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

Jonathan Cohn
University of Texas at Austin

Travis L. Johnson
University of Texas at Austin

Zack Liu
University of Houston

Malcolm Wardlaw
University of Georgia

August 2023

Abstract

This paper assesses different approaches to testing the cross-sectional valuation effects of an event such as a regulatory change for firms with different characteristics. Standard cross-sectional return regressions typically reject at the 1% significance level more than 20% of the time in *non-event* periods, suggesting that the bar for rejecting in these tests is far too low. Clustering standard errors results in only a modest reduction in excess rejection. Using the time-series distribution of cross-sectional OLS coefficients from regressions in pre-event windows to conduct inference addresses the excess rejection problem but typically results in low-power tests. We propose an alternative approach involving a time-series of cross-sectional GLS regressions, using principal component analysis of pre-event returns to estimate the covariance matrix, and show that this approach offers substantial improvements in power.

*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 Sheridan Titman and seminar participants at The University of Texas at Austin for valuable feedback. We thank Christopher McDonald for excellent research assistance.

If observations with similar covariate values in a regression have correlated model errors, then the corresponding standard error is typically biased downward, resulting in excess rejection under the null (i.e., Type I errors). This issue is endemic in empirical corporate finance research, where firms with similar characteristics likely have common exposure to economic forces and therefore cross-correlated outcomes. A lack of cross-sectional independence potentially compromises inference based on any analysis of data with a cross-sectional component, even if only within variation is used for identification (as in a firm fixed effects regression). Researchers in corporate finance typically address concerns about cross correlations, if at all, by clustering standard errors at the level of a cross-sectional grouping such as industry or location. However, both the severity of the excess rejection problem and whether clustering is sufficient to account for cross correlations more generally is unclear.

This paper studies the cross-correlation problem and potential solutions in the context of cross-sectional event studies of stock returns around a market-moving event such as an unanticipated regulatory shock. Our main takeaway is that using standard errors from the event-window cross-sectional regression itself to conduct statistical testing, as is standard, produces substantial excess rejection under the null, even when clustering by industry and/or location. In contrast, using the distribution of coefficients from a time-series of cross-sectional regressions outside of the event window to conduct statistical testing produces approximately the correct rejection rate under the null, though we show that care must be taken with the choice of test. We introduce a novel GLS-based variant of this approach and show that it offers substantial improvements in power over a more conventional OLS-based approach. More broadly, our analysis suggests that cross-correlated errors may be a greater problem and cross-sectional clustering less effective at addressing it than is commonly understood.

Researchers use cross-sectional event studies of the type we analyze to test predictions about the differential value implications of a given event for different types of firms. This empirical strategy has been a part of the corporate finance toolkit for decades and, noting that a return represents a change in a firm's valuation, represents one of the earliest forms of difference-in-differences estimation in finance. To gain insight into how researchers establish statistical significance in these tests in practice, we conduct an informal survey of papers in top-3 finance journals implementing these tests and find that 87% report standard errors produced by the event-window cross-sectional regression itself. Of these, 35% report unadjusted standard errors; 25% report robust standard errors, which address concerns about heteroskedasticity but not cross correlations; 35% report standard errors clustered by industry; and 5% report standard errors clustered by geographic location.

How serious is the cross correlation problem in these tests? To answer this question, we reproduce the data sets used in several recent cross-sectional event studies published in top-3 finance journals. For each, we estimate a series of cross-sectional regressions of returns over short windows (1 day or 5 days) on the associated covariates of interest in the period 1991-2021, excluding the event window itself. We then compute the fraction of regressions in which the coefficient of interest is statistically significant at different significance levels. Since the event in question does not occur in these non-event windows, this fraction provides an estimate of the rejection rate under the null that the event does not differentially affect firm valuations. We also conduct the same exercise using several covariates commonly analyzed in finance but not, to our knowledge, analyzed in a specific cross-sectional event study, noting that there is no event window to exclude in this case.

For typical covariates, tests in non-event windows reject at the 1% significance level more than 20% of the time. The problem is especially acute for covariates closely related to firm fundamentals (e.g., firm size), where rejection rates at the 1% level in non-event windows can exceed 50%. While clustering by industry reduces excess rejection rates, tests still typically reject at the 1% level more than 10% of the time in non-event windows, with rejection rates exceeding 30% for covariates closely related to firm fundamentals. The problem caused by rejection rates this high outside of a specific event being studied is that they undermine how much we can learn from a test of the differential effect of the event itself. Conditional on rejecting the null of no relationship between returns in an event window and a covariate, considerable weight must be given to the possibility that the test would reject even absent the event, especially if the power of the test is relatively low.

An alternative strategy to significance testing that sidesteps this problem is to compare an event-window coefficient to the distribution of coefficient obtained from estimating the same cross-sectional regression in a series of non-event windows. This distribution approximates the sampling distribution of the coefficient under the null hypothesis. Statistical significance is then established by testing whether the event-window coefficient is "large" relative to the distribution of non-event window coefficients. We refer to this approach in general as "time-series OLS," since the approach typically involves estimating OLS regressions (e.g., Sefcik and Thompson, 1986). We show that the time-series OLS approach effectively eliminates excess Type I errors, subject to one important condition. Because the distribution of non-event window coefficients typically exhibits fat tails, parametric statistical tests such as a z-test still reject too often under the null – about 3% of the time at the 1% level. In contrast, using the empirical cumulative distribution function (CDF) of non-event window coefficients to compute a p-value for the event-window coefficient gets the Type I error rate approximately right.

A more serious issue with the time-series OLS approach is that it may have low power because it does not exploit information about cross correlations (Chandra and Balachandran, 1992). The lack of power is important, as most events that researchers study are likely to cause differences in returns across firms with different characteristics that are modest in size relative to the noise in returns, making these differences difficult to detect with low-powered tests. We confirm this lack of power for several covariates by introducing a known relationship between returns and a covariate and showing that time-series OLS often fails to detect this relationship. To address power concerns, we propose a new approach, which we term “time-series GLS,” that substitutes GLS for OLS regressions in both event windows and non-event windows. Implementing GLS requires an estimate of the covariance matrix of errors. Our approach uses principal component analysis (PCA) to encode the most important cross correlation patterns into the covariance matrix. The logic behind this approach is that we have no way of knowing *ex ante* the dimensions on which returns covary, so we extract as much information about cross correlations as we can from the data.[1]

We compare the efficiency of this time-series GLS approach to that of the time-series OLS approach by again introducing a known relationship between returns and a covariate and computing the frequency with which each approach detects the relationship. The boost in efficiency is large: Time-series GLS detects the induced relationship nearly twice as often as time-series OLS in most cases. The boost in efficiency is especially large when the event window is longer than one day, which is important since most papers implementing cross-sectional event studies of returns use windows longer than one day to allow for uncertainty about the timing with which information about the event is impounded into stock prices. While the time-series GLS approach we describe is more complex than traditional tests in corporate finance, we provide a turn-key Stata module that implements the approach (as well as the time-series OLS approach) in seconds in most cases.[2]

Finally, to gain more insight into the nature of return cross-correlations and why cross-sectional clustering is not more effective, we further evaluate the PCs of returns. Our analysis shows that cross correlation patterns are complex and not well-explained by membership in common cross-sectional groups such as industry or location, which explains why cross-sectional clustering strategies are not more effective. Specifically, we show that many PCs are required to summarize the cross-sectional variance in returns and that the economic factors captured by the first few PCs vary considerably over time, are often period-specific,

[1] Note that we are not trying to explain the cross-section of returns.
[2] This module is available at https://github.com/MalcolmWardlaw/csestudy).

and are difficult to map to traditional economic factors known to explain returns.

Our paper contributes to a growing 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 et al., 2004). Petersen (2009) shows that, in finance, the Fama-MacBeth procedure may be preferable to clustering in panel data when cross-sectional clustering is more important than time-series clustering. The time-series approaches we describe in this paper are close in spirit to the Fama-MacBeth approach. Abadie et al. (2023) consider more flexible approaches to modeling error structure and show that both robust and clustered standard errors can be severely *inflated*. In contrast, our evidence suggests that standard errors can be severely *deflated* due to cross correlated model errors, even when clustering standard errors cross-sectionally.

Our paper also contributes to the literature in finance on event study analysis of returns. The literature has explored the challenges that return cross-correlations create for inference in studies of abnormal *mean* event returns (Collins and Dent, 1984; Bernard, 1987; Lyon et al., 1999; Brav et al., 2000; Mitchell and Stafford, 2000; Jegadeesh and Karceski, 2009; Kolari and Pynnönen, 2010). However, we are aware of only two prior, older papers exploring the challenges they pose for inference in cross-sectional analysis of returns around an event.[3] Sefcik and Thompson (1986) describe these challenges and propose a time-series OLS approach as a solution. Chandra and Balachandran (1992) use simulated data to show that time-series approach using weighted least squares or assuming constant correlations within and across industries produce more efficient estimates than the time-series OLS approach. The recommendations of both papers have largely been ignored, perhaps in part because of the overhead associated with implementing the recommended approaches but also in part because of the near-universal adoption of clustering in empirical corporate finance to address concerns about cross-correlations.

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

In addition to reminding readers of the inference problem associated with using standard errors from cross-sectional return regressions around an event to conduct statistical testing, our paper makes several unique contributions. First, we provide the first full quantification of the severity of the cross-correlation problem in corporate finance and show that the problem can indeed be severe. Second, we demonstrate that cross-sectional clustering of standard errors may be largely ineffective at addressing the problem. Third, we quantify the power deficiencies of the time series OLS approach. Fourth, and most importantly, we propose a novel solution and show that our solution delivers substantial efficiency gains. Finally, we show that return cross-correlations are complex, time-varying, and difficult to summarize using observable factors. While we cannot extrapolate from our analysis to other settings, the fact that returns should reflect innovations to beliefs about a firm's fundamentals suggests that the problems we highlight may be severe in any setting where the outcome variable is related to firm fundamentals, which encompasses most empirical corporate finance.

1 Approaches

This section presents various approaches to estimating the cross-sectional effect of an event on stock returns. We begin by presenting a general framework describing cross-sectional event study analysis and some practical considerations.

1.1 Preliminaries

Consider an event E that induces a treatment effect in the market values of a set of firms. One object that may be of interest is the average treatment effect among the set of firms. The average return of firms in the event window represents an estimate of this average treatment effect. Another object that may be of interest – and the one that is the focus of this paper – is the heterogeneous treatment effect associated with the event across firms. A researcher may posit that the event in question affects the valuation of different firms differently. For example, a regulation may affect firms with different regulatory records differently, an election outcome may affect firms with different political connections differently, or a macroeconomic shock may affect firms with different financial structures differently.

Formally, let $\mathbf{X}$ denote an $N \times J$ matrix of firm characteristics with elements x_{ij}, where $i \in N$ indexes firms and $j \in J$ indexes characteristics. These characteristics include both variables of interest and any control variables. Let $\tau(\mathbf{x_i})$ represent the treatment effect associated with the event for firm i, where $\mathbf{x_i}$ is the $1 \times J$ vector of firm i's characteristics, with the first element equal to one. Typically, the researcher assumes that the treatment effect is linear in $\mathbf{x_i}$ – that is, that $\tau(\mathbf{x_i}) = \mathbf{x_i}\boldsymbol{\beta}$, where $\boldsymbol{\beta} = \{\beta_1, \beta_2, ..., \beta_J\}$ is a $J \times 1$ vector of coefficients. The objective of cross-sectional event study analysis is typically to estimate and conduct hypothesis testing on the elements of $\boldsymbol{\beta}$.

One important consideration in cross-sectional event study analysis is the choice of an "event window." This window represents the period of time over which market participants becomes aware of the event. In some cases, the market becomes aware of the event at a discrete, well-defined point in time. In these cases, the event window is typically a single day. In other cases, the exact time at which the market becomes aware of the event is less clear, and an event window longer than one day may be appropriate. Event windows of one to five days are common. Researchers sometimes also study longer event windows, with the idea that either the event itself unfolds slowly over time or the market needs time to fully digest the repercussions of the event. For example, studies of the cross section of returns in the early stages of the COVID-19 pandemic often focus on periods of approximately one month starting sometime in March 2020.

1.2 Cross-sectional OLS regressions

The standard approach to estimating the cross-sectional valuation effects of an event is to regress returns during the event window on firm characteristics using OLS. Let r_i denote the buy-and-hold return of firm i's stock during the event window. The standard cross-sectional event study regression equation is

$$r_i = \alpha + \mathbf{x_i}\boldsymbol{\beta} + \epsilon_i. \tag{1}$$

One important challenge in conducting hypothesis testing on the β_j coefficients is that errors in returns are likely to be both heteroskedastic and cross-sectionally correlated. These features generally make default standard errors from OLS estimation of (1) incorrect. Cross-sectional correlation in particular is likely to make standard errors too small, since firms that are similar on observable characteristics are likely exposed to similar economic forces in general and therefore to experience positive comovement in stock prices. Put differently, any correlation between returns and firm characteristics in the event window may simply reflect more general co-movement in the returns of firms with similar characteristics that would have occurred even absent the event.

To address these concerns, most cross-sectional event studies report either White (1980) adjusted standard errors, which account for heteroskedasticity, or industry-clustered standard errors, which account for both heteroskedasticity and cross-sectional correlation in errors within industry. However, White-adjusted standard errors do not account for cross-correlations at all, and it is unclear whether clustering at the industry level adequately accounts for cross correlations. More generally, clustering at any group level requires specifying a group structure *a priori*. Given the complexity of return correlations, it is unclear whether any pre-specified group-level clustering is adequate.

1.3 Time-series OLS

An alternative approach to hypothesis testing in cross-sectional event study analysis is to estimate (1) for both the event window and a set of pre-event windows using OLS and then use the distribution of pre-event window coefficients to test the statistical significance of a corresponding event-window coefficient.[4] We refer to this general approach as times-series OLS (TS-OLS for short).[5] This approach implicitly treats the non-event window coefficients as draws from a placebo data-generating process that is comparable to the event-window data-generating process but without any differential treatment effect associated with the event. A standard error based on the time-series of non-event coefficients maps neatly into the textbook definition of a standard error as "a measure of the statistical accuracy of an estimate, equal to the standard deviation of the theoretical distribution of a large population of such estimates."[6] Under the assumption that return cross correlations are time-invariant, this approach fully accounts for any cross-correlations by using a benchmark for hypothesis testing that also reflects the effects of cross correlations. Note that, for a binary characteristic, this approach is effectively a difference-in-differences approach.

Formally, let Z denote a set of pre-event windows, each of the same length as the event window E. Let r_{it} denote firm i's return in window t for each $t \in \{E, Z\}$. In addition, let $\mathbf{x_{it}}$ denote the $1 \times J$ vector of characteristics of firm i measured at the start of window t. Consider the following cross-sectional regression for a given window t:

$$r_{it} = \alpha_t + \mathbf{x_{it}}\boldsymbol{\beta}_t + \epsilon_{it} \tag{2}$$

Estimation of this regression for each $t \in \{E, Z\}$ yields a time series of estimated coefficient

[4]In principle, one could use a post-event window period instead. However, a pre-event window seems more natural if pre-event return data is available.

[5]Sefcik and Thompson (1986) show that this approach is tantamount to forming portfolios with weights determined by the distribution of the explanatory characteristics and then comparing portfolio returns in and out of the event window. This approach is therefore sometimes referred to as "Portfolio OLS."

[6]Source: https://doc.sitespect.com/knowledge/sitespect-statistics

vectors $\hat{\boldsymbol{\beta}}_{\mathbf{t}}$, where $\hat{\beta}_{jt}$ denotes the estimated coefficient j for day t. The time series of estimated coefficients $\hat{\beta}_{jt}$ for $t \in Z$ form a basis for testing the statistical significance of the estimated event-window coefficient $\hat{\beta}_{jE}$.

There are two possible null hypotheses that one might naturally test with respect to β_{jE}. One is that $\beta_{jE} = 0$. This is the hypothesis that cross-sectional regressions of the type described in Section 1.2 test. The other is that β_{jE} is equal to the mean of the non-event window coefficient. This hypothesis allows for the possibility that returns may be systematically correlated with firm characteristics. Given the evidence of such correlations from the asset-pricing literature, we focus on this second null hypothesis.

There are multiple sub-approaches to testing the significance of event-window coefficients using the non-event window time-series. The first is to estimate the second-stage regression

$$\hat{\beta}_{jt} = \delta_j + \gamma_j e_t + \nu_{jt} \tag{3}$$

separately for each characteristic x_j, where e_t is an indicator variable equal to 1 if $t = E$ and 0 if $t \in Z$. The coefficient $\hat{\gamma}_t$ represents an estimate of the difference between the first-stage event-window coefficient $\hat{\beta}_{jE}$ and the mean of the first-stage pre-event window coefficients. The standard error of $\hat{\gamma}_j$ can be used to test the statistical significance of this difference.

The second sub-approach is to compute a z-score for the difference between the event-window coefficient and non-event window coefficients as

$$z_j = \frac{\hat{\beta}_{jE} - \mu(\hat{\beta}_{jt,t\in Z})}{\sigma(\hat{\beta}_{jt,t\in Z})}, \tag{4}$$

where $\mu()$ and $\sigma()$ represents the empirical mean and standard deviation of a random variable, respectively. Both of the first two approaches impose assumptions on the distribution of coefficients that may not be satisfied in practice.

The third sub-approach is to compute a p-value based on the empirical cumulative distribution function (CDF) of the pre-event day coefficients. Formally, this p-value is

$$p_h = \frac{1}{T_Z} \sum_{t \in Z} I\{|\hat{\beta}_{jt} - \mu(\hat{\beta}_{jt,t \in Z})| > |\hat{\beta}_{jE} - \mu(\hat{\beta}_{jt,t \in Z})|\}, \tag{5}$$

where T_Z is the number of non-event windows and $I\{\}$ is an indicator function taking a value of one if the condition in brackets is true and zero if the condition is false. The advantage of the third approach over the first two is that it imposes no distributional assumptions on the time series of the cross-sectional coefficients.

