JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 59, No. 3, May 2024, pp. 1003–1030

doi:10.1017/S0022109023000133

Double Machine Learning: Explaining the Post-Earnings Announcement Drift

Jacob H. Hansen
Aarhus University Department of Economics and Business Economics, CREATES
jacob.hald.hansen@gmail.com (corresponding author)

Mathias V. Siggaard
Aarhus University Department of Economics and Business Economics, CREATES
siggaard@econ.au.dk

Abstract

We demonstrate the benefits of merging traditional hypothesis-driven research with new methods from machine learning that enable high-dimensional inference. Because the literature on post-earnings announcement drift (PEAD) is characterized by a “zoo” of explanations, limited academic consensus on model design, and reliance on massive data, it will serve as a leading example to demonstrate the challenges of high-dimensional analysis. We identify a small set of variables associated with momentum, liquidity, and limited arbitrage that explain PEAD directly and consistently, and the framework can be applied broadly in finance.

I. Introduction

Linear regression is a simple and powerful technique: It provides interpretable coefficients, and its asymptotic properties are well-established and known. It is undoubtedly the default model when working with financial data, but its simplicity becomes its weakness in high-dimensional settings (see Gu, Kelly, and Xiu (2020), Christensen, Siggaard, and Veliyev (2023)). Relying on standard linear regressions with few explanatory variables in high-dimensional settings induces endless design combinations and ultimately leads to countless conclusions, and the new era in finance (governed by massive data sets and increasing computational capacity) is exacerbating the problem. Therefore, we reason that moving toward high-dimensional methods and data-driven approaches is to some degree inevitable.

This article demonstrates the possibility of conducting high-dimensional inferences by combining knowledge and theory from finance with modern statistical and computational techniques. By harnessing the pioneering work of

We thank participants of the 2022 Duke Financial Econometrics Seminar, 2022 Aarhus University Econometrics-Finance Seminar Series, and Nordic Finance Network Workshop for their many useful comments and questions. We also thank Tim Bollerslev, Kim Christensen, Tom Engsted, Jonas Nygaard Eriksen, Campbell R. Harvey, Jae Hoon Kim, Stig Vinther Møller, Nicolaj Søndergaard Mühlbach, Thomas Quistgaard Pedersen, Jylhä Petri, and Allan Timmermann for their valuable comments and insightful suggestions. Special thanks are due to Jennifer Conrad (the editor) and Vitaly Meursault (the referee) for detailed comments and helpful suggestions, which helped improve many aspects of this article.

Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018), that is, the double machine learning (DML) procedure, we conduct valid statistical inference in a high-dimensional setting while controlling for a large set of explanatory variables. As a direct result, we achieve a data-driven approach that reduces researcher dependency. The main objective is to construct a partially linear model with a modified moment condition. This modification ensures consistency and interpretability of the coefficient of interest equivalent to the familiar linear beta coefficient. In addition to allowing for valid inferences, the method permits a complex relationship between the dependent variable and the set of controls through a possible highly nonlinear and high-dimensional function. Because we rely on a generic approach, the proposed procedure can be applied more broadly in finance, such as to explain fund flows, preannouncement earnings drift, dividend announcements, initial public offerings, mergers and acquisitions, debt and equity issues, and stock splits. To demonstrate the benefits of the procedure, we revisit an unresolved question in finance (i.e., the origins of post-earnings announcement drift (PEAD)).

Known as PEAD, the tendency of stock prices to drift in the direction of an earnings surprise has attracted extensive attention since its discovery by Ball and Brown (1968). Its importance and relevance are revealed clearly by Fink (2020), who reviewed 216 published papers on the phenomenon. As emphasized by Fama (1998), PEAD is one of the most robust and persistent financial anomalies, and it has resulted in what we (in the spirit of Cochrane (2011)) term a "zoo of controls." The expansion of this zoo has been amplified by the reliance on high-dimensional empirical data, which are noisy and subject to omitted-variable bias (Fink (2020)). The PEAD literature is a classic example of a line of research that relies on massive sample sizes and a zoo of variables. This forces researchers to hand-pick a small set of controls while being impeded by a lack of academic consensus on this choice.

We demonstrate how inference is highly sensitive to the choice of controls, and we note that conclusions can favor the researchers' hypothesis if the "right set" of controls is chosen. Moreover, by relying on massive sample sizes, smaller and more complex statistical effects can be detected, but often with little or no practical relevance (see McCloskey and Ziliak (1996), Lin, Lucas Jr., and Shmueli (2013), Kim and Ji (2015), and Kim (2017)). Consequently, if the set of choices is large enough (see Harvey, Liu, and Zhu (2016)), statistical significance can be found even if there is no meaningful effect. Therefore, we reason that the PEAD literature provides a suitable context in which to illustrate the benefits of using high-dimensional data-driven methods to reduce researcher dependency and strengthen the credibility of explanations.

This article makes four main contributions. First, we showcase the advantages of combining high-dimensional methods with finance knowledge and theory to improve our understanding of PEAD. Second, by taking numerous potential controls generated by existing theories and relying on a new testing framework, we isolate a small set of variables that explain PEAD directly and consistently. Third, we demonstrate that if the high degree of researcher dependency is not mitigated, then incorrect conclusions can be drawn. Fourth, by exploring a large set of potential controls from the cross-section of stock returns literature, we find a more prominent role in price trends than that suggested in the PEAD literature.

The majority of studies examining PEAD seek to identify a variable that can explain cross-sectional differences in the drift. These studies rely on the earnings response coefficient framework, where the cumulative abnormal return after the announcement is regressed on i) SUE (the main surprise in earnings), ii) a variable of interest, iii) their interaction, and iv) a set of control variables.[1] To date, it has been standard to investigate a single variable of interest and rely on simple linear regression. However, our aim herein is to go further, with the empirical section investigating 20 different variables of interest previously related to the PEAD literature. Furthermore, we ensure a close link to the existing literature and robustness of the results because we test the DML framework against three standard ordinary least squares (OLS) specifications commonly applied in the literature. The first OLS regression often serves as a motivating regression, where the abnormal returns are regressed on SUE and a variable of interest. The second OLS regression allows for a small subset of control variables chosen ex ante, whereas the third OLS regression includes a multitude of controls. However, it is still doubtful whether this set of variables guarantees correct model specification and valid statistical inference. Therefore, we leverage the new capabilities of the DML procedure, where we extend the set of control variables to include 73 stock-specific variables from Green, Hand, and Zhang (2017), all first-order interactions, and a large set of fixed effects, totaling 2,836 controls. By harnessing the capabilities of post-lasso developed by Belloni and Chernozhukov (2013), we assume that a sparse linear combination of controls can approximate PEAD, and we allow the potential set to be high-dimensional. In this way, we can adhere to the idea that PEAD is approximated well by a sparse linear combination and also reduce researcher dependency. The instability of linear regression in high-dimensional settings becomes clear when its coefficients are compared to those of the DML procedure. Out of 20 variables, 17 are statistically significantly associated with either variation in PEAD or cumulative abnormal returns for the simplest model. The same issue arises in a "kitchen sink" approach controlling for a long list of variables (see Whited, Swanquist, Shipman, and Moon (2022)). Yet when leveraging the high-dimensional capabilities of the DML procedure by using all 2,836 controls, we find a 28% reduction in the number of variables explaining PEAD. Specifically, the procedure finds the following variables to explain variations in PEAD with statistical significance: The reporting lag between quarter end and announcement date, a loss indicator showing whether a past announcement was negative, an indicator for announcements in the same fiscal year, past returns over the last year, and the amount of other earnings news. These results imply that familiar variables such as firm size, preannouncement returns, and trading volume are not associated with PEAD when accounting for the large set of controls. Considering how inferences have changed with time, the DML procedure identifies a small set of variables that are consistently statistically significant, and the remaining variables are sporadically significant.

We illustrate the generality of the framework by testing whether any of the additional 73 stock-specific variables of Green et al. (2017) explain PEAD.

[1]When the interaction between the variable of interest and the SUE (surprise in earnings) variable can explain the cumulative abnormal return with statistical significance, it is said to explain the variations in PEAD.

Interestingly, the statistically significant variables are mainly those associated with price trends, liquidity, and volatility, consistent with the findings of Gu et al. (2020), who explore return predictability. To ensure the stability of the inferences from our main analysis, we consider several different popular quantile ranks of the variables and demonstrate how inferences are stable across different model specifications, estimation methods, and variable definitions.

Our study contributes to several strands of the literature. We are among the first to use the advantages of the DML procedure to overcome omitted variables, model misspecification, and nonlinearities in a finance application.[2] In recent years, machine learning methods have become popular because of their attractive ability in handling complex and high-dimensional data. The main reason for the success of these new methods is their ability to balance the bias–variance tradeoff. However, by allowing for bias, the interpretability of estimates becomes challenging, therefore much effort has been made to de-bias coefficients and generates confidence intervals when, for instance, relying on regularization methods such as ridge or lasso (e.g., see Belloni, Chen, Chernozhukov, and Hansen (2012), Nickl and Van De Geer (2013), Belloni, Chernozhukov, and Hansen (2014b), Javanmard and Montanari (2014), Lee, Sun, Sun, and Taylor (2016), and Athey, Imbens, and Wager (2018)). Although these are novel methods, the present article goes even further in that we rely on the more general DML procedure by Chernozhukov et al. (2018). Its generality ensures interpretability in high-dimensional settings via a broad array of machine learning methods such as lasso, random forests, boosted trees, and neural networks. Second, we contribute to the literature explaining PEAD by systematically evaluating multiple explanations while accounting for a high-dimensional set of controls. The origins of this literature can be traced back to Foster, Olsen, and Shevlin (1984) and Bernard and Thomas (1989), who find a negative relationship between PEAD and firm size, which is one of the best-established factors. Countless other explanations exist, such as limited arbitrage (Mendenhall (2004)), information uncertainty (Kormendi and Lipe (1987), Jiang, Lee, and Zhang (2005), and Francis, Lafond, Olsson, and Schipper (2007)), illiquidity (Bhushan (1994), Chordia, Goyal, Sadka, Sadka, and Shivakumar (2009)), and under-reaction to earnings news (Mendenhall (1991), DellaVigna and Pollet (2009), and Hirshleifer, Lim, and Teoh (2009)). We refer to Fink (2020) for an excellent and more-extensive review of the PEAD literature. Third, the present article adds to the new line of research addressing the issue of omitting relevant variables, which is a prevalent problem in finance (e.g., see Harvey et al. (2016), Feng, Giglio, and Xiu (2020), Freyberger, Neuhierl, and Weber (2020), and Giglio and Xiu (2021)).[3] Yet, in the empirical accounting literature, omitted variables have been addressed mainly in relation to causal relationships (e.g., see Roberts and Whited (2013), Gow, Larcker, and Reiss (2016)). Furthermore, issues with reliance on *p*-values are well-established in the finance literature (e.g., see Keuzenkamp and Magnus (1995), McCloskey and Ziliak (1996), Kim and Ji (2015), and Harvey et al. (2016)). However, few papers

[2]Yang, Chuang, and Kuan (2020) study the DML procedure using simulations and investigate the Big N audit quality effect.

