

# Valid Inference in Single-Firm, Single-Event Studies

Jonah B. Gelbach, *University of Pennsylvania Law School*, Eric Helland, *Robert Day School, Claremont McKenna College*, and Jonathan Klick, *University of Pennsylvania Law School*

Send correspondence to: Jonah B. Gelbach, University of Pennsylvania Law School, USA; E-mail: gelbach@gmail.com

Single-firm event studies play an important role in both scholarship and litigation despite the general invalidity of standard inference. We use a broad cross-section of 2000–2007 CRSP data and find that the standard approach performs poorly in terms of both Type I and Type II error rates. We discuss a simple-to-use alternative, the SQ test, based on sample quantiles of the empirical distribution of pre-event fitted excess returns, which has correct asymptotic Type I error rate. Results suggest that the test will be useful in studying the impact of firm-specific events such as regulation, anti-trust rulings, and corporate or securities litigation. (*JEL*: C12, C14, G00, G14, K00, K22)

## 1. Introduction

Event studies have been used widely to examine proposed mergers, evaluate takeover policy, and assess the effects of a wide range of laws and regulations affecting corporations.<sup>1</sup> Prominent scholars writing in law and

---

We thank the editor and two anonymous referees, as well as Martijn Cremers, Ezra Friedman, Kei Hirano, Al Klevorick, David Tabak, Justin Wolfers, Josh Wright, and participants at the 2007 Future of Securities Fraud Litigation conference at Claremont McKenna College, Northwestern University Law School, Stanford Law School, the University of Illinois Law School, and the University of Pennsylvania Law School for helpful comments.

1. Other recent applications go beyond standard corporate finance topics: [Dellavigna and La Ferrara \(2010\)](#) uses event studies as a forensic tool to assess whether

American Law and Economics Review

doi:10.1093/aler/aht009

Advance Access publication May 12, 2013

© The Author 2013. Published by Oxford University Press on behalf of the American Law and Economics Association. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

finance have emphasized the empirical value of event-study methodology, with [Bhagat and Romano \(2002, p. 141\)](#) writing that “Event studies are among the most successful uses of econometrics in policy analysis,” and [Fama \(1991\)](#) attributing most of what we know empirically about corporate finance to event studies. Most event study applications involve many firms or many events, for which standard statistical inference methods are appropriate.

This paper concerns single-firm event studies, which are especially important in the context of securities litigation. A plaintiff alleging securities fraud under SEC Rule 10b-5 must establish six basic elements: presence of a material misrepresentation or omission; scienter; connection with the sale or purchase of a security; reliance; economic loss; and loss causation (See *Dura Pharmaceuticals v. Broudo*, 544 U.S. pp. 336, 341–42). Event studies can be used to address directly the materiality and loss causation elements. Additionally, financial economics is highly relevant to establishing reliance, which can be tightly linked to the appropriateness of using event studies to address materiality and loss causation. We discuss these issues in Section 2.1.

Below we discuss statistical reasons why the typical approach to event-study methodology is flawed in important cases. The standard approach in testing for statistical significance of event effects involves comparing test statistics to critical values associated with the standard normal distribution. This methodology is justified when excess returns themselves come from a normal distribution, but there is considerable evidence against normality.<sup>2</sup> Alternatively, the standard approach is justified when there are many event dates, and when the researcher is comfortable confining attention to average effects across these many dates. But in the application on which we focus, this fact may be unavailing. For example, securities lawsuits may involve only a relatively small number of event dates on which corrective disclosures or alleged misrepresentations occurred. To the extent that plaintiffs must separately establish significant securities price movements on each of

---

firms violate United Nations weapons embargoes, and [Dube et al. \(2011\)](#) study the effects of coups and related events on stock returns for affected companies.

2. For an early reference on this point, see [Brown and Warner \(1985\)](#); [Ford and Kline \(2006\)](#) present a more recent discussion. We present extensive new evidence to this effect in Appendix A.

these dates, they will have to conduct a collection of single-event studies for each date of interest. For such applications, the standard approach may lead to substantial inferential errors—which is what we find to be the case in the data we use below. Our focus in this paper is on evaluating the performance of an alternative approach to inference. We call this alternative the “SQ” test, since it involves using sample quantiles of the estimated excess returns distribution to estimate the critical value for the event effect. The SQ test is simple to use and has excellent theoretical properties. Most notably, it allows analysts to fix the Type I error rate at any chosen level by choosing an appropriate sample quantile to estimate.

To illustrate the SQ test for the Type I error rate  $\alpha = 0.05$ , suppose a firm discloses that its past quarterly earnings were substantially below the level claimed in an earlier earnings statement. A group of shareholders then files an action under SEC rule 10b-5. To establish materiality and loss causation, the plaintiffs must present evidence sufficient to convince the court that the corrective disclosure reduced the value of the firm’s stock. To use the SQ test, an expert on either side would obtain data on the security’s daily return and the market return for both the event date and a set of, say,  $n = 100$  pre-event observations. She would then use ordinary least squares to estimate the regression of the firm’s return on a constant, the market return, and an event dummy, so that the estimated coefficient on the event dummy is the estimated event effect. All of these steps are taken in both the standard approach and in ours.

The standard approach involves comparing the event dummy’s  $t$ -statistic to critical values based on the standard normal distribution, or the Student’s  $t$  distribution (in practice there will be little difference for standard sample sizes). To test the null hypothesis of a zero event effect against the lower-tailed alternative, an analyst using the standard approach would reject at level 0.05 if the  $t$ -statistic were less than  $-1.64$  (i.e., negative and larger in magnitude than 1.64). By contrast, implementing the SQ test involves calculating the fitted residuals from the estimated model, sorting them, and finding the fifth most negative value among the nonevent dates. The analyst would reject the null hypothesis if the coefficient on the event dummy were less than or equal to this value.

This simple example illustrates how easy the SQ test is to use in practice. Intuitively, the test works because the fifth-most negative element in

a sample of 100 randomly drawn observations is the sample 0.05-quantile, and sample quantiles are consistent estimators of population quantiles (see, e.g., Walker, 1968). Unsurprisingly, then, the SQ test's asymptotic size, or Type I error rate, equals the desired significance level, in this case 0.05. While Conley and Taber (2011) provide formal results establishing this fact, our paper is the first to systematically explore the performance of the SQ test in the corporate finance-event study context. In addition, we offer both interesting empirical results and what we believe to be new analytical results concerning the SQ test's power relative to the standard approach.

The rest of this paper proceeds as follows. We discuss both the event study literature and the econometric literature related to the SQ test in Section 2. In Section 3, we then describe our data, which come from the Center for Research in Security Performance's (CRSP's) database for the years 2000–2007. In Section 4, we introduce a variant of the typical statistical model used in event studies undertaken in securities lawsuits. We use analytical arguments to illustrate the importance of normality of the distribution of excess returns for achieving valid inference, even asymptotically. We quantify the poor Type I error rate performance of the standard approach, showing that this performance is systematically related to estimated quantiles of the firms' estimated excess returns distributions.

In Section 5, we provide a general discussion of the SQ test and then present Monte Carlo results concerning its Type I error rate. We find that the SQ test performs very well in samples with 100 pre-event dates. We then turn, in Section 6, to the issue of power. We show in Section 6.1 that the SQ test has relatively high power: it rejects the null hypothesis of no event effect with relatively substantial frequency when there really is an event effect. This good power performance is important because there is sometimes a tradeoff between size and power: controlling the probability that a test wrongly rejects may force analysts to reject less often when the null hypothesis is actually false. Interestingly, we show analytically that on a size-corrected basis, the standard approach and the SQ test have the same asymptotic power. Thus, as a theoretical matter, the standard approach can have better power than the SQ test only in cases when the standard approach's true Type I error rate differs from the desired significance level. In Section 6.2, we present evidence that, in our data, the standard approach's downward size distortions cause it to reject true effects less frequently than

does the SQ test. Thus, in the empirically relevant case when no size correction is made to the standard approach, the SQ test performs better in our data.

In Section 7, we discuss two potential extensions, involving multiple firms and multiple events. We conclude in Section 8.

## 2. Relationship to Previous Literature

We begin this section by discussing the literature on event studies and corporate finance. We then discuss the relationship of our proposed sample quantile test to existing work in the statistics and econometric literatures.

### 2.1. Event Studies and Securities Litigation

Event studies have been used in the academic literature to analyze many corporate finance issues, including the effects of earnings restatements and the adoption of various corporate governance mechanisms on firm value. They also play a prominent role in merger analysis and antitrust policy in both the academic and regulatory spheres. For example, law and economics scholars have used event studies to examine the effects of state-level legal changes (e.g., takeover statute enactments), as well as federal regulatory changes (e.g., The Sarbanes–Oxley Act and the Private Securities Litigation Reform Act). [Khotari and Warner \(2007\)](#) offer an excellent recent review of the event-study literature, while [Campbell et al. \(1997, p. 149\)](#) provide a very useful textbook discussion.

The popularity of event studies derives from their simple and elegant method of controlling for general market effects and, possibly, other relevant covariates, thereby isolating the causal effects of events such as a law's passage, corporate governance adoption, and so on. Event-study methodology also provides a framework for determining whether estimated effects lie outside the range that could be expected due to ordinary random variation in stock returns, allowing researchers to determine whether the measured effect of an event is statistically significant.

As large a role as event studies play in empirical financial economics and policy analysis, their importance in litigation (e.g., under SEC Rule 10b-5), may be even greater. In *Basic v. Levinson* (1988), 485 U.S. 224, the Supreme

Court held that there is a (rebuttable) presumption that public securities markets are informationally efficient. Under this presumption, new public information is absorbed rapidly into securities prices. Thus, any fraudulent statement or omission is presumed to be quickly capitalized into the price of publicly traded securities. As a consequence, unless the presumption of informational efficiency is rebutted, plaintiffs can claim reliance on the allegedly fraudulent statement or omission simply by stating that they relied on the market price, sparing them the obligation of establishing direct reliance on the allegedly fraudulent statement or omission (the sufficiency of asserting reliance on the market price is the source of the term “fraud-on-the-market”).

For an omitted fact (or misrepresentation) to be material under Supreme Court precedent, “there must be a substantial likelihood that the disclosure of the omitted fact would have been viewed by the reasonable investor as having significantly altered the ‘total mix’ of information made available” (*Basic Inc. v. Levinson*, 485 U.S. 224, 231–32 (quoting *TSC Indus., Inc. v. Northway, Inc.*, 426 U.S. 438, 449).) Loss causation requires that whatever loss the plaintiff alleges she has suffered was actually caused by the allegedly material mis-statement or omission, rather than by something else. Reliance, materiality, and loss causation are linked, because each is connected to the question of whether an alleged misrepresentation affected a security’s price:<sup>3</sup>

- The *Basic* Court held that the presumption of reliance is rebutted if a defendant can show that the alleged misrepresentation did not affect

---

3. We do not mean to suggest that reliance, materiality, and loss causation are indistinguishable elements for Rule 10b-5 purposes. Indeed, the Supreme Court in *Erica P. John Fund, Inc. v. Halliburton Co.*, 131 S. Ct. 2179, 2185–6 (2011) (holding that loss causation need not be proved at the class certification stage) emphasized that it is possible for plaintiffs to gain the benefit of the fraud-on-the-market theory’s rebuttable presumption of reliance without also establishing loss causation. To establish the rebuttable presumption, plaintiffs need show only that the alleged misrepresentations were public, that the security in question traded in an informationally efficient market, and that transactions in question occurred after the alleged misrepresentation but before revelation of the true state of affairs. On the other hand, since transactions might fit this pattern even when fraud revelation has no effect on a security’s price, it is possible to establish reliance via the *Basic* presumption without being able to use an event study to establish loss causation or materiality.

the price received or paid by the plaintiff, or the plaintiff's decision to trade.<sup>4</sup>

- If the security price was unaffected by an alleged misrepresentation, then the misrepresentation is likely not the sort of thing a reasonable investor would view as important information, and is therefore not material.<sup>5</sup>
- If the security price was unaffected by an alleged misrepresentation, then the alleged misrepresentation cannot have caused any loss suffered by the plaintiff.

Plaintiffs seeking to meet the reliance, materiality, and loss causation elements of a securities fraud case can use event study evidence to establish that a security's price movement was associated with allegedly fraudulent statements. According to one Arizona federal district court judge, "[t]he tool most often used by experts to isolate the economic losses caused by the alleged fraud is the event study," (*In re Apollo Group Inc. Securities Litigation*, 509 F.Supp.2d 837, 844 (D.Ariz., 2007)), and some courts have even effectively required the use of an event study for these purposes (*See, for example, In re Oracle Securities Litigation*, 829 F. Supp. 1176, 1181 (N.D. Cal, 1993); *In re Executive Telecard, Ltd. Sec. Litig.*, 979 F. Supp. 1021 (S.D.N.Y. 1997)). Finally, the Supreme Court emphasized in *Dura Pharmaceuticals, Inc. v. Broudo* that plaintiffs must "adequately allege and prove

---

4. *Basic Inc. v. Levinson*, 485 U.S. 224, 248–9 ("For example, if petitioners could show that the 'market makers' were privy to the truth about the merger discussions here with Combustion, and thus that the market price would not have been affected by their misrepresentations, the causal connection could be broken: the basis for finding that the fraud had been transmitted through market price would be gone. Similarly, if, despite petitioners' allegedly fraudulent attempt to manipulate market price, news of the merger discussions credibly entered the market and dissipated the effects of the misstatements, those who traded Basic shares after the corrective statements would have no direct or indirect connection with the fraud.") (footnotes omitted).

5. See, e.g., *Oran v. Stafford*, 226 F.3d 275, 282 (2000, CA 3rd Cir) ("when a stock is traded in an efficient market, the materiality of disclosed information may be measured post hoc by looking to the movement, in the period immediately following disclosure, of the price of the firm's stock. Because in an efficient market the concept of materiality translates into information that alters the price of the firm's stock, if a company's disclosure of information has no effect on stock prices, it follows that the information disclosed . . . was immaterial as a matter of law.") (citation and quotation marks omitted).

the traditional elements of causation and loss” (*Dura*, 544 U.S. 336, 346 (2005)), which would entail alleging and proving “that *Dura*’s share price fell significantly after the truth became known” (*Id.*, at 347). The word “significantly” suggests that event studies are a natural way to meet these requirements, both for their ability to estimate the magnitude of any event effects and for inferential purposes (i.e., testing statistical significance).

