Sector Index Selection in Securities Litigation Event Studies: R-Squared and Model Stability

Steven P. Feinstein, Ph.D. (Babson College)[*]

Luke C.D. Stein, Ph.D. (Babson College)[†]

O. Miguel Villanueva, Ph.D. (Boston University)[‡]

Abstract

An event study regression estimates the return a stock would have generated in the absence of company-specific news. The sector index in the regression model can affect the predicted return and thus the measured abnormal return and statistical significance. In a securities litigation context, the sector choice can thus affect determinations of market efficiency, price impact, loss causation, and damages. We examine one narrow selection rule: choosing from a set of candidate sector indices the model with the highest in-sample R-squared, a measure of fit. We compare that choice with the sector designation made by S&P Dow Jones Indices, which serves as an external benchmark (though not a uniquely "correct" event-study control).

For 30 Dow Jones Industrial Average firms, the R-squared-maximizing index differs from the S&P Dow Jones-designated index for 15 firms in 2023 and 18 firms in 2024. Restricting the comparisons to each firm's designated index and just one predetermined, plausible alternate reduces disagreement, and the same candidate has higher R-squared in both years for 25 of 30 firms. By contrast, when the candidate set comprises all ten sector indices, the index with the highest R-squared changes between years for 18 firms, and is unstable across repeated random holdout periods. These results show that, while R-squared can be an informative metric, it should not be relied upon exclusively. A defensible process should emphasize economic reasoning. When fit is considered, the candidate set should be limited and disclosed, mechanical own-stock contamination should be removed, and the fit should be evaluated out-of-sample and for stability.

A previous draft of this paper was circulated under the title "Assessing the Reliability of Using R-Squared for Selecting the Sector Index in Securities Litigation Event Studies."

[*] feinstein@babson.edu. Babson College; Babson Park, MA 02457.

[†] lcdstein@babson.edu. Babson College; Babson Park, MA 02457.

[‡] villanue@bu.edu. Boston University; 1010 Commonwealth Ave.; Boston, MA 02215.

1. Introduction

A critical element in securities litigation is quantifying the effect of a company-specific disclosure on the subject company's stock price, if any. Event studies are used for this purpose.[1] To isolate the effect of company-specific information, an event study must identify and control for marketwide and industry-sector effects. Broad-market and sector news that affects the subject stock at the same time as company news must be accounted for when estimating the abnormal return. The market factor is akin to a tide that raises or lowers many stocks together; the sector factor captures movement associated with industry conditions. A conