $\beta = 74\%$.[147] In other words, the "price" of insisting on a confidence level of 95% is that event study tests will miss three out of every four frauds of magnitude $50 million with a $5 billion low-volatility firm.

Suppose the SEC decides this is too high a price to pay and instead believes that it is important to ensure that there is a $\pi^{min} = 50\%$ chance of detecting that size fraud. Then in Figure 3's terms, this policy maker must choose the point of intersection between the bottom error rate combination curve and the horizontal dotted line indicating the Type II error rate is $\beta = 50\%$. The dotted line drawn down to the horizontal axis from this point of intersection shows that the corresponding Type I error rate is 16%. That means the SEC should mandate a significance level of $\alpha^* = 16\%$, rather than the 5% typically used by courts now, when a $5 billion firm with $\sigma = 1\%$ is sued. Another way to say it is that the SEC should insist on a confidence level of $CL^* = 84\%$ rather than the value of 95% typically required now.

What if the firm has higher volatility? The top error rate combination curve (dashed line) in Figure 3 plots the Type II error rate associated with

[147] Recall that the Type II error rate satisfies $\beta = 100\% - \pi$; *see* Table 5.

each Type I error rate for a firm with standard deviation $\sigma = 3\%$, retaining the assumption that the magnitude of fraud of interest is \$50 million and the firm's market value is \$5 billion, so that the SEC again sets $\gamma^* = 1\%$. The figure shows that the line where the Type II error rate is $\beta = 50\%$ intersects the error rate combination curve at a Type I error rate of 37%. Thus, if the SEC adopts a policy of mandating power of at least $\pi = 50\%$ for a \$50 million fraud, Figure 3 indicates the SEC should set the significance level to $\alpha = 37\%$—equivalently, set the confidence level to CL = 63%—when a \$5 billion high-volatility firm ($\sigma = 3\%$) is sued.

Figure 3 thus shows two key points. First, holding constant the defendant firm's market capitalization at \$5 billion, the SEC should require a confidence level substantially below the academic choice of 95% if it wants to ensure the probability of a correct decision will be at least 50% when a fraud of \$50 million has occurred. This is true even if the firm's stock has low volatility. Second, the degree of stock price volatility matters a lot. For a low-volatility firm ($\sigma = 1\%$), obtaining power of $\pi = 50\%$ for a \$50 million fraud would entail reducing the confidence level to 84%, which means a threshold stock-price drop of 1% be statistically significant.[148] The threshold price drop is the same, 1%, for a high-volatility firm ($\sigma = 3\%$), but because that firm's stock returns are more variable, setting minimum required power equal to 50% implies the confidence level is just CL = 63%. In other words, although the threshold price drop is the same regardless of the firms' volatility, that threshold is much more commonly achieved for the high-volatility firm.[149] This result reflects one of our key points: because the test has lower power for a firm with higher volatility, all else equal we should be willing to accept a higher Type I error rate for a higher volatility firm.

This example assumed that the minimum required power for detecting a \$50 million fraud would be $\pi^{min} = 50\%$. As we vary the minimum required power, we necessarily change the required confidence level, CL^*. Increases in minimum required power will lead to reduced CL^* values, and vice-versa.

---

[148] This can be seen by using the equation in footnote 89, setting $\Phi\left(z_{1\text{-}CL} + \frac{\gamma}{\sigma}\right) = 0.5$ and solving for CL. Because $\Phi(0) = 0.5$, we must have $z_{1\text{-}CL} + \frac{\gamma}{\sigma} = 0$, which implies $z_{1\text{-}CL} = -\frac{\gamma}{\sigma}$. Because γ and σ both equal 1 when the policy-salient event effect magnitude is 1% of market capitalization and the firm has low volatility, we have $z_{1\text{-}CL} = -1$. This means the threshold drop in stock price to find a significant effect must be 1%. The value of $1 - CL$ for which this holds is 0.16 (that is, the 0.16-quantile of the standard normal distribution has value -1), so the implied confidence level is $100\% - 16\% = 84\%$.

[149] On the same argument as used in the previous footnote, we again have $z_{1\text{-}CL} = -\frac{\gamma}{\sigma}$. We still have $\gamma = 1\%$, but with our high-volatility firm we now have $\sigma = 3\%$. It follows that $z_{1\text{-}CL} = -1/3$. The value of $1 - CL$ for which this holds is 0.37 (that is, the 0.37-quantile of the standard normal distribution has value -1), so the implied confidence level is $100\% - 37\% = 63\%$. Because the high-volatility firm has standard deviation $\sigma = 3\%$, the 37th percentile of its stock-return distribution is 3 times the corresponding percentile for the standard normal distribution, that is, $3\% \times z_{0.37} = 3\% \times \left(-\frac{1\%}{3\%}\right) = -1\%$, so the threshold is again a price drop of 1% for the high-volatility firm.

