Past is Prologue: Inference from the Cross Section of Returns Around an Event *

Jonathan B. Cohn
University of Texas at Austin

Travis L. Johnson
University of Texas at Austin

Zack Liu
University of Houston

Malcolm I. Wardlaw
University of Georgia

June 2024

Abstract

The possibility of contemporaneous events hampers inference about differential effects of a quasi-experimental event across firms. We show that this possibility consistently causes high false positive rates in tests of differences in short-term event-driven stock returns – nearly 50% at the 1% significance level in some cases. Clustering standard errors (e.g., by industry) is an inadequate solution. Researchers should instead use the distribution of pre-event return relationships to test for significance. We introduce a novel GLS-based variant of this testing strategy, show that it increases power substantially over OLS-based variants, and provide a Stata module that implements both.

---

*Jonathan Cohn: jonathan.cohn@mccombs.utexas.edu, (512) 232-6827. Travis Johnson: travis.johnson@mccombs.utexas.edu, (512) 232-6824. Zack Liu: zliu@bauer.uh.edu, (713) 743-4764. Malcolm Wardlaw: malcolm.wardlaw@uga.edu, (706) 204-9295.

We thank Paul Irvine (Lone Star Discussant), Minjoo Kim, Sheridan Titman, Steven Urry, Zirui Wang, James Weston (FRA Discussant) and seminar participants at Auburn University, Chinese University of Hong Kong, Clemson University, Financial Research Association (FRA) Conference, Hong Kong University, Hong Kong Science and Technology University, Lone Star Conference, Nanyang Technological University, National University of Singapore, University of Houston, University of Massachusetts at Amherst, University of Oklahoma, and University of Texas at Austin for valuable feedback. We thank Christopher McDonald and George Rao for excellent research assistance.

A common identification strategy in corporate finance examines differences in the effects of a quasi-experimental shock (e.g., a surprise election outcome) across firms differing on some dimension of predicted exposure (e.g., connectedness to the winning party). One threat to causal inference in such analysis, even if the shock is truly exogenous, is the possibility of other, contemporaneous events whose effects on firms differ along similar dimensions. These confounding shocks can cause false positives, erroneously suggesting that the focal event differentially affects firms in a hypothesized manner when, in reality, it does not. This study investigates the practical importance of this problem in a specific context – testing differences in the valuation effects of an event as reflected in short-term stock returns. In addition to allowing us to shed new light on a widely-used methodology in corporate finance,[1] this context affords a rare opportunity to assess the counterfactual distribution of outcome differences using many samples where the focal event is absent.

To fix ideas, consider three events from 2019 – a tentative agreement on Phase 1 of a trade deal between the U.S. and China reached on October 11, a UK House vote against a no-deal Brexit on March 13, and a "Blood Moon" lunar eclipse, which many cultures view as a bad omen, on January 21. The first two events plausibly portended a more favorable environment for investment and the third a less favorable environment. Supporting these hypotheses, the relationships between a firm's event-day return and its investment level are positive for the first two events and negative for the third, all statistically significant at the 5% level based on standard t tests. This evidence might be convincing but for the fact that the relationship between investment and one-day returns is significant at the 5% level on 49% of *all* trading days in 2019, and larger in magnitude than on any of these three days more

[1] See Leftwich (1981) for an early example. Recent examples in top-three finance journals include Acemoglu, Hassan, and Tahoun (2018), Amihud and Stoyanov (2017), Bizjak, Kalpathy, Mihov, and Ren (2022), Brown and Huang (2020), Child, Massoud, Schabus, and Zhou (2021), Cremers, Litov, and Sepe (2017), Gilje and Taillard (2017), Licht, Poliquin, Siegel, and Li (2018), Liu, Shu, and Wei (2017), Schoenherr (2019), and Zeume (2017), which collectively have more than 2,000 google scholar cites at the time of this writing. See Eisfeldt, Schubert, and Zhang (2023), Jiang, Parsons, Sun, and Titman (2023), and Lu and Ye (2023) for current working paper examples.

than 40% of the time (Figure 1). The prevalence of these relationships muddies inference by making it difficult to attribute an event-day relationship to the event in question.

[Figure 1 about here]

With this motivation in mind, we systematically evaluate relationships between short-term (one- and five-day) returns and various firm characteristics over a long sample period. We present evidence that the threat of false positives when studying differences in returns around a single quasi-experimental shock is generally severe. Specifically, we show that coefficients from regressions of short-term returns on firm characteristics, including characteristics analyzed in published papers using this methodology, are typically statistically significant at the 1% level more than 10% of the time, up to nearly 50% of the time for the characteristics size, market-to-book ratio, and cash holdings. Treating the distribution of relationships over time as the counterfactual distribution in a defined event period, one would need to double or even triple conventional standard errors in most cases to achieve the correct rate of false positives if the focal event has no effect (the null hypothesis). Clustering standard errors at a group level such as industry reduces rates of statistical significance, but these rates remain as high as 34% at the 1% significance level.

Further consideration of Figure 1 suggests a potential solution to the problem: assess statistical significance by comparing the event-period relationship to the distribution of the relationship in non-event periods. We should only attribute a relationship in the event period to the event itself if the relationship in that period is in the tails of the distribution of the relationship in proximate periods. Sefcik and Thompson (1986) describes a portfolio-based version of this strategy. We describe a more intuitive and flexible variation: regress returns on a characteristic in the event period and in a series of pre-event periods, and then test whether the event-period coefficient is in the tails of the coefficient distribution.[2] Despite

[2] In principle, one could use post-event days as well. In practice, it is more conventional to use pre-event information as a benchmark to avoid any risk of look-ahead bias.

its intuitive appeal, statistical testing based on the distribution of relationships in pre-event periods remains rare – only 13% of papers analyzing differences in short-term returns around a quasi-experimental event that we identify in top-three finance journals in the past ten years use this method, with the remaining 87% using cross-sectional standard errors from the event period regression itself.[3]

We consider two specific approaches to implementing this testing strategy. The first uses OLS to estimate both the event-period and pre-event period coefficients and is comparable to the approach of Sefcik and Thompson (1986). The second is novel and uses GLS to estimate both instead. In theory, GLS boosts efficiency by weighting observations more for pairs of firms whose returns comove less and therefore contain more independent information. More precise estimates of both event and pre-event period relationships should allow for statistically more powerful tests – an important consideration when the differential effect of an event may be modest relative to the noise in returns.

While the need to estimate a covariance matrix of model errors makes GLS difficult to implement in most corporate finance applications, the daily frequency and approximate serial independence of stock returns make it viable here. Following Giglio and Xiu (2021), our approach uses principal component analysis (PCA) of daily returns over a period prior to a given day to encode information about return comovement driven by common exposure to day-to-day news into an estimated covariance matrix for that day. We use the first 100 principal components (PCs) in our analysis but show that our conclusions are not sensitive to the number of PCs we use.

We assess the performance of both the OLS and GLS approaches. In doing so, we use 199 pre-event periods but show that our conclusions are not sensitive to the number of pre-event periods we use. We show that, for either approach (OLS or GLS), using a t-statistic to test

[3]Of the 87% that use standard errors from the event regression, 35% use unadjusted standard errors; 25% robust standard errors, which address concerns about heteroskedasticity but not cross correlations; 35% standard errors clustered by industry; and 5% standard errors clustered by geographic location.

differences between event-period and pre-event period relationships still overstates statistical significance slightly because of fat tails in the distribution of relationships between short-term returns and most characteristics. Testing significance by computing a p-value for the event-period relationship based on the empirical cumulative density function (CDF) of the relationship corrects the problem. The relationships between event returns and investment level for the three events described in paragraph 2 cease to be significant at even the 10% level when we use either the OLS or GLS approach to conduct statistical testing.

While both the OLS and GLS approaches are effective at addressing concerns about false positives, we show that the GLS approach increases statistical power substantially relative to the OLS approach. When we introduce an artificial relationship between short-term returns and a characteristic in a given period and treat this as an event period, the GLS approach generally detects the relationship nearly twice as often as the OLS approach at the 5% significance level and up to three times as often at the 1% level, across multiple characteristics, event-period lengths, and artificial relationship magnitudes. To encourage adoption of this general testing strategy, we have written and made available a turnkey Stata module that implements both the OLS- and GLS-based approaches.[4]

Finally, we further analyze the PCs that we use in our GLS implementation to better understand correlations in exposure to day-to-day news. Like Lopez-Lira and Roussanov (2023), we find that many PCs are required to summarize the cross-sectional variance in returns and that these PCs are difficult to map to traditional factors related to expected returns such as those in Fama and French (2015). We additionally find that the return factors captured by the first few PCs vary considerably over time and are often period-specific. For example, sensitivity to finance industry returns is important for explaining return correlations in 2008, while sensitivity to meme stock returns is important for explaining these correlations in 2021. The complexity and instability of short-term return correlations explain the limited

[4] `https://github.com/MalcolmWardlaw/csestudy`.

effectiveness of clustering in correcting standard errors. Time-varying correlations may also reduce the statistical power of approaches using pre-event periods for comparison but should not cause excess false positives.

Circling back to the more general threat posed by contemporaneous events, while the low frequency and serial correlation of most outcome variables in corporate finance preclude the type of analysis we perform in this paper, it is reasonable to expect that the problems we document are severe for other outcome variables as well. Our paper then suggests that relying on analysis of differences in outcomes after a single event for identification may be even more problematic than previously appreciated. This problem may affect studies of staggered events as well if the events cluster close together in time. Our paper also suggests an unappreciated advantage of analyzing differences in short-term returns around an event rather than longer-term outcomes: the observability of returns at high frequency makes it possible to address concerns about contemporaneous events.

Our paper contributes to the literature on practical issues in computing standard errors when regression errors are not i.i.d. Several papers show that clustering can substantially alter standard errors (Moulton, 1986, 1987; Bertrand, Duflo, and Mullainathan, 2004). Petersen (2009) shows that, in finance, the Fama and MacBeth (1973) procedure may be preferable to clustering by firm in panel data when cross-sectional clustering is more important than time-series clustering. The approach we describe is close in spirit to the Fama-MacBeth approach. Abadie, Athey, Imbens, and Wooldridge (2023) considers flexible approaches to modeling error structure and shows that clustering can result in *inflated* standard errors and hence low statistical power. In contrast, our evidence suggests that even clustered standard errors can be severely *deflated* when studying differences in outcomes in response to a quasi-experimental event due to common exposure to contemporaneous events.

Our paper also contributes to the literature on event study analysis of returns. The literature has primarily explored the challenges that return correlations create for inference in studies of abnormal *mean* event returns (Collins and Dent, 1984; Bernard, 1987; Lyon, Barber, and Tsai, 1999; Brav, Geczy, and Gompers, 2000; Mitchell and Stafford, 2000; Jegadeesh and Karceski, 2009; Kolari and Pynnönen, 2010). We are not aware of any prior analysis of the practical challenge caused by return correlations for *cross-sectional* analysis of returns around an event – a more common methodology in modern corporate finance.[5] Sefcik and Thompson (1986) describes this challenge conceptually but does not assess its practical importance in real-world applications. The approach that they propose using pre-event relationships to conduct statistical testing has largely been ignored. Our novel GLS variant of this approach offers substantial improvements in statistical power in real-world applications. Finally, we resolve practical challenges in implementing these approaches by providing a turnkey Stata module.

1. Data and Sample

Our analysis involves regressing firm-level stock returns over periods of one and five days on various firm characteristics. Our sample period is 1991–2021, which contains 7,811 trading days. This period is long enough to allow us to estimate the distributions of relationships between returns and characteristics accurately and to explore variation in these distributions over time. Our analysis uses daily firm-level stock return data from CRSP and annual financial statement data from Compustat. Applying standard data filters, we only include stocks traded on the NYSE, AMEX, and NASDAQ, with share codes equal to 10 and 11, and drop financial and utility firms (1-digit SIC equal to 4 or 6).

We analyze eight firm characteristics. The first four are some of the most commonly studied characteristics in corporate finance and capture important firm attributes. These are $Size$, computed daily as the natural log of market equity, which is the product of daily closing stock price and number of shares outstanding from CRSP; M/B (market-to-book

[5]See Kothari and Warner (2007) for a survey of the literature on the econometrics of event study analysis.

ratio), computed monthly as the log of market value of equity at the end of the prior month to book value (Compustat CEQ) measured at the end of the prior fiscal year (we exclude the small number of observations where CEQ is zero or negative); $Profit$, computed annually as gross profit (Compustat GP) divided by total assets (Compustat AT); and $Invest$, computed annually as capital expenditures (Compustat $CAPX$) divided by total assets.

The other four characteristics we analyze are from recently-published papers that focus on analyzing differences in stock returns around important quasi-experimental events. We choose these specific characteristics because they are easy to replicate. The characteristics, all measured annually, are $Cash/AT$, which is cash and short-term investments (Compustat CHE) divided by total assets; $Debt/AT$, which is the sum of long-term debt (Compustat $DLTT$) and debt in current liabilities (Compustat DLC), divided by total assets; $TaxRate$, which is 100 times income taxes paid (Compustat $TXDP$) divided by the difference between pre-tax income (Compustat PI) and special items (Compustat SPI), set to 0 if $PI < 0$; and $NYHQ$, which is an indicator variable equal to 1 if a firm is headquartered in New York (Compustat $STATE$ equal to "NY") and 0 otherwise.[6]

The primary unit of observation is a firm-day, though we also analyze returns over periods of five days. We associate with each observation the value of each characteristic as of the most recent available date prior to the date of the observation. The resulting sample consists of 21,060,248 firm-days belonging to 11,549 unique firms. Table 1 presents summary statistics for the sample.

[Table 1 about here]

[6]Fahlenbrach, Rageth, and Stulz (2021) studies return differences with the first two variables around the arrival of the COVID-19 pandemic in 2020 to test the effects of financial structure on resilience; Wagner, Zeckhauser, and Ziegler (2018) studies return differences with the third variable around the resolution of the 2016 U.S. Presidential election to test whether returns are higher for firms likely to benefit more from a lower-corporate tax rate regime; and Acemoglu, Johnson, Kermani, Kwak, and Mitton (2016) studies return differences with the fourth variable around the announcement of Timothy Geithner as nominee for Treasury Secretary in November 2008 to test whether better-connected firms benefited from this appointment.