One practical consideration in implementing this approach is the choice of a pre-event window period to use as a benchmark. The length of this window involves a tradeoff. A longer pre-event window allows a larger sample of pre-event window coefficient observations to use in inference, but it also increases the risk posed by time-varying return correlations. Instability of return correlations between the event window and pre-event window periods weakens the rationale for using the pre-event window period as a benchmark for hypothesis testing. Nevertheless, even if return correlations change over time, using the time-series of regression coefficients as a benchmark is almost certainly better than ignoring information about the distribution of return correlations contained in pre-event windows. In our analysis, we use a 252-trading day (approximately one-year) pre-event window period, which seems like a reasonable compromise in terms of period length.

Another practical consideration is when to measure the characteristics in $\mathbf{x_{it}}$ 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) window t. One option is to use the characteristics measured on the most recent available date prior to the earliest pre-event window. While this approach is simple, it may be inefficient if characteristics change over time. The other option is to measure the characteristics as of the most recent available date prior to window t, using more recent information where available for more recent windows.

For example, if characteristic j is an annual financial variable, then x_{ijt} would be the value of x_j for firm i as of the most recent fiscal year end prior to the start of window t. Given the potential efficiency gain, we adopt this latter approach in our analysis.

1.4 Time-series GLS

One disadvantage of TS-OLS is that it does not exploit information about cross correlations that might allow for more efficient estimates and hence greater statistical power. Statistical power is typically critical in testing the cross-sectional return effects of an event, as most events would be expected to produce modest cross-sectional differences in returns, making it difficult to distinguish differences in returns attributable to an event from noise. We propose a more powerful alternative to TS-OLS that we call time-series GLS (TS-GLS for short). Like TS-OLS, this approach uses the time series of cross-sectional coefficients from pre-event window regressions to conduct hypothesis testing. However, it uses GLS rather than OLS to estimate these cross-sectional regressions.

GLS achieves efficiency gains relative to OLS by using the inverse of the covariance matrix of the regression errors to weight observations. Because the diagonal elements of the covariance matrix measure the variance of the errors, this weighting addresses concerns about heteroskedasticity by down-weighting observations with high-variance errors. Because the off-diagonal elements measure the cross-sectional covariance among errors, this weighting also address concerns about correlated errors by down-weighting observations with correlated errors. Intuitively, the more correlated the errors of two observations, the less independent information they contain, making them less informative. More efficient cross-sectional estimates of both event-window and pre-event window coefficients should make tests based on TS-GLS more powerful than those based on TS-OLS.

Estimating every element of the return covariance matrix individually is generally infeasible, as doing so requires at least as many days of data as firms in the sample. Even if a long enough time-series of returns were available, attempting to estimate every element of the covariance matrix would likely result in overfitting, potentially making GLS less efficient than OLS. Instead, we propose capturing the most important quanta of return covariation using principal component analysis (PCA) and using the principal components (PCs) to construct an estimate of the covariance matrix.

Formally, consider an $N \times T$ matrix of firm-day returns $\mathbf{R}$ and the first $K \leq T$ PCs of $\mathbf{R}$. The first PC, $\mathbf{c_1}$, is the $N \times 1$ vector such that the projection of $\mathbf{R}$ onto $\mathbf{c_1}$ explains as much of the variation in $\mathbf{R}$ as possible. One can find $\mathbf{c_1}$ by minimizing the sum of the square of the distances from the elements of $\mathbf{R}$ to the projected points along $\mathbf{c_1}$. Alternatively and equivalently, one can find $\mathbf{c_1}$ by maximizing the variance of the projected points along $\mathbf{c_1}$. Note that $\mathbf{c_1}$ is the first eigenvector of $\mathbf{R}$.

Next, consider an $N \times T$ matrix $\mathbf{R_1}$ that contains the residuals from the projection of $\mathbf{R}$ onto $\mathbf{c_1}$ – that is, the orthogonalization of $\mathbf{R}$ to $\mathbf{c_1}$. The second PC, $\mathbf{c_2}$, is the $N \times 1$ vector chosen such that the projection of $\mathbf{R_1}$ onto $\mathbf{c_2}$ explains as much of the variation in $\mathbf{R_1}$ as possible. The vector $\mathbf{c_2}$ is the second eigenvector $\mathbf{R}$. Repeating this process K times results in K PCs (i.e., eigenvectors of $\mathbf{R}$), $\mathbf{c_1}, \mathbf{c_2}, ..., \mathbf{c_K}$. By construction, these PCs are orthogonal to each other. If $K = T$, then the PCs together will explain 100% of the variation in $\mathbf{R}$.

Now, consider a set of K portfolios consisting of the N stocks whose returns are given by $\mathbf{R}$, with the weights of portfolio $k \in K$ given by the elements of the vector $\mathbf{c_k}$. Since the K PCs are mutually orthogonal, we can treat these K portfolios as asset pricing factors. We can then use these factors and the loadings on them to construct a covariance matrix. Let the $N \times 1$ vector $\mathbf{r_t}$ denote the vector of returns for period t (the tth column of $\mathbf{R}$), and let

$f_{k,t} = \mathbf{c_k}'\mathbf{r_t}$. Assuming a linear factor structure, we can write

$$r_{it} = \phi_i + \sum_{k=1}^{K} \lambda_{ik} f_{kt} + \epsilon_{it}, \tag{6}$$

$$\mathrm{Cov}(\epsilon_{it}, \epsilon_{jt}) = \begin{cases} \sigma_i^2 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases}. \tag{7}$$

Given this structure, we can compute the estimated covariance matrix of returns as

$$\hat{\Omega}_{ii'} \equiv \mathrm{Cov}(r_{it}, r_{i't}) = \begin{cases} \sum_{k=1}^{K} \lambda_{ik}\lambda_{i'k}\mathrm{Var}(f_{it}) + \sigma_i^2 & \text{when } i = i' \\ \sum_{k=1}^{K} \lambda_{ik}\lambda_{i'k}\mathrm{Var}(f_{it}) & \text{when } i \neq i' \end{cases}. \tag{8}$$

Note that rather than specifying factors *ex ante*, as is typical in asset pricing, we use PCA to construct these factors. Our objective is not to explain the cross section of returns using economically meaningful factors but rather to estimate covariances as accurately as possible. It therefore better to allow the data to determine the factors rather than imposing them *ex ante*. As we will see in Section 3, the first constructed factor is effectively the market factor by construction, but the remaining constructed factors overlap little with other standard asset pricing factors, and the factors with which they overlap vary over time.

One practical consideration when implementing TS-GLS is the number of PCs to use in constructing the covariance matrix. The number of PCs can be any whole number between 0 and T. Using 0 PCs is equivalent to using WLS rather than GLS to estimate the cross-sectional return regressions and accounts for heteroskedasticity but not for 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 estimates using the TS-GLS approach, at least up to a point. However, beyond a certain point, adding more PCs results in over-fitting, which can reduce efficiency. As part of our analysis, we explore the optimal number of PCs to use.

2 Analysis of Different Approaches

In this section, we analyze the statistical properties of tests based on the approaches described in Section 1.

2.1 Data and sample

Our analysis involves regressing firm-level stock returns over short windows of time on firm characteristics. Our sample period is 1991-2021. We focus on this period because it is long yet relatively recent and because the data necessary to construct the firm characteristics we analyze is well-populated during this period. We begin by collecting daily firm-level stock returns from CRSP for the period 1990-2021. We use return data starting one year prior to the sample period because we require one year of return data prior to a given window of time when we implement the time-series approaches described in the previous section.

We analyze eight firm characteristics, which we compute based on data from CRSP and Compustat. Four of these are characteristics commonly studied in finance that capture elements of a firm's fundamentals. We compute $Log(size)$ daily as the natural log of market equity, which is the product of daily closing stock price and number of shares outstanding from CRSP. We compute B/M daily as the log of the ratio of book value (Compustat ceq), measured at the prior fiscal year end, to market equity. We compute $Profit$ as annual gross profit (Compustat GP) divided by total assets (Compustat AT). We compute $Invest$ as the annual growth rate of total assets (Compustat AT divided by prior-year AT minus 1). We obtain B/M, $Profit$, and $Invest$ from the open-source asset pricing project (Chen and Zimmermann, 2021).

The other four characteristics we analyze are variables used in recent cross-sectional event studies. We choose these four characteristics because they are easy to replicate using Compustat data. We measure all characteristics annually. We compute $Cash/AT$ as cash and short-term investments (Compustat CHE) divided by total assets. We compute $Debt/AT$ as the sum of long-term debt (Compustat $DLTT$) and debt in current liabilities (Compustat DLC), divided by total assets. Fahlenbrach, Rageth, and Stulz (2021) study return differences with these two variables around the arrival of the COVID-19 pandemic in 2020. We compute $TaxRate$ as 100 times income taxes paid (Compustat $TXDP$) divided by the sum of pre-tax income (Compustat PI) and special items (Compustat SPI), set to 0 if $PI < 0$. Wagner, Zeckhauser, and Ziegler (2018) study return differences with this variable around the resolution of the 2016 U.S. Presidential election. We define $NYHQ$ as an indicator variable equal to 1 if a firm is headquartered in New York (Compustat $STATE$ equal to "NY") and 0 otherwise. Acemoglu, Johnson, Kermani, Kwak, and Mitton (2016) study return differences with this variable around the announcement of Timothy Geithner as nominee for Treasury Secretary in November 2008.

