The SEC's Short-Sale Experiment: Evidence on Causal Channels and Reassessment of Indirect Effects

Finance Working Paper N° 813/2022
August 2024

Bernard S. Black
Northwestern University and ECGI

Hemang Desai
Southern Methodist University

Kate Litvak
Northwestern University

Woongsun Yoo
Central Michigan University

Jeff Jiewei Yu
University of Arizona

ECGI Working Paper Series in Finance

The SEC's Short-Sale Experiment: Evidence on Causal Channels and Reassessment of Indirect Effects

Working Paper N° 813/2022
August 2024

Bernard S. Black
Hemang Desai
Kate Litvak
Woongsun Yoo
Jeff Jiewei Yu

Abstract

During 2005–2007, the Securities and Exchange Commission conducted a randomized trial in which it removed short-sale restrictions from one third of the Russell 3000 firms. Early studies found modest market microstructure effects of removing the restrictions but no effect on short interest, stock returns, volatility, or price efficiency. More recently, many studies have attributed a wide range of indirect outcomes to this experiment, mostly without assessing the causal channels for those outcomes. We examine the three most often cited causal channels for these indirect effects: short interest, returns, and managerial fear. We find no evidence to support these channels. We then reexamine the principal findings in four recent studies using a sample that closely matches the actual experiment and a common research design and find minimal support for the reported indirect effects. Our findings highlight the importance of confirming a causal channel or an economic mechanism and show that sample selection and specification choices can produce statistical significance even without an underlying economic mechanism.

Keywords: natural experiments; causal channels; specification choice; Regulation SHO; SEC experiment

Bernard S. Black*
Nicholas D. Chabraja Professor
Northwestern University
2211 Campus Drive, Evanston
IL 60208, United States
phone: +1 847 491 5049
e-mail: bblack@kellogg.northwestern.edu

Hemang Desai
Distinguished Professor of Accounting
Southern Methodist University
Dallas, TX 75275-0333, USA
phone: +1 214 768 3185
e-mail: hdesai@smu.edu

Kate Litvak
Professor of Law
Northwestern University
375 E Chicago Ave.
Chicago, IL 60611, USA
phone:
e-mail: k-litvak@northwestern.edu

Woongsun Yoo
Assistant Professor of Accounting
Central Michigan University
1101 S. Franklin St.
Mount Pleasant, MI 48859, USA
phone: +1 (989) 774-1933
e-mail: yoo1w@cmich.edu

Jeff Jiewei Yu
Associate Professor of Accounting
University of Arizona
1130 E. Helen St.
Tucson, Arizona 85721-0108
e-mail: jeffyu@arizona.edu

*Corresponding Author

informs
https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 70, No. 8, August 2024, pp. 5131–5156
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

The SEC's Short-Sale Experiment: Evidence on Causal Channels and Reassessment of Indirect Effects

Bernard S. Black,[a,b] **Hemang Desai,**[c] **Katherine Litvak,**[a] **Woongsun Yoo,**[d,]* **Jeff Jiewei Yu**[e]

[a] Pritzker School of Law, Northwestern University, Chicago, Illinois 60611; [b] Kellogg School of Management, Northwestern University, Evanston, Illinois 60208; [c] Cox School of Business, Southern Methodist University, Dallas, Texas 75275; [d] College of Business Administration, Central Michigan University, Mount Pleasant, Michigan 48859; [e] Eller College of Management, University of Arizona, Tucson, Arizona 85721

*Corresponding author

Contact: bblack@northwestern.edu, https://orcid.org/0000-0002-6499-0600 (BSB); hdesai@smu.edu, https://orcid.org/0000-0003-0541-0618 (HD); k-litvak@law.northwestern.edu, https://orcid.org/0009-0001-7408-2104 (KL); yoo1w@cmich.edu, https://orcid.org/0000-0001-8428-5645 (WY); jeffyu@arizona.edu, https://orcid.org/0000-0002-2986-2777 (JJY)

Received: January 6, 2022
Revised: February 13, 2023
Accepted: March 31, 2023
Published Online in Articles in Advance: December 13, 2023

https://doi.org/10.1287/mnsc.2023.4918

Abstract. During 2005–2007, the Securities and Exchange Commission conducted a randomized trial in which it removed short-sale restrictions from one third of the Russell 3000 firms. Early studies found modest market microstructure effects of removing the restrictions but no effect on short interest, stock returns, volatility, or price efficiency. More recently, many studies have attributed a wide range of indirect outcomes to this experiment, mostly without assessing the causal channels for those outcomes. We examine the three most often cited causal channels for these indirect effects: short interest, returns, and managerial fear. We find no evidence to support these channels. We then reexamine the principal findings in four recent studies using a sample that closely matches the actual experiment and a common research design and find minimal support for the reported indirect effects. Our findings highlight the importance of confirming a causal channel or an economic mechanism and show that sample selection and specification choices can produce statistical significance even without an underlying economic mechanism.

History: Accepted by David Sraer, finance.

Supplemental Material: The internet appendix and data are available at https://doi.org/10.1287/mnsc.2023.4918.

Keywords: natural experiments • Regulation SHO • causal channels • specification choice • indirect effects

A fragile inference is not worth taking seriously (Leamer 1985).

1. Introduction

In July 2004, the Securities and Exchange Commission (SEC) announced a randomized trial (contained in Regulation SHO) in which it suspended short-sale restrictions (price tests) for one third of the firms ("pilot" firms) in the Russell 3000 Index (R3000) that traded on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), or Nasdaq national market. Specifically, for the pilot firms, the SEC suspended the uptick rule for the NYSE and AMEX firms and the similar but less restrictive bid test for the Nasdaq firms during a roughly two-year period (May 2, 2005, through July 5, 2007). These tests were designed to restrict short selling in declining markets. The SEC left some but not all of the prior restrictions in place for the remaining roughly 2,000 firms ("controls"). The SEC's objective in conducting the experiment was to study the effects of relatively unrestricted short selling on market volatility, price efficiency, and liquidity and to monitor trading behavior (Securities and Exchange Commission 2004).

The NYSE uptick rule, rule 440B, was based on SEC rule 10a-1, adopted in 1938. It was intended to prevent short selling from exacerbating market declines (Macey et al. 1989). The NYSE rule prohibited short sales at a price lower than the previous price (called a "minus tick") or at the same price as the previous price if the last different price was higher ("zero-minus tick"). For Nasdaq, rule 3350 (adopted in 1998) prohibits short sales at or below the current bid when that bid is lower than the previous bid ("bid test," Alexander and Peterson 2008).[1]

Diether et al. (2009) provide an example to illustrate the NYSE uptick rule. Say the last sale was at $28.05 and the current quotes are \$28 (bid) and \$28.05 (ask). If a short seller submits a market sell order, the order will not execute at the bid (\$28.00) because this would be below the previous sale price (\$28.05). The order instead is treated as a limit order to sell at \$28.05. This restriction delays execution of short sales in declining markets (Alexander and Peterson 1999). This delay increases the cost of short selling. The Nasdaq bid test was weaker than the NYSE uptick rule. It required only that short sales be executed at a price at least as high as the previous bid.

That short selling is costly has important implications. Short selling is generally viewed as improving market efficiency by allowing bearish information to be promptly reflected in price. A substantial body of work provides evidence that short sellers are informed (heavily shorted firms underperform) and an increase in short interest is a bearish signal (e.g., Aitken et al. 1998, Desai et al. 2002, Asquith et al. 2005). Conversely, short-sale restrictions, by restricting trading by individuals with bearish information, can result in upward-biased prices (Miller 1977). Using a rational expectations framework, Diamond and Verrecchia (1987) argue that costly short selling does not bias the price, but reduces the speed of price adjustment to bad news. In their model, an unexpected increase in short interest conveys bad news to investors by revealing the existence of negative private information.

Though price tests delay short-sale execution and thereby increase the cost of short selling, delays because of the NYSE uptick rule at the time of the experiment were measured in minutes (Alexander and Peterson 1999). Christophe et al. (2004) find that the Nasdaq bid test had little impact on short-sale execution speed. It is unclear whether these delays were important for fundamentals-based short sellers, who typically take short positions based on the firms' long-term prospects and hold those positions for substantial time periods (Desai et al. 2002). Even back in the 1980s, the effectiveness of the uptick rule in limiting short selling was questioned (Barclay 1989).[2]

Several developments over the years further diminished the effectiveness of the price tests in constraining short selling. First, the progressive reductions in tick size from \$0.125 to a penny by the time of the experiment, reduced any effect of the uptick rule. Alexander and Peterson (2002) find that execution times for short sales declined when the tick size was reduced to 1/16th (\$0.0625) in 1997. Decimalization, beginning in 2001, further reduced execution times.

Second, regional exchanges, which traded the shares of larger NYSE firms, did not impose the uptick rule, and Archipelago and INET, which accounted for around 40% of trading volume in Nasdaq firms, did not impose the bid test.[3] This further weakened any effect for firms traded on these venues. Many larger firms also had traded options, which are an alternate way to express bearish sentiment, with low incremental cost compared with direct short sales (Battalio and Schultz 2006).

By the time of the experiment in 2005, widespread belief among practitioners was that the price tests no longer presented a meaningful impediment to valuation-based short selling. Consistent with this belief, initial studies of the experiment found little to no direct impact of removing the price tests on short interest, returns, volatility, or price efficiency (Securities and Exchange Commission 2007, Alexander and Peterson 2008, Diether et al. 2009). Prior work also found higher short-sale volume for NYSE firms and somewhat weaker evidence for Nasdaq firms, suggesting that removing short-sale constraints led to increased arbitrage trading and/or market making activity for these firms.[4] Based on these findings, the SEC in 2007 removed short-sale restrictions for all firms.[5]

Despite reasons to doubt that the price tests were a meaningful impediment to valuation-based short selling and no evidence of a direct impact of the Regulation SHO experiment on short interest or returns, more than 60 papers in accounting, finance, and economics report that suspension of the price tests had wide-ranging indirect effects on pilot firms. These studies claim that pilot firm managers responded to the experiment by reducing capital expenditures and R&D outlays, reducing discretionary accruals, changing disclosure policies and compensation plans, reducing investment in workplace safety and more (see the internet appendix, Table IA-1, for a summary). Some find that the experiment affected the behavior of third parties, such as auditors and analysts.

The broad range of indirect effects attributed to the short-sale experiment is surprising for a number of reasons. First, to credibly attribute indirect effects to the experiment, there should be evidence supporting a causal channel (an economic mechanism through which removing the price tests could generate those indirect effects). If the price tests significantly constrained fundamentals-based short selling, then removing them should result in an increase in short interest (in the presence of heterogenous expectations), which should then result in lower prices (Miller 1977). However, early studies of the experiment found that removing the price tests did not meaningfully affect short interest, returns, volatility, price efficiency, etc. Most indirect effects studies do not examine or provide evidence of a causal channel to support their findings. Second, firm responses to this experiment should be proportional to the extent to which the price tests restricted short selling. If the price tests did not meaningfully constrain short selling, one would not expect removing them to generate wide-ranging indirect effects. This raises the possibility that the indirect effects found in later studies could be false positives (Harvey 2017). Third, when the experiment ended, the SEC removed the price tests for all firms. Thus, any gap between pilot and control firms caused by the experiment should close after the experiment. Few of the indirect effects studies test for such a reversal; those that do test often find no reversal.

This discussion suggests that the evidence of indirect effects of the Regulation SHO experiment may not be robust. In this paper, we carefully examine the principal causal channels that could support indirect effects. We study the three channels most commonly cited in the indirect effects literature: (i) increased short selling proxied by higher short interest, (ii) negative impact on pilot firms' share prices or returns, and (iii) managers' fear of bear raids. For a fourth channel, price efficiency, we rely on prior work finding no evidence that the experiment impacted price efficiency (Alexander and Peterson 2008, Diether et al. 2009, De Angelis et al. 2017) or the speed with which share prices respond to negative information (Bai 2008, De Angelis et al. 2017).

As noted, prior work found higher short sale volume but no increase in short interest or a drop in share prices for pilot firms. However, Grullon et al. (2015) (GMW) report evidence for an effect on short interest between the experiment announcement in 2004 and launch in 2005 (a period not previously studied) and a drop in share prices for small pilot firms (which they define as below median in assets) before the experiment announcement (again, a period not previously studied). We, therefore, extend prior work studying short interest and returns by using a more comprehensive sample that closely matches the actual pilot and control firms and a longer time period that spans the period before the experiment started and after it ended. Because the price tests were eliminated for all firms in July 2007, we also examine whether short interest at control firms increased relative to pilot firms after that. We find no evidence that short interest increased for pilot firms relative to control firms between experiment announcement and launch or during the experiment, nor a relative increase in short interest for control firms after the experiment ended. Also, prior work studies only differences in mean short interest for pilot versus control firms. We extend that work by examining the distribution of short interest across firms. We again find no evidence of differences between pilot and control firms.

We then examine stock returns of pilot firms relative to control firms over a much longer period than prior work, including the preannouncement period in which GMW report evidence of a drop in pilot firm prices. If the price tests had meaningfully restricted bearish information from being reflected in prices, then suspension of these tests for pilot firms could be expected to result in higher short interest and corresponding negative relative returns for pilot firms. Similarly, once the restrictions were removed from all firms, one should expect an increase in short interest and negative returns for control firms relative to pilot firms. We do not find evidence of differential returns for pilot firms relative to control firms either before, during, or after the experiment. For the GMW finding of negative returns for small pilot firms (below median in assets) before the experiment was announced, this result is not robust to how one defines "small" (based on assets versus market capitalization or trading volume). There result also lacks a causal mechanism because it comes from a period when the list of pilot firms was unknown.

The third commonly cited channel is managerial fear. The papers relying on this mechanism argue that even if the Regulation SHO experiment did not actually affect short interest or returns, pilot firm managers could have feared being targeted by short sellers and taken preemptive actions (for example, Fang et al. 2016 (FHK)). We cannot directly study this channel, but we can assess its plausibility. If firm managers were fearful that relaxing the price tests would adversely affect them, one might expect them to voice concerns in various ways: speaking with business news reporters, writing to the SEC when it sought public comments, or seeking meetings with SEC officials to express opposition. For example, when the Financial Accounting Standards Board (FASB) proposed expensing employee stock options in 1993, there was major opposition to this proposed rule from firm managers, including appeals to Congress, and extensive media coverage of manager opposition. In contrast, the short-sale experiment generated very little publicity with no evidence of managerial opposition. We searched the business press during 2003 when the rule was proposed, in 2004 when the experiment was announced, in 2006 when the SEC proposed repeal, and in 2007 when the rule was repealed. We also reviewed all comment letters received by the SEC for the experiment and the proposed repeal. We found no evidence of manager opposition in either of these sources at any of these times. We cannot rule out the managerial fear channel, but our analysis of the media coverage (or lack thereof) of the Regulation SHO experiment and the comment letters raises doubts as to the plausibility of the channel and whether it can generate the wide-ranging changes in firm policies documented in the indirect effect studies.

The lack of evidence supporting the principal causal channels asserted in the indirect effects studies raises concern that the reported results may be false positives. We, therefore, sought to assess the robustness of the reported effects. To keep the task reasonable, we reexamined the principal outcomes from four studies, which we chose for several reasons. First, all are published in top journals and, thus, are likely to represent the best of the indirect effects literature. Second, they rely on data from standard sources, which eased the task of reexamination. Third, they rely on different causal channels and provide evidence for an array of economically important effects. We summarize the major conjectures tested in each study as well as the causal channels on which they relied.

FHK rely on the managerial fear channel. They conjecture that, in response to a greater threat of short selling, pilot firms' managers reduced earnings management to preemptively deter short sellers. They measure earnings management using performance-matched discretionary accruals (PMDA).

GMW primarily rely on the causal channels running through short interest and returns. They conjecture that short-sale constraints result in overvaluation of firms, which can induce overinvestment, and that removing these constraints reduces share prices, and firms then reduce investment. GMW report that small pilot firms reduced investment, experienced lower growth, and raised less capital during the experiment period.

