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

January 2024

Abstract

Correlated exposure to confounding events causes frequent Type I errors in cross-sectional event studies: estimated relationships between returns and typical covariates are significant at the 1% level on more than 20% of all days. Cross-sectional event-study regression standard errors would need to be doubled or tripled to produce correct Type I error rates. Clustering standard errors helps but is inadequate. Testing significance by comparing event-day relationships to a distribution of pre-event day relationships achieves correct Type I error rates. We introduce a novel GLS-based variant of this approach and show that it improves power over OLS approaches.

*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), Sheridan Titman, 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.

The field of corporate finance has a rich tradition of exploring the link between firm-specific stock returns and various firm attributes in the context of significant events like regulatory changes or political developments.[1] This cross-sectional event study methodology holds considerable promise for validating economic theories, and its findings frequently inform policy debates and legal discussions. Nevertheless, there is a persistent concern about its propensity to produce Type I errors. This issue arises because firms sharing similar traits often also share exposure to other, unrelated events, leading to potentially misleading associations between event-period returns and firm characteristics. Although this problem is understood theoretically (Sefcik and Thompson, 1986), its actual significance in practical scenarios remains unknown. This uncertainty complicates evaluation of cross-sectional event study methodology's dependability.

In this study, we leverage daily stock return data at the firm level, spanning from 1991 to 2021, to evaluate the frequency of Type I errors in cross-sectional event studies. Our methodology involves examining the significance of the connection between stock returns and a variety of firm characteristics – some of which are subjects in existing cross-sectional event studies – on a day-to-day basis. This technique enables us to estimate Type I error rates under the assumption that the particular event under investigation does not influence the relationship between returns and a given characteristic. We propose that this null hypothesis, rather than a null of no cross-sectional relationship on the event day, is the actual target of research inquiries. Indeed, it is the possibility of statistically significant cross-sectional relationships on many days that potentially confounds significance testing in cross-sectional event studies.

Our evidence suggests that the cross-correlation problem is severe: For typical covariates, coefficients are statistically significant at the 1% level on more than 20% of the days in our sample period. The problem is especially acute for covariates closely related to firm

[1] The first such study we are aware of is Leftwich (1981).

fundamentals (e.g., firm size and book-to-market ratio), where significance rates at the 1% level can exceed 50%. Our estimates suggest that inflating event-period return regression standard errors 2 to 3 times would be necessary to restore the Type I error rate to the intended size of the test. While clustering standard errors by industry, location, or other dimension can help, tests based on clustered standard errors still typically result in significance at the 1% level on more than 10% of days, with rates exceeding 30% for covariates closely related to firm fundamentals. Overall, clustering standard errors appears to be an inadequate solution to the cross-correlation problem in cross-sectional event studies.

Motivated by concerns about excess Type I errors, Sefcik and Thompson (1986) proposes testing the statistical significance of the relationship between returns and a characteristic on the event day by comparing the absolute magnitude of that relationship to the magnitude of the same relationship on a series of days prior to the event ("pre-event days").[2] The logic behind this approach is that we should only confidently attribute a relationship on the event day to the event itself if the relationship on that day is exceptional in the sense that its magnitude is larger than the magnitude of the same relationship on most other days. If it is not, then it is hard to rule out the possibility that the observed event-day relationship would have occurred by chance, even absent the event. This approach, which we refer to as "time-series OLS," sidesteps concerns about correlated exposure to other events because these exposures should be reflected in the distribution of observed non-event day relationships as well. Yet, adoption of this alternative approach to statistical testing remains limited. We find in an informal survey of papers employing cross-sectional event studies involving a single event appearing in top-3 finance journals in the last ten years that only 13% use this approach.[3]

[2]Sefcik and Thompson (1986) shows that this test can be implemented using a portfolio approach. We abstract away from this specific implementation for the sake of providing a more intuitive explanation of the approach.

[3]Of the 87% that use conventional regression standard errors to conduct statistical tests, 35% report unadjusted standard errors; 25% report robust standard errors, which address concerns about heteroskedas-

As with tests based on cross-sectional regression standard errors, we estimate the frequency of Type I errors using a time-series OLS approach by computing the fraction of days with statistically significant relationships between 1991 and 2021. We show that this frequency is approximately correct, with one note of caution. The distributions of relationships between daily returns and most characteristics are generally fat-tailed, causing a t test to find statistically significant relationships slightly too often (2-4% at the 1% significance level). In contrast, a p-value approach akin to bootstrapping, based on comparison to the empirical cumulative density function (CDF) of pre-event day relationships, ensures approximately the correct frequency of statistically significant relationships.

While time-series OLS addresses concerns about excess Type I errors, there are concerns about its statistical power (e.g., Chandra and Balachandran, 1992). We assess the power of this approach by adding relationships between returns and firm characteristics on artificial event dates and then determining how frequently time-series OLS detects these relationships at a statistically significant level. We find some support for concerns about power. Time-series OLS detects an added return of 50 basis points per one standard deviation change in a characteristic at the 5% significance level more than 75% of the time for many characteristics. However, detection rates drop to close to 50% for characteristics such as size and book-to-market ratio that are closely linked to firm fundamentals. Intuitively, these relationships are harder to detect for the same reason that they are subject to frequent Type I errors in tests based on cross-sectional regression standard errors: realized returns are meaningfully related to these characteristics on many days, making it difficult to tease out a relationship associated with any specific event.

One might argue that we should continue to rely on cross-sectional regression standard errors to conduct statistical testing at least in some cases given concerns about the power of

ticity but not cross correlations; 35% report standard errors clustered by industry; and 5% report standard errors clustered by geographic location.

the time-series OLS approach. However, this logic is misleading. Using cross-sectional regression standard errors may result in higher rates of statistical significance. However, it will do so largely by producing Type I errors. The abundance of Type I errors makes it impossible to determine how much information a statistically significant coefficient actually conveys about the differential effect of an event on firms with different characteristics. Nevertheless, concerns about power are important, as many event-induced relationships between returns and characteristics may be modest in magnitude. With that concern in mind, we introduce a new approach, which we call "time-series GLS," that substitutes feasible generalized least squares (GLS) for OLS to estimate both event-day and non-event day relationships.

GLS uses the inverse of an estimated covariance matrix of regression errors to weight observations, potentially increasing power by downweighting pairs of observations whose errors are more correlated and which therefore provide less independent information about the relationship in question. Our approach uses principal component analysis (PCA) of daily returns to encode information about return correlations into the estimated covariance matrix. In asset pricing parlance, we model the covariance between individual stock returns as arising from common exposures to latent factors, which we extract from return data using PCA, following Giglio and Xiu (2021) and Lopez-Lira and Roussanov (2023).[4] This approach allows us to capture information about return correlations along many dimensions without needing to specify those dimensions *a priori*. More precise estimates of both event-day and pre-event day relationships should make it easier to detect a relationship specific to the event.

We show that, like time-series OLS, time-series GLS produces approximately the correct rate of Type I errors. However, time-series GLS outperforms time-series OLS in terms of power by a wide margin in many cases, especially when the effect would otherwise be difficult to detect. For example, it detects a relatively small relationship of 25 basis points of

[4]Richer machine learning approaches, for example those in Kelly et al. (2019) or Lettau and Pelger (2020), are needed to identify which factors are "priced" in the sense that they explain differences in expected, rather than realized, returns. However, our objective is not to identify priced factors.

additional return per one standard deviation change in a characteristic at the 5% significance level 1.2-2.9 times as often as time-series OLS. The additional power may make it feasible to test for differences in returns associated with an event that would otherwise be impossible to test with any reliability. To make implementing the time-series approaches easier, we provide a turnkey Stata module (`csestudy`) that implements both.[5]

Finally, we shed light on the nature of return cross correlations and why clustering is not more effective at correcting standard errors by examining the principal components (PCs) that we use in our time-series GLS implementation. Our analysis shows that cross-correlation patterns reflected in the PCs are complex and not well-captured by membership in common cross-sectional groups such as industry or location, which explains why cross-sectional clustering strategies are not more effective. Like Lopez-Lira and Roussanov (2023), we find that many PCs are required to summarize the cross-sectional variance in returns and 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.

The high frequency and low serial correlation of returns allows us to estimate excess Type I error rates in cross-sectional event studies of returns. Researchers also often conduct cross-sectional event studies of longer-term outcome variables such as capital investment or profitability, frequently using difference-in-differences methodology. While this methodology accounts for differences in means between treatment and control groups before the event, it does not account for a lack of within-group independence due to common exposure to other events. While we cannot readily estimate Type I error rates for these outcomes, we have every reason to believe that correlated exposures to confounding events are just as much a problem with longer-term outcomes as with returns. This possibility points to an advantage of analyzing returns rather than longer run outcomes around an event: approaches

[5] https://github.com/MalcolmWardlaw/csestudy

that reliably account for correlated exposures to other events are feasible when the outcome variable is returns.