We match the stock return data to the four common variables based on permno and to the four previously-analyzed variables using matched CRSP-Compustat data. The unit of observation is a firm-day. We associate with each firm-day observation the value of each characteristic as of the most recent available date prior to the day of the observation. The resulting sample consists of 35,559,001 firm-days belonging to 16,766 unique firms. Table 1 presents summary statistics for the sample.

[Table 1 about here]

2.2 Cross-sectional regressions

If the returns of firms with similar characteristics are correlated in general, then a test of the cross-sectional effect of a specific event on event-window returns may reject the null of no relationship too often relative to the intended size of the test. We assess the rate of excess rejection by estimating cross-sectional regressions for each window of a given length in our sample period and calculating the rejection rate for each characteristic at the 1% and 5% significance levels. For each of the four variables analyzed in prior studies, we exclude windows overlapping with the event window as defined in the study.

We compute rejection rates based on default standard errors, White-adjusted standard errors, and standard errors clustered at the Fama-French 49-category industry level. In addition, we compute rejection rates based on industry-clustered standard errors where we also control for Fama-French 49-category industry fixed effects since the combination of industry fixed effects and industry clustering is common in cross-sectional event-window return regressions. We begin by estimating univariate regressions of 1-day returns on each characteristic separately. Figure 1 presents the results.

[Figure 1 about here]

Rejection rates at the 1% and 5% significance level based on default standard errors average 33.0% and 43.4%, respectively, across the eight characteristics. Rejection rates are greater for characteristics more closely linked to fundamentals. They are highest for $Log(size)$ and B/M and lowest for the $NYHQ$ indicator. This conclusion is not surprising since day-to-day innovations in expected future cash flows are likely to exhibit more commonality among firms with similar fundamentals than among firms headquartered in the same state. These results suggest that rejection rates based on default standard errors are commonly an order of magnitude larger than the intended rejection rates under the null of no differential treatment effect associated with the event.

Non-event day rejection rates based on White-corrected standard errors are generally smaller than those based on default standard errors, and those based on standard errors clustered at the industry level smaller still. However, rejection rates based on clustered standard errors are still far higher than the intended rejection rates under the null, averaging 18.8% and 29.5% at the 1% and 5% significance levels, respectively. Adjusting standard errors to account for cross-sectional correlation in errors at the industry level does not appear to adequately account for return correlations. Including industry fixed effects in addition to clustering at the industry level has little impact on excess rejection rates, decreasing these rates in some cases but increasing them in others. Overall, the evidence in Figure 1 suggests that tests based on standard cross-sectional event study regressions cannot reliably separate out a hypothesized differential treatment effect of an event from ambient cross-sectional return correlations.

In this baseline case, we estimate separate univariate regressions for each characteristic and use 1-day event windows. In practice, researchers analyzing cross-sectional differences in returns around an event often estimate multivariate regressions and, in some cases, use multi-day event windows to account for uncertainty about the exact time at which the market learned about the event or slow market reaction to the event. We next present mean rejection rates across the eight characteristics for all four combinations of univariate and multivariate regressions and 1-day and 5-day event windows. Table 2 presents the results.

[Table 2 about here]

Rejection rates are higher for 5-day event windows than for 1-day event windows. Rejection rates based on default or White-adjusted standard errors are lower for multivariate regressions than for univariate regressions. However, those based on industry-clustered standard errors, either with or without industry fixed effects included, do not differ much between univariate and multivariate regressions. The smaller differences for industry-clustered standard errors are likely attributable to the fact that controlling for other observable characteristics accounts for at least some of the cross-correlation in returns associated with a given characteristic. Overall, rejection rates are considerably higher than the intended size of the tests in all cases, suggesting that cross-sectional regressions are highly unreliable in assessing differences in treatment effects associated with an event in general.

2.3 Time-series OLS

We begin our analysis of TS-OLS by computing rejection rates for these tests. We estimate OLS regressions of returns on characteristics separately for each day in the sample period. Then, for each day in 1991-2021, we conduct hypothesis testing by comparing the coefficient for that day to the distribution of the coefficients on the 252 trading days prior. We analyze each of the three specific sub-approaches to implementing these tests described in Section 1. Figure 2 presents the results.

[Figure 2 about here]

As anticipated, rejection rates using TS-OLS are much closer to the intended rejection rates under the null hypothesis than those based on standard cross-sectional regressions. However, rejection rates are still too high when we conduct hypothesis testing using the 2-stage regression or z-score approaches. The excess rejection rates of these tests arise from fat tails in the distribution of coefficients that these tests fail to take into account. In contrast, rejection rates based on p-values using the empirical CDF of pre-event window coefficients are only slightly higher than the intended rate. Because of the fat-tail issue, we recommended relying on p-values based on the empirical CDF for determining statistical significance when using the TS-OLS approach and consider only this specific sub-approach for the remainder of the analysis.

We next assess the power of TS-OLS with 1- and 5-day event windows. One at a time, for each window of the specified length in the period 1991-2021 and each characteristic, we add an artificial cross-sectional "effect" to returns of 25bp per one-standard deviation change in the characteristic, creating an artificial event window. We then estimate cross-sectional

OLS regressions for that window and for each window of the same size in a pre-event period consisting of the 252 days prior, where there is no added effect. Finally, we compute a p-value for the artificial event window coefficient based on the empirical CDF of the pre-event window coefficient time series and use that p-value to determine statistical significance at the 5% level. We also conduct the same exercise introducing a larger 50bp return effect per one-standard deviation change in the characteristic. Figure 3 presents detection rates based on these tests.

[Figure 3 about here]

Overall, detection rates are fairly low, suggesting that TS-OLS may lack the power to reliably detect differences in returns associated with an event. With a 1-day event window, the mean detection rate across the eight characteristics is 40.2% when the added return is 25bp per standard deviation change in a characteristic and 72.5% when the added return is 50bp. Detection rates are generally higher for characteristics where excess rejection rates from standard cross-sectional tests (Figure 1) are lower. Intuitively, stronger return correlations among firms that are similar on a characteristic add more noise to the time-series of pre-event day coefficients, which makes detecting a relationship on the artificial event day more difficult.

Detection rates are considerably lower with a 5-day event window than with a 1-day window, averaging only 15.0% for a 25bp effect and 32.5% for a 50bp effect. The degree to which detection rates decrease with the length of the event window is a serious concern, since papers often analyze multi-day event windows, especially when the exact timing of the event is difficult to determine. While the limited power of TS-OLS overall is a concern, this approach still represents a substantial upgrade over standard cross-sectional tests since it at least gets the size of the test approximately right.

2.4 Time-series GLS

We assess the size and power of TS-GLS the same way we do for TS-OLS. However, before doing so, we first need to specify the number of PCs that we will use to estimate the return covariance matrix. The optimal number of PCs, which maximizes efficiency, is unclear *a priori*. We use two approaches to gain insight into the optimal number of PCs to use. The first approach involves analyzing 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 for returns Ω based on PCA, as specified in equation (8).[7] Given an estimated covariance matrix for returns on t, $\hat{\Omega}_t$, the minimum variance portfolio's weights $w_{mvp,t}$ are specified by

$$w_{mvp,t} = \frac{\hat{\Omega}_t^{-1}\mathbf{1}}{\mathbf{1}'\hat{\Omega}_t^{-1}\mathbf{1}}, \tag{9}$$

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

The relation between K and the volatility of the minimum variance portfolio is informative about the incremental information content of each additional PC for forecasting the inverse of the next-day covariance matrix – exactly the object we use for GLS. This approach also has the advantage of not being specific to any firm characteristic. Figure 4 plots this relationship. The variance of the minimum variance portfolio decreases sharply with the addition of first several PCs. The variance flattens out around 50 PCs and is largely invariant until it begins sharply increasing around $K = 245$. As K approaches 252, the maximum available given our choice of a $T = 252$-day window for daily returns in the pre-period, the PCA approach over-fits the covariance matrix and produces a worse out-of-sample forecast for the true covariance matrix.

[7] Clarke et al. (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.

[Figure 4 about here]

The second approach involves analyzing the relationship between the ratio of daily GLS and OLS coefficients and the number of PCs. This approach is informative about the degree to which GLS reduces estimation noise relative to OLS by increasing estimation precision. Specifically, for each characteristic, we estimate cross-sectional OLS and GLS regressions for each day in the sample period, using the 252 trading days prior to a given day to construct the covariance matrix we use in the GLS regressions. We then compute the ratio of the time-series standard deviation of GLS coefficients to the time-series standard deviation of OLS coefficients. We repeat this exercise varying the number of PCs we use to construct the covariance matrix between 1 and 250. For each characteristic, Figure 5 plots the relationship between the standard deviation ratios and the number of PCs used.