While many practitioners have used the standard approach to inference in this context, several authors have recently noted and attempted to address the challenges of conducting inference with only a small number of events. For example, in his study of American Express’s conversion to limited-liability status, [Weinstein \(2008\)](#) considers parametric alternatives to assuming normality of the excess-returns distribution. [Klick and Sitkoff \(2008\)](#) use a Monte Carlo, re-sampling approach in the spirit of permutation and bootstrap testing. Finally, [Hein and Westfall \(2004\)](#) and [Ford and Kline \(2006\)](#) use bootstrap re-sampling methods, which solves the standard approach’s asymptotic inference problems via re-sampling techniques.<sup>6</sup>

## 2.2. Related Statistical and Econometric Literature

The SQ test is related to several strands of statistical and econometric research: outlier detection, predictive tests of structural change, end-of-sample instability tests, permutation and randomization inference, and bootstrap-based inference. As we discuss in Appendix A.3, when there is only one event, the estimated coefficient on the event-dummy in a market model equals the predicted residual from estimating the model without the event-date observation. The ratio of this coefficient estimate to the reported standard error is the usual *t*-statistic. In the literature on regression diagnostics, this ratio is known as the studentized residual and is often used as a measure of an observation’s “leverage” in estimating slope coefficients. The purpose of evaluating leverage is typically to decide whether to omit an observation or otherwise address outlier influence for purposes of improving the performance of slope-coefficient estimation. In the litigation-event study context, this concern does not arise since the event date’s

---

6. We discuss the relationship between the bootstrap and the SQ test in Appendix B of an earlier version of this paper, [Gelbach et al. \(2011\)](#).



degree of leverage is interesting only for purposes of testing the size of the event effect.<sup>7</sup>

In the econometrics literature, the single-firm, single-event context can be thought of as a special case of a class of models considered by [Conley and Taber \(2011\)](#). Conley and Taber's model involves a treatment group with a small number of cross-sectional observations affected by a policy change and a large number of cross-sectional comparison-group observations not affected. Our event date is analogous to their treatment group, and our set of pre-event observations is analogous to their comparison group. Their size results can easily be shown to apply to our context, and a specialization of their Proposition 2 justifies the SQ test.<sup>8</sup>

However, Conley and Taber are primarily interested in estimating effects in different contexts from ours, such as the effects of state-level policy reforms on labor market outcomes. Thus, one contribution of ours is to show that their theoretical results work well in the important context of event studies involving securities returns. An additional contribution of ours involves the analysis of power. While Conley and Taber report interesting simulation results concerning power (see their Table 3), they do not derive any analytical results. In Section 6.1, we show that for single-event studies, the SQ test and the standard approach have identical size-corrected power. This means that if the SQ test has lower power than the standard approach, it is only because the standard approach suffers from size distortions. We also show that in our sample, observed power (i.e., the power of the test without correcting size distortions) is generally better for the SQ test than for the standard approach.

Besides [Conley and Taber \(2011\)](#), the econometric literature most closely related to the SQ test is the literature on structural change. An early example in this literature is [Chow \(1960\)](#), who shows how to test the null

---

7. For a discussion of leverage and influence, see [Belsley et al. \(2004\)](#).

8. We became aware of this aspect of their work after writing an earlier draft of this paper that included our own proof that the SQ test's asymptotic size is correct. An additional paper of which we became aware after writing earlier drafts of this paper is [Simpson and Hosken \(1998\)](#), who actually deploy the SQ test in an FTC working paper concerning one industry-specific empirical application, though without exploring its statistical properties. According to the FTC website, parts of this paper were subsequently published in [Hosken and Simpson \(2001\)](#) and [Simpson \(2001\)](#). Neither of those published papers appears to use the SQ test.

hypothesis that the (linear) regression relationship between  $y$  and  $X$  for the next  $m_2$  observations is the same as for the first  $m_1$  observations, given the normality of regression residuals;<sup>9</sup> Chow's focus was on testing whether the coefficients are the same in the two periods. Since he assumes normality of all residuals, he effectively assumes away the problem we confront here.

A variety of authors subsequently have explored the problem of testing for structural breaks when the break point is unknown; for example, see [Andrews \(1993\)](#). While this literature involves some similar statistical issues to our present context, the break point of interest generally is known in litigation-relevant event studies. An early paper focusing on cases with known break points and allowing for nonnormality is [Dufour et al. \(1994\)](#). Dufour et al consider a more general econometric framework than ours, allowing for both multiple equations and nonlinearity. However, as [Andrews \(2003\)](#) has noted, their three approaches to estimating critical values all have disadvantages (assumed normality, asymptotic conservativeness, and the need to choose values of ancillary parameters). [Andrews \(2003\)](#) develops a test statistic using predicted end-of-sample residuals. Like the SQ test, his test's critical values are estimated using the empirical distribution of predicted residuals from earlier in the sample. While his focus is on two-sided testing, his test and theoretical arguments could easily be modified to accommodate one-sided testing as well.

Another related literature involves permutation and randomization inference. Results in this literature rest on the fact that under the null hypothesis of no event effect, the event-date excess return comes from the same distribution as pre-event excess returns. For more on this literature, see [Rosenbaum \(2002\)](#).

A final related literature involves bootstrap-based inference. Bootstrap test statistics and critical values are computed by replacing an unknown distribution with an empirical distribution function that consistently estimates the unknown distribution. [Hein and Westfall \(2004, HW\)](#) evaluate bootstrap procedures proposed by [Chou \(2004\)](#), [Hein et al. \(2001\)](#), and [Kramer \(2001\)](#). As we do, HW focus on inference in the single-event case, raising

---

9. Chow cites [Mood \(1950, pp. 304–05\)](#) as containing this result for the special case  $m_2 = 1$ .

similar concerns to ours vis-a-vis the standard approach.<sup>10</sup> HW provide both solid heuristic arguments and Monte Carlo evidence that the procedures they evaluate perform well in the single-firm case under both normal and some parametric nonnormal data-generating processes. While HW provide evidence on the S&P 500 and several insurance-based subindexes, we provide comprehensive empirical evidence for a wide array of securities. As noted above, we also derive results that allow analytical comparisons, across the standard approach and the SQ test, of both estimated asymptotic Type I error rates and asymptotic power.<sup>11</sup> For a discussion of the formal relationship between the SQ test and the specific procedure HW use, which is a form of residual bootstrap, see Appendix B of our earlier working paper, [Gelbach et al. \(2011\)](#).

### 3. Data

We use data on securities returns from the widely used Center for Research in Security Performance (CRSP) database, available for academic use through the website of the University of Pennsylvania's Wharton Research Data Services. In order to conduct the Monte Carlo study detailed below, we downloaded all daily observations from the CRSP database for the years 2000–2007. According to p. 10 of the CRSP Data Description Guide, which is available at the Wharton Research Data Services website, the securities included in this sample include common stocks certificates, American Depositary Receipts (ADRs), shares of beneficial interest units (depository units, units of beneficial interest, units of limited partnership interest, depository receipts, etc.), closed-end mutual funds, foreign stocks on NYSE, AMEX, NASDAQ, and NYSE Arca, Americus trust components (primes and scores), HOLDRs trusts, and REITs (real estate investment trusts).<sup>12</sup>

10. Much of HW's interest lies in the single-event, multi-firm case. While our focus is primarily on the single-firm case, we do discuss the multi-firm case in Section 7.1.

11. HW offer analytical results on the asymptotic Type I error rate of the standard approach when the number of firms is large and all firms have the same excess returns distribution. Such results are interesting, but they offer no guidance in the single-firm case that interests us here.

12. We have also re-computed our results for a subset of observations that includes only common stocks ever traded on the NYSE, AMEX, or NASDAQ during our period.

Our initial data draw includes 14,587,459 daily observations. We kept only observations for which the daily returns ( $r_{et}$ ) and value-weighted returns including dividend ( $vwret_d$ ) variables were nonmissing, eliminating 219,774 observations. We then drew, at random, five million observations. Of these observations, we kept only those associated with securities for which at least 500 daily observations remained; this criterion eliminated 303,746 observations. The resulting sample includes 4,696,254 daily returns observations on 3,050 securities, for an average of 1,540 observations per security.

We then calculated security-specific market-return coefficient (commonly known as the “beta”) from a simple market model including a constant and the market return variable. The daily fitted excess return is the difference between the actual daily return and its predicted value based on the market model, described in Section 4. The sample mean of fitted excess returns in our sample is 0 by construction. The sample standard deviation of all fitted excess returns is 0.040: shifting the distribution one standard deviation to the left entails a reduction in a security’s value by roughly 4%. In much of our analysis, we standardize fitted excess returns by the standard deviation of firm-specific excess returns. This standardization imposes mean-zero, standard deviation-one fitted excess returns at the firm level, facilitating comparisons both across firms and to the standard normal distribution.

#### 4. The Basic Framework With One Firm and One Event

We begin our discussion in Section 4.1 by introducing the standard model for daily securities returns; throughout, we use the terms “firm,” “stock,” and “security” interchangeably. We focus in this section on the case in which there is a single firm and a single event (we discuss extensions to

---

These results were qualitatively very similar to those reported below, so we do not include them in the paper; we will provide these results on request. In addition, we found that adding one or two lags of the market return to the market model—one way to address nonsynchronous trading—does not eliminate the substantial variation across firms in sample quantiles of the firm-specific excess distributions, which is the key driver of our empirical results.

the multiple-firm and multiple-event cases in Section 7). In Section 4.2, we describe the standard approach to inference in event studies and detail the inference problem that plagues it in the one-firm, one-event case.

#### 4.1. The Basic Framework

Daily event studies involve a security's daily return; we focus on the log form, defined as

$$r_s \equiv \ln P_s - \ln P_{s-1}, \quad (1)$$

where  $P_s$  is the firm's stock price on day  $s$ .<sup>13,14</sup> Event studies typically use a model like this one for firms' daily returns:

$$r_s^j = X_s^j \beta^j + A_s^j, \quad (2)$$

where the superscript  $j$  indexes firms, the row-vector of data  $X_s^j$  includes 1 and a measure of the market return for day  $s$ , and possibly other variables that might vary by firm;  $\beta^j$  is a vector of parameters that must be estimated; and  $A_s^j$  is firm  $j$ 's day- $s$  excess return, the component of the observed return that cannot be explained by  $X_s^j$  given the value of  $\beta^j$ . Other variables sometimes included in so-called factor models are measures of firm size, the firm's book-to-market equity, and momentum.<sup>15</sup> For exposition, we focus on the simple market model here, so we include only the CRSP value-weighted portfolio as a nonconstant regressor. For concreteness, write the market return variable as  $r_{\text{mkt},s}$ , so that  $X_s = [1, r_{\text{mkt},s}]$ . For reference, Table 1 lists and defines the notation just described, as well as other notation introduced below. For the moment, we focus only on (2) as applied to a single firm, so we suppress the superscript  $j$ .

To account for the possibility of event effects, suppose that we have data on  $r_s$  and  $r_{\text{mkt},s}$  for dates  $s = 1, 2, \dots, n$ . On date  $e = n + 1$ , an event occurs.

13. We have omitted notation concerning split factors and dividends from (1); the CRSP data we use do account for these factors.

14. Our results were qualitatively similar whether we used as the dependent variable  $r_s$  or  $R_s \equiv (P_s - P_{s-1})/P_{s-1} = \exp[r_s] - 1$ , so we report only the log-form results. We refer readers to our earlier working paper, Gelbach et al. (2011), for results using  $r_s$  as the dependent variable.

15. The three-factor model includes size and book-to-market variables in addition to the market return (see Fama and French, 1992; Fama and French, 1993). Carhart (1997) added a momentum variable, resulting in the four-factor model.

**Table 1.** List and Description of Variables Used

$n$	Number of pre-event observations in a generic event study
$s, e$	Generic date ( $s$ ), and event date ( $e = n + 1$ )
$r_s^j$	Firm $j$ 's return on date $s$
$r_{\text{mkt},s}$	Market return on date $s$
$X_s$	The row vector $(1, r_{\text{mkt},s})$
$\beta^j = (\beta_0^j, \beta_1^j)'$	Usual column vector of coefficients
$D_s$	Dummy variable indicating whether $s = n + 1 = e$ (the event date)
$\gamma^j$	Event effect for firm $j$
$n_j$	Number of pre-event observations on firm $j$ in our full sample, $\hat{S}_j$
$A_s^j$	Firm $j$ 's composite excess return, including any event effect: $A_s^j = D_s^j \gamma + a_s^j$ , for $s = 1, 2, \dots, n_j + 1$
$a_s^j$	Excess return unrelated to event effect, $s = 1, 2, \dots, n_j + 1$
$\hat{a}_s^j$	Fitted excess return: OLS estimate of $a_s^j$
$\hat{u}_s$	Standardized fitted excess return: $\hat{a}_s^j / \hat{\sigma}_a^j$
$F^j$	True distribution of $a_s^j$
$\hat{F}^j$	Empirical distribution function of $\hat{a}_s^j$ in a sample of size $n$
$\hat{F}_{n_j}^j$	Empirical distribution function of $\hat{a}_s^j$ in full sample of $n_j$ observations (i.e., $\hat{S}_j$ )
$G^j, \hat{G}^j, \hat{G}_{n_j}^j$	Analogous to $F^j, \hat{F}^j$ , and $\hat{F}_{n_j}^j$ except that we define the distributions using standardized fitted excess returns, $\hat{u}_s^j \equiv \hat{a}_s^j$ .
$\hat{S}_j$	Sample of all $n_j$ observations on $\hat{a}_s^j$ : $\hat{S}_j = \{\hat{a}_s^j\}_{s=1}^{n_j}$
$\hat{\beta}^j, \hat{\gamma}^j$	OLS estimates of $\beta^j$ and $\gamma^j$
$\sigma_a^j, \hat{\sigma}_a^j$	Standard deviation of $a_s^j$ and usual estimate based on OLS market model estimation
$\sigma_\gamma^j, \hat{\sigma}_\gamma^j$	True standard error of $\gamma^j$ and usual estimate based on OLS market model estimation
$\alpha$	Desired (nominal) significance level
$z_\alpha$	$\alpha$ -quantile of standard normal distribution
$y_\alpha^j$ and $w_\alpha^j$	$\alpha$ -quantiles of $F^j$ and $G^j$
$\hat{y}_\alpha^j$ and $\hat{w}_\alpha^j$	Sample $\alpha$ -quantiles of $\hat{F}^j$ and $\hat{G}^j$
$\lceil x \rceil$	Integer $c$ such that $x - 1 < c \leq x$ .
$\bar{\rho}_j$	Monte Carlo rejection rate for firm $j$ given desired $\alpha$
$c(\alpha, n)$	Index of order statistic corresponding to sample $\alpha$ -quantile from sample of size $n$ : $c(\alpha, n) = \lceil \alpha \times (n + 1) \rceil$ .

