Internet Appendix

A Methodology

A.1. DML

The section aims to explain in more detail the theory behind the DML procedure and how valid inference can be obtained by changing the moment condition. We assume the same partial linear model (PLM) as in equations (1) and (2). Firstly, as illustrated in the following naive example, by applying machine learning for approximating $g_0(\boldsymbol{Z})$, a convergence rate slower $1/\sqrt{n}$ can be caused, which results in a regularization bias. To illustrate this point, assume the sample size N is split up into two equal sized parts, the main part $n = N/2$ and an auxiliary part of size $N - n$, which is denoted by $i \in I^c$. Then estimating $\hat{g}_0(\boldsymbol{Z})$ by machine learning in the auxiliary sample and using the main sample to estimate $\hat{\theta}_0$ by least squares, as:

$$\hat{\theta}_0 = \left(\frac{1}{n}\sum_{i\in I} X_i^2\right)^{-1}\left(\frac{1}{n}\sum_{i\in I} X_i\left(Y_i - \hat{g}_0\left(\boldsymbol{Z}_i\right)\right)\right).$$

The slower convergence rate can be seen by decomposing the scaled estimation error as follows:

$$\sqrt{n}(\hat{\theta}_0 - \theta) = \underbrace{\left(\frac{1}{n}\sum_{i\in I} X_i^2\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} X_i\varepsilon_i}_{:=a} + \underbrace{\left(\frac{1}{n}\sum_{i\in I} X_i^2\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} X_i\left(g_0\left(\boldsymbol{Z}_i\right) - \hat{g}_0\left(\boldsymbol{Z}_i\right)\right)}_{:=b}.$$

Under relatively mild conditions, a is asymptotically normally distributed. Hence, if b converges to zero in probability, we have the desired properties. However, the challenge occurs since b, which is given by:

$$b = \left(E\left[X_i^2\right]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I} m_0(\boldsymbol{Z}_i)\left(g_0\left(\boldsymbol{Z}_i\right) - \hat{g}_0\left(\boldsymbol{Z}_i\right)\right) + o_p(1),$$

is typically a sum of n terms that do not have mean zero and hence, by dividing with $\sqrt{n}$, it will not converge in probability. To clarify, when dealing with high-dimensional or highly complex data sets, using regularization is key for informative learning. Balancing the trade-off between variance and bias using regularization is often assumed to be the main driver behind machine learning's superior performance in complex data sets. However, by relying on regularization, a converge rate of $n^{-\varphi}$ with $\varphi < 1/2$ is often triggered. In this naive example, it would lead to a "regularization bias", given by $g_0(\boldsymbol{Z}_i) - \hat{g}_0(\boldsymbol{Z}_i)$, of order $O_p(n^{-\varphi_g})$, and with $\varphi_g < 1/2$, which will cause the expectation for b to be of stochastic order $\sqrt{n}n^{-\varphi_g} \to \infty$. With this "inferior" rate of convergence of θ_0, it is clear that an alternative structure has to be considered compared to the naive approach.

To overcome the "inferior" rate of convergence, Chernozhukov et al. (2018) propose an alternative procedure named DML for estimating θ_0. By reformulating the loss function, the regularization bias can be accounted for, and a consistent estimate of θ_0 can be obtained. The main idea is to partial out the effect of $\boldsymbol{Z}$ from both Y and X. Proof and simulation results can be found in Chernozhukov et al. (2018), but the following part will highlight the main and most important points.

Consider the new representation of the PLM given by equation (3) in Section II.A and defining $V_i = X_i - m_0(\boldsymbol{Z}_i) = \nu_i$ and $W_i = Y_i - \ell_0(\boldsymbol{Z}_i)$, then an estimate of θ_0 can be found by:

$$\check{\theta}_0 = \left(\frac{1}{n}\sum_{i\in I}\hat{V}_i^2\right)^{-1}\left(\frac{1}{n}\sum_{i\in I}\hat{V}_i\hat{W}_i\right).$$

In contrast to the naive approach, this orthogonalization will be root-n consistent and approximately Gaussian under a very mild set of conditions. In line with the naive approach, it is possible to decompose the scaled estimation error of $\check{\theta}_0$:

$$\sqrt{n}\left(\check{\theta}_0-\theta_0\right)=\left(\frac{1}{n}\sum_{i\in I}V_i^2\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i\in I}\left[\hat{V}_i\hat{W}_i-V_i\left(W_i-\varepsilon_i\right)\right]\right).$$
$$=a^*+b^*+c^*.$$

In line with the arguments for the naive approach, the leading term given by a^* will under mild conditions be asymptotically normally distributed. The regularization bias term b^*, given as:

$$b^*=\left(E\left[X_i^2\right]\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I}\left(\hat{m}_0\left(Z_i\right)-m_0\left(Z_i\right)\right)\left(\hat{g}_0\left(\boldsymbol{Z}_i\right)-g_0\left(\boldsymbol{Z}_i\right)\right),$$

differs from the naive approach since it depends on the product of the estimator errors from both $\hat{m}_0\left(\boldsymbol{Z}_i\right)$ and $\hat{g}_0\left(\boldsymbol{Z}_i\right)$. Consistent with the arguments above, the convergence rates of $\hat{m}_0$ and $\hat{g}_0$ are respectively $n^{-\varphi_m}$ and $n^{-\varphi_g}$, causing b^* to have an upper bound of $\sqrt{n}n^{-\left(\varphi_m+\varphi_g\right)}$. Although φ_m and φ_g are found to be below 1/2, the product is typically found to be above, and hence, $\check{\theta}_0$ has good properties even under relatively slow rates of convergence of m_0 and g_0. The final term, c^*, includes terms as:

$$\left(\frac{1}{n}\sum_{i\in I}V_i^2\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i\in I}V_i\left(\hat{g}_0\left(\boldsymbol{Z}_i\right)-g_0\left(\boldsymbol{Z}_i\right)\right),$$

where correlation can induce an overfitting bias. As suggested in Chernozhukov et al. (2018), the use of sample splitting can ensure that c^* vanishes in probability. By estimating $\hat{g}_0$ in the auxiliary sample, and ε_i and ν_i are the errors in the main sample, the equation would vanish in probability by Chebyshev's inequality.

By applying the orthogonalized formulation and using data splitting, both the obstacle with regularization and overfitting bias can be accounted for, which results in $\sqrt{n}\left(\check{\theta}_0-\theta_0\right)$ being asymptotically normally distributed. However, by applying data splitting, a loss of efficiency is caused by the substantial loss of data when estimating $\hat{\theta}$. As suggested in Chernozhukov et al. (2018), full efficiency can be gained by flipping the role of the main and auxiliary samples and obtaining two estimates of the parameter of interest. Since these two estimators are approximately independent, full efficiency is obtained when averaging.

A.2. Machine Learning Methods

A.2.1 Lasso

Relying on a linear regression model in settings with a low signal-to-noise ratio, then increasing the number of predictors will eventually make the model fit noise rather than relevant information, also known as overfitting. To alleviate overfitting and improve the fit, a penalty term can be added to the least squares loss function, $\mathcal{L}(\beta)$:

$$\tilde{\mathcal{L}}(\beta;\gamma) = \mathcal{L}(\beta) + \phi(\beta;\gamma), \tag{13}$$

where $\phi(\beta;\gamma)$ is the penalty term, $\beta = (\beta_0, \beta_1, ..., \beta_P)'$ is a $P+1$ vector, P is the number of covariates, and γ is a scalar or a vector of hyperparameters.

One of the most widely used penalty methods is lasso by Tibshirani (1996). However, to ensure theoretical justification of the performance, we consider the lasso estimator with a data-driven penalty loading:

$$\phi(\beta;\gamma) = \frac{\lambda}{n}\left\|\hat{\Psi}\beta\right\|_1, \tag{14}$$

where $\|\beta\|_1 = \sum_{p=1}^{P}\left|\beta_p\right|$, $\hat{\Psi} = \text{diag}\left(\hat{\psi}_1 ..., \hat{\psi}_p\right)$ is a diagonal matrix of penalty loadings, and λ controls the amount of shrinkage. This penalty term will shrink irrelevant covariates to zero and thereby generate sparsity. When using lasso, a well-documented finite-sample bias will occur. Instead, Belloni and Chernozhukov (2013) suggest using post-lasso, which performs at least as well as lasso in terms of rate of convergence and has a smaller bias. As post-lasso has more desirable statistical properties, the main analysis of this paper will be based on post-lasso. Specifically, it is a two-step procedure, where in the first step, lasso is applied as a selection model, selecting $\hat{T}$ variables, hence:

$$\hat{T} = \text{support}\left(\hat{\beta}\right) = \left\{p \in \{1, ..., P\} : \left|\hat{\beta}\right| > 0\right\}.$$

Hereafter, a second step where an unpenalized estimate based on least squares, using only variables selected to have nonzero coefficients, is re-fitted, hence $\beta_p = 0$ if $\hat{\beta}_p = 0$.

The coefficient estimates heavily relies on the amount of penalization, and therefore the choice of the penalization parameter λ becomes critical. In this paper, we follow Belloni et al. (2012) and set the penalty loading to $\hat{\psi}_p = \sqrt{E_n\left[z_{i,p}^2 \hat{\varepsilon}_i^2\right]}$, while allowing for potential heteroscedasticity in ε_i. The penalty level is given as:

$$\lambda = 2c\sqrt{n}\Phi^{-1}\left(1 - \gamma/\left(2P\right)\right),$$

where $c = 1.1$ and $\gamma = 0.1/\log(n)$. The estimates of the residuals $\hat{\varepsilon}_i$ are initialized by running least squares of Y_i on five regressions, which are found to be highly correlated with Y_i. Hereafter, $\hat{\varepsilon}_i = Y_i - \boldsymbol{Z}_i\hat{\beta}$ is updated and the procedure is repeated 15 times. In contrast to cross-validation, the data-driven penalty loadings have theoretically justified performance.