[Figure 5 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 few PCs and is essentially flat between 50 and 250 principal components. Based on the results in Figures 4 and 5 and in the spirit of choosing a round number, we use 100 PCs in the remainder of the analysis and recommend this as the default number of PCs. However, the results are virtually unchanged if we use any number of PCs between 50 and 240.

To compare the statistical power of the TS-GLS and TS-OLS approaches, we repeat the tests where we introduce a 25bp or 50bp return effect per one standard deviation change in a characteristic, using GLS instead of OLS to estimate cross-sectional regressions. Table 3 presents a comparison of the mean detection rates across the eight characteristics at a 5% significance level using TS-GLS and TS-OLS.

[Table 3 about here]

TS-GLS performs much better than TS-OLS at detecting the introduced effect. For a 1-day event window and a 25bp effect, the mean detection rate is 70.4% using TS-GLS, compared to 40.2% for the OLS approach. The improvement in performance appears to be approximately proportionate to the TS-OLS rejection rate, though naturally the improvement is constrained when TS-OLS detection rates are already high. The benefits of using TS-GLS rather than TS-OLS then appear to be greatest when the TS-OLS detection rates are in an intermediate range. Overall, it appears that explicitly accounting for cross-sectional return correlations increases the statistical power of tests based on the time series of coefficients substantially.

3 Meaning of Principal Components

In this section, we analyze the information about drivers of return cross correlations embedded in the PCs. We begin by plotting the fraction of total variation in return that the first 1, 5, 25, and 50 PCs explain by year. Figure 6 presents these plots.

[Figure 6 about here]

Two observations are worth making. First, the first few PCs explain a relatively small fraction of returns. For example, in most years, the first five 5 PCs explain less than 30% of the variation in returns, while the first 25 PCs explain less than 50% of the variation. Second, the fraction of total return variation that the first few PCs explain varies considerably over time. It is higher in years in which large market-moving shocks occurred. For example, the first few PCs explain a larger fraction of the return variation during the financial crisis (2008–09) and in the aftermath of the onset of the COVID-19 pandemic (2020–2021).

PCs of stock returns can be interpreted as portfolio weights and used to construct factor portfolios. We next examine the relationship between PC-based factor portfolio returns, calculated by implementing PCA on a balanced panel of daily individual stock returns in each calendar year, and the returns on eight pre-specified factors in the years 2008, 2013, 2020, and 2021. We choose 2008, 2020, and 2021 because these were years in which major market-moving events occurred. We choose 2013 to provide a relatively quiescent year for comparison.

Four of the pre-specified factor portfolios are based on well-known factors from the asset pricing literature. These are the equity market portfolio ($MktRf$), small-minus-big portfolio (SMB), high-minus-low portfolio (HML), and up-minus-down portfolio (UMD). The other four factor portfolios are constructed as long-only equal-weighted combinations of stocks or portfolios 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; 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 were available to trade on each day.

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 PCs and each of the pre-specified factors. Table 4 presents the results. Panels A, B, C, and D present the results for 2008, 2013, 2020, and 2021, respectively.

[Table 4 about here]

The first PC-weighted portfolio return is highly correlated with the equity market portfolio in all four years. By construction, the first PC-weighted portfolio is approximately the equal-weighted market portfolio. The correlation is less than one because the equity market portfolio return is value-weighted. Because of this distinction, the first PC-weighted portfolio return is also correlated with the SMB factor portfolio return.

For 2008, the second and fourth 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 third PC-weighted portfolio return is highly correlated with the HML and SMB factor portfolio returns, while the fourth 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 fourth is highly correlated with the HML, $Finance$, and $Tech$ factor portfolio returns. For 2013, a relatively quiescent year, none of the second through fifth PC-weighted portfolios are strongly correlated with any of the factor portfolio returns.

One conclusion from this analysis is that the factors driving cross-sectional correlations in returns vary substantially from year to year and often represent factors unique to a period. Another is that it is often difficult to determine what drives cross sectional correlations in any given period. These conclusions both further suggest that specifying dimensions of return correlation *a priori* – for example, by clustering on a dimension like industry – is likely to do a poor job of accounting for important sources of cross-sectional correlation in returns.

4 Conclusions

Our results suggest that standard cross-sectional regressions of returns around an event on firm characteristics reject the null hypothesis far too frequently given the intended size of the test to be reliable for conducting hypothesis testing and that common adjustments to standard errors are inadequate in addressing this problem. A time-series approach based on cross-sectional OLS regressions addresses the problem with excess rejection rates but provides limited power to detect cross-sectional differences in event-window returns, especially over longer windows. A times-series approach based on cross-sectional GLS regressions, using principal component analysis to encode the most important sources of cross-sectional correlation into the covariance matrix, appears to offer substantially more statistical power. While this approach is more complex, we plan to provide a Stata module that will implement both this approach and the time-series approach based on cross-sectional OLS regressions automatically.

Our first set of results could indicate broader problems with clustering standard errors in empirical corporate finance. Corporate finance researchers rely heavily on clustering to address concerns about correlated regression errors. It is difficult in general to assess the effectiveness of clustering in accounting for correlations in errors. Because we observe returns at a high frequency and they are largely serially uncorrelated, we are able to assess the effectiveness of clustering in cross-sectional event studies. Our results suggest that clustering is ineffective in accounting for correlated errors because return correlations are too complex for pre-specified clusters to capture. While we can only speculate, it seems likely that the same issue would arise with any regressions where the dependent variable is connected to firm fundamentals.

References

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

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.

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

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

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.

Chandra, R. and B. V. Balachandran (1992). More powerful portfolio approaches to regressing abnormal returns on firm-specific variables for cross-sectional studies. *Journal of Finance 47*(5), 2055–2070.

Chen, A. Y. and T. Zimmermann (2021). Open source cross-sectional asset pricing. *Critical Finance Review, Forthcoming*.

Clarke, R. G., H. De Silva, and S. Thorley (2006). Minimum-variance portfolios in the us equity market. *The 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*, 48–84.

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.

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

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.

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.

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*, 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: journal of the Econometric Society*, 817–838.

Figure 1: Cross-Sectional OLS Rejection Rates

This figure depicts rejection rates from univariate cross-sectional OLS regressions of 1-day returns on each of eight characteristics. The sample period is the 7,811 trading days from 1991–2021. For the four characteristics from papers studying the cross section of returns around an event (*Cash*/*AT*, *Debt*/*AT*, *TaxRate*, and *NYHQ*), days in the event window analyzed by the paper are excluded. The figure shows rejection rates based on default standard errors, White-adjusted standard errors, Fama-French 49-category industry clustered standard errors, and industry clustered standard errors from regressions where we also include industry fixed effects. The top panel shows rejection rates at the 1% level, while the bottom panel shows rejection rates at the 5% level.