[3]Feng et al. (2020) used double selection to mitigate an omitted-variable bias with lasso estimation to investigate the marginal importance of factors related to cross section of stock returns.

address the risk of model misspecification (e.g., see Feng et al. (2020)), which is our main focus.

The remainder of this article is organized as follows: Section II sets the theoretical foundation by introducing the DML procedure. Section III describes the data and defines important variables. Section IV demonstrates the capabilities of DML and the issues with least squares in high dimensions. Finally, Section V concludes the article.

II. Methodology

Selecting the correct covariates is challenging when theory about doing so is scarce and when relations are many and may be complex and nonlinear. In such settings, it is infeasible to assume that researchers can choose the correct set of controls unambiguously to ensure correct model specification. Therefore, the possibility of misspecifying the model and thereby creating omitted-variable bias cannot be neglected. When exploring PEAD, it is of utmost importance to select the set of explanatory variables consistently associated with PEAD in order to ensure valid inference of new variables. One strategy for doing so is to run a standard lasso regression assuming sparsity and define all nonzero variables as important; however, this naive method does not ensure valid inference because the coefficients will be biased, and so it is applicable only for prediction tasks. To overcome this, Nickl and Van De Geer (2013) explore the possibility of constructing confidence intervals in high-dimensional settings, and Javanmard and Montanari (2014) propose an efficient algorithm for constructing confidence intervals and *p*-values. Lee et al. (2016) propose a general approach to conduct valid inference after model selection and a variable relevance test.[4] Furthermore, several novel findings and methods have been proposed in the statistical literature on estimating treatment effects (see Belloni et al. ((2012), (2014b)), Belloni, Chernozhukov, and Hansen (2014a), Belloni, Chernozhukov, Fernández-Val, and Hansen (2017), and Athey et al. (2018)). However, the present article relies on the more general and flexible DML procedure of Chernozhukov et al. (2018), which can determine the correct set of explanatory variables in a purely data-driven way. The method enables estimation of the familiar linear beta coefficient with valid confidence intervals using a broad set of machine learning methods. Unlike other high-dimensional approaches, the DML framework is not restricted to a linear functional form because the procedure relies on a partial linear structure. The loss function is modified to ensure valid statistical inference of the variable of interest.

The DML Procedure

For the DML procedure, as in Chernozhukov et al. (2018), we consider the partially linear model (PLM) given by

[4] Additional studies include Van de Geer, Bühlmann, Ritov, and Dezeure (2014), Zhang and Zhang (2014), and Lei, G'Sell, Rinaldo, Tibshirani, and Wasserman (2018).

$$(1) \qquad Y = X\theta_0 + g_0(\mathbf{Z}) + \varepsilon, \qquad E[\varepsilon|\mathbf{Z}, X] = 0,$$

$$(2) \qquad X = m_0(\mathbf{Z}) + v, \qquad E[v|\mathbf{Z}] = 0,$$

where θ_0 is the parameter of interest, X is the variable of interest, Y is the outcome variable, ε and v are disturbance terms, and $\mathbf{Z} \in \mathbb{R}^P$ contains P covariates used as control variables. The main takeaways from the model are the following: First, θ_0 can be interpreted as a standard beta coefficient with valid standard errors in a linear regression. In relation to the PEAD literature, θ_0 is the parameter explaining the association between the variable of interest and PEAD. Second, the functional form by which $\mathbf{Z}$ affects X and Y can be nonlinear and high-dimensional. Third, we can use a broad set of machine learning methods to uncover these high-dimensional nonlinear relationships between $\mathbf{Z}$ and X and between $\mathbf{Z}$ and Y.

To clarify, equation (1) is the main equation, describing the relationship between the outcome variable and the variable of interest, X. Equation (2) is not of direct interest, but it accounts for the dependency between the controls and X; it is critical when removing the regularization bias, akin to omitted-variable bias. The PLM structure makes no parametric assumptions about $g_0(\mathbf{Z})$ or $m_0(\mathbf{Z})$, thus it induces fewer researcher-dependent choices and limits the risk of model misspecification. Instead, the two functions are treated as high-dimensional nuisance functions permitting nonlinear effects of $\mathbf{Z}$. The term "nuisance function" refers to a function that is not of immediate interest but is necessary for ensuring correct model specification. Estimates of both g_0 and m_0 are needed to partial out the impact of $\mathbf{Z}$ on X and Y, in the spirit of Frisch–Waugh–Lovell, ensuring an unbiased estimate of θ_0 (see Chernozhukov et al. (2018)).

A straightforward albeit naive approach would be to estimate g_0 in a separate part of the data set (known as the auxiliary part) via machine learning, then partial out the effect of $\mathbf{Z}$ on Y, and finally estimate θ_0 by using the least squares of X on $Y - \widehat{g}_0(\mathbf{Z})$. However, as discussed by Chernozhukov et al. (2018), this naive approach will induce overfitting and regularization bias, triggered by $\widehat{\theta}_0$ not being root-n consistent. To overcome the "inferior" rate of convergence, Chernozhukov et al. (2018) proposed an orthogonalized formulation of the PLM; this accounts for regularization bias resulting in root-n consistency of θ_0. Thus, the new representation of the PLM is given as

$$(3) \qquad Y - \ell_0(\mathbf{Z}) = (X - m_0(\mathbf{Z}))\theta_0 + \varepsilon,$$

where the estimate of the conditional expectation functions $\ell_0(\mathbf{Z}) = E[Y|\mathbf{Z}] = m_0(\mathbf{Z})\theta_0 + g_0(\mathbf{Z})$ and $m_0(\mathbf{Z}) = E[X|\mathbf{Z}]$ are nonparametric regression tasks estimated in an auxiliary sample via machine learning yielding $\widehat{\ell}_0$ and $\widehat{m}_0$. With the primary object of predicting conditional expectations, machine learning is a good candidate for the task in hand. The impact by which $\mathbf{Z}$ affects X and Y is then partialled out in the main sample, followed by the double-residualized regression of equation (3). Following Chernozhukov et al. (2018), 2-fold cross-fitting is applied to ensure that an overfitting bias from using machine learning will not distort the θ_0 estimates. Specifically, the data are split into two parts: the main part $n = N/2$ and an auxiliary part of size $N-n$. We use the auxiliary part to obtain an estimate of $E[Y|\mathbf{Z}]$ and $E[X|\mathbf{Z}]$, enabling us to achieve a valid estimate of θ_0 in the main part. This cross-fitting strategy yields a root-n consistent estimate of θ_0 for a large set of machine learning methods. As argued by Chernozhukov et al. (2018), the 2-fold cross-fitting procedure reduces the efficiency of the estimate because the data are split into two. Efficiency can be restored by flipping the role of the main and auxiliary samples.[5] The main results rely on the post-lasso estimator of Belloni and Chernozhukov (2013) when estimating the nuisance functions. In a robustness check, we also consider the random forest estimator. We provide further detail on the DML procedure and post-lasso estimator in Appendix A of the Supplementary Material.

III. Data

The sample of stocks used herein includes all listed securities from the Center for Research on Security Prices (CRSP) database with share codes 10 and 11. We require that all securities are listed on the New York Stock Exchange, NASDAQ, or AMEX and report earnings announcements in the merged CRSP/Compustat database.[6] Before matching across variables, the entire sample spans from July 1971 to the end of Mar. 2020, totaling 781,975 firm-quarters across 19,253 unique companies.

Variable Construction

We construct 20 control variables previously used in the empirical literature; the construction of each variable is described in Appendix B of the Supplementary Material. To improve numerical optimization and ease the interpretation of coefficients, we standardize the continuous value of each control. In a robustness check reported in Section IV.E, we consider various quantile ranks of the variables. Summary statistics for the variables are given in Table 1. To control for additional stock-specific covariates, we consider 96 stock-specific variables from the cross-section of the stock returns literature as used by Green et al. (2017).[7] After matching with our data, we remove all variables for which our data set has more than 50,000 observations with missing values, which yields 73 variables. We measure stock-specific variables in the same month as the event date.[8] These stock-specific predictive characteristics have not necessarily been linked to PEAD but have been found to explain the cross-section of stock returns. We provide a list of the stock-specific variables in Appendix B of the Supplementary Material.[9] To control for possible industry and time effects, we include i) industry dummies

[5]The two estimates will be approximately independent; thus, averaging will restore full efficiency.

[6]We link firm PERMNO of the CRSP/Compustat database to CUSIP codes used in the Thomson Reuters IBES and Datastream database using the linking suite from Wharton Research Data Services.

[7]See the corresponding paper for details about the variables. We thank the authors for making their SAS code available; it can be found on Jeremiah Green's website: https://sites.google.com/site/jeremiahrgreenacctg/home.

[8]Results are robust to measuring the stock-specific variables in the prior month. The main results are reported using this measurement period in Appendix C of the Supplementary Material.

[9]Because of data availability, we do not consider all the variables used by Green et al. (2017).

TABLE 1

Summary Statistics of Variables of Interest

In Table 1, variables are scaled for comparability. Variable definitions are provided in Appendix B of the Supplementary Material.

Variable	Type	Scaling	Mean	25th	50th	75th	Std. Dev.
CAR	Value	10^2	−0.71	−11.27	−0.61	9.63	22.90
SUE	Decile	–	5.55	3.00	6.00	8.00	2.77
RUNUP	Value	10^2	0.27	−6.28	0.16	6.57	14.22
PASTRET	Value	10^2	6.21	−9.59	4.17	18.07	31.93
EARET	Value	10^2	0.17	−2.27	0.05	2.50	5.99
PRICE	Value	10^{-2}	3.64	0.09	0.17	0.33	96.53
SIZE	Value	–	6.78	5.40	6.68	8.01	1.88
DOLVOL	Value	10^{-5}	9.29	0.18	0.94	4.54	90.94
VOL	Value	10^{-4}	2.53	0.14	0.53	1.95	6.73
BM	Value	–	0.53	0.28	0.46	0.70	0.66
ANALYST	Value	–	7.35	2.00	5.00	10.00	6.52
LEV	Value	–	0.21	0.04	0.18	0.31	0.20
TURNOVER	Value	–	0.26	0.07	0.14	0.26	0.97
DECR	Dummy	–	0.39	0.00	0.00	1.00	0.49
LOSS	Dummy	–	0.15	0.00	0.00	0.00	0.36
NRANK	Decile	–	5.86	3.00	6.00	8.00	2.94
REPLAG	Value	10^{-1}	3.59	2.40	3.10	4.30	1.75
SAMEFIS	Dummy	–	0.75	0.00	1.00	1.00	0.43
EXPRISK	Value	10^3	0.16	0.03	0.07	0.16	0.30
ARBRISK	Value	10^3	1.14	0.36	0.68	1.34	1.48
BETA	Value	–	1.16	0.75	1.09	1.49	0.58
ILLIQ	Value	10^7	3.00	0.01	0.04	0.34	35.46

constructed using French's 48 industry classifications and ii) the year, month, and day of the week dummies. After matching across variables, we have 97 fixed effects.