To illustrate, suppose we decide that the minimum required power for detecting a $50 million fraud need be only $\pi^{min} = 30\%$, rather than 50%. For a company worth $5 billion, the policy-salient event effect magnitude is thus $\gamma^{*} = 1\%$. This is one unit of standard deviation for a low-volatility firm ($\sigma = 1\%$), but only a third of a unit of standard deviation for a high-volatility firm ($\sigma = 3\%$). Accordingly, any given confidence level will correspond to lower power for the high-volatility firm than for the low-volatility firm. When the true effect is $\gamma^{*} = 1\%$, power of $\pi^{min} = 30\%$ is achieved with a confidence level of 94% for a low-volatility firm—essentially the typical current practice. But for a high-volatility firm ($\sigma = 3\%$), power of $\pi^{min}=30\%$ is achieved only if the confidence level is CL = 80%, so that the plaintiff faces a less demanding standard,[150] though it is more demanding than the CL = 63% value we saw with minimum required power of $\pi^{min}=50\%$.

We have also held the policy-salient magnitude of fraud and market capitalization value constant, which together imply that the effect size γ^* is constant. Suppose we hold the policy-salient magnitude of fraud constant but consider a firm with a lower market capitalization, for example, $2.5 billion rather than $5 billion. When it is expressed in percentage terms, the policy-relevant event effect magnitude of interest is $\gamma^* = 2\%$, rather than 1%. An increase in γ allows a higher (more demanding) confidence level at the same minimum required power of $\pi^{min} = 50\%$. For a high-volatility firm ($\sigma = 3\%$), the confidence level is $CL^{*} = 75\%$. For a low-volatility firm ($\sigma = 1\%$), the confidence level is $CL^{*} = 98\%$. Notice that this is a *more* demanding standard than would be required using the conventional approach.

C. Adopting the Method to Account for Non-Normality

One complication for the method in Part IV.B is that abnormal returns typically do not have a normal distribution. We flagged this point in Part II. If the SEC adopted our proposal, it would need to account for non-normality of abnormal returns. This section shows how. The discussion here is unavoidably technical, and readers may reasonably skip this section.

The importance of normality is a well-known fact from probability theory: the shape of the abnormal returns probability distribution is entirely determined by the level of volatility, σ. This turns out to mean that the threshold for finding statistical significance depends only on the confidence level and the volatility, σ. When abnormal returns have a non-normal distribution, the abnormal returns distribution cannot be fully summarized using σ, so the threshold for finding statistical significance depends on the confidence level and the full shape of the abnormal returns distribution. In other

150 With a confidence level of 80%, the high-volatility firm's threshold value for rejecting the null hypothesis becomes roughly a 2.6% drop in stock price, rather than the threshold drop of 4.9% using a confidence level of 95%.

words, if we call the abnormal returns distribution F, the threshold for statistical significance is now written Threshold(F,CL).

Although it requires some different notation, non-normality is not an impediment to our method. It can be shown that there is still a definite, and useful, mathematical relationship between power (π), the confidence level (CL), the event effect magnitude (γ), and the abnormal returns distribution (now characterized as F rather than summarized with σ).[151]

By using this relationship, we can determine the confidence level value implied when the SEC sets the policy-salient event effect and minimum required power. Let AR(p) be the p-quantile of the abnormal return distribution, which means that the fraction p of abnormal returns are less than the value AR(p). Then given the SEC's choice of minimum required power π^{min} for the policy-salient event effect magnitude γ^*, the required confidence level is given by the equation[152]:

$$\alpha^* = F(AR(\pi^{min}) - \gamma^*).$$

Once it has chosen a policy-salient event effect magnitude γ^* and a corresponding minimum required power π^{min}, the SEC can determine the required significance level α^* using the following procedure:

1. Compute the π^{min}-quantile of the abnormal return distribution, $AR(\pi^{min})$.
2. Subtract the policy-salient event effect γ^* from this quantile.
3. Set α^* equal to the share of abnormal return observations that can be expected to have a value less than the difference from step 2.
4. The required confidence level CL^* is thus 100% minus α^* (with the latter being formatted in percentage terms, for example, if $\alpha^* = 0.16$, we have $CL^* = 100\% - 16\% = 84\%$).