Hope et al. (2017) (HHZ) also rely on a causal channel running through short interest and returns and test for an effect of the short-sale experiment on behavior of firms' external auditors. They conjecture that the suspension of price tests increased litigation risk for auditors because pilot firms would face significantly higher short-selling threats, and because short sales drive stock prices down, some investors may launch securities lawsuits and name the auditors as defendants, resulting in higher litigation risk. The auditors would then preemptively respond to expected higher litigation risk by increasing audit fees.

Lin et al. (2019) (LLS) rely on a price efficiency channel. They posit that share prices of pilot firms became more informative and that, with more informative prices to guide managerial actions, shareholders perceive lower need for direct CEO performance incentives, so pilot firms should have lower CEO wealth-to-performance sensitivity (WPS). They also hypothesize that managers rely on the more informative prices to guide business decisions, which makes investment more sensitive to Tobin's *q*.

We summarize these four papers in Table 1. To reexamine these papers, we use a sample that closely follows the SEC's exclusion rules for experiment. Our sample is larger than the samples of all four studies we reexamine; and therefore more representative of the experiment. We defined preexperiment ("pre"), during experiment ("during"), and postexperiment ("post") periods and made other specification choices consistently across the reexamined papers. We did so before running any regressions.[6]

With our sample and specification, we find no support for the principal conjectures in these four studies. None of the outcomes is statistically significant; several coefficients have the wrong sign. We then study "best match" specifications in which we did our best to match each paper's sample and design based on the descriptions in each paper. Two results (one for GMW, the other for LLS) become statistically significant with the predicted sign, but those results are fragile and, for LLS, are driven by nonparallel trends in the pretreatment period. Finally, we conduct technical replication of FHK and HHZ, for which the authors provided us with their sample and code. The reported results technically replicate but are not robust to minor changes in specification.

Our findings have important implications for academics and policy makers. First, we fill a gap in the literature by carefully examining the principal causal channels that could support the indirect effects attributed to the Regulation SHO experiment. Our failure to find support for these causal channels and our finding that the results in the four reexamined papers are not robust highlight the importance of providing evidence on a credible causal channel as a basis for reliable inference. The Heath et al. (2023) critique of multiple hypothesis testing strengthens those concerns.[7] Second, because the experiment involved a regulatory change, the results have policy implications. Our evidence that the suspension of the price tests had little direct impact suggests that these restrictions did not prevent information-based short selling or result in increased short-selling threats to pilot firms. This is an important finding as information-based short selling makes markets more efficient. A plethora of indirect effects studies conjecture that removing the tick tests increased short-selling threats and generated significant negative externalities. Our results show that the evidence, in at least the studies we reassess, is not robust. This suggests that regulators should be cautious in inferring policy implications from indirect effects studies absent evidence for a causal channel. Finally, even when researchers begin with a randomized trial, they must make many specification choices. We provide evidence that, even for a randomized trial, some specification choices can produce significant results when other reasonable choices would not.[8]

This paper proceeds as follows. Section 2 discusses the evidence on causal channels. Section 3 summarizes our reexamination methodology. Section 4 presents our reexamination results. Section 5 discusses broader implications from our project and concludes. Given the scope of this project, which both examines several causal channels and reexamines four papers, we relegate many supporting results to the internet appendix.

2. Evidence on Causal Channels

2.1. Importance of a Causal Channel in Testing Indirect Effects of the SEC Experiment

We view evidence supporting a causal channel as central to the credibility of the indirect effects studies. We examine the three causal channels most often cited in the indirect effects literature: short interest, returns, and managerial fear. Aside from the manager fear channel, any indirect effects from the Regulation SHO experiment should follow from a causal channel that begins with the direct effects of the experiment on short selling. The most natural channel involves greater short selling at pilot firms, which should lead to higher

Table 1. Summary of the Four Re-examined Papers

Paper	Conjecture and main result	Posited causal channel	Plausibility of causal channel	Our evidence
Fang et al. (2016)	Pilot firms reduced earnings management, measured using performance matched discretionary accruals, calculated as discretionary accruals of sample firms less discretionary accruals of matching firms, selected based on lagged ROA.	Pilot firms' managers feared being targeted by short sellers and hence reduced earnings management.	No evidence to support manage fear or other causal channels. Absent support for a causal channel, unclear why manages would take costly action to reduce earnings.	No significant change in PMDA with either balanced or unbalanced sample. Even with their exact sample and specification, no significant results if one selects matching firm based on current instead of lagged ROA.
Grullon et al. (2015)	Short-sale constraints result in overvaluation, which leads to overinvestment. Thus, removing short-sale constraints should reduce share prices and investment. Prices of small pilot firms (below sample median in assets) fell in the two weeks *before* experiment announcement. Small pilot firms reduced investment and raised less capital.	Lower share prices lead to lower investment, lower growth, and hence less need to raise capital.	Share price drop before SEC list of pilot firms was made public is implausible and sensitive to sample choice. Unclear why a small price drop in 2004 would lead to major change in investment over 2005–2007.	No support for share price drop. No significant change in the Grullon et al. (2015) outcomes.
Hope et al. (2017)	Increased short selling increases auditor securities litigation risk. Auditors respond by increasing audit fees.	Auditor fear that short sellers will drive down share prices, leading to increased litigation risk; auditors, therefore, increase audit fees.	No evidence for main causal channels. Authors do not study whether litigation increased.	Following the authors' exact sample selection steps, we construct a larger sample and find no significant change in audit fees.
Lin et al. (2019)	Pilot firm share prices become more informative because of reduced short-sale constraints. Shareholders and boards perceive lower need for direct incentives, so CEO pay-performance sensitivity falls. Managers rely more heavily on prices to guide decisions, so investment sensitivity to Tobin's q rises.	Fewer short-sale constraints lead to more informative prices; boards and managers believe prices are more informative.	The authors do not assess whether prices become more informative, and other studies find no significant effects of the experiment on price efficiency.	No significant change in any of the Lin et al. (2019) outcomes.

short interest at pilot firms relative to control firms. A second plausible channel involves the impact of greater short selling on pilot firms' share prices. Effects on short interest or price might, in turn, affect manager behavior or the behavior of third parties, such as auditors and analysts. However, the early studies of the experiment found no significant change in short interest or pilot firm returns (Alexander and Peterson 2008, Diether et al. 2009).

Despite the lack of evidence that the experiment affected short interest or returns, many indirect effects papers invoke greater short selling as a causal channel without discussing this evidence, sometimes citing evidence for an increase in short-sale trading volume. However, higher short-sale volume without higher short interest is consistent with increased market making and/or arbitrage trading and need not imply an increase in fundamentals-based short selling. Many indirect effects papers invoke a share price channel because of presumed increased short selling without discussing or providing evidence on the experiment's effect on prices or returns.

2.2. Sample Selection

The SEC announced the short-sale experiment on July 28, 2004. To assess the GMW results on whether the experiment affected stock returns around the announcement or change in short interest between the announcement and experiment launch, the appropriate sample is all pilot and control firms at the time of announcement, which we term the "2004 announcement sample." The SEC assigned one third of the firms in the R3000 index as of June 25, 2004, to be pilot firms, effectively at random. The SEC's list of pilot firms includes only 986 firms (which the SEC called "category A" firms) instead of 1,000 because the SEC excluded firms that were not listed on the NYSE, AMEX, or Nasdaq national market and firms that became public after April 30, 2004.

The SEC never published a list of all control firms, only those in category A (pilot) or category B (control firms in the Russell 1000). We construct the 2004 announcement sample by starting with the R3000 list as of June 30, 2004, and carefully following the SEC's exclusion rules. For details, see the internet appendix. The resulting sample includes 2,954 firms: 985 pilot (the SEC's 986 firms minus one firm that was delisted on June 28, 2004) and 1,969 control.

2.3. Causal Channel 1: Impact on Short Interest

To examine the impact of the Regulation SHO experiment on short interest, we study midmonth short interest, reported on Compustat, from July 2003 to December 2007. We treat July 2003 through July 2004 as the preannouncement period, August 2004 through April 2005 as the prelaunch period, May 2005 through June 2007 as the experiment period, and July to December 2007 as the postexperiment period.[9]

2.3.1. Graphic Evidence. We start with graphic evidence. In Figure 1, we plot monthly short interest of pilot and control firms from July 2003 through December 2007. Panel (a) shows results for the 2004 announcement sample (2,954 firms). Panel (b) shows results using a narrower "2005 analysis sample," which was created by updating the 2004 announcement sample to the start of the experiment period (May 2005) and excluding financial and utility firms. The 2005 analysis sample comprises 2,115 firms. Panel (c) shows results using the GMW best match sample. Vertical lines indicate the experiment start and end; a dotted line shows the experiment announcement. For all three samples, there is no evidence of higher short interest for pilot firms during either the prelaunch or the experiment period. This is consistent with the price tests not meaningfully affecting short sellers' ability to trade on bearish sentiment.

GMW report evidence of an increase in short interest during the prelaunch period, which is stronger for small firms (below-median assets). We, therefore, also report, in panel (d), a similar comparison for small firms within the GMW best match sample. Here, too, there is no meaningful separation between pilot and control firms. It is also unclear to us why one would expect, on theoretical grounds, a rise in short interest before the experiment began.

2.3.2. Difference-in-Difference (DiD) Regression Results for Short Interest. Next, we undertake a more formal analysis of whether the SEC experiment affected short interest. We use a DiD approach to estimate the causal effect of the experiment on pilot firms' short interest during both the prelaunch and the experiment period:

$$y_{its} = \alpha_s + f_{is} + \lambda_{ts} + \gamma_s^a * Ann_t + \gamma_s^e * During_t + (\beta_s^a * P_i * Ann_t) + (\beta_s^e * P_i * During_t) + \epsilon_{its}. \quad (1)$$

Here, y is short interest as percentage of shares outstanding, P_i is a pilot firm dummy (= 1 for pilot firms), Ann_t and $During_t$ are dummy variables for the postannouncement but prelaunch period (August 2004 through April 2005) and the experiment period (May 2005 through June 2007), i indexes firms, t indexes time in months, s indicates the sample, and the f_{is} and λ_{ts} are firm and month fixed effects (FE). We cluster standard errors (s.e.) on firm, following Alberto et al. (2023), who recommend clustering at the level of treatment assignment. The coefficients of principal interest are on the interaction terms: β_s^a for the prelaunch period and β_s^e for the experiment period.

In Table 2, we present DiD regressions, following Equation (1), for the four samples shown in Figure 1 and three additional samples: small firms in the 2004 announcement sample (based on market capitalization) and small firms in the GMW best match sample (based on market capitalization and trading volume). We use market capitalization and trading volume as alternative measures of "small" because R3000 membership is based on market capitalization and the SEC ranked firms based on trading volume when selecting pilot firms. For both the prelaunch and experiment periods, all coefficients on the interaction terms are small and statistically insignificant.[10]

2.3.3. Distribution of Short Interest. Although there is no evidence for higher mean short interest for pilot firms, the SEC experiment could still have allowed short sellers to target some firms more heavily. We, therefore, also test for a difference in the distribution of short interest between pilot and control firms and report results in Internet Figure IA-3. There is no visual evidence of a difference in the distribution of short interest between pilot and control firms during the preannouncement, prelaunch, or experiment periods. A Kolmogorov–Smirnov test fails to reject the null of equal distributions of short interest for pilot and control firms.

In sum, there is no evidence that the Regulation SHO experiment led to higher short interest for pilot firms and, thus, no evidence for a causal channel that runs through short interest.

2.4. Causal Channel II: Impact of Short-Selling Restrictions on Share Prices

We next examine evidence for a second possible channel: lower returns for pilot firms. Prior studies found no effect of the short-sale experiment on returns (Diether et al. 2009) or extreme price changes (Securities and Exchange Commission 2007). However, GMW report small but statistically significant negative abnormal returns to small

Figure 1. (Color online) Monthly Open Short Interest

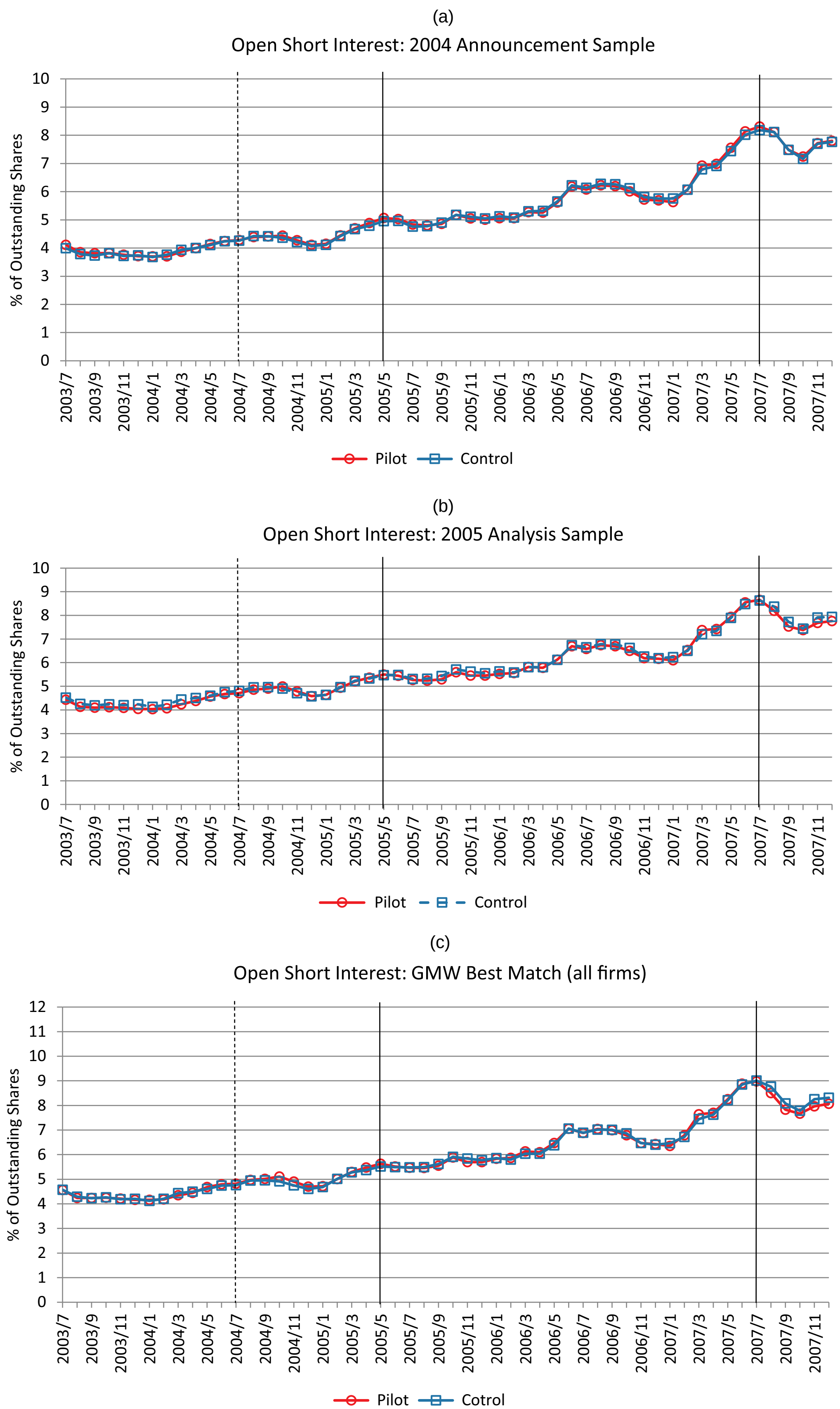
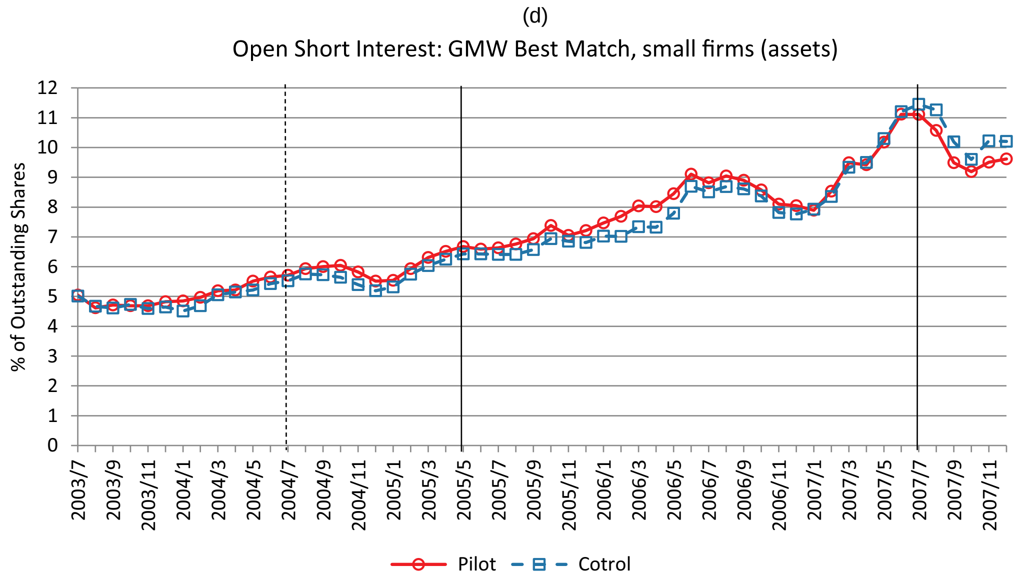


Figure 1. (Continued)



Notes. Mean midmonth short interest as a percentage of outstanding shares over July 2003–December 2007 for pilot firms and control firms for the 2004 announcement sample (panel (a)), 2005 analysis sample (panel (b)), Grullon et al. (2015) best match sample (panel (c)), and Grullon et al. (2015) best match sample, (small firms, below-median assets) (panel (d)). The SEC experiment was announced on July 28, 2004, and ran from May 2, 2005, through July 5, 2007. Dotted vertical line shows announcement date; solid vertical lines separate prelaunch, experiment, and postexperiment periods. Not all firms have data on short interest in each month.

pilot firms (based on assets) relative to control firms during the two weeks before the experiment was announced. We, therefore, reconsider the evidence for an effect on returns, focusing on this period. We report "raw" rather than abnormal returns (as do GMW) because there should be only chance imbalances in market model βs between pilot and control firms; we confirm preexperiment balance (see Internet Table IA-4). Thus, a comparison of raw returns to pilot versus control firms should be unbiased.