A. Surprise in Earnings

Our main results are based on the earnings surprise measure using analyst forecasts and actual earnings rather than time series forecasts (see the discussion in Livnat and Mendenhall (2006)).[10] We define the main surprise in earnings (SUE) variable for firm i in quarter t as

$$\text{(4)} \qquad \text{SUE}_{i,t} = \frac{\text{EPS}_{i,t} - \widehat{\text{EPS}}_{i,t}}{\text{PRICE}_{i,t}},$$

where $\text{EPS}_{i,t}$ is the actual earnings per share reported in IBES, $\widehat{\text{EPS}}_{i,t}$ is the median analyst expectation, and $\text{PRICE}_{i,t}$ is the price reported in IBES. All stocks with a share price less than $1 and observations with actual or forecasted earnings exceeding the share price are removed to reduce the noise of SUE deflators (see Livnat and Mendenhall (2006)). After matching across all variables and removing those missing, the sample spans from the start of Apr. 1984 to the end of Mar. 2020, totaling 170,719 firm-quarters across 5,315 unique companies.

To address the existence of outliers and nonlinearity in the earnings surprise-return relation, we replace the continuous value of SUE with its decile rank

[10]For completeness, we also consider standardizing using the standard deviation of analyst forecasts and a time series measure of SUE (see Jegadeesh and Livnat (2006), Livnat and Mendenhall (2006)). We again find a small set of variables explain PEAD. These results are included in Appendix D of the Supplementary Material.

computed using yearly breakpoints. We use decile ranks because PEAD is strongest for relatively extreme announcements (see Bernard and Thomas (1989), Bernard and Thomas (1990), Bhushan (1994), and Bartov, Radhakrishnan, and Krinsky (2000)).

B. Abnormal Returns

We calculate the abnormal return using the period after the earnings announcement (occurring at time $t=0$) from time $t+2$ to $t+61$, $\text{CAR}^{[2,61]}$, because Bernard and Thomas (1989) show that PEAD is strongest in the 3 months after the earnings announcement. The firm-specific cumulative abnormal return is determined as the actual return minus the return predicted by the CAPM.[11] The model uses a fixed estimation window of $t-231$ to $t-31$ days prior to the announcement to avoid potential short-term trends in returns (see MacKinlay (1997)).[12]

IV. Results

A. Model Specifications

The majority of studies examining PEAD seek to identify a variable that can explain cross-sectional differences in the drift. These studies rely on the earnings response coefficient framework, where the next 3 months' abnormal returns after an announcement are regressed on SUE, a variable of interest, their interaction, and a set of control variables. The aim is to find a significant relation between the variable of interest and PEAD. The main focus is therefore on the interaction term, which gauges how the stock market response varies across firms with different levels of the variable of interest.

We create four different model specifications to test the sensitivity of inferences from the ex ante model chosen by the researcher and illuminate potential issues associated with omitted-variable bias. First, we estimate the cross-sectional differences in $\text{CAR}^{[2,61]}$ by including SUE, the variable of interest, X, and its interaction; often used as a motivating regression in empirical studies (see DellaVigna and Pollet (2009), Hirshleifer et al. (2009)). We consider L variables of interest, $\boldsymbol{X}_i \in \mathbb{R}^{\text{L}}$, where each variable of interest is estimated in a separate regression, totaling L regressions. The general equation for the lth variable is as follows:

$$(5)\quad \text{CAR}_i^{[2,61]} = \theta_{l,0} + \text{SUE}_i\theta_{l,1} + X_{l,i}\theta_{l,2} + (X_{l,i} \times \text{SUE}_i)\theta_{l,3} + \varepsilon_{l,i}, \qquad 1 \le l \le L,$$

where $X_{l,i}$ is the variable of interest for announcement i, $\theta_{l,2}$, and $\theta_{l,3}$ are parameters of interest, and $\varepsilon_{l,i}$ is an error term.[13] Thus, $\theta_{l,3}$ measures the effect of SUE

[11] Following MacKinlay (1997), we regress the firm-specific return for u days in the estimation window on the market return, that is $R_{u,i} - r_f = \alpha_i + \beta_i\left(R_{u,m} - r_f\right)$, where $R_{u,m}$ is the CRSP value-weighted market return. The abnormal return at time t for firm i is given as $\text{AR}_{t,i} = R_{t,i} - r_f - \widehat{\alpha}_i - \widehat{\beta}_i\left(R_{t,m} - r_f\right) + \varepsilon_i$. We require that at least 140 observations are present within the estimation window.

[12] The results are robust to changing the measurement horizon to $\text{CAR}^{[2,21]}$ or $\text{CAR}^{[2,41]}$ and to risk-adjustments of CAR using the Fama–French 3-factor model and its extension with momentum (see Carhart (1997)).

[13] When predicting the cross-sectional differences in $\text{CAR}^{[2,61]}$ and SUE_i, we run the following regression: $\text{CAR}_i^{[2,61]} = \theta_0 + \theta_{l,1}\text{SUE}_i + \varepsilon_{l,i}$.

on $\text{CAR}^{[2,61]}$ for different values of $X_{l,i}$ and is therefore the parameter of primary interest. The parameter of secondary interest is the estimate on the variable of interest, namely $\theta_{l,2}$, which measures its association with $\text{CAR}^{[2,61]}$. In addition to SUE, we consider 20 variables previously used in the empirical literature, hence $L=20$.

For the second specification, a subset of control variables chosen ex ante, $\mathbf{x}_i \in \mathbf{X}_{(-l),i}$ and the corresponding interactions, $\mathbf{x}_i \times \text{SUE}_i$, are added. We define $\mathbf{Z}_i^{(1)} = \{\mathbf{x}_i, \mathbf{x}_i \times \text{SUE}_i\}$. In line with the first specification, the general equation for the lth variable is as follows:

$$(6)\quad \text{CAR}_i^{[2,61]} = \theta_{l,0} + \text{SUE}_i\theta_{l,1} + X_{l,i}\theta_{l,2} + (X_{l,i} \times \text{SUE}_i)\theta_{l,3} + \mathbf{Z}_i^{(1)}\beta + \varepsilon_{l,i}, 1 \le l \le L,$$

where $\mathbf{Z}_i^{(1)}$ denotes the set of control variables chosen ex ante and $\beta = (\beta_1, \beta_2, \ldots, \beta_P)$ is a P vector. This is the most common specification used in the empirical literature explaining PEAD (see, e.g., Bhushan (1994), Bartov et al. (2000), and Mendenhall (2004)). However, it is notable that this specification often relies on 5–10 control variables with little justification for including or excluding particular variables (Whited et al. (2022)). Consistent with the literature, we consider only a subset of common control variables among the 20 variables studied, that is, the abnormal 1-month prior return (RUNUP), returns over the past year (PASTRET), stock price (PRICE), firm size (SIZE), dollar trading volume (DOLVOL), book-to-market (BM), explained risk (EXPRISK), arbitrage risk (ARBRISK), and illiquidity (ILLIQ), hence $P = 9 \times 2 = 18$.[14] We consider the impact of altering the set of controls in Section "The Sensitivity of OLS."

The third specification mimics a "kitchen sink" approach in which a large set of controls is included to avoid potential review comments (see Whited et al. (2022)). We include the full set of control variables, that is, SUE, all 20 variables of interest, 47 industry fixed effects, and 50 time fixed effects (year, month, and day of the week), defined as $\mathbf{D}_i$. We define $\mathbf{Z}_i^{(2)} = \left\{\mathbf{X}_{(-l),i}, \mathbf{X}_{(-l),i} \times \text{SUE}_i, \mathbf{D}_i\right\}$ and the equation for the lth variable as

$$(7)\quad \text{CAR}_i^{[2,61]} = \theta_{l,0} + \text{SUE}_i\theta_{l,1} + X_{l,i}\theta_{l,2} + (X_{l,i} \times \text{SUE}_i)\theta_{l,3} + \mathbf{Z}_i^{(2)}\beta + \varepsilon_{l,i}, 1 \le l \le L,$$

where $\mathbf{Z}_i^{(2)}$ contains the set of controls and $\beta = (\beta_1, \beta_2, \ldots, \beta_P)$ is a P vector. A total of $P = 19 + 19 + 97 = 135$ variables are included in $\mathbf{Z}_i^{(2)}$.

The fourth specification removes the need for ex ante variable selection by utilizing the DML procedure, and it also allows for a more flexible PLM structure. We consider the following equations for the lth variable:

$$(8)\quad \text{CAR}_i^{[2,61]} = \text{SUE}_i\theta_{l,0} + X_{l,i}\theta_{l,1} + (X_{l,i} \times \text{SUE}_i)\theta_{l,2} + g_{l,0}\left(\mathbf{Z}_i^{(3)}\right) + \varepsilon_{l,i}, 1 \le l \le L,$$

$$(9)\quad \text{SUE}_i = m_{l,0}\left(\mathbf{Z}_i^{(3)}\right) + v_{l,i}^0,$$

[14]Note that if one of the controls is the variable of interest, then $P = 16$.

$$X_{l,i} = m_{l,1}\left(\mathbf{Z}_i^{(3)}\right) + v_{l,i}^1, \tag{10}$$

$$(X_{l,i} \times \text{SUE}_i) = m_{l,2}\left(\mathbf{Z}_i^{(3)}\right) + v_{l,i}^2, \tag{11}$$

where $\theta_{l,1}$ and $\theta_{l,2}$ are the parameters of interest, $\mathbf{Z}_i^{(3)}$ contains the set of controls, and $\varepsilon_{l,i}$, $v_{l,i}^0$, $v_{l,i}^1$, and $v_{l,i}^2$ are error terms. Because $\theta_{l,2}$ measures if the effect of SUE on $\text{CAR}^{[2,61]}$ varies significantly for different values of the variable of interest, it is our primary parameter of interest. We follow the method described in Section II and use the post-lasso method to estimate $\widehat{\ell}_{l,0}$, $\widehat{m}_{l,0}$, $\widehat{m}_{l,1}$, and $\widehat{m}_{l,2}$ (see equation (3)).[15] To leverage the high-dimensional capabilities of the post-lasso estimator, we also consider 73 stock-specific variables, $\mathbf{S}_i$, from Green et al. (2017). Therefore, $\mathbf{Z}_i^{(2)}$ is extended by $\mathbf{S}_i$, $\mathbf{S}_i^2$, and all first-order interaction terms, hence $\mathbf{Z}_i^{(3)}$ encompasses 135 main variables, 73 stock-specific variables, and 2,628 derived variables.