To account for event effects, we re-write the day- $s$  excess return variable as  $A_s = D_s \gamma + a_s$ , where  $D_s$  is the event dummy variable (i.e.,  $D_s = 1(s = e)$ ), and  $1(\cdot)$  is the indicator function that equals one when its argument is true and zero otherwise. The parameter  $\gamma$  is the true effect of an event on the level of the firm's daily return, which could be either positive, negative, or zero. The  $a_s$  term represents the part of the excess return that is unrelated to

the event, so that  $A_s$  can be viewed as an event date-specific location shift of  $a_s$ , with  $\gamma$  being the shift parameter.

We follow common practice and assume that all excess returns for pre-event dates are iid conditional on the full set of regressors and come from the same distribution, which we name  $F$ .<sup>16</sup> Below, we will allow these true excess returns distributions to vary across firms, in which case we call firm  $j$ 's distribution  $F^j$ . Throughout, we will assume that excess returns are continuously distributed, so that  $F$  is strictly increasing on its entire support. We refer to the standard deviation of the excess returns distribution  $F$  as  $\sigma_a$ . Given the common assumption that events affect only the level of the daily return, it follows that  $a_e$ , the part of the event-date excess return that is unrelated to the event, also has distribution  $F$ . In sum, for any value  $y$ , we have the model

$$\begin{aligned} r_s &= X_s \beta + D_s \gamma + a_s, & \Pr(a_s \leq y | \text{data}, ) &= F(y), \\ s &= 1, 2, \dots, n + 1. \end{aligned} \tag{3}$$

#### 4.2. The Standard Approach to Inference

Using the so-called regression approach to estimating event effects, a researcher estimates  $\beta$  and  $\gamma$  jointly using ordinary least squares (OLS) estimation. She then evaluates the event's effect by testing the null hypothesis  $H_0 : \gamma = 0$  against some alternative hypothesis; for exposition, we focus on the lower-tailed case, in keeping with our example of testing whether a corrective disclosure reduces a firm's market valuation.

The standard approach to carrying out hypothesis tests involves estimating (3) by OLS and comparing the usual  $t$ -statistic for  $\hat{\gamma}$  to critical values based on the standard normal distribution (or the Student's  $t$  distribution with the appropriate degrees of freedom, though with large  $n$ , the difference will be trivial). Let  $\hat{\beta}$  and  $\hat{\gamma}$  be the OLS estimates of  $\beta$  and  $\gamma$ , and let  $\hat{\sigma}_\gamma$  be the square-root of the estimated variance of  $\hat{\gamma}$ . Then the  $t$ -statistic,  $\hat{t}$ , for testing the null hypothesis of no effect is the ratio of  $\hat{\gamma}$  to  $\hat{\sigma}_\gamma$ . A researcher using the standard approach will reject the null hypothesis against the alternative

16. Results in Andrews (2003) and Conley and Taber (2011) show that the SQ test retains its good properties under a wide class of non-iid processes.

hypothesis of a negative effect, at level  $\alpha$ , if  $\hat{t}$  is less than the  $\alpha$ -quantile of the standard normal distribution,  $z_\alpha$ .

In conventional settings, this procedure “works,” in the sense of having either finite-sample or asymptotic Type I error rate equal to  $\alpha$ , for one of two reasons. First, if  $a_s$  is normally distributed, then it can be shown that  $\hat{\gamma}$  is normal as well. In this case,  $\hat{t}$  has a Student’s  $t$  distribution with  $n - 2$  degrees of freedom under the null hypothesis that  $\gamma = 0$ .

Second, if  $\hat{\gamma}$  is not exactly normal, but  $\sqrt{n}(\hat{\gamma} - \gamma)$  has an asymptotically normal distribution, then the distribution of  $\hat{t}$  under  $H_0$  is well approximated by the standard normal distribution when  $n$  is reasonably large. In such cases, it is common and generally appropriate to treat  $\hat{t}$  as if it were standard normal. The statistical argument for this result requires the applicability of a central limit theorem (CLT) to the behavior of the parameter  $\hat{\gamma}$ . CLT results typically hold in econometrics applications because a statistic can be written as a sample mean of a large number of observations, since sample means are asymptotically normal under extremely broad conditions. But when there is only one event date, or if separate effects are estimated for each of a set of multiple events, then  $\hat{\gamma}$  cannot be written as a sample mean of many observations. Instead, under the usual assumptions, the asymptotic distribution of  $(\hat{\gamma} - \gamma)$  is the same as the distribution of  $a_e$ , the true event-date excess return. Thus, when the null hypothesis of zero event effect is true, so  $\gamma = 0$ , and the number of pre-event observations is large,  $\hat{\gamma}$  is approximately distributed according to the distribution  $F$ . Because the proof of this fact is detailed and depends entirely on well-known facts, we relegate it to Appendix A.3.

The fact that  $\hat{\gamma} - \gamma$  has asymptotic distribution  $F$  means that when we have a large number of pre-event observations,  $\Pr(\hat{\gamma} \leq y)$  is approximately equal to  $F(y - \gamma)$ . When the null hypothesis is true,  $\gamma$  is zero, and thus the event effect estimated using least-squares estimation of (3) behaves just like a random draw from the firm’s excess returns distribution  $F$ , given large  $n$ .

Next, consider the  $t$ -statistic. Because  $\hat{\sigma}_\gamma$  is consistent for  $\sigma_\gamma$ , the asymptotic distribution of the conventional  $t$ -statistic is the same as the distribution of  $(a_e + \gamma)/\sigma_a$ . Thus, for large  $n$ , the  $t$  statistic will behave like a scaled and location-shifted random draw from the distribution of standardized excess returns. For any  $w$  in the support of this scaled-shifted distribution,



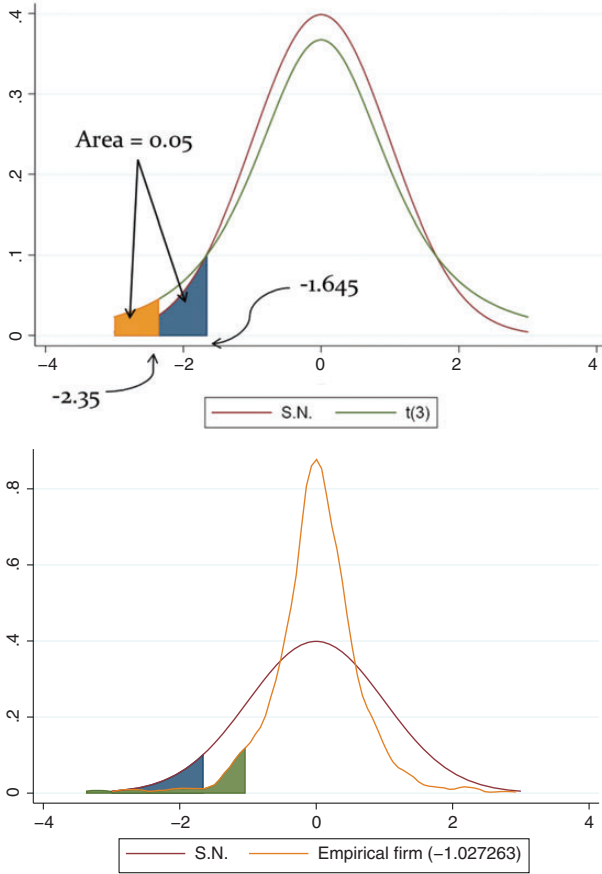
we thus have

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr(\hat{t} \leq w) &= \Pr\left(\frac{a_e + \gamma}{\sigma_a} \leq w\right) \\ &= \Pr\left(\frac{a_e}{\sigma_a} \leq w - \frac{\gamma}{\sigma_a}\right) \\ &= F(\sigma_a w - \gamma). \end{aligned} \quad (4)$$

This result means that the  $t$ -statistic inherits its large-sample distributional properties from the distribution of a single random draw from the excess returns distribution. Notably, the large-sample behavior of the  $t$ -statistic generally will be normal if and only if the excess returns distribution  $F$  is itself normal. Thus, the standard approach has asymptotically correct Type I error rate only in the special case of normally distributed excess returns.

In Appendix A, we test two null hypotheses concerning the distributions of firm-specific excess returns in our data. We first test and easily reject the null hypothesis that all 3,050 firms' excess returns come from the same normal distribution. We then test whether each firm's excess returns come from a normal distribution, with the variances allowed to differ across firms. Again we reject easily; few firms' distributions are consistent with normality. These results imply that the standard approach will yield erroneous Type I error rates for at least some, and possibly all, firms. The direction and magnitude of these size errors can be estimated consistently using the sample quantiles we discuss in Section 5, though these errors would be unknown using only the results provided via the standard approach.

What can be said, as an analytical matter, about the standard approach's Type I error rate in the general case of nonnormal excess returns? Under the null hypothesis,  $\gamma = 0$ , so as  $n$  grows, we have that  $\Pr(\hat{t} \leq z_\alpha)$  converges to  $F(\sigma_a z_\alpha)$ . Define the  $\alpha$ -quantile of the true excess returns distribution as  $y_\alpha$ , so that by construction,  $\Pr(a_s \leq y_\alpha) = F(y_\alpha) = \alpha$ . When  $y_\alpha < \sigma_a z_\alpha$ , a lower-tailed test using the standard approach will reject more than  $100 \times \alpha\%$  of the time, even with large  $n$ . When the opposite holds, the standard approach will reject less than that percentage. Only when  $y_\alpha$  happens to equal  $\sigma_a z_\alpha$  will the standard approach reject at the desired Type I error rate  $\alpha$ .



**Figure 1.** Illustrating Over- and Under-Rejection With Nonnormality.

For concreteness, Figure 1 illustrates the two possible cases when  $y_\alpha \neq \sigma_\alpha z_\alpha$ . In the top panel, we illustrate the case  $y_\alpha < \sigma_\alpha z_\alpha$ , plotting the density from a standard normal distribution together with the density from a Student's- $t$  distribution with three degrees of freedom. The standard normal's 0.05-quantile is  $-1.645$ , while the  $t(3)$  distribution's 0.05-quantile is  $-2.35$ . Thus, if excess returns follow a  $t(3)$  distribution but one uses the  $-1.645$  critical value for  $\hat{t}$ , one will reject the null hypothesis of no event effect greater than five percent of the time. In this case, the standard approach will over-reject. This situation is good for plaintiffs and bad for

defendants, since it would lead fact finders to erroneously accept the plaintiffs' position more often than the intended standard.

In the figure's bottom panel, we illustrate the case when  $y_\alpha > \sigma_a z_\alpha$ . We again plot the density from the standard normal distribution and then add the density of standardized fitted excess returns that we estimated for one of the firms in our sample, using Stata's *kdensity* command. To construct this density estimate, we first calculated  $\hat{a}_s$  for all 882 dates  $s$  for which we have data for this firm. We then constructed the standardized fitted excess return for generic date  $s$   $\hat{u}_s = \hat{a}_s / \hat{\sigma}_a$ , where  $\hat{\sigma}_a$  is the sample standard deviation of  $\hat{a}_s$ . We will refer to the distribution of the corresponding standardized residual  $u_s \equiv a_s / \sigma_a$  as  $G$ . Thus, by construction we have  $G(w) = \Pr(u_s \leq w) = \Pr(a_s \leq \sigma_a w) = F(\sigma_a w)$ ; equivalently,  $F(y) = G(y / \sigma_a)$ . If the firm's true excess returns came from a normal distribution, then the quantiles of  $G$  would be the standard-normal quantiles, and  $\hat{u}_s$  would follow a Student's- $t$  distribution with degrees of freedom equal to the number of observations minus one. Given the large number of observations we have on the firm in the picture, the standard normal 0.05-quantile would be a good approximation to the true 0.05-quantile of  $\{\hat{u}_s\}$  under normality of  $F$ . However, our estimated sample-0.05 quantile was  $-1.027$ , much closer to the origin than the standard normal distribution value of  $-1.645$ . As the figure shows, using the standard normal critical value of  $-1.645$  would yield a Type I error rate considerably below 0.05 for this firm.<sup>17</sup>

Below, we provide evidence that in the CRSP data for 2000–2007, such systematic under-rejection is important. Such a result is beneficial in one sense, since reducing Type I errors is generally a good thing. However, it comes at a cost. As we show in Section 6.2, fact finders using the standard approach with a firm like the one in the bottom panel of Figure 1 would very infrequently accept plaintiffs' position even when it is correct. That is, for a firm like this one, the Type II error rate will tend to be very high, so that power is low.<sup>18</sup>

17. We emphasize that we chose the firm in this picture precisely because its sample 0.05-quantile is especially close to the origin, allowing easy visualization of the source of under-rejection when using the standard approach. Among the firms in our sample, 99% had a sample 0.05-quantile of standardized fitted excess returns to the left of this firm's value of  $-1.027$ .

18. The optimal tradeoff between size and power from the point of view of securities law is an interesting question, though one that lies beyond the scope of this paper.

As a first pass at quantifying the importance of nonnormality in evaluating the standard approach's Type I error rate, we can estimate the asymptotic Type I error rate of the standard approach, under the assumption that all 3,050 securities' returns come from the same distribution (i.e.,  $F^j = F$  for every  $j = 1, 2, \dots, 3050$ ). Under this null hypothesis, the empirical distribution function of our pooled sample of excess returns,  $\hat{F}^{\text{pooled}}$ , is consistent for  $F$ . As above, the standard approach's true asymptotic Type I error rate for a level- $\alpha$  test against the lower-tailed alternative is  $F(\sigma_a z_\alpha)$ . The sample standard deviation of our excess returns sample is 0.040, so for  $\alpha = 0.05$ , our estimate of  $\sigma_a z_\alpha$  is  $0.040 \times -1.64 = -0.066$ . The asymptotic Type I error rate for a level-0.05 test based on the standard approach is consistently estimated by  $\hat{F}^{\text{pooled}}(-0.066)$  (i.e., the fraction of estimated standardized fitted excess returns that are less than  $-0.066$ ). Of the 4,696,254 excess returns in our pooled sample, 145,836 are less than or equal to  $-0.066$ . Our estimate of the Type I error rate based on a desired level-0.05 test using the standard approach on pooled data is thus  $145836/4696254 = 0.031$ . In other words, given the assumption of a common excess returns distribution, with data pooled across firms, the standard approach rejects only about 60% as often as the desired level of 0.05. Similar calculations show actual rejection rates of 0.022 and 0.049 for desired significance levels  $\alpha = 0.025$  and 0.10.