Table 2. DiD Regressions for Mid-Month Short Interest

	2004 announcement			GMW best match			
Sample	All firms (1)	Small firms (based on market cap) (2)	2005 analysis (3)	All firms (4)	Small firms (based on assets) (5)	Small firms (based on market cap) (6)	Small firms (based on trading volume) (7)
Prelaunch × Pilot	−0.022 (−0.19)	0.098 (0.51)	0.089 (0.61)	0.088 (0.59)	0.248 (1.08)	0.331 (1.32)	−0.041 (−0.25)
Experiment × Pilot	0.038 (0.27)	0.251 (1.04)	0.141 (0.79)	0.102 (0.55)	0.408 (1.26)	0.323 (0.99)	−0.114 (−0.63)
Prelaunch dummy	**0.814*** (7.87)**	**1.474*** (9.11)**	**0.843*** (6.38)**	**0.816*** (5.97)**	**1.270*** (5.65)**	**1.665*** (7.81)**	**0.876*** (6.06)**
Experiment dummy	**4.014*** (25.25)**	**5.633*** (22.63)**	**3.960*** (20.57)**	**4.237*** (21.62)**	**5.879*** (19.35)**	**6.502*** (21.39)**	**4.709*** (20.14)**
Constant	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.670	0.671	0.669	0.679	0.677	0.672	0.651
Observations	129,284	64,746	94,280	86,956	44,736	43,519	44,384
Pilot (control) firms	875 (1,735)	428 (897)	614 (1,237)	564 (1,133)	278 (603)	266 (591)	288 (570)
Number of firms	2,610	1,325	1,851	1,697	881	857	858

Notes. Regressions of midmonth short interest (as percentage of outstanding shares) with firm and month fixed effects over July 2003–December 2007 for indicated samples on prelaunch dummy (= 1 for August 2004 through April 2005), experiment dummy (= 1 for May 2005–June 2007), prelaunch × pilot and experiment × pilot interaction terms, and constant term. *t*-statistics with standard errors clustered on firm in parentheses. Significant results at 5% or better in boldface.

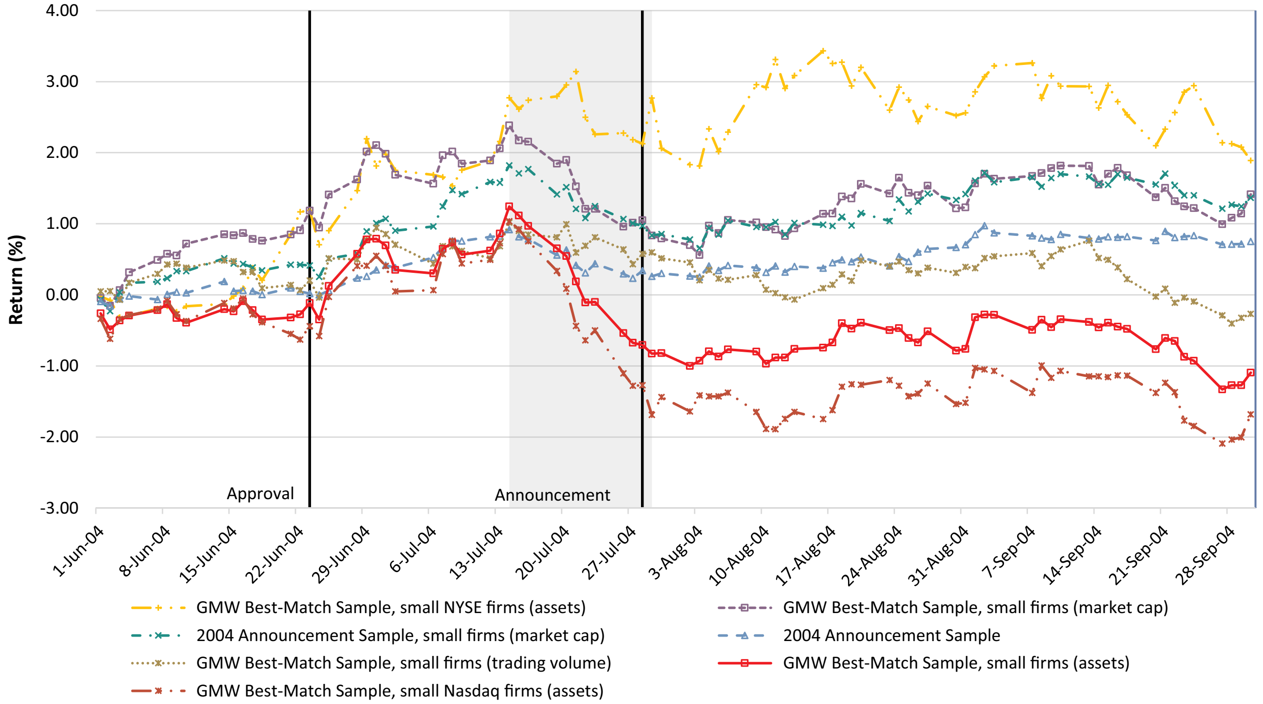
*, **, *** indicates statistical significance at the 10%, 5%, and 1% levels, respectively.

2.4.1. Graphic Evidence. GMW report both graphical and regression evidence of modestly negative relative returns to small pilot firms over a (−10, +1) window around the announcement date for the experiment. In Figure 2, we report buy and hold relative returns (BHRRs; buy and hold returns of pilot firms minus those for control firms) over June 1, 2004, through September 30, 2004. GMW use the same period and a similar measure, which they call buy and hold-abnormal return equivalent. We present results for (i) the 2004 announcement sample and (ii) small firms within this sample (based on market capitalization). For comparison with GMW, we also study several samples based on the GMW best match sample: small firms (based on assets, market capitalization, or trading volume), small NYSE firms (based on assets), and small Nasdaq firms (based on assets). We exclude two firms with share price <\$1 on June 30, 2004. Vertical lines indicate the experiment approval and formal announcement dates. Shading indicates the (−10, +1) window.

In Figure 2, there is evidence of a relative drop in BHRRs over the (−10, +1) event window for only two of the seven samples: small firms (based on assets) and small Nasdaq firms within the GMW best match sample. For both samples, pilot firm returns are initially higher than control firm returns following SEC approval of the experiment on June 23 but then decline during the two weeks before the formal announcement. Considering all seven samples, the negative BHRR reported by GMW is not robust to sample choice. It is not found for the preferred, larger sample (the 2004 announcement sample) nor for NYSE firms even though the experiment was expected to have a larger impact on NYSE firms because the Nasdaq bid test was weaker than the NYSE uptick rule (Christophe et al. 2004).[11]

2.4.2. Regression Evidence on BHRRs. In Table 3 we report BHRRs for the samples shown in Figure 2 over several windows around experiment announcement: (i) days −10 to +1, (ii) days −1 to +1, and (iii) day 0. We also report *t*-statistics for a two-sample difference in means between pilot and control firms. The BHRRs for the (−1, +1) window and the announcement date are small and statistically insignificant across samples. For the longer (−10, +1) window, the BHRRs are negative and statistically significant only for two of the seven samples: small firms (based on assets) and small Nasdaq firms within the GMW best match sample. For all GMW best match sample small firms (based on assets), the BHRR is −1.55% and barely statistically significant

Figure 2. (Color online) BHRRs for Different Samples



Notes. Figure shows BHRRs for pilot versus control firms in indicated samples over June 1, 2004–September 30, 2004. Firms are equally weighted. Vertical lines indicate SEC announcement of experiment approval on June 23, 2004, and the SEC formal experiment announcement on July 28, 2004. Shaded area indicates principal event period used by GMW: (−10, +1) relative to the formal announcement date. Samples are 2004 announcement sample; 2004 announcement sample, small firms (based on market capitalization); GMW best match sample small firms (based on assets); GMW best match sample small firms (based on market capitalization); GMW best match sample small firms (based on trading volume); GMW best match sample small firms (based on assets), NYSE firms; and GMW best match sample small firms (based on assets), Nasdaq firms.

Table 3. Buy-and-Hold Relative Returns Around SEC Experiment Announcement

Event window	(−10, +1)	(−1, +1)	Day 0
Sample			
2004 announcement, all firms	−0.0060	−0.0011	0.0011
	(1.75)	(0.56)	(0.98)
2004 announcement, small firms (based on market cap)	−0.0067	−0.0055	−0.0001
	(1.16)	(1.61)	(0.07)
GMW best match small firms (based on assets)	**−0.0155****	−0.0042	0.0000
	(2.02)	(1.01)	(0.00)
GMW best match small firms (based on market cap)	−0.0096	−0.0032	0.0001
	(1.27)	(0.78)	(0.06)
GMW best match small firms (based on trading volume)	−0.0006	−0.0022	0.0015
	(0.10)	(0.69)	(0.85)
GMW best match small NYSE firms (based on assets)	0.0044	0.0050	0.0000
	(0.41)	(0.80)	(0.00)
GMW best match small Nasdaq firms (based on assets)	**−0.0224****	−0.0077	0.0001
	(2.31)	(1.50)	(0.04)
GMW reported (all firms)	−0.0152	−0.0017	0.0004
GMW reported (small firms, based on assets)	−0.0235	−0.0092	−0.0021

Notes. BHRRs for pilot firms relative to control firms for indicated samples over event windows covering trading days (−10, +1) and (−1, +1) relative to SEC announcement on July 28, 2004, plus the announcement date (day 0) for the indicated samples. *t*-statistics (in parentheses) are for two-sample difference in means. Bottom two rows report BHRR's reported by GMW. They do not report statistical significance. Significant results at 5% level or better in boldface.
*, **, *** indicates significance at the 10%, 5% and 1% levels, respectively.

($t = 2.02$). By comparison, GMW report a BHRR of −2.35% for small firms (based on assets).[12]

Overall, we find limited evidence for a share price drop for pilot firms prior to the experiment announcement, which is not robust to sample choice and is driven by negative returns to small Nasdaq firms even though the Regulation SHO experiment was expected to have a larger impact on NYSE firms. Thus, we do not find support for a causal channel that runs through stock returns.

2.4.3. Access to the SEC's List of Pilot Firms Prior to Public Announcement. There is also a causal channel question in attributing negative preannouncement returns of the pilot firms to the experiment. To trade prior to the announcement, someone would need access to the SEC's list of pilot firms prior to public release of this list. This seems implausible.[13] Further, we find no evidence of a price reaction either over a narrower (−1, +1) window or on the day when the SEC released the actual list. This further suggests that one cannot treat the modest negative preannouncement relative returns for some but not other samples of small firms as a true response to the experiment.

2.5. Causal Channel III: Managerial Fear

FHK and a number of other studies rely, as a causal channel, on the assertion that managers changed their behavior because of fear of increased short selling.[14] This channel cannot be directly tested, but we can assess the plausibility of this channel and whether the fear was likely to be widespread. Managers would have to believe that the price tests were an important impediment to fundamentals-based short selling. This would be contrary to widespread beliefs at the time, discussed earlier, that, by 2004, the price tests had little effect on fundamentals-based short selling.

But suppose many managers worried (enough to drive the results that FHK and others report) nonetheless. How would they be likely to react? If managers felt that the experiment would have an important negative effect on their firms, some of them would have written to the SEC expressing their opposition—either individually or through their counsel or a trade group—and there would have been significant media coverage of manager opposition. An FASB proposal to require expensing of the fair value of employee stock options provides a relevant example. This proposal provoked major, well-publicized opposition from many firm managers, including extensive press coverage and an appeal to Congress to override the proposed rule.[15] Thus, we undertook a detailed examination of comment letters to the SEC on the proposed experiment and press coverage of the experiment.

We searched the business press (including the Dow Jones News Service (DJNS), Bloomberg, *The Wall Street Journal* (WSJ), and *The New York Times* (NYT)) for news stories and other information about the experiment during the prelaunch period (September 2003–April 2005) and the experiment period (May 2005–July 2007) and present results of the search in Internet Table IA-7, panel A. We found very little business press attention to the experiment. The SEC's June 23, 2004, approval of the experiment was covered in a short WSJ story, which does not mention manager opposition. The story quotes the head of the SEC's division of market regulation as saying that the SEC had expanded the pilot program relative to the proposing release because of overwhelming comment from the industry supporting an expanded program. The formal announcement of the pilot was not covered in the standard business news sources. Coverage between the announcement and the launch was sparse and contained only technical explanations of how the experiment would work. The experiment launch, in May 2005, was noted in a DJNS story a few days earlier with a WSJ summary the next day. In 2006, the SEC extended the experiment, originally planned for one year, for a second year with minimal press attention and no apparent controversy. The SEC's December 2006 proposal to repeal the short-sale rule also attracted no apparent opposition. A *New York Times* story about the repeal explains:[16] "You may not have read of this proposal. It was virtually ignored by the news media, and if any companies are upset about it, they have not made themselves known. A pilot program that exempted some companies from the so-called uptick rule starting in 2005 drew little attention." This is the only *New York Times* story about the experiment we found.

The SEC formally approved the repeal in June 2007. A *Wall Street Journal* story on the repeal explained that the rule had become "more of an annoyance than a hindrance" to short sellers, discussed researchers' view that "the uptick rule's usefulness has disappeared," and did not mention any opposition to the repeal.[17] The official repeal announcement, on July 3, 2007, received no press coverage.[18]

As a further check on whether there was substantial managerial concern, we reviewed all comment letters the SEC received for its November 2003 proposal for the short-sale experiment.[19] Internet Table IA-7, panel C, lists all comment letters from organizations (as opposed to individuals) and indicates whether they supported the experiment without changes, supported the experiment but proposed change, or opposed the experiment.[20]

As the SEC notes in the adopting release, most of the comments it received supported the experiment.[21] Of the 23 comments that the SEC received on the experiment from organizations, 20 were favorable.[22] The only negative letters came from the NYSE, the NYSE's specialist association, and a small quantitative trading firm, Susquehanna International Group. Nasdaq supported the experiment as did several other exchanges and trading platforms (Chicago Board Options Exchange, Chicago Stock Exchange, Archipelago Holdings), major banking and investment banking firms (Charles Schwab, Citigroup, Goldman Sachs, JPMorgan Securities, Merrill Lynch, Morgan Stanley, UBS Securities), and mutual fund groups (Investment Company Institute, Managed Funds Association). Many comments supported expanding the number of firms to be covered by the experiment—a change that the SEC adopted. There were no comments from individual firms or business groups other than the favorable comments from financial firms.