As noted, we systematically examine each variable of interest, which yields L unique separate regressions. This allows for more flexibility in the DML method, compared to simply expanding $X_{l,i}$ to $\mathbf{X}_i \in \mathbb{R}^{\mathrm{L}}$ in equation (8). First, such an expansion will considerably restrict the model by assuming no dependency and no nonlinearities between the variable of interest $X_{l,i}$ and $\mathbf{X}_{-l,i}$. Second, a linear relation between $\mathbf{X}_{-l,i}$ and $\text{CAR}_i^{[2,61]}$ is imposed by the researcher. Therefore, to reduce researcher dependency and enable a more data-driven approach, we examine each variable sequentially, as opposed to performing a joint examination.[16] To increase flexibility and allow for potential nonlinearities, we repeat the analysis using the random forest method to estimate the nuisance function in Section IV.E.1. The random forest method not only allows for nonlinearity but also implicitly accommodates interaction effects between explanatory variables, which enables us to exclude all first-order and second-order interaction terms in $\mathbf{Z}_i^{(3)}$.

B. Empirical Results of the Model Specifications

Inferences for both the primary interaction term estimates and the total number of significant estimates across the four model specifications are summarized in Table 2 and the corresponding coefficient estimates for the interaction terms are presented in Table 3. The number of significant variables are reported using a 1% standard significance level and also with a Benjamini–Hochberg correction, which controls for the error rate when conducting multiple testing (Benjamini and Hochberg (1995)). We cluster standard errors by announcement day and firm

[15]As discussed in Section II, we ensure full efficiency and valid standard errors by employing 2-fold cross-fitting. The cross-fitting procedure introduces randomness but to limit the effect of this variability, we obtain 100 estimates of each parameter of interest and apply the median method, see Chernozhukov et al. (2018).

[16]Note that we cannot compute meaningful and comparable R^2 values between OLS and the DML procedure. As seen in equation (3), the resulting R^2 from the DML specification will only be based on the residual regression after partialling out.

TABLE 2

Summary of Inferences Across Model Specifications

The first 2 rows of Table 2 report the numbers of significant interaction terms – which are the parameters of primary interest – across the 4 model specifications at a 1% standard significance level and a Benjamini–Hochberg corrected level. The bottom 2 rows report the total numbers of significant interaction terms and variables of interest (noninteraction terms).

	OLS			DML w. Post-Lasso
Separate Regressions Using	No Controls	Selected Controls	Full Set of Controls	Nuisance Function
Significant interaction terms (1%)				
Standard significance level	8/20	8/20	7/20	5/20
B–H corrected	7/20	5/20	6/20	4/20
Significant variables (1%)				
Standard significance level	23/41	23/41	21/41	13/41
B–H corrected	23/41	20/41	20/41	12/41

TABLE 3

Estimates of Interaction Terms Across Model Specifications

In Table 3, variables are listed based on the magnitude of the differences in coefficient estimates from the first and fourth columns. * denotes significance at the 1% level, and ** indicates changes in significance between the first and fourth columns. The specifications use all firm-quarters in the sample and each control variable is continuous but standardized, except for SUE, which is in deciles. The first numerical column reports estimates (in percentage) from separate regressions with no controls for the parameters of primary interest. The second column adds the set of controls chosen ex ante and their interactions with SUE, and the third column uses the full set of controls. The fourth column reports estimates using the high-dimensional nuisance function. The estimates for the variables of interest themselves (noninteraction terms) are reported in Table C1 in the Supplementary Material. Standard errors are clustered by day and firm.

	OLS			DML w. Post-Lasso
Separate Regressions Using:	No Controls	Selected Controls	Full Set of Controls	Nuisance Function
SUE × REPLAG	−3.588*	−3.264*	−3.693*	−2.051*
SUE × SIZE	0.138	−1.792*	−2.969*	−1.338
SUE × LOSS**	0.077	−0.343	−0.270	−1.170*
SUE × VOL	0.405	1.111*	1.162*	−0.817
SUE × DELVOL	0.732	0.893*	0.651	−0.418
SUE × LEV	0.435	−0.018	0.167	−0.499
SUE × EXPRISK**	−1.231*	0.563	0.642	−0.383
SUE × DECR	0.197	0.173	0.060	−0.624
SUE × EARET**	1.333*	0.905*	0.861*	0.548
SUE × ARBRISK**	−1.534*	−1.464	−1.419	−0.886
SUE × ANALYST	−0.317	0.043	−0.281	−0.873
SUE × ILLIQ	0.399	0.178	0.084	−0.078
SUE × SAMEFIS	1.504*	1.654*	1.117*	1.925*
SUE × BETA**	−1.480*	−0.710	−0.833	−1.183
SUE × PASTRET	3.242*	3.225*	3.381*	2.973*
SUE × BM	−0.608	−0.506	−0.645	−0.751
SUE × TURNOVER	0.143	0.369	0.146	0.009
SUE × PRICE	1.086	1.226	1.199	0.969
SUE × RUNUP	0.656	0.746	0.714	0.719
SUE × NRANK	1.317*	0.766*	−1.680*	1.317*
No. of obs.	170,719	170,719	170,719	170,719
Fixed effects			Yes	Yes
No. of variables	3	21	138	2,839

(Petersen (2009)) to account for potential time-series and cross-sectional dependencies. After partialling out the effect of the nuisance functions, we compute cluster robust standard errors for the DML procedure.

The first column of Table 2 reports the results for the first specification, which depends on a single variable (see equation (5)). In total, 8 out of 20 interaction terms can significantly explain the variation in PEAD. Interestingly, we observe only a minimal drop in the number of significant variables when considering the more conservative Benjamini–Hochberg corrected p-values. It is difficult to justify the assumption that the model specification is correct, because inferences are only based on SUE, the variable of interest, and their interaction. If this naive specification is perfectly specified and no omitted-variable bias is present, then including more controls will have a limited effect on the coefficient estimates and their standard errors.

Reported respectively in the second and third columns, the second and third specifications show little or no change in the numbers of significant variables, tentatively suggesting that there is no omitted-variable bias in the first specification.[17] However, when investigating the coefficient estimates below, a profound inconsistency is detected across columns.

The fourth specification makes no ex ante model selection choices. Instead, it relies on the DML procedure and the extended data set (see equations (8)–(11)). Only 1 out of 4 interaction terms is significant at a 1% level, equivalent to at least a 28% decrease compared to the first, second, and third specifications. The decrease is even more pronounced for all variables, reported in the bottom 2 rows of Table 2.

Table 3 reports the coefficient estimates of primary interest, that is, the interactions between SUE and each variable of interest. We also report estimation results for the variables of interest themselves (noninteraction terms) in Table C1 in the Supplementary Material. As stated in Section IV.A, we estimate a single regression for each variable, so each row of Table 3 represents 20 separate regressions. Variables are listed based on the magnitude of differences in coefficient estimates between the first and fourth columns. * indicates significance at the 1% level, and ** indicates changes in significance between the first and fourth columns. We provide a visual summary of the results in Figures C1 and C2 in the Supplementary Material.

As reported in the first column of Table 3, 8 out of 20 variables are statistically significant, whereas 15 out of 20 are significant in the first column of the noninteraction terms estimates in Table C1 in the Supplementary Material. The ease of detecting small and complex effects with the first specification is clearly illustrated when merging the results of the 2 columns, because 17 out of 20 variables can statistically significantly explain either PEAD or the abnormal return at a 1% level. It is important to note that the ease of detecting statistical significance is amplified by the reliance on over 170,000 earnings announcements.[18]

Table 3 shows substantial changes in the coefficient estimates reported across columns. This suggests the presence of an omitted-variable bias, which was not evident from the summary in Table 2. For instance, the estimate of SUE $\times$ NRANK changes sign from 1.317 in the first column to -1.680 in the third column, but it is statistically significant in both cases. This example demonstrates the failure of linear regression in high-dimensional settings. However, by relying on the high-dimensional capabilities of DML, we find support for the hypothesis of a positive association between the amount of other distracting earnings news (NRANK) and PEAD (see Hirshleifer et al. (2009)).

[17]The second specification uses a set of controls chosen ex ante, namely RUNUP, PASTRET, PRICE, SIZE, DOLVOL, BM, EXPRISK, ARBRISK, and ILLIQ, as explained in Section IV.A.

[18]The high number of observations used herein is consistent with the PEAD literature (e.g., see Bhushan (1994), Livnat and Mendenhall (2006), and DellaVigna and Pollet (2009)).

The DML results in the fourth column confirm the statistical significance of the interaction terms of reporting lag (REPLAG), a same fiscal quarter indicator (SAMEFIS), PASTRET, and NRANK. Diving further into the DML procedure, results show that a loss indicator if the last announcement was negative (LOSS) is positive and significant at the 1% level, with a larger magnitude than in the OLS specifications. In contrast to the OLS specifications, the DML procedure does not detect statistical significance for SIZE, number of shares traded (VOL), or announcement date abnormal returns (EARET). The nonsignificance of SIZE is surprising, because this is one of the oldest explanations of PEAD. In Section IV.C, we document how SIZE has been significant in subsamples spanning from 1998 to 2005, which may explain its earlier importance. As reported in Table 3 and illustrated in Figure C1 in the Supplementary Material, when comparing DML to the OLS specifications, we observe a noticeable change in coefficient magnitudes and increased standard errors.

To quantify the differences in coefficient estimates, we compute the standardized difference between two means using Cohen's d. Comparing the third and fourth columns, almost 60% (41%) of the variables show a Cohen's d above 1 (2), so they have a difference between the effects of more than 1 (2) standard deviations.

From Table 3, it is difficult to gauge whether inferences change because of the high-dimensional set of controls used or the DML procedure. Therefore, we restrict DML to incorporate the full set of controls, $\mathbf{Z}_i^{(2)}$, without stock-specific variables or second-order interaction terms (see Table C2 in the Supplementary Material). We find that standard errors generally increase because of this procedure, whereas coefficients change because of the exponential increase in controls, indicating that coefficient estimates should be more precise when accounting for the high-dimensional set of controls. With a more flexible structure and cross-fitting, it is not surprising that DML yields higher standard errors. Overall, DML provides a strong conservative baseline for statistical significance in high dimensions compared to traditional linear regression methods. Therefore, we argue that the credibility of these variables is thus strengthened.

The Sensitivity of OLS

To illustrate the sensitivity of the standard linear regression, we randomly permute the set of controls chosen ex ante. Figure 1 plots t-statistics from a linear regression where the set of controls is selected randomly (either 5, 10, or 20 different controls) from the 20 variables of interest and the 73 stock-specific variables.[19] The figure reports the results for SUE, SIZE × SUE, ANALYST × SUE, and BM × SUE using 500 different permutations of controls. Therefore, this test assumes that the researcher has access to the full set of variables but has no prior knowledge about which to choose. This illustration simulates a common hurdle that researchers face in many aspects of financial research.