The size distortions for  $\alpha = 0.05$  and 0.10 raise serious concerns for the standard approach. However, we arrived at them only after imposing the assumption that all 3,050 of the securities in our sample have the same underlying distribution of excess returns (i.e.,  $F^j = F$  for all  $j = 1, 2, \dots, 3050$ ). We show in Appendix A that formal testing clearly rejects this assumption. We also show that evidence rejects the null hypothesis that all firm-specific excess returns distributions are normal. Thus, we now provide evidence allowing for variation across firms in their excess returns distributions.

### 4.3. Estimated Type I Error Rates Using the Standard Approach

As above, we will work with the standardized fitted excess returns, defined as  $\hat{u}_s^j = \hat{a}_s^j / \hat{\sigma}_a^j$ . Let  $G^j$  be the true distribution of standardized excess returns for firm  $j$ . If we write the  $\alpha$  quantile of this distribution as

$w_\alpha$ , then by definition of quantiles we have  $G^j(w_\alpha) = \alpha$ . Because the standard approach rejects whenever the  $t$ -statistic is less than  $z_\alpha$ , it has true Type I error rate  $G^j(z_\alpha)$ . In general,  $G^j(z_\alpha) = \alpha$  only if  $G^j$  is normal, which is easy to reject for nearly all firms in our sample.<sup>19</sup>

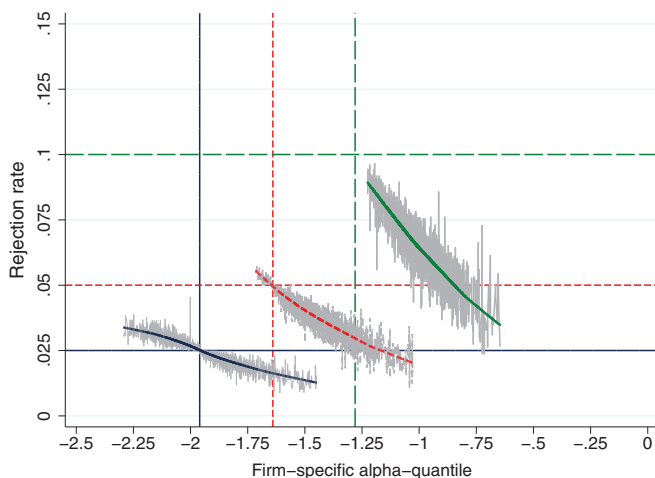
We can estimate the firm-specific error rate directly using our sample data. Denote the empirical distribution function associated with our sample of  $n_j$  standardized fitted excess returns for firm  $j$  as  $\hat{G}_{n_j}^j(w)$ . This function tells us the fraction of standardized fitted excess returns observations in our sample whose value does not exceed the level  $w$  (i.e., the sample Type I error rate using the standard approach). For clarity, we use the notation  $\hat{\alpha}_{sa}^j(\alpha) \equiv \hat{G}_{n_j}^j(z_\alpha)$  to refer to the observed error rate for security  $j$ , using the standard approach, when the desired error rate is  $\alpha$ .

To estimate the asymptotic Type I error rate of the standard approach for each firm given  $\alpha = 0.05$ , we thus use  $\hat{\alpha}_{sa}^j(0.05) = \hat{G}_{n_j}^j(-1.645)$ , which is just the fraction of firm  $j$ 's standardized fitted excess returns that are more negative than  $-1.645$ . The analytical discussion above suggests that the estimated Type I error rate for the standard approach should be higher for values of  $y_{0.05}^j$  that are further to the left of the origin (i.e., very negative); lower for values of  $y_{0.05}^j$  closer to the origin; and should generally fall as  $y_{0.05}^j$  increases toward the origin.

In Figure 2, we plot the estimated asymptotic Type I error rates at three desired significance levels,  $\alpha \in \{0.025, 0.05, 0.10\}$ ; these significance levels are indicated by horizontal lines. We indicate the standard normal distribution quantiles corresponding to these levels using vertical lines at  $-1.960$ ,  $-1.645$ , and  $-1.282$ . The lighter, jagged lines plot the estimated asymptotic Type I error rate for the standard approach at each  $\alpha$  (i.e.,  $\hat{\alpha}_{sa}^j(\alpha)$ ), the share of each firm's standardized fitted excess returns that fall below the standard normal distribution's  $\alpha$ -quantile. The darker lines plot smoothed, nonparametric (lowess) estimates of the average rejection rate at each value of the firm-specific sample  $\alpha$ -quantile, which is given by the horizontal axis.

We note that for all three choices of  $\alpha$ , we have included data on only those firms whose  $\hat{w}_{0.05}^j$  lies in the middle 98% of the cross-firm distribution of  $\hat{w}_\alpha^j$ ; this sample restriction avoids visual noise related to extreme outliers.

19. It is of course possible that two different distributions will have some  $\alpha$ -quantiles in common, but unless the distributions are the same, that will happen only by accident in practice and is not predictable *a priori*.



**Figure 2.** Asymptotic Type I Error Rates for the Standard Approach at Levels  $\alpha \in \{0.025, 0.05, 0.10\}$ .

The left-most jagged series plots the estimated asymptotic Type I error rate for the standard approach at desired significance level  $\alpha = 0.025$ , with the middle and right-most series plotting the estimated asymptotic Type I error rates for  $\alpha = 0.05$  and  $0.10$ .

The average value of  $\hat{\alpha}_{sa}^j(0.025)$  across our 3,050 firms is 0.0248, essentially equal to the desired level of 0.025. Based on this result, most researchers would conclude that the standard approach performs well for  $\alpha = 0.025$ . However, there is heterogeneity across firms' quantiles (as expected given our results in Appendix A). Roughly half of firms (1,520 of 3,050) in our sample have  $\hat{w}_{0.025}^j < -1.96$ , so that their estimated asymptotic Type I error rate exceeds 0.025, with the other half having estimated asymptotic Type I error rates below  $\alpha = 0.025$ .

Turning to  $\alpha = 0.05$ , the average value of  $\hat{\alpha}_{sa}^j(0.025)$  across our 3,050 firms is 0.038, which is a nontrivial average size distortion. Only 4.7% of firms (144 of 3,050) have sample 0.05-quantile to the left of  $-1.64$ . Thus, the standard approach would under-reject a true null of zero effect for the vast majority of firms.

Finally, we consider  $\alpha = 0.10$ . The average value of  $\hat{\alpha}_{sa}^j(0.10)$  across our 3,050 firms is only 0.065. In addition, only four of our 3,050 firms have a sample 0.10-quantile below  $z_{0.10} = -1.28$ . This explains why the entire

upper series lies to the right of the vertical line at  $-1.28$ , since we have graphed the rejection rate for only the middle 98% of firms as measured by firm-specific sample 0.10-quantile. As a consequence, every firm represented in the graph has an estimated asymptotic Type I error rate below the desired level of 0.10.

In sum, the evidence shows that the standard approach leads to substantial under-rejection for levels 0.05 and 0.10, with actual size below its desired level for the vast majority of firms. For level 0.025, the standard approach leads to correct size when averaging across firms, but the actual firm-specific Type I error rate can differ substantially from the desired level at  $\alpha = 0.025$ . An important feature of Figure 2 concerns the negative slopes of the smoothed estimated asymptotic Type I error rates. As we noted above, the graph of rejection rates against  $\hat{w}_\alpha^j$  for a procedure with correct size would be a horizontal line at  $\alpha$ , up to sampling error. Not only do all three collections of error-rate lines primarily lie below their desired  $\alpha$  values, the error rates clearly fall as firms' sample  $\alpha$ -quantiles rise toward zero.

## 5. The SQ Test

In this section, we explain the SQ test. We then turn to Monte Carlo and other empirical evidence to assess the SQ test's performance in typically sized samples.

### 5.1. Deriving and Characterizing the Test

In Section 4, we explained why the estimated coefficient on the event dummy,  $\hat{\gamma}$ , converges in probability to  $\gamma + a_e$ .<sup>20</sup> As a consequence,  $\hat{\gamma} - \gamma$  has the same asymptotic distribution,  $F$ , as the event-date excess return,  $a_e$ . Equation (4) thus shows that under the null hypothesis of no event effect,  $\lim_{n \rightarrow \infty} \Pr(\hat{\gamma} \leq y) = F(y) = \Pr(a_e \leq y)$ . The SQ test is based on this simple but powerful fact.

Suppose momentarily that we knew the value of the quantiles of  $F$ . That is, for any  $\alpha$  between 0 and 1, suppose we could determine  $y_\alpha$  that satisfies

$$\Pr(a_e \leq y_\alpha) = F(y_\alpha) = \alpha. \quad (5)$$

20. We drop the notation indexing firms for simplicity throughout this section.

Since  $\hat{\gamma}$ 's asymptotic null distribution is  $F$ , we could then use  $y_\alpha$  as a critical value to test  $H_0$  at level  $\alpha$ : we would reject the null against the lower-tailed alternative whenever  $\hat{\gamma} \leq y_\alpha$ . The key challenge to asymptotically valid inference at level  $\alpha$  is therefore to find a way to consistently estimate the  $\alpha$ -quantile of  $F$ . Under the assumptions commonly made in event studies, this is a surprisingly simple task, requiring only a trivial amount of additional work beyond that necessary for the standard approach. The following procedure characterizes the SQ test, which achieves asymptotic Type I error rate equal to  $\alpha$  in testing  $H_0$  against the lower-tailed alternative  $H_l$ . Note that we define the procedure in terms of nonstandardized fitted excess returns (i.e.,  $\hat{a}_s$ ) rather than  $\hat{u}_s$ . Because standardizing does not change the order statistics of a sample, the procedure involves identical steps and results when it is used with standardized fitted excess returns.

- PROCEDURE 1 (The SQ Test Against  $H_l : \gamma_l < 0$ )
1. Estimate  $\hat{\beta}$  and  $\hat{\gamma}$  using OLS estimation of the market model (3).
  2. For each nonevent date  $s \in \{1, 2, \dots, n\}$ , calculate the fitted excess return  $\hat{a}_s = r_s - X_s \hat{\beta}$ .
  3. Sort  $\hat{a}_s$  from least to greatest. Let the  $i$ th order statistic be written  $\hat{a}_{(i)}$ , so that

$$\hat{a}_{(1)} \leq \hat{a}_{(2)} \leq \dots < \hat{a}_{(n)}.$$

4. Next, define the ceiling operator  $\lceil \cdot \rceil$  such that  $\lceil x \rceil$  returns the integer  $c$  with the property that  $x < c \leq x + 1$ . Define  $c(\alpha, n) = \lceil \alpha \times n \rceil$ , and find the  $c(\alpha, n)$  order statistic of  $\{\hat{a}_s\}$ ; call this value  $\hat{y}_\alpha$ , and note that it is the sample  $\alpha$ -quantile of the realized  $\hat{a}_s$  values. For example, in the case of  $\alpha = 0.05$  and  $n = 100$ , we have  $c(0.05, 100) = 5$ , so we find the 5th least (i.e., most negative) value of  $\hat{a}_s$ .
5. Reject  $H_0$  against  $H_l$  if and only if  $\hat{\gamma} < \hat{y}_\alpha$ .

It can be shown that as the number of pre-event observations  $n$  grows, the probability that Procedure 1 rejects  $H_0$  when it is true converges to  $\alpha$ , which confirms that the SQ test has asymptotically correct size. We refer readers interested in a formal proof to Proposition 2 of Conley and Taber (2011) or to our earlier paper, Gelbach et al. (2011). It will be helpful, though, to



give a heuristic discussion of the result. Define the empirical distribution function  $\hat{F}$  based on nonevent date data:

$$\hat{F}(y) = \frac{1}{n} \sum_{s=1}^n 1(\hat{a}_s \leq y). \quad (6)$$

The empirical distribution function,  $\hat{F}$  tells us the share of all fitted excess returns that are no greater than some arbitrarily chosen  $y$ . Note that  $\hat{F}$  involves sample, rather than “true”, information about excess returns in two ways. First,  $\hat{F}$  is defined using observations on fitted, rather than true, excess returns. Second, we have only a sample of  $n$  pre-event dates, rather than the entire population of excess returns values. Thus, even if we knew each  $a_s$ , we would not know the population distribution  $F$ , but rather only the empirical distribution function using true, rather than fitted, excess returns, defined as

$$F_n(y) = \frac{1}{n} \sum_{s=1}^n 1(a_s \leq y), \quad (7)$$

To make these two sources of error more explicit, observe that we can decompose the error we commit in using  $\hat{F}(y)$  rather than  $F(y)$  as follows:

$$\hat{F}(y) - F(y) = \underbrace{\hat{F}(y) - F_n(y)}_{\text{Error 1}} + \underbrace{F_n(y) - F(y)}_{\text{Error 2}}. \quad (8)$$

Error 1 arises due to the fact that we use  $\hat{a}_s$  rather than  $a_s$ , given that we have only a sample. This error vanishes as  $n$  grows, because  $\hat{a}_s$  and  $a_s$  differ only according to whether  $X'\hat{\beta}$  or  $X'\beta$  is subtracted from the firm’s observed daily return, and since  $\hat{\beta} \xrightarrow{p} \beta$ , this difference is irrelevant asymptotically. Error 2 arises because even if we could observe each day’s excess return, we would have only a sample of  $n$  observations; thus, even then we could observe only  $F_n$  rather than the population distribution  $F$ . As a consequence of the Glivenko–Cantelli theorem, Error 2 also converges to zero (e.g., see [van der Vaart, 1998](#)). Because convergence of a cumulative distribution function and convergence of its associated quantiles are equivalent, convergence of  $\hat{F}(y)$  to  $F(y)$  under the null hypothesis is sufficient for each sample quantile  $\hat{y}_\alpha$  to converge to its corresponding population quantile  $y_\alpha$ . It then follows that  $\hat{F}(\hat{y}_\alpha) \xrightarrow{p} F(y_\alpha) = \alpha$ .