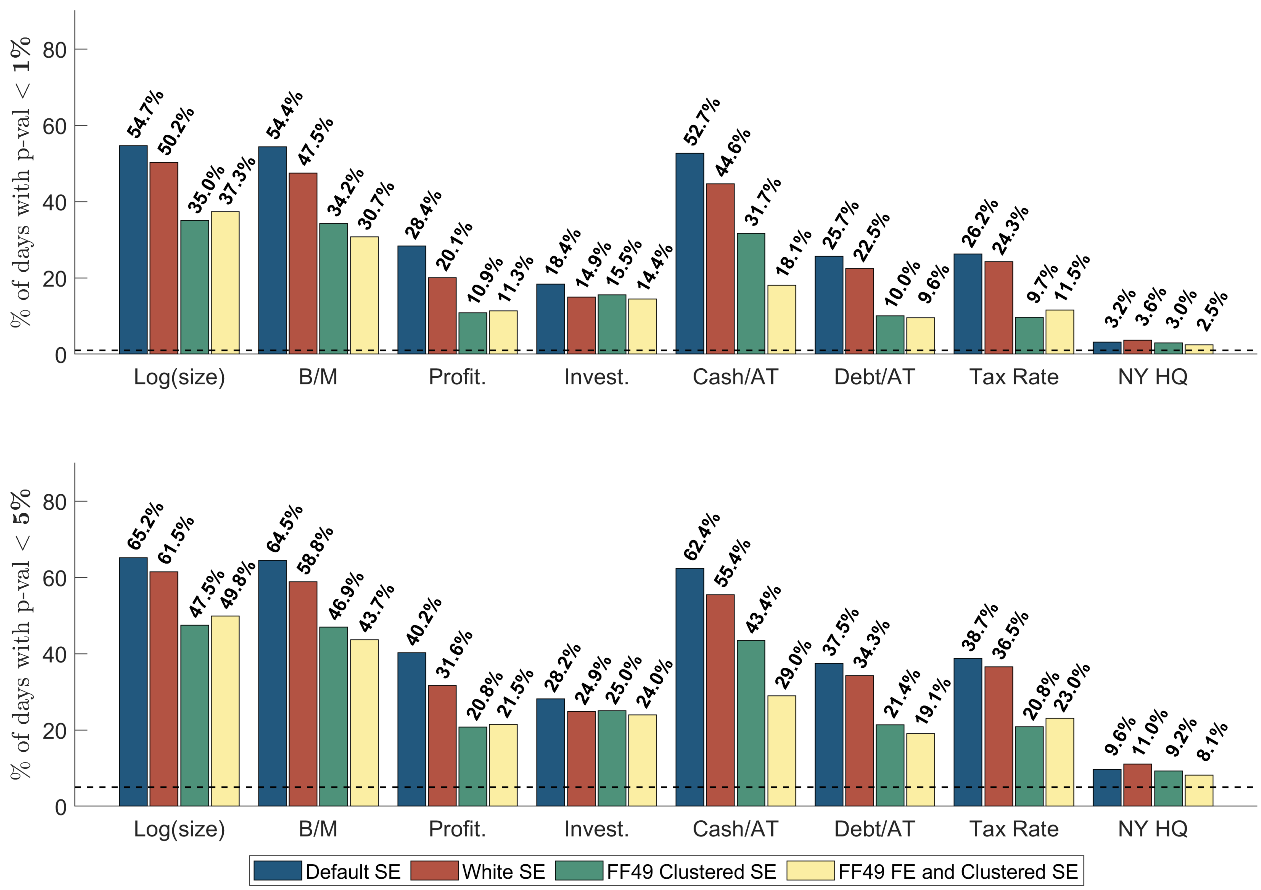

Figure 2: Time-Series OLS Rejection Rates

This figure depicts rejection rates from time-series OLS regressions of 1-day returns on each of eight characteristics. The sample period is the 7,811 trading days from 1991–2021. For the four characteristics from papers studying the cross section of returns around an event ($Cash/AT$, $Debt/AT$, $TaxRate$, and $NYHQ$), days in the event window analyzed by the paper are excluded. The figures shows rejection rates based on the three specific approaches described in Section 1.3. The top panel shows rejection rates at the 1% level, while the bottom panel shows rejection rates at the 5% level.

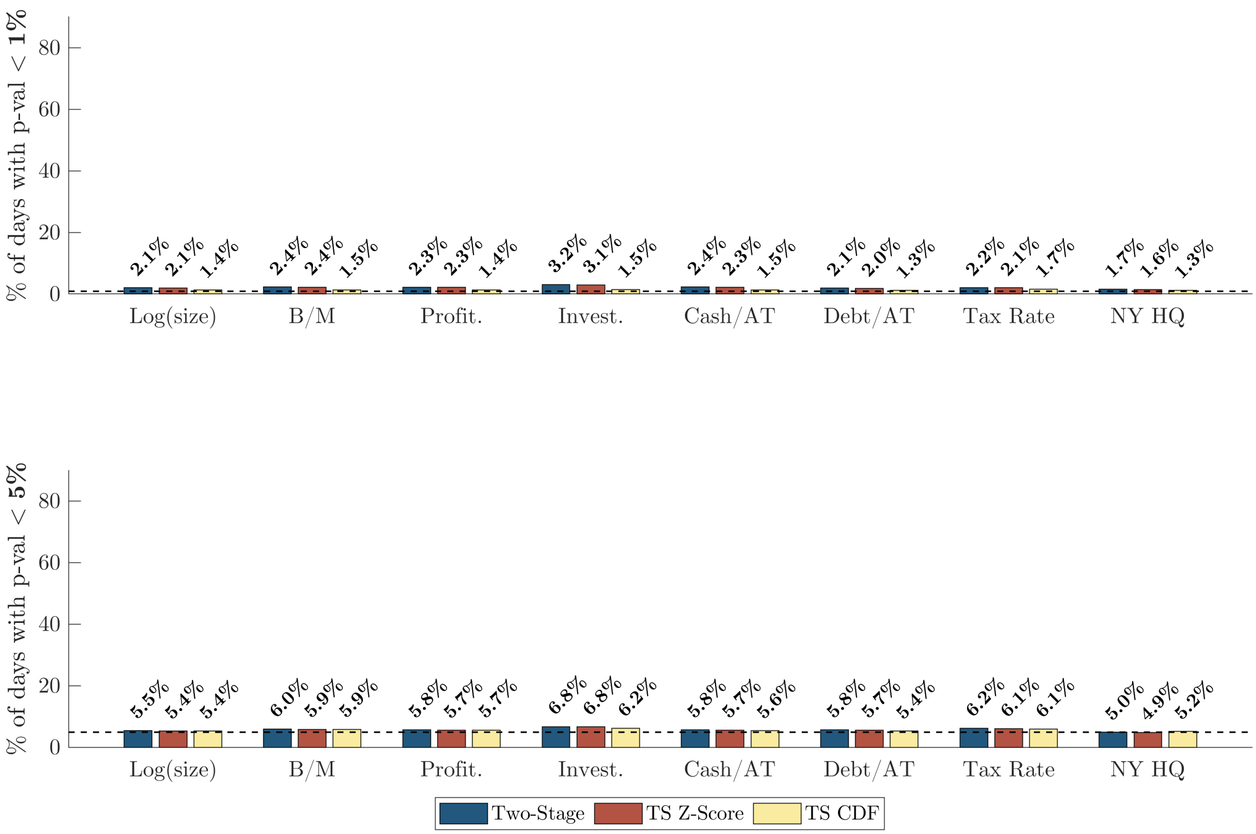

Figure 3: Time-Series OLS Detection Rates with Added Effects

This figure depicts detection rates at the 5% significance level from time-series OLS regressions of returns on each of eight characteristics, with an effect of either 25bp or 50bp added to returns for each one standard deviation increase in the given characteristic in each artificial event window. The sample period is the 7,811 trading days from 1991–2021. For the four characteristics from papers studying the cross section of returns around an event ($Cash/AT$, $Debt/AT$, $TaxRate$, and $NYHQ$), days in the event window analyzed by the paper are excluded. The top panel shows rejection rates when the added effect is 25bp, while the bottom panel shows rejection rates when the added effect is 50bp.

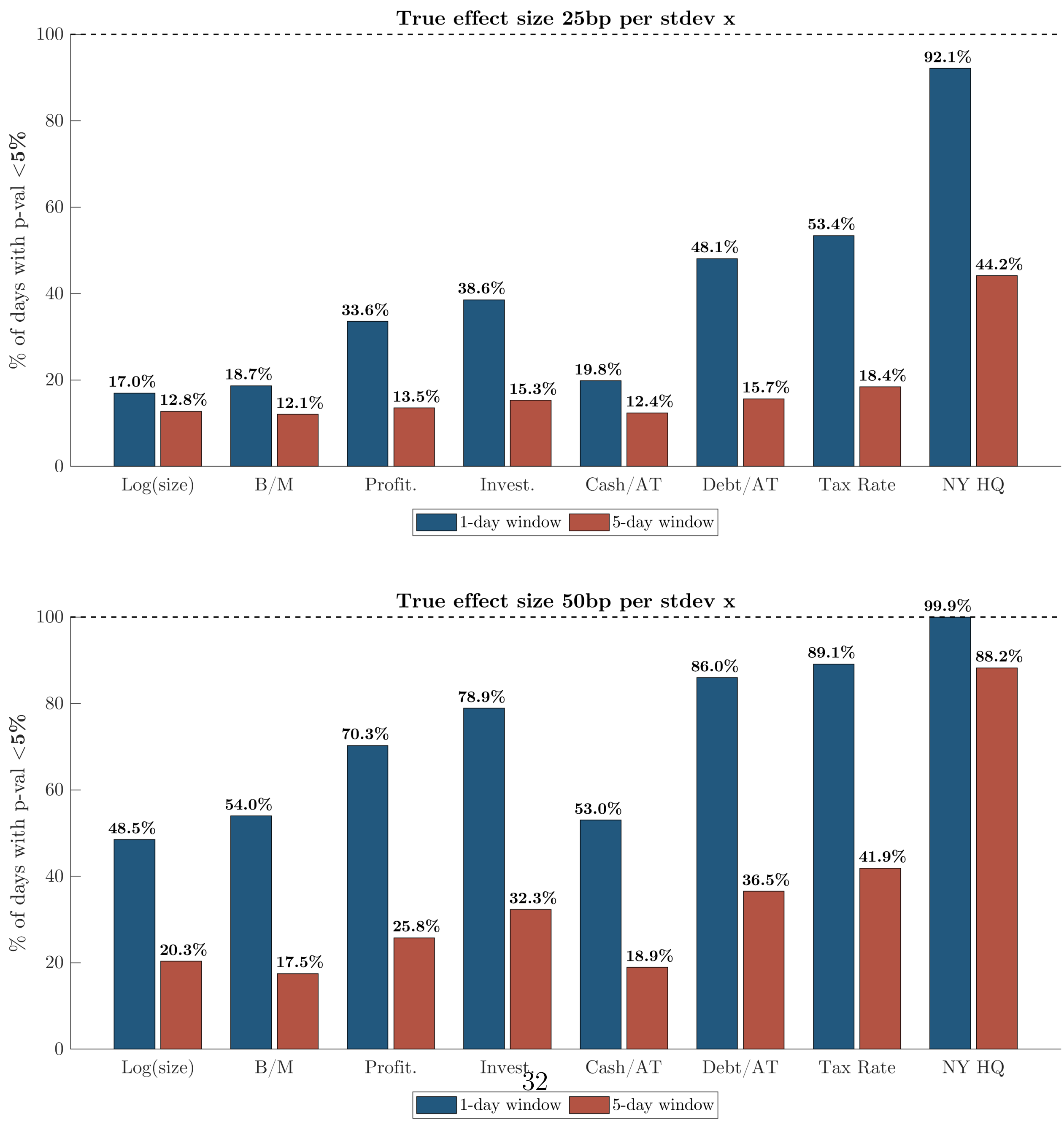

Figure 4: Minimum Variance Portfolio Variances and # of Principal Components

This figure depicts the standard deviation of realized returns for a minimum-variance portfolio of US equities as a function of the number of principal components used to form an out-of-sample forecast for the covariance matrix of returns (K). Our covariance matrix forecasts are constructed using the first K PCs from implementing PCA on balanced panel of daily returns over the prior 252 trading days, assuming that other than via these PCs each stock's return is uncorrelated. The sample period is the 7,811 trading days from 1991–2021.