We also reviewed all comments received by the SEC on its 2006 proposal to repeal the rule.[23] Internet Table IA-7, panel C, summarizes the nine comments from organizations; all are favorable. All supported the repeal, including the NYSE, which had originally opposed the pilot.[24] Two comments recommended retaining the uptick rule for small-cap firms, which had not been part of the experiment. Once again, the SEC received no comments from individual firms or business groups.

We view this examination of comment letters as confirming financial industry support for the experiment, no evidence of opposition from firm managers, and more broadly lack of evidence that firm managers viewed the short-sale rule as important. This analysis cannot disprove the existence of managerial concern but is consistent with lack of widespread managerial fear.

Moreover, even if managers were initially worried, they should have realized by 2006 and even more so in 2007 that there had been no increase in short interest and no news reports suggesting a higher frequency of bear raids against pilot firms. This suggests that initial effects on pilot firms, if any, would likely diminish over time. Yet each of the four reexamined papers finds larger effects in 2007, and the statistical significance of the reported results depends on the 2007 magnitudes.[25] It is difficult to causally attribute this pattern to the experiment, which increases the likelihood that the reported results are false positives.

In sum, there are several substantive reasons to discount the likely magnitude of a potential fear channel. These include minimal news attention to the experiment, no news stories suggesting manager opposition, no comment letters from firm managers to the SEC opposing the experiment, and observed treatment effects that are often stronger in 2007 than in 2005 or 2006.

2.6. Other Possible Causal Channels

The three channels discussed are the principal ones relied on by the indirect effects studies, but other causal channels are possible. LLS and a few other papers posit a price efficiency channel, in which more short selling makes market prices more informative. The managers realize this and, therefore, rely more heavily on market prices when making decisions. We do not reassess this channel as several prior studies, using a variety of proxies for price efficiency, did not find evidence of improved price efficiency for pilot firms. Moreover, limited manager attention to the experiment generally, discussed for the manager fear channel, makes it unlikely that managers would change their behavior based on assumed greater price efficiency.

3. Reexamination Methodology

Having reviewed the evidence on causal channels, we turn to reexamining core results from the four studies mentioned earlier. This section presents our methodology, and Section 4 presents results.

3.1. Sample Selection for Reexamination of FHK, GMW, HHZ, and LLS

To study the effects of the actual experiment, one must update the 2004 announcement sample, which we use in Section 2 to study the effects of the experiment announcement, to account for firms that ceased trading between experiment announcement and launch (principally because of acquisitions). We also exclude financial and utility firms. These changes led to a 2005 analysis sample, comprising 2,115 firms, which closely follows the SEC's experiment design and is substantially larger than those in the studies we reexamine. This may be one reason why our results diverge from the papers we reexamine.[26] HHZ, GMW, and FHK all impose additional restrictions that reduce sample size. HHZ and GMW require firms to be in the R3000 in both 2004 and 2005, and FHK use a balanced sample of firms with data from 2001 to 2010. LLS do not explain their sample selection process.

3.2. Specification Choices

We constructed the 2004 announcement and 2005 analysis samples closely following the SEC's rules. We necessarily made many specification choices, but made them before running any regressions and consistently across the four reexaminations.

3.2.1. Firm and Year Fixed Effects. We organize firm-year observations based on fiscal year and use firm and fiscal year FE, which is a standard DiD specification. GMW and LLS also use firm and fiscal year FE; HHZ use firm FE but not year FE; FHK do not use firm FE. FHK's main model also does not include year FE, but they report similar results with year FE.

3.2.2. Sample Without Covariates and Are Covariates Appropriate? The reexamined papers use a variety of covariates and also report either univariate results (FHK, HHZ, GMW) or results with firm and year FE but without covariates (LLS). We do not include time-varying covariates in our preferred specification. This increases sample size and avoids potential bias from using covariates that might be affected by the experiment. All of the covariates used in the four studies are potentially endogenous; some are directly so under the authors' own hypotheses.[27] Given the initial randomization, one can obtain unbiased estimates without covariates. In practice, including their covariates makes very little difference in coefficient estimates or precision (see the internet appendix).[28]

3.2.3. Balanced vs. Unbalanced Panel. Given the randomized nature of the experiment, we prefer an unbalanced panel. This increases sample size and is appropriate because there is no reason to expect differential attrition between pilot and control firms. We confirm that there is no differential attrition.[29] HHZ, GMW, and LLS all use unbalanced samples. FHK use a balanced panel of firms with data available throughout 2001–2010. They state that they found similar results with an unbalanced panel, but we show below, their results do not survive with an unbalanced panel.

3.2.4. Sample Periods and Transition Period. Each reexamined paper makes a different choice regarding the pre, during, and post periods and whether to exclude a transition period (roughly 2004) between experiment announcement and launch (FHK, HHZ, and LLS do so) and include a post period in the study (FHK and HHZ do so).

We use fiscal years 2001–2010 as our sample period. We define which firm fiscal years are in this period using the Compustat convention under which, if the fiscal year-end month is January–May, the fiscal year is (calendar year $-$ 1), and if the fiscal year-end month falls in June through December, the fiscal year is the calendar year in which the fiscal year ends. Thus, our sample period includes fiscal year ends from June 2001 through May 2011.

The logic for including a post period is compelling. Any treated-minus-control difference found during the experiment should reverse once the experiment ends and the price tests are removed for all firms. Observing whether the reversal occurs is an important robustness check. Because the experiment started in May 2005, it is unclear if 2004 should be excluded; this may depend on the posited causal channel.[30]

Judgment is needed on which firm fiscal years fall within each period. The experiment ran from May 2005 to July 2007.[31] We treat firm fiscal years for which six or more months fall within the experiment period as within the during period (fiscal years ending October 2005 through December 2007). Earlier fiscal years are part of the preperiod, and later fiscal years are part of the postperiod. We include the period from experiment announcement in July 2004 through its launch in May 2005 in the preperiod. In contrast, FHK, HHZ, and LLS exclude this period from the analysis.

3.2.5. Winsorization. We follow common practice in the accounting literature and winsorize outcomes and most covariates at the 1%/99% level across all sample years; this reduces the impact of outlier observations and can address data errors in Compustat (Hribar and Nichols 2007). FHK, GMW, and HHZ also winsorize at this level; LLS winsorize one of their outcomes at this level, but not their other outcomes. See Internet Table IA-4 for winsorization details.

3.3. Methodology for FHK and Measures of Earnings Management

FHK use only one proxy for earnings management: PMDA. The use of this measure was advocated by Kothari et al. (2005) when studying accruals for firms with extreme performance to account for the correlation between performance and accruals. This measure takes residuals from the modified Jones model of accruals (Dechow et al. 1995), which are also termed discretionary accruals or abnormal accruals (AA), and subtracts corresponding residuals of a matching firm selected based on industry and performance (return on assets (ROA)). However, the original and modified Jones models explain only about 10%–12% of the variation in firm-level accruals. Hence, the PMDA measure is likely to add noise to the measure of discretionary accruals (Dechow et al. 2010). Consistent with this intuition, Dechow et al. (2011) show that these models have lower power to detect earnings management relative to unadjusted measures of accruals, such as operating and total accruals. Furthermore, PMDA has the lowest power among the various measures of accruals to detect earnings management, and hence, its use is recommended only when studying a sample of firms with extreme performance (Dechow et al. 2010).

Given this limitation of PMDA, we also study three simpler measures of accruals—operating accruals, total accruals (comprehensive measure of accruals), and AA—and report results in the internet appendix. Given the initial randomization, the pilot and control samples are balanced on firm characteristics, and hence, these simpler measures are unbiased. At a minimum, they provide useful robustness checks.

To compute AA and PMDA, we follow FHK and estimate accruals cross-sectionally within each fiscal year and Fama–French 48 industry, using the modified Jones model:

$$\frac{TA_{i,t}}{AT_{i,\,t-1}} = \beta_0 + \beta_1 \frac{1}{AT_{i,\,t-1}} + \beta_2 \frac{\Delta REV_{i,t}}{AT_{i,\,t-1}} + \beta_3 \frac{PPE_{i,t}}{AT_{i,\,t-1}} + \varepsilon_{i,\,t}, \tag{2}$$

where i indexes firms and t indexes fiscal years. TA_t is earnings before extraordinary items (IBC) minus operating cash flows (OANCF minus XIDOC) for year t. $AT_{t\text{-}1}$ is total assets at the end of fiscal year $t-1$. ΔREV_t is the change in sales revenue (SALE) from year $t-1$ to t. PPE_t is gross property, plant, and equipment (PPEGT) at the end of year t. We estimate Equation (2), requiring a minimum of 10 industry-year observations and use the estimated coefficients to calculate normal accruals $NA_{i,t}$:

$$NA_{i,t} = \widehat{\beta}_0 + \widehat{\beta}_1 \frac{1}{AT_{i,\,t-1}} + \widehat{\beta}_2 \frac{(\Delta REV_{i,t} - \Delta AR_{i,t})}{AT_{i,\,t-1}} + \widehat{\beta}_3 \frac{PPE_{i,t}}{AT_{i,\,t-1}}, \tag{3}$$

where $\Delta AR_{i,t}$ is the change in accounts receivables (RECT). We then calculate firm-year abnormal accruals and PMDA as[32]

$$\begin{aligned} &AA_{i,t} = (TA_{it}/AT_{i,t-1}) - NA_{i,t}. \\ &PMDA_{i,t} = AA_{i,t} - AA_{j,t}\, for\ matched\ firm\ j. \end{aligned} \tag{4}$$

To find the matching firm j, we follow Kothari et al. (2005) and match each firm-year observation with an observation from another firm in the same industry with the closest same-year return on assets, defined as net income divided by total assets (AT_t).[33]

3.4. DiD Specification

3.4.1. Simple DiD Specification. To test whether pilot firms had lower accruals than control firms during the experiment period, we estimate the following DiD model for each of the accrual measures over 2001–2010:

$$y_{i,t} = \beta_0 + \gamma_t + f_i + \beta_1\ Pilot_i * During_t + \beta_2\ Pilot_i * Post_t + \varepsilon_{i,t}. \tag{5}$$

Here, $y_{i,t}$ is the accruals measure, $Pilot_i = 1$ for pilot (treated) firms and 0 for control firms, *During* is a dummy variable for the experiment period, *Post* is a dummy variable for the postexperiment period, and the γ_t and f_i are year and firm FE. Noninteracted terms are omitted because they are absorbed by the firm and year FE. A negative coefficient β_1 on *Pilot* × *During* provides evidence that pilot firms reduce earnings management during the experiment period. The expected coefficient β_2 on *Pilot* × *Post* should be close to zero because short-sale restrictions were removed for all firms following the experiment. We can also test for a sign reversal in the postperiod, relative to the experiment period by replacing *Pilot* × *During* with *Pilot* × (*During* or *Post*) in Equation (5). With this specification, in Equation (6), One should expect a positive coefficient on *Pilot* × *Post*, similar in magnitude to the negative coefficient on *Pilot* × *During* in Equation (5).

$$y_{i,t} = \beta_0 + \gamma_t + f_i + \beta_1\ Pilot_i * (During\ or\ Post)_t + \beta_3\ Pilot_i * Post_t + \varepsilon_{i,t}. \tag{6}$$

We cluster standard errors by firm. In the internet appendix, we also use a specification with covariates in which we add $\boldsymbol{\lambda} \times \mathbf{x}_{i,t}$ to Equations (5) and (6), where $\mathbf{x}_{i,t}$ is a vector of covariates (for each i, t) and $\boldsymbol{\lambda}$ is a coefficient vector.

3.4.2. Annual Differences and Leads-and-Lags Specification. To both assess whether pretreatment trends are parallel and allow for treatment effect to emerge gradually during the treatment period, we use two graphical approaches. The first is a plot of univariate differences in means for pilot versus control firms. When data are available, we extend these plots back to 1998. A longer pretreatment period is useful in identifying nonparallel pretreatment trends and in assessing random variation in means between the two groups during the preexperiment period. The second is a leads-and-lags specification, in which we estimate a separate treatment effect for each year before, during, and after the experiment period and plot annual coefficients and 95% confidence intervals. We present univariate graphs in the text and lead-and-lags graphs in the internet appendix; inference from both is similar.

3.5. GMW Specifications

GMW study five outcomes, capital expenditures (capex)/assets, (capex + R&D)/assets, percent asset growth, equity issuance, and debt issuance, for all firms and for small firms. We discuss in the text the three outcomes for which they report the strongest results: capex/assets, percent asset growth, and equity issuance. We discuss their other measures in the internet appendix. Our DiD specifications use Equations (5) and (6). GMW do not examine reversal of the treatment effect after the experiment ends. We study both the during and post periods.

3.6. HHZ Specifications

HHZ study only one outcome, *ln*(audit fees). Our DiD specifications use Equations (5) and (6), with *ln*(audit fees) as the dependent variable. Our first specification uses firm and fiscal year FE but no covariates. A good case can be made for controlling for firm size because size is a known, powerful predictor of audit fees even though an effect on growth is one possible outcome of the experiment. Therefore, in a second regression model, we control for *ln*(assets). We also report models that include, respectively, the short list of covariates in HHZ, table 5, model 1 (which controls for *ln*(sales) as a measure of size) and the longer list in HHZ, table 5, model 2.

3.7. LLS and Triple Difference Specification

LLS study WPS using a DiD specification and also R&D/assets and (capex + R&D)/assets using a triple difference specification with $Pilot \times During$ interacted with Tobin's q. We discuss WPS and (capex + R&D)/Assets in the text and discuss R&D/Assets in the internet appendix. For WPS, we again rely on Equations (5) and (6). For the other outcomes, we replace Equation (5) with a triple difference specification:

$$y_{i,t} = \beta_0 + \gamma_t + f_i + \beta_1 * q_{i,t} + \boldsymbol{\gamma} * \boldsymbol{DBL_{i.t}} + \beta_2 * q_{i,t} * Pilot_i * During_t + \beta_3 * q_{i,t} * Pilot_i * Post_t + \varepsilon_{i,t}. \quad (7)$$

Here, q is Tobin's q, **DBL** is a vector of the double interactions: $Pilot_i \times During_t$, $Pilot_i \times Post_t$, $q_{i,t} \times Pilot_i$, $q_{i,t} \times During_t$, and $q_{i,t} \times Post_t$ and $\boldsymbol{\gamma}$ is the corresponding vector of coefficients. The core coefficients are those on the triple interactions, β_2 and β_3. We make analogous changes to Equation (6) to measure sign reversal. LLS do not study whether results during the experiment period reverse after the experiment ends. We study both the during and the post periods.

4. Reexamination Results

In this section, we reexamine core results from each paper. Our focus is on the robustness of the reported results to alternative specifications. We present results using (i) our research design and sample and (ii) best match samples and specifications that match those from each paper as closely as we could based on what was stated in each paper. We constructed the FHK and HHZ best match specifications without knowing the authors' actual choices and samples, which we obtained later from the authors. We briefly discuss the technical replication of FHK and HHZ in Section 4.6. Using their data sets and code, we are able to technically replicate FHK's PMDA result and HHZ's audit fees result. However, these results are fragile. We discuss the robustness of the FHK and HHZ results in the internet appendix.

4.1. Summary Statistics and Pretreatment Balance

We define variables and confirm that pilot and control firms are similar for both pretreatment outcomes and covariates in Internet Table IA-4. Good balance was expected given the initial randomization.