2. Testing Differences in Returns Around Quasi-Experimental Events

In this section, we present a framework for understanding the consequences of concurrent events when testing hypotheses about the differential effect of a quasi-experimental event on the valuations of firms with different characteristics. We then explain why traditional statistical tests based on standard errors from cross-sectional event-period return regressions are likely to produce excessive false positives. Finally, we present evidence that the threat of excess false positives in these tests is generally severe.

2.1. Conceptual Framework

We begin by presenting a simple framework that models a discrete quasi-experimental event, with both event-period return effects that depend on a firm characteristic as well as stochastic "routine news" with return effects that also depend on the same characteristic. We focus on a one-day event period and a single characteristic here for simplicity but extend the framework to multi-day events and multiple firm characteristics in Appendix A. Let $r_{i,t}$ denote firm i's stock return on day t. Suppose that a researcher wishes to study differences in the effect of a quasi-experimental event occurring on day $t = \tau$ (the "event day") on firm value across firms that differ on characteristic $x_{i,\tau-1}$. The researcher uses event-day stock return $r_{i,\tau}$ to measure the value impact of the event on firm i.

In addition, suppose that firm i's return on day t depends on stochastic routine news n_t, which affects firms with different values of $x_{i,t-1}$ differently. News n_t represents a composite of all events on day t that have different implications for the valuation of firms with different values of x. It is observable to market participants, but a researcher cannot measure it except through its effect on returns. It is drawn independently on each day t from a distribution with mean 0 and variance σ_n^2. The exact shape of the distribution is not important. Routine news arrives on the event day, $t = \tau$, just like it does on other days. So, on the event day, there are potentially two pieces of news that affect the returns of firms with different values of $x_{i,\tau-1}$ differently – the event that the researcher wishes to study and routine news n_τ.

Because researchers typically estimate the differential effect of an event using linear regressions, and for the sake of simplicity, we assume that returns $r_{i,t}$ are linearly related to the interactions of $x_{i,t-1}$ with both the event, when it occurs, and routine news. Formally, the data generating process is

$$r_{i,t} = m_t + (n_t + e\mathbb{1}(t = \tau))(x_{i,t-1} - \bar{x}_{t-1}) + \delta_{i,t}, \tag{1}$$

where e is the differential effect of the quasi-experimental event that the researcher is studying and is the object of interest, $\mathbb{1}()$ is the indicator function, $\bar{x}_{t-1}$ is the cross-sectional mean of $x_{i,t-1}$, m_t is a mean-$\bar{m}$ variable that captures the aggregate return implications of any news on the day and is independent of n_t, and $\delta_{i,t}$ is a mean-0 i.i.d. error term. Since $\mathbb{E}[n_t] = 0$, $\mathbb{E}[r_{i,t}] = \bar{m} + e\mathbb{1}(t = \tau)x_{i,t-1}$.[7]

2.2. Testing for Differences Using Cross-Sectional Regressions

Because the researcher cannot observe routine news n_t, and the returns on day τ reflect the effects of both the quasi-experimental event and routine news, there is no way for the researcher to estimate e and n_τ separately. The best the researcher can do is to estimate $e + n_\tau$. We rewrite Eq. (1) for day $t = \tau$ in estimable form as

$$r_{i,\tau} = a_\tau + b_\tau x_{i,\tau-1} + \delta_{i,\tau}, \tag{2}$$

where $a_\tau = m_\tau - (n_\tau + e)\bar{x}_{i,\tau-1}$ and $b_\tau = e + n_\tau$. This is the simplest version of the standard regression equation that researchers typically estimate to test for differences in short-term returns around a quasi-experimental event.

[7] The general framework in Appendix A allows $x_{i,t-1}$ to affect $\mathbb{E}[r_{i,t}]$ and permits correlations among $\delta_{i,\tau}$.

Let $\hat{b}_\tau$ denote a regression estimate of b_τ. Since $\delta_{i,\tau}$ is independent of $x_{i,\tau-1}$ across i, the standard error of $\hat{b}_\tau$ from the regression is valid for assessing the precision of $\hat{b}_\tau$ as an estimate of b_τ and testing the null hypothesis $b_\tau = 0$. Moreover, because $\mathbb{E}[n_\tau x_{i,\tau-1}] = \mathbb{E}[n_\tau]\mathbb{E}[x_{i,\tau-1}] = 0$, $\mathbb{E}[\hat{b}_\tau] = e$ – that is, $\hat{b}$ is an unbiased estimate of e. Thus, the possibility of news that causes no differences in average returns as a function of x does not bias an event-period return relationship as an estimate of the differential event effect. However, even if the news does not cause differences in returns on average, it does cause differences – positive or negative – on any given day, including the day of the event. That is, $\mathbb{E}[\hat{b}_\tau | n_\tau] = e + n_\tau \neq e$ almost surely. This possibility breaks the assumption that errors are i.i.d. when using $\hat{b}_\tau$ to test hypotheses about e. More specifically, errors of firms with similar (dissimilar) values of $x_{i,\tau-1}$ are positively (negatively) correlated with each other. This can easily be seen by rewriting Eq. (1) for day τ as

$$r_{i,\tau} = m_t + e(x_{i,\tau-1} - \bar{x}_{t-1}) + \nu_{i,\tau}, \tag{3}$$

where $\nu_{i,\tau} = n_\tau(x_{i,\tau-1} - \bar{x}_{\tau-1}) + \delta_{i,\tau}$, and observing that

$$Cov(\nu_{i,\tau}, \nu_{j,\tau}) = E[\nu_{i,\tau}\nu_{j,\tau}] = (x_{i,\tau-1} - \bar{x}_{\tau-1})(x_{j,\tau-1} - \bar{x}_{\tau-1})\sigma_n^2, \tag{4}$$

which follows from the fact that $\mathbb{E}[\delta_{i,\tau}] = \mathbb{E}[\delta_{j,\tau}] = \mathbb{E}[\delta_{i,\tau}\delta_{j,\tau}] = 0$. Thus, the returns of firms with high values of x will tend to comove positively with each other, and the returns of firms with low values of x will tend to comove positively with each other. The intensity of comovement increases with σ_n^2, which captures the quantity of routine news that differentially affects firms with different values of x. We can also readily see the effect of stochastic news n_τ on the standard error of $\hat{b}_\tau$ as an estimate of e:

$$Var(\hat{b}_\tau - e) = Var(\hat{b}_\tau - b_\tau + b_\tau - e) = Var(\hat{b}_\tau - b_\tau + n_\tau) = Var(\hat{b}_\tau - b_\tau) + \sigma_n^2, \tag{5}$$

where the last step follows from the fact that the sampling error $\hat{b}_\tau - b_\tau$ is independent of the news n_τ. This is the key result from the analysis in this section.

Result 1. *The variance of estimation errors for b_τ as an estimate of e is $Var(\hat{b}_\tau - e) = Var(\hat{b}_\tau - b_\tau) + \sigma_n^2$.*

Note that the sampling error $Var(\hat{b}_\tau - b_\tau)$ would be present even absent routine news that affects firms with different values of $x_{i,\tau-1}$ differently. While sampling error tends towards 0 as the sample size grows, a larger sample does not reduce the additional variance from the impact of routine news, which drives a wedge between the true b_τ and e that would remain even in an infinite sample. Since the standard error of the estimate $\hat{b}_\tau$, $\sqrt{Var(\hat{b}_\tau - b_\tau)}$, overstates the precision of $\hat{b}_\tau$ as an estimate of e, a statistical test of $\hat{b}_\tau$ based on this standard error rejects the null hypothesis that the event has no differential value effects across firms differing on characteristic x. The severity of this problem scales with σ_n^2.

Intuitively, using the standard error of $\hat{b}_\tau$ obtained from estimating the regression Eq. (2) tests the wrong null hypothesis. This standard error would be appropriate for testing the null that there is no relationship between returns and characteristic x on day τ – i.e., that $b_\tau = 0$. However, the researcher's objective when estimating (2) is to test the null that the focal event does not *cause* a relationship between event-day returns and x – i.e., that $e = 0$. These are different nulls since there may well be a genuine relationship (i.e., one not caused by sampling error) between returns and x on the event day due to routine news n_τ, even if the focal event itself does not cause any differences in returns.

While the possibility of other news that differentially affects firms with different values of x is a problem theoretically, its practical importance is unclear. Moreover, the severity of this problem may vary across different firm characteristics. We explore the practical implications of this problem next.

2.3. Observed Rates of Statistical Significance

To assess the practical risk of false positives when using standard errors from a regression of event-day returns on a characteristic to test significance in real-world applications, we estimate long sequences of short-term return regressions. Specifically, for each of the eight characteristics we describe in Section 1 and each of the 7,811 trading days in our sample period 1991–2021, we estimate OLS regressions of returns on the characteristic. We then compute the fraction of regressions in which the coefficient on the characteristic is statistically significant at the 1% and 5% levels based on the coefficient's standard error in the regression.[8] The more frequently the coefficient for a particular characteristic is statistically significant in these tests, the harder it is to attribute a significant coefficient on the day of a specific event to the event itself.

In practice, researchers sometimes compute White (1980) adjusted standard errors to account for heteroskedasticity or "cluster" standard errors to account for correlated regression errors when testing differences in event returns. Since the threat of false positives caused by routine news when testing for differences in the event effect manifests in correlated regression errors (Eq. (4)), clustering in particular might be effective in alleviating this threat. The most commonly-specified dimension of clustering in event studies of returns is 49 industry categories from Fama and French. We therefore begin by computing statistical significance rates using default standard errors, White (1980) adjusted standard errors, and standard errors clustered at the Fama-French 49-category industry level. In addition, we compute these rates based on industry-clustered standard errors where we include Fama-French 49-category industry fixed effects, since the combination of industry fixed effects and industry clustering is sometimes used in practice. Figure 2 presents the results from this analysis.

[Figure 2 about here]

[8] We focus on 2-sided tests throughout the paper since these are the convention in finance research.

Statistical significance rates at the 1% and 5% significance level based on default standard errors average 31.6% and 42.2%, respectively, across the eight characteristics. These rates are greatest for $Size$, M/B, and $Cash/AT$, suggesting that these attributes capture important differences in exposure to day-to-day news. That descriptors such as size and market-to-book ratio that are often used to characterize firms are statistically significantly related to returns on more days than, say, headquarters location is not surprising. Day-to-day news is more likely to have a similar effect on the valuations of firms of similar size or expected growth rates than on those headquartered in the same state.

Significance rates based on White-corrected standard errors are generally smaller than those based on default standard errors, and those based on industry-clustered standard errors are smaller still. However, even significance rates based on industry-clustered standard errors average 17.7% and 29.2% at the 1% and 5% significance levels, respectively. Including industry fixed effects in addition to clustering at the industry level reduces significance rates further in most but not all cases.

The stark takeaway from Figure 2 is that statistically significant relationships between returns on the day of a specific event and a firm characteristic are likely to be common, even if the event does not differentially affect the returns of firms that differ on that characteristic. As a result, the risk of false positives regarding the differential effect of the event based on tests of such relationships is likely to be too high to allow for reliable inference. Moreover, adjusting standard errors to account for correlations in errors at the industry level does not appear to be a sufficient cure for this problem, suggesting that the return correlation patterns caused by routine news are too complex for industry membership to capture. We return to this last issue in Section 4.

In this baseline analysis, we estimate separate univariate regressions for each characteristic without controls and use daily raw returns. In practice, researchers often include control variables and/or analyze multiple characteristics of interest. We therefore extend our analysis to the cases of multivariate regressions. In addition, researchers sometimes measure returns over event periods longer than one day, either because the event plays out over a period longer than one day or because of uncertainty about exactly when the market learns about and processes the event. We extend our analysis to returns measured over five days instead of one to understand the implications of using longer event periods. Finally, we extend our analysis to other measures of returns that are sometimes used in practice, including risk-adjusted and logged returns.

Table 2 reports average statistical significance rates across the eight characteristics we study for all four combinations of univariate and multivariate regressions and one-day and five-day event periods. In the multivariate regressions, the other seven characteristics serve as control variables for the characteristic of interest. Panel A reports results based on the different ways of computing standard errors shown in Figure 2, while Panel B reports results where we use alternative return measures.

[Table 2 about here]

Statistical significance rates are even higher for five-day event periods than for one-day event periods. While we assume that n_t is mean-zero and i.i.d., in practice there is a small amount of autocorrelation in the impact of routine news on returns-characteristic relationships. Thus, if returns are related to a characteristic on the first day of a five-day period, they are slightly more likely than average to have a relationship with the same sign on subsequent days, resulting in more statistically significant relationships with five-day returns than with one-day returns. Significance rates are generally lower for multivariate regressions than for univariate regressions. Intuitively, control variables account for common exposure to day-to-day news that overlaps to some degree with common exposure based on the characteristic of interest. This result suggests an unappreciated benefit of including control variables when analyzing differences in an outcome variable after a quasi-experimental event: doing so can reduce the risk of false positives due to contemporaneous events. Note that this rationale for including control variables is different than the standard rationale of absorbing variation in regression errors that might otherwise be correlated with the characteristic of interest and hence cause bias.

Panel B of Table 2 shows that risk-adjusting returns decreases significance rates slightly. Intuitively, when factor loadings are correlated with firm characteristics, risk-adjusted returns will be less correlated with firm characteristics than unadjusted returns. Appendix Table C1 reports results for additional clustering approaches and shows that significance rates remain elevated for alternative clustering choices as well.

Ultimately, regardless of how returns are measured, how standard errors are computed, and whether control variables are included, Table 2 shows average significance rates remain stubbornly high. The prevalence of statistically significant relationships would make it hard to conclude with confidence from a statistically significant relationship in the period around a specific event that the event causes differences in returns.