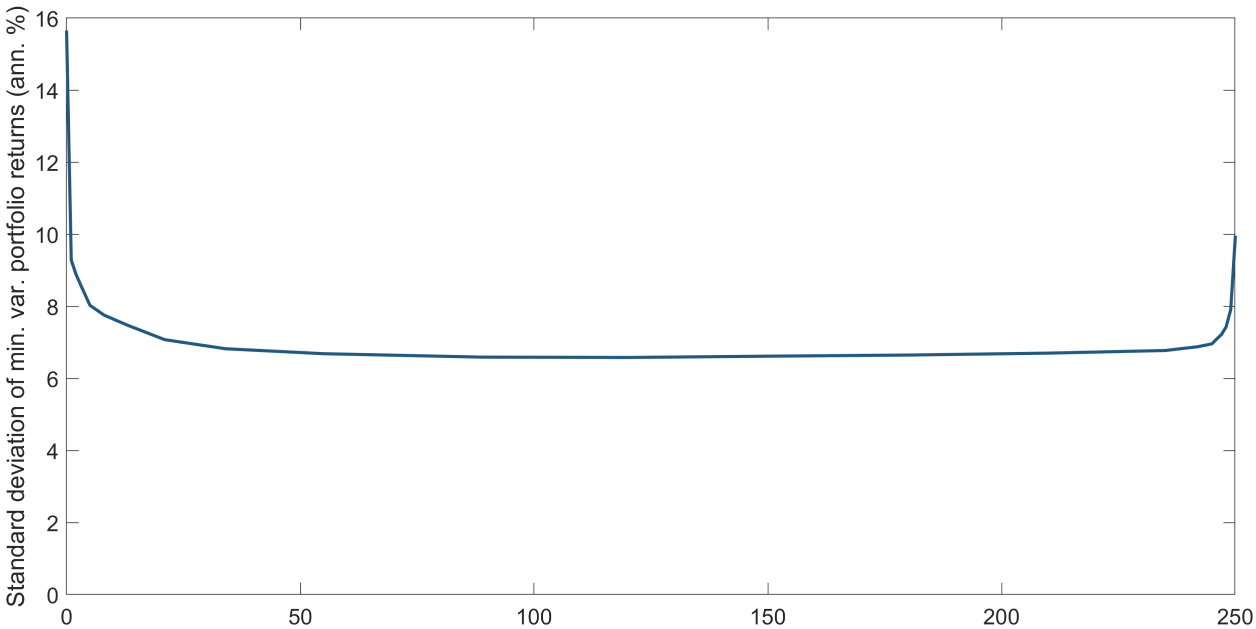

Figure 5: TS-GLS/TS-OLS Coefficient Standard Deviations and # of Principal Components

This figure plots the ratio of the time-series standard deviation of cross-sectional GLS coefficients to the time-series standard deviation of cross-sectional OLS coefficients from regressions of 1-day returns on each of the characteristics against the number of PCs we use in the time-series GLS regressions. The sample period is the 7,811 trading days from 1991–2021. For the four characteristics from papers studying the cross section of returns around an event ($Cash/AT$, $Debt/AT$, $TaxRate$, and $NYHQ$), days in the event window analyzed by the paper are excluded.

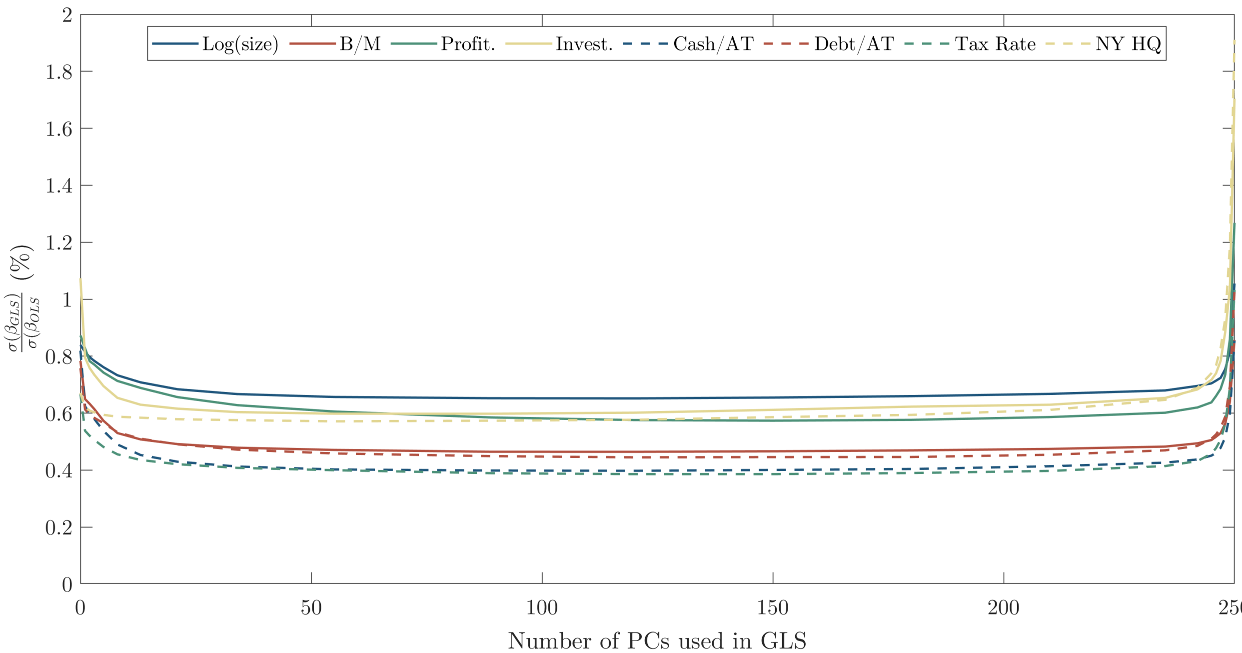

Figure 6: Explanatory Power of Principal Components

This figure plots the mean percent of the daily cross-sectional variance of returns that a given number of principal components explains by calendar year, for 1, 5, 25, and 100 principal components. The sample period is the 7,811 trading days from 1991–2021.