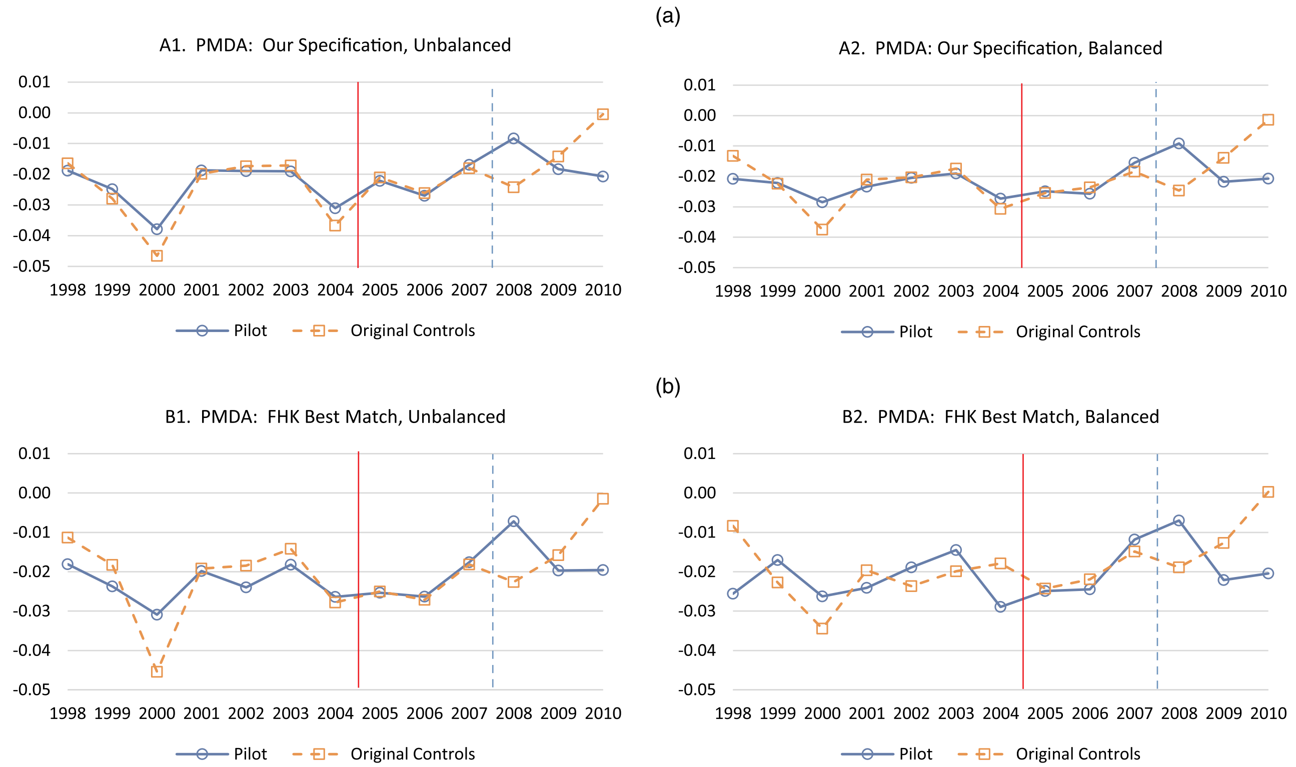
4.2. FHK: Results for Accruals

For FHK, we present results in the text for PMDA, using both unbalanced and balanced panels. FHK report results using a balanced panel but assert in a footnote that unbalanced panel results are similar. For the balanced panel results, we require the FHK covariates (*ln*(Total Assets), Market-to-book ratio, Return on assets, and Leverage) to be nonmissing for all sample years but do not include them in regressions. We present in the internet appendix results for operating accruals, total accruals, and AA with both balanced and unbalanced panels.

4.2.1. Graphical and Regression Evidence. We begin our reexamination with graphical evidence. Given the initial randomization with no evidence of differential attrition, a simple comparison of means can provide valuable information. In Figure 3, we provide annual means for PMDA, separately for pilot and control firms, for both our specification and the FHK best match specification. There is no evidence of a treatment effect with either specification.

Table 4, panel A, presents regression results for PMDA for both unbalanced and balanced panels using our specification. We report coefficients for $Pilot \times During$, $Pilot \times Post$, and sign reversal (the change in the coefficient from during to post). Panel B presents results using the

Figure 3. (Color online) Fang et al. (2016) Annual Means for Performance-Matched Discretionary Accruals



Notes. Figures show annual means separately for pilot and control firms for PMDA over 1998–2010. Left-hand figures use unbalanced panel; right-hand figures use balanced panel. Panel (a) is our specification. Panel (b) is the Fang et al. (2016) best match specification. Solid and dashed vertical lines separate pre, during, and post periods.

FHK best match specification. Panel C presents the FHK reported result for comparison. In both panels A and B, the coefficients on *Pilot* × *During* are insignificant. Moreover, the coefficients have a positive sign (opposite from predicted) for the unbalanced panel (both specifications) and for a balanced panel with our specification. Thus, neither graphical nor regression evidence supports the FHK conjecture with either our specification or the FHK best match specification.

The results in Table 4 also bear on the economic significance of the reported results. A true decrease in earnings management of 1% of total assets per year would be economically meaningful. However, the coefficient estimates vary substantially across reasonable specifications, ranging from +0.0075 (opposite from predicted) to −0.01 (FHK's reported coefficient). This large range, including change in sign, suggests caution in treating the reported coefficient as reliable[34]

Further reason to doubt economic significance comes from Internet Figure IA-7. This figure shows that there is little change in AA of the pilot or control firms during the experiment period. The negative PMDA coefficient in FHK is driven by differences in the AA of the matching firms, and even those differences disappear if one selects matching firms based on concurrent ROA as opposed to lagged ROA.

4.2.2. One- vs. Two-Way Clustering.

In Table 4, we report s.e. that are clustered on firm and standard deviations (s.d.) based on randomization inference (explained subsequently). FHK cluster standard errors on both firm and fiscal year. The other three papers use standard errors clustered on firm as does our specification. FHK have a short time dimension, hence a small number of year clusters, and clustered standard errors with a small number of clusters can be downward biased (Cameron et al. 2008). We use randomization inference to assess the validity of the s.e. that FHK report, which are much lower than ours (0.004 versus our 0.0087).

To implement randomization inference for each sample of pilot and control firms in panel A of Table 4, we drew at random the correct number of pilot firms. For example, with our specification and a balanced panel, the PMDA sample includes 1,398 firms (492 pilot and 906 control). We drew 492 of these firms at random, assigned them to be pseudo-pilot firms, and computed the pseudo-coefficients on *Pilot* × *During* and *Pilot* × *Post* using the same regression specification as in Table 4 (from Equation (5)). We repeated this procedure 1,000 times to obtain an empirical distribution of the pseudo-coefficients and measured the standard deviation. The randomization inference s.d. are very close to s.e. with firm clusters and roughly twice as large as the s.e. that

Table 4. Fang et al. (2016) Accruals Measures

Outcome: Performance matched discretionary accruals			
Panel	Predicted sign	Unbalanced	Balanced
Panel A. Our specification			
Pilot × During	Negative	0.0029	0.0075
s.d. (rand. inf.)		(0.0079)	(0.0087)
s.e. (cluster on firm)		[0.0080]	[0.0087]
Pilot × Post	Near zero	−0.0037	−0.0007
s.d. (rand. inf.)		(0.0081)	(0.0083)
s.e. (cluster on firm)		[0.0081]	[0.0084]
Sign Reversal	Positive	−0.0066	−0.0081
s.e. (cluster on firm)		(0.0082)	(0.0083)
Firm-year observations		18,375	13,980
Pilot (control) firms		698 (1,401)	492 (906)
R^2 (within)		0.1%	0.1%
Panel B. Fang et al. (2016) best match specification			
Pilot × During	Negative	0.0038	−0.0022
s.d. (rand. inf.)		(0.0084)	(0.0094)
s.e. (2-way cluster)		.	[0.0030]
Pilot × Post	Near zero	0.0018	−0.0079
s.d. (rand. inf.)		(0.0087)	(0.0090)
s.e. (two-way cluster)		[0.0084]	[0.0086]
Sign reversal	Positive	−0.0020	−0.0057
s.e (cluster on firm)		[0.0086]	[0.0091]
Firm-year observations		16,074	11,808
Pilot (control) firms		697 (1,400)	466 (846)
Adjusted R^2		−0.0%	−0.0%
Panel C. Fang et al. (2016) results as reported			
Pilot × During			**−0.010****
s.e. (two-way cluster)			**[0.004]**
Pilot × Post			0.004
s.e. (two-way cluster)			[0.004]
Firm-year observations			9,873
Pilot (control) firms			388 (709)
Adjusted R^2			0.10%

Notes. (Panel A) Regressions, using our specification with firm and fiscal year fixed effects of performance matched discretionary accruals PMDA on *Pilot × During*, *Pilot × Post*, and constant term over fiscal years 2001–2010, following Equation (5). During and Post periods are defined in the text. (Panel B) Regressions are similar to panel A, but use Fang et al. (2016) best match specification. Two-way clustering is on firm and calendar year, using cluster2.ado. (Panel C) Fang et al. (2016) reported results (their table III, model (1)). All panels. Sign reversal = coefficient. on (*Pilot × Post*) minus coefficient on (*Pilot × During*), using Equation (6). Coefficient on constant term is suppressed. Unbalanced panel uses all nonmissing firm-year observations. Balanced panel requires outcome to be nonmissing for all sample years. Covariates (must be nonmissing for balanced panel but not included in regressions) are *ln(Total Assets)*, *Market-to-book ratio*, *ROA*, and *Leverage*. Randomization inference s.d.s based on 1,000 repetitions in parentheses; s.e. with indicated clustering in brackets. Significant results at 5% level or better in boldface.

*, **, *** indicates significance at the 10%, 5%, and 1% level, respectively.

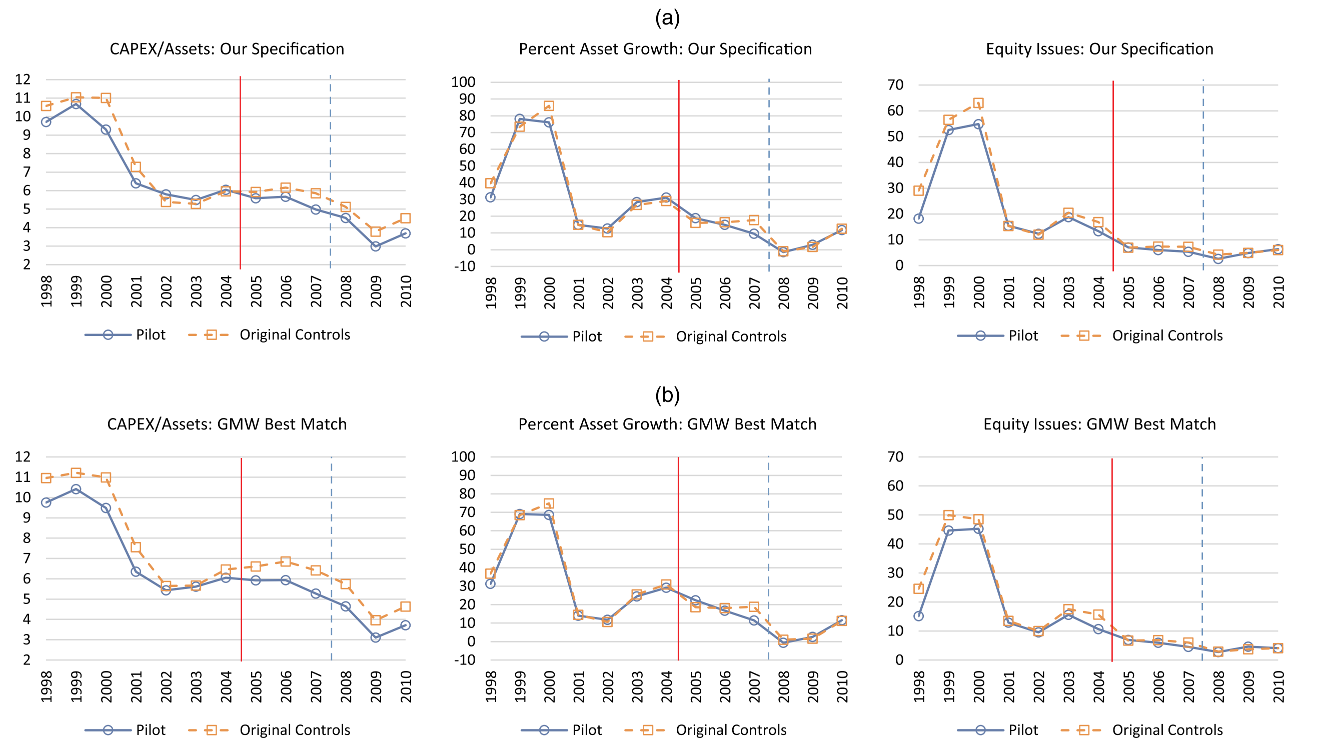
FHK report.[35] Thus, downward bias in two-way clustered s.e. drives the statistical significance for their reported coefficient.

4.3. GMW: Firm Investment and Growth

GMW report evidence of lower investment (measured by capex/assets), lower asset growth, and lower equity issuance for small firms (based on assets). We begin our reexamination with graphical evidence. In Figure 4, we present the annual means for small firms, using both our specification (top graphs) and the GMW best match specification (bottom graphs). Both sets of graphs are visually similar. We find some evidence for reduced investment and no evidence for slower growth or less equity issuance.

Whereas there is some evidence for a reduction in investment, several findings are not easily reconciled with a direct effect of the experiment. First, the effect is largest in 2007—more than two years after the experiment started and, for the many firms with fiscal-year

Figure 4. (Color online) Grullon et al. (2015) Annual Means for Outcomes for Small Firms (Based on Assets)



Notes. Figures show annual univariate means for indicated outcomes separately for small pilot and control firms (below median in assets in most recent fiscal year ending prior to July 28, 2004). Panel (a) is our specification (2005 analysis sample). Panel (b) is the Grullon et al. (2015) best match specification. In both panels, solid and dashed vertical lines separate pre, during, and post periods.

Table 5. Grullon et al. (2015) Investment and Capital Raising

Dependent variable (%)	Predicted sign	CAPEX/assets (1)	Percentage asset growth (2)	Equity issues (3)
		Panel A. Our Specification		
Pilot × During	Negative	−0.496*	−2.423	−0.723
		(1.68)	(0.91)	(0.46)
Pilot × Post	Near zero (implied)	−0.383	1.210	1.249
		(1.10)	(0.49)	(0.76)
Firm-year observations		9,083	9,115	8,921
Pilot (control) firms		337 (719)	337 (719)	332 (711)
R^2 (within)		0.050	0.044	0.034
		Panel B. Grullon et al. (2015) best match specification, no covariates		
B1. With firm and fiscal year FE	Negative	**−0.644****	−0.532	0.505
		(2.06)	(0.20)	(0.37)
B2. Without firm or fiscal year FE	Negative	−0.612*	−0.903	0.527
		(−1.86)	(−0.34)	(0.39)
Pilot (control) firms		308 (660)	308 (660)	302 (652)
		Panel C. Grullon et al. (2015) univariate results as reported		
Pilot × During (univariate, no firm or fiscal year FE)	Negative	**−0.97**	**−6.12**	−1.61
		(2.88)***	**(2.55)****	(1.82)*
Pilot (control) firms		313 (629)	306 (622)	302 (606)

Notes. (Panel A) Regressions with firm and fiscal year fixed effects of indicated dependent variables (defined in Internet Table IA-4) on Pilot × During, Pilot × Post, and constant term, using our specification and small firm sample (using Grullon et al. (2015) definition, below-median in assets in most recent fiscal year ending prior to July 28, 2004). During and Post periods are defined in the text. (Panel B) Same but uses the Grullon et al. (2015) best match specification. (Panel C) Grullon et al. (2015) univariate results from their table 5. All panels. Coefficient on constant term is suppressed. We follow Grullon et al. (2015) and multiply by 100 to convert amounts to percentages. *t*-statistics with standard errors clustered on firm in parentheses. Significant results at 5% level or better in boldface.

*, **, *** indicates significance at the 10%, 5%, and 1% level, respectively.

ends in December, well after the experiment had ended. Second, there is a similar gap in investment between treated and control firms in the 2000–2001 period. Third, this gap does not close once the experiment is reversed. For percentage asset growth, pilot and control firms are similar in 2005 and 2006 with pilot firms showing a relative dip in 2007. As discussed, it is difficult to causally attribute effects that emerge or increase in 2007 to the experiment.