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 et al., 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 time-series approach we describe is close in spirit to the Fama-MacBeth approach. Abadie et al. (2023) considers flexible approaches to modeling error structure and shows that clustering at too high a level can result in severely *inflated* standard errors and hence low statistical power. 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 on event study analysis of returns. The literature has primarily 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). We are not aware of any prior analysis of the practical challenge caused by return cross correlations for cross-sectional analysis of returns around an event.[6] Sefcik and Thompson (1986) describes this challenge conceptually but do not assess its practical importance in real-world applications. The time-series approach that they propose has largely been ignored. Our analysis suggests that care must be taken in implementing time-series approaches to avoid excess rejection due to fat tails in the distribution of return-characteristic relationships. In addition, we propose a more efficient GLS estimator and resolve the practical issues in implementing such approaches by providing a turnkey Stata

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

module.

1. Conceptual Framework and Econometric 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.

Suppose that stock returns are a function of firm characteristics and random noise, and that the relation between returns and firm characteristics varies period-by-period depending on the news that arrives. An example firm characteristic is a firm's market capitalization. Even though the average relationship between daily returns and market capitalization is near-zero, news arrives on some days that meaningfully and statistically increases the value of large cap firms relative to small cap firms, leading to a positive cross-sectional relation between size and returns on those days. On other days, the news leads to the opposite relation.

We formalize this data generating process as follows:

$$r_{i,t} = \lambda' x_{i,t-1} + \delta_{i,t}, \tag{1}$$

where $x_{i,t-1}$ is a $K \times 1$ vector of firm characteristics observed prior to t that includes a constant, λ is a $K \times 1$ vector containing each characteristic's relation with average returns, and the residuals satisfy:

$$\delta_{i,t} = \left(n_t + \frac{e_\tau}{N_\tau} \mathbb{1}(t \in \tau) \right)' x_{i,t-1} + \nu_{i,t}, \tag{2}$$

$$\mathbb{E}(n_t) = \mathbb{E}(\nu_{i,t}) = \mathbb{E}(x_{i,t-1}\nu_{i,t}) = 0, \tag{3}$$

$$\mathrm{Var}(\nu_{i,t}) = \Omega_\nu. \tag{4}$$

The vector n_t contains mean-zero random variables expressing the news arriving in period t in terms of its impact on returns as a function of firm characteristics. The total incremental effect of $x_{i,t-1}$ on $r_{i,t}$ due to the event occurring during "event window" τ, our coefficients of interest, are in the vector e_τ. N_τ is the number of days in τ.

This framework features two distinct sources of correlations in unexpected returns $\delta_{i,t}$. The first is that firms with similar characteristics $x_{i,t-1}$ have correlated returns due to common exposure to news arrival. The second is that the part of returns orthogonal to $x_{i,t-1}$, $\nu_{i,t}$, may still have some cross-correlations due to omitted firm characteristics or a latent factor structure. These latter correlations are reflected in the off-diagonal elements of Ω_ν.

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 of a few days may be appropriate. Researchers also sometimes 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.

1.1. Cross-sectional OLS estimates

The standard methodology to estimate e_τ and conduct hypothesis tests, which we refer to as **cross-sectional OLS**, uses an OLS regression of event period returns on a vector of firm characteristics $x_{i,\tau-1}$, including a constant, variables of interest, and controls:

$$\sum_{t\in\tau} r_{i,t} = b'_\tau x_{i,\tau-1} + \epsilon_{i,\tau}, \tag{5}$$

where $x_{i,\tau-1}$ are characteristic values observed prior to the beginning of the event window τ.[7] In this subsection, we show that while any bias in $\hat{b}_\tau$ as an estimate of e_τ (the object of interest) is likely to be tiny, conventional standard errors for $\hat{b}_\tau$ may significantly understate the estimation error in e_τ.

The conceptual challenge in conducting hypothesis testing on the $\hat{b}_\tau$ coefficients is that regression errors $\epsilon_{i,\tau}$ in returns are likely to be both heteroskedastic and cross-sectionally correlated. These features generally make default standard errors from OLS estimation of Eq. (5) 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, many recent 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*, which requires knowing the important dimensions on which returns cluster. Given the complexity of return correlations, it is unclear whether any pre-specified group-level clustering is adequate.

We formalize these conceptual issues in the context of the above data generating process. Substituting Eq. (2) into Eq. (1) and assuming characteristics $x_{i,t}$ are constant for each i

[7] The outcome variable is also often the average or cumulative buy-and-hold return in the event window τ, both of which face same issues we identify here.

during the event window, we have:

$$r_{i,t} = \left(\lambda + n_t + \frac{e_\tau}{N_\tau}\right)' x_{i,t-1} + \nu_{i,t} \ \forall\, t \in \tau, \tag{6}$$

$$\Rightarrow \sum_{t\in\tau} r_{i,t} = \left(N_\tau \lambda + \sum_{t\in\tau} n_t + e_\tau\right)' x_{i,\tau-1} + \sum_{t\in\tau} \nu_{i,t}, \tag{7}$$

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

$$b_\tau = e_\tau + N_\tau \lambda + \sum_{t\in\tau} n_t, \tag{8}$$

$$\epsilon_{i,\tau} = \sum_{t\in\tau} \nu_{i,T}. \tag{9}$$

Proposition 1 (Bias and excess variance of the cross-sectional OLS approach)**.** *Cross-sectional OLS estimates of b_T in Equation (5) have the following relations to the true event-specific effect e_T:*

$$\mathbb{E}(\hat{b}_\tau - e_\tau) = \underbrace{N_\tau \lambda}_{bias}, \tag{10}$$

$$\begin{aligned} Var(\hat{b}_\tau - e_\tau) &= N_\tau \left[Var(\hat{b}_\tau - b_\tau) + N_\tau\, Var(n_t) \right] \\ &= N_\tau \left[\underbrace{(X'X)^{-1}(X'\Omega_\nu X)(X'X)^{-1}}_{\textit{Correlation-corrected OLS variance}} + \underbrace{Var(n_T)}_{\textit{News variance}} \right], \end{aligned} \tag{11}$$

where X is an $N \times K$ matrix stacking all the $x'_{i,\tau-1}$ vectors. The $\mathbb{E}(\cdot)$ and $Var(\cdot)$ operators condition on $x_{i,\tau-1}$ and draw random $\nu_{i,t}$ and n_t for all $t \in \tau$.

Proofs are in Appendix A.

In words, Proposition 1 shows that using cross-sectional OLS estimates of b_τ to measure e_τ may result in bias and is likely to understate sampling error. Any relationship between the firm characteristics and average returns ($\lambda \neq 0$) will bias b_τ relative to e_τ. In practice, this effect is likely to be extremely small for daily event windows ($N_\tau = 1$) because differences in average daily returns as a function of firm characteristics are dwarfed by return volatility. For larger event windows, especially those greater than a month, this bias could have a larger impact.

The more important result in practical settings is that the estimation error in $\hat{b}_\tau$ relative to e_τ – the key to statistical inference – is likely to be inflated for two reasons. First, residuals in each day $t \in \tau$ ($\nu_{i,t}$) may have cross-correlations due to news unrelated to $x_{i,\tau-1}$. This manifests in the Ω_ν term in Eq. (11), and can in principle be corrected by using cluster-robust standard errors. In practice, however, these clusters may miss important cross-cluster correlations. Second, the news that would have arrived without the event, $\sum_{t\in\tau} n_t$, drives a wedge between the true b_τ and e_τ that would remain even in an infinite sample without estimation error. This effect always increases $\mathrm{Var}(\hat{b}_\tau - e_\tau)$ and cannot be corrected using any standard methodology such as clustering that focuses on estimating the correct Ω_ν. Like the bias in $\hat{b}_\tau$, the variance increases in N_τ, meaning longer event windows result in less-precise estimates of e_τ, making it only possible to reliably detect economically large effects.