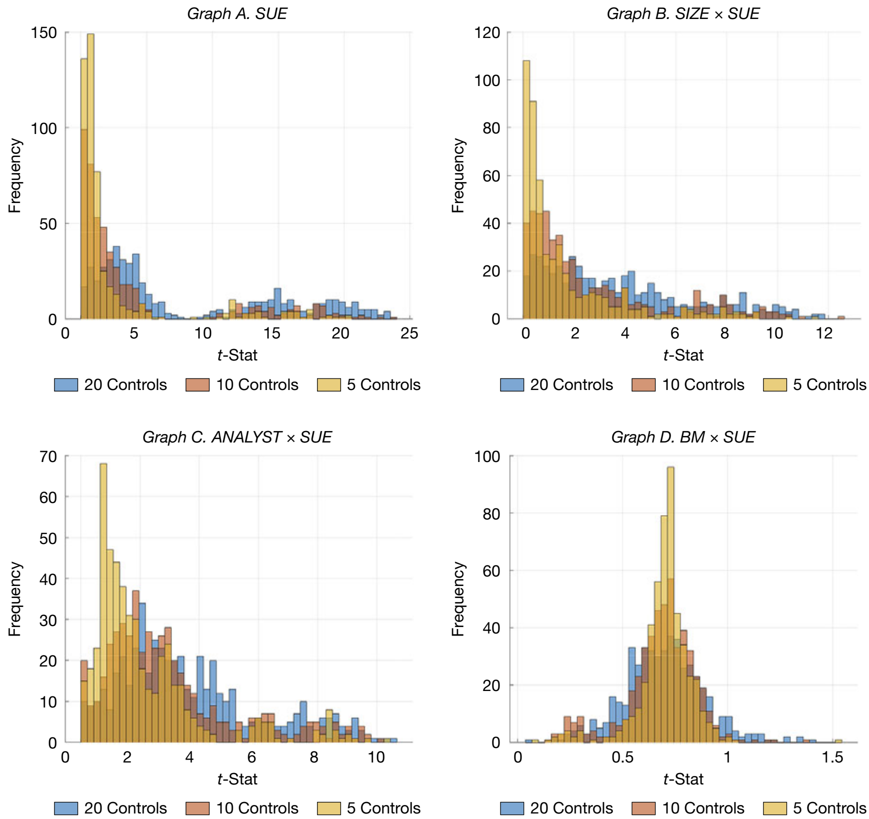
Figure 1 shows the instability of inferences when randomly permuting the set of controls. As demonstrated in Graph A, the inferences for SUE are very volatile

[19]The results remain similar when selecting randomly from the set of 20 variables of interest.

FIGURE 1

Sensitivity of OLS Inferences

Figure 1 plots *t*-statistics obtained with OLS for 4 different estimates using 500 different permutations of controls. Graphs A to D show the *t*-statistics for the listed coefficients from OLS while randomly selecting 5, 10, or 20 different controls from the variables of interest and the set of stock-specific variables, *S*.



and vary from nonsignificant to highly significant, with *t*-statistics well above 10. The *t*-statistics are found more frequently to be close to 0 when relying on 5 controls compared to 10 or 20. Conditioning on 5 controls, 19% of the permutations are significant at the 1% level, in contrast to 74% when conditioning on 20 controls. The same sensitivity is evident in Graphs B and C for SIZE × SUE and ANALYST × SUE, respectively. Even though the number of analysts following a stock (ANALYST) was found to be nonsignificant in all four specifications in Table 3, Figure 1 reveals that ANALYST can statistically significantly explain PEAD for 60% of the permutations using 20 variables.

Finally, Graph D of Figure 1 documents how BM × SUE is detected consistently as nonsignificant. This small illustration demonstrates a major challenge in the high-dimensional setting, that is, when relying on traditional statistical methods, hand-picking the "correct" set of controls can lead to unreliable inferences.

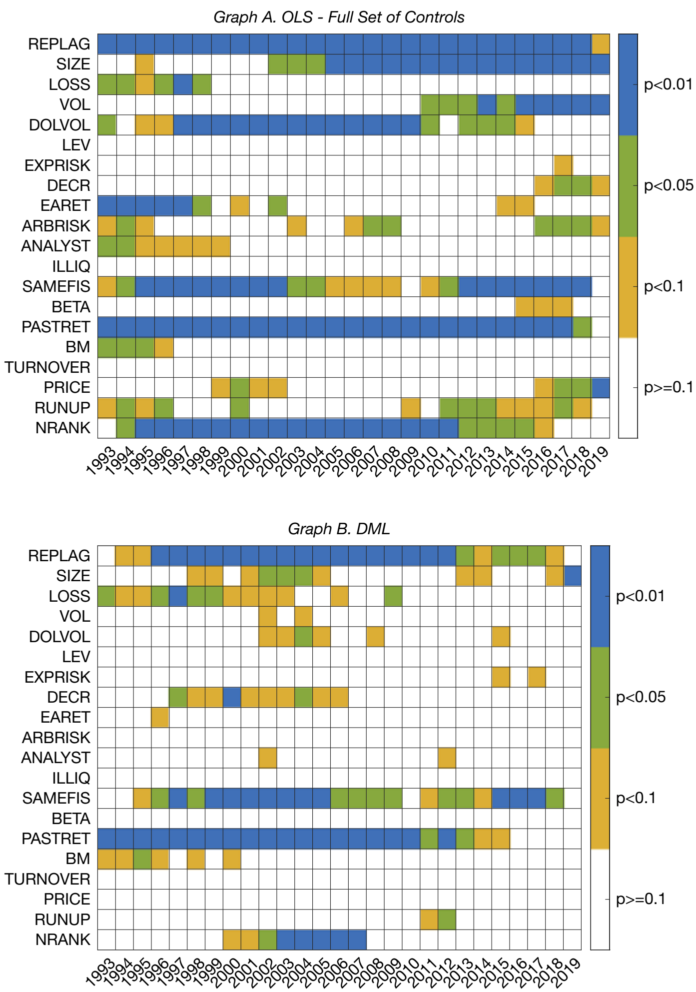
C. Inferences Through Time

A natural concern is the reliance on a sample spanning from 1984 to 2020, which includes years that were unavailable for researchers in the 1990s and 2000s. Therefore, this section examines systematically the consistency of variables' relevance through time to detect whether pockets of significance have occurred. We compute inferences based on 10 years of data and roll the window from 1993 to 2019. Graph A in Figure 2 reports significance using the full set of controls, whereas

FIGURE 2

Inferences Through Time

Figure 2 reports p-values from separate regressions based on 10 years of data. We use a rolling window from 1993 to 2019. Graph A uses OLS with the full set of controls, and Graph B uses DML. Both graphs consider the respective interaction term between SUE and the variable listed on the y-axis. Variables are listed according to Table 3. The colors illustrate the significance level: 1% (blue), 5% (green), 10% (yellow), and nonsignificant (white).



Graph B reports significance using DML.[20] The graphs show the p-value for each corresponding interaction term displayed on the vertical axis.

Graphs A and B of Figure 2 show that a small set of variables has consistently explained PEAD through all subsamples. Across both graphs, REPLAG, SAMEFIS, and PASTRET are highly statistically significant for the majority of the sample period, supporting the full sample evidence in Table 3. A direct comparison between the graphs reveals more conservative estimates for DML. For example, SIZE, DOLVOL, and NRANK are significant across the majority of the subsamples in Graph A, but not those in Graph B. Although Table 3 identifies LOSS and NRANK as significant based on DML, Graph B reveals only pockets of significance. Furthermore, pockets of significance are also found for SIZE, DOLVOL, a decreasing earnings between quarters indicator (DECR), and BM, which can explain earlier findings in the literature (see, e.g., Foster et al. (1984), Narayanamoorthy (2006), and Shivakumar (2006)).

Overall, the results in Graph B of Figure 2 supports that a small set of variables has consistently explained PEAD through time.

D. Controls that Matter

This subsection investigates the advantages of relying on the more conservative inferences obtained from DML when exploring a new high-dimensional set of potential variables. We question whether any of the 73 stock-specific variables from Green et al. (2017) are associated with PEAD. We expect a low signal-to-noise ratio because we investigate a high-dimensional set of potential controls with no former association to PEAD, and therefore we propose a 2-step scheme. First, a preinference step is conducted with the object of reducing dimensionality and increasing computational feasibility. We employ lasso's ability to perform variable selection on the following regression:

$$\text{CAR}_i^{[2,61]} = \mathbf{Z}_i\beta + \varepsilon_i, \tag{12}$$

where $\mathbf{Z}_i = \{1, \text{SUE}_i, \mathbf{X}_i, \text{SUE}_i \times \mathbf{X}_i, \mathbf{S}_i, \text{SUE}_i \times \mathbf{S}_i\}$ and $\beta = (\beta_1, \beta_2, \ldots, \beta_P)'$ is a P vector. A total of $P = 2 + 2 \times (20 + 73) = 188$ variables are included in $\mathbf{Z}_i$. As variables of interest, we consider all variables for which either the interaction term or the variable itself has a nonzero coefficient.[21]

In the second step, we follow the same procedure as in Section IV.B, where we conduct inferences on each nonzero variable from the preinference step using all four model specifications. Results for the nonzero stock-specific variables are summarized in Table 4, and coefficient estimates are reported in Table 5. Because each selected variable is estimated in a separate regression and conditioned on the same explanatory variables as in Section IV.B, the 20 estimated variables from that section have identical coefficients and are therefore excluded from the table.

[20] Results for the first and second model specifications and for both the interaction terms and variables themselves are presented in Figure C3 in the Supplementary Material.

[21] Note a naive approach relying on the estimated coefficients from lasso will be biased because of regularization.

TABLE 4

Summary of Inferences from Controls that Matter

The first 2 rows of Table 4 report the numbers of significant interaction terms – which are the parameters of primary interest – across the 4 model specifications at a 1% standard significance level and a Benjamini–Hochberg corrected level. The bottom 2 rows report the total numbers of significant interaction terms and variables of interest (noninteraction terms).

	OLS			DML w. Post-Lasso
Separate Regressions Using	No Controls	Selected Controls	Full Set of Controls	Nuisance Function
Significant interaction terms (1%)				
Standard significance level	4/11	3/11	2/11	5/11
B–H corrected	4/11	3/11	1/11	3/11
Significant variables (1%)				
Standard significance level	14/22	13/22	13/22	11/22
B–H corrected	14/22	13/22	12/22	8/22

TABLE 5

Controls that Matter

In Table 5, variables are listed based on the magnitude of the differences in coefficient estimates from the first and fourth columns. * denotes significance at the 1% level, and ** indicates changes in significance between the first and fourth columns. The specifications use all firm-quarters across the sample, and all control variables are continuous but standardized except for SUE, which is in deciles. The first numerical column reports estimates (in percentage) from separate regressions with no controls for the parameters of primary interest. The second column adds the set of controls chosen ex ante and their interactions with SUE, and the third column uses the full set of controls. The fourth column reports estimates using the high-dimensional nuisance function. The estimates for the variables of interest themselves (noninteraction terms) are reported in Table C3 in the Supplementary Material. Standard errors are clustered by day and firm.