B Variable Description

Variable Description

Variable Name	Description	Data Source
CAR$^{[2,H]}$	Cumulative abnormal return at time $\tau + 2$ to $\tau + H$, where τ denotes the time of the announcement and H days after the announcement. The abnormal return is defined as the excess return of that predicted by a simple market model. The estimation window is from time $\tau - 231$ to $\tau - 31$ relative to the announcement with a minimum of 140 observations.	CRSP/Compustat
SUE	Surprise in earnings defined as the I/B/E/S actual earnings per share minus the corresponding median earnings per share forecast, scaled by the corresponding I/B/E/S reported price per share (Livnat and Mendenhall (2006)).	I/B/E/S
SUE_SD	Surprise in earnings defined as the I/B/E/S actual earnings per share minus the corresponding median earnings per share forecast, scaled by the standard deviation of the analyst forecasts (Jegadeesh and Livnat (2006)).	I/B/E/S
SIZE	Log of the market capitalization of the firm as measured in the month-end prior to the earnings announcement. Market capitalization is the dollar share price multiplied by the number of ordinary shares in issue (Bhushan (1994)).	Datastream
BM	Book-to-market value measured as the book value of equity divided by the market value of equity measured in the previous quarter to the announcement.	Datastream
RUNUP	Measures the cumulative abnormal return in the month prior to the announcement, namely, $\tau - 30$ to $\tau - 1$.	CRSP/Compustat
PASTRET	The cumulative returns from $\tau - 230$ to $\tau - 30$.	CRSP/Compustat
DOLVOL	The dollar trading volume measured as the average monthly dollar trading volume in the previous calendar year in millions (Bhushan (1994)).	Datastream
VOL	The total number of shares traded in month $t - 1$.	Datastream

Table B1: Variable Description.
The table reports the variables of interest previously used in the empirical literature.

Variable Name	Description	Data Source
TURNOVER	Share turnover defined as volume over the entire prior fiscal year, divided by shares outstanding at the end of that fiscal year.	Datastream
ANALYST	Number of analysts who provide earnings per share forecasts in the I/B/E/S database (Bhushan (1994)).	I/B/E/S
PRICE	Stock price in the end of the previous calendar year ranked from 1 to 5 based on the share price range of Bhushan (1994).	I/B/E/S
LEV	Leverage, long-term debt plus current portion of long-term debt to total assets measured at end of month $t-1$ used as a control in Shivakumar (2006).	CRSP/Compustat
SAMEFIS	An indicator if earnings announcement t is in same fiscal year as announcement $t-1$ used as a control in Narayanamoorthy (2006).	I/B/E/S
LOSS	An indicator if negative earnings was announced in previous quarter used as a control in Narayanamoorthy (2006).	I/B/E/S
REPLAG	Lag between the quarter-end and earnings announcement date used in Dehaan et al. (2017).	I/B/E/S
NRANK	The decile rank of the number of firms announcing earning news on the same date (Hirshleifer et al. (2009)).	CRSP/Compustat
DECR	An indicator of one if earnings decreased between quarter $t-2$ to $t-1$ (Narayanamoorthy (2006)).	I/B/E/S
EARET	The two-day cumulative abnormal return measured using a market model on the earnings announcement date, $t-1$ to t. (Garfinkel and Sokobin (2006)).	CRSP/Compustat
ARBRISK	Residual variance from a market model regression estimated over the last 200 trading days ending 30 days prior to the earnings announcement (Mendenhall (2004)).	Datastream
EXPRISK	Explained variance from a market model regression estimated over the last 200 trading days ending 30 days prior to the earnings announcement (Mendenhall (2004)).	Datastream

Variable Name	Description	Data Source
ILLIQ	Amihud (2002) illiquidity measure computed using daily data in the month of the announcement used in Sadka (2006). Data from Green et al. (2017).	CRSP/Compustat
BETA	Estimated market beta from weekly returns and equal weighted market returns for 3 years ending one-month prior to the month of the announcements with at least 52 weeks of returns. Data from Green et al. (2017).	CRSP/Compustat

Additional Explanatory Variables

Table B1: Description of Stock-Specific Variables.
The table reports the list of additional control variables that have been shown to explain the cross-section of stock returns. We refer the reader to the Appendix of Green et al. (2017) for construction of the individual variables.

Acronym	**Characteristic**	**Data Source**
	Variables from Green et al. (2017)	
ABSACC	Absolute accruals	CRSP/Compustat
ACC	Working capital accruals	CRSP/Compustat
AEAVOL	Abnormal earnings announcement volume	CRSP/Compustat
AGE	# years since first Compustat coverage	CRSP/Compustat
AGR	Asset growth	CRSP/Compustat
BASPREAD	Bid-ask spread	CRSP/Compustat
BETASQ	Beta squared	CRSP/Compustat
BM_IA	Industry-adjusted book to market	CRSP/Compustat
CASH	Cash holdings	CRSP/Compustat
CASHDEBT	Cash flow to debt	CRSP/Compustat
CASHPR	Cash productivity	CRSP/Compustat
CFP	Cash flow to price ratio	CRSP/Compustat
CFP_IA	Industry-adjusted cash flow to price ratio	CRSP/Compustat
CHATOIA	Industry-adjusted change in asset turnover	CRSP/Compustat
CHCSHO	Change in shares outstanding	CRSP/Compustat
CHEMPIA	Industry-adjusted change in employees	CRSP/Compustat
CHINV	Change in inventory	CRSP/Compustat
CHMOM	Change in six-month momentum	CRSP/Compustat
CHPMIA	Industry-adjusted change in profit margin	CRSP/Compustat
CHTX	Change in tax expense	CRSP/Compustat
CINVEST	Corporate investment	CRSP/Compustat
CONVIND	Convertible debt indicator	CRSP/Compustat
CURRAT	Current ratio	CRSP/Compustat
DEPR	Depreciation/PP&E	CRSP/Compustat
DIVI	Dividend initiation	CRSP/Compustat
DIVO	Dividend omission	CRSP/Compustat
DY	Dividend to price	CRSP/Compustat
EGR	Growth in common shareholder equity	CRSP/Compustat
EP	Earnings to price	CRSP/Compustat
GMA	Gross profitability	CRSP/Compustat
GRCAPX	Growth in capital expenditures	CRSP/Compustat
GRLTNOA	Growth in long-term net operating assets	CRSP/Compustat
HERF	Industry sales concentration	CRSP/Compustat
HIRE	Employee growth rate	CRSP/Compustat
IDIOVOL	Idiosyncratic return volatility	CRSP/Compustat
INDMOM	Industry momentum	CRSP/Compustat
INVEST	Capital expenditures and inventory	CRSP/Compustat
LGR	Growth in long-term debt	CRSP/Compustat

Acronym	Characteristic	Data Source
MAXRET	Maximum daily return	CRSP/Compustat
MOM12M	12-month momentum	CRSP/Compustat
MOM1M	1-month momentum	CRSP/Compustat
MOM36M	36-month momentum	CRSP/Compustat
MS	Financial statement score	CRSP/Compustat
MVE_IA	Industry-adjusted size	CRSP/Compustat
NINCR	Number of earnings increases	CRSP/Compustat
OPERPROF	Operating profitability	CRSP/Compustat
ORGCAP	Organizational capital	CRSP/Compustat
PCHCAPX_IA	Industry-adjusted % change in capital expenditures	CRSP/Compustat
PCHCURRAT	% change in current ratio	CRSP/Compustat
PCHDEPR	% change in depreciation	CRSP/Compustat
PCHGM_PCHSALE	% change in gross margin - % change in sales	CRSP/Compustat
PCHQUICK	% change in quick ratio	CRSP/Compustat
PCHSALE_PPCHINVT	% change in sales - % change in inventory	CRSP/Compustat
PCHSALE_PCHRECT	% change in sales - % change in A/R	CRSP/Compustat
PCHSALE_PCHXSGA	% change in sales - % change in SG&A	CRSP/Compustat
PCHSALEINV	% change in sales-to-inventory	CRSP/Compustat
PCTACC	Percent accruals	CRSP/Compustat
PRICEDELAY	Price delay	CRSP/Compustat
PS	Financial statement score	CRSP/Compustat
QUICK	Quick ratio	CRSP/Compustat
RD	R&D increase	CRSP/Compustat
RD_MVE	R&D to market capitalization	CRSP/Compustat
RD_SALE	R&D to sales	CRSP/Compustat
REALESTATE	Real estate holdings	CRSP/Compustat
RETVOL	Return volatility	CRSP/Compustat
ROAQ	Return on assets	CRSP/Compustat
ROAVOL	Earnings volatility	CRSP/Compustat
ROEQ	Return on equity	CRSP/Compustat
ROIC	Return on invested capital	CRSP/Compustat
RSUP	Revenue surprise	CRSP/Compustat
SALECASH	Sales to cash	CRSP/Compustat
SALEINV	Sales to inventory	CRSP/Compustat
SALEREC	Sales to receivables	CRSP/Compustat
SECURED	Secured debt	CRSP/Compustat
SECUREDIND	Secured debt indicator	CRSP/Compustat
SGR	Sales growth	CRSP/Compustat
SIN	Sin stocks	CRSP/Compustat
SP	Sales to price	CRSP/Compustat

Acronym	**Characteristic**	**Data Source**
STD_DOLVOL	Volatility of liquidity (dollar trading volume)	CRSP/Compustat
STD_TURN	Volatility of liquidity (share turnover)	CRSP/Compustat
STDACC	Accrual volatility	CRSP/Compustat
STDCF	Cash flow volatility	CRSP/Compustat
TANG	Debt capacity/firm tangibility	CRSP/Compustat
TB	Tax income to book income	CRSP/Compustat
ZEROTRADE	Zero trading days	CRSP/Compustat

C Additional Analysis

Table C1: Estimates Across Model Specifications - Non-Interaction Terms. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and fourth columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. 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 percent) from separate regressions with no controls for the parameters of secondary 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. Standard errors are clustered by day and firm.