1.2. Time-series OLS estimates

An alternative approach to hypothesis testing in cross-sectional event study analysis, pioneered by Sefcik and Thompson (1986), is to use the cross-sectional regressions in non-event periods to approximate the distribution of $\hat{b}$ under the null hypothesis. We refer to this approach as **time-series OLS** (TS-OLS for short).[8] The detailed procedure is:

1. Estimate single-day cross-sectional regressions:

$$r_{i,t} = b_t' x_{i,t-1} + \epsilon_{i,t}, \tag{12}$$

[8]Sefcik and Thompson (1986) shows 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."

for a sample of all days in the event window and L days spanning both the event window τ and a set of pre-event windows τ_{-1}, τ_{-2}, ..., τ_{-L}, each of the same length as the event window.[9]

2. Estimate the event-specific effect as:

$$\hat{e}_{\tau}^{\text{TSOLS}} = \hat{b}_{\tau} - \frac{1}{L}\sum_{l=1}^{L}\hat{b}_{\tau_{-l}}, \tag{13}$$

$$\hat{b}_{\tau} = \sum_{t\in\tau} b_t, \quad \hat{b}_{\tau_{-l}} = \sum_{t\in\tau_{-l}} b_t \tag{14}$$

3. Calculate the standard error and p-value for $\hat{e}_{\tau}^{\text{TSOLS}}$ using the distribution of estimated $\hat{b}_{\tau_{-l}}$ across pre-event windows.

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.”[10] 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 also that this approach is similar to the Fama and MacBeth (1973) methodology in asset pricing, and, for a binary characteristic, is effectively a difference-in-differences approach.

To formalize the intuition for TS-OLS, note that outside the event window the return

[9] In principle, one could use a post-event window period instead. However, a pre-event window is less likely to include any post-announcement drift or follow-on events that muddy the coefficients interpretation as unrelated to the event in question.

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

generating equation (1) simplifies to:

$$r_{i,t} = (\lambda + n_t)' x_{i,t-1} + \nu_{i,t}. \tag{15}$$

If we estimate the regression in Eq. (5) in a non-event window τ_{-l}, we therefore have:

$$b_{\tau_{-l}} = N_\tau \lambda + \sum_{t \in \tau_{-l}} n_t, \tag{16}$$

$$\Rightarrow \mathbb{E}(\hat{b}_{\tau_{-l}}) = N_\tau \lambda, \tag{17}$$

$$\mathrm{Var}(\hat{b}_{\tau_{-l}}) = N_\tau \left[\mathrm{Var}(\hat{b}_{\tau_{-l}} - b_{\tau_{-l}}) + N_\tau \mathrm{Var}(n_t) \right]. \tag{18}$$

This leads to Proposition 2, which specifies the TS-OLS approach and shows the distribution of estimates based on it in the population.

Proposition 2 (Unbiasedness and correct variance of the time-series OLS approach)**.** *Assuming stationary distributions for estimation error $\hat{b}_{\tau_{-l}} - b_{\tau_{-l}}$ and news n_t, $\hat{e}_\tau^{TSOLS}$ has the following relations to the true event-specific effect e_τ:*

$$\mathbb{E}(\hat{e}_\tau^{TSOLS} - e_\tau) = \underbrace{0}_{\textit{No bias}}, \tag{19}$$

$$Var(\hat{e}_\tau^{TSOLS} - e_\tau) = \underbrace{Var(\hat{b}_{\tau_{-l}})}_{\textit{Time-series variance}}, \tag{20}$$

$$\Phi(\hat{e}_\tau^{TSOLS} - e_\tau) = \underbrace{\Phi(\hat{b}_{\tau_{-l}})}_{\textit{Time-series CDF}}, \tag{21}$$

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

There are two sub-approaches to testing the significance of event-window coefficients using the non-event window time-series. The first leverages Eq. (20) and computes a t-statistic for element j of coefficient vector b as follows:

$$\hat{t}_j \equiv \frac{\hat{e}^{\mathrm{TSOLS}}_{j,\tau}}{\sqrt{\mathrm{Var}(\hat{e}^{\mathrm{TSOLS}}_{j,\tau} - e_{j,\tau})}} = \frac{\hat{b}_\tau - \frac{1}{L}\sum_{l=1}^{L} \hat{b}_{\tau_{-l}}}{\sqrt{\mathrm{Var}(\hat{b}_{\tau_{-l}})}}. \tag{22}$$

One can then compare $\hat{t}_j$ to critical values from the standard normal, Student's T, or other cumulative distribution function. This approach is identical to estimating a second-stage regression of the time-series of coefficients $\hat{b}_{j,\tau_{-L}}, \ldots, \hat{b}_{j,\tau_{-1}}, \hat{b}_{j,\tau}$ on a constant and an indicator for the true event window. The coefficient on the indicator equals $\hat{b}_\tau - \frac{1}{L}\sum_{l=1}^{L} \hat{b}_{\tau_{-l}}$ while the standard error equals $\sqrt{\mathrm{Var}(\hat{b}_{\tau_{-l}})}$ when a small-sample correction is applied to the variance. $\hat{t}_j$ is therefore the standard t-statistic for this second-stage regression.

A second sub-approach 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 (21). Formally, this p-value is

$$\hat{p}_j = \frac{1}{L}\left[\sum_{l=1}^{L} \mathbb{1}\left(\left|\hat{b}_{j,\tau_{-l}} - \mu(\hat{b}_j)\right| > \left|\hat{b}_{j,\tau} - \mu(\hat{b}_j)\right|\right)\right], \tag{23}$$

$$\mu(\hat{b}_j) \equiv \frac{1}{L}\sum_{l=1}^{L} \hat{b}_{j,\tau_{-l}}. \tag{24}$$

The advantage of this approach over the first is that it imposes no distributional assumptions on the time series of the cross-sectional coefficients. The disadvantage is that it produces only p-values and not standard errors, though standard errors can be inferred from the p-values with the addition of a distributional assumption.

Another small disadvantage of this approach is that $\hat{p}_j$ only has the correct test size under the null, in the sense that $\mathbb{P}(\hat{p}_j \leq p^*) = p^*$, when $L+1$ is divisible by $\frac{1}{p^*}$. Under the null, the expected rejection rate is

$$\mathbb{P}(\hat{p}_j \leq p^*\%) = \frac{\lceil p^* \cdot (1+L) \rceil}{(1+L)}, \tag{25}$$

where $\lceil \cdot \rceil$ is the ceiling function that rounds up to the nearest integer.[11] Because the critical values we care about in practice include 1%, we need L to be one less than a multiple of 100.

One practical consideration in implementing this approach is the length of the pre-event window period to use as a benchmark. There is a tradeoff here. A longer pre-event window affords 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 199-trading day (approximately 10 months) pre-event window period, which seems like a reasonable compromise in terms of period length and satisfies Eq. (25) for $p^*\% = 1\%$.

Another practical consideration is when to measure the characteristics in x_τ 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 τ. 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 the beginning of the event window and each pre-event window, using more recent

[11] Suppose $L = 252$ and the cross-sectional coefficients $\hat{b}$ are i.i.d. in both the pre-event and event window. In this case, $\hat{p}_j \leq 1\%$ only when the event-window effect is in the top three of the 253 estimated coefficients. However, $\mathbb{P}(\hat{p}_j \leq 1\%) = \frac{3}{253} > 1\%$, and thus the test will not have the correct size due to the discreteness of the placebo periods.

information where available for more recent windows. For example, if characteristic j is an annual financial variable, then $x_{ij\tau}$ would be the value of x_j for firm i as of the most recent fiscal year end prior to the start of window τ. Given the potential efficiency gain, we adopt this latter approach in our analysis.

1.3. 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 critical in testing the cross-sectional return effects of an event, as many events would be expected to produce only moderately large 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 addresses 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.

In our setting, it is important that the GLS approach account for return comovement orthogonal to x (Ω_ν in the specification above) and comovement due to latent news arrival causing day-to-day variation in the relation between x and returns (n_t in the specification above). We do so by using the covariance matrix of demeaned returns in implementing GLS.

Estimating every element of the return covariance matrix individually is generally infeasible, as doing so with fewer days of data than firms in the sample produces a rank-deficit matrix that cannot be used for OLS. 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 follow Giglio and Xiu (2021) and capture the important quanta of return covariation using principal component analysis (PCA), using the principal components (PCs) to construct an estimate of the covariance matrix.[12]

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

$$r_{it} = \phi_i + \sum_{k=1}^{K} \lambda_{ik} f_{kt} + \epsilon_{it}, \tag{26}$$

$$\mathrm{Cov}(\epsilon_{it}, \epsilon_{jt}) = \begin{cases} \sigma_i^2 & \text{when } i = j \\ 0 & \text{when } i \neq j \end{cases}. \tag{27}$$

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

$$\hat{\Omega}_{r,ij} \equiv \mathrm{Cov}(r_{it}, r_{jt}) = \sum_{k=1}^{K} \lambda_{ik}\lambda_{jk}\mathrm{Var}(f_{kt}) + 1(i = j)\sigma_i^2 \tag{28}$$

Note that rather than specifying factors *ex ante*, as is common 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

[12]While most readers will be familiar with PCA, we provide a simple primer on PCA in Appendix B. See Appendix C for an explanation of why we use PCA instead of a machine-learning based technique.