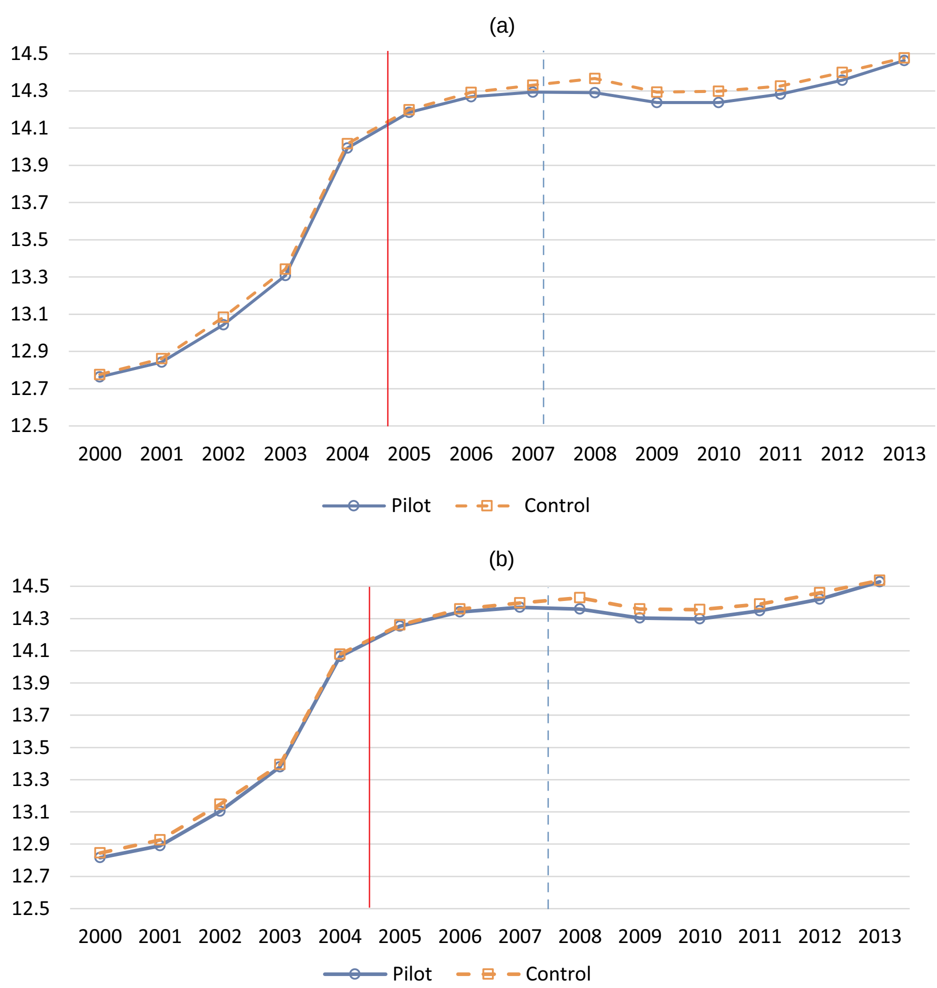
We provide DiD regression results for small firms in Table 5. Panel A provides results for our specification, panel B switches to the GMW best match specification, and panel C provides the GMW reported results for comparison.[36]

With our specification, all three outcomes in Table 5 are insignificant although with negative coefficients. With the best match specification, the coefficient for capex/assets is negative and mildly significant (t-statistic = 2.06). However, significance is driven by a gap that increases in 2006 and again in 2007 with no postexperiment reversal (Figure 4). Percentage asset growth is insignificant, and equity issuance takes a positive coefficient (opposite from predicted).

To better understand whether smaller firms reduced investment and growth, we reran the DiD regressions after dividing the sample into size quintiles based on assets (see Internet Table IA-26). The coefficient on capex/assets is negative and significant with the GMW best match specification only for the smallest quintile (coefficient = −1.22, t-statistic = 2.60) but is insignificant if quintiles are chosen based on market capitalization. Furthermore, of the 115 pilot firms in the smallest quintile (based on assets), 108 are Nasdaq firms. Yet, as discussed earlier, the suspension of the price tests had little impact on Nasdaq firms. No significant drop in asset growth or equity issuance is observed for any of the size quintiles. Thus, the quintile analysis provides little support for a causal effect of the experiment on investment.

With regard to economic significance, GMW report a point estimate of a 0.97% reduction in capex/assets ratio and 6.1% lower asset growth for small pilot firms

Figure 5. (Color online) Hope et al. (2017) Annual Means for *ln*(Audit Fees)



Notes. Figure shows univariate means separately for treated and control firms of *ln*(audit fees), winsorized at 1%/99%, using our specification (panel (a)) and the Hope et al. (2017) best match specification (panel (b)). Solid and dashed vertical lines separate pre, during, and post periods.

relative to control firms during the experiment period. Given that the mean capex/assets ratio (percentage of asset growth) for pilot firms prior to the experiment was 5.7% (20.4%), a decline of 0.97% (6.1%) seems large as a response to their estimate of a −2.35% announcement period return for small pilot firms.[37]

In sum, considering both our specification and GMW best match specification and both graphic and regression evidence, there is very limited evidence for the GMW hypothesis of lower investment and slower growth by small pilot firms (based on total assets only). Moreover, the only significant result is driven by Nasdaq firms, for whom the direct impact of the suspension of the bid test was minimal, making it difficult to causally attribute the result to the experiment.

4.4. HHZ: Auditing Fees

HHZ report that pilot firms have about 5% higher audit fees during the experiment period with bare statistical significance (insignificant for univariate difference in means, $t = 1.99$ with limited covariates and $t = 1.98$ with more extensive covariates).

In Figure 5, we provide annual means for *ln*(audit fees), separately for pilot and control firms, for both our specification (panel (a)) and the HHZ best match specification (panel (b)). With both specifications, the two sets of means move closely in parallel with each other with no evidence of a treatment effect. The pilot firm mean is somewhat below the control firm mean for 2007, but as discussed earlier, it is difficult to causally attribute an effect in 2007 to the experiment. Moreover, the gap does not close in the post period.

We present regression results in Table 6, using an unbalanced panel as do HHZ. We present results for our specification in panel A, results with the HHZ best match specification in panel B, and the HHZ reported results in panel C. Column (1) with firm and fiscal year FE but no covariates is our preferred specification; column (2), which controls for *ln*(assets) (which correlate strongly with audit fees) is a sensible alternative. In column (3), we

Table 6. Hope et al. (2017) Audit Fees

	Dependent variable = *ln*(Audit Fees)	Predicted sign	(1)	(2)	(3)	(4)
Panel	Hope, Hu, and Zhao covariates		No	No	Limited	Full
	ln(Assets) as covariate		No	Yes	No	No
Panel A. Our specification						
	Pilot × *During*	Positive	−0.0046	−0.0088	−0.0086	−0.0132
			(−0.24)	(−0.48)	(−0.48)	(−0.75)
	Pilot × *Post*	Near zero	−0.0214	−0.0140	−0.0177	−0.0274
			(−0.85)	(−0.65)	(−0.85)	(−1.34)
	Sign reversal	Negative	−0.0167	−0.0053	−0.0090	−0.0141
			(−1.00)	(−0.36)	(−0.61)	(−0.97)
	Firm-year observations		17,973	17,964	17,651	16,985
	Pilot (control) firms		683 (1,375)	683 (1,375)	682 (1,368)	666 (1,335)
	R^2 (within)		0.717	0.756	0.761	0.772
Panel B. Hope et al. (2017) best match specification						
	Pilot × *During*	Positive	0.0151	0.0107	0.0135	0.0116
			(0.62)	(0.48)	(0.62)	(0.54)
	Pilot × *Post*	Near zero	−0.0025	0.0120	0.0012	−0.0051
			(−0.09)	(0.50)	(0.05)	(−0.23)
	Firm-year observations		18,924	18,924	18,924	18,924
	Pilot (control) firms		615 (1,233)	615 (1,233)	615 (1,233)	615 (1,233)
	R^2 (within)		0.716	0.774	0.775	0.789
Panel C. Hope et al. (2017) results as reported						
	Pilot × *During*		0.048	Not reported	**0.0476**	**0.0465**
			(1.09)		**(1.99)****	**(1.98)****
	Pilot × *Post*		−0.014		0.0192	0.0146
			(−0.36)		(0.76)	(0.61)
	Pilot (control) firms		538 (1,072)		538 (1,072)	538 (1,072)
	Adjusted R^2				0.915	0.920

Notes. (Panel A) Regressions with firm and fiscal year fixed effects of *ln*(audit fees) on *Pilot* × *During*, *Pilot* × *Post*, constant term, and indicated covariates. Sample is our 2005 analysis sample (unbalanced panel) merged with audit fee data from Audit Analytics. During and Post periods are defined in the text. (Panel B) Regressions similar to panel A but using Hope et al. (2017) best match specification. (Panel C) Hope et al. (2017) univariate results from their table 4 (for column (1)) and multivariate results from their table 5 (for columns (3) and (4)). All panels column (2) controls for *ln*(total assets). Column (3) uses the "limited" covariates specified in Hope et al. (2017) table 5, model (1) (ln(sales), leverage, book-to-market ratio, and ROA). Column (4) uses full covariates from Hope et al. (2017) table 5, model (2). All continuous variables are winsorized at 1%/99%. Coefficients on covariates and constant term are suppressed. *t*-statistics, using standard errors clustered on firm in parentheses. Significant results at 5% level or better in boldface.

*, **, *** indicates significance at the 10%, 5%, and 1% level, respectively.

use HHZ's short list of covariates and in column (4) their full list. With our specification, all coefficients are small and slightly negative (opposite from predicted).[38] With the HHZ best match specification, the coefficients are positive but far lower than those reported by HHZ and insignificant.

As to economic significance, the coefficient estimates vary substantially across specifications from −0.0132 (opposite sign from predicted) to +0.0476. Moreover, even using the highest point estimate, a 0.0476 increase in *ln*(audit fees), economic magnitude is small. Relative to mean preexperiment fees of \$820,000, this implies an ($e^{0.0476}$-1) × \$820,000 = \$40,000 increase in mean audit fees.

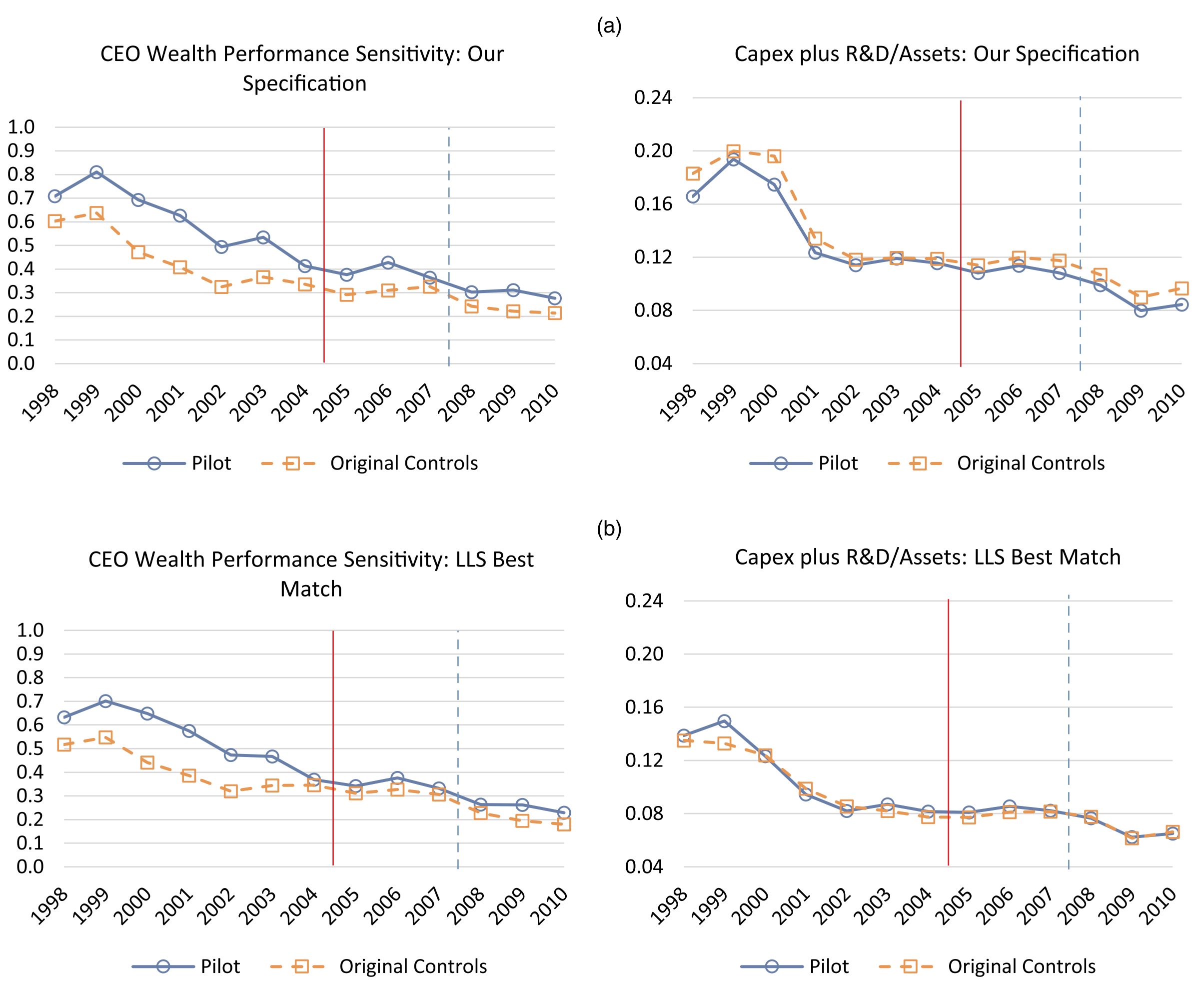
In sum, neither graphical nor regression evidence supports the HHZ hypothesis of higher audit fees for pilot firms.

4.5. LLS: Sensitivity of Investment and CEO Wealth to Share Price

We present results for CEO WPS and (Capex + R&D)/Assets. We begin with graphical evidence and show annual means for pilot and control firms in Figure 6, using both our specification (top graphs) and the LLS best match specification (bottom graphs). For WPS, with both specifications, WPS is higher for pilot firms in the preexperiment period, but the gap shrinks by 2004 and is roughly constant over 2004–2007. Thus, there is no evidence of a treatment effect. However, the nonparallel pretreatment trends drive a spurious negative coefficient on *Pilot* × *During* in a DiD regression, especially one that drops 2004 as LLS do.[39] For (Capex + R&D)/Assets, both graphs show a small relative decline for pilot firms (opposite from predicted) over 2005–2007 and, thus, no evidence of a treatment effect.

We turn to regression results in Table 7. We present results for our specification in panel A, results with the LLS best-match specification in panel B, and the LLS reported results in panel C. Consider WPS, for which LLS predict a drop for pilot firms during the experiment period. We find an insignificant drop with our specification and a barely significant coefficient of −0.129 (t = 2.05) with the LLS best match specification. However, the

Figure 6. (Color online) Lin et al. (2019) Annual Means for Outcomes



Notes. Figures show annual univariate means separately for pilot and control firms for indicated Lin et al. (2019) outcomes for our specification (panel (a)) and the Lin et al. (2019) best match specification (panel (b)). Solid and dashed vertical lines separate pre, during, and post periods.