Another way to quantify the threat posed by false positives due to routine news in this setting is to determine by how much standard errors would need to be inflated to achieve the intended rate of false positives absent the event, assuming that the time-series distribution of a relationship is the distribution of the counterfactual event-period relationship. We refrain from tabulating the required rates of inflation since the information is largely redundant with that in Table 2, with higher significance rates in that table translating into greater need to inflate standard errors. However, to provide some context, the average default standard error across all eight characteristics in a univariate regression with a one-day event window would need to be multiplied by 3.4 (4.1) to achieve the intended rate of false positives at the 5% (1%) level, while the average Fama-French 49 industry-clustered standard error would need to be multiplied by 2.2 (2.5).

3. Using Pre-Event Period Relationships to Test Significance

In this section, we consider an alternative methodology to test for a relationship between event-period returns and a firm characteristic that involves comparing this relationship to the time-series distribution of the same relationship estimated in a series of pre-event periods. We first detail specific approaches to implementing this methodology, including a novel GLS-based approach, and discuss some practical considerations when using these approaches. We show that these approaches, if implemented correctly, produce approximately the correct rate of false positives. We then present evidence that our GLS-based approach offers substantial increases in statistical power over an OLS-based approach.

3.1. Time-Series OLS and GLS

Comparing an event-period relationship to a distribution of pre-event period relationships effectively uses the pre-event distribution to approximate the distribution of the event-period relationship under the null hypothesis that the event itself has no differential effect. Returning to the framework described in Section 2.1, this testing strategy uses the distribution of pre-event relationships to approximate the distribution of $\hat{b}$ under the null that $e = 0$. We describe a procedure for implementing this methodology in the case of a single-day event period and then show how to extend it to a multi-day event period in Appendix A. The procedure involves three steps. The first two are:

Step 1: Estimate single-day cross-sectional regressions:

$$r_{i,t} = b_t x_{i,t-1} + \epsilon_{i,t}, \tag{6}$$

for the event day, $t = \tau$, and for a series of L pre-event days, $t = \tau_{-1}, \tau_{-2}, \ldots, \tau_{-L}$.[9]

[9] A natural choice of pre-event days is the L days immediately preceding the event, $t = \tau - L, \tau - (L - 1), \ldots, \tau - 1$. However, one could build in a lag $d > 0$ and use days $t = \tau - L - d, t = \tau - (L-1) - d, \ldots \tau - 1 - d$

Step 2: Estimate the event-specific effect as:

$$\hat{e} = \hat{b}_\tau - \mu(\hat{b}_{\tau_{-\ell}}), \qquad \mu(\hat{b}_{\tau_{-\ell}}) = \frac{1}{L}\sum_{\ell=1}^{L} \hat{b}_{\tau_{-\ell}}, \tag{7}$$

where $\hat{b}_{\tau_{-\ell}}$ is the estimated coefficient for pre-event day $\tau_{-\ell}$.[10]

The third and final step involves testing the statistical significance of $\hat{e}$. We describe two specific approaches to this step. Before describing these approaches, we derive some useful properties of $\hat{e}$ (see Appendix A for the proof):

Result 2. *Assuming stationary distributions for sampling error* $\hat{b}_{\tau_{-\ell}} - b_{\tau_{-\ell}}$ *and news* n_t, $\hat{e}$ *has the following relations to the true event-specific effect* e*:*

$$\mathbb{E}(\hat{e} - e) = \underbrace{0}_{\textit{No bias}}, \tag{8}$$

$$Var(\hat{e} - e) = \underbrace{Var(\hat{b}_{\tau_{-\ell}})}_{\textit{Time-series variance}}, \tag{9}$$

$$\Phi(\hat{e} - e) = \underbrace{\Phi(\hat{b}_{\tau_{-\ell}})}_{\textit{Time-series CDF}}, \tag{10}$$

where $\Phi(x)$ *is the cumulative density function evaluated at* x*.*

First, $\hat{e}$ is an unbiased estimator of e, which is the object of interest. Second, the distribution of sampling error in $\hat{e}$ is simply the time-series distribution of the coefficients from the pre-event day regressions, which is useful for statistical testing. The first specific approach to testing $\hat{e} = 0$ leverages Eq. (9) and computes a t-statistic, using the standard deviation of the pre-event window coefficients to compute the standard error of $\hat{e}$.

---

if the risk of information leakage prior to the event is high. In principle, one could also use post-event days as well as or instead of pre-event days. However, a pre-event window is less likely to include any post-announcement drift or follow-on events that might muddy interpretation.

[10]One could refrain from subtracting $\mu(\hat{b}_{\tau_{-\ell}})$ since the mean relationship between returns and characteristics is generally close to zero. In practice, subtracting $\mu(\hat{b}_{\tau_{-\ell}})$ has little effect on the results that follow.

Step 3(t): Compute the t-statistic for testing $\hat{e} = 0$ as

$$\hat{t} \equiv \frac{\hat{e}}{\sqrt{\text{Var}(\hat{e} - e)}} = \frac{\hat{b}_\tau - \mu(\hat{b}_{\tau_{-\ell}})}{\sqrt{\text{Var}(\hat{b}_{\tau_{-\ell}})}}, \tag{11}$$

and compare $\hat{t}$ to critical values from the standard normal, Student's T, or other cumulative distribution function.[11]

The second approach to testing $\hat{e} = 0$ is to compute a p-value based on the empirical cumulative distribution function (CDF) of the pre-event day coefficients - i.e., the empirical counterpart to (10).

Step 3(CDF): Compute a p-value for testing $\hat{e} = 0$ as:

$$\hat{p} = \frac{1}{L+1}\left[\sum_{\ell=1}^{L} \mathbb{1}\left(\left|\hat{b}_{\tau_{-\ell}} - \mu(\hat{b})\right| > \left|\hat{b}_\tau - \mu(\hat{b})\right|\right)\right], \tag{12}$$

and compare it to the significance threshold percentile p^*.

Note that this is a two-tailed test since it compares the absolute value of the (de-meaned) event-day coefficient to the distribution of absolute values on pre-event days. The advantage of this approach over the first is that it imposes no distributional assumptions on the time series of pre-event relationships that is used as a benchmark. As we show shortly, the time-series distribution of relationships between short-term returns and firm characteristics tends to exhibit fat tails, which can cause excess false positives when using a t statistic for testing.[12]

[11] This approach is identical to estimating a second-stage regression of the time-series of coefficients $\hat{b}_{\tau_{-L}}, \ldots, \hat{b}_{\tau_{-1}}, \hat{b}_\tau$ on a constant and an indicator equal to 1 on day τ and 0 otherwise. The coefficient on the indicator equals $\hat{b}_\tau - \mu(\hat{b}_{\tau_{-\ell}})$, while the standard error equals $\sqrt{\text{Var}(\hat{b}_{\tau_{-\ell}})}$ when a small-sample correction is applied to the variance. $\hat{t}_j$ is therefore the standard t-statistic for this second-stage regression. Note the similarity of this approach to the Fama and MacBeth (1973) methodology in asset pricing.

[12] While the p-value approach does not produce standard errors directly, standard errors can easily be inferred with the addition of a distributional assumption.

The statistical significance threshold level p^* imposes a mild technical constraint on the number of pre-event periods L that the researcher can use with the CDF approach. Specifically, it requires that $L+1$ be divisible by $\frac{1}{p^*}$. Otherwise, no pre-event day absolute coefficient deviation falls exactly at the p^*th percentile of the distribution.[13] Because the critical values researchers care about in practice often include 1%, it is natural to choose L to be one less than a multiple of 100.

We consider two specific approaches to this broad methodology – an OLS-based approach akin to the Sefcik and Thompson (1986) "portfolio OLS" approach, and a novel GLS-based approach. The only difference between the two approaches is in Step 1 (estimation of cross-sectional regressions for event and pre-event days). The OLS-based approach estimates the regressions in Step 1 using OLS. We refer to this approach as **time-series OLS** (TS-OLS for short).[14] The GLS approach, in contrast, estimates both the event-day and pre-event day regressions in Step 1 using GLS instead. We refer to the GLS-based approach as **time-series GLS** (TS-GLS for short).

The GLS approach requires estimating a covariance matrix of returns on each day, the inverse of which is used to weight observations in the regression. The potential advantage of this approach is that it increases the efficiency of both event-day and non-event day estimates by placing less weight on pairs of firms whose returns are more correlated with each other and therefore contain less independent information. The combination of tightening the distribution of pre-event day coefficients by removing noise and more precisely estimating the event-day coefficient should increase statistical power, making differential effects of an event easier to detect.

[13] Suppose $L = 252$. In this case, $\hat{p} \leq 1\%$ only when $\hat{b}_\tau$ is in the top three of the 253 estimated coefficients. This implies $\mathbb{P}(\hat{p} \leq 1\%) = \frac{3}{253} > 1\%$, and thus an event-day observation that is a random draw from the empirical pre-event distribution will be significant at the 1% level more often than 1% of the time.

[14] Sefcik and Thompson (1986) shows that this implementation is tantamount to forming portfolios with weights determined by the distribution of the explanatory characteristics and then comparing event and pre-event portfolio returns. Their approach maps into the t statistic approach to implementing TS-OLS but is not amenable to computing a p value based on the empirical CDF of the coefficient distribution.

Reliably estimating a covariance matrix is typically challenging in corporate finance applications since doing so requires many draws of the data with independent variation. Because short-term stock returns are observable daily and approximately serially uncorrelated, doing so is potentially feasible in the context of short-term stock returns. Estimating every element of the return covariance matrix individually using sample covariance, however, is generally infeasible, as doing so with fewer days of data than firms produces a rank-deficit matrix. Instead, we use PCA to embed the most important information about return comovement into the estimated covariance matrix. We demean returns on each day prior to conducting PCA so the resulting covariance estimates better represent the covariance of residuals in cross-sectional regressions that include a period-specific intercept. This approach to estimating a covariance matrix of returns is similar to the approaches in Giglio and Xiu (2021) and Lopez-Lira and Roussanov (2023).

Writing $f_{k,t}$ as the realization of the kth PC of returns on day t, we specify the covariance matrix of returns by assuming an arbitrage pricing theory (APT) structure:

$$r_{i,t} = \phi_i + \sum_{k=1}^{K} \lambda_{i,k} f_{k,t} + \epsilon_{i,t}, \tag{13}$$

$$\mathrm{Cov}(\epsilon_{i,t}, \epsilon_{j,t}) = \begin{cases} \sigma_i^2 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases}. \tag{14}$$

Given this structure, we can estimate the elements of the covariance matrix of returns as:

$$\hat{\Omega}_{r,ij} \equiv \mathrm{Cov}(r_{i,t}, r_{j,t}) = \sum_{k=1}^{K} \lambda_{i,k} \lambda_{j,k} \mathrm{Var}(f_{k,t}) + \mathbb{1}(i = j)\sigma_i^2. \tag{15}$$

We choose an approach that maximizes use of the data to construct factors rather than specifying factors *ex ante*, as is common in asset pricing, because our objective is to estimate covariances as accurately as possible and not to explain the cross section of returns using economically meaningful factors. As we will see in Section 4, the first constructed factor is often essentially the market factor, but the remaining constructed factors overlap little with other standard asset pricing factors, and the factors with which they overlap vary over time. Thus, a set of factors chosen *ex ante* is likely to produce a much less accurate estimate of the covariance matrix.

3.2. Practical Considerations

Implementing either TS-OLS or TS-GLS requires two practical choices. The first choice is L, the number of pre-event days to use. A longer series of pre-event days affords a larger sample of pre-event coefficient observations to use in statistical testing. However, using pre-event days farther back in time from the event increases the likelihood that changes over time in the distribution of news affecting firms differentially as a function of firm characteristics (i.e., n_t) reduce comparability. In addition, as discussed above, using the CDF approach requires that the number of pre-event periods equal one less than a multiple of 100. In our analysis, we use 199 pre-event trading days, which seems like a reasonable compromise in terms of period length and satisfies the restriction for using the p-value approach, though we explore the implications of using different numbers of pre-event days.

The second practical choice is when to measure firm characteristics if these characteristics are time-varying. To avoid look-ahead bias, the characteristics should always be measured prior to a given (event or pre-event) day. One option is to use the characteristics measured on the most recent available date prior to the earliest pre-event day in the comparison set. While this approach is simple, it may be inefficient if characteristics change over time. Another option is to measure a characteristic for each day as of the most recent prior date for which it is available. For example, if a characteristic is an annual financial variable, then $x_{i,t-1}$ would be the value of x for firm i as of the most recent fiscal year end prior to day t. Given the potential efficiency gain, we adopt this latter approach in our analysis.

Implementing the TS-GLS approach requires two additional choices – K, the number of PCs to use in constructing the covariance matrix, and T_K, the number of days to use in computing the PCs. The number of PCs must be a whole number satisfying $0 \leq K \leq T_K$. Using 0 PCs is equivalent to using weighted least squares (WLS) rather than GLS and accounts for heteroskedasticity but not return correlations. Using more PCs allows for more precise estimates of the return covariances. This increased precision should increase the efficiency of cross-sectional estimates and hence of the statistical power of the TS-GLS approach, at least up to a point. However, beyond a certain point, adding more PCs may result in overfitting, which reduces efficiency. A larger number of days T_K allows for estimating more PCs and therefore a richer covariance matrix, but using data from farther in the past increases the likelihood that changes in covariances over time reduce comparability. We choose $T_K = 200$, a round number comparable to our choice of L that, again, seems like a reasonable compromise relative to the tradeoff.

We use two approaches to gain insight into the optimal number of PCs to use. First, we analyze the relationship between the number of PCs K and the variance of the minimum variance portfolio constructed using an ex-ante estimate of the covariance matrix of returns Ω based on PCA, as specified in Eq. (15).[15] Given an estimated covariance matrix for returns on day t, $\hat{\Omega}_t$, the minimum variance portfolio's weights $w_{mvp,t}$ are

$$w_{mvp,t} = \underset{w \;\; s.t. \; w'\mathbf{1}=1}{\arg\min} \; Var_{t-1}(w'r_t) = \frac{\hat{\Omega}_t^{-1}\mathbf{1}}{\mathbf{1}'\hat{\Omega}_t^{-1}\mathbf{1}}, \tag{16}$$

where $\mathbf{1}$ is a vector of ones, and the denominator assures that the $w_{mvp,t}$ sum to one.