	OLS	OLS	OLS	DML w. post-lasso
Separate regressions using:	*No controls*	*Selected controls*	*Full set of controls*	*Nuisance function*
PASTRET	**−30.213**	**−28.785**	**−29.036**	**−12.797**
BM	**12.169**	**8.622**	**8.459**	**6.330**
ARBRISK	**−3.231**	**−3.595**	**−6.056**	**−8.208**
<u>SUE</u>	−0.440	**3.078**	**2.791**	**4.529**
<u>DECR</u>	**4.242**	**3.111**	**2.919**	0.050
<u>RUNUP</u>	**−3.145**	**−2.736**	**−2.969**	0.612
EARET	**5.411**	**3.604**	**3.381**	**2.121**
<u>ILLIQ</u>	**3.350**	**1.733**	**1.535**	0.608
<u>SAMEFIS</u>	**−2.739**	**−2.035**	−0.629	−0.220
TURNOVER	**−0.834**	0.367	0.043	**1.559**
<u>LOSS</u>	**2.831**	**3.516**	**2.475**	0.530
<u>REPLAG</u>	**0.987**	−0.220	−1.324	−0.866
<u>VOL</u>	0.073	**2.815**	**1.817**	**1.851**
PRICE	−0.272	−0.593	−0.615	−1.041
<u>ANALYST</u>	**0.633**	**6.997**	**6.826**	1.383
NRANK	−0.606	0.104	0.056	0.118
EXPRISK	−0.564	1.042	1.552	−1.177
SIZE	**−6.169**	**−3.620**	**−10.144**	**−5.758**
<u>BETA</u>	**−1.333**	−0.564	**−1.342**	−0.996
<u>LEV</u>	**0.909**	**1.804**	**1.759**	1.104
DOLVOL	−0.012	**0.663**	−0.138	0.058
Observations	170,719	170,719	170,719	170,719
Fixed effects			X	X
No. of variables	3	21	138	2,839

Figure C1: Visualization of Estimates - Interaction Terms.
The figure visualizes the results for the primary interaction term estimates from Table 3 across the four model specifications. The dots indicate the standardized coefficient estimate and the tails indicate 1% confidence intervals of the estimates. The tails are colored black if the variable is significant at a 1% level. Variables are listed on the y-axis according to Table 3.

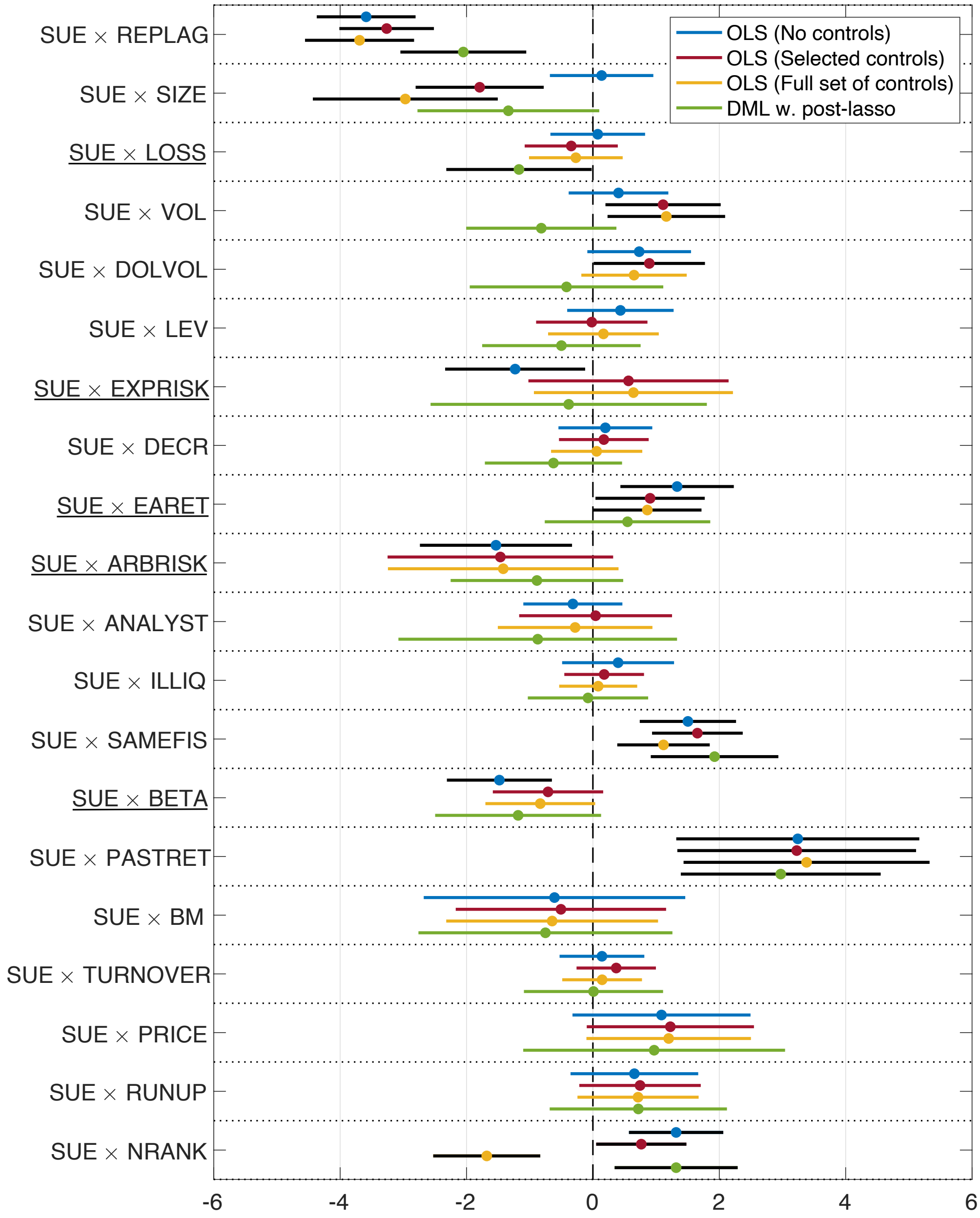


Figure C2: Visualization of Estimates - Non-Interaction Terms.
The figure visualizes the results for the variables of interest themselves from Table C1 across the four model specifications. The dots indicate the standardized coefficient estimate and the tails indicate 1% confidence intervals of the estimates. The tails are colored black if the variable is significant at a 1% level. To ensure proper visualization of the results, we limit the x-axis to cover the interval between -0.16 to 0.16. The PASTRET results from the OLS specifications are outside this interval, and are therefore excluded. Variables are listed on the y-axis according to Table C1.

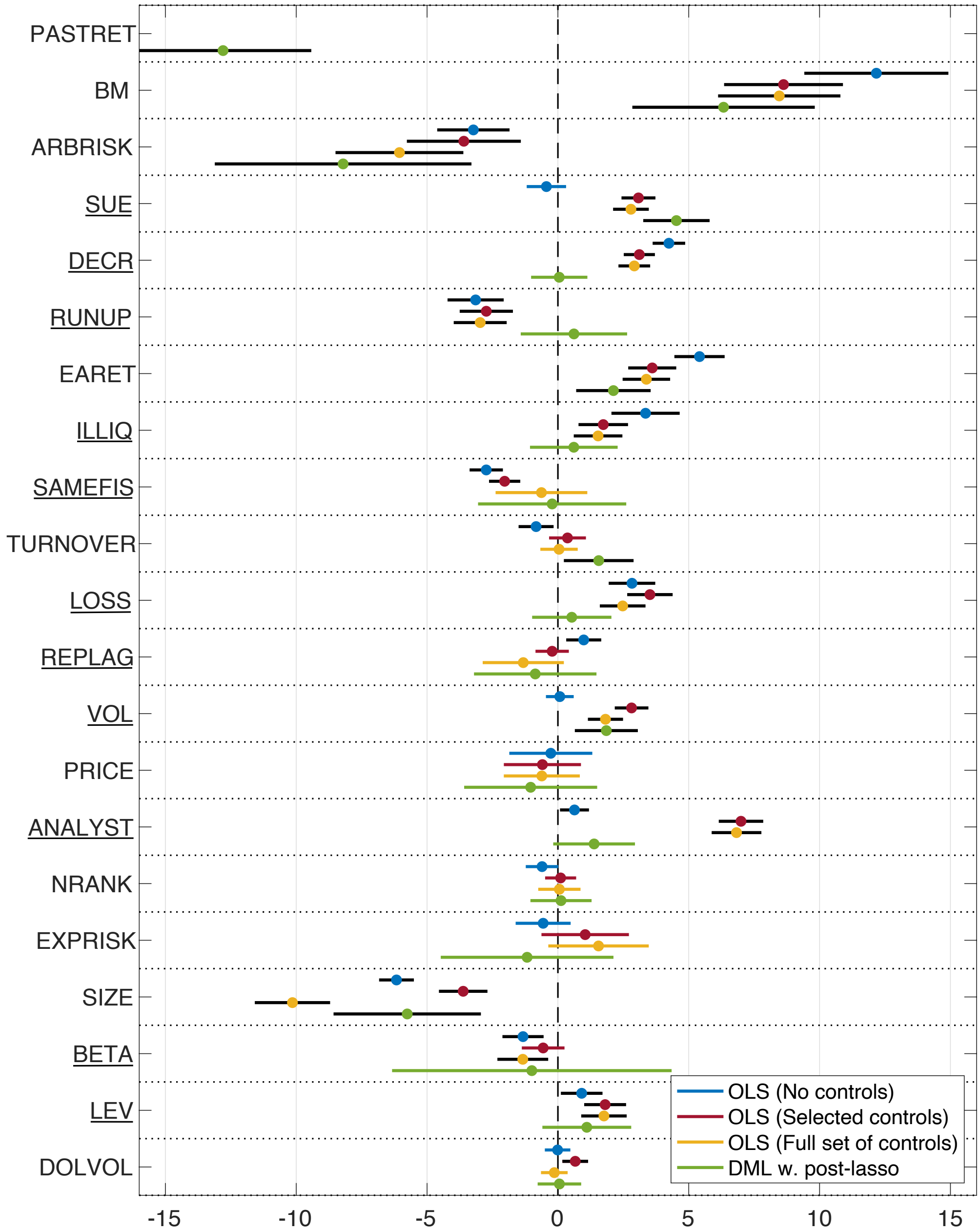


Figure C3: Inferences Through Time - All Variables.
This figure reports p-values for stock-specific covariates from separate regressions based on ten years of data. We use a rolling window from 1993 to 2019. Panel A uses OLS with the full set of controls and panel B uses DML. Both panels consider both the interaction term and the variable itself on the y-axis, where the most significant of the two is reported. Variables are listed according to Table C1. Colors illustrate levels of significance, 1% (blue), 5% (green), 10% (yellow), and non-significant (white).