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.

Putting the pieces together, the steps for conducting time-series GLS are:

1. Estimate Ω_r, the covariance matrix of de-meaned returns, as specified in Equation (28) with latent factors from PCA on a sample of past daily returns demeaned by firm.

2. Estimate single-day cross-sectional using GLS regressions:

$$r_{i,t} = b_t' x_{i,t-1} + \epsilon_{i,t}, \tag{29}$$

$$\hat{b}_t^{\text{GLS}} = (X_{t-1}' \Omega_r X_{t-1})^{-1} X_{t-1}' \Omega R_t, \tag{30}$$

   where X_{t-1} is a matrix containing all $x_{i,t-1}$ and R_t contains all $r_{i,t}$, for a sample of all days in the event window and L days spanning both the event window τ and a set of pre-event windows τ_{-1}, τ_{-2}, …, τ_{-L}, each of the same length as the event window.

3. Estimate the event-specific effect as:

$$\hat{e}_\tau^{\text{TSGLS}} = \hat{b}_\tau^{\text{GLS}} - \frac{1}{L} \sum_{l=1}^{L} \hat{b}_{\tau_{-l}}^{\text{GLS}}, \tag{31}$$

$$\hat{b}_\tau^{\text{GLS}} = \sum_{t \in \tau} \hat{b}_t^{\text{GLS}}, \quad \hat{b}_{\tau_{-l}}^{\text{GLS}} = \sum_{t \in \tau_{-l}} b_t^{\text{GLS}} \tag{32}$$

4. Calculate the standard error and p-value for $\hat{e}_\tau^{\text{TSGLS}}$ using the distribution of estimated $\hat{b}_{\tau_{-l}}^{\text{GLS}}$ across pre-event windows.

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.

To gain insight into the optimal number of PCs to use, 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 for returns Ω based on PCA, as specified in equation (28).[13] Given an estimated covariance matrix for returns on t, $\hat{\Omega}_t$, the minimum variance portfolio's weights $w_{mvp,t}$ are specified by

$$w_{mvp,t} = \frac{\hat{\Omega}_t^{-1}\mathbf{1}}{\mathbf{1}'\hat{\Omega}_t^{-1}\mathbf{1}}, \tag{33}$$

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. Figure 1 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 $K = 199$.

[Figure 1 about here]

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 195. At the end of Section 2,

[13] 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.

we assess the sensitivity of the precision gain from GLS relative to OLS to the number of PCs.

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.[14] We use return data starting in 1990 even though the sample period starts in 1991 because we require 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. The first 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

[14]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).

prior-year AT minus 1).

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) studies 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) studies 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) studies 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 21,060,248 firm-days belonging to 11,549 unique firms. Table 1 presents summary statistics for the sample.

[Table 1 about here]

2.2. Cross-sectional regressions

We estimate the frequency of Type I errors in standard cross-sectional event study regressions for each of the eight characteristics by estimating regressions of returns on the characteristic for all 1-day and 5-day windows in our sample period and computing the fraction of these windows in which the relationship is statistically significant at the 1% and 5% levels.[15] Researchers sometimes use event windows longer than one day in cross-sectional event studies of returns because of uncertainty about the timing with which the market learns and processes news of the event. We include 5-day windows in our analysis to understand the implications of using a longer event window. For each of the four variables analyzed in prior studies, we exclude windows overlapping with the event window as defined in the study, though the effects of this exclusion are immaterial.

We compute the fraction of relationships that are statistically significant based on default standard errors, White-adjusted standard errors, and standard errors clustered at the Fama-French 49-category industry level – standard errors commonly reported in practice. In addition, we compute this fraction 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 sometimes used in cross-sectional event-window return regressions. We begin by estimating univariate regressions of 1-day returns on each characteristic separately. Figure 2 presents the results.

[Figure 2 about here]

Statistical significance rates at the 1% and 5% significance level based on default standard errors average 28.1% and 38.3%, respectively, across the eight characteristics. These rates are greater for characteristics more closely linked to fundamentals. They are highest

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

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 Type I errors based on default standard errors are commonly an order of magnitude larger than the intended size of the test.

Statistical significance 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, rates based on industry-clustered standard errors are still far higher than the intended size of the test, averaging 18.1% and 29.3% 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 2 suggests that tests based on standard cross-sectional event study regressions do not allow for reliable tests of the differential return effects of a specific event on firms with different characteristics.

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 statistical significance 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. Panel A presents results for different approaches to computing standard errors. In addition to the clustering strategies described above, we also include results based on standard errors clustered at different industry levels, by headquarters state level, and by 5-

by-5 size × book-to-market bins and 5-by-5-by-5 size × book-to-market × gross profitability bins from Fama and French (2015). Panel B presents results based on default standard errors using different risk-adjusted return measures.

[Table 2 about here]

Statistical significance rates are higher for 5-day event windows than for 1-day event windows. Rates are generally lower for multivariate regressions than for univariate regressions. This result suggests that one benefit of including control variables in a cross-sectional event study regression is a reduction in Type I errors. Clustering by broader industry categories or size and book-to-market buckets result in the lowest statistical significance rates. However, it is important to note that a low rate of statistical significance in these tests does not indicate that a given clustering scheme is superior. Clustering with a small number of clusters can also substantially reduce power (Abadie et al., 2023). Risk-adjusting returns lowers significance rates slightly as well. However, the bottom line is that average statistical significance rates are considerably higher than the intended size of the tests in all cases.

Another way to assess the severity of the cross-correlation problem is to determine by how much standard errors would need to be inflated to achieve the intended Type I error rate (i.e., the size of the test). We refrain from tabulating these required rates, since the information is largely redundant with that in Table 2, with higher rejection rates in that table translating into greater need to inflate standard errors. However, to provide some context, the average White standard error across all eight characteristics in a univariate regression with a one-day event window would need to be multiplied by 3.0 (3.75) to achieve the target rate of statistical significance at the 5% (1%) level. The average FF49 industry-clustered standard error would need to be multiplied by 2.3 (2.85).

2.3. 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 1-day and 5-day window in our sample period and computing the fraction of windows in which the estimated relationship is statistically significant at the 1% and 5% levels. We analyze both of the specific sub-approaches to implementing these tests described in Section 1. Figure 3 presents the results.

[Figure 3 about here]

As anticipated, statistical significance rates using the time-series approaches are much closer to the intended Type I error rate under the null hypothesis than those based on standard cross-sectional regressions. However, these rates are still too high when we conduct hypothesis testing using the t-statistic approach, especially at the 1% significance level. These excess rates arise from fat tails in the distribution of coefficients that these tests fail to take into account. In contrast, statistical significance 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 a time-series approach and consider only this specific sub-approach for the remainder of the analysis.

We next assess the power of TS-OLS and TS-GLS 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 25 bp per one-standard deviation change in the characteristic, creating an artificial event window. We then estimate cross-sectional regressions for that window and for each window of the same size in a pre-event period consisting of the 199 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 1% and 5% level. Figure 4 presents detection rates based on these tests.

[Figure 4 about here]

Using either time-series approach, detection rates vary widely and are generally smaller for characteristics such as size and book-to-market ratio that are more directly related to a firm's fundamentals. Note that these are the same characteristics for which cross-sectional regressions produce the highest rates of statistically significant relationships (Figure 2). This is not a coincidence. The existence of strong relationships in many non-event windows makes it more difficult to detect the differential effect of an event. Not surprisingly, detection rates are higher when we introduce larger artificial effects. However, TS-OLS detection rates at the 5% level are less than 60% for 3 of the 8 characteristics when we add a 50 bp effect, suggesting that concerns about the power of the TS-OLS approach are valid.

In every case, the TS-GLS approach detects the added effect more frequently than the TS-OLS approach, with large increases in detection rates in many cases. For example, TS-GLS detects 25 basis points 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 power gains occurs for characteristics where TS-OLS power is low.

To allow for a more formal and comprehensive comparison, Table 3 reports the average detection rates across all eight characteristics for TS-OLS and TS-GLS for the 25 and 50 bp added effects separately for 1-day and 5-day event windows. Average detection rates across the eight characteristics are 1.25-1.92 times as high for TS-GLS as for TS-OLS. Detection rates are much lower for 5-day event windows than for 1-day event windows. Intuitively, the noise in returns is greater when the window over which they are measured is longer, making effects more difficult to detect. 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 not the focus of this paper, the results suggest that the returns to precisely pinning down the timing with which the market responds to an event and therefore being able to use a narrower event window are high.