	OLS			DML w. Post-Lasso
Separate Regressions Using	No Controls	Selected Controls	Full Set of Controls	Nuisance Function
SUE × NINCR**	−0.237	0.154	0.142	1.132*
SUE × INDMOM	1.126*	0.305	0.470	2.189*
SUE × BASPREAD**	−1.996	−0.841	−0.863	−0.964
SUE × MOM12	3.735*	2.455*	2.523*	2.766*
SUE × MOM1M**	0.546	−0.015	0.337	1.367*
SUE × IDIOVOL	−0.522	0.290	0.667	0.150
SUE × CHTX	−0.235	−0.197	−0.168	0.437
SUE × MOM36M	−0.332	0.179	0.072	0.241
SUE × AGR	−0.685	−0.303	−0.361	−0.434
SUE × STD_DOLVOL	0.988*	1.518*	1.116*	1.152*
SUE × MS	0.638	0.881*	0.386	0.670
No. of obs.	170,719	170,719	170,719	170,719
Fixed effects			Yes	Yes
No. of variables	3	21	138	2,839

The nonzero variables chosen by lasso in the first preinference step are associated with price trends, liquidity, and arbitrage risk. Specifically, the 11 newly selected stock-specific variables are the number of earnings increases (NINCR), industry momentum (INDMOM), bid–ask spread (BASPREAD), 12-month momentum (MOM12M), 1-month momentum (MOM1M), idiosyncratic volatility (IDIOVOL), change in tax expense (CHTX), 36-month momentum (MOM36M), asset growth (AGR), volatility of dollar trading volume (STD_DOLVOL), and financial statement score (MS).[22] The selected variables are consistent with findings of Gu et al. (2020), who find prominent roles for momentum, liquidity, and

[22]From the set of the initial 20 variables of interest SUE, PASTRET, EARET, SIZE, VOL, BM, ANALYST, leverage (LEV), share turnover (TURNOVER), NRANK, REPLAG, SAMEFIS, EXPRISK, and ARBRISK are found to be nonzero.

volatility in explaining general equity risk premia. Therefore, relying on the pre-inference step, we document a clear link between the variables explaining general return patterns and PEAD. In addition to the 4 dominant momentum variables, we find joint evidence of two well-established PEAD hypotheses, that is, illiquidity (see Chordia et al. (2009)) and limited arbitrage (see Mendenhall (2004)). Consistent with the first hypothesis, lasso detects VOL, TURNOVER, BASPREAD, and STD_DOLVOL, which proxy illiquidity (see Amihud and Mendelson (1989), Chordia, Subrahmanyam, and Anshuman (2001)). The second hypothesis of limited arbitrage is supported by the detection of the variables ARBRISK and IDIOVOL. Furthermore, in line with Gu et al. (2020), we find the last group of important variables to be valuation ratios and fundamental signals.

There is a possibility that the small set of variables selected by lasso could be a direct result of its sparsity assumption, that is, if two collinear variables predictive of the outcome are present, lasso tends to set one of them to 0. To address this concern, we rerun the analysis relying on the ridge method (Hoerl and Kennard (1970)) for the preinference step to ensure robustness. We find strong alignment between the 11 chosen variables from lasso and the size of the coefficients from ridge, suggesting lasso does not exclude any potentially essential variables. In addition to ridge, adaptive lasso can be a compelling alternative given its oracle properties (Zou (2006)). We find roughly identical results when comparing the set of nonzero variables between the 2 methods.

The summary for the second step is reported in Table 4, where DML identifies 5 of the 11 selected interaction terms as significant, which is slightly more than the OLS specifications. Because a minor set of interaction terms survives the 2-step procedure, significance for the remaining 93% stock-specific variables cannot be claimed.

Comparing DML with the OLS specifications in Table 5, we see a notable difference in the coefficient estimate of the significant variables (e.g., NINCR, INDMOM, and MOM1M). Although little consistency in the coefficient estimate is detected, 2 variables stand out with consistent estimates across model specifications, namely MOM12M and STD_DOLVOL. With large coefficient estimates for the momentum variables INDMOM, MOM12M, and MOM1M in the fourth column, we find support for the importance of price trends in explaining PEAD using DML. Hence, firms whose stock price has momentum prior to the earnings announcement experience a larger drift. Together, these results provide important insights into how inferences can be conducted when investigating the link between PEAD and a high-dimensional set of new potential explanations with no former association.

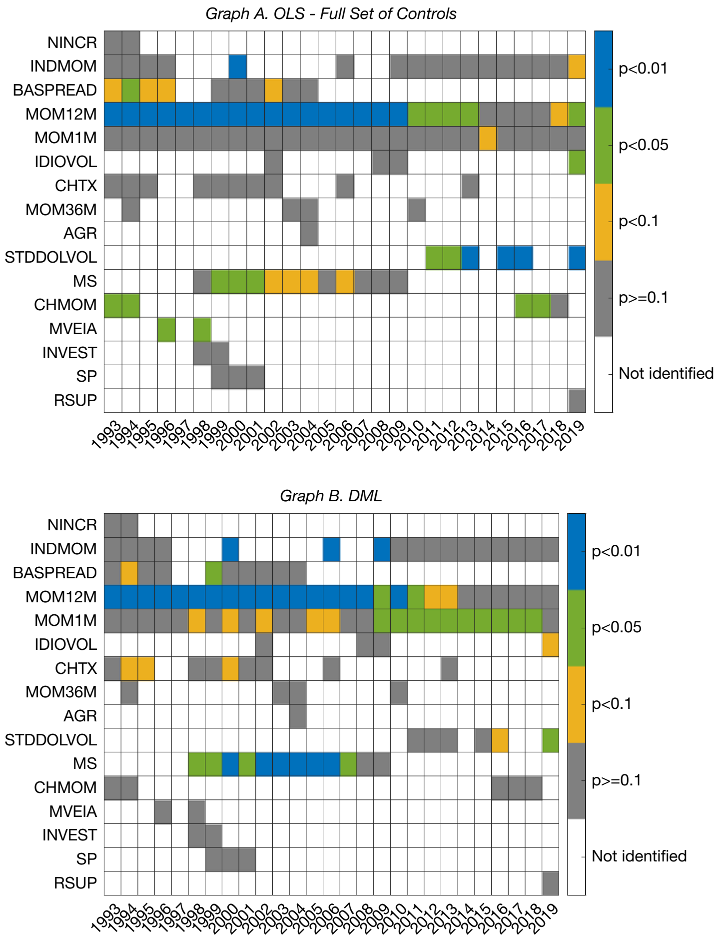
Controls that Matter Through Time

To strengthen the robustness of the results in the previous section, we examine whether either this new set of significant variables is only sporadically significant or a true association can be concluded. We conduct the same 2-step scheme as in Section IV.D, where inferences are obtained over 10 years and rolled from 1993 to 2019. Figure 3 reports the results based on OLS using the full set of controls in

FIGURE 3

Controls that Matter Through Time

Figure 3 reports *p*-values for stock-specific covariates from separate regressions based on 10 years of data using a rolling window from 1993 to 2019. Graph A uses OLS with the full set of controls, and Graph B uses DML. Both graphs consider the respective interaction terms between SUE and the variables listed on the vertical axis. Variables are listed according to Table 5. Colors illustrate levels of significance: 1% (blue), 5% (green), 10% (yellow), nonsignificant (gray), and not selected by lasso (white). We use 2-fold cross-fitting and obtain estimates using two splits.



Graph A and those based on DML in Graph B.[23] Graphs A and B show the significance of each interaction term between the variable of interest and SUE. White color denotes that the variable is not selected in a given subsample, and for selected variables, blue denotes significance at the 1% level, green denotes 5%, yellow denotes 10%, and gray denotes selected but not significant.

When considering the preinference step by lasso, a slight increase in the number of nonzero variables is reported when we split the data into 27 subsamples.

[23]The results for the first and second model specifications and for both the interaction terms and variables themselves are reported in Figure C7 in the Supplementary Material.

In contrast to the 11 variables detected in the full sample, 16 stock-specific variables are found to be nonzero in the 27 subsamples.[24] The new variables include change in 6-month momentum (CHMOM), industry-adjusted size (MVEIA), capital expenditures and inventory (INVEST), sales to price (SP), and revenue surprises (RSUP). Although a more extensive set of variables is found to be nonzero, 85% of the 73 variables are constantly set to 0, emphasizing that the majority of variables are detected as not important by lasso.

Turning to the second inference step of the 2-step scheme, some interesting patterns emerge. As demonstrated by the colored squares, variables consistently chosen by the first step are generally highly significant, whereas variables selected sporadically are generally not significant. Focusing on the significant variables in Graph B, MOM12M and MOM1M are consistently detected as significant, which is in line with the full sample results in Table 5. The variable MS was identified as nonsignificant in Section IV.D, but Figure 3 identifies it as significant from 1998 to 2009. Overall, compared to Section IV.D, we identify an even smaller set of variables that are consistently associated with PEAD, and these are related mainly to price trends.

E. Stability of Inferences

In this section, we examine the stability of our main results and test whether our main conclusion directly results from the specific model setup. First, we replace the post-lasso estimator with a random forest estimator when estimating the nuisance function to examine sensitivity to the choice of machine learning algorithm. Second, we investigate the impact of using various quantile ranks of variables, which is a method commonly used to facilitate interpretation of the effects.[25]

1. A Nonlinear Nuisance Function

Although we assume a PLM structure (see equation (1)), a linear relation between covariates is still imposed when estimating the nuisance function with the post-lasso method. This subsection addresses this drawback by allowing for a more flexible machine-learning model by replacing the post-lasso method with the random forest method.[26] We rely on the random forest method's ability to implicitly account for nonlinearities and interaction effects between explanatory variables.

By leveraging its capabilities, we are able to omit a priori construction of interaction terms from $\mathbf{Z}_i^{(3)}$ (see Section IV.A), so $\mathbf{Z}_{i,\mathrm{RF}}^{(3)}$ encompasses $135 + 73 = 208$ explanatory variables (i.e., 135 main variables and 73 stock-specific variables).[27] We summarize the results in Table 6 and report estimates in Table 7. To facilitate comparison to the results for OLS specifications and DML using post-lasso, the

[24]The original 20 variables of interest selected by the preinference step are shown in Figure C5 in the Supplementary Material.

[25]We report results using different measures of SUE, measurement horizons, and risk-adjustments of abnormal returns, and also consider changing the measurement period of stock-specific variables, in Appendix D of the Supplementary Material.

[26]The DML procedure provides valid inferences for a broad range of machine learning methods, such as random forest, boosted trees, and neural networks, as shown by Chernozhukov et al. (2018).

[27]We estimate random forest using its plain vanilla settings of Breiman (2001) where the number of trees is set to 400.

TABLE 6

Summary of Inferences from Random Forest

The first 2 rows of Table 6 report the numbers of significant interaction terms – which are the parameters of primary interest – across the 4 model specifications at a 1% standard significance level and a Benjamini–Hochberg corrected level. The bottom 2 rows report the total numbers of significant interaction terms and variables of interest (noninteraction terms).