The relationship between K and the volatility of the minimum variance portfolio is informative about the incremental information content of each additional PC for forecasting

[15] Clarke, De Silva, and Thorley (2006) shows that using PCA to estimate the covariance matrix and form a minimum-variance portfolio of US equities results in substantial risk reduction with little or no reduction in average returns.

the inverse of the next-day covariance matrix – exactly the object we use for GLS. Figure 3 plots this relationship. The variance of the minimum variance portfolio decreases sharply with the addition of the first several PCs. The variance flattens out around $K = 50$ and is largely invariant until $K = 199$.

[Figure 3 about here]

Second, for different numbers of PCs, K, and different characteristics, we estimate daily return regressions using both GLS and OLS and then analyze the tightness of the GLS coefficient distribution relative to the OLS coefficient distribution. Since the time-series approaches involve comparing the event-day relationship to the pre-event relationship distribution, a tighter distribution implies greater ease in detecting an event-related effect. To quantify the relative tightness, we compute the ratio of the standard deviations of the daily GLS and OLS coefficients. Note that the OLS coefficient standard deviation is constant for a given characteristic in this calculation since we are only varying the number of PCs used to estimate GLS. We compute the ratio because it is informative about the increase in precision from using GLS instead of OLS. For each characteristic we study, Figure 4 plots this ratio against the number of PCs used.

[Figure 4 about here]

As with the relationship between minimum variance and number of PCs, the ratio of GLS to OLS coefficient standard deviations declines sharply with the first twenty PCs and is essentially flat between 50 and 195 PCs. This conclusion adds to the evidence that the power gains from TS-GLS are largely insensitive to the number of PCs beyond the first twenty. In the spirit of choosing a round number, we use $K = 100$ in the remainder of the analysis, though we also explore the sensitivity of results to using different numbers of PCs.

One potential concern with both the TS-OLS and TS-GLS approaches is that the distribution of n_t may be time-varying and, in particular, may differ between the event period and pre-event periods. If the distribution in the event period is more variable than in pre-event periods, then both TS-OLS and TS-GLS could produce excess false positives. This difference in variability could occur because the focal event is part of a set of connected events, of which others affect different firms in ways similar to the focal event. For example, an unexpected change in corporate tax rates may be part of a broader package of tax reforms announced all at once, which would make it difficult conceptually to determine whether differences in event-period returns are attributable to the change in corporate tax rates or another part of the overall tax reform. It is therefore important to investigate the broader context in which an event occurs before treating the event as a clean quasi-natural experiment. Time variation in the distribution of n_t could also add noise to the estimated covariance matrix used in the TS-GLS approach, which would weaken the power of this approach.

3.3. Performance of Time-Series Approaches

Mirroring our analysis of cross-sectional tests, we begin our analysis of the TS-OLS and TS-GLS approaches by applying these approaches for each characteristic to each one-day and five-day period in our sample period and computing the fraction of periods in which the estimated relationship is statistically significant at the 1% and 5% levels. We compute these fractions for both the t-statistic and CDF approaches to statistical testing described in Section 3.1. Figure 5 presents the results.

[Figure 5 about here]

As anticipated, statistical significance rates for a given threshold significance level using both time-series approaches are much closer to the threshold significance level than those based on standard cross-sectional regressions. However, these rates are still slightly higher than the threshold when we use the t-statistic approach, especially at the 1% significance level. These excess rates arise from fat tails in the distribution of coefficients that a t test fails to take into account. In contrast, statistical significance rates based on p-values using the empirical CDF of coefficients are approximately equal to the threshold significance level, more or less by construction. We therefore focus exclusively on the CDF approach for the remainder of our analysis.

We next assess the statistical power of the TS-OLS and TS-GLS approaches. One at a time, for each day in the period 1991–2021 and each characteristic, we create an artificial event effect by introducing a relationship between returns and the characteristic of 25 basis points (bp) per one-standard deviation increase in the characteristic. We do not add this effect on pre-event days relative to the artificial event day. We then use the TS-OLS and TS-GLS approaches to estimate and test the significance of differences in returns with the characteristic on the artificial event day at the 1% and 5% significant levels. Finally, for each characteristic, we compute the fraction of artificial event days on which the test is able to detect the artificial event effect at each of the two significance levels. We repeat this process using an artificial event effect of 50 bp return per one standard deviation increase in the characteristic. Figure 6 presents detection rates based on these tests.

[Figure 6 about here]

Using either time-series approach, detection rates vary widely and are generally smaller for the characteristics size, market-to-book ratio, and cash holdings. Note that these are the same characteristics for which tests based on standard errors from cross-sectional event-period regressions produce the highest rates of statistically significant relationships (Figure 2). This is not a coincidence. The frequent existence of strong relationships even absent the event in question makes it more difficult to detect the differential effect of the event.

Not surprisingly, detection rates are higher when we introduce a larger artificial effect. More importantly, the TS-GLS approach detects the artificial effect more frequently than the TS-OLS approach in every case, with large increases in detection rates in most cases. For example, TS-GLS detects 25 bp of additional return per one standard deviation change in a characteristic at the 5% significance level 1.2-2.9 times as often as TS-OLS, depending on the characteristic. The greatest detection rate gains occur for characteristics where TS-OLS detection rates are relatively low.

To allow for a more comprehensive comparison, Table 3 reports average detection rates across all eight characteristics for TS-OLS and TS-GLS for the 25 and 50 bp artificial effects separately for one-day and five-day event periods. Average detection rates across the eight characteristics are 1.3-2.1 times and 1.6-3.7 times as high for TS-GLS as for TS-OLS at the 5% and 1% significance levels, respectively. Detection rates are much lower for five-day event periods than for one-day event periods. Intuitively, the noise in returns is increasing in period length, making a given effect size more difficult to detect. The degree to which detection rates deteriorate with the length of the event period is a serious concern, since papers often analyze multi-day event periods. While not the focus of this paper, the results suggest that the benefit of precisely pinning down the timing with which the market responds to a discrete event and therefore being able to use a short event period is large.

[Table 3 about here]

As discussed in Section 3.2, a researcher implementing TS-GLS needs to choose L, the number of pre-event days for testing, and K, the number of PCs used to construct the covariance matrix. In the analysis above, we set $L = 199$ and $K = 100$ because we believe that 199 pre-event days allow us to determine the distribution of the coefficients without losing comparability and that 100 PCs capture the most important aspects of return covariances. In Appendix Table C2, we examine detection rates using TS-GLS for different choices of L and K and find that detection rates are similar for reasonable choices of these parameters.

We conclude from this analysis that using either time-series approach instead of a standard error from the event-period regression to test significance effectively addresses the threat of excessive false positives caused by contemporaneous events. However, TS-GLS provides a big improvement in statistical power over TS-OLS. Our `csestudy` Stata module can use either approach at the user's discretion.

4. Principal Component Analysis

The PCs employed in the TS-GLS approach capture information about covariances in returns. In this section, we further analyze the PCs to learn more about the nature of these covariances. We begin by plotting the fraction of total variation in returns that the first one, five, 25, and 50 PCs explain by year. Figure 7 presents these plots. It also plots the average explanatory power of the Fama-French 49-category industries for comparison.

[Figure 7 about here]

Two observations are worth making. First, confirming the conclusions of Lopez-Lira and Roussanov (2023), it takes many PCs to explain the variation in returns. The PCs are essentially factors chosen one-by-one specifically to maximize incremental explanatory power. One might imagine that only a few factors chosen with so many degrees of freedom would be needed to explain a large fraction of variation in returns. Yet, the first 5 PCs explain less than 30% of the variation in returns in most years, and even the first 25 PCs explain less than 50% of the variation in most years. Going from 25 to 100 PCs nearly doubles the fraction of variation explained. It therefore appears that correlation patterns in returns are extremely complex.

Second, the Fama-French 49-category industries combined only explain approximately the same fraction of cross-sectional return variation as the first PC does. In most years, just the first five PCs explain more than twice the variation that industry categories do. The limited ability of industry to explain the cross section of returns explains why clustering by industry has limited ability to correct standard errors in purely cross-sectional tests. Return correlations are too complex to summarize well with virtually any set of pre-chosen variables.

We next analyze the extent to which the first five PCs in each year map into identifiable economic factors that might plausibly drive return correlations. We present this analysis for four specific years for the purposes of illustration – 2008, 2020, and 2021, in which major market-moving events occurred, and 2010 to provide a relatively quiescent year for comparison. We choose four well-known factors from the asset pricing literature: the equity market portfolio ($MktRf$), small-minus-big portfolio (SMB), high-minus-low portfolio (HML), and up-minus-down portfolio (UMD). We construct four other factors as long-only equal-weighted portfolios of stocks designed to capture period-specific conditions. The $Tech$ factor portfolio is constructed from the Software, Hardware, and Chips Fama-French 49 industries; the $Finance$ portfolio from Banks, Real Estate, and Finance industries; and the $Covid$ portfolio from Meals, Healthcare, and Drugs industries. The $Memes$ factor portfolio combines whichever subset of GameStop (GME), AMC (AMC), Bed Bath and Beyond (BBBY), and Blackberry (BB) stocks was available to trade on each day. We orthogonalize the non-$MktRf$ factors with respect to $MktRf$ within each calendar year.

For each of the first five PCs in each year, we compute the returns on a portfolio where the weights are the elements of the PC. We then compute the absolute values of the correlations between each of these returns and each of the pre-specified factors. Table 4 presents the results. Panels A, B, C, and D present the results for 2008, 2010, 2020, and 2021, respectively.

[Table 4 about here]

The first PC-weighted portfolio return is highly correlated with the equity market portfolio in three of the four years. Because we de-mean returns within each stock-day prior to conducting PCA, the first PC does not mechanically capture the simple average return on each day. Instead, the first PC is often, but not always, correlated with $MktRf$ because it captures differences in daily returns driven by cross-sectional differences in market betas. Furthermore, because PCA weights stocks equally, the resulting first PC is effectively equal-weighted, while $MktRf$ is value-weighted. For this reason, the first PC is also somewhat correlated with the SMB factor portfolio return in all four years.

For 2008, the second and fifth PC portfolio-weighted returns are both highly correlated with the $Finance$, HML, and UMD factor portfolio returns. These two PCs both appear to pick up common exposure to the financial crisis. For 2020, the fourth PC is correlated with the *Covid* factor portfolio return. For 2021, the third PC-weighted portfolio return is highly correlated with the $Memes$ factor portfolio return, while the fifth is correlated with the HML, $Finance$, and $Tech$ factor portfolio returns. For 2010, a relatively quiescent year, none of the second through fifth PC-weighted portfolios are strongly correlated with any of the factor portfolio returns.

We broaden our analysis by calculating the fraction of all return variances explained by the factors presented in Table 4 in each year of the sample period. The results, presented in Figure 8, show that each factor is an important determinant of return variances in some years and largely irrelevant in others.[16] While this time variation is most dramatic for the $Finance$ and $Memes$ factors, whose explanatory powers peak in 2008 and 2021, respectively, even standard asset pricing factors such as SMB vary significant over time in their importance. Figure 8 also shows that there are many years, particularly early in our sample period, where none of the example factors is important in driving return variances.[17]

[Figure 8 about here]

Overall, the analyses in Table 4 and Figure 8 indicate that factors driving return comovement vary substantially from year to year and often represent factors unique to a period.

[16] As a point of reference, SMB has the highest average percent explained, 2.2%, comparable to the average for PC3. The other seven factors' average percent explained are all between 0.65% and 1%, comparable to the average for PCs 32 and 10, respectively.

[17] In untabulated analysis, we find that there are many years where at least one of the first five PCs does not strongly correlate with any of the 170+ factors from Chen et al. (2022).

Furthermore, it is often difficult to determine what drives comovement in any given period. Correlation patterns are complex and time-varying, necessitating the use of many statistically-extracting components to power our GLS approach.

5. Conclusions

Our results suggest that traditional tests of the relationship between returns around a quasi-experimental event and firm characteristics produce too many false positives to allow for reliable inference about the differential treatment effect of the event. Furthermore, it appears that clustering strategies are inadequate to address this problem because correlation patterns are too complex to summarize with a handful of group clusters. The solution to this problem is to use the distribution of relationships on pre-event days to conduct statistical testing. An event-day relationship should stand out before we conclude that the event causes differences in returns. Our analysis shows that a GLS-based approach to estimating event and pre-event day relationships offers substantial improvements in statistical power over a more conventional OLS-based approach. It may therefore allow for the detection of even modest differential event effects that might otherwise be difficult to test.

While our analysis focuses on returns because the daily frequency and approximate serial independence of returns make it possible to assess the risk of false positives due to correlated exposure to other events, it potentially has broader implications. There is every reason to believe that the same problems exist with other outcome variables commonly studied in corporate finance applications. Our analysis suggests that we should be cautious about reaching strong conclusions based on analysis in which clustering of standard errors on cross-sectional dimensions such as industry or location is used to address concerns about cross-sectionally correlated regression errors. We leave further exploration of these problems in the context of other outcome variables to future work.

Appendix A. General Conceptual Framework and Proofs

In this appendix, we present a formal conceptual framework for our analysis that generalizes the framework in Sections 2.1, 2.2, and 3.1 to allow for J different firm characteristics; N-day event windows; non-zero relationships between firm characteristics and average returns; and a general covariance matrix for returns.

The general data generating process is:

$$r_{i,t} = a_t + \left[\lambda + n_t + \frac{e}{N}\mathbb{1}(t \in \tau)\right]' (x_{i,t-1} - \bar{x}_{t-1}) + \delta_{i,t}, \tag{17}$$

$$\mathbb{E}[n_t] = \mathbb{E}[\delta_{i,t}] = \mathbb{E}[x_{i,t-1}\delta_{i,t}] = \mathbb{E}[n_t\delta_{i,t}] = 0, \mathrm{Var}(\delta_t) = \Omega_\delta, \tag{18}$$

where a_t is a random variable capturing the aggregate return implications of any news on day t, λ is a $J \times 1$ vector containing each characteristic's relation with average returns, n_t is a $J \times 1$ vector containing mean-zero random variables expressing how the news arriving in period t affects returns as a function of firm characteristics, e is a $J \times 1$ vector containing the differential effects of the event on per-day stock returns, τ is the set of days in the event period, $x_{i,t}$ is a $J \times 1$ vector of firm characteristics, $\bar{x}_{t-1}$ is the mean $x_{i,t}$, $\delta_{i,t}$ is a mean-zero error term, and δ_t is an $M \times 1$ vector containing all $\delta_{i,t}$ with covariance matrix Ω_δ.