[Table 3 about here]

Overall, the power advantage of TS-GLS over TS-OLS seems large enough, especially when power is low to begin with, to warrant using the former, even if the approach is more complex. Note that our `csestudy` Stata module automates both time-series approaches and so is equally easy to implement using TS-GLS as TS-OLS. For standard data sets, it produces output for either within seconds. However, even the TS-OLS approach is superior to standard cross-sectional approaches since it produces approximately the correct rate of Type I errors, while cross-sectional approaches often produce highly excessive Type I errors.

2.4. Revisiting the number of Principal Components to use in TS-GLS

We now briefly revisit our choice to use 100 PCs to estimate the covariance matrix used to weight observations in the GLS regressions in our TS-GLS approach. Specifically, we analyze the sensitivity of the average improvement in precision over the full sample period from using GLS instead of OLS to the number of PCs we use for each characteristic. This increase in precision is the source of power gains from the TS-GLS approach. To quantify the increase in precision, we compute the ratio of the standard deviations of the daily GLS and OLS coefficients. The lower this ratio, the tighter the distribution of GLS coefficients relative to OLS coefficients. Figure 5 plots the relationships between the standard deviation ratios and the number of PCs used for each characteristic.

[Figure 5 about here]

As with the relationship between minimum variance and number of PCs (Figure 1), the ratio of GLS to OLS coefficient standard deviations declines sharply with the first few PCs and is essentially flat between 50 and 195 PCs. It does not appear then that the power gains from TS-GLS are sensitive to the number of PCs in this range, further justifying our choice of 100 PCs.

3. Principal Component Analysis

In this section, we further analyze the PCs of returns that we use to construct the covariance weighting matrix for our GLS approach. By construction, these PCs capture the most important quanta of return cross correlations. 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 produce too many Type I errors to allow for reliable hypothesis testing and that common adjustments to standard errors are inadequate in addressing this problem. A time-series approach that involves comparing event-window relationships to a time-series of pre-event window relationships addresses the problem with excess rejection rates. When these relationships are estimated via OLS, statistical power can be low. Using the GLS-based alternative that we introduce increases power substantially and may therefore allow for detection of more modest relationships between event returns and characteristics that might otherwise be difficult to test.

Our first set of results hints at the possibility of 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 the effectiveness of clustering at accounting for cross-sectionally correlated errors is disappointing 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. We leave exploration of this problem in the context of other dependent variables for future work.

Appendix A. Proofs of Propositions

Proposition 1 (Bias and excess variance of the cross-sectional OLS approach)**.** *Cross-sectional OLS estimates of b_T in Equation (5) have the following relations to the true event-specific effect e_T:*

$$
\begin{aligned}
\mathbb{E}(\hat{b}_\tau - e_\tau) &= \underbrace{N_\tau \lambda}_{bias}, && (10)\\
Var(\hat{b}_\tau - e_\tau) &= N_\tau \left[Var(\hat{b}_\tau - b_\tau) + N_\tau \, Var(n_t) \right] \\
&= N_\tau \left[\underbrace{(X'X)^{-1}(X'\Omega_\nu X)(X'X)^{-1}}_{\textit{Correlation-corrected OLS variance}} + \underbrace{Var(n_T)}_{\textit{Excess variance}} \right], && (11)
\end{aligned}
$$

where X is an $N \times K$ matrix stacking all the $x'_{i,\tau-1}$ vectors. The $\mathbb{E}(\cdot)$ and $Var(\cdot)$ operators condition on $x_{i,\tau-1}$ and draw random $\nu_{i,t}$ and n_t for all $t \in \tau$.

Proof. Because $\hat{b}_T$ is an OLS estimate, we have:

$$
\hat{b}_T - b_T = (X'X)^{-1}X'\nu, \tag{34}
$$

where ν is a $N \times 1$ vector of all $\nu_{i,T}$. This implies:

$$
\begin{aligned}
\hat{b}_T - e_T &= \hat{b}_T - b_T + b_T - e_T \\
&= (X'X)^{-1}X'\nu + \lambda + n_T, && (35)
\end{aligned}
$$

which immediately gives Eqs. (10) and (11). □

Proposition 2 (Unbiasedness and correct variance of the time-series OLS approach)**.** *Write $b_{\tau_{-l}}$ for the vector of coefficients relating $\sum_{t\in\tau_{-l}} r_{i,t}$ to $x_{i,\tau_{-l}}$ in non-event window τ_{-l}, and b_τ*

for the coefficients in the event window. The time-series OLS estimates of e_τ is:

$$\hat{e}_\tau^{TSOLS} = \hat{b}_\tau - \frac{1}{L}\sum_{l=1}^{L}\hat{b}_{\tau_{-l}}, \tag{14}$$

Assuming stationary distributions for estimation error $\hat{b}_{\tau_{-l}} - b_{\tau_{-l}}$ and news n_t, $\hat{e}_\tau^{TSOLS}$ has the following relations to the true event-specific effect e_τ:

$$\mathbb{E}(\hat{e}_\tau^{TSOLS} - e_\tau) = \underbrace{0}_{No\ bias}, \tag{19}$$

$$Var(\hat{e}_\tau^{TSOLS} - e_\tau) = \underbrace{Var(\hat{b}_{\tau_{-l}})}_{Time\text{-}series\ variance}, \tag{20}$$

$$\Phi(\hat{e}_\tau^{TSOLS} - e_\tau) = \underbrace{\Phi(\hat{b}_{\tau_{-l}})}_{Time\text{-}series\ CDF}. \tag{21}$$

Proof. Because e_T has no effect when $t \neq T$, the stationarity of $\hat{b}_t - b_t$ and news n_t imply:

$$\mathbb{E}(\hat{b}_T - e_T) = \mathbb{E}(\hat{b}_t) \text{ for } t \neq T, \tag{36}$$

$$\mathrm{Var}(\hat{b}_T - e_T) = \mathrm{Var}(\hat{b}_t) \text{ for } t \neq T. \tag{37}$$

$\hat{e}_T^{\mathrm{TSOLS}}$ therefore satisfy:

$$\mathbb{E}(\hat{e}_T^{\mathrm{TSOLS}} - e_T) = \mathbb{E}(\hat{b}_T - \mathbb{E}(b_t) - e_T) = \lambda + e_T - \lambda - e_T = 0, \tag{38}$$

$$\begin{aligned}\mathrm{Var}(\hat{e}_T^{\mathrm{TSOLS}} - e_T) &= \mathrm{Var}(\hat{e}_T^{\mathrm{TSOLS}} - \hat{b}_T + \hat{b}_T - e_T) = \mathrm{Var}(\mathbb{E}(b_t) + \hat{b}_T - e_T), \\ &= \mathrm{Var}(\hat{b}_t).\end{aligned} \tag{39}$$

□

Appendix B. Primer on PCA of returns

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 $\mathbf{R}$.

Appendix C. Why PCA instead of "machine learning" techniques

Initial empirical tests of the APT in Chamberlain and Rothschild (1983), Connor and Korajczyk (1986), and others used PCA to estimate the unobserved factor structure, as we do here. Some more-recent asset pricing papers use techniques from the machine learning literature to extract latent factors, citing specific limitations of the PCA approach. Kelly et al. (2019) and Gu et al. (2021) point out that the time-invariant and linear relation between factors and realized returns in Equation (28) is unlikely to hold for individual stocks in longer samples. Lettau and Pelger (2020) points out that PCA misses factors that explain only a small fraction of realized returns but are nevertheless important for explaining expected returns. Neither of these limitations are salient for us because we are interested in finding all factors – priced or unpriced – that explain realized returns. This allows us to use a short time-series relative to typical asset pricing studies because estimating average returns requires much more data than estimating a covariance matrix.

A parallel body of recent research provides support for the continued use of PCA for estimating the latent factors that explain realized returns. Lopez-Lira and Roussanov (2023) uses PCA to show there are many unpriced latent factors in individual stock returns, and that hedging these factors substantially increases the Sharpe ratio of standard priced factors. Like us, they find that many (50 or more) principal components are necessary to explain a substantial fraction of realized stock returns, many of which are orthogonal to the canonical factors. Pelger (2020) uses PCA on high-frequency intraday individual stock returns to find latent systematic risk factors. Giglio and Xiu (2021) uses PCA on individual stock returns to model the correlation of return residuals, as we do. They use this technique to increase statistical power and robustness when estimating the characteristic-return relation.

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., 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 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.