This result implies that researchers can use the SQ test to fix the asymptotic Type I error rate of the test for a zero event effect at any chosen level. The procedure thus provides the basis for asymptotically valid inference, even though the procedure amounts to little more than sorting some fitted values.<sup>21</sup>

## 5.2. Monte Carlo Evidence on the Small-Sample Size of the SQ Test

Given that the SQ test has asymptotically correct size, a natural next question concerns its small-sample behavior. It is always possible that an asymptotically justified method requires enormous sample sizes in practice, and event studies typically involve large but not massive samples. We thus present Monte Carlo evidence in this section for pre-event samples of size  $n = 100$ .

We investigate the SQ test's small-sample performance with significance levels  $\alpha = 0.025, 0.05,$  and  $0.10$ , though for simplicity we explain the procedure only once, using  $\alpha = 0.05$ . Our Monte Carlo experiment consists of  $m = 1, 2, \dots, 1,000$  repetitions of the following procedure:

1. For each firm  $j$ , we draw  $n + 1 = 101$  values from the  $n_j$  observed values of  $(r_s^j, X_s)$  in our data; we use random sampling with

---

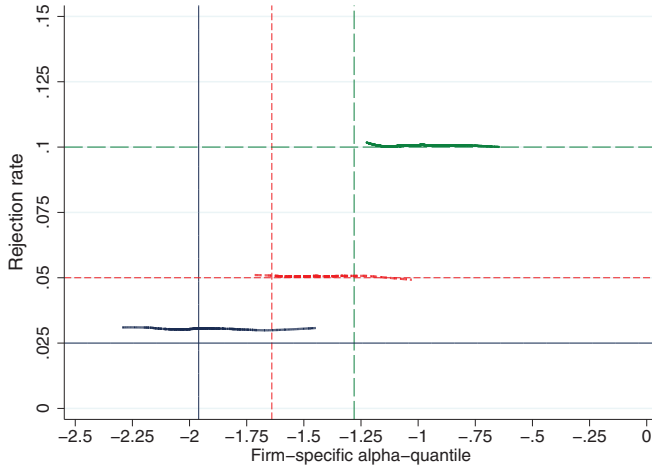
21. An interesting potential extension to the analysis below would involve nonlinear generalizations of (3). An example would be to modify the main equation in (3) to be  $H(y_s, X_s, \theta) = a_s$  for some outcome variable  $y_s$ , regressors  $X_s$ , parameter vector  $\theta$ , nonlinear function  $H$ , and residual  $a_s$ . If  $H$  is known and continuous, and if  $\theta$  can be consistently estimated, then by the continuous mapping theorem,  $\hat{a}_s - a_s = H(y_s, X_s, \hat{\theta}) - H(y_s, X_s, \theta)$  converges in probability to zero; therefore, the asymptotic equivalence lemma (White, 2001, Lemma 4.7, p. 67) implies that for the event date,  $\hat{a}_e$  and  $a_e$  will have the same asymptotic distribution. We conjecture that in this situation, the sample quantiles of the pre-event residuals will also be consistent. Thus, we believe the SQ test should be valid for this class of nonlinear models (concerning this generalized model, see also Dufour et al., 1994). Even if  $H$  is unknown, finite-dimensional  $\theta$  often is still consistently estimable via semi-parametric methods (see, e.g., Horowitz, 1998), so  $a_s = H(y_s, X_s, \theta)$  should be consistently estimable under suitable regularity conditions; thus the SQ test might sometimes be valid then, too. A final question is whether the SQ test might be valid when  $\theta$  is infinite-dimensional, so that the model is nonparametric. We have not investigated this case, which would involve some technical issues, but we would not be surprised if forms of the SQ test were appropriate in at least some such cases.

- replacement to make these draws. Observation  $s$  for firm  $j$  on Monte Carlo repetition  $m$  is  $(r_{s,m}^j, X_{s,m})$ ,  $s \in \{1, 2, \dots, n+1\}$ , where  $X_{s,m} = (1, r_{\text{mkt},s,m})$ . For all firms, we set the “event” dummy  $D_{s,m}$  equal to 1 for  $s = n+1$  and 0 for  $1 < s \leq n$ .
2. We estimate the model in (3). The fitted excess return for day  $s \leq n$  in Monte Carlo iteration  $m$  is  $\hat{a}_{s,m}^j = r_{s,m}^j - X'_{s,m} \hat{\beta}^{j,m}$ , where  $\hat{\beta}^{j,m}$  is the OLS estimate of  $\beta$  for firm  $j$  on the  $m$ th Monte Carlo repetition. The corresponding estimated event effect can be calculated as  $\hat{\gamma}^{j,m} = r_{101,m}^j - X'_{101,m} \hat{\beta}^{j,m}$  (see Appendix A.3).
  3. We calculate the sample 0.05-quantile,  $\hat{\gamma}_{0.05}^{j,m}$ , of the first  $n = 100$  realizations of  $\hat{a}_{s,m}^j$ . This is the fifth order statistic of the fitted residuals for the  $j$ th firm on the  $m$ th Monte Carlo repetition. For Monte Carlo repetition  $m$ , we reject the null hypothesis against a lower-tailed alternative based on the SQ approach if and only if the estimated event effect is less than the sample quantile (i.e.,  $\hat{\gamma}^{j,m} < \hat{\gamma}_{0.05}^{j,m}$ ). We set  $\rho_{j,m} = 1$  if we reject for firm  $j$  on Monte Carlo iteration  $m$ , and we set  $\rho_{j,m} = 0$  if we do not reject.

For each security  $j$ , we then calculate the Monte Carlo rejection rate (MCRR) over the 1,000 repetitions of this experiment (i.e.,  $\bar{\rho}_j = 1000^{-1} \sum_{m=1}^{1000} \rho_{j,m}$ ). If a test has correct size for a given  $j$ , then it should exhibit a rejection rate of 0.05, up to Monte Carlo simulation error. Using only 1,000 Monte Carlo repetitions may yield an imprecise rejection rate for any given  $j$ , so we use two approaches to dealing with this imprecision.

First, we can investigate the grand mean Monte Carlo rejection rate when we pool over all 3,050 firms (i.e.,  $\bar{\rho} \equiv 3050^{-1} \sum_j \bar{\rho}_j$ ). For  $\alpha \in \{0.05, 0.10\}$ , the Monte Carlo estimates of  $\bar{\rho}$  are 0.051 and 0.101, very close to the desired levels; we discuss the case of  $\alpha = 0.025$  below. Based on these pooled-across-firm results, 100 pre-event observations is a large enough value of  $n$  for the SQ test to deliver an actual Type I error rate that is essentially indistinguishable from  $\alpha \in \{0.05, 0.10\}$  for practical purposes.

Second, we again use local smoothing techniques to exploit the large number of securities in our sample, which averages out the simulation error. Thus, we expect the graph of the MCRR to be horizontal at approximately level  $\alpha$ , up to simulation error. Figure 3 plots the lowest smoothing of the results for desired significance levels  $\alpha \in \{0.025, 0.05, 0.10\}$ . As



**Figure 3.** Monte Carlo Size Results for the SQ Test,  $\alpha \in \{0.025, 0.05, 0.10\}$ .

with Figure 2 for the standard approach, for reference we include dashed horizontal lines at each desired significance level and vertical lines at the corresponding standard-normal quantiles. We do not plot the underlying, firm-specific Monte Carlo rejection rates.<sup>22</sup>

A first result displayed in Figure 3 is that the estimated Type I error rate in firm-specific samples of size  $n = 100$  does not vary with firm-specific sample  $\alpha$ -quantiles. Thus, the SQ test's rejection probabilities appear to be independent of the  $\alpha$ -quantiles in firm-specific standardized fitted excess returns distributions, both asymptotically and in typical sample sizes. This finding is to be expected, given our theoretical discussion, and it is an important improvement over the standard approach. A second key result in Figure 3 is that the estimated Type I error rate is essentially indistinguishable from  $\alpha$  for  $\alpha \in \{0.05, 0.10\}$ .

For  $\alpha = 0.025$ , Figure 3 suggests that the SQ test rejects more frequently than the desired level. The raw mean rejection rate confirms this suggestion: the SQ test's average Monte Carlo rejection rate across firms is 0.0304 at desired level 0.025. Note, though, that in a sample of size  $n = 100$ ,

22. Even with 1,000 Monte Carlo repetitions, it can be shown that the standard deviation of a single-firm-specific rejection rate across simulations is roughly 0.01 when using  $\alpha = 0.10$ . Thus, plotting roughly 3,000 individual firms' rejection rates yields a lot of visual noise, so we plot only the lowest smoothing.

the critical value used for the SQ test when  $\alpha = 0.025$  is the  $c(0.025, 100) = \lceil 2.5 \rceil$  order statistic. Since  $\lceil 2.5 \rceil = 3$ , the SQ test in a sample with  $n = 100$  uses the same critical value for a level-0.025 test as for a level-0.03 test. This property is not a defect of the test, but rather of the chosen sample size, 100.

Given a desired significance level  $\alpha$ , this example shows the importance of choosing  $n$  so that  $\alpha \times n$  is an integer; when it is not, the SQ test's small-sample performance will yield upward distortions in actual size.<sup>23</sup> To illustrate this point, we ran the same Monte Carlo procedure described above for  $\alpha = 0.025$  and  $n \in \{40, 80, 200\}$ , with only 100 Monte Carlo repetitions in each of these three cases. We chose these pre-event sample sizes because each has the property that  $c(0.025, n)$  is integer-valued:  $c(0.025, 40) = 1$ ,  $c(0.025, 80) = 2$ , and  $c(0.025, 200) = 5$ . The average Monte Carlo rejection rate across firms was 0.0267 for  $n = 40$ , 0.0253 for  $n = 80$ , and 0.0252 for  $n = 200$ , confirming that the SQ test works well when  $c(\alpha, n)$  is integer-valued.

## 6. Asymptotic Power of the SQ Test

We now investigate the asymptotic power properties of the SQ test. In the litigation context, low-power tests have the property that they will fail to reject the null of no event impact even when the true impact on firm value is negative. Thus, power is important in the litigation context because it tells us how often the plaintiffs win when some unlawful action has reduced the value of the security in question. A common measure of a test's power is whether the test is unbiased. An unbiased test has power greater than its size, so that the test rejects more frequently under the alternative than under the null. Unbiasedness is a minimal power criterion: a test that rejects a false null less often than a true one is obviously of limited use. We show below analytically that (i) the SQ test is asymptotically unbiased, while (ii) the

---

23. An alternative approach in noninteger cases would be to interpolate between  $c(\alpha, n)$  and  $c(\alpha, n) - 1$  so as to smooth the rejection rate. This approach requires an interpolation algorithm, which could be difficult to choose in practice, given that excess returns exhibit nonnormality of unknown form. Since  $n$  is usually a choice variable, we do not pursue this approach here.

standard approach may be either asymptotically biased or unbiased, depending on the shape of the excess returns distribution  $F$ .<sup>24</sup>

In addition to proving the analytical results just discussed, in the rest of this subsection we estimate the asymptotic power of the SQ test and the standard approach for each firm in our sample. We show below that in our data, the standard approach has relatively poor power properties by comparison to the SQ test.

### 6.1. Asymptotic Unbiasedness

Consider the SQ test's power against the lower-tailed alternative,  $H_l: \gamma < 0$ . This application would arise in testing whether a firm's corrective disclosure caused a reduction in its stock price, for example. Since the true effect  $\gamma$  is a scalar constant,  $\Pr(\hat{\gamma} \leq y) = \Pr(\hat{\gamma} - \gamma \leq y - \gamma)$ . The SQ test's finite-sample rejection probability given  $\gamma$  and level  $\alpha$  is  $\Pr(\hat{\gamma} - \gamma \leq \hat{y}_\alpha - \gamma)$ . Since  $\hat{y}_\alpha$  is consistent for  $y_\alpha$ , and since the asymptotic distribution of  $\hat{\gamma} - \gamma$  is  $F$ , it follows that  $\Pr(\hat{\gamma} - \gamma \leq \hat{y}_\alpha - \gamma)$  converges to  $F(y_\alpha - \gamma)$ . Therefore, the SQ test's asymptotic rejection probability against a lower-tailed alternative is  $F(y_\alpha - \gamma)$ . Since  $\gamma < 0$  when the alternative hypothesis is true, and since  $F$  is strictly increasing given the assumption of continuously distributed excess returns, we have  $F(y_\alpha - \gamma) > F(y_\alpha) = \alpha$ . As a result, the SQ test's asymptotic probability of rejecting the null against a true lower-tailed alternative is greater than its asymptotic size, which establishes unbiasedness.

Next we determine whether the standard approach is unbiased. Recall that the standard approach rejects  $H_l$  if and only if  $\hat{\gamma} < \hat{\sigma}_a z_\alpha$ , which is asymptotically equivalent to the condition that  $\hat{\gamma} < \sigma_a z_\alpha$ . As above, we work with standardized excess returns, so the standard approach's asymptotic rejection rate given a true effect of  $\gamma$  is given by

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr \left( \frac{\hat{\gamma}}{\hat{\sigma}_a} \leq z_\alpha \right) &= \Pr(a_e \leq \sigma_a z_\alpha - \gamma) \\ &= F(\sigma_a z_\alpha - \gamma). \end{aligned}$$

24. Another, more demanding, power criterion is test consistency. A consistent test has asymptotic power equal to 1, rather than simply exceeding its size. Many commonly used tests are consistent, e.g., it can be shown that a test of the null hypothesis that a firm's beta is 0 is consistent. Neither the SQ test nor the standard approach can be consistent, even under normality of  $F$ , since the number of event dates equals 1 regardless of  $n$ . Test consistency is simply not possible with a fixed number of events.

This probability exceeds  $\alpha$  only when  $F(\sigma_a z_\alpha - \gamma) > F(y_\alpha)$ , which is violated when  $y_\alpha > \sigma_a z_\alpha - \gamma$ . The standard approach may yield either asymptotically biased or unbiased tests, depending on the magnitude of the event effect and the discrepancy between  $y_\alpha$  and its counterpart under normality,  $\sigma_a z_\alpha$ . Biased tests will result for more firms when there are smaller true effects (i.e.,  $\gamma$  closer to zero), and less dispersion of the excess returns distribution relative to normality.

To compare power for the two tests, consider first the case when the standard approach has correct asymptotic size, so that  $y_\alpha = \sigma_a z_\alpha$ . In this case, the standard approach's asymptotic rejection rate for given  $\gamma$  is  $F(y_\alpha - \gamma)$ , exactly equal to the SQ test's asymptotic power. Thus, when the standard approach provides correct asymptotic size, the two tests have equal asymptotic power.