The general framework in Eq. (17) simplifies to the framework studied in the body of the paper when $N = J = 1$, $\lambda = 0$, and Ω_δ is a diagonal matrix with constant variance. Proofs in this general framework therefore apply to the corresponding results in the simplified model.

Suppose a researcher conducts a cross-sectional OLS regression of total event-period returns on a constant and pre-event firm characteristics $x_{i,\tau-1}$:

$$\sum_{t\in\tau} r_{i,t} = a_\tau + b_\tau' x_{i,\tau-1} + \epsilon_{i,t}, \tag{19}$$

and uses the resulting $\hat{b}_\tau$ as an estimate of e.

Result 1 (Variance of b as an estimate of e)**.** *Assuming characteristics $x_{i,t}$ are constant for each i during the event window, cross-sectional OLS estimates of b_τ in Eq. (19) have the following relationships to the true event-specific effect e:*

$$\mathbb{E}[\hat{b}_\tau - e] = \underbrace{N\lambda}_{bias}, \tag{20}$$

$$Var(\hat{b}_\tau - e) = \underbrace{Var(\hat{b}_\tau - b_\tau)}_{OLS\ estimation\ error} + \underbrace{N\,Var(n_t)}_{Excess\ variance}. \tag{21}$$

where the $\mathbb{E}[\cdot]$ and $Var(\cdot)$ operators condition on $x_{i,\tau-1}$ and draw random $\delta_{i,t}$ and n_t for all $t \in \tau$.

Proof. Summing Eq. (17) across event window days yields:

$$\sum_{t\in\tau} r_{i,t} = \sum_{t\in\tau} a_t - \left(N\lambda + \sum_{t\in\tau} n_t + e\right)' \bar{x}_{t-1} + \left(N\lambda + \sum_{t\in\tau} n_t + e\right)' x_{i,\tau-1} + \sum_{t\in\tau} \delta_{i,t}, \tag{22}$$

which implies that in the population (with no estimation error):

$$a_\tau = \sum_{t\in\tau} a_t - \left(N\lambda + \sum_{t\in\tau} n_t + e\right)' \bar{x}_{t-1} \tag{23}$$

$$b_\tau = N\lambda + \sum_{t\in\tau} n_t + e, \tag{24}$$

$$\epsilon_{i,\tau} = \sum_{t\in\tau} \delta_{i,t}. \tag{25}$$

From, this we get:

$$\text{Var}(\hat{b}_\tau - e) = \text{Var}(\hat{b}_\tau - b_\tau + b_\tau - e) = \text{Var}\left(\hat{b}_\tau - b_\tau + N\lambda + \sum_{t\in\tau} n_t\right). \tag{26}$$

Because $\hat{b}_\tau$ is an OLS estimate,

$$\begin{bmatrix} \hat{a}_\tau - a_\tau \\ \hat{b}_\tau - b_\tau \end{bmatrix} = (X'X)^{-1}X'\epsilon_\tau, \tag{27}$$

where X is an $M \times (J+1)$ matrix containing a constant and $x_{i,\tau-1}$ for all i, and ϵ_τ is an $M \times 1$ vector containing all $\epsilon_{i,t}$. Using $E(n_t \delta_{i,t}) = 0$, Eq. (27) therefore implies that $\hat{b}_\tau - b_\tau$ is uncorrelated with n_t, and Eq. (26) simplifies to:

$$\text{Var}(\hat{b}_\tau - e) = \text{Var}\left(\hat{b}_\tau - b_\tau\right) + N\text{Var}(n_t). \tag{28}$$

We complete the proof by noting, from Eq. (25) and the unbiasedness of OLS, that $\mathbb{E}[\hat{b}_\tau] - e = b_\tau - e = N\lambda$. □

We now turn to the time-series OLS approach articulated in Section 3.1.

Result 2 (Unbiasedness and correct variance of the time-series OLS approach)**.** *Write $b_{\tau_{-\ell}}$ for the vector of coefficients relating $\sum_{t\in\tau_{-\ell}} r_{i,t}$ to $x_{i,\tau_{-\ell}}$ in non-event window $\tau_{-\ell}$, and b_τ for the coefficients in the event window. The time-series OLS estimates of e is:*

$$\hat{e} = \hat{b}_\tau - \frac{1}{L}\sum_{\ell=1}^{L} \hat{b}_{\tau_{-\ell}}, \tag{7}$$

Assuming stationary distributions for estimation error $\hat{b}_{\tau_{-\ell}} - b_{\tau_{-\ell}}$ and news n_t, $\hat{e}_\tau$ has the *following distribution across samples as $L \to \infty$ for a given e:*

$$\mathbb{E}[\hat{e} - e] = \underbrace{0}_{\textit{No bias}}, \tag{8}$$

$$\underbrace{Var(\hat{e} - e)}_{\textit{Sampling variance}} = \underbrace{Var(\hat{b}_{\tau_{-\ell}})}_{\textit{Time-series variance}}, \tag{9}$$

$$\underbrace{\Phi(\hat{e} - e)}_{\textit{Sampling CDF}} = \underbrace{\Phi(\hat{b}_{\tau_{-\ell}})}_{\textit{Time-series CDF}}. \tag{10}$$

Proof. Because e has no effect when $t \notin \tau$, the stationarity of $\hat{b}_{\tau_{-\ell}} - b_{\tau_{-\ell}}$ and news n_t imply:

$$\Phi(\hat{b}_\tau - e) = \Phi(\hat{b}_{\tau_{-\ell}}) \Rightarrow \tag{29}$$

$$\mathbb{E}[\hat{b}_\tau - e] = \mathbb{E}[\hat{b}_{\tau_{-\ell}}], \tag{30}$$

$$\text{Var}(\hat{b}_\tau - e) = \text{Var}(\hat{b}_{\tau_{-\ell}}), \tag{31}$$

where all moments are across samples unless otherwise stated. We therefore have:

$$\mathbb{E}[\hat{e} - e] = \mathbb{E}\left[\hat{b}_\tau - \frac{1}{L}\sum_{\ell=1}^{L} \hat{b}_{\tau_{-\ell}} - e\right] = \mathbb{E}[\hat{b}_\tau - e] - \mathbb{E}[\hat{b}_{\tau_{-\ell}}] = 0. \tag{32}$$

As $L \to \infty$, time-series estimates of Φ and variance converge to their across-sample counterparts. Combined with Eqs. (31) and (29), this gives us Eqs. (9) and (10). □

References

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2023). When should you adjust standard errors for clustering? *Quarterly Journal of Economics 138*(1), 1–35.

Acemoglu, D., T. A. Hassan, and A. Tahoun (2018). The power of the street: Evidence from Egypt's Arab Spring. *The Review of Financial Studies 31*(1), 1–42.

Acemoglu, D., S. Johnson, A. Kermani, J. Kwak, and T. Mitton (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics 121*(2), 368–391.

Amihud, Y. and S. Stoyanov (2017). Do staggered boards harm shareholders? *Journal of Financial Economics 123*(2), 432–439.

Bernard, V. L. (1987). Cross-sectional dependence and problems in inference in market-based accounting research. *Journal of Accounting Research 25*(1), 1–48.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Bizjak, J. M., S. L. Kalpathy, V. T. Mihov, and J. Ren (2022). Ceo political leanings and store-level economic activity during the Covid-19 crisis: effects on shareholder value and public health. *Journal of Finance 77*(5), 2949–2986.

Brav, A., C. Geczy, and P. A. Gompers (2000). Is the abnormal return following equity issuances anomalous? *Journal of Financial Economics 56*(2), 209–249.

Brown, J. R. and J. Huang (2020). All the president's friends: Political access and firm value. *Journal of Financial Economics 138*(2), 415–431.

Chen, A. Y., T. Zimmermann, et al. (2022). Open source cross-sectional asset pricing. *Critical Finance Review 11*(2), 207–264.

Child, T. B., N. Massoud, M. Schabus, and Y. Zhou (2021). Surprise election for Trump connections. *Journal of Financial Economics 140*(2), 676–697.

Clarke, R. G., H. De Silva, and S. Thorley (2006). Minimum-variance portfolios in the US equity market. *Journal of Portfolio Management 33*(1), 10–24.

Collins, D. W. and W. T. Dent (1984). A comparison of alternative testing methodologies used in capital market research. *Journal of Accounting Research 22*(1), 48–84.

Cremers, K. M., L. P. Litov, and S. M. Sepe (2017). Staggered boards and long-term firm value, revisited. *Journal of Financial Economics 126*(2), 422–444.

Eisfeldt, A. L., G. Schubert, and M. B. Zhang (2023). Generative AI and firm values. National Bureau of Economic Research Working Paper.

Fahlenbrach, R., K. Rageth, and R. M. Stulz (2021). How valuable is financial flexibility when revenue stops? Evidence from the COVID-19 crisis. *Review of Financial Studies 34*(11), 5474–5521.

Fama, E. F. and K. R. French (2015). A five-factor asset pricing model. *Journal of Financial Economics 116*(1), 1–22.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*(3), 607–636.

Giglio, S. and D. Xiu (2021). Asset pricing with omitted factors. *Journal of Political Economy 129*(7), 1947–1990.

Gilje, E. P. and J. P. Taillard (2017). Does hedging affect firm value? Evidence from a natural experiment. *Review of Financial Studies 30*(12), 4083–4132.

Jegadeesh, N. and J. Karceski (2009). Long-run performance evaluation: Correlation and heteroskedasticity-consistent tests. *Journal of Empirical Finance 16*(1), 101–111.

Jiang, H., C. Parsons, L. Sun, and S. Titman (2023). Skilled foreign labor, urban agglomeration, and value creation. Working Paper.

Kolari, J. W. and S. Pynnönen (2010). Event study testing with cross-sectional correlation of abnormal returns. *Review of Financial Studies 23*(11), 3996–4025.

Kothari, S. P. and J. B. Warner (2007). Econometrics of event studies. In *Handbook of Empirical Corporate Finance*, pp. 3–36. Elsevier.

Leftwich, R. (1981). Evidence of the impact of mandatory changes in accounting principles on corporate loan agreements. *Journal of Accounting and Economics 3*(1), 3–36.

Licht, A. N., C. Poliquin, J. I. Siegel, and X. Li (2018). What makes the bonding stick? A natural experiment testing the legal bonding hypothesis. *Journal of Financial Economics 129*(2), 329–356.

Liu, L. X., H. Shu, and K. J. Wei (2017). The impacts of political uncertainty on asset prices: Evidence from the Bo scandal in China. *Journal of Financial Economics 125*(2), 286–310.

Lopez-Lira, A. and N. L. Roussanov (2023). Do common factors really explain the cross-section of stock returns? Jacobs Levy Equity Management Center for Quantitative Financial Research Paper.

Lu, R. and Z. Ye (2023). Roe v. rates: Reproductive healthcare and public financing costs. Working Paper.

Lyon, J. D., B. M. Barber, and C.-L. Tsai (1999). Improved methods for tests of long-run abnormal stock returns. *Journal of Finance 54*(1), 165–201.

Mitchell, M. L. and E. Stafford (2000). Managerial decisions and long-term stock price performance. *Journal of Business 73*(3), 287–329.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*(3), 385–397.

Moulton, B. R. (1987). Diagnostics for group effects in regression analysis. *Journal of Business & Economic Statistics 5*(2), 275–282.

Petersen, M. A. (2009). Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies 22*(1), 435–480.

Schoenherr, D. (2019). Political connections and allocative distortions. *The Journal of Finance 74*(2), 543–586.

Sefcik, S. E. and R. Thompson (1986). An approach to statistical inference in cross-sectional models with security abnormal returns as dependent variable. *Journal of Accounting Research 24*(2), 316–334.

Wagner, A. F., R. J. Zeckhauser, and A. Ziegler (2018). Company stock price reactions to the 2016 election shock: Trump, taxes, and trade. *Journal of Financial Economics 130*(2), 428–451.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica 48*(4), 817–838.

Zeume, S. (2017). Bribes and firm value. *Review of Financial Studies 30*(5), 1457–1489.

Figure 1: Daily Relationship Between Returns & Investment in 2019

This figure presents estimates of the cross-sectional relationship between firm-level returns and *Investment* for each day in 2019, where *Investment* is capital expenditures divided by total assets. Each point depicts the absolute value of the the difference between the coefficient from a regression of one-day returns on *Investment* for the day to which it corresponds on the x-axis and the average coefficient for the year. A blue diamond indicates that the raw coefficient is statistically significant at the 5% level based on a t test using White (1980) standard errors, while a red circle indicates insignificance at the 5% level. The estimates for January 22 (the day after the Blood Moon), March 13, and October 11 are enclosed in larger diamonds for emphasis.