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.

Chamberlain, G. and M. Rothschild (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica 51*(5), 1281–1304.

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.

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.

Connor, G. and R. A. Korajczyk (1986). Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of Financial Economics 15*(3), 373–394.

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.

Gu, S., B. Kelly, and D. Xiu (2021). Autoencoder asset pricing models. *Journal of Econometrics 222*(1), 429–450.

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

Kelly, B. T., S. Pruitt, and Y. Su (2019). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics 134*(3), 501–524.

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.

Lettau, M. and M. Pelger (2020). Factors that fit the time series and cross-section of stock returns. *Review of Financial Studies 33*(5), 2274–2325.

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

Pelger, M. (2020). Understanding systematic risk: A high-frequency approach. *Journal of Finance 75*(4), 2179–2220.

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 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.

Figure 1: 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 199 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.

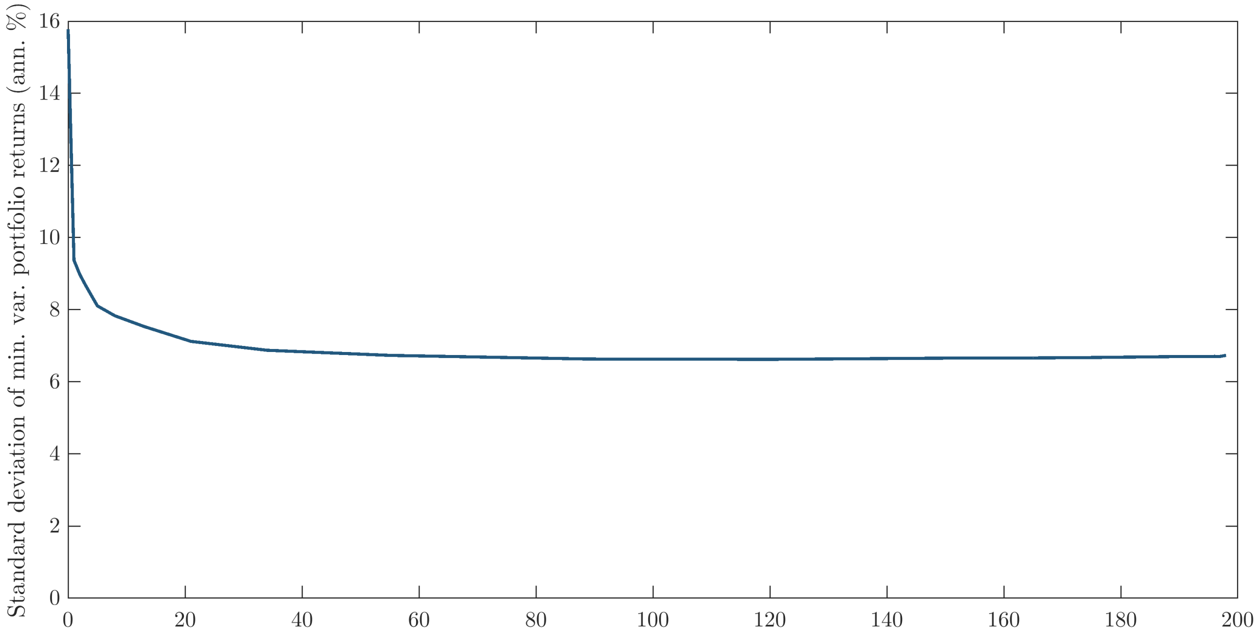

Figure 2: Cross-Sectional OLS Statistical Significance Rates

This figure depicts statistical significance rates from univariate cross-sectional OLS regressions of 1-day returns on each of eight characteristics. The figure shows rates based on default standard errors, White-adjusted (robust) 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 rates at the 1% significance level, while the bottom panel shows rejection rates at the 5% significance level. The sample period is the 7,811 trading days from 1991–2021.

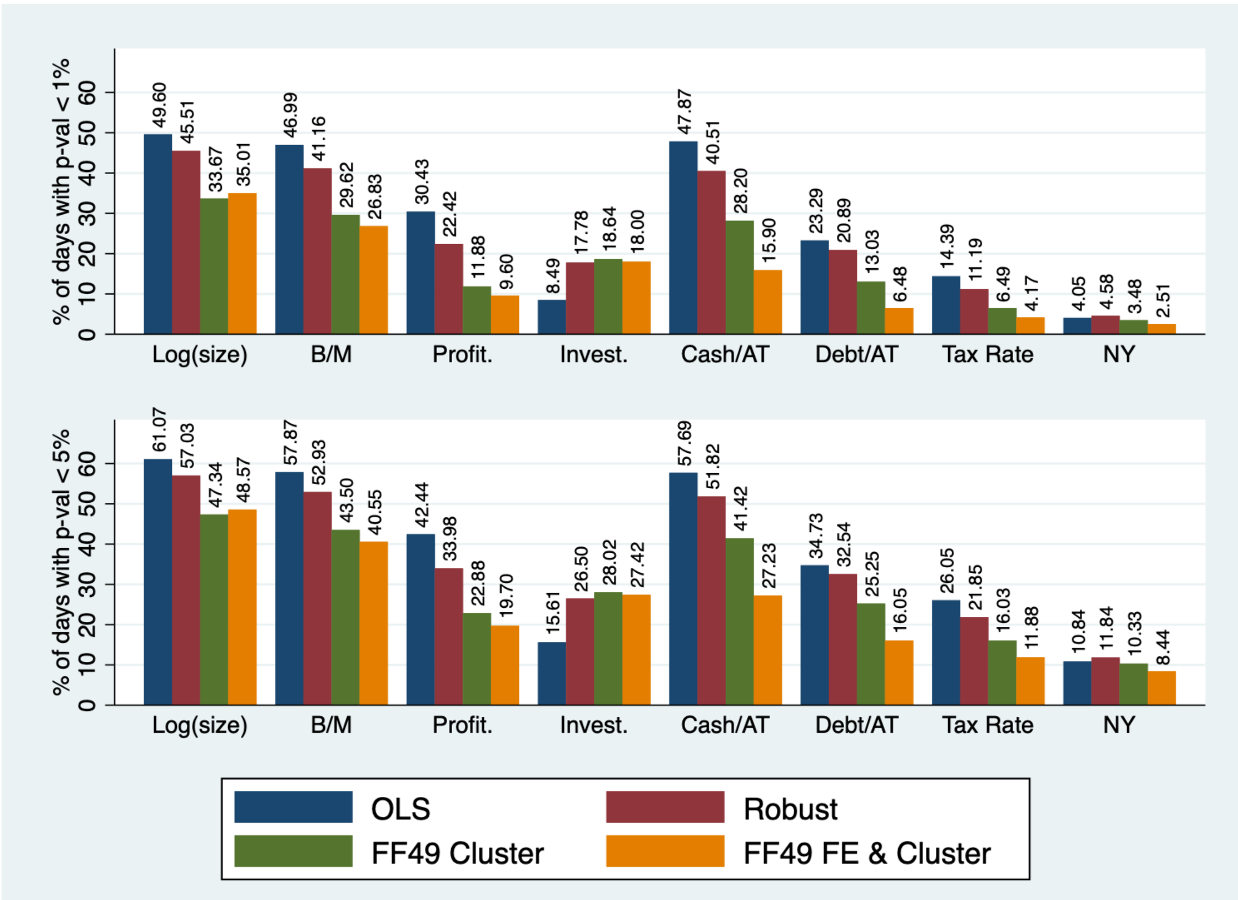

Figure 3: Time-Series OLS and GLS Statistical Significance Rates

This figure depicts statistical significance rates for time-series OLS and GLS approaches on 1-day returns for each of eight characteristics. The sample period is the 7,811 trading days from 1991–2021. Rates of significance based on both the empirical cumulative distribution function (CDF) and t-statistic approaches described in Section 1.2 are shown. The top panel shows rates at the 1% significance level, while the bottom panel shows rejection rates at the 5% significance level.