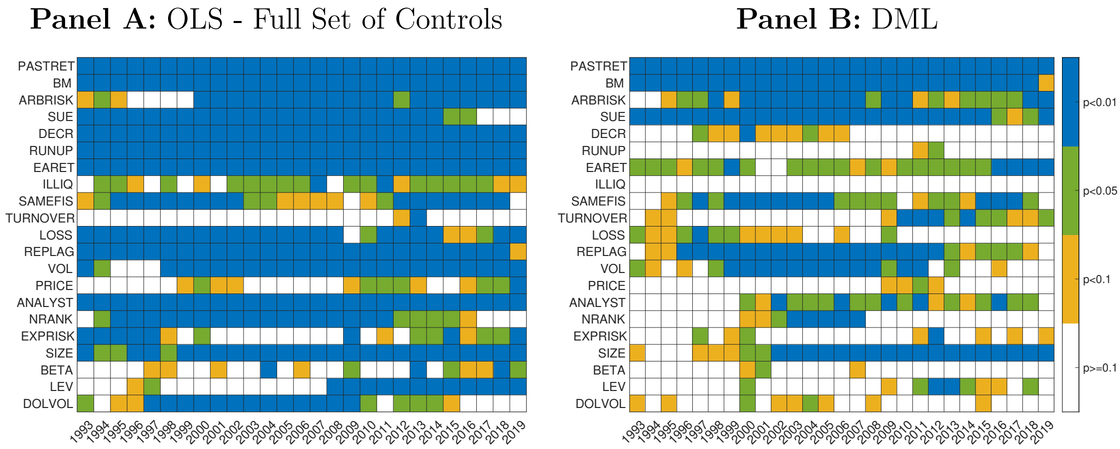


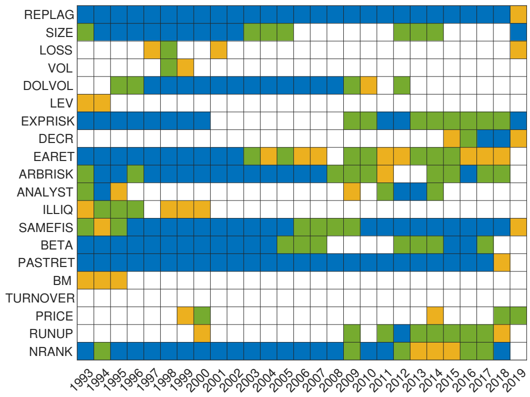
Table C2: Estimates for Post-Lasso Using the Full Set of Variables.
Variables are listed based on the magnitude of the differences in coefficient estimates from the first and fourth columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. 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 two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		−0.440		**3.078**		**2.791**		**2.159**
EXPRISK	**−1.231**	**−0.564**	0.563	**1.042**	0.642	**1.552**	0.491	**1.309**
REPLAG	**−3.588**	**0.987**	**−3.264**	**−0.220**	**−3.693**	**−1.324**	**−2.223**	**−1.121**
ARBRISK	**−1.534**	**−3.231**	−1.464	**−3.595**	−1.419	**−6.056**	−0.172	**−5.826**
SIZE	0.138	−6.169	**−1.792**	−3.620	**−2.969**	−10.144	−0.845	−10.401
BETA	**−1.480**	**−1.333**	−0.710	**−0.564**	−0.833	**−1.342**	−0.619	**−1.207**
LEV	0.435	0.909	−0.018	**1.804**	0.167	1.759	−0.183	1.736
EARET	**1.333**	5.411	**0.905**	**3.604**	**0.861**	**3.381**	0.772	**3.491**
DECR	0.197	**4.242**	0.173	**3.111**	0.060	**2.919**	−0.254	**2.892**
ILLIQ	0.399	**3.350**	0.178	**1.733**	0.084	**1.535**	0.014	**1.512**
DOLVOL	0.732	**−0.012**	**0.893**	**0.663**	0.651	**−0.138**	0.358	**−0.122**
LOSS	0.077	**2.831**	−0.343	3.516	−0.270	2.475	−0.276	2.479
PASTRET	**3.242**	**−30.213**	**3.225**	**−28.785**	**3.381**	**−29.036**	**2.932**	**−29.095**
SAMEFIS	**1.504**	**−2.739**	**1.654**	**−2.035**	**1.117**	**−0.629**	**1.763**	**−0.685**
VOL	0.405	0.073	**1.111**	2.815	**1.162**	1.817	0.162	1.863
ANALYST	−0.317	**0.633**	0.043	6.997	−0.281	6.826	−0.166	6.764
TURNOVER	0.143	**−0.834**	0.369	**0.367**	0.146	0.043	0.276	0.148
NRANK	**1.317**	−0.606	**0.766**	0.104	**−1.680**	0.056	**1.442**	0.252
BM	−0.608	**12.169**	−0.506	**8.622**	−0.645	**8.459**	−0.557	**8.704**
RUNUP	0.656	**−3.145**	0.746	−2.736	0.714	**−2.969**	0.698	−2.887
PRICE	1.086	**−0.272**	1.226	**−0.593**	1.199	**−0.615**	1.105	−0.727
Observations	170,719		170,719		170,719		170,719	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

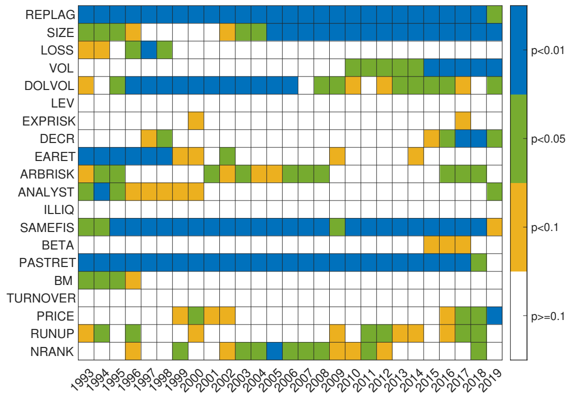
Figure C4: Inferences Through Time - OLS Specification 1 and 2.
This figure reports p-values for stock-specific covariates from separate regressions based on ten years of data. We use a rolling window from 1993 to 2019. Panels A and B consider the interaction term between the variable listed on the y-axes and SUE. Panels C and D consider both the estimate on the variable itself or its interaction with SUE, the most significant of the two is reported. Panels A and C uses OLS with no controls and panels B and D uses OLS with selected control variables. Variables in the top panels are sorted according to Table 3 and bottom panels to Table C1. Colors illustrate levels of significance, 1% (blue), 5% (green), 10% (yellow), and non-significant (white).

Interaction Terms Only:

Panel A: OLS - No Controls

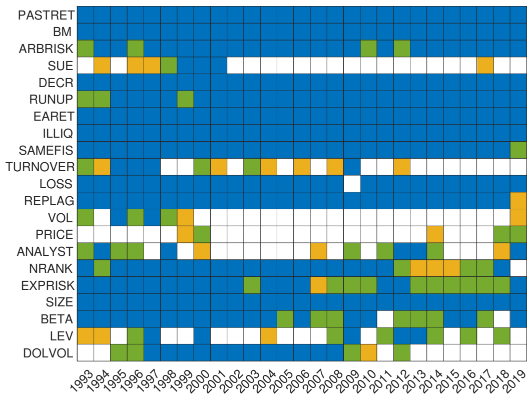


Panel B: OLS - Selected Controls



Variable and Interaction Terms:

Panel C: OLS - No Controls



Panel D: OLS - Selected Controls

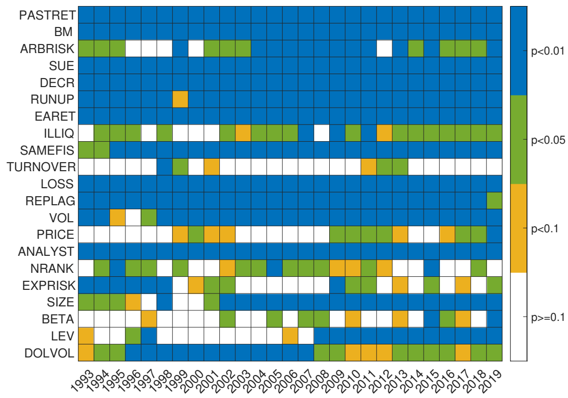


Figure C5: Selection by Lasso.
This figure reports when a variable of interest or a stock-specific variable is chosen by lasso. The model is estimated using a rolling-window containing ten years of data and with the end year of the window given by the x-axis. The y-axis considers both the estimate on the variable itself or its interaction with SUE. The area is shaded gray when lasso selects a variable.