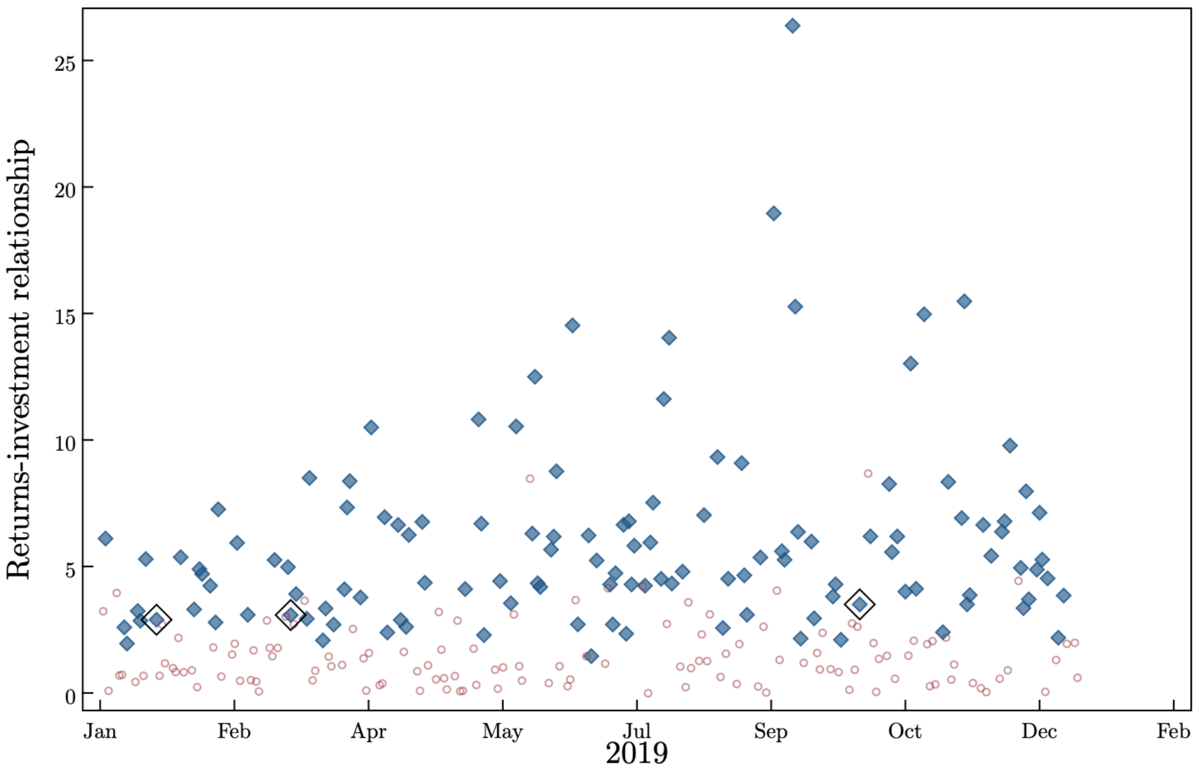

Figure 2: Significance Rates Based on Cross-Sectional Standard Errors

This figure depicts statistical significance rates from univariate cross-sectional OLS regressions of 1-day returns on each of eight characteristics. The bars show the percentage of days for which the coefficient from the regression is statistically significant based on default standard errors, White (1980) (robust) standard errors, Fama-French 49-category industry clustered standard errors, and industry clustered standard errors where we also include industry fixed effects. The top panel shows rates of significance at the 1% significance level, while the bottom panel shows rates of significance at the 5% significance level. Our sample is the 7,811 trading days in 1991 through 2021.

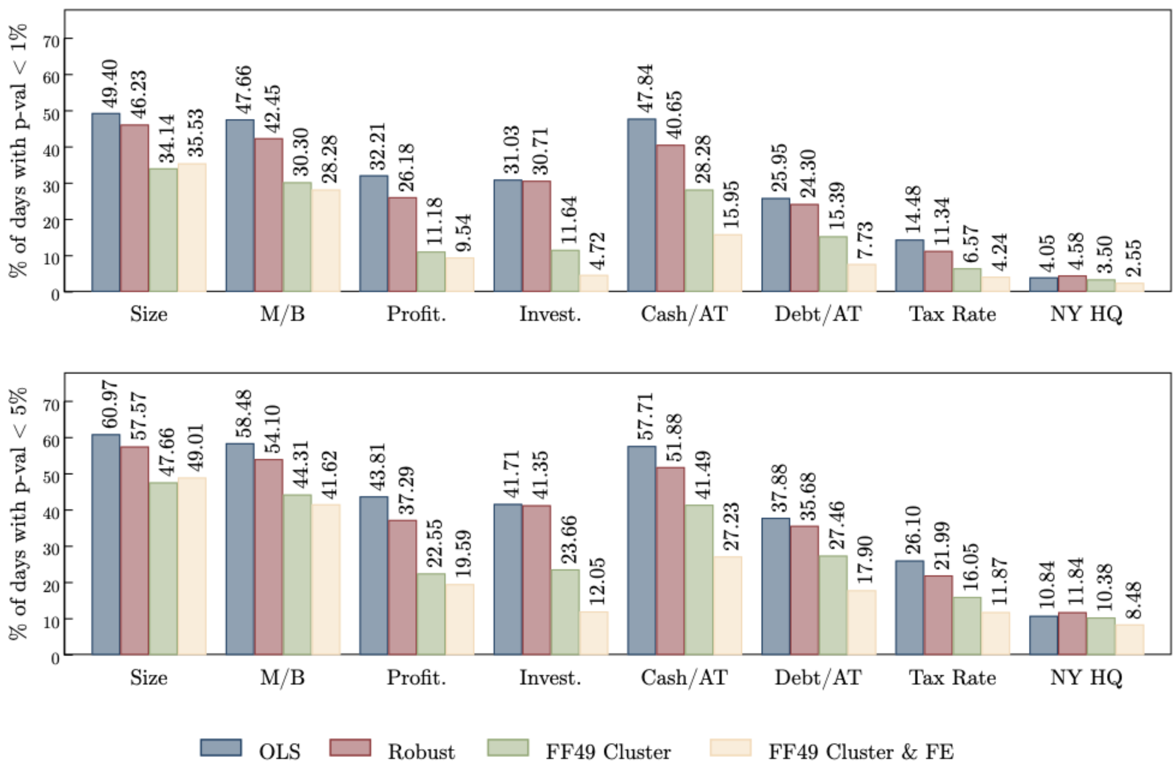

Figure 3: Minimum Variance Portfolio Variances and # of Principal Components

This figure depicts the average annualized standard deviation of daily returns for a minimum-variance portfolio of US equities as a function of the number of principal components (PCs), K, used to form an out-of-sample forecast for the covariance matrix of returns. Our covariance matrix forecasts are constructed using the first K PCs from implementing principal component analysis on balanced panel of daily returns over the prior 199 trading days, assuming that other than via these PCs each stock's return is uncorrelated. See Section 3.1 for more details. Our sample is the 7,811 trading days in 1991 through 2021.

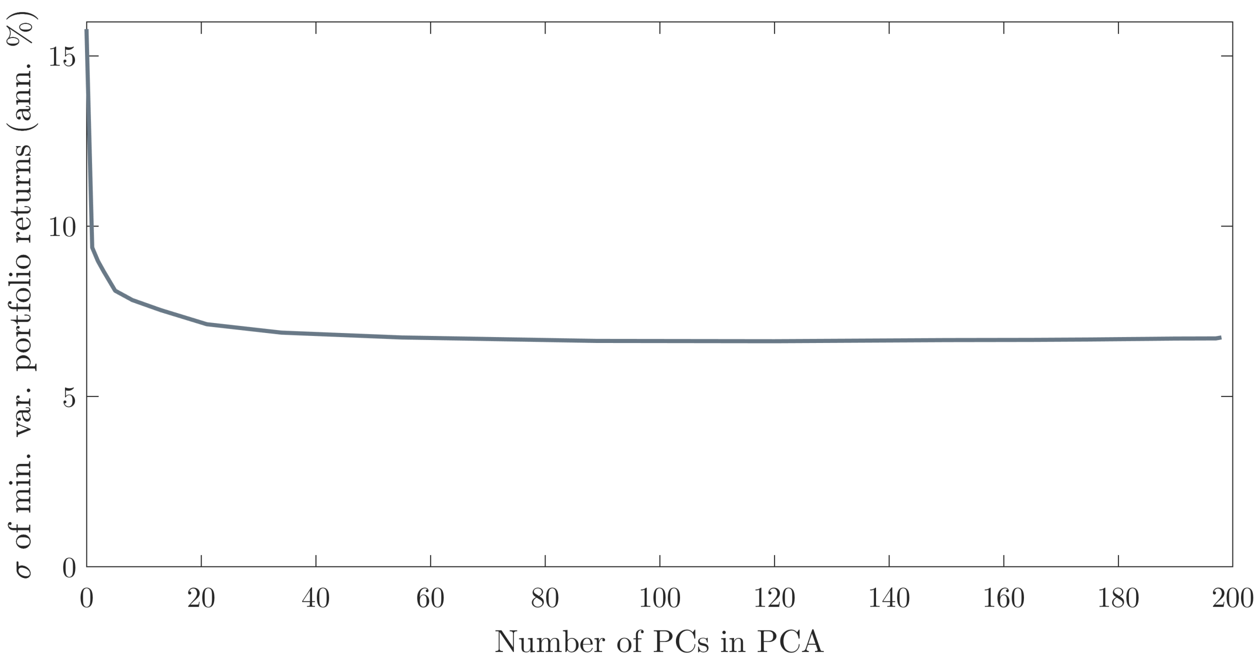

Figure 4: GLS/OLS Coefficient Standard Deviation Ratios and # of Principal Components

This figure plots the ratio of the time-series standard deviations of cross-sectional GLS and OLS coefficients from cross-sectional regressions of one-day returns on each of eight characteristics against the number of PCs we use to estimate the covariance matrix for the GLS regressions. See Section 3.1 for more details. Our sample is the 7,811 trading days in 1991 through 2021.

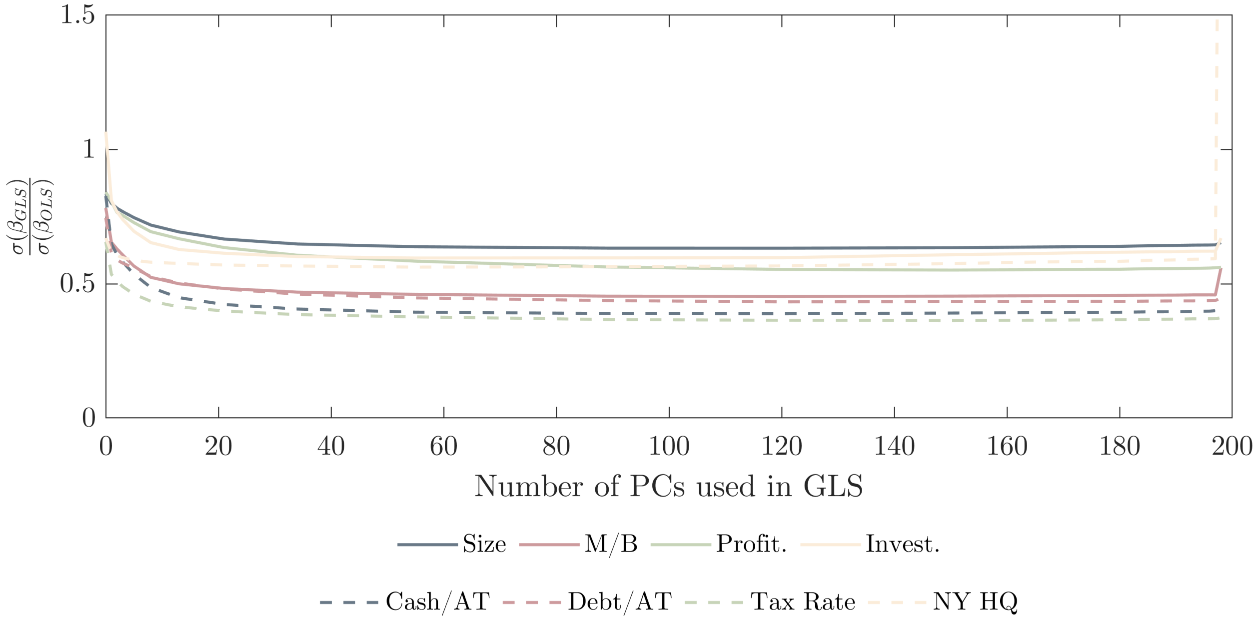

Figure 5: Significance Rates Based on TS-OLS and TS-GLS

This figure depicts statistical significance rates for the relationship between one-day returns and each of eight characteristics using the time-series OLS (TS-OLS) and time-series GLS (TS-GLS) approaches described in Section 3.1. We use 199 pre-event periods and, for each covariance matrix used in TS-GLS, 100 principal components extracted from the 200 preceding trading days. The bars show the percentage of days for which the coefficient from the regression is statistically significant based on t-statistics using the time-series of coefficients as standard errors ('T-Test') or based on a p value from comparison to the empirical time-series cumulative distribution function of coefficients ('CDF'). The top panel shows rates of significance at the 1% significance level, while the bottom panel shows rates of significance at the 5% significance level. Our sample is the 7,811 trading days in 1991 through 2021.