Although this might seem different from the procedure we defined in Part IV.B, the two are closely linked. When abnormal returns follow a normal distribution, knowing the standard deviation of the abnormal return dis-

[151] Written in terms of the significance level α, this relationship is $\pi = F(AR(\alpha)+\gamma)$, where $AR(p)$ is the p^{th} quantile of the abnormal return distribution F, that is, the fraction p of randomly drawn abnormal returns are less than $AR(\alpha)$. Thus, the fraction α of abnormal returns are less than $AR(\alpha)$. The relationship in note 89 for the normal distribution case is a special case of this relationship. To see this, observe that when the abnormal return distribution F is normal with standard deviation σ, as assumed above, $AR(\alpha) = \sigma z_a$, where z_a is the α quantile of the standard normal distribution. Then power is $F(\sigma z_a + \gamma)$, and it can be shown that this is the same as $\Phi\left(z_\alpha + \frac{\gamma}{\sigma}\right)$, where Φ is once again the cumulative distribution function of the standard normal distribution.

[152] In general, it can be shown using the equation in the previous footnote that $\alpha = F(AR(\pi) - \gamma)$. This follows because the functions AR and F are inverses, so that (i) that equation may be rewritten as $AR(\alpha) = AR(\pi) - \gamma$, and (ii) applying the function F to both sides yields the equation $\alpha = F(AR(\pi)-\gamma)$. The version of this equation in the text then follows by setting $\pi = \pi^{min}$ and $\gamma = \gamma^*$, and then denoting the resulting significance level α^*.

tribution is enough to determine any quantile of that distribution.[153] That fact means that the standard deviation of the firm's stock returns tells us all we need to know to determine the required significance level α^*. Thus, this procedure is simply a more general version of the one we introduced in Part IV.B; the additional generality allows us to account for non-normality in the shape of the abnormal return distribution.

Our more general procedure requires that we estimate (i) the abnormal return distribution quantile $\mathrm{AR}(\pi^{\min})$ and (ii) the share of abnormal returns that can be expected to have a value less than the difference $\mathrm{AR}(\pi^{\min}) - \gamma^*$. Both quantities in (i) and (ii) may be estimated validly using the set of estimates of abnormal returns from the estimated regression on which the event study is based, that is, they may be estimated validly using the set of $\widehat{\mathrm{AR}}$ values we defined in Part II.A.[154]

This discussion shows that it is practical to implement our proposed approach without assuming that abnormal returns are normally distributed. This is important given the extensive evidence indicating that abnormal returns are not, generally, normally distributed.[155] The discussion shows that all we need is the ability to estimate a sample quantile of the abnormal return distribution, together with the value of the abnormal return distribution at a particular point, which can be done using estimated abnormal return values from the event study regression model.[156]

---

[153] When abnormal returns are normally distributed, their quantiles necessarily satisfy the relationship $AR(p) = \sigma z_p$, where z_p is the p-quantile of the standard normal distribution (that is, the normal distribution having mean 0 and standard deviation 1). This is a general property of the normal distribution having mean 0 and standard deviation σ.

[154] By "validly estimated," we mean "consistently estimated," which means that in a large enough sample the estimated value can be expected to be very close to the true value. As discussed in Gelbach et al., *supra* note 69, the quantile $AR(p)$ may be consistently estimated by first sorting all the estimated $\widehat{AR}$ values and then finding the least-valued one that has at least the fraction p of observations below it; call this value, $\widehat{AR}(p)$. Thus, we estimate the quantile $AR(\pi^{min})$ using $\widehat{AR}(\pi^{\min})$, which provides the first quantity described in the text above.

The second quantity we must estimate is $F(AR(\pi^{min})\text{-}\gamma^*)$. Even if we do not know the true abnormal return distribution F, we may validly estimate it using the set of estimated residuals, $\widehat{AR}$. This is based on the empirical distribution function, $\widehat{F}$, which is the distribution of the observed estimated abnormal returns $\widehat{AR}$. A result known as the Glivenko-Cantelli theorem implies that the share of $\widehat{AR}$ values that are less than x—written $\widehat{F}(x)$—is a valid estimate of the share of actual abnormal returns with values less than x, which is written $F(x)$. Thus, if we knew the π^{min} quantile $AR(\pi^{min})$, we could estimate the second needed quantity using $\widehat{F}(\mathrm{AR}(\pi^{\min})\text{-}\gamma^*)$, which is the share of estimated abnormal returns, $\widehat{AR}$, that have value less than $AR(\pi^{min}) - \gamma^*$. In practice we do not know $AR(\pi^{min})$. We can solve this problem by first consistently estimating the quantile $AR(\pi^{min})$ as described above, and then plugging the estimate into our function, so that we use $\widehat{F}(\widehat{AR}(\pi^{\min}) - \gamma^*)$ to estimate $F(AR(\pi^{min})\text{-}\gamma^*)$. This can be shown to be a consistent estimator.