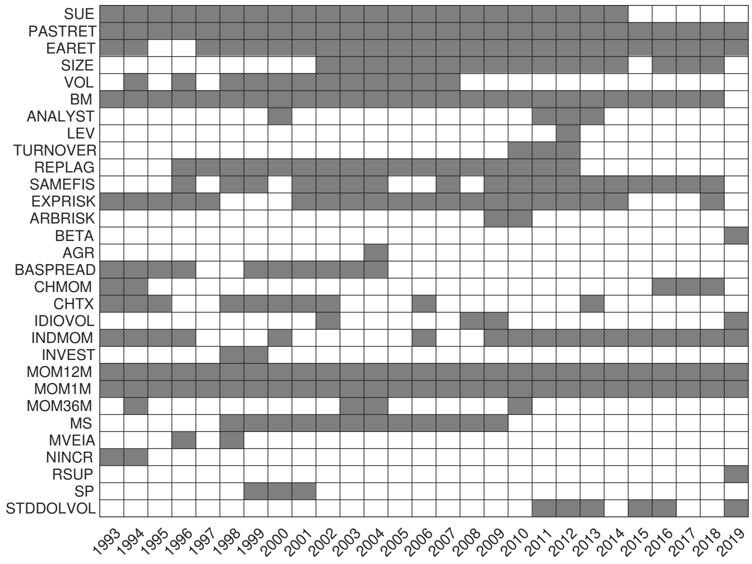


Table C3: Controls That Matter - Non-Interaction Terms.
Variables are listed based on the magnitude of the differences in coefficient estimates from the first and fourth columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. 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 percent) from separate regressions with no controls for the parameters of secondary 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. Standard errors are clustered by day and firm.

	OLS	OLS	OLS	DML w. post-lasso
Separate regressions using:	*No controls*	*Selected controls*	*Full set of controls*	*Nuisance function*
INDMOM	**−12.124**	**−4.916**	**−6.609**	**3.079**
MOM12M	**−37.662**	**−29.570**	**−30.396**	**−30.411**
<u>NINCR</u>	**−5.644**	**−4.526**	**−3.614**	−0.249
<u>CHTX</u>	**−5.557**	**−4.643**	**−4.164**	−0.986
BASPREAD	**4.997**	**6.038**	**6.678**	**8.592**
<u>STD_DOLVOL</u>	**2.380**	**−1.849**	**−1.845**	−0.575
IDIOVOL	**−5.199**	**−6.819**	**−11.934**	**−2.954**
<u>MOM36M</u>	**1.010**	0.456	**1.771**	−0.722
AGR	0.053	**−1.077**	**−0.867**	−1.013
MOM1M	**−5.934**	**−2.263**	**−2.227**	**−5.280**
MS	**1.703**	**3.439**	**4.218**	**1.598**
Observations	170,719	170,719	170,719	170,719
Fixed effects			X	X
No. of variables	3	21	138	2,839

Figure C6: Controls That Matter Through Time - All Variables.
This figure reports p-values for stock-specific covariates from separate regressions based on ten years of data. We use a rolling window from 1993 to 2019. Panel A uses OLS with the full set of controls and panel B uses DML. Both panels consider the interaction term between SUE and the variable itself listed on the y-axis, the most significant of the two is reported. Variables are listed according to Table 5. Colors illustrate levels of significance: 1% (blue), 5% (green), 10% (yellow), non-significant (gray), and not selected by lasso (white). We use two-fold cross-fitting and obtain estimates using two splits.

Panel A: OLS - Full Set of Controls

Panel B: DML

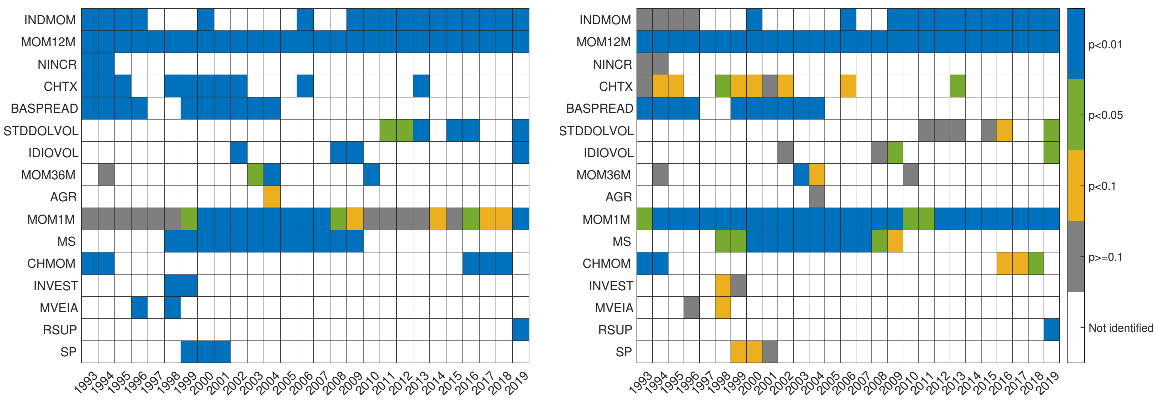
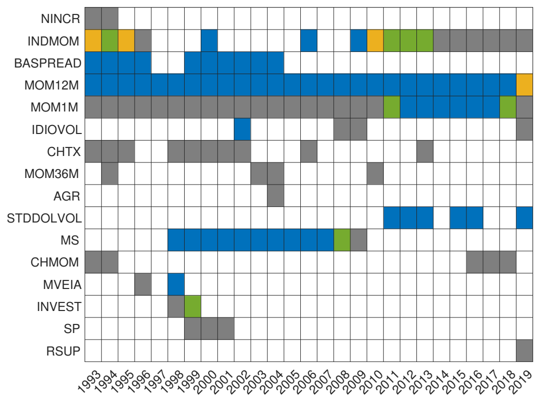


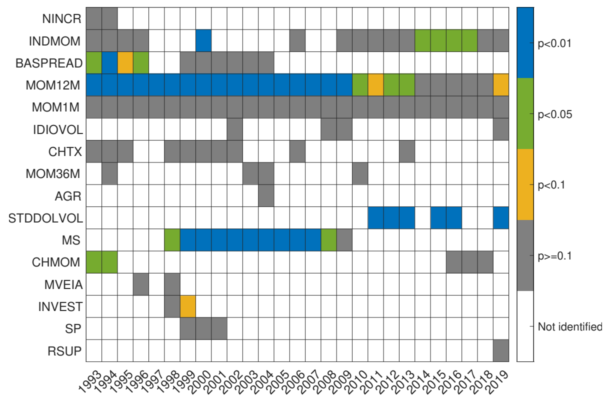
Figure C7: Controls That Matter Through Time - OLS Specification 1 and 2.
This figure reports p-values for stock-specific covariates from separate regressions based on ten years of data. We use a rolling window from 1993 to 2019. Panels A and B consider the interaction term between the variable listed on the y-axes and SUE. Panels C and D consider both the estimate on the variable itself or its interaction with SUE, the most significant of the two is reported. Panels A and C uses OLS with no controls, while panels B and D uses OLS with selected control variables. Variables in the top panels are sorted according to Table 5 and bottom panels to Table C3. Colors illustrate levels of significance: 1% (blue), 5% (green), 10% (yellow), non-significant (gray), and not selected by lasso (white). We use two-fold cross-fitting and obtain estimates using two splits.

Interaction Terms Only:

Panel A: OLS - No Controls

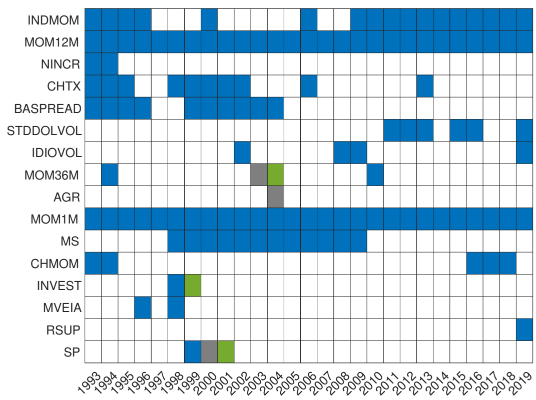


Panel B: OLS - Selected Controls



Variable and Interaction Terms:

Panel C: OLS - No Controls



Panel D: OLS - Selected Controls

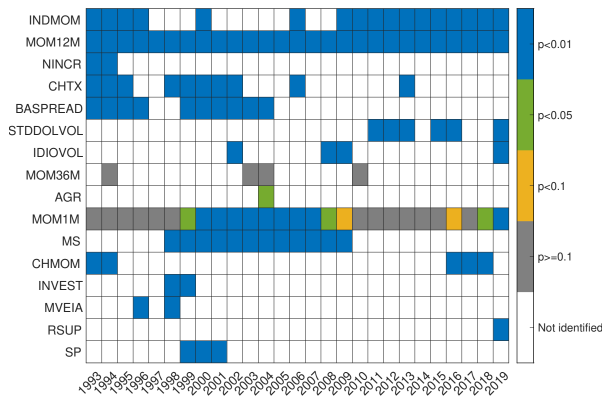


Table C4: Estimates Using Random Forest - Non-Interaction Terms.
Variables are listed based on the magnitude of the differences in coefficient estimates from the first and third columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. 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 column reports estimates (in percent) from separate regressions with the set of controls chosen ex ante for the parameters of secondary interest. The second column reports estimates using the high-dimensional nuisance function estimated using post-lasso and the third column using random forest. Standard errors are clustered by day and firm.