Separate Regressions Using	OLS Selected Controls	DML w. Post-Lasso Nuisance Function	DML w. Random Forest Nuisance Function
Significant interaction terms (1%)			
Standard significance level	8/20	5/20	5/20
B–H corrected	5/20	4/20	4/20
Significant variables (1%)			
Standard significance level	23/41	13/41	11/41
B–H corrected	20/41	12/41	9/41

TABLE 7

Estimates of Interaction Terms Using Random Forest

In Table 7, variables are listed based on the magnitude of the differences in coefficient estimates from the first and third columns. * denotes significance at the 1% level, and ** indicates changes in significance between the first and fourth columns. The specifications use all firm-quarters in the sample, and each control variable is continuous but standardized except for SUE, which is in deciles. The first numerical column reports estimates (in percentage) from separate regressions with the set of ex ante chosen controls and their interactions with SUE for the parameters of primary interest. The second column reports the estimates using the high-dimensional nuisance function estimated using post-lasso, and the third column is estimated using random forest. The estimates for the variables of interest themselves (noninteraction terms) are reported in Table C4 in the Supplementary Material. Standard errors are clustered by day and firm.

Separate Regressions Using	OLS Selected Controls	DML w. Post-Lasso Nuisance Function	DML w. Random Forest Nuisance Function
SUE × REPLAG	−3.264*	−2.051*	−1.627*
SUE × EXPRISK	0.563	−0.383	0.278
SUE × ARBRISK	−1.464	−0.886	−0.213
SUE × BETA	−0.710	−1.183	−0.409
SUE × EARET**	0.905*	0.548	0.390
SUE × SIZE**	−1.792*	−1.338	−0.697
SUE × LOSS	−0.343	−1.170*	−0.746
SUE × VOL**	1.111*	−0.817	−0.393
SUE × RUNUP**	0.746	0.719	1.411*
SUE × ANALYST**	0.043	−0.873	−1.020*
SUE × NRANK**	0.766*	1.317*	0.624
SUE × DECR	0.173	−0.624	−0.303
SUE × PASTRET	3.225*	2.973*	2.758*
SUE × LEV	−0.018	−0.499	0.070
SUE × ILLIQ	0.178	−0.078	0.046
SUE × DOLVOL**	0.893*	−0.418	0.396
SUE × SAMEFIS	1.654*	1.925*	1.816*
SUE × TURNOVER	0.369	0.009	−0.156
SUE × PRICE	1.226	0.969	0.826
SUE × BM	−0.506	−0.751	−0.412
No. of obs.	170,719	170,719	170,719
Fixed effects		Yes	Yes
No. of variables	21	2,839	211

first and second columns of Table 6 are identical to the second and fourth columns, respectively, of Table 2.

As reported in Table 6, post-lasso and random forest identify the same number of significant interactions terms (i.e., 1 out of 4). Therefore, these results support how the conservatism of the DML procedure is invariant to the choice of machine learning method.

When investigating the coefficient estimates in Table 7, we find variations between random forest and post-lasso results. Although post-lasso and random forest agree about the significance of REPLAG, PASTRET, and SAMEFIS, random forest does not identify LOSS or NRANK as significant. Instead, RUNUP and ANALYST are identified as highly significant. Interestingly, these variables are not found to be significant by any of the OLS specifications. In line with post-lasso, random forest finds a large set of well-established variables to be nonsignificant, such as firm beta (BETA), SIZE, or BM. The inconsistency between the post-lasso and random forest methods is not surprising: Post-lasso assumes that a sparse set of control variables linearly affects the variable of interest and the explanatory variable, whereas random forest allows these relationships to be highly complex and nonlinear.

Although the two machine learning methods estimate the nuisance function differently, they agree about a more conservative number of significant variables compared to the more traditional OLS specifications.

2. Quantile Rank of Variables

To facilitate interpretation of the magnitude of estimated PEAD effects, it is common to transform data into deciles (see, e.g., Bhushan (1994), Mendenhall (2004), Garfinkel and Sokobin (2006), and Hirshleifer et al. (2009)). Therefore, Tables 8 and 9 investigate whether variables remain significant when relying on decile ranks instead of continuous values. Note that some information will be lost when such a transformation is conducted, which may impact the conclusions. Following the literature, we form deciles using yearly decile breakpoints to allow for potential time trends.

Comparing the first 3 numerical columns of Table 8 to the main results in Table 2, we observe strong variation in the number of significant variables across all OLS specifications. Whereas the first column of Table 2 finds 8 variables to be significant, Table 8 reports a notable 62% increase, with the first specification identifying 13 variables as significant. Inconsistent results are also reported for the second and third columns, where a decrease of at least 25% is detected between Tables 2 and 8 for the two specifications (from 8 to 6 and 7 to 5 variables, respectively).

We find similar results when transforming the data into quintile ranks and when restricting the data to the 10th and 1st deciles (see Tables C6 and C7 in the Supplementary Material). In contrast, strong robustness is detected for the DML

TABLE 8

Summary of Inferences Using Deciles

The first 2 rows of Table 8 report the numbers of significant interaction terms – which are the parameters of primary interest – across the 4 model specifications at a 1% standard significance level and a Benjamini–Hochberg corrected level. The bottom 2 rows report the total numbers of significant interaction terms and variables of interest (noninteraction terms).

	OLS			DML w. Post-Lasso
Separate Regressions Using	No Controls	Selected Controls	Full Set of Controls	Nuisance Function
Significant interaction terms (1%)				
Standard significance level	13/20	6/20	5/20	5/20
B–H corrected	13/20	5/20	5/20	4/20
Significant variables (1%)				
Standard significance level	30/41	23/41	23/41	16/41
B–H corrected	30/41	22/41	22/41	15/41

https://doi.org/10.1017/S0022109023000133 Published online by Cambridge University Press

TABLE 9

Estimates Using Deciles

In Table 9, interaction-term estimates for each decile-transformed control variable. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and fourth columns. * denotes significance at the 1% level, and ** indicates changes in significance between the first and fourth columns. The specifications use all firm-quarters in the sample, and deciles are computed using yearly breakpoints. The first column reports estimates (in percentage) from separate regressions with no controls for the parameters of primary interest. The second column adds the set of controls chosen ex ante and their interactions with SUE, and the third column uses the full set of controls. The fourth column reports estimates using the high-dimensional nuisance function. The estimates for the variables of interest themselves (noninteraction terms) are reported in Table C6 in the Supplementary Material. Standard errors are clustered by day and firm.

	OLS			DML w. Post-Lasso
Separate Regressions Using	No Controls	Selected Controls	Full Set of Controls	Nuisance Function
SUE × PASTRET	1.929*	1.530*	1.540*	3.299*
SUE × VOL**	−1.227*	−1.584	−1.877	0.117
SUE × REPLAG	−3.753*	−3.403*	−3.462*	−2.417*
SUE × LOSS**	0.077	−0.557	−0.599	−1.092*
SUE × EXPRISK**	−1.464*	−0.111	0.016	−0.524
SUE × ILLIQ**	0.890*	1.285	0.556	−0.038
SUE × EARET**	1.293*	0.342	0.451	0.404
SUE × DECR	0.197	0.158	0.111	−0.619
SUE × BETA**	−1.696*	−0.267	−0.465	−0.885
SUE × DOLVOL**	−1.048*	0.033	0.499	−0.383
SUE × LEV	0.302	−0.213	0.075	−0.307
SUE × PRICE	−0.300	−1.167*	−1.976*	0.273
SUE × ANALYST**	−1.114*	−0.651	−0.641	−0.604
SUE × SIZE	−0.302	−0.098	0.425	0.181
SUE × BM	−0.471	−0.621	−0.773	−0.948
SUE × ARBRISK**	−1.382*	−1.750*	−0.803	−0.906
SUE × TURNOVER**	−1.138*	−0.131	0.172	−0.743
SUE × SAMEFIS	1.504*	1.805*	1.199*	1.890*
SUE × NRANK	1.317*	0.714*	−1.207*	1.246*
SUE × RUNUP	0.786	0.376	0.608	0.841
No. of obs.	170,719	170,719	170,719	170,719
Fixed effects			Yes	Yes
No. of variables	3	21	138	2,839

procedure, where a quarter of variables are identified as statistically significant, identical to the main results. The coefficient estimates in the fourth column of Table 9 show even greater robustness because estimates vary only slightly compared to the main results in Table 3. Overall, we find that across time and transformations, only PASTRET, REPLAG, and SAMEFIS have consistently been significant for DML when using both post-lasso and random forest. Hence, 17 variables do not survive the more conservative DML procedure and our broad set of robustness checks. Interestingly, well-established variables such as BETA, SIZE, and BM do not survive.

V. Conclusion

This article combines a high-dimensional inference technique based on the machine-learning literature with traditional hypothesis-driven research to demonstrate how valid statistical inference can be conducted in a high-dimensional setting. As well as inferences in high dimensions, the DML procedure of Chernozhukov et al. (2018) can control for a large set of covariates and reduce researcher dependency. As a prominent example, we revisit an unresolved question in finance namely the origins of the PEAD. The empirical literature striving to explain PEAD has uncovered what we term a "zoo of controls." There is little academic consensus on model design, and researchers rely on data sets comprising thousands of earnings announcements, which has led to an environment in which researchers can detect complex effects with little practical use. After more than 60 years of research and over 216 published papers (Fink (2020)), taming the zoo has so far been neglected. Therefore, PEAD serves as a strong showcase of the need to move toward high-dimensional methods to reduce researcher dependency and strengthen the credibility of explanations.

In the reported study, we conduct a comprehensive comparison between DML and three standard OLS specifications. To ensure comparability, each chosen OLS specification has been applied broadly in the PEAD literature, that is, i) an OLS regression with zero controls, ii) a regression that allows for a small subset of control variables chosen ex ante, and iii) a final regression that includes a multitude of controls. In contrast to the general PEAD literature and to ensure the robustness of our results, we conduct a comprehensive study of 20 different variables of interest, which we compare across models. Our concern becomes clear when investigating the inferences from the three OLS specifications. Of 20 variables, 17 were statistically significantly associated with either variation in PEAD or cumulative abnormal returns for the simplest model. To stress our concern further, we demonstrate how inferences are highly sensitive to the choice of controls by permuting the set in a linear regression. This shows that researchers can either implicitly or explicitly choose a set of controls to support their hypothesis. By leveraging the high-dimensional capabilities of the DML procedure, we extended the set of controls by adding 73 stock-specific variables from the cross-section of the stock returns literature (Green et al. (2017)), all first-order interaction terms, and fixed effects, yielding a total of 2,836 controls. With a 28% reduction in the number of variables explaining the drift, we demonstrate how the more conservative DML procedure can strengthen the credibility of a small set of factors. A direct comparison between DML and the OLS specifications illustrates that the coefficient estimates change considerably for several variables. Thus, we conjecture that an omitted-variable bias distorts the coefficient estimates from the standard OLS procedure. A vast set of robustness checks is conducted to strengthen our results, including analyzing inferences through time, using various quantile ranks of variables, and allowing for nonlinearities in the DML procedure.