Table 7. Lin et al. (2019) Investment to Price and CEO Wealth to Performance Sensitivity

Dependent variable	Predicted sign	(Capex + R&D)/Assets (1)	Predicted sign	CEO wealth-performance sensitivity (2)
		Panel A. Our specification		
Q × Pilot × During	Positive	−0.0061*		
		(0.0033)		
Pilot × During		0.0097	Negative	−0.0748
		(0.0061)		(0.0514)
Q × Pilot × Post	Negative (implied)	0.0002		
		(0.0046)		
Pilot × Post	Near 0 (implied)	−0.0016	Near zero (implied)	−0.0768
		(0.0074)		(0.0593)
Sign reversal (Post – During)	Negative (implied)	0.0063	Positive (implied)	−0.0020
		(0.0043)		(0.0332)
Q, double interactions		Yes		No
Firm-year observations		17,302		12,382
Pilot (control) firms		694 (1,394)		517 (999)
R^2		0.746		0.697
		Panel B. Lin et al. (2019) best match specification		
Q × Pilot × During	Positive	0.0022		
		(0.0029)		
Pilot × During		−0.0021	Negative	**−0.1292****
		(0.0063)		**(0.0627)**
Firm-year observations		8,554		9,582
Pilot (control) firms		637 (1,246)		672 (1,310)
R^2 (within)		0.772		0.725
		Panel C. Lin et al. (2019) results as reported		
Q × Pilot × During	Positive	**0.004****		
		(0.002)		
Pilot × During		−0.006	Negative	**−0.174****
		(0.005)		**(0.077)**
Firm-year observations		8,267		9,400
R^2		0.845		0.709

Notes. Regressions, with firm and fiscal year fixed effects of indicated dependent variables (defined in Internet Table IA-4) on indicated variables and constant term. (Panel A) Uses our specification. All outcomes are winsorized at 1%/99%. During and Post periods are defined in the text. (Panel B) Uses Lin et al. (2019) best match specification, including winsorizing CEO wealth-performance sensitivity at 1%/99% but not winsorizing (Capex + R&D)/assets). (Panel C) Lin et al. (2019) results as reported. They report s.e.'s; we infer statistical significance from the ratio of coefficient to standard error. All panels. Coefficients of principal interest are the triple interaction terms for Capex + R&D, and *Pilot* × *During* and *Pilot* × *Post* for wealth-performance sensitivity. Coefficients on Q, double interactions (Q × During, Q × Post, Q × Pilot), and constant term are suppressed. Standard errors clustered on firm in parentheses. Significant results at 5% level or better in boldface.

*, **, *** indicates significance at the 10%, 5%, and 1% level, respectively.

negative coefficients are driven by nonparallel pretreatment trends. For (capex + R&D)/assets, LLS predict a positive sign on the triple interaction *Pilot* × *During* × Tobin's *q*. We instead find a negative coefficient (marginally significant) with our specification, which becomes positive but insignificant with the LLS best match specification. In sum, considering both graphical and the regression evidence and both specifications, there is no support for the LLS conjectures.

4.6. Technical Replication of FHK and HHZ

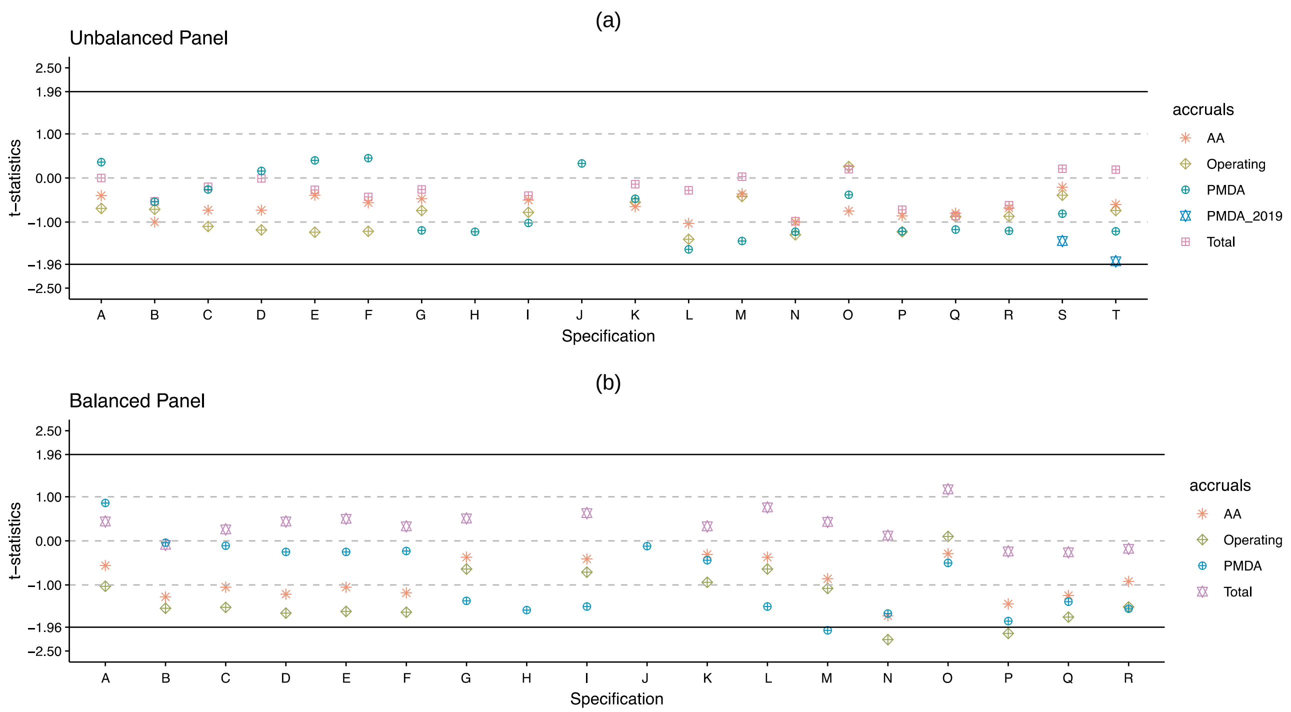
Our objective in this paper is to assess the robustness of the indirect effects attributed to the Regulation SHO experiment in the four studies that we reexamine, using a sample that closely follows the SEC's exclusion rules and a range of reasonable specifications. This reassessment was motivated by the lack of evidence for a causal channel to support the indirect effects reported in these studies. Our objective was not technical replication of these studies.

However, in response to an earlier version of this paper, FHK posted their sample and code for their PMDA result, and HHZ provided their data and code to us for their audit fees result. In the internet appendix, we confirm technical replication with their exact samples and specifications but find that both the FHK and HHZ results are highly sensitive to minor departures from their exact specifications, including, for FHK, using simpler measures of accruals.[40]

4.7. Additional Evidence on Robustness Across Specifications

With our specifications, across the outcomes reported in Tables 4–7, all coefficients on *Pilot* × *During* are statistically insignificant, and several have the opposite sign

Figure 7. (Color online) Fang et al. (2016) Accruals Measures and Statistical Significance Across Specifications



Notes. Panel (a) shows t-statistics, using standard errors clustered on firm for coefficients on *Pilot* × *During* for indicated accruals measures (AA, operating accruals, PMDA, and total accruals) for unbalanced panel. Specifications are as follows (see tables listed below for additional details): A. Our specification (Internet Table IA-9, panel A). B–F (from Internet Table IA-11). B. A but use calendar year periods. C. B but remove 2004 from preperiod. D. C but use Fang et al. (2016) sample. E. D but remove firm FE. F. Fang et al. (2016) best match specification. G. Fang et al. (2016) PMDA specification (Internet Table IA-12). H–M (from Internet Table IA-13). H. G but correct Fang, Huang, and Karpoff matching error. I. G but also correct Fang et al. (2016) duplicates error. J. G but use current year ROA to find matching firm. K. H except winsorize operating accruals at 1%/99% within fiscal year instead of excluding outliers. L. I but use firm and fiscal year fixed effects. M. G except include 2004 in preperiod. N–R use Fang et al. (2016) operating accruals specification (Internet Table IA-14). N. Fang et al. (2016) operating accruals specification. O. N but winsorize operating accruals at 1%/99% within fiscal year instead of excluding outliers. P. N but winsorize in second stage regressions. Q. N but impute covariate values in determining sample. R. N but both winsorize in second stage regressions and impute covariate values. S. Specification from Fang et al. (2019), table 15, but corrects their postperiod error (Internet Table IA-15). T. M but add firm FE. Panel B. Similar to A except using balanced panel. Omits S and T, which apply only to unbalanced panel. Both panels. Solid horizontal lines show 5% significance (t-statistics= ±1.96).

from the predicted signs. With the best match specifications, two coefficients on *Pilot* × *During* are statistically significant with the predicted sign (one for GMW, the other for LLS), but both have t-statistics barely above two, and neither is convincing when viewed together with the graphic results.

In Figures 7 and 8, we further assess sensitivity to specification for FHK and HHZ in the spirit of Simonsohn et al. (2020), who suggest that authors should report results across a range of reasonable but meaningfully different specifications. Figure 7 shows variations for FHK. It reports t-statistics for four accruals measures for both unbalanced and balanced panels across the specifications reported in the text and the additional specifications in the internet appendix. Solid horizontal lines show 5% significance ($t = \pm 1.96$). Across 142 regressions, only four coefficients are significant (two for PMDA, two for operating accruals) and only mildly so (t = 1.96, 2.03, 2.10, and 2.24). Twenty-three coefficients are positive (opposite sign from predicted). In Figure 8, we summarize the t-statistics for the HHZ outcome across the specifications and covariate choices in the text and the internet appendix. Only 4 of 78 coefficients are significant and barely so (t = 1.98, 1.99, 2.00, and 2.01). These figures provide evidence that the results in both papers lack robustness across a range of samples and reasonable specifications.

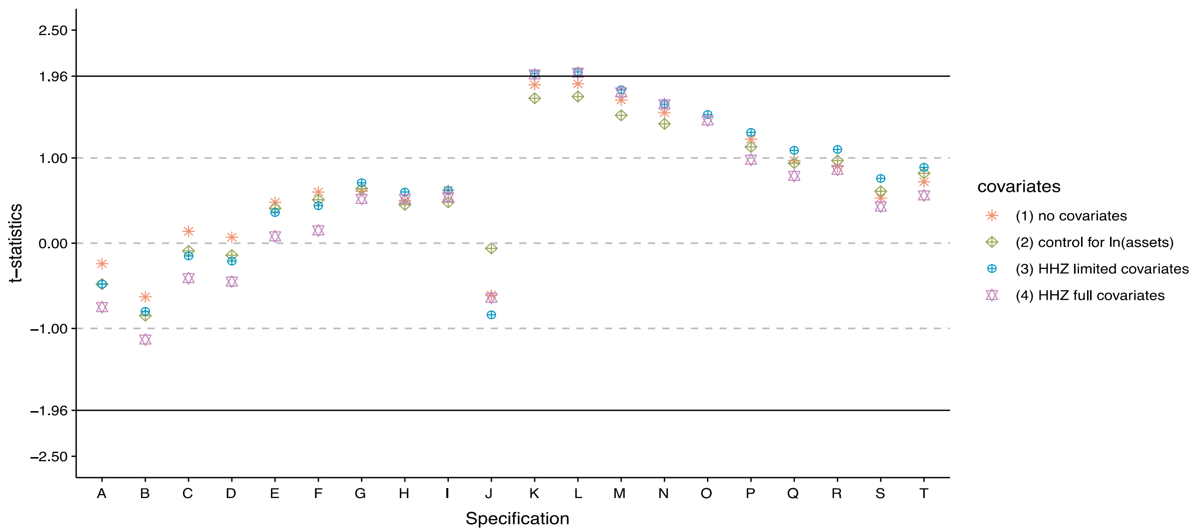
4.8. Summary

Overall, across a range of specifications, including the many additional results reported in the internet appendix, there is no evidence for a treatment effect for FHK, HHZ, or LLS and minimal evidence for GMW. Thus, there is no reliable evidence that the SEC's short-sale experiment, which had minor direct effects on trading markets, had statistically significant or economically important indirect effects on firms.

5. Discussion and Conclusions

During 2005–2007, the SEC conducted a randomized trial in which it suspended short-sale price tests for

Figure 8. (Color online) (Hope et al. (2017) Statistical Significance Across Specifications



Notes. Figure shows *t*-statistics with standard errors clustered on firm for the coefficients on *Pilot* × *During* for regressions of *ln*(audit fees) for the specifications in Tables 6, IA-19, IA-20, and IA-21. Specifications are as follows (see these tables for additional details). Regressions use unbalanced panel except as indicated: A. Our specification from Table 6, panel A. B. Our specification but for balanced panel, from Internet Table IA-20. C–H are from Internet Table IA-19, panels B–G. C. Same as A but remove fiscal 2004 from preperiod. D. Same as E but extend post period to include 2011–2013. E. Same as D but switch to Hope et al. (2017) definition of pre, during, and post periods. F. Same as E except replace fiscal year FE with during and post dummies. G. Same as F but require firms to be in R3000 in June 2005. H. Same as G but require data on Hope et al. (2017) full covariates for all models. I. Hope et al. (2017) best match specification from Table 6, panel B. J. Best match specification but for balanced panel from Internet Table IA-20. K–T are from Internet Table IA-21, panels A–J. K. Hope et al. (2017) exact sample and specification. L. Corrects Hope et al. (2017) data error for audit fees for four firms in 2000. M. Same as K but map Audit Analytics fiscal years correctly to Compustat fiscal years. N. Same as M but add fiscal year fixed effects. O. Same as M but use covariate values from Compustat. P. Start from O but obtain additional observations from Audit Analytics and Compustat. Q. Start from P but use historical CIK values to match to Audit Analytics. R. Similar to Q but start with Hope et al. (2017) list of 1,899 firms before their matching to Audit Analytics and Compustat. S. Start from R but remove Hope et al. (2017) requirement that firms be included in Russell 3000 index in June 2005. T. Start from S but exclude firms that ceased trading before experiment start. In all panels, solid horizontal lines show 5% significance (*t*-statistic = ±1.96).

approximately 1,000 pilot firms. Initial studies of the experiment found little or no impact of removing these restrictions on short interest or pilot firm returns. Based on these studies, the SEC removed short-sale price tests for all firms when the experiment ended in 2007. Since then, many papers have documented a wide range of indirect effects of the experiment on pilot firms but generally without providing evidence of a causal channel or an underlying economic mechanism to support these indirect effects.

In this study, we examine the most commonly asserted causal channels: impact on short interest, impact on returns, and managerial fear. We do so using a sample that closely matches the actual experiment and uses a longer time period than prior studies. We also provide evidence on a causal channel that is relied on by several indirect effect studies but not closely examined: managerial fear. Our analysis fails to find support for any of these causal channels. We rely on prior studies to rule out a fourth posited channel: running through price efficiency.

The lack of a natural causal channel raises the likelihood that the reported indirect effects may be false positives. We thus reexamine the principal findings in four recent papers: FHK, GMW, HHZ, and LLS. We find minimal support for the reported results in these papers with either our specification or best match specifications.

Our findings are important for a number of reasons. First, a number of indirect effects studies directly or indirectly draw policy implications on the impact of short-sale restrictions without documenting a credible causal channel. The lack of evidence to support the principal causal channels plus the nonrobustness of the results in the four studies that we reexamine suggest that regulators should infer policy implications from these and other indirect effects studies with caution.

Second, a significant body of research documents the positive role short selling plays in the capital markets by making prices more informative and markets more efficient. In contrast, the findings from the indirect effect studies imply that removing even modest restrictions on short selling can have major negative effects on firm behavior. Our results provide a counterweight to these results by casting doubt on the reliability of the reported indirect effects.

Third, we highlight the importance of close investigation of the evidence for a causal channel as a basis for reliable inference. Evidence on indirect effects that is not supported by evidence for a causal channel may not be robust, as we find for the four reexamined papers. Finally, even when researchers begin with a randomized trial, they must make many specification choices. Our study provides evidence that these choices matter and one specification choice may produce significant results when other reasonable choices would not.

Acknowledgments

The authors thank David Sraer (department editor), an anonymous associate editor and two anonymous referees for their detailed comments and suggestions. The authors thank Parth Lalkiya and Lauren Fiotakis for research assistance. The authors thank Gauri Bhat, Ian Gow, Doug Hanna, Ankit Jain (discussant), Pab Jotikasthira, Sanjay Kallapur, Umang Khetan (discussant), Srini Krishnamurthy, Charles Lee, Dan Millimet, Shiva Rajgopal, Mehrdad Samadi, Holger Spamann, Shyam Sundar, Jake Thomas, Sorabh Tomar, and Ram Venkataraman as well as workshop participants at American Law and Economics Association (2021), Bar Ilan University, Carnegie Mellon University, Columbia Business School, Conference on Empirical Legal Studies (2022), Corporate Law Academic Webinar Series (2020), Financial Management Association (2021), Florida State University, Haifa University, Hebrew University of Jerusalem, Indian School of Business, Midwest Finance Association, Pace University, Southern Methodist University, Stanford Law School, Tel Aviv University, University of Central Florida, and University of Texas at Arlington as well as participants at American Law and Economics Association, Conference on Empirical Legal Studies, Midwest Finance Association, and Financial Management Association Annual Meetings for comments and discussions on an earlier draft.