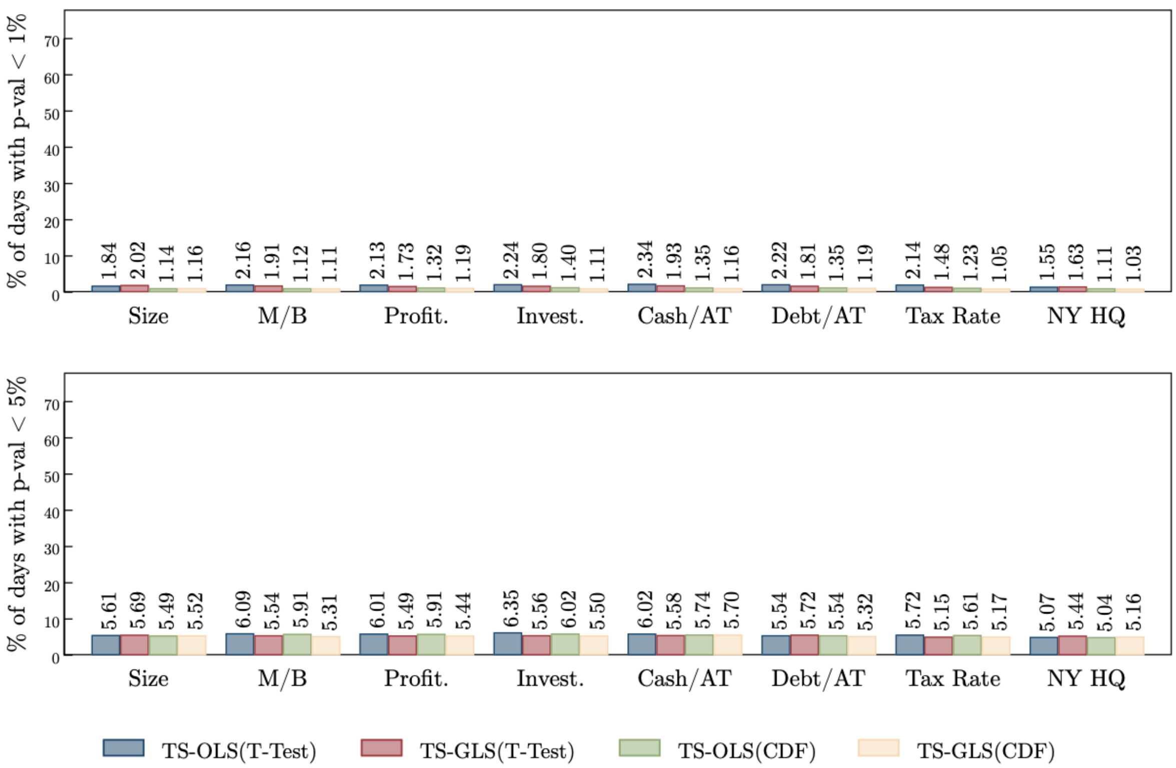

Figure 6: Detection Rates in Artificial Event Periods: TS-OLS vs TS-GLS

This figure depicts statistical significance rates for the relationship between one-day returns and each of eight characteristics using the time-series OLS (TS-OLS) and time-series GLS (TS-GLS) approaches described in Section 3.1. For each day, we create an artificial event by adding either 25 or 50 bp to a firm's return for each one standard deviation increase in the corresponding characteristic on the x axis. We use 199 pre-event days (with no added relationship) and, for each covariance matrix used in TS-GLS, 100 principal components extracted from the 200 preceding trading days. The bars show the percentage of days for which the coefficient from the regression is statistically significant at the 5% level based on a comparison to the empirical time-series cumulative distribution function of coefficients ('CDF'). The top panel shows results for a 25 bp added relationship, while the bottom panel shows results for a 50 bp added relationship. Our sample is the 7,811 trading days in 1991 through 2021.

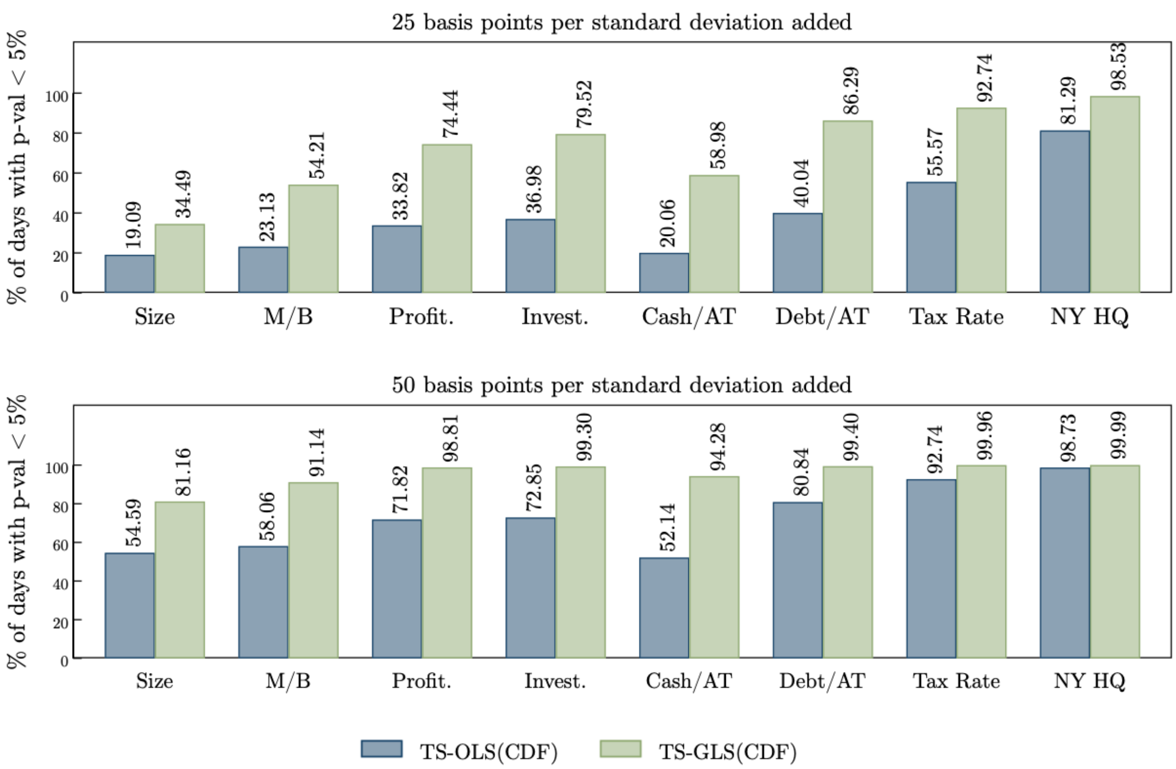

Figure 7: Explanatory Power of Principal Components of Daily Returns

This figure plots the average percent of the daily cross-sectional variance of returns that the first K principal components (PCs) combined explain by calendar year, for $K = 1$, 5, 25, and 100. It also shows the average percent explained by Fama-French 49-category industry portfolios combined for comparison. Our sample is the 7,811 trading days in 1991 through 2021.

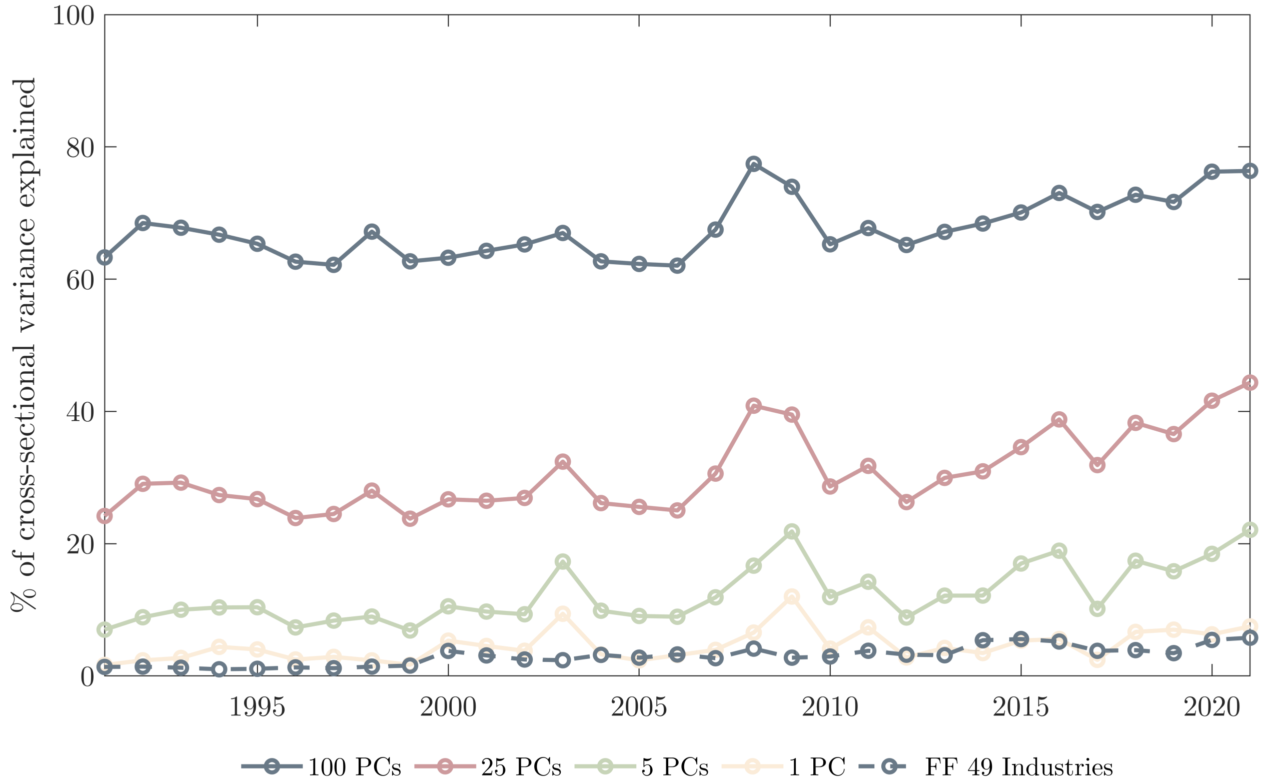

Figure 8: Explanatory Power of Example Factors for Daily Returns

This figure plots the average percent of the daily cross-sectional variance of returns that seven example factors explain by calendar year. The factors are size (SMB), value (HML) and momentum (UMD) factors, as collected from Ken French's data library; $Tech$, the equal-weighted average return of Software, Hardware, and Chips Fama-French 49 industries; $Finance$, the average return of Banks, Real Estate, and Finance industries; $Covid$, the average return of Meals, Healthcare, and Drugs industries; and $Memes$, the average return of whichever subset of GameStop (GME), AMC (AMC), Bed Bath and Beyond (BBBY), and Blackberry (BB) stocks is available to trade on each day. All factors are orthogonalized with respect to the excess market return ($MktRf$) in each calendar year. Our sample is the 7,811 trading days in 1991 through 2021.

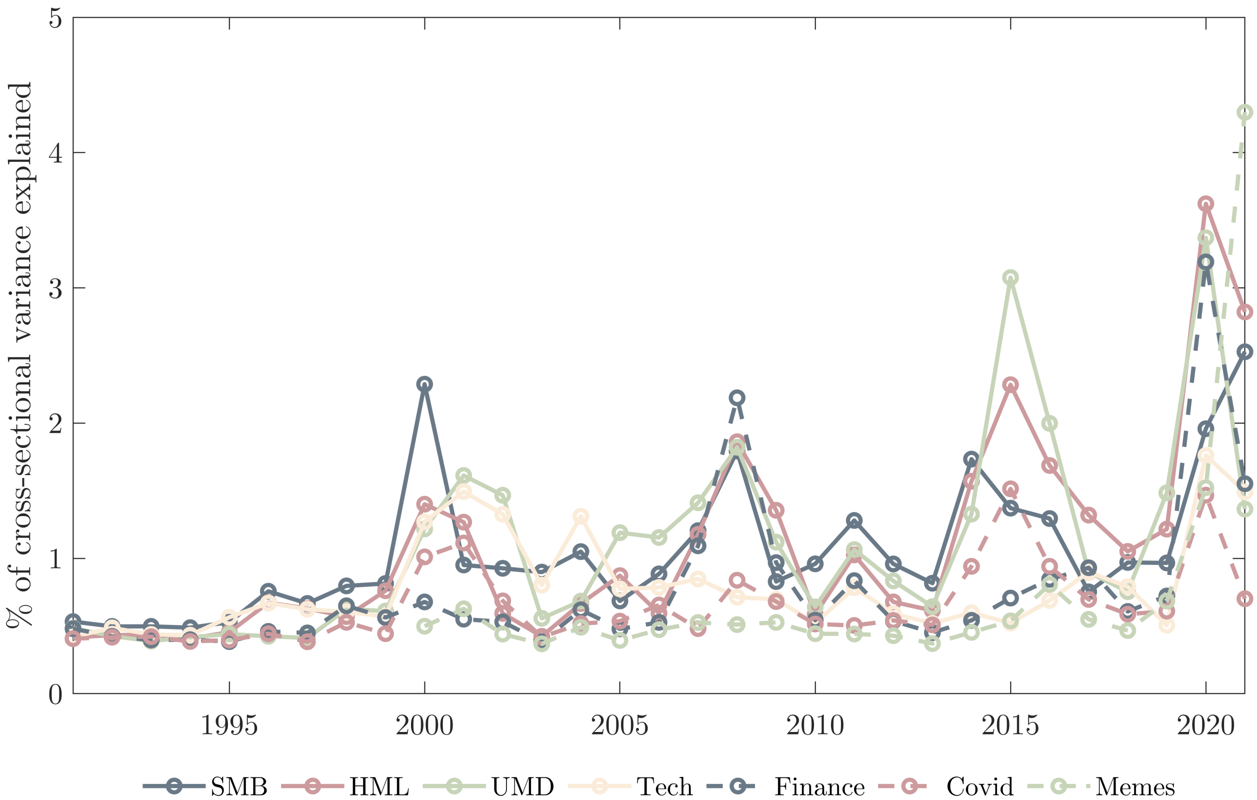

Table 1: Summary Statistics

This table presents summary statistics for the sample of firm-day observations used in our analysis. *Size* is the natural log of market equity, which is the product of the daily closing stock price on the previous day and the number of shares outstanding from CRSP. M/B is the log of market value of equity at the end of the prior month to book value (Compustat CEQ) measured at the end of the prior fiscal year (observations for which book value is zero or negative are excluded). $Profit.$ is gross profit (Compustat GP) divided by total assets (Compustat AT). $Invest.$ is capital expenditures (Compustat $CAPX$) divided by total assets. $Cash/AT$ is cash and short-term investments (Compustat CHE) divided by total assets. $Debt/AT$ is the sum of long-term debt (Compustat $DLTT$) and debt in current liabilities (Compustat DLC), divided by total assets. *Tax Rate* is 100 times income taxes paid (Compustat $TXDP$) divided by the difference between pre-tax income (Compustat PI) and special items (Compustat SPI), set to 0 if $PI - SPI < 0$. $NY\ HQ$ is an indicator variable equal to 1 if a firm is headquartered in New York (Compustat $STATE$ equal to "NY") and 0 otherwise. Our sample is the 7,811 trading days in 1991 through 2021.