When exploring the additional 73 stock-specific variables using a 2-step scheme, we find prominent roles for momentum, liquidity, and volatility, consistent with the findings of Gu et al. (2020). Therefore, these variables are crucial in explaining not only general equity risk premia but also abnormal returns around earnings announcements.

To the best of our knowledge, this study is the first to leverage the capabilities of DML to strengthen the credibility of a small set of PEAD explanations. Our hope is that this article highlights the dangers of hand-picking a small set of controls and helps to establish a new standard for testing new explanations of PEAD. Furthermore, we envision this article joining a new line of papers combining statistical and computational techniques with hypothesis-driven research to increase the credibility of existing findings and theory.

Supplementary Material

To view supplementary material for this article, please visit http://doi.org/10.1017/S0022109023000133.

References

Amihud, Y., and H. Mendelson. "The Effects of Beta, Bid–Ask Spread, Residual Risk, and Size on Stock Returns." *Journal of Finance*, 44 (1989), 479–486.

Athey, S.; G. W. Imbens; and S. Wager. "Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 80 (2018), 597–623.

Ball, R., and P. Brown. "An Empirical Evaluation of Accounting Income Numbers." *Journal of Accounting Research*, 6 (1968), 159–178.

Bartov, E.; S. Radhakrishnan; and I. Krinsky. "Investor Sophistication and Patterns in Stock Returns After Earnings Announcements." *Accounting Review*, 75 (2000), 43–63.

Belloni, A.; D. Chen; V. Chernozhukov; and C. Hansen. "Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain." *Econometrica*, 80 (2012), 2369–2429.

Belloni, A., and V. Chernozhukov. "Least Squares After Model Selection in High-Dimensional Sparse Models." *Bernoulli*, 19 (2013), 521–547.

Belloni, A.; V. Chernozhukov; I. Fernández-Val; and C. Hansen. "Program Evaluation and Causal Inference with High-Dimensional Data." *Econometrica*, 85 (2017), 233–298.

Belloni, A.; V. Chernozhukov; and C. Hansen. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives*, 28 (2014a), 29–50.

Belloni, A.; V. Chernozhukov; and C. Hansen. "Inference on Treatment Effects After Selection Among High-Dimensional Controls." *Review of Economic Studies*, 81 (2014b), 608–650.

Benjamini, Y., and Y. Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society. Series B (Methodological)*, 57 (1995), 289–300.

Bernard, V. L., and J. K. Thomas. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research*, 27 (1989), 1–36.

Bernard, V. L., and J. K. Thomas. "Evidence that Stock Prices Do Not Fully Reflect the Implications of Current Earnings for Future Earnings." *Journal of Accounting and Economics*, 13 (1990), 305–340.

Bhushan, R. "An Informational Efficiency Perspective on the Post-Earnings Announcement Drift." *Journal of Accounting and Economics*, 18 (1994), 45–65.

Breiman, L. "Random Forests." *Machine Learning*, 45 (2001), 5–32.

Carhart, M. M. "On Persistence in Mutual Fund Performance." *Journal of Finance*, 52 (1997), 57–82.

Chernozhukov, V.; D. Chetverikov; M. Demirer; E. Duflo; C. Hansen; W. Newey; and J. Robins. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *Econometrics Journal*, 21 (2018), C1–C68.

Chordia, T.; A. Goyal; G. Sadka; R. Sadka; and L. Shivakumar. "Liquidity and the Post-Earnings-Announcement Drift." *Financial Analysts Journal*, 65 (2009), 18–32.

Chordia, T.; A. Subrahmanyam; and V. R. Anshuman. "Trading Activity and Expected Stock Returns." *Journal of Financial Economics*, 59 (2001), 3–32.

Christensen, K.; M. Siggaard; and B. Veliyev. "A Machine Learning Approach to Volatility Forecasting." *Journal of Financial Econometrics*, forthcoming (2023).

Cochrane, J. H. "Presidential Address: Discount Rates." *Journal of Finance*, 66 (2011), 1047–1108.

DellaVigna, S., and J. M. Pollet. "Investor Inattention and Friday Earnings Announcements." *Journal of Finance*, 64 (2009), 709–749.

Fama, E. F. "Market Efficiency, Long-Rerm Returns, and Behavioral Finance." *Journal of Financial Economics*, 49 (1998), 283–306.

Feng, G.; S. Giglio; and D. Xiu. "Taming the Factor Zoo: A Test of New Factors." *Journal of Finance*, 75 (2020), 1327–1370.

Fink, J. "A Review of the Post-Earnings-Announcement Drift." *Journal of Behavioral and Experimental Finance*, 29 (2020), 100446.

Foster, G.; C. Olsen; and T. Shevlin. "Earnings Releases, Anomalies, and the Behavior of Security Returns." *Accounting Review*, 59 (1984), 574–603.

Francis, J.; R. Lafond; P. Olsson; and K. Schipper. "Information Uncertainty and Post-Earnings Announcement Drift." *Journal of Business Finance & Accounting*, 34 (2007), 403–433.

Freyberger, J.; A. Neuhierl; and M. Weber. "Dissecting Characteristics Nonparametrically." *Review of Economic Studies*, 33 (2020), 2326–2377.

Garfinkel, J. A., and J. Sokobin. "Volume, Opinion Divergence, and Returns: A Study of Post-Earnings Announcement Drift." *Journal of Accounting Research*, 44 (2006), 85–112.

Giglio, S., and D. Xiu. "Asset Pricing with Omitted Factors." *Journal of Political Economy*, 129 (2021), 1947–1990.

Gow, I. D.; D. F. Larcker; and P. C. Reiss. "Causal Inference in Accounting Research." *Journal of Accounting Research*, 54 (2016), 477–523.

Green, J.; J. R. Hand; and X. F. Zhang. "The Characteristics that Provide Independent Information About Average U.S. Monthly Stock Returns." *Review of Economic Studies*, 30 (2017), 4389–4436.

Gu, S.; B. Kelly; and D. Xiu. "Empirical Asset Pricing via Machine Learning." *Review of Economic Studies*, 33 (2020), 2223–2273.

Harvey, C. R.; Y. Liu; and H. Zhu. "… and the Cross-Section of Expected Returns." *Review of Economic Studies*, 29 (2016), 5–68.

Hirshleifer, D.; S. S. Lim; and S. H. Teoh. "Driven to Distraction: Extraneous Events and Underreaction to Earnings News." *Journal of Finance*, 64 (2009), 2289–2325.

Hoerl, A. E., and R. W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems." *Technometrics*, 12 (1970), 55–67.

Javanmard, A., and A. Montanari. "Confidence Intervals and Hypothesis Testing for High-Dimensional Regression." *Journal of Machine Learning Research*, 15 (2014), 2869–2909.

Jegadeesh, N., and J. Livnat. "Post-Earnings-Announcement Drift: The Role of Revenue Surprises." *Financial Analysts Journal*, 62 (2006), 22–34.

Jiang, G.; C. M. Lee; and Y. Zhang. "Information Uncertainty and Expected Returns." *Review of Accounting Studies*, 10 (2005), 185–221.

Keuzenkamp, H. A., and J. R. Magnus. "On Tests and Significance in Econometrics." *Journal of Econometrics*, 67 (1995), 5–24.

Kim, J. H. "Stock Returns and Investors' Mood: Good Day Sunshine or Spurious Correlation?" *International Review of Financial Analysis*, 52 (2017), 94–103.

Kim, J. H., and P. I. Ji. "Significance Testing in Empirical Finance: A Critical Review and Assessment." *Journal of Empirical Finance*, 34 (2015), 1–14.

Kormendi, R., and R. Lipe. "Earnings Innovations, Earnings Persistence, and Stock Returns." *Journal of Business*, 60 (1987), 323–345.

Lee, J. D.; D. L. Sun; Y. Sun; and J. E. Taylor. "Exact Post-Selection Inference, with Application to the Lasso." *Annals of Statistics*, 44 (2016), 907–927.

Lei, J.; M. G'Sell; A. Rinaldo; R. J. Tibshirani; and L. Wasserman. "Distribution-Free Predictive Inference for Regression." *Journal of the American Statistical Association*, 113 (2018), 1094–1111.

Lin, M.; H. C. Lucas Jr.; and G. Shmueli. "Research Commentary – Too Big to Fail: Large Samples and the p-Value Problem." *Information Systems Research*, 24 (2013), 906–917.

Livnat, J., and R. R. Mendenhall. "Comparing the Post-Earnings Announcement Drift for Surprises Calculated from Analyst and Time Series Forecasts." *Journal of Accounting Research*, 44 (2006), 177–205.

MacKinlay, A. C. "Event Studies in Economics and Finance." *Journal of Economic Literature*, 35 (1997), 13–39.

McCloskey, D. N., and S. T. Ziliak. "The Standard Error of Regressions." *Journal of Economic Literature*, 34 (1996), 97–114.

Mendenhall, R. R. "Evidence on the Possible Underweighting of Earnings-Related Information." *Journal of Accounting Research*, 29 (1991), 170–179.

Mendenhall, R. R. "Arbitrage Risk and Post-Earnings-Announcement Drift." *Journal of Business*, 77 (2004), 875–894.

Narayanamoorthy, G. "Conservatism and Cross-Sectional Variation in the Post-Earnings Announcement Drift." *Journal of Accounting Research*, 44 (2006), 763–789.

Nickl, R., and S. Van De Geer. "Confidence Sets in Sparse Regression." *Annals of Statistics*, 41 (2013), 2852–2876.

Petersen, M. A. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." *Review of Economic Studies*, 22 (2009), 435–480.

Roberts, M. R., and T. M. Whited, "Endogeneity in Empirical Corporate Finance." In George M. Constantinides, and Milton Harris, eds. *Handbook of the Economics of Finance*, Vol. 2. Amsterdam: Elsevier (2013), 493–572.

Shivakumar, L. "Accruals, Cash Flows and the Post-Earnings-Announcement Drift." *Journal of Business Finance & Accounting*, 33 (2006), 1–25.

Van de Geer, S.; P. Bühlmann; Y. Ritov; and R. Dezeure. "On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models." *Annals of Statistics*, 42 (2014), 1166–1202.

Whited, R. L.; Q. T. Swanquist; J. E. Shipman; and J. R. Moon. "Out of Control: The (over) Use of Controls in Accounting Research." *Accounting Review*, 97 (2022), 395–413.

Yang, J.; H. Chuang; and C. Kuan. "Double Machine Learning with Gradient Boosting and its Application to the Big N Audit Quality Effect." *Journal of Econometrics*, 216 (2020), 268–283.

Zhang, C., and S. S. Zhang. "Confidence Intervals for Low Dimensional Parameters in High Dimensional Linear Models." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 76 (2014), 217–242.

Zou, H. "The Adaptive Lasso and its Oracle Properties." *Journal of the American Statistical Association*, 101 (2006), 1418–1429.