As we showed above, the standard approach does not generally have correct size, so the fact that the two tests have equal size-corrected power is primarily of theoretical interest. As an empirical matter, what matters is whether the size distortions we documented above bring along especially low or high power. This point is especially relevant in the context of litigation and policy decisions. In each of these cases, the financial consequences of inferential outcomes are potentially very large. Thus, we will use our CRSP data to compare the estimated asymptotic rejection rate of the uncorrected standard approach with the SQ test's estimated asymptotic rejection rate.

## 6.2. Empirical Estimates of Asymptotic Power

The SQ test's asymptotic power for security  $j$  is  $F(y_\alpha^j + \sigma_a)$  when the event effect is a one standard deviation reduction in value ( $\gamma = -\sigma_a$ ), and  $F(y_\alpha^j + \sigma_a/2)$  when the event effect is a one-half standard deviation reduction ( $\gamma = -\sigma_a/2$ ). Equivalently, the SQ test's asymptotic power in the two cases is  $G(w_\alpha^j + 1)$  and  $G(w_\alpha^j + \frac{1}{2})$ , since the quantiles of  $F$  and  $G$  are related by the identity  $y_\alpha^j/\sigma_a = w_\alpha^j$ . Letting  $\hat{G}_{n_j}^j$  be the empirical distribution function of standardized fitted excess returns for security  $j$  based on all  $n_j$  available observations, we estimate the relevant probabilities using  $\hat{G}_{n_j}^j(\hat{w}_{\alpha,n_j}^j + 1)$  and  $\hat{G}_{n_j}^j(\hat{w}_{\alpha,n_j}^j + 0.5)$ . That is, we calculate the share of firm  $j$ 's standardized fitted excess returns that are less than  $\hat{w}_{\alpha,n_j}^j + 1$  for a one standard deviation reduction, and the share that are less than

**Table 2.** Asymptotic Power Under Normality and Empirical Excess Returns Distributions

	Value of $\alpha$		
	0.025	0.05	0.10
$\gamma = -0.5$			
Theoretical, under normality of $F$	0.072	0.126	0.217
Empirical, standard approach	0.053	0.087	0.140
Empirical, SQ Test	0.056	0.111	0.228
$\gamma = -1$			
Theoretical, under normality of $F$	0.169	0.260	0.389
Empirical, standard approach	0.110	0.175	0.320
Empirical, SQ Test	0.117	0.249	0.500

$\hat{w}_{\alpha, n_j}^j + 0.5$  for a one-half standard deviation reduction. The results discussed above ensure that these probabilities are consistent for  $G(w_\alpha^j + 1)$  and  $G(w_\alpha^j + 0.5)$  as  $n_j$  grows. Note that this evidence involves actual sample information only, rather than a Monte Carlo experiment.

As a basis for comparison, suppose we knew that  $F$  were normal, so that standardized excess returns had a standard normal distribution. In this case, we could compute the asymptotic power of the standard approach analytically at each choice of  $\alpha$ , since then  $\Pr(\text{Reject} \mid \gamma, \sigma_a, \alpha) = \Pr(\hat{\gamma}/\sigma_a \leq z_\alpha)$ , which converges to  $\Phi(z_\alpha - \gamma/\sigma_a)$ , where  $\Phi$  is the standard normal cumulative distribution function. Note that since the SQ test is asymptotically identical to the standard approach when  $F$  is normal, this is also the asymptotic power for each test under normality of  $F$ . (We stress that since excess returns distributions are clearly nonnormal, the only point of using the normality comparison is to fix a baseline level of power that could be considered reasonably attainable.)

Table 2 reports asymptotic power for each choice of  $\alpha$  and  $\gamma$  under normality of  $F$ . It also reports the cross-firm average estimated asymptotic power for the standard approach and for the SQ test. The table shows that the standard approach's power given the empirical excess returns distributions is about 70% of the power that would be achieved under normality in all but one case (when  $\gamma = -1$  and  $\alpha = 0.10$ , in which case the standard approach's power is still less than would be achieved under normality). The SQ test's power exceeds the standard approach's power in all six cases, and



substantially so in several of them. Table 2 shows that the SQ test has good asymptotic power even in highly nonnormal data.

Figure 4 graphs firm-specific estimated asymptotic power against firm-specific sample  $\alpha$ -quantiles. The top graphs involve  $\gamma = -0.5$ , while the bottom ones involve  $\gamma = -1$ . The graphs on the left-hand side are for the standard approach, while those on the right-hand side are for the SQ test. In each graph, we consider  $\alpha \in \{0.025, 0.05, 0.10\}$ . As in the size figures above, we plot the rejection rate on the vertical axis and the firm-specific sample  $\alpha$ -quantile,  $\hat{y}_{0.05,n_j}^j$ , on the horizontal axis. Dashed horizontal lines represent the average rejection rate for each choice of  $\alpha$  (these averages are the power figures from Table 2). We also continue to restrict attention to the middle 98% of firms as measured by sample  $\alpha$ -quantile values. As above, the jagged, lighter series are firm-specific rejection rates, while the smoother series are lowest estimates of the rejection rate given the firm-specific sample  $\alpha$ -quantile.

The results in Figure 4 illustrate two important facts in addition to those demonstrated in Table 2. First, holding constant the significance level  $\alpha$  and the true effect size  $\gamma$ , the standard approach's power falls as the firm-specific  $\alpha$ -quantile rises toward zero. Second, the opposite is true for the SQ test: power increases as the firm-specific  $\alpha$ -quantile rises toward zero, holding constant  $\alpha$  and  $\gamma$ .<sup>25</sup>

### 6.3. Summary of Empirical Power Results

We can summarize our empirical power results with two points. First, the SQ test has substantial power, especially against an effect size as large

25. There is a simple way to explain this combination of results. Since we are working with standardized fitted excess returns, the distribution's shape becomes more compressed as the 0.05-quantile  $\hat{w}_{0.05,n_j}^j$  rises toward zero. As a general matter, then, a value of  $\hat{w}_{0.05,n_j}^j$  closer to zero implies a greater mass between  $\hat{w}_{0.05,n_j}^j$  and  $\hat{w}_{0.05,n_j}^j + \delta$  for any positive  $\delta$ . This means that a fixed  $\delta = -\gamma$  will tend to move us further into the distribution when a security has a value of  $\hat{w}_{0.05,n_j}^j$  that is closer to zero. The result is a greater asymptotic rejection rate for the SQ test as we move to the right in the figures above. By contrast,  $z_\alpha$  does not increase as  $\hat{w}_{0.05,n_j}^j$  does: the standard approach uses the same critical value for all securities, regardless of the quantiles of their excess returns distributions. Therefore, its asymptotic rejection rate falls as  $\hat{w}_{0.05,n_j}^j$  moves closer to zero. Thus the pattern shown in the graphs on the left side of Figure 4 is precisely what we should expect to see.

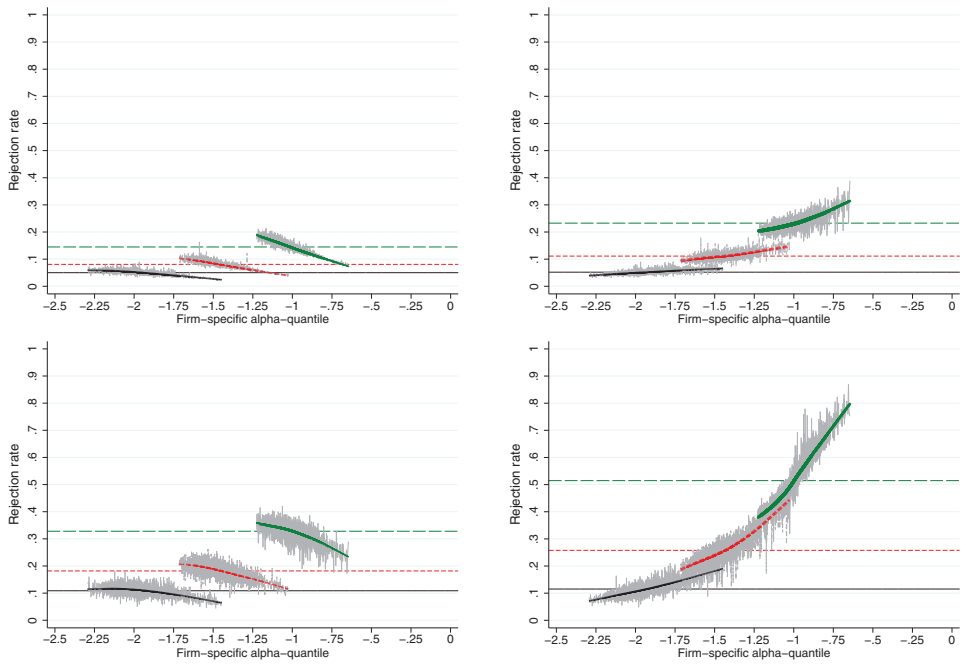


Figure 4. Asymptotic Power for the Standard Approach and the SQ Test,  $\gamma \in \{-0.5, -1\}$ ,  $\alpha \in \{0.025, 0.05, 0.10\}$ .

as one standard deviation in magnitude. Second, the standard approach's asymptotic power is considerably lower than the SQ test's power for much of the range of the data. Substantively, our results for the standard approach using data from 2000–2007 suggest the presence of a potentially severe bias against finding an event effect, especially for  $\alpha = 0.05$  or  $0.10$ . Among other things, this suggests the potential for considerable anti-plaintiff bias in the context of securities litigation.

## 7. Extensions to Accommodate Multiple Firms and Multiple Events

In this section, we extend the above results to accommodate multiple firms and multiple events.

### 7.1. Multiple Firms

One interesting extension involves the possibility of dealing with multiple firms experiencing an event on a single day. An interesting example is the effects on Microsoft's competitors of the June 7, 2000, order breaking up the company.<sup>26</sup> An additional, and litigation-relevant, example could involve a firm that is sued by multiple competitors that all allege anti-competitive acts on a given day.

The SQ approach can be extended to the multiple-firm case by specifying the basic returns model (3) for each of  $m > 1$  firms. As before,  $D_s^j$  is a dummy variable that equals 1 on the event date and 0 on all other dates. The parameter  $\gamma^j$  is firm  $j$ 's event effect. To implement the SQ approach, one estimates the  $m$  firm-specific equations, yielding  $\hat{\gamma} = (\hat{\gamma}^1, \hat{\gamma}^2, \dots, \hat{\gamma}^m)$ , the vector of the  $m$  firm-specific estimated event effects. Under the null hypothesis that all event effects are 0, the element of  $\hat{\gamma}$  corresponding to firm  $j$  will have asymptotic marginal distribution  $F^j$ , the same as the firm's distribution of excess returns on nonevent dates.

Working with the vector  $\hat{\gamma}$ , or functions of it, requires that we derive its asymptotic distribution. Let  $a_s = (a_s^1, a_s^2, \dots, a_s^m)$  be a random draw from the joint distribution of firms' excess returns. If firms' excess returns

26. See [Bittlingmayer and Hazlett \(2000\)](#) for more on Microsoft, antitrust enforcement, and event studies.

are mutually independent on any date  $s$ , then the asymptotic distribution of  $\hat{\gamma}$  satisfies  $F_{0a} = \times_{j=1}^m F^j$ , (i.e., the joint distribution is the product of marginals). When there is within-day, cross-firm dependence, this relationship does not hold, and we simply define the joint distribution of  $a_s$  as  $F_{0a}$ . In either case, it remains true that  $\hat{\gamma} \xrightarrow{P} \gamma + a_e$ . Therefore,  $\hat{\gamma} - \gamma$  has asymptotic distribution equal to the asymptotic distribution of the vector  $a_e$ , which is simply  $F_{0a}$ . Testing joint hypotheses involving multiple firms is thus a straightforward generalization of the single-firm case.

For example, let  $\varphi$  be some real-valued function of the vector of firm-specific excess returns that is bounded and continuous with probability 1. Under the null hypothesis that the event has zero effect on all  $m$  firms, the asymptotic distribution of  $\varphi(\hat{\gamma})$  is the distribution of  $\varphi(a_s)$ . This latter distribution can be estimated consistently using  $\hat{F}_{0\varphi}(y) = (n + 1)^{-1} \sum_{s=1}^n 1(\varphi(\hat{a}_s) \leq y)$ , where  $\hat{a}_s$  is the vector of  $m$  fitted excess returns (standardized or not) for any nonevent date. Our single-firm results on asymptotic size generalize easily to this case, though power properties must be established on a case-by-case basis.<sup>27,28</sup>

27. It is easy to establish that the generalized SQ test is unbiased whenever  $\varphi(a) = G(c'a)$ , where each element  $c_j$  of  $c = (c_1, \dots, c_m)'$  is positive and  $G$  is continuous and strictly increasing. To see why, observe that the inter-day independence assumption implies  $\Pr(\varphi(a_e) \leq y) = \Pr(\varphi(a_s) \leq y) = F_{0\varphi}(y)$  for all  $s = 1, 2, \dots, n$ . Since  $\hat{\gamma} - \gamma \xrightarrow{P} a_e$ , it follows that  $\Pr(\varphi(\hat{\gamma}) \leq y) = \Pr(c'\hat{\gamma} \leq y) = \Pr(c'(\hat{\gamma} - \gamma) \leq y - c'\gamma)$ , which converges to  $F_{0\varphi}(y - c'\gamma)$ . Let  $y_\alpha$  be the  $\alpha$ -quantile of the distribution of  $\varphi(a_s)$ , so that  $F_{0\varphi}(y_\alpha) = \alpha$ . Since all elements of  $c$  are strictly positive, and since the lower-tailed alternative is  $H_l: \gamma^j < 0$  for all  $j$ , under  $H_l$  we have  $c'\gamma < 0$ , and thus  $y_\alpha - c'\gamma > y_\alpha$ . It then follows from the fact that  $F_{0\varphi}$  is strictly increasing that under  $H_l$ ,  $F_{0\varphi}(y_\alpha - c'\gamma) > F_{0\varphi}(y_\alpha) = \alpha$ , which establishes that the test's power is greater than its significance level, establishing unbiasedness. An obvious choice for the  $j$ th element of the column vector  $c$  is to use  $c_j = \sigma_{a_j}^{-1}$ , the inverse standard deviation of excess returns for the  $j$ th security. This standard deviation is unknown, but it can be estimated from the pre-event excess returns. Using this choice of  $c$  has the effect of equalizing the scale of  $c_j \hat{\gamma}^j$  across  $j$ , so that no subset of securities dominates the test statistic's distribution. Note that this test would perform less well for  $H_l: \gamma^j < 0$  for some, rather than all,  $j$ , since firms with negative effects could be masked by firms with positive effects, which is allowed under this choice of  $H_l$ .