	OLS	DML w. post-lasso	DML w. random forest
Separate regressions using:	*Selected controls*	*Nuisance function*	*Nuisance function*
PASTRET	**−28.785**	**−29.095**	**−8.534**
ARBRISK	**−3.595**	**−5.826**	**−13.509**
SIZE	**−3.620**	**−10.401**	**−14.759**
<u>EARET</u>	**3.604**	**3.491**	1.255
<u>DECR</u>	**3.111**	**2.892**	0.121
BM	**8.622**	**8.704**	**8.318**
SUE	**3.078**	**2.159**	**3.282**
<u>ILLIQ</u>	**1.733**	1.512	0.771
<u>LOSS</u>	**3.516**	**2.479**	0.289
<u>RUNUP</u>	**−2.736**	**−2.887**	−0.720
VOL	**2.815**	**1.863**	**2.249**
NRANK	0.104	0.252	1.539
<u>SAMEFIS</u>	**−2.035**	−0.685	−0.918
REPLAG	−0.220	−1.121	−0.646
TURNOVER	0.367	0.148	0.702
<u>ANALYST</u>	**6.997**	**6.764**	1.965
<u>DOLVOL</u>	**0.663**	−0.122	1.177
BETA	−0.564	−1.207	−2.025
EXPRISK	1.042	1.309	−1.164
PRICE	−0.593	−0.727	0.151
<u>LEV</u>	**1.804**	**1.736**	0.989
Observations	170,719	170,719	170,719
Fixed effects		X	X
No. of variables	21	2,839	211

Table C5: Estimates Using Deciles - Non-Interaction Terms.
Non-interaction term estimates for each variable measured as deciles, computed using yearly quantile breakpoints. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and fourth columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first numerical column reports estimates (in percent) from separate regressions with no controls for the parameters of secondary 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. Standard errors are clustered by day and firm.

	OLS	OLS	OLS	DML w. post-lasso
Separate regressions using:	*No controls*	*Selected controls*	*Full set of controls*	*Nuisance function*
PASTRET	**−28.263**	**−21.591**	**−20.908**	**−9.004**
DOLVOL	**1.014**	**38.785**	**40.980**	**12.786**
VOL	**−2.509**	0.372	**2.313**	**3.605**
BM	**17.773**	**11.523**	**12.939**	**11.870**
ARBRISK	**−1.674**	**−13.562**	**−14.525**	**−6.462**
<u>SUE</u>	−0.440	**3.575**	**3.349**	**4.086**
<u>DECR</u>	**4.242**	**2.121**	**2.189**	−0.143
EARET	**4.790**	**2.164**	**2.170**	**1.295**
<u>RUNUP</u>	**−2.804**	**−0.976**	**−1.048**	0.685
<u>EXPRISK</u>	**−1.965**	**1.283**	**3.013**	0.555
<u>SAMEFIS</u>	**−2.739**	**−1.724**	**3.457**	−0.283
<u>REPLAG</u>	**1.113**	−0.537	**−1.280**	−0.831
<u>LOSS</u>	**2.831**	**1.890**	0.746	0.968
ILLIQ	**7.844**	**25.130**	**26.370**	**9.160**
SIZE	**−7.352**	**−16.355**	**−18.052**	**−6.048**
PRICE	**−4.428**	**−4.846**	**−6.997**	**−5.626**
NRANK	−0.606	0.309	0.084	0.345
LEV	**1.023**	0.340	0.387	**1.723**
ANALYST	0.297	**2.864**	**3.040**	0.067
<u>BETA</u>	**−1.075**	**−2.478**	**−2.438**	−0.846
TURNOVER	−0.104	**−4.001**	**−5.007**	−0.064
Observations	170,719	170,719	170,719	170,719
Fixed effects			X	X
No. of variables	3	21	138	2,839

Table C6: Estimates Using Quintiles.

Estimates for each variable measured as quintiles, computed using yearly quantile breakpoints. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		−0.393		**3.259**		**2.820**		**4.004**
REPLAG	**−3.470**	**1.182**	**−3.112**	**−0.419**	**−3.181**	**−1.008**	**−2.188**	−0.594
VOL	**−1.090**	**−2.544**	−1.237	**0.432**	−1.295	**0.500**	0.154	**2.406**
LOSS	0.008	**2.828**	−0.423	**2.111**	−0.443	**1.282**	**−1.040**	**0.911**
PASTRET	**1.921**	**−27.218**	**1.469**	**−22.259**	**1.524**	**−21.879**	**2.931**	**−8.043**
EXPRISK	**−1.496**	**−1.881**	−0.226	**0.619**	−0.140	**1.957**	−0.521	0.041
ILLIQ	**0.797**	**7.731**	1.238	**17.599**	0.822	**18.057**	−0.079	**4.559**
BETA	**−1.635**	**−1.112**	−0.107	−2.244	−0.233	−2.407	−0.811	**−1.387**
EARET	**1.239**	**4.621**	0.467	**2.357**	0.554	**2.385**	0.436	**1.358**
DECR	0.135	4.242	0.147	**2.340**	0.081	**2.358**	−0.638	−0.151
ARBRISK	**−1.536**	**−1.506**	**−1.690**	−10.322	−0.987	−10.840	−0.864	−4.959
LEV	0.176	1.087	−0.346	**0.152**	−0.091	**0.267**	−0.447	1.238
DOLVOL	**−0.927**	**0.961**	−0.144	**27.405**	0.317	**26.803**	−0.351	7.171
TURNOVER	**−1.208**	**−0.032**	−0.353	**−1.765**	−0.131	**−2.315**	−0.648	0.997
ANALYST	**−0.983**	**0.293**	−0.613	3.670	−0.661	3.990	−0.487	0.181
BM	−0.498	**17.475**	−0.583	12.007	−0.672	13.516	−0.982	11.296
SIZE	−0.218	**−7.264**	0.208	**−10.783**	0.680	**−10.621**	0.262	−3.070
SAMEFIS	**1.442**	**−2.742**	**1.682**	−1.961	**1.112**	**2.101**	**1.850**	−0.325
PRICE	−0.204	**−4.426**	**−1.095**	**−4.656**	**−1.714**	**−6.322**	0.160	**−5.711**
NRANK	**1.152**	**−0.710**	0.644	**0.049**	**−1.141**	**−0.164**	**1.125**	0.248
RUNUP	0.689	**−2.577**	0.279	**−1.255**	0.493	**−1.385**	0.697	**0.860**
Observations	170,719		170,719		170,719		170,719	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

Table C7: Estimates Using the 1st and 10th SUE Decile.

Estimates using SUE in the top or bottom decile and each control variable is continuous but standardized. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		0.643		**5.616**		**4.998**		**6.723**
EXPRISK	−1.666	5.142	0.975	6.638	1.264	**3.231**	1.722	−0.847
DOLVOL	2.556	**−2.379**	1.854	**−1.061**	3.275	**−3.350**	4.984	**−4.730**
REPLAG	**−4.763**	**0.652**	**−4.505**	**−0.382**	**−4.797**	**0.872**	**−2.688**	0.456
ARBRISK	−2.199	1.409	−2.187	−2.730	−2.048	−6.084	−0.291	−8.325
BETA	−1.424	**1.384**	−0.977	**0.704**	−0.890	**1.426**	−0.002	**0.742**
LOSS	0.175	2.741	−0.988	2.409	−1.148	1.448	−1.071	0.551
SIZE	0.506	−9.610	−1.334	**−7.022**	−1.805	−11.623	−0.680	−5.567
RUNUP	−0.814	**−1.469**	0.098	**−1.545**	−0.180	**−2.474**	0.055	**0.764**
ANALYST	−1.284	**2.101**	−1.527	**7.630**	−1.838	**7.038**	−0.435	1.843
TURNOVER	0.361	−0.102	0.856	0.931	0.480	0.532	1.183	1.299
ILLIQ	0.415	2.911	−0.065	0.709	−0.090	0.529	−0.365	0.390
PASTRET	**3.424**	**−29.669**	**3.266**	**−28.055**	**3.161**	**−28.128**	2.777	−13.847
DECR	0.759	**4.181**	0.556	**3.037**	0.536	2.560	0.174	0.209
LEV	0.878	−0.773	0.207	1.605	0.264	1.596	0.298	0.855
VOL	0.552	−0.490	1.302	2.412	1.416	0.676	1.060	0.706
SAMEFIS	**2.376**	**−3.036**	**2.341**	**−2.624**	1.394	−0.968	2.025	0.400
BM	−1.221	**9.838**	−1.061	**7.071**	−1.149	6.862	−1.518	6.059
NRANK	**1.814**	−0.025	1.166	0.466	**−2.196**	**1.731**	1.553	**1.978**
EARET	0.348	5.758	0.338	3.929	0.560	3.484	0.226	2.842
PRICE	2.983	**−0.235**	**3.225**	−0.635	**3.143**	−0.653	2.862	−1.107
Observations	29,853		29,853		29,853		29,853	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

Table C8: Estimates Using Stock-Specific Variables Measured One-Month Prior.
The table reports the main results across the four model specifications using the stock-specific variables of Green et al. (2017) measured one-month prior to the earnings announcement date. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		−0.417		**3.077**		**2.745**		**4.617**
REPLAG	**−3.582**	**0.933**	**−3.253**	**−0.275**	**−3.672**	**−1.442**	**−2.026**	**−1.313**
SIZE	0.079	**−6.188**	**−1.869**	**−3.624**	**−3.048**	**−10.110**	−1.374	**−6.438**
LOSS	0.135	**2.855**	−0.280	**3.560**	−0.231	**2.514**	−1.049	**0.488**
VOL	0.319	0.067	**1.050**	2.807	**1.106**	1.816	−0.858	1.905
DOLVOL	0.710	**−0.010**	**0.901**	**0.661**	0.668	**−0.143**	−0.424	**0.055**
EXPRISK	**−1.329**	−0.408	0.486	**1.146**	0.575	1.660	−0.338	−0.447
LEV	0.455	0.897	0.014	**1.797**	0.187	**1.759**	−0.412	**0.953**
DECR	0.280	**4.276**	0.261	**3.159**	0.149	**2.952**	−0.521	**0.072**
ARBRISK	**−1.570**	**−3.145**	−1.484	**−3.567**	−1.460	**−5.921**	−0.852	−8.521
EARET	**1.260**	**5.483**	0.846	**3.670**	0.806	**3.458**	0.587	2.411
ANALYST	−0.369	**0.609**	0.073	6.965	−0.259	6.786	−0.950	**1.247**
ILLIQ	0.459	**3.383**	0.229	**1.756**	0.141	**1.548**	−0.030	0.622
PASTRET	**3.271**	**−30.237**	**3.251**	**−28.811**	**3.408**	**−29.068**	**2.898**	−11.289
BETA	**−1.445**	−1.353	−0.644	−0.647	−0.786	−1.485	−1.202	−1.781
SAMEFIS	**1.531**	**−2.716**	**1.665**	−1.999	**1.120**	−0.595	**1.768**	−0.095
TURNOVER	0.154	**−0.852**	0.393	**0.339**	0.179	0.026	−0.049	1.515
BM	−0.622	12.193	−0.527	8.645	−0.653	8.490	−0.785	6.373
PRICE	0.941	**−0.427**	1.086	**−0.746**	1.067	**−0.764**	0.792	**−1.238**
RUNUP	0.712	**−3.024**	0.796	−2.639	0.769	**−2.850**	0.801	−2.968
NRANK	**1.339**	**−0.562**	**0.782**	**0.155**	**−1.659**	**0.086**	**1.313**	0.108
Observations	170,719		170,719		170,719		170,719	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