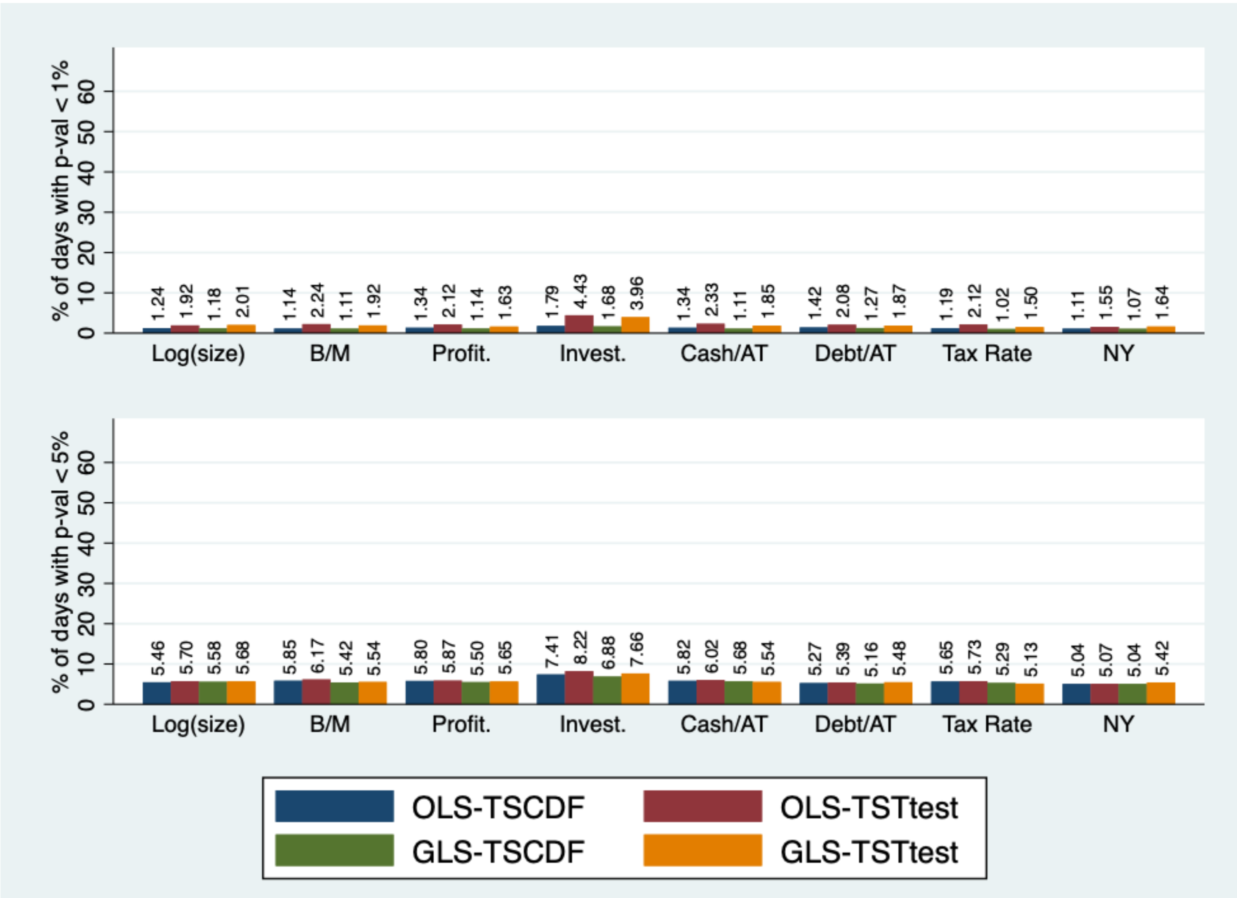

Figure 4: Time-Series OLS Detection Rates with Added Effects

This figure depicts detection rates of time-series OLS and GLS approaches on 1-day returns for each of eight characteristics with an effect of 25 bp or 50 bp added to returns for each one standard deviation increase in the given characteristic on the artificial event day. The detection rates are based on the empirical cumulative distribution function (CDF) approach described in Section 1.2. The top panel shows detection rates for the 25 bp effect, while the bottom panel shows detection rates for the 50 bp effect. The sample period is the 7,811 trading days from 1991–2021.

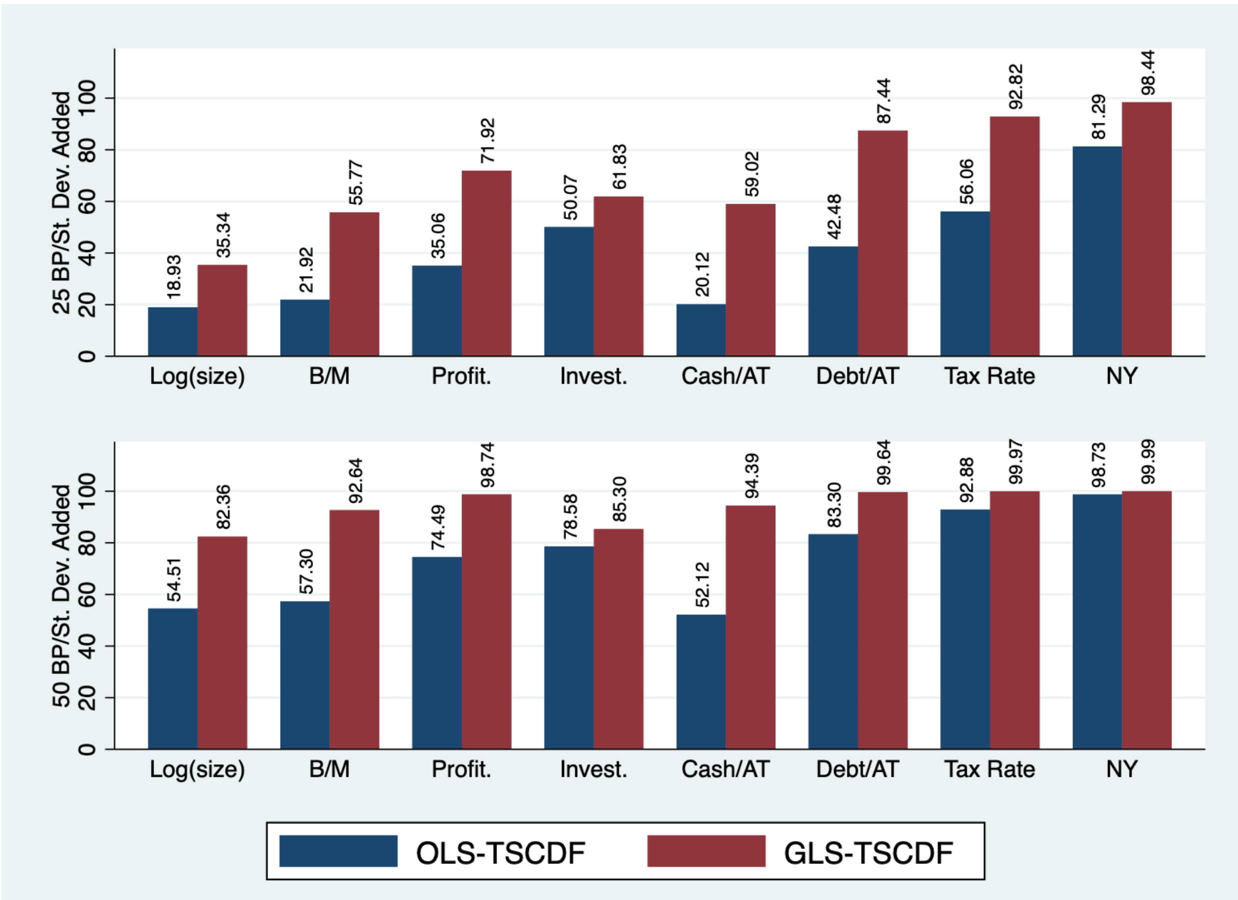

Figure 5: TS-GLS/TS-OLS Coefficient Standard Deviations and # of Principal Components

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.