Endnotes

[1] We follow Diether et al. (2009) and generally refer to both rules as "price tests". When we want to distinguish between them, we call them the NYSE uptick rule and the Nasdaq bid test. The American Stock Exchange (AMEX) rule is similar to the NYSE rule. We generally refer to both NYSE and AMEX firms as NYSE firms.

[2] Macey et al (1989) argue that execution delays are more important for index arbitrageurs, who typically seek to short many firms at once. For example, if futures are trading at discount to the cash price, an arbitrageur would buy the futures and seek to short the underlying shares, but the price tests could delay execution of this trade for some of the firms in the index.

[3] Furthermore, the Nasdaq bid test applies to Nasdaq National Market stocks but not Nasdaq SmallCap stocks. Christophe et al. (2004) find no difference in shorting activity during down markets for Nasdaq National Market versus Nasdaq SmallCap stocks.

[4] Diether et al. (2009) and Alexander and Peterson (2008) find smaller trade size and higher short sale volume for NYSE firms. For Nasdaq firms, suspension of the bid test had limited impact on pilot firms, consistent with the bid test being less restrictive than the NYSE uptick rule.

[5] During the experiment, in addition to fully suspending price tests for pilot firms, the SEC also suspended the NYSE uptick rule for control firms in the Russell 1000 (R1000) for after-hours trading (from 4:15 p.m. until the opening on the next trading day). The SEC also suspended the NYSE uptick rule for all firms at times when the consolidated transaction reporting system was off (generally from 8:00 p.m. until 4:00 a.m. the next day). The Nasdaq bid test always only applied during limited hours (generally 9:30 a.m. to 4:00 p.m.). These factors reduced the differences in the rules that applied to pilot versus control firms.

[6] We discuss details of our sample selection and other specification choices in our prespecified analysis plan (Black et al. 2019), available on SSRN at https://ssrn.com/abstract=3415529.

[7] Our concerns with robustness involve papers reporting indirect effects of the SEC experiment on the behavior of managers, firms, and third parties such as auditors. They do not apply to market microstructure papers, which study direct effects of the experiment on trading markets.

[8] We have posted the Stata code and data sets needed to generate the results in this paper and the SAS code we used to generate starting data sets on Professor Black's website.

[9] The experiment was announced on July 28, 2004, so July 2004 mid-month short interest precedes the announcement. All trading days in May 2005 are in the experiment period. The experiment ended on July 5, 2007. Regression results for the experiment period are not sensitive to whether we treat July 2007 as part of the experiment or the post-experiment period or study month-end instead of mid-month short interest.

[10] In Internet Figure IA-2, we report leads-and-lags graphs showing monthly estimated treatment effects during the sample period. Across the four samples, there are no apparent trends in either the prelaunch or experiment periods, and all monthly coefficients are statistically insignificant.

[11] In Internet Figure IA-5, we report weekly BHRRs over 2003–2005 for the 2004 announcement sample and small firms within this sample (based on market capitalization). A few weekly relative returns are significant at the 5% level but no more than would be expected from chance alone.

[12] GMW report BHRR but report statistical significance only for a regression-based measure of relative returns. We consider this measure in Internet Table IA-6; results are consistent with Table 3.

[13] There were important details in the SEC's selection of pilot firms that were not known prior to the announcement, so no one could have known which specific firms would be designated as pilot firms. The SEC often holds material market-sensitive information, including information about investigations of particular firms. We are unaware of instances of leakage or theft of this information. Billett et al. (2020) note that SEC staff held meetings with market participants during June and July 2004 to discuss the experiment logistics. We can think of no reason why, at those meetings, the SEC staff would have provided nonpublic information about which firms would be pilot firms.

[14] FHK (at p.1255) assert that "The decision to eliminate all short-sale price tests prompted a huge backlash from managers and politicians." However, as support, FHK cite stories from the financial crisis period in 2008, long after the experiment. GMW (at p.1739) assert that "In public comments, NYSE officials, specialists, and member firms all expressed support for short-sale restrictions." However, as support, they cite only comment letters from the NYSE and its specialist association. As we discuss, member firms supported the experiment. Most other studies that rely on a manager fear channel either simply assert the potential for manager fear or cite FHK or GMW.

[15] The first FASB proposal to require expensing of employee stock options, in 1993, produced more than 700 comment letters opposing the requirement and significant media coverage. The companies lobbied Congress, and SEC Chair Arthur Levitt urged FASB to ease off. See, for example, Bodie et al. (2003) and Gleason and Glendening (2019).

[16] See "Floyd Norris, 70 Years Later, A Scapegoat Gets a Break," *The New York Times* (December 8, 2006). The NYT also published an op-ed article in October 2006, supporting repeal of the short-sale

rule: see Richard Sauer, Bring on the Bears," *The New York Times* (October 6, 2006).

[17] See Spencer Jakab, "Short-Sellers May Owe ETFs Some Thanks – Dropping of 'Uptick' Rule by SEC Comes as Growth of Stock Baskets Is Soaring," *The Wall Street Journal* (June 15, 2007).

[18] See SEC release 34-55, 970 (July 3, 2007).

[19] See SEC release 34-48709, 68 Federal Register 62972 (November 6, 2003). The comment letters are available at https://www.sec.gov/rules/proposed/s72303.shtml.

[20] Of 43 comments from individuals, 36 supported the experiment.

[21] See SEC release 34-50103 (July 28, 2004), 69 Federal Register (August 6, 2004), at 48008 and 48012.

[22] We include in organizational letters a comment from Prof. James Angel of Georgetown University, an expert on securities markets. We count a joint submission by Citigroup, Goldman Sachs, Merrill Lynch, and Morgan Stanley as four comments and a joint submission by JPMorgan Securities and UBS Securities as two comments.

[23] See SEC release 34-54891, 71 Federal Register 75068 (December 13, 2006). The comment letters are available at https://www.sec.gov/comments/s7-21-06/s72106.shtml.

[24] The SEC also received eight comments from individuals; of these, four supported and four opposed repeal.

[25] For FHK, see Internet Figure IA-6, panel D3; for GMW, see Figure 4; for HHZ (who rely on auditor fear), see Figure 5; for LLS (who rely on a price efficiency channel), see Figure 6.

[26] See the internet appendix for details on the steps to generate the 2005 analysis sample. FHK, GMW, and HHZ also exclude financial and utility firms; LLS do not.

[27] For example, FHK study accruals but control for ROA, which is directly affected by accruals. GMW study investment but control for ln(assets), which is directly affected by investment.

[28] Because of space constraints, we present results for FHK, GMW, and LLS with covariates in the internet appendix. We present results for HHZ (who study a single outcome) in the text both with and without covariates.

[29] If there were differential attrition, using an unbalanced sample might ameliorate the potential bias. There would still be no reason to prefer a balanced sample.

[30] For some research questions, excluding 2004 is not appropriate. For example, LLS conjecture that prices become more informative after the price tests are eliminated. For this causal channel, 2004 should be included in the preperiod.

[31] See SEC release 34-55970 (July 3, 2007) (ending the short-sale restrictions effective July 6, 2007).

[32] Following Hribar and Nichols (2007), we winsorize the variables used to estimate AA and PMDA at 1%/99% within each fiscal year. We winsorize again, across all sample years, in estimating Equations (5) and (6).

[33] We note, however, that using a time-varying match is not advisable in a "causal" project. Because accruals affect ROA, one should match, if at all, on pretreatment values of ROA, determined in 2004.

[34] We thank an anonymous referee for suggesting the discussion of economic significance.

[35] For the FHK best match specification, standard errors clustered on firm (not reported) are very close to the randomization inference s.d.'s reported in Table 4.

[36] GMW report univariate results without firm or year FE and regression results with firm and fiscal year FE and covariates. They do not report regression results with firm and fiscal year FE but without covariates, which would be more directly comparable to our results. GMW also present triple-difference results in which the third difference is between large and small firms. This specification does not make sense to us. The DiD and triple-difference specifications produce identical results in a model with interactions between small firm dummy and the year dummies, the constant term, and the covariates.

[37] GMW are aware of this issue and compare the reduction in capex/assets to a long-term point estimate of −9% relative returns for small pilot firms over two years (GMW at p.1740). However, these long-term relative returns are not statistically significant and cannot be causally attributed to the experiment.

[38] In Internet Table IA-20, we report results for HHZ with a balanced panel as a robustness check. All regression coefficients are negative (opposite from predicted).

[39] For the LLS sample period, the DiD coefficient compares the pilot-minus-control average for 2005–2007 to the pilot-minus-control average for 2002–2003. It is apparent from Figure 6 that this coefficient will be negative even though there is no meaningful change in the pilot-minus-control difference over 2004–2007. Note that LLS rely on a price efficiency channel, which depends on increased short selling for pilot firms. That can only start after the price tests were suspended in May 2005. Thus, for the LLS conjecture, 2004 should be included in the preperiod.

[40] For example, the statistical significance of the FHK result disappears with an unbalanced panel or using s.e.'s clustered on firm or randomization inference. Moreover, the FHK coefficient is close to zero if one chooses a matching firm based on concurrent ROA as recommended by the PMDA developers (Kothari et al. 2005) rather than on lagged ROA. For HHZ, the coefficient on *Pilot* × *During* shrinks substantially and loses statistical significance if we use their list of firms but add observations that they missed because of incomplete matching to CRSP, Compustat, and Audit Analytics.

References

Aitken MJ, Frino A, McCorry MS, Swan PL (1998) Short sales are almost instantaneously bad news: Evidence from the Australian Stock Exchange. *J. Finance* 53(6):2205–2223.

Alberto A, Athey S, Imbens GW, Wooldridge J (2023) When should you adjust standard errors for clustering? *Quart. J. Econom.* 138(1):1–35.

Alexander GJ, Peterson MA (1999) Short selling on the New York Stock Exchange and the effects of the uptick rule. *J. Financial Intermediation* 8(1–2):90–116.

Alexander GJ, Peterson MA (2002) Implications of a reduction in tick size on short-sell order execution. *J. Financial Intermediation* 11(1):37–60.

Alexander GJ, Peterson MA (2008) The effect of price tests on trader behavior and market quality: An analysis of regulation SHO. *J. Financial Markets* 11(1):84–111.

Asquith P, Pathak P, Ritter J (2005) Short interest, institutional ownership and stock returns. *J. Financial Econom.* 78(2):243–276.

Bai L (2008) The uptick rule of short sale regulation: Can it alleviate downward price pressure from negative earnings shocks. *Rutgers Bus. Law J.* 5(1):1–63.

Barclay MJ (1989) Comment on Macey, Mitchell & Netter. *Cornell Law Rev.* 74(5):836–840.

Battalio R, Schultz P (2006) Options and the bubbler. *J. Finance* 61(5):2071–2102.

Billett MT, Liu F, Tian X (2020) Information spillovers and cross monitoring between the stock market and loan market. Kelley School of Business Research Paper No. 16–20.

Black B, Desai H, Litvak K, Yoo W, Yu JJ (2019) Pre-analysis plan for the reg SHO reanalysis project. Preprint, submitted August 5, https://dx.doi.org/10.2139/ssrn.3415529.

Bodie Z, Kaplan RS, Merton RC (2003) For the last time: Stock options are an expense. *Harvard Bus. Rev.* 81(3):62–71.

Cameron AC, Gelbach JB, Miller DL (2008) Bootstrap-based improvements for inference with clustered errors. *Rev. Econom. Statist.* 90(3):414–427.

Christophe SE, Ferri MG, Angel JJ (2004) Short-selling prior to earnings announcement. *J. Finance* 59(4):1845–1876.

De Angelis D, Grullon G, Michenaud S (2017) The effects of short-selling threats on incentive contracts: Evidence from an experiment. *Rev. Financial Stud.* 30(5):1627–1659.

Dechow PM, Ge W, Schrand C (2010) Understanding earnings quality: A review of the proxies, their determinants and their consequences. *J. Accounting Econom.* 50(2–3):344–401.

Dechow PM, Sloan R, Sweeney A (1995) Detecting earnings management. *Accounting Rev.* 70(2):193–225.

Dechow PM, Ge W, Larson CR, Sloan RG (2011) Predicting material accounting misstatements. *Contemporary Accounting Res.* 28(1): 17–82.

Desai H, Ramesh K, Thiagarajan R, Balachandran BV (2002) An investigation of the informational role of short interest in the Nasdaq market. *J. Finance* 57(5):2263–2287.

Diamond D, Verrecchia RE (1987) Constraints on short selling and asset price adjustment to private information. *J. Financial Econom.* 18(2):277–311.

Diether K, Lee K-H, Werner I (2009) It's SHO time! Short-sale price-tests and market quality. *J. Finance* 64(1):37–73.

Fang VW, Huang A, Karpoff J (2016) Short selling and earnings management: A controlled experiment. *J. Finance* 71(3):1251–1293.

Fang VW, Huang A, Karpoff J (2019) Reply to "The Reg SHO Reanalysis Project: Reconsidering Fang, Huang and Karpoff (2016) on Reg SHO and Earnings Management" by Black et al. Preprint, submitted January 17, 2020, https://dx.doi.org/10.2139/ssrn.3507033.

Gleason CA, Glendening M (2019) Lobbying and opposition to SFAS No. 123(R): An examination of campaign contributions from CEOs and PACs. *Accounting Horizons* 33(1):103–124.

Grullon G, Michenaud S, Weston J (2015) The real effects of short-selling constraints. *Rev. Financial Stud.* 28(6):1737–1767.

Harvey CR (2017) Presidential address: The scientific outlook in financial economics. *J. Finance* 72(4):1399–1440.

Heath D, Ringgenberg MC, Samadi M, Werner IM (2023) Reusing natural experiments. *J. Finance* 78(4):2329–2364.

Hope O-K, Hu D, Zhao W (2017) Third-party consequences of short-selling threats: The case of auditor behavior. *J. Accounting Econom.* 63(2–3):479–498.

Hribar P, Nichols DC (2007) The use of unsigned earnings quality measures in tests of earnings management. *J. Accounting Res.* 45(5):1017–1053.

Kothari SP, Leone AJ, Wasley CE (2005) Performance matched discretionary accrual measures. *J. Accounting Econom.* 39(1):163–197.

Leamer EE (1985) Sensitivity analyses would help. *Amer. Econom. Rev.* 75(3):308–313.

Lin T-C, Liu Q, Sun B (2019) Contractual managerial incentives with stock price feedback. *Amer. Econom. Rev.* 109(7):2446–2468.

Macey RJ, Mitchell M, Netter J (1989) Restrictions on short sales: An analysis of the uptick rule and its role in view of the October 1987 stock market crash. *Cornell Law Rev.* 74(5):799–835.

Miller EM (1977) Risk, uncertainty and divergence of opinion. *J. Finance* 32(4):1151–1168.

Securities and Exchange Commission (2004) Short Sales. 17 C.F.R. Parts 240, 241 and 242, Release No. 34–50103; File No. S7–23–03, https://www.govinfo.gov/content/pkg/FR-2004-08-06/pdf/04-17571.pdf.

Securities and Exchange Commission (2007) Economic analysis of the short sale price restrictions under the regulation SHO pilot. Office of Economic Analysis, Washington, DC. Preprint, Submitted February 6, https://www.sec.gov/news/studies/2007/regshopilot020607.pdf.

Simonsohn U, Simmons JP, Nelson LD (2020) Specification curve analysis. *Nature Human Behav.* 4:1208–1214.

about ECGI

The European Corporate Governance Institute has been established to improve *corporate governance through fostering independent scientific research and related activities.*

The ECGI will produce and disseminate high quality research while remaining close to the concerns and interests of corporate, financial and public policy makers. It will draw on the expertise of scholars from numerous countries and bring together a critical mass of expertise and interest to bear on this important subject.

The views expressed in this working paper are those of the authors, not those of the ECGI or its members.

Electronic Access to the Working Paper Series

The full set of ECGI working papers can be accessed through the Institute's Web-site (www.ecgi.global/content/working-papers) or SSRN:

Finance Paper Series	http://www.ssrn.com/link/ECGI-Fin.html
Law Paper Series	http://www.ssrn.com/link/ECGI-Law.html