28. One choice for  $\varphi$  would be  $\varphi(a) = a'\hat{\Omega}^{-1}a$ , where  $\hat{\Omega}$  is a consistent estimate of  $\Omega$ , the  $m \times m$  variance matrix of  $a_s$ . This typical choice of  $\varphi$  would be appropriate only for two-sided alternatives, since it leads to tests that reject whenever  $\hat{\gamma}^j$  is too far from zero in either direction, for some  $j$ . Moreover, this choice of  $\varphi$  for the SQ approach may not yield an unbiased test unless each  $a_s$  has a multivariate normal distribution. Under the null, this test has the same asymptotic power as the standard approach of using  $\chi_m^2$

## 7.2. Multiple Events

The case of multiple events is also easy to address. For simplicity, we assume there is only one firm, since the previous section shows that it is easy to accommodate more than one firm, and this way we can drop the firm superscript for clarity. With  $m$  event dates, the model is now  $r_s = X_s \beta + \sum_{j=1}^m D_s^j \gamma^j + a_s$ , where  $D_s^j$  is a date  $j$ -specific dummy variable and  $\gamma^j$  is the effect of the date- $j$  event.

We focus here on the case in which  $m = 2$ , so that there are two event dates.<sup>29</sup> This case is of particular interest in the litigation context given some readings of Justice Breyer's opinion in *Dura*, that both the date of the initial fraudulent act and the date of the corrective disclosure may require event analysis in future fraud-on-the-market cases.<sup>30</sup>

Under the null hypothesis,  $\gamma^1 = \gamma^2 = 0$ . An alternative hypothesis relevant for a case like *Dura* might involve  $\gamma^1 > 0 > \gamma^2$ : on the date of the fraudulent act, the stock rises an unusually large amount, later to fall an unusually large amount on the date of the corrective disclosure. The arguments above establish that  $(\hat{\gamma}^1 - \gamma^1) - a_{e_1} \xrightarrow{P} 0$  and  $(\hat{\gamma}^2 - \gamma^2) - a_{e_2} \xrightarrow{P} 0$ , where  $a_{e_j}$  is the excess return on the  $j$ th event date. Under  $H_0$ ,  $\gamma^j = 0$ , so  $\hat{\gamma}^j - a_{e_j} \xrightarrow{P} 0$ . It then follows, as before, that the asymptotic null distribution of  $\hat{\gamma}^j$  is  $F(y) = \Pr(a_s \leq y)$ .

---

critical values, which are unbiased under normality. However, we have not so far been able to establish that this choice of  $\varphi$  yields an unbiased test when normality does not hold. (The usual power results derived in standard textbooks lean heavily on the normality of  $\hat{\gamma}$  under the alternative. Without normality, the quadratic form version of  $\varphi$  is not generally distributed  $\chi^2$ . Since  $E[a_e] = 0$ , it is easy to show that  $E[\hat{\gamma}' \Omega^{-1} \hat{\gamma}]$  converges in probability to  $a_e' \Omega^{-1} a_e + \gamma' \Omega^{-1} \gamma$ , but this fact alone is insufficient to generate an analytical result on power without normality of  $a_e$ .) Moreover, the resulting quadratic-form test is sensitive to all departures of  $|\gamma^j|$  from 0, including those in the direction opposite to a one-tailed alternative of interest. This is an undesirable feature in testing against a one-sided alternative. For example, it could lead us to reject the null when a firm experiences a positive, rather than negative, event effect on the date of a corrective disclosure.

29. It is straightforward to generalize to more dates. We consider only the two-date case both for brevity and because of the obvious application to securities litigation explained here.

30. For example, this is the reading favored by Dunbar and Mayer (2006), whose analysis would generally require an event study assessing the effects of both the initial alleged mis-statement and the subsequent corrective disclosure.

As noted above, it is common in event studies, especially those used in litigation, to assume conditional independence of returns across days. Under this assumption, the asymptotic joint distribution of  $(\hat{\gamma}^1, \hat{\gamma}^2)$  is simply the product of marginals,  $F(y_1)F(y_2)$ . We can use this fact to construct a level- $\alpha$  test of the null  $\gamma^1 = \gamma^2 = 0$  against the alternative that  $\gamma^1 > 0 > \gamma^2$  as follows. Let  $y_{\delta_1}$  and  $y_{\delta_2}$  be the  $\delta_1$ - and  $\delta_2$ -quantiles that satisfy  $F(y_{\delta_1}) = \delta_1$  and  $F(y_{\delta_2}) = \delta_2$ . Since  $\hat{\gamma}^1$  and  $\hat{\gamma}^2$  are asymptotically independent, it follows that

$$\lim_{n \rightarrow \infty} \Pr(\hat{\gamma}^1 > y_{\delta_1} \text{ and } \hat{\gamma}^2 < y_{\delta_2}) = [1 - F(y_{\delta_1})]F(y_{\delta_2}) \quad (9)$$

$$= (1 - \delta_1)\delta_2. \quad (10)$$

For a level- $\alpha$  test, we choose  $\delta_1$  and  $\delta_2$  so that  $\alpha = (1 - \delta_1)\delta_2$ . A natural requirement is that in testing the joint null, we hold the two event effects to the same probabilistic standard, in which case  $1 - \delta_1 = \delta_2 = \delta$ . Thus,  $\delta = \sqrt{\alpha}$ . So, for a level- $\alpha$  test, define  $\hat{y}_{\sqrt{\alpha}}$  and  $\hat{y}_{1-\sqrt{\alpha}}$  as the sample  $\sqrt{\alpha}$ - and  $(1 - \sqrt{\alpha})$ -quantiles of the distribution of fitted excess returns. The rejection rule for the SQ test is simple: reject the null that both event effects are 0 against the alternative hypothesis, of a positive date-1 and a negative date-2 effect, if and only if both  $\hat{\gamma}^1 > \hat{y}_{1-\sqrt{\alpha}}$  and  $\hat{\gamma}^2 < \hat{y}_{\sqrt{\alpha}}$ . This makes intuitive sense: we reject the joint null whenever the two event dates' estimated effects are simultaneously far from the origin in the directions that the alternative hypothesis specifies.

It is interesting to note that this result is very different from the naive approach of conducting two separate level- $\alpha$  tests, one for each of  $\hat{\gamma}^1$  and  $\hat{\gamma}^2$ . To illustrate, suppose that  $F$  were actually normal, with standard deviation  $\sigma_a$ . In that case, the critical values for a level- $\alpha$  test would be  $0.76\sigma_a$  for  $\hat{\gamma}^1$  and  $-0.76\sigma_a$  for  $\hat{\gamma}^2$ : any time  $\hat{\gamma}^1 > 0.76$  and  $\hat{\gamma}^2 < -0.76$  both occur, we would reject the null at level 0.05. By contrast, the naive approach would reject only if  $\hat{\gamma}^1 > 1.64\sigma_a$  and  $\hat{\gamma}^2 < -1.64\sigma_a$ . Given that  $F$  is normal, the probability of this joint event is only  $0.05^2 = 0.0025$ . Thus, the naive approach would radically under-reject the null. Requiring such a test is tantamount to changing the rules of the litigation game against plaintiffs. Justice Breyer's opinion in *Dura* says nothing of changing the standard (i.e., the significance level), needed for proof. Rather, it concerns the facts that must be established for any given standard of proof. We believe it would be

mistaken to implicitly read a new, much more stringent, standard of proof into a decision that is entirely silent on the point.

Finally, we note that exactly the opposite issue arises if the alternative hypothesis is that only one of the two event-date excess returns is nonzero. In this case, we must choose critical values that account for the fact that we get two draws from  $F$  under the null. The typical approach in the statistics literature is to use the Bonferonni correction for multiple draws, which involves using test-specific significance levels  $\delta = \alpha/2$ , or 0.025 in the level-0.05 case.<sup>31</sup>

## 8. Conclusion

In this paper, we have demonstrated the advantages of using the SQ test in single-firm, single-event studies. The SQ test's asymptotic Type I error rate always equals the analyst's desired significance level, and our Monte Carlo results suggest that the same holds with real-world sample sizes of pre-event observations. In addition, the test has considerable asymptotic power for empirically relevant excess returns distributions, and it gives up no power to the standard approach when the standard approach is valid.

An additional contribution of this paper has been to document the systematic and substantial errors of inference likely to result from inappropriate use of the standard approach to single-firm, single-event studies. The fact of nonnormality of excess returns distributions, and the basic problems associated with it for event studies have been known and discussed elsewhere. To our knowledge, though, ours is the first study to document the extent of this problem for event studies using a broad cross-section of firm-level data. Moreover, we believe our findings that (i) the standard approach's asymptotic Type I error rates are frequently biased downward as an empirical matter, and (ii) that its power is correspondingly low, are new. In the context of

---

31. Because they ignore the possibility that both date-specific test statistics might cause the test to reject, Bonferonni corrections are slightly conservative. The exact value for  $\delta$  is 0.0253 in the case  $m = 2$ . In general, this will make little difference asymptotically, though with relatively small  $n$ , it could matter slightly in practice. For instance, if  $n = 120$ , then  $0.025n = 2.975$ , so we would use the third order statistic of  $\hat{F}$  as our estimated critical value, while  $0.0253n = 3.0107$ , so we would use the fourth order statistic.

securities fraud litigation that helps motivate this paper, our evidence suggests that over the period 2000–2007, use of the standard approach might have led to substantial anti-plaintiff bias in some cases.

## Appendix A. Empirical Evidence on Firm-Specific Excess Returns Distributions

### A.1. Is the Distribution of Pooled Excess Returns Normal?

We first consider whether the pooled sample of excess returns can plausibly be normal. A simple test of normality is the Jarque–Bera test based on two properties that hold for all normal distributions: they are symmetric, so that they have skewness equal to zero, and they have kurtosis equal to 3.<sup>32</sup> The pooled distribution of excess returns in our sample has sample skewness equal to  $-0.11$  and sample kurtosis equal to  $94.4$ . The Jarque–Bera test statistic,  $JB$ ,<sup>33</sup> is based on the sample skewness and kurtosis values and is distributed  $\chi^2$  with two degrees of freedom under normality, with a level-0.05 critical value of  $5.99$ . The sample skewness and kurtosis values just listed yield a  $JB$  value of greater than 1.6 billion, so we can reject the null hypothesis that the pooled distribution of excess returns is normal.

### A.2. Do All Securities' Excess Returns Come from the Same Distribution?

Our next test concerns whether firms' excess returns come from the same (nonnormal) distribution. Given the strength of the null hypothesis, that all 3,050 distributions are the same, one could devise a variety of tests. We focus on two tests that concern the behavior of the security-specific sample 0.05-quantiles, which we call  $\hat{y}_{0.05, n_j}^j$ . We choose this statistic as the basis of our tests because variation in the true  $\alpha$ -quantiles guarantees incorrect size for some securities, as discussed above.

32. A distribution's skewness is the ratio of its third central moment to the cube of its standard deviation. A distribution's kurtosis is the ratio of its fourth central moment to the fourth power of its standard deviation.

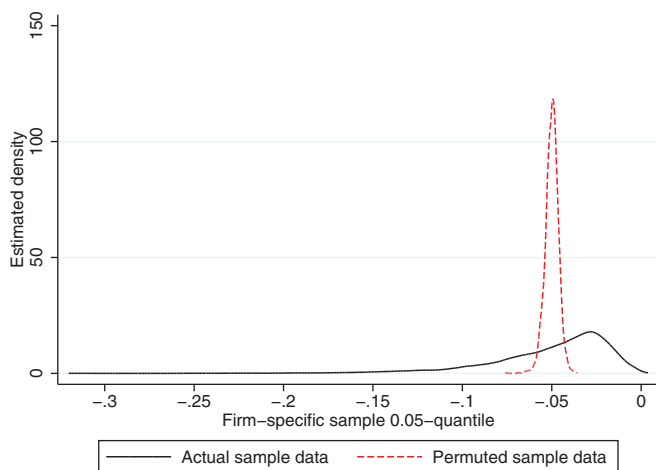
33. The  $JB$  statistic for distribution  $j$  equals  $(n_j/6)(sk_j^2 + (1/4)(\kappa_j - 3)^2)$ , where  $n_j$  is the sample size,  $sk_j$  is the sample skewness, and  $\kappa_j$  is the sample kurtosis.



It is of course possible that sample variation in key firm-specific sample  $\alpha$ -quantiles is driven by random noise. To see how such a situation could occur, imagine that all securities' excess returns came from the same underlying distribution,  $F$ . Given that we have data on 3,050 securities in our sample, we will wind up with 3,050 values of  $\hat{y}_{0.05, n_j}^j$  in any random sample drawn from this distribution. With so many draws, some of the security-specific sample 0.05-quantiles may appear to lie relatively far from the true population 0.05-quantile. That is, even if all securities have the same true sample 0.05-quantile  $y_{0.05}$ , there will be some set of firms whose realized sample quantiles are quite far from  $y_{0.05}$ . Thus, one needs a metric to evaluate whether the variation in sample 0.05-quantiles exceeds the variation to be expected under the null hypothesis of homogeneous excess returns distributions. We offer two approaches.

Our first approach is based on a permutation exercise. Under the null hypothesis that all firms' excess-returns distributions are the same, randomly re-ordering fitted excess returns across firm-day observations will not systematically affect the distribution of 3,050 firm-specific sample 0.05-quantiles,  $\{\hat{y}_{0.05, n_j}^j\}_{j=1}^{3050}$ . Figure A1 plots two kernel density estimates of the cross-firm distribution of  $\hat{y}_{0.05, n_j}^j$ . The first estimate, given by the solid line, is a density estimate for the sample 0.05-quantiles using the fitted excess returns based on the roughly five million actual returns we observe for our 3,050 firms. The second estimate, given by the dashed line, is a density estimate for the same statistics generated after randomly permuting all observed fitted excess returns across firm-day observations. If firms' excess returns came from the same underlying excess returns distribution, these two density estimates would be equal up to random variation induced by the permutation.

The figure suggests two important conclusions. First, much of the mass of the actual excess returns distribution of sample 0.05-quantiles lies considerably to the right of the permutation distribution. This means that firm-specific 0.05-quantiles are systematically closer to the origin than they would be if there were no heterogeneity in firms' excess returns distributions. As a consequence, the standard approach will reject less often at level 0.05 than it would in the absence of cross-firm distributional heterogeneity, even given the same pooled excess returns distribution. This result is consistent with the low rejection rates we find in the main text.