D Alternative Measures

Alternative Surprise in Earnings Measures

To assess the impact of using the earnings surprise measure based on time series forecasts, we define a time series measure of SUE where quarterly earnings follow a seasonal random walk with a drift process (Jegadeesh and Livnat (2006)). The parameters of the model are estimated over rolling windows between quarters $t-21$ through $t-1$ using historical data. Thus, each quarter the following model is estimated:

$$X_{i,t} = \delta_{i,t} + X_{i,t-4} + \varepsilon_{i,t}, \tag{15}$$

where $X_{i,t}$ is the actual I/B/E/S reported earnings per share, while $\delta_{i,t}$ and $\varepsilon_{i,t}$ are drift and random noise terms, respectively. The standardized time series SUE measure is defined as in Livnat and Mendenhall (2006):

$$\text{SUE\_TS}_{i,t} = \frac{X_{i,t} - \hat{\delta}_{i,t} - X_{i,t-4}}{\text{PRICE}_{i,t}}, \tag{16}$$

where $\text{PRICE}_{i,t}$ is the I/B/E/S reported share price. The main advantage of the time series measure is no requirement of analysts to follow a stock or provide guidance. Yet, other problems remain as discussed in Livnat and Mendenhall (2006) and Jegadeesh and Livnat (2006), which is why we rely on the analyst measure in our analysis. The results using the time series measure is reported in Table D1.

We consider changing the deflator in our main specification with the standard deviation of analyst forecasts as used by Jegadeesh and Livnat (2006):

$$\text{SUE\_SD}_{i,t} = \frac{\text{EPS}_{i,t} - \hat{\text{EPS}}_{i,t}}{STD(\hat{\text{EPS}})_{i,t}}. \tag{17}$$

We report the results in Table D2.

Table D1: Estimates Using SUE_TS.

Estimates using SUE_TS in deciles, and each control variable is continuous but standardized. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE_TS, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
<u>SUE</u>		**−12.023**		**−5.673**		**−5.638**		0.560
<u>ARBRISK</u>	**−4.285**	**2.201**	**−3.057**	**2.880**	**−3.215**	**2.400**	−0.238	−1.692
<u>EXPRISK</u>	**−4.035**	**−0.475**	−0.361	**−2.123**	−0.491	**−2.118**	−0.667	**−3.243**
<u>BETA</u>	**−2.140**	**−0.342**	0.060	**−1.800**	−0.769	**−2.263**	0.068	**−2.555**
<u>REPLAG</u>	**−2.497**	1.363	**−2.248**	−0.111	**−2.616**	−0.721	−1.198	−0.514
NRANK	0.827	**−1.064**	0.496	**−0.468**	−1.257	**−0.279**	−0.400	**−0.268**
<u>LOSS</u>	**1.589**	4.263	**2.184**	**2.557**	**2.199**	2.656	0.517	0.537
<u>EARET</u>	**2.011**	6.052	**1.509**	**4.249**	**1.519**	**3.975**	1.085	2.304
ILLIQ	1.076	**4.998**	1.007	**2.516**	0.932	**2.891**	0.210	**2.429**
VOL	−0.843	**−0.470**	1.017	**2.255**	0.761	**2.463**	−0.007	1.805
RUNUP	1.119	**−1.980**	**2.139**	**−2.349**	**2.260**	**−2.645**	1.934	**−0.258**
LEV	0.556	1.305	0.395	2.165	0.363	2.977	−0.107	2.899
PASTRET	**3.088**	**−29.223**	**2.493**	**−27.691**	**2.613**	**−27.571**	**2.653**	−12.963
TURNOVER	−0.123	**−0.319**	0.330	**0.383**	0.229	**0.529**	0.265	1.443
DECR	0.188	**2.264**	0.602	2.147	−0.139	2.221	−0.157	0.239
DOLVOL	−0.337	**−0.115**	0.276	0.796	0.025	0.006	−0.071	0.384
ANALYST	−1.074	**−1.494**	0.291	**4.801**	−0.791	**5.451**	−0.817	1.538
BM	−1.373	12.969	−1.314	8.654	−1.141	8.783	−1.518	6.979
SIZE	0.119	−8.274	**−2.591**	−2.573	−1.905	−10.086	−0.008	−6.275
SAMEFIS	1.038	−4.592	**1.315**	**−3.913**	1.012	**−3.268**	1.153	−3.060
PRICE	0.120	**0.111**	0.084	**−0.237**	0.005	**−0.315**	0.029	−0.180
Observations	74,216		74,216		74,216		74,216	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

Table D2: Estimates Using SUE_SD.

Estimates using SUE_SD in deciles, and each control variable is continuous but standardized. Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE_SD, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		**−3.107**		1.194		0.587		0.864
REPLAG	**−3.699**	**1.452**	**−1.773**	**−0.095**	**−2.306**	**−1.406**	−1.202	−1.324
NRANK	**1.984**	**−1.416**	0.466	**−0.574**	−1.583	**−0.877**	0.005	**−0.464**
SIZE	**2.612**	**−6.944**	−0.042	**−2.662**	−2.333	**−11.069**	0.737	**−5.291**
BETA	**−1.724**	−0.917	0.136	−2.184	0.062	−2.920	0.046	−3.455
ARBRISK	**−1.907**	**0.232**	−1.845	**0.022**	−2.062	**−0.996**	−0.255	**−4.879**
EXPRISK	**−2.280**	0.625	0.444	0.402	0.757	1.138	−0.777	−2.115
ANALYST	−0.701	0.613	0.684	**6.134**	1.180	**6.788**	0.680	1.371
LEV	0.014	**1.434**	0.095	**2.375**	−0.017	**2.300**	−1.147	**1.822**
DECR	−0.617	4.048	−0.162	**3.293**	−0.121	**3.342**	0.117	0.259
EARET	0.721	**4.997**	−0.812	**3.641**	−0.926	**3.419**	0.017	2.460
SAMEFIS	**2.250**	−4.248	**1.675**	−3.220	1.272	−0.243	1.931	−0.099
BM	1.134	**11.965**	0.657	**8.178**	0.487	**7.891**	0.826	6.065
VOL	0.037	**0.393**	0.545	**2.676**	0.208	**2.366**	0.335	1.462
TURNOVER	0.036	**−0.302**	0.204	0.271	0.292	0.390	0.282	1.489
LOSS	−0.531	**2.957**	0.015	3.061	0.098	2.352	−0.321	0.682
RUNUP	0.286	**−2.851**	0.374	**−2.652**	0.921	−2.944	0.454	1.008
PRICE	−0.132	−0.112	0.037	−0.379	−0.105	−0.374	−0.289	−0.728
ILLIQ	0.101	4.098	0.326	2.424	0.316	2.512	0.153	1.480
DOLVOL	0.159	0.075	−0.037	**0.800**	−0.141	**−0.224**	0.204	−0.043
PASTRET	**2.708**	**−29.861**	2.037	**−28.623**	2.121	**−28.946**	2.720	−13.153
Observations	83,225		83,225		83,225		83,225	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

Alternative Cumulative Abnormal Return Measures

We consider two other measurement horizons of the cumulative abnormal return from time $t + 2$ to $t + 41$, $\text{CAR}^{[2,41]}$, and from time $t + 2$ to $t + 21$, $\text{CAR}^{[2,21]}$, in Tables D3 and D4, respectively. Additionally, we report results using $\text{CAR}^{[2,61]}$ risk-adjusted with the Fama-French 3-factor model and the 3-factor model extended with momentum (Carhart (1997)) rather than the market model. These results are reported in Tables D5 and D6, respectively.

Table D3: Estimates Using CAR$^{[2,41]}$ as Dependent Variable.

Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
<u>SUE</u>		−0.056		**3.023**		**2.701**		**4.484**
SIZE	0.055	**−5.341**	**−1.877**	**−3.446**	**−2.404**	**−9.660**	−1.289	−5.671
REPLAG	**−2.675**	**0.856**	**−2.386**	**−0.187**	**−2.492**	**−1.938**	**−1.388**	**−1.494**
VOL	0.140	**−0.188**	0.742	**2.271**	0.870	**1.520**	−0.898	**1.329**
LOSS	0.276	2.457	0.019	3.298	0.090	1.984	−0.759	0.094
DOLVOL	0.501	**−0.059**	0.669	**0.563**	0.600	**−0.135**	−0.455	**0.107**
LEV	0.364	0.887	0.049	**1.660**	0.138	1.352	−0.463	0.376
<u>EARET</u>	**1.225**	4.586	0.863	**3.016**	0.785	**2.876**	0.489	**1.769**
DECR	0.426	**3.807**	0.400	**2.814**	0.233	**2.579**	−0.259	**0.123**
<u>ARBRISK</u>	**−1.416**	−3.650	−1.867	**−4.077**	−1.813	**−5.901**	−0.739	−7.028
<u>NRANK</u>	**1.289**	**−0.077**	**0.816**	**0.532**	−0.828	**0.635**	0.693	0.320
EXPRISK	−0.729	**−1.037**	1.240	0.977	1.378	1.231	−0.158	**−1.778**
<u>BETA</u>	**−1.575**	**−1.320**	**−0.925**	**−0.379**	**−1.066**	**−1.225**	−1.268	−2.082
ILLIQ	0.349	**2.452**	0.198	**1.109**	0.099	**0.899**	0.048	0.628
ANALYST	−0.605	0.480	−0.450	6.063	−0.749	6.056	−0.409	1.197
BM	−0.249	**10.575**	−0.189	7.468	−0.330	**7.456**	−0.423	5.620
PASTRET	**2.931**	**−26.554**	**3.056**	**−25.251**	**3.200**	−25.356	**2.761**	−11.365
<u>SAMEFIS</u>	**0.984**	−1.917	**1.153**	−1.315	0.674	−1.050	1.070	−0.985
PRICE	0.523	**−0.216**	0.671	**−0.467**	0.653	**−0.475**	0.446	**−0.861**
RUNUP	0.625	**−3.055**	0.739	−2.627	0.748	**−2.736**	0.688	0.293
TURNOVER	0.107	**−0.975**	0.314	**0.118**	0.234	0.004	0.049	1.335
Observations	170,719		170,719		170,719		170,719	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

Table D4: Estimates Using CAR$^{[2,21]}$ as Dependent Variable.

Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		**1.013**		**3.257**		**2.957**		**4.775**
<u>SIZE</u>	−0.335	**−3.654**	**−2.481**	**−1.914**	**−2.305**	**−5.981**	**−1.697**	−3.347
ARBRISK	−1.439	**−3.274**	−2.299	**−2.774**	−2.168	**−5.147**	−0.336	**−4.516**
<u>VOL</u>	−0.411	**−0.141**	0.311	**1.455**	0.738	**0.704**	**−1.387**	**0.762**
EXPRISK	−0.464	−1.440	1.700	−0.163	1.946	**1.797**	0.425	−1.018
DOLVOL	0.247	**0.013**	0.609	**0.353**	0.650	**−0.123**	−0.624	**0.200**
<u>REPLAG</u>	**−1.153**	1.016	**−0.979**	0.294	−0.600	−0.717	−0.332	−0.441
<u>NRANK</u>	**1.099**	−0.260	**0.762**	**0.192**	0.083	**0.826**	0.356	0.387
LEV	0.206	**0.016**	−0.108	**0.613**	−0.154	**0.880**	−0.499	**0.509**
DECR	0.692	**2.767**	0.660	**2.001**	0.475	**1.988**	−0.007	0.094
LOSS	0.469	1.571	0.320	2.407	0.388	1.713	−0.142	0.246
<u>BETA</u>	**−1.684**	−0.942	**−1.284**	0.192	**−1.392**	−0.716	−1.177	−2.822
<u>ANALYST</u>	**−1.165**	**0.738**	−0.961	**4.731**	−1.005	**4.716**	−0.754	1.257
EARET	0.968	**3.679**	0.726	**2.491**	0.730	**2.375**	0.581	1.625
<u>SAMEFIS</u>	**0.948**	−0.659	**1.098**	−0.201	**0.792**	0.495	0.798	0.635
RUNUP	0.370	**−2.668**	0.524	−2.312	0.449	−2.378	0.468	−0.595
BM	−0.526	**8.617**	−0.560	6.396	−0.685	6.608	−0.620	5.367
TURNOVER	−0.431	**−0.414**	−0.242	0.450	−0.261	0.194	−0.510	1.073
ILLIQ	−0.231	**1.847**	−0.338	**0.843**	−0.396	**0.905**	−0.180	1.180
PRICE	−0.306	**−0.742**	−0.178	−0.922	−0.185	−0.941	−0.345	−1.229
PASTRET	**2.048**	**−20.558**	**2.253**	−19.576	**2.297**	−19.685	**2.035**	−8.272
Observations	170,719		170,719		170,719		170,719	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

Table D5: Estimates - Risk-Adjusted Abnormal Returns Using Fama-French 3-Factor Model.
Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		−0.160		**3.094**		**2.704**		**4.117**
VOL	0.455	**0.068**	**0.936**	**3.052**	**1.055**	**2.030**	−0.940	2.078
LOSS	−0.012	**2.547**	−0.375	**4.293**	−0.251	**2.998**	**−1.378**	**1.015**
DOLVOL	0.772	**0.001**	0.819	**0.714**	0.642	**−0.142**	−0.435	**0.098**
REPLAG	**−3.480**	0.672	**−3.189**	−0.348	**−3.740**	−1.290	**−2.342**	−0.834
SIZE	0.226	**−5.168**	**−1.475**	**−3.831**	**−2.474**	**−9.838**	−0.859	**−5.930**
LEV	0.396	0.807	−0.032	**1.583**	0.151	1.594	−0.463	1.052
DECR	0.207	4.313	0.201	**3.351**	0.112	**2.841**	−0.510	**0.098**
EARET	**1.178**	**4.671**	**0.920**	**3.125**	0.847	**3.015**	0.556	**1.815**
SAMEFIS	**1.475**	**−1.541**	**1.649**	**−0.912**	**1.186**	**−0.526**	**2.060**	−0.177
ILLIQ	0.407	**3.056**	0.239	**1.838**	0.106	**1.601**	−0.115	0.716
ARBRISK	**−1.296**	**−6.724**	−1.210	−6.368	−1.137	−8.309	−0.863	**−10.915**
NRANK	**1.182**	**−0.260**	0.679	**0.357**	**−1.771**	**−0.004**	**1.574**	0.200
TURNOVER	0.022	**−0.700**	0.204	**0.578**	0.015	**0.182**	−0.191	1.555
EXPRISK	−0.993	−3.927	0.570	−0.411	0.624	−0.948	−0.787	−1.917
BM	−0.646	**10.857**	−0.532	7.801	−0.699	8.032	−0.756	6.135
ANALYST	−0.310	**0.598**	−0.235	**6.580**	−0.498	6.379	−0.251	1.403
PASTRET	**2.874**	**−27.394**	**3.018**	−25.858	**3.076**	−25.532	**2.821**	−10.954
PRICE	1.054	**−0.313**	1.187	**−0.513**	1.187	**−0.583**	1.004	**−0.910**
RUNUP	0.386	**−3.447**	0.565	−2.773	0.584	−2.737	0.429	0.386
BETA	**−1.431**	**−1.922**	−0.810	**0.471**	−0.811	**−0.503**	**−1.420**	−2.442
Observations	170,719		170,719		170,719		170,719	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	

Table D6: Estimates - Risk-Adjusted Abnormal Returns Using Fama-French 3-Factor Model and Momentum.
Variables are listed based on the magnitude of the differences in coefficient estimates from the first and seventh columns. Underscored variable names denote changes in significance, and bold coefficient values denote significance at the 1% level. The specifications use all firm-quarters in the sample. The first two columns reports estimates (in percent) from separate regressions with no controls for interaction terms and the variable itself, respectively. The third and fourth columns add the set of controls chosen ex ante and their interactions with SUE, and the fifth and sixth columns uses the full set of controls. The last two columns reports estimates using the high-dimensional nuisance function using only the full set of controls. Standard errors are clustered by day and firm.

	OLS		OLS		OLS		DML w. post-lasso	
Separate regressions using:	*No controls*		*Selected controls*		*Full set of controls*		*Nuisance function*	
	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself	Interaction term	Variable itself
SUE		−0.440		**3.078**		**2.791**		**4.529**
REPLAG	**−3.588**	**0.987**	**−3.264**	**−0.220**	**−3.693**	**−1.324**	**−2.051**	−0.866
SIZE	0.138	**−6.169**	**−1.792**	**−3.620**	**−2.969**	**−10.144**	−1.338	**−5.758**
LOSS	0.077	**2.831**	−0.343	**3.516**	−0.270	**2.475**	**−1.170**	**0.530**
VOL	0.405	0.073	**1.111**	2.815	**1.162**	1.817	−0.817	1.851
DOLVOL	0.732	**−0.012**	**0.893**	**0.663**	0.651	**−0.138**	−0.418	**0.058**
LEV	0.435	0.909	−0.018	**1.804**	0.167	1.759	−0.499	1.104
EXPRISK	**−1.231**	−0.564	0.563	**1.042**	0.642	**1.552**	−0.383	**−1.177**
DECR	0.197	**4.242**	0.173	**3.111**	0.060	**2.919**	−0.624	**0.050**
EARET	**1.333**	**5.411**	**0.905**	**3.604**	**0.861**	**3.381**	0.548	2.121
ARBRISK	**−1.534**	**−3.231**	−1.464	**−3.595**	−1.419	**−6.056**	−0.886	−8.208
ANALYST	−0.317	**0.633**	0.043	6.997	−0.281	6.826	−0.873	**1.383**
ILLIQ	0.399	**3.350**	0.178	**1.733**	0.084	**1.535**	−0.078	0.608
SAMEFIS	**1.504**	**−2.739**	**1.654**	**−2.035**	**1.117**	**−0.629**	**1.925**	−0.220
BETA	**−1.480**	−1.333	−0.710	−0.564	−0.833	−1.342	−1.183	−0.996
PASTRET	**3.242**	**−30.213**	**3.225**	−28.785	**3.381**	−29.036	**2.973**	−12.797
BM	−0.608	**12.169**	−0.506	**8.622**	−0.645	8.459	−0.751	6.330
TURNOVER	0.143	−0.834	0.369	0.367	0.146	0.043	0.009	1.559
PRICE	1.086	**−0.272**	1.226	**−0.593**	1.199	**−0.615**	0.969	**−1.041**
RUNUP	0.656	**−3.145**	0.746	−2.736	0.714	**−2.969**	0.719	0.612
NRANK	**1.317**	**−0.606**	**0.766**	**0.104**	**−1.680**	**0.056**	**1.317**	0.118
Observations	170,719		170,719		170,719		170,719	
Fixed effects					X		X	
No. of variables	3		21		138		2,839	