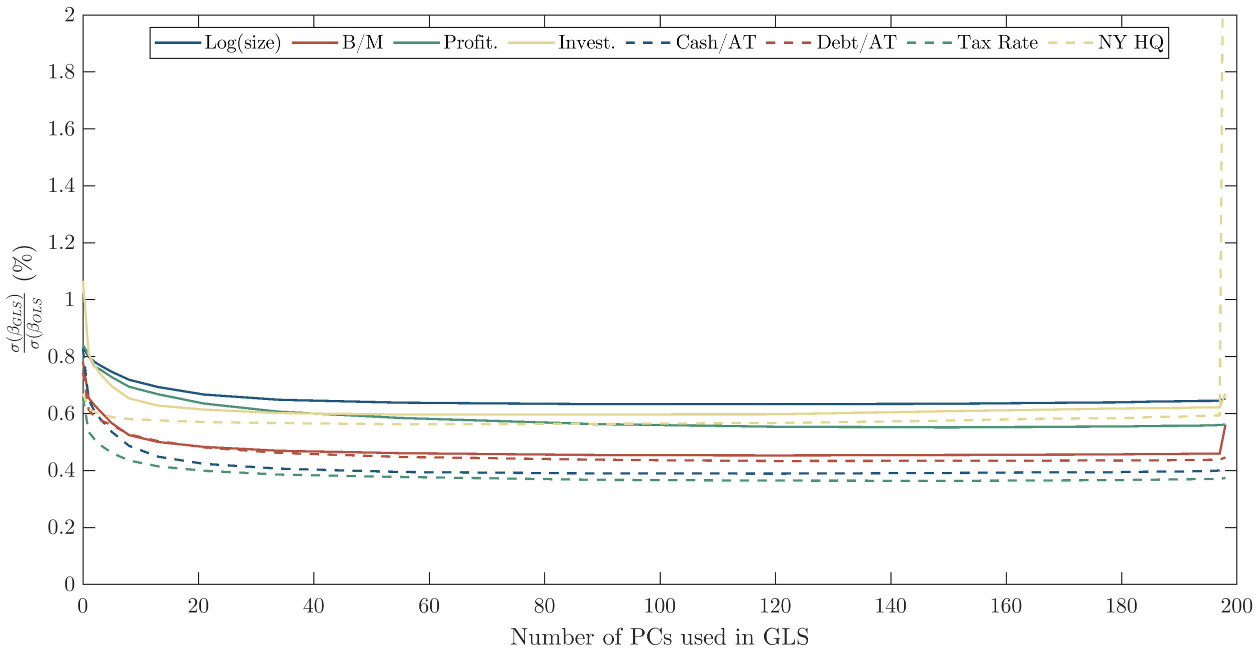

Figure 6: Explanatory Power of Principal Components

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.

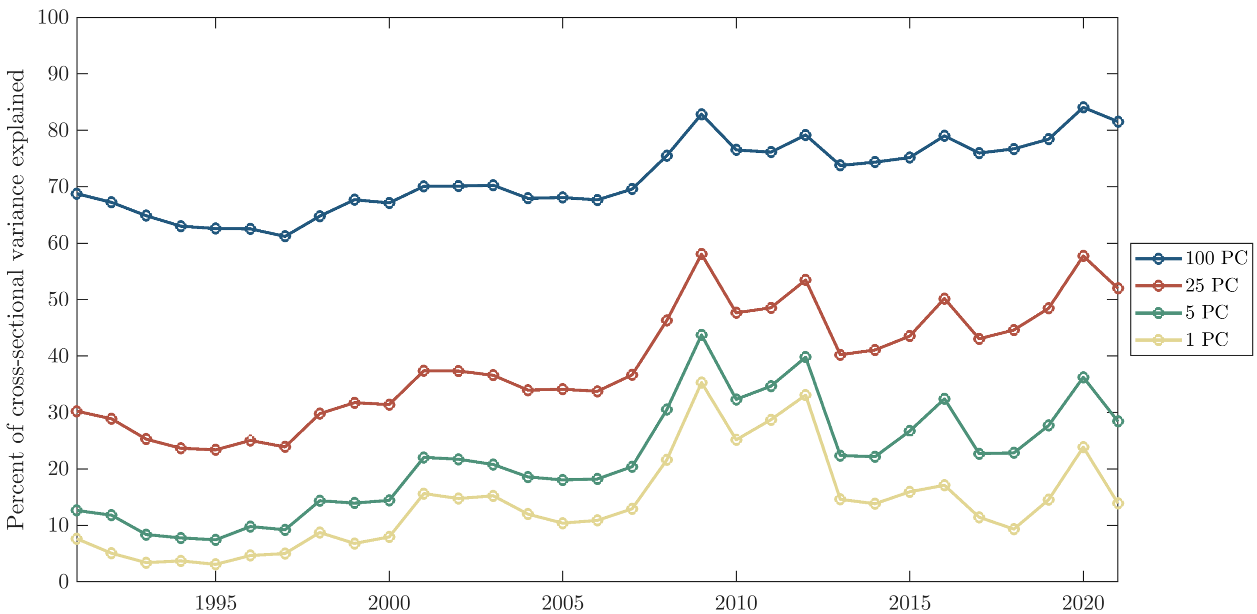

Table 1: Summary Statistics

This table presents summary statistics for the sample of firm-day observations used in our analysis. The sample period is the 7,811 trading days from 1991–2021. $Log(size)$ is the natural log of market equity, which is the product of daily closing stock price on the previous day 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	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
Log(size)	21,051,745	11,548	2,695	6.14	6.01	1.96	1.81
B/M	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.	15,780,343	7,968	2,020	0.36	0.05	30.27	30.26
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: Cross-Sectional OLS Statistical Significance Rates

This table presents mean statistical significance rates at the 1% and 5% statistical significance levels from cross-sectional OLS regressions of returns on each of eight firm characteristics. The first four columns report rates at the 1% significance level and the second four columns report rates at the 5% significance level. In each row, we report rates for univariate and multivariate regressions (where all eight characteristics are included as explanatory variables) for one-day and five-day return windows. In Panel A, we present rates for different clustering choices. In Panel B, we present rates for different risk-adjusted return measures. The sample period is the 7,811 trading days from 1991–2021.

Panel A: Standard Error Clustering

	1% Significance				5% Significance			
	Univariate		Multivariate		Univariate		Multivariate	
	1d	5d	1d	5d	1d	5d	1d	5d
Vanilla SE	28.1 %	34.5 %	17.6 %	28.1 %	38.3 %	44.6 %	27.9 %	38.3 %
Robust SE	25.5 %	32.1 %	16.3 %	25.5 %	36.1 %	42.8 %	26.4 %	36.1 %
FF49 Cluster	18.1 %	22.4 %	12.7 %	18.1 %	29.3 %	34.0 %	22.8 %	29.3 %
FF49 FE & Cl	14.8 %	19.6 %	11.0 %	14.8 %	25.0 %	30.4 %	20.0 %	25.0 %
FF30 Cluster	20.2 %	24.6 %	14.1 %	20.2 %	29.5 %	34.0 %	22.9 %	29.5 %
FF10 Cluster	11.1 %	13.9 %	8.4 %	11.1 %	23.4 %	27.1 %	19.6 %	23.4 %
SIC4 Cluster	20.9 %	26.0 %	14.0 %	20.9 %	31.8 %	37.3 %	24.1 %	31.8 %
State Cluster	28.3 %	34.0 %	20.8 %	28.3 %	40.0 %	45.7 %	31.9 %	40.0 %
5 Size x 5 BM	8.0 %	23.9 %	12.8 %	8.0 %	18.8 %	36.2 %	23.3 %	18.8 %
... x 5 Profit.	7.5 %	28.3 %	15.0 %	7.5 %	17.5 %	39.6 %	25.2 %	17.5 %
FF49 x State	16.2 %	24.9 %	16.1 %	16.2 %	28.3 %	36.1 %	27.3 %	28.3 %
FF30 x State	15.3 %	24.0 %	15.0 %	15.3 %	26.6 %	35.5 %	26.2 %	26.6 %

Panel B: Alternative Returns

	1% Significance				5% Significance			
	Univariate		Multivariate		Univariate		Multivariate	
	1d	5d	1d	5d	1d	5d	1d	5d
CAPM	17.3 %	20.9 %	10.8 %	17.3 %	28.3 %	32.6 %	20.0 %	28.3 %
FF3	22.7 %	26.9 %	13.6 %	22.7 %	34.0 %	38.3 %	23.1 %	34.0 %
FF4	22.5 %	26.5 %	14.0 %	22.5 %	33.8 %	38.0 %	23.6 %	33.8 %
LogRet	18.2 %	22.7 %	10.8 %	18.2 %	29.3 %	34.3 %	19.9 %	29.3 %

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

This table presents mean detection rates at the 1% and 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 25 bp and 50 bp per standard deviation change in a characteristic. The detection rates are based on the empirical cumulative distribution function (CDF) approach described in Section 1.2. 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 top panel shows rates at the 1% significance level, while the bottom panel shows rejection rates at the 5% significance level. The sample period is the 7,811 trading days from 1991–2021.

Panel A: 1% Significance Rate

Window	Effect (bp)	TS-OLS	TS-GLS	Difference	Ratio
1 day	25	24.21	52.71	28.49	2.18
1 day	50	60.07	87.00	26.93	1.45
5 day	25	6.63	14.12	7.50	2.13
5 day	50	21.66	43.71	22.05	2.02

Panel B: 5% Significance Rate

Window	Effect (bp)	TS-OLS	TS-GLS	Difference	Ratio
1 day	25	43.36	71.53	28.17	1.65
1 day	50	75.76	94.51	18.75	1.25
5 day	25	14.79	28.45	13.65	1.92
5 day	50	35.85	60.94	25.09	1.70

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%