**Figure A1.** The Cross-Firm Distribution of Actual Versus Permutation-Based Sample 0.05-Quantiles for Fitted Excess Returns.

Second, there is much more dispersion in the actual cross-firm distribution of sample 0.05-quantiles than in the permutation-based distribution. For example, the sample standard deviation of  $\hat{y}_{0.05,n_j}^j$  across  $j$  is 0.035 in the actual data, roughly an order of magnitude greater than the permutation distribution's 0.0037. This finding implies that different firms will have much more variation in rejection rates using the standard approach on any random sample than they would if firms had the same underlying excess returns distributions—just as we should expect if firms have systematically differing true excess returns distributions.

While the visual evidence in Figure A1 is overwhelming, it does not constitute a formal test. To carry one out, we use the fact that sample quantiles from continuous distributions are asymptotically normally distributed, regardless of the distribution generating them. Formally, if  $F^j = F$  for all firms  $j$ , then  $\sqrt{n_j}(\hat{y}_{\alpha,n_j}^j - y_\alpha^0) \xrightarrow{d} N(0, V_0)$ , where  $y_\alpha^0$  is the true  $\alpha$ -quantile of the common-across-firms excess returns distribution,  $V_0 = \alpha(1 - \alpha)/[f_0(y_\alpha^0)]^2$ , and  $f_0$  is the density function associated with  $F$ . Under the null, we can estimate  $y_\alpha$  using the sample  $\alpha$ -quantile of the pooled distribution, which in our data is  $\hat{y}_{0.05,n_j}^0 = -0.050$  (in other words, the sample 0.05-quantile of the pooled excess returns distribution is a reduction in firm

value of 5%). We use Stata's `kdensity` command to estimate the pooled density at this value, with the result that  $\hat{f}_0(\hat{y}_{0.05}) = 1.67$ .

Next, we define  $\hat{Z}_{0.05}^j = (n_j/\hat{V}_0)^{1/2}(\hat{y}_{0.05,n_j}^j - \hat{y}_{0.05,n_j}^0)$ , with  $\hat{V}_0 = 0.05 \times 0.95/1.67^2 = 0.017$ . Under the null hypothesis of a common excess returns distribution, the sample  $\{\hat{Z}^j\}_{j=1}^{3050}$  must come from an approximately standard normal distribution. However, the actual sample mean of this distribution is 0.26, and the sample SD is 8.94, strongly suggesting that the underlying distribution does not have mean 0 and SD 1. As with the permutation figure, these results show that the cross-firm distribution of sample 0.05-quantiles is shifted to the right and is much more dispersed than would be true if all excess returns came from the same distribution. In addition, the sample skewness of  $\{\hat{Z}^j\}$  is  $-1.13$ , and its sample kurtosis is 4.70, casting doubt on whether the distribution of standardized sample 0.05-quantiles is normal at all. Indeed, the resulting JB statistic is 1,019, greatly exceeding the critical value of 5.99 for a level-0.05 test. These results clearly reject the null hypothesis that all firms' excess returns come from a single distribution.

### A.3. Are Security-Specific Excess Return Distributions Normal?

The preceding analysis shows that the pooled distribution of excess returns is not normally distributed, as well as that there is heterogeneity across firms in  $F^j$ , the underlying excess returns distributions. But it remains possible that each security's excess returns distribution is normal, with variances differing across securities. This combination would cause both of the above results, and it would also be unproblematic for the standard approach in event studies involving a single firm, since each excess return would come from some normal distribution.

To evaluate this possibility, we calculated the sample skewness and kurtosis values for excess returns within the size- $n_j$  sample for each security  $j$ . All but two of the 3,050 securities in our sample have a value of  $JB > 5.99$ , and the mean is over 75,000. Thus, there is overwhelming evidence against firm-specific normality of excess returns. Combined with our theoretical results, the foregoing empirical findings suggest that the standard approach may involve substantial Type I distortions, when applied to firms one at a time.

## Appendix B. Proof that the Asymptotic Null Distribution of $\hat{\gamma}$ is $F$

The proof that  $\hat{\gamma} - \gamma \rightarrow F$  is based on the following two facts:

*Fact 1.* The event-date fitted excess return,  $\hat{a}_e = r_e - X_e \hat{\beta} - \hat{\gamma}$ , exactly equals 0, which implies  $\hat{\gamma} = r_e - X_e \hat{\beta}$ .

This fact is well known in the outlier-detection and predictive-test literatures, e.g., see [Belsley et al. \(2004\)](#) or [Dufour \(1980\)](#). To prove it, observe that the OLS estimation criterion is to choose  $g$  and  $b$  to minimize the sum of squared estimated residuals,  $(r_e - X_e b - g)^2 + \sum_{s=1}^n (r_s - X_s b)^2$ . Whatever value  $b$  takes, we can always make  $\hat{a}_e^2 = 0$  by setting  $\hat{\gamma}(b) = g = r_e - X_e b$ . Thus,  $\hat{a}_e = 0$  is a necessary condition for minimization of the OLS criterion.

We note two related consequences of Fact 1. First, the OLS estimate  $\hat{\beta}$  from estimating (3) exactly equals the estimate that would be obtained by estimating (2) with the event date excluded. This is true because, with  $\hat{a}_e$  always equal to 0, the OLS objective function for choosing  $b$  is the same when we estimate (3) with the event date included or estimate (2) with the event date excluded. Second, the estimated standard error of the regression,  $\hat{\sigma}_a$ , is the same regardless of whether we estimate equation (2) with the event date excluded or (3) with the event date included. This result follows by observing that because  $\hat{\beta}$  is the same in each case, and the estimated event-date residual is zero, the sum of squared fitted residuals is identical in each case. Any difference would therefore have to come from the denominator used to estimate  $\sigma_a^2$ . When we estimate (2) rather than (3), we add one observation but also one parameter, leaving the number of degrees of freedom unaffected at  $n - 2$ .

*Fact 2.* The estimated standard error of the event effect,  $\hat{\sigma}_\gamma$ , converges in probability to  $\sigma_a$ , the standard deviation of the excess returns distribution  $F$ .

To prove Fact 2, we begin by noting that  $\hat{\sigma}_\gamma^2 = \hat{\sigma}_a^2 (D' M_x D)^{-1}$ . Note that the term  $D' M_x D$  equals  $1 - X_e (X' X)^{-1} X_e'$ , whose second term equals  $\text{trace}[X_e (X' X)^{-1} X_e'] = \text{trace}[(X' X/n)^{-1} X_e' X_e/n]$ . By a law of large numbers,  $(X' X/n)^{-1}$  converges in probability to its expectation, which is finite given standard moment conditions on  $X_s$ . Since there is only one event date, and it is fixed,  $X_e$  does not change with  $n$ , and  $X_e' X_e/n$  converges to zero. Therefore, the second term in  $D' M_x D$  converges to 0 in probability, so

$D'M_xD$  converges to 1. Since  $\hat{\sigma}_a^2$  is consistent for  $\sigma_a^2$ ,  $\hat{\sigma}_\gamma^2 = \hat{\sigma}_a^2(D'M_xD)^{-1}$  converges to  $\sigma_a^2$  in probability, and thus  $\hat{\sigma}_a \xrightarrow{p} \sigma_a$ .

From these facts, it follows that the distribution of  $\hat{\gamma}$  becomes arbitrarily close to the distribution of  $a_e + \gamma$ , with probability converging to 1. This result can be shown to hold as a special case of Corollary 3.1 of [Dufour et al. \(1994\)](#). However, it is easy to prove it directly using (3) and Fact 1, which together show that  $\hat{\gamma} = r_e - X_e\hat{\beta} = \gamma + a_e - X_e(\hat{\beta} - \beta)$ . Since  $\hat{\beta}$  is consistent for  $\beta$ , the probability limit of  $\hat{\gamma} - \gamma - a_e$  equals 0. A basic result in large-sample statistics, called the asymptotic equivalence lemma by [White \(2001, Lemma 4.7, p. 67\)](#), holds that if  $\text{plim}(\hat{\gamma} - \gamma) - a_e = 0$ , then the asymptotic distribution of  $(\hat{\gamma} - \gamma)$  and  $a_e$  must be the same. The distribution of  $a_e$  is obviously unaffected by the number of pre-event observations we choose to consider, so its asymptotic distribution is simply  $F$  from (3). The asymptotic equivalence lemma thus tells us that  $\lim_{n \rightarrow \infty} \Pr(\hat{\gamma} - \gamma \leq y) = F(y)$ , proving the result.

## References

- Andrews, D. 1993. "Tests for parameter instability and structural change with unknown change point," *Econometrica* 821–856.
- Andrews, D. W. K. 2003. "End-of-sample instability tests," 71 *Econometrica* 1661–94.
- Belsley, D. A., E. Kuh, and R. E. Welsch. 2004. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley-Interscience.
- Bhagat, S., and R. Romano. 2002. "Event studies and the law: Part i: Technique and corporate litigation," 4 *American Law and Economics Review* 141–68.
- Bittlingmayer, G., and T. W. Hazlett. 2000. "Dos kapital: Has antitrust action against Microsoft created value in the computer industry?," 55 *Journal of Financial Economics* 329–59.
- Brown, S. J., and J. B. Warner. 1985. "Using daily stock returns: The case of event studies," 14 *Journal of Financial Economics* 3–31.
- Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton, NJ: Princeton University Press.
- Carhart, M. M. 1997. "On persistence in mutual fund performance," 52 *The Journal of Finance* 57–82.
- Chou, P.-H. 2004. "Bootstrap tests for multivariate event studies," 23 *Review of Quantitative Finance and Accounting* 275–90.

- Chow, G. C. 1960. “Tests of equality between sets of coefficients in two linear regressions,” 28 *Econometrica* 591–605.
- Conley, T. G., and C. R. Taber. 2011. “Inference with difference in differences with a small number of policy changes,” 93 *Review of Economics and Statistics* 113–25.
- Dellavigna, S., and E. La Ferrara. 2010. “Detecting illegal arms trade,” 2 *American Economic Journal: Economic Policy* 26–57.
- Dube, A., E. Kaplan, and S. Naidu. 2011. “Coups, corporations, and classified information,” 126 *Quarterly Journal of Economics* 1375–409.
- Dufour, J.-M. 1980. “Dummy variables and predictive tests for structural change,” 6 *Economics Letters* 241–47.
- Dufour, J.-M., E. Ghysels, and A. Hall. 1994. “Generalized predictive tests and structural change analysis in econometrics,” 35 *International Economic Review* 199–229.
- Dunbar, F. C., and M. K. Mayer. 2006. *Dura and the New Vocabulary of Litigation Under Rule 10b-5*, Technical report, NERA.
- Fama, E. 1991. “Efficient capital markets ii,” 46 *Journal of Finance* 1575–617.
- Fama, E. F., and K. R. French. 1992. ‘The cross-section of expected stock returns’, 47 *Journal of Finance* 427–65.
- . 1993. “Common risk factors in the returns on stocks and bonds,” 33 *Journal of Financial Economics* 3–56.
- Ford, G. S., and A. D. Kline. 2006. “Event studies for merger analysis: An evaluation of the effects of non-normality on hypothesis testing” in Patrick Moriati I., ed., *Antitrust Policy Issues*, 135–56 (chapter 8). New York: Nova Science Publishers. Available at <http://www.worldcat.org/title/antitrust-policy-issues/oclc/65978494>.
- Gelbach, J. B., E. Helland, and J. Klick. 2011. Valid inference in single-firm, single-event studies, Technical report, Available at SSRN. Available at: <http://ssrn.com/abstract=1442222>.
- Hein, S. E., and P. Westfall. 2004. “Improving tests of abnormal returns by bootstrapping the multivariate regression model with event parameters,” 2 *Journal of Financial Econometrics* 451–71.
- Hein, S. E., P. Westfall, and Z. Zhang. 2001. Improvements on event study tests: Bootstrapping the multivariate regression model, Technical report, Texas Tech University. Working Paper.
- Horowitz, J. 1998. *Semiparametric Methods in Econometrics*. New York: Springer.
- Hosken, D., and J. D. Simpson. 2001. “Have supermarket mergers raised prices? an event study analysis,” 8 *International Journal of the Economics of Business* 329–42.
- Khotari, S., and J. B. Warner. 2007. “Econometrics of event studies” in B. Espen Eckbo, ed., *Handbook of Corporate Finance: Empirical Corporate Finance*.

- Vol. 1, 3–36 (chapter 1). Amsterdam: Elsevier/North-Holland. Available at <http://store.elsevier.com/product.jsp?isbn=9780444508980&pagename=search>.
- Klick, J., and R. Sitkoff. 2008. “Agency costs, charitable trusts, and corporate control: Evidence from Hershey’s kiss-off,” 108 *Columbia Law Review* 749–838.
- Kramer, L. A. 2001. “Alternative methods for robust analysis in event study applications” in Cheng-Few Lee, ed., *Advances in Investment Analysis and Portfolio Management*. Vol. 8, 109–32. Oxford: JAI. Available at: <http://www.worldcat.org/title/advances-in-investment-analysis-and-portfolio-management/oclc/223376058>.
- Mood, A. M. 1950. *Introduction to the Theory of Statistics*. New York: McGraw-Hill.
- Rosenbaum, P. R. 2002. *Observational Studies*. 2nd ed. New York: Springer.
- Simpson, J. D. 2001. “Did may company’s acquisition of associated dry goods corporation reduce competition? an event study analysis,” 18 *Review of Industrial Organization* 351–62.
- Simpson, J. D., and D. Hosken. 1998. Are retailing mergers anticompetitive? an event study analysis, Technical Report 216, FTC Bureau of Economics Working Paper.
- van der Vaart, A. W. 1998. *Asymptotic Statistics*. Cambridge, UK: Cambridge University Press.
- Walker, A. M. 1968. “A note on the asymptotic distribution of sample quantiles,” 30 *Journal of the Royal Statistical Society. Series B (Methodological)* 570–75. Available at: <http://www.jstor.org/stable/2984261>.
- Weinstein, M. I. 2008. “Don’t buy shares without it: Limited liability comes to American Express,” 37 *Journal of Legal Studies* 189–228.
- White, H. 2001. *Asymptotic Theory for Econometricians*. Revised ed. San Diego, CA: Academic Press.