Variable	Firm-Days	Firms	Firms/Day	Mean	Median	σ	Within-day σ
Return (%)	21,060,248	11,549	2,696	0.12	0.00	4.00	3.81
Size	21,051,745	11,548	2,695	6.14	6.01	1.96	1.81
M/B	20,477,556	11,403	2,622	1.01	0.92	0.98	0.96
Profit.	21,059,937	11,548	2,696	0.36	0.34	0.32	0.32
Invest.	20,887,703	11,510	2,674	0.06	0.04	0.07	0.07
Cash/AT	21,059,996	11,549	2,696	0.21	0.12	0.24	0.24
Debt/AT	20,986,859	11,543	2,687	0.21	0.16	0.22	0.22
Tax Rate	14,387,738	8,090	1,842	23.97	23.66	16.64	16.35
NY HQ	21,060,248	11,549	2,696	0.07	0.00	0.25	0.25

Table 2: Significance Rates Based on Cross-Sectional Standard Errors

This table presents statistical significance rates for coefficients from cross-sectional OLS regressions of short-term stock returns on each of eight firm characteristics based on cross-sectional standard errors. Each cell reports the percentage of all one- or five-day periods in which the regression coefficient is statistically significant, averaged across the eight characteristics. Each row reports the percentages for different combinations of 1% and 5% significance, univariate and multivariate regressions (where all eight characteristics are included as explanatory variables), and one-day and five-day return periods. Panel A presents rates of significance based on default standard errors, White (1980) adjusted standard errors, and standard errors clustered by Fama-French 49-category industry, with the fourth row based on regressions that also include Fama-French 49-category industry fixed effects. Panel B presents rates of significance for alternative return measures, including CAPM, Fama-French 3-factor, and Fama-French 4-factor adjusted returns and logged returns ($log(1 + return)$). Our sample is the 7,811 trading days in 1991 through 2021.

Panel A: Different Standard Errors

		1% Significance Level				5% Significance Level			
		Univariate		Multivariate		Univariate		Multivariate	
Return	Specification	1d	5d	1d	5d	1d	5d	1d	5d
Unadj.	Default SE	31.6%	38.7%	21.5%	28.4%	42.2%	49.2%	32.0%	38.8%
Unadj.	Robust SE	28.3%	35.5%	19.0%	25.6%	39.0%	46.2%	29.3%	36.1%
Unadj.	FF49 Cluster	17.7%	22.0%	13.6%	18.8%	29.2%	33.7%	24.1%	30.0%
Unadj.	FF49 Cl & FE	13.6%	18.2%	10.8%	17.2%	23.5%	29.0%	20.1%	27.8%

Panel B: Alternative Returns

		1% Significance Level				5% Significance Level			
		Univariate		Multivariate		Univariate		Multivariate	
Return	Specification	1d	5d	1d	5d	1d	5d	1d	5d
CAPM	FF49 Cl & FE	16.7%	20.4%	10.4%	15.5%	28.0%	32.2%	19.7%	25.6%
FF3	FF49 Cl & FE	13.1%	16.1%	8.4%	13.4%	24.1%	27.3%	17.4%	23.6%
FF4	FF49 Cl & FE	12.4%	15.3%	7.5%	12.6%	23.2%	26.6%	16.4%	22.6%
LogRet	FF49 Cl & FE	17.7%	22.3%	10.6%	15.9%	29.2%	34.0%	20.0%	26.4%

Table 3: Average Detection Rates in Artificial Event Periods: TS-OLS vs TS-GLS

This table presents statistical significance rates for the relationship between short-term (one- or five-day) returns and each of eight characteristics using the time-series OLS (TS-OLS) and time-series GLS (TS-GLS) approaches described in Section 3.1. For each one- or five-day period, we create an artificial event by adding either 25 or 50 bp to a firm's return for each one standard deviation increase in the characteristic in question, spreading the effect equally across days in the case of five-day periods. We use 199 pre-event periods (with no added relationship) and, for each covariance matrix used in TS-GLS, 100 principal components extracted from the 200 preceding trading days. Each row reports the percentage of all artificial event periods in which the estimated event-period relationship is statistically significant using TS-OLS and TS-GLS based on a p value from comparison to the empirical time-series cumulative distribution function of coefficients as well as the difference and ratio of these percentages. Different rows report these percentages for different combinations of event-period length (one- or five-day) and size of added return per one standard deviation increase in a characteristic (25 or 50 bp). The top panel shows significance rates at the 1% level, while the bottom panel shows significance rates at the 5% level. Our sample is the 7,811 trading days in 1991 through 2021.

Panel A: 1% Significance Level

Period	Effect (bp)	TS-OLS	TS-GLS	Difference	Ratio
1 day	25	17.26%	48.18%	30.92%	2.79
1 day	50	51.67%	86.04%	34.36%	1.66
5 day	25	3.13%	11.69%	8.55%	3.73
5 day	50	13.05%	40.48%	27.43%	3.10

Panel B: 5% Significance Level

Period	Effect (bp)	TS-OLS	TS-GLS	Difference	Ratio
1 day	25	38.75%	72.40%	33.65%	1.87
1 day	50	72.72%	95.50%	22.78%	1.31
5 day	25	12.85%	27.07%	14.22%	2.11
5 day	50	31.68%	61.14%	29.46%	1.93

Table 4: Evolving Correlations With Principal Components

This table presents the correlations between portfolios representing the first five principal components of individual stock returns in a calendar year and a variety of other factors constructed from stock returns. The first four are the market ($MktRf$), size (SMB), value (HML) and momentum (UMD) factors. The second four are period-specific factors: $Tech$, the equal-weighted average return of Software, Hardware, and Chips Fama-French 49 industries; $Finance$, the average return of Banks, Real Estate, and Finance industries; $Covid$, the average return of Meals, Healthcare, and Drugs industries; and $Memes$, the average return of whichever subset of GameStop (GME), AMC (AMC), Bed Bath and Beyond (BBBY), and Blackberry (BB) stocks is available to trade on each day. We orthogonalize the non-$MktRf$ factors with respect to $MktRf$ in each calendar year. Panel A presents results for 2008, Panel B for 2010, Panel C for 2020, and Panel D for 2021.

Panel A: 2008

	'Mkt'	'Crisis'	?	?	?
PC	1	2	3	4	5
% x-sectional var. explained	6.6%	3.1%	2.8%	2.2%	2.0%
$\|\rho(PC_i, MktRf)\|$	89%	22%	10%	1%	10%
$\|\rho(PC_i, SMB)\|$	29%	27%	25%	6%	4%
$\|\rho(PC_i, HML)\|$	11%	56%	8%	11%	43%
$\|\rho(PC_i, UMD)\|$	10%	45%	19%	12%	53%
$\|\rho(PC_i, Tech)\|$	2%	1%	7%	2%	12%
$\|\rho(PC_i, Finance)\|$	21%	61%	20%	11%	39%
$\|\rho(PC_i, Covid)\|$	0%	19%	15%	4%	1%
$\|\rho(PC_i, Memes)\|$	2%	3%	5%	9%	9%
R^2	91%	55%	12%	3%	31%

Panel B: 2010

	'Mkt'	?	?	?	?
PC	1	2	3	4	5
% x-sectional var. explained	4.1%	3.9%	1.4%	1.3%	1.2%
$\|\rho(PC_i, MktRf)\|$	93%	5%	13%	5%	3%
$\|\rho(PC_i, SMB)\|$	29%	1%	16%	18%	14%
$\|\rho(PC_i, HML)\|$	3%	3%	26%	13%	17%
$\|\rho(PC_i, UMD)\|$	8%	3%	31%	21%	19%
$\|\rho(PC_i, Tech)\|$	0%	2%	1%	4%	0%
$\|\rho(PC_i, Finance)\|$	7%	9%	4%	24%	1%
$\|\rho(PC_i, Covid)\|$	2%	1%	18%	4%	14%
$\|\rho(PC_i, Memes)\|$	1%	1%	11%	8%	8%
R^2	95%	1%	22%	19%	10%

Panel C: 2020

	'Mkt'	?	'Covid'	?	?
PC	1	2	3	4	5
% x-sectional var. explained	6.3%	4.6%	2.9%	2.5%	2.1%
$\lvert\rho(PC_i, MktRf)\rvert$	60%	6%	6%	5%	6%
$\lvert\rho(PC_i, SMB)\rvert$	28%	36%	30%	8%	13%
$\lvert\rho(PC_i, HML)\rvert$	54%	45%	36%	3%	9%
$\lvert\rho(PC_i, UMD)\rvert$	47%	47%	36%	13%	2%
$\lvert\rho(PC_i, Tech)\rvert$	33%	29%	22%	2%	20%
$\lvert\rho(PC_i, Finance)\rvert$	55%	38%	17%	2%	0%
$\lvert\rho(PC_i, Covid)\rvert$	26%	6%	39%	21%	11%
$\lvert\rho(PC_i, Memes)\rvert$	31%	21%	23%	6%	17%
R^2	73%	29%	47%	11%	12%

Panel D: 2021

	?	'Quant'	'Memes'	?	'Tech'
PC	1	2	3	4	5
% x-sectional var. explained	7.5%	5.7%	4.0%	2.7%	2.2%
$\lvert\rho(PC_i, MktRf)\rvert$	11%	37%	16%	7%	34%
$\lvert\rho(PC_i, SMB)\rvert$	23%	44%	27%	8%	43%
$\lvert\rho(PC_i, HML)\rvert$	1%	46%	15%	48%	61%
$\lvert\rho(PC_i, UMD)\rvert$	5%	40%	10%	5%	8%
$\lvert\rho(PC_i, Tech)\rvert$	22%	17%	9%	25%	44%
$\lvert\rho(PC_i, Finance)\rvert$	8%	18%	3%	38%	56%
$\lvert\rho(PC_i, Covid)\rvert$	4%	11%	12%	12%	5%
$\lvert\rho(PC_i, Memes)\rvert$	44%	15%	81%	2%	7%
R^2	25%	64%	70%	25%	65%

Table C1: Significance Rates Based on Alternative Cross-Sectional Standard Errors

This table presents mean statistical significance rates for coefficients from cross-sectional OLS regressions of short-term stock returns on each of eight firm characteristics based on alternative cross-sectional standard errors. Each cell reports the percentage of all 1- or 5-day periods from 1991 through 2021 in which the regression coefficient is statistically significant. Each row reports the percentages for different combinations of 1% and 5% significance, univariate and multivariate regressions (where all eight characteristics are included as explanatory variables), and one-day and five-day return periods. Panel A presents rates of significance based on standard errors clustered at different industry levels, at the state level, at the level of size quintile by book-to-market quintile bin, and at the level of size quintile by book-to-market quintile by gross profitability quintile bin. Panel B presents rates of significance based on standard errors clustered by both industry and size.

Panel A: Alternative Clustering

		1% Significance Level				5% Significance Level			
		Univariate		Multivariate		Univariate		Multivariate	
Return	Specification	1d	5d	1d	5d	1d	5d	1d	5d
Unadj.	SIC4 Cluster	21.3%	26.7%	15.6%	21.3%	32.5%	38.1%	25.9%	32.2%
Unadj.	FF49 Cluster	17.7%	22.0%	13.6%	18.8%	29.2%	33.7%	24.1%	30.0%
Unadj.	FF30 Cluster	20.2%	24.6%	15.4%	20.4%	29.2%	33.6%	24.2%	29.6%
Unadj.	State Cluster	30.0%	36.0%	23.1%	29.3%	42.0%	48.1%	34.8%	41.0%
Unadj.	5 Size x 5 B/M	20.2%	25.5%	15.2%	21.0%	31.9%	37.9%	26.2%	32.7%
Unadj.	... x 5 Profit.	23.8%	25.5%	17.6%	24.0%	35.0%	37.9%	28.1%	34.7%

Panel B: Two-way clustering

		1% Significance Level				5% Significance Level			
		Univariate		Multivariate		Univariate		Multivariate	
Return	Specification	1d	5d	1d	5d	1d	5d	1d	5d
Unadj.	FF49 & State	20.6%	24.4%	17.3%	22.3%	31.8%	35.9%	28.7%	34.4%
Unadj.	FF30 & State	19.7%	23.5%	16.0%	20.7%	31.0%	35.2%	27.6%	33.0%

Table C2: Average TS-GLS Detection Rates in Artificial Event Periods: Robustness

This table presents mean rates with which the times-series GLS (TS-GLS) approach to significance testing described in Section 3.1 successfully detects relationships between one-day returns and each of eight characteristics added to create artificial event days for different numbers of pre-event days (L) and principal components (K) used to construct covariance matrices. We use the preceding 200 trading days to compute principal components. Each row reports the percentage of all artificial event periods from 1991 to 2021 in which the estimated event-period relationship is statistically significant for different numbers of principal components. Different rows report these percentages for different combinations of number of pre-event days and size of added return per one standard deviation increase in a characteristic (25 or 50 bp). The top panel shows significance rates at the 1% level, while the bottom panel shows significance rates at the 5% level.

Panel A: 1% Significance Level

			1 day			5 day	
Effect (bp)	L	$K = 50$	$K = 100$	$K = 150$	$K = 50$	$K = 100$	$K = 150$
25	149	40.07%	41.36%	40.44%	10.07%	10.99%	11.03%
25	199	47.11%	48.18%	47.12%	10.82%	11.69%	11.71%
25	249	43.71%	44.84%	43.76%	8.99%	9.82%	9.83%
50	149	80.26%	81.31%	80.82%	35.57%	39.37%	39.62%
50	299	85.35%	86.04%	85.81%	37.07%	40.48%	41.33%
50	249	83.36%	84.15%	83.87%	34.06%	37.19%	37.86%

Panel B: 5% Significance Level

			1 day			5 day	
Effect (bp)	L	$K = 50$	$K = 100$	$K = 150$	$K = 50$	$K = 100$	$K = 150$
25	149	71.18%	71.91%	71.09%	25.57%	26.91%	27.12%
25	199	71.80%	72.40%	71.62%	25.57%	27.07%	27.22%
25	249	71.06%	71.64%	70.81%	24.50%	26.00%	25.98%
50	149	95.09%	95.22%	95.18%	57.64%	60.92%	61.64%
50	199	95.37%	95.50%	95.38%	57.78%	61.14%	61.95%
50	249	95.21%	95.34%	95.17%	56.93%	60.25%	60.95%