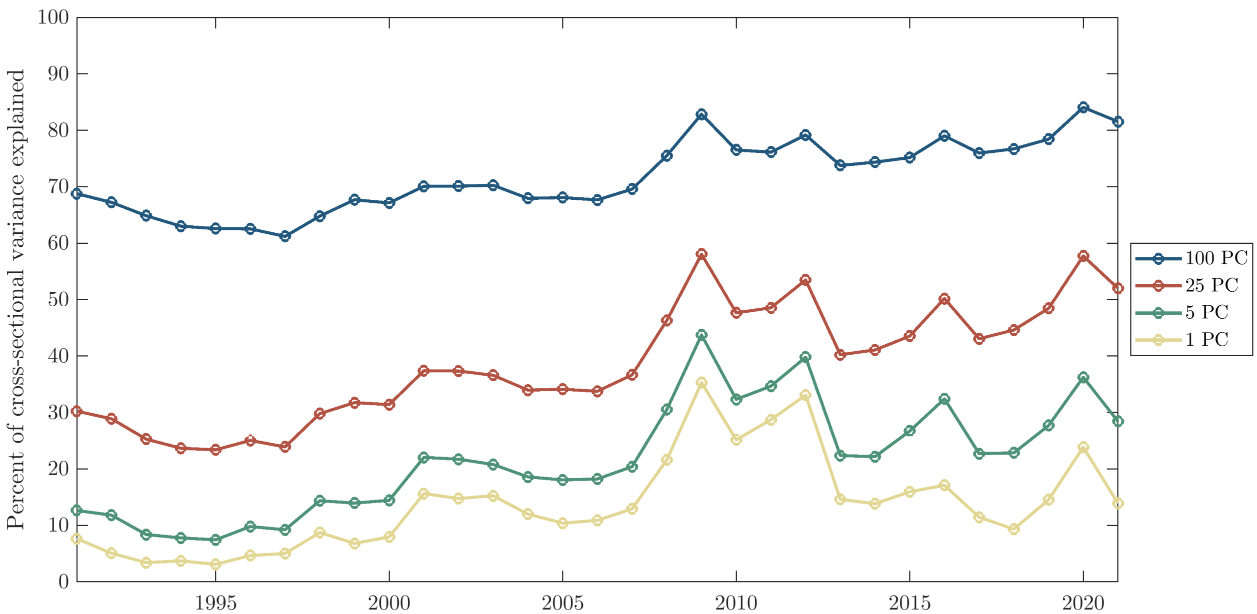

Table 1: Summary Statistics

This table presents summary statistics for the sample of firm-day observations we use in our analysis. $Log(size)$ is the natural log of market equity, which is the product of daily closing stock price and number of shares outstanding from CRSP. B/M is the log of the ratio of book value (Compustat CEQ), measured at the prior fiscal year end, to market equity. $Profit$ is gross profit (Compustat GP) divided by total assets (Compustat AT). $Invest$ is the ratio of capital expenditures (Compustat $CAPEX$) to total assets (Compustat AT). $Cash/AT$ 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. $TaxRate$ is 100 times income taxes paid (Compustat $TXDP$) divided by the sum of pre-tax income (Compustat PI) and special items (Compustat SPI), set to 0 if $PI < 0$. $NYHQ$ is an indicator variable equal to 1 if a firm is headquartered in New York (Compustat $STATE$ equal to "NY") and 0 otherwise.

Variable	Firm-Days	Unique Firms	Firms/Day	Mean	Median	σ	Within-day σ
Return (%)	35,559,001	16,766	4,410	0.09	0.00	4.92	4.41
Log(size)	34,092,227	15,506	4,365	19.37	19.25	2.21	2.09
B/M	32,799,415	15,249	4,199	-0.76	-0.68	1.03	1.01
Profitability	27,676,880	12,588	3,543	0.33	0.31	0.37	0.37
Investment	31,396,391	14,339	4,020	0.19	0.06	1.96	1.19
Cash/AT	34,081,926	15,496	4,363	0.18	0.09	0.22	0.22
Debt/AT	33,933,548	15,478	4,344	0.22	0.17	0.21	0.21
Tax Rate	33,510,699	15,484	4,290	16.61	7.62	21.23	20.82
NY HQ	34,092,227	15,506	4,365	0.08	0.00	0.27	0.28

Table 2: Cross-Sectional OLS Regression Rejection Rates

This table presents mean rejection rates at the 1% and 5% statistical significance levels from cross-sectional OLS regressions of returns on each of eight firm characteristics. It shows rejection rates based on default standard errors, White-adjusted standard errors, Fama-French 48-category industry clustered standard errors, and industry clustered standard errors where industry fixed effects are included in the regression for 1-day and 5-day return windows and for univariate and multivariate regressions (where all eight characteristics are included as explanatory variables). The sample period is the 7,811 trading days from 1991–2021.

Mean 1% rejection rate

Window	Regression	Default SE	White SE	Clustered SE	Clust SE + FE
1 day	Univariate	33.0	28.5	18.8	16.9
1 day	Multivariate	22.7	18.8	16.0	14.6
5 days	Univariate	31.9	27.6	18.2	16.4
5 days	Multivariate	22.0	18.1	15.5	14.2

Mean 5% rejection rate

Window	Regression	Default SE	White SE	Clustered SE	Clust SE + FE
1 day	Univariate	43.4	39.4	29.5	27.3
1 day	Multivariate	33.3	29.3	26.5	24.3
5 days	Univariate	41.9	38.0	28.5	26.4
5 days	Multivariate	32.2	28.4	25.7	23.5

Table 3: Detection Rates for Added Cross-Sectional Effects: TS-OLS vs TS-GLS

This table presents mean detection rates at the 5% statistical significance levels from univariate TS-OLS and TS-GLS regressions with artificially added cross-sectional event-window return effects across eight firm characteristics (explanatory variables). Return effect sizes are 25bp and 50bp per standard deviation change in a characteristic. For each artificial event window, the non-event period is the 252 trading days prior to the start of that window, and we use the 252 trading days prior to the event window to construct the principal component-based covariance matrix that we use in estimating GLS regressions. The table shows detection rates for TS-OLS and TS-GLS in percents, the difference in these rates, and the ratio of these rates. The sample period is the 7,811 trading days from 1991–2021.

Window	Effect (bp)	TS-OLS	TS-GLS	Diff	Ratio
1 day	25	40.2	70.4	30.2	175%
1 day	50	72.5	94.2	21.7	130%
5 days	25	15.0	29.0	14.0	193%
5 days	50	32.5	59.6	27.1	183%

Table 4: Evolving Correlations With Principal Components

This table presents the correlations between factor portfolios constructed from the the first five principal components of individual stock returns in a calendar year and a variety of other factors. constructed from individual stock returns. The first four are the market ($MktRf$), size (SMB), value (HML) and momentum (UMD) factors, as collected from Ken French's data library. We also compute correlations with four 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 are available to trade on each day. Panel A presents results for 2008, Panel B for 2013, Panel C for 2020, and Panel D for 2021.

Panel A: 2008

	'Mkt'	'Crisis'	?	?	?
PC	1	2	3	4	5
% x-sectional var. explained	19.7%	2.8%	2.4%	2.0%	1.9%
$\lvert\rho(PC_i, MktRf)\rvert$	96%	7%	7%	7%	1%
$\lvert\rho(PC_i, SMB)\rvert$	24%	34%	24%	4%	4%
$\lvert\rho(PC_i, HML)\rvert$	8%	50%	4%	50%	6%
$\lvert\rho(PC_i, UMD)\rvert$	11%	38%	15%	58%	8%
$\lvert\rho(PC_i, Tech)\rvert$	2%	5%	4%	18%	8%
$\lvert\rho(PC_i, Finance)\rvert$	13%	56%	14%	47%	3%
$\lvert\rho(PC_i, Covid)\rvert$	6%	14%	13%	18%	1%
$\lvert\rho(PC_i, Memes)\rvert$	2%	2%	5%	9%	6%
R^2	98%	46%	9%	39%	3%

Panel B: 2013

	'Mkt'	?	?	?	?
PC	1	2	3	4	5
% x-sectional var. explained	8.2%	3.9%	2.3%	1.9%	1.4%
$\lvert\rho(PC_i, MktRf)\rvert$	93%	8%	1%	2%	2%
$\lvert\rho(PC_i, SMB)\rvert$	33%	0%	11%	7%	7%
$\lvert\rho(PC_i, HML)\rvert$	3%	9%	4%	7%	16%
$\lvert\rho(PC_i, UMD)\rvert$	0%	9%	7%	3%	6%
$\lvert\rho(PC_i, Tech)\rvert$	5%	3%	0%	5%	12%
$\lvert\rho(PC_i, Finance)\rvert$	6%	8%	5%	1%	5%
$\lvert\rho(PC_i, Covid)\rvert$	1%	7%	6%	5%	1%
$\lvert\rho(PC_i, Memes)\rvert$	1%	2%	2%	6%	1%
R^2	98%	4%	3%	2%	7%

Panel C: 2020

	'Mkt'	?	'Value'	'Covid'	?
PC	1	2	3	4	5
% x-sectional var. explained	20.5%	4.0%	2.6%	2.4%	2.1%
$\lvert\rho(PC_i, MktRf)\rvert$	88%	6%	21%	25%	0%
$\lvert\rho(PC_i, SMB)\rvert$	36%	25%	2%	31%	8%
$\lvert\rho(PC_i, HML)\rvert$	34%	18%	79%	4%	8%
$\lvert\rho(PC_i, UMD)\rvert$	31%	23%	73%	1%	18%
$\lvert\rho(PC_i, Tech)\rvert$	25%	13%	44%	1%	1%
$\lvert\rho(PC_i, Finance)\rvert$	34%	12%	68%	20%	4%
$\lvert\rho(PC_i, Covid)\rvert$	19%	5%	2%	46%	18%
$\lvert\rho(PC_i, Memes)\rvert$	23%	7%	19%	37%	9%
R^2	98%	13%	77%	41%	11%

Panel D: 2021

	'Mkt'	?	'Memes'	'Value'	?
PC	1	2	3	4	5
% x-sectional var. explained	12.2%	6.8%	3.7%	3.5%	2.4%
$\lvert\rho(PC_i, MktRf)\rvert$	74%	6%	17%	23%	11%
$\lvert\rho(PC_i, SMB)\rvert$	53%	28%	27%	7%	6%
$\lvert\rho(PC_i, HML)\rvert$	1%	1%	6%	84%	19%
$\lvert\rho(PC_i, UMD)\rvert$	24%	8%	13%	26%	3%
$\lvert\rho(PC_i, Tech)\rvert$	14%	22%	14%	45%	6%
$\lvert\rho(PC_i, Finance)\rvert$	19%	8%	3%	58%	13%
$\lvert\rho(PC_i, Covid)\rvert$	16%	3%	11%	3%	14%
$\lvert\rho(PC_i, Memes)\rvert$	15%	45%	81%	10%	1%
R^2	86%	26%	70%	78%	7%