[155] *See* Gelbach et al., *supra* note 69.

[156] We note that the discussion in this Part points the way to addressing concerns raised in Fisch et al., *supra* note 7 about how experts compute the threshold abnormal return value for rejecting the null hypothesis. In particular, the approach could easily be adopted to allow for dynamic changes in volatility, using, for example, the GARCH approach that Fisch et al. suggest. All this method requires is an approach that allows consistent estimation of the abnormal return distribution. Thus, it may be applied to virtually any appropriately designed event study. An implication is that it may be implemented using a data-driven approach to determining

D. Discussion of Our Proposed Method

The method we suggest in Parts IV.B and IV.C started with the assumption that the SEC chooses a policy-salient event effect dollar magnitude. Given the defendant firm's market capitalization, this determines a value of γ^*. Once the SEC chooses a minimum acceptable power level, π^{min}, the confidence level, CL^*, can be calculated (given the defendant firm's volatility level, σ). This is true because of the direct mathematical relationship between power, confidence level, effect magnitude, and volatility.[157]

Our suggested approach effectively reverses the conventional hypothesis testing method of starting with a confidence level of 95% and letting the chips fall where they may with respect to power. Instead, it starts with power and allows it to determine the confidence level. But unlike the conventional hypothesis testing method, our method is not arbitrary. The conventional approach of using a 95% confidence level is not based on the determination that this confidence level satisfies a particular policy objective. By contrast, our approach asks the SEC to reflect on its policy goals and then to announce a standard of evidence consistent with those goals.

We emphasize that our approach has the twin implementation virtues of flexibility and objectivity. The approach is flexible because it allows the required confidence level CL^* to vary with the characteristics of the defendant firm. For defendants with lower market capitalization, a given policy-salient event effect will imply a greater value of γ^*, which allows a given power level to be achieved using a higher confidence level. Similarly, the required confidence level will vary with the structure of firms' abnormal return distributions.[158] Further, as we discussed in Part IV.B, the SEC can use its experience and data to make appropriate choices about policy-salient event effect magnitudes and minimum required power in whole *classes* of cases based on firms' market capitalization, volatility, or other characteristics. Thus, if it adopted our approach, the SEC would be using a standard of evidence that responds flexibly—and desirably—to the facts of the case.

We claim the virtue of objectivity as well, at least in part. The method requires two pieces of case-specific information: the defendant firm's market value and the structure of the defendant firm's abnormal return distribution. There is little basis to question the objectivity of market capitalization, because it can be determined from public stock prices. To the extent that one might argue over whether market value should be determined using pre- or post-event data, the SEC could answer that question as part of its rulemak-

---

which firms are used to construct a peer-firm index used as a regressor in the event study regression model; *see* Baker & Gelbach, *supra* note 62.

[157] *See supra* note 89.

[158] This is true whether one assumes normality, so that the standard deviation s is sufficient to capture the structure of the abnormal return distribution, or whether one instead takes the more flexible approach to the abnormal return distribution in Part IV.C. Both approaches amount to estimating a firm-specific abnormal return distribution.

ing. There is more scope for argument over the structure of the defendant firm's abnormal return distribution, because this must be estimated by experts.[159] If there turned out to be substantial within-case argument by parties over the value of σ, the SEC could also announce standards for the conduct of event studies.[160]

Conclusion

Event studies are a virtual necessity in securities litigation. The typical study used by an expert witness employs a 95% confidence level to test for statistical significance, for no reason other than scholarly convention: it is what those writing in the academic literature do. As a positive matter, this represents a disconnect with legal standards of proof. As a normative matter, it may cause courts to reject cases much more often than would be beneficial.

In this paper, we demonstrate why that is the case, extending the nascent literature showing why event-study standard operating procedure can be problematic in securities litigation. We then argue that the SEC, rather than courts or Congress, should develop litigation standards designed specifically to trade off Type I and Type II error rates—equivalently, confidence level and power. Finally, we offer a novel and feasible framework that the SEC might implement. This approach is based on ensuring that a minimum level of power is obtained for a benchmark fraud magnitude. Given knowledge of the defendant firm's market capitalization and abnormal returns distribution, it is straightforward to determine the maximum confidence level (minimum significance level) consistent with the minimum required power of detecting a fraud of the benchmark magnitude.

[159] Again, this is true whether the normality assumption is maintained or not. Either way, the firm's abnormal return distribution is estimated on the basis of the results of an expert's econometric event study.

[160] For suggestions along those lines, see Baker & Gelbach, *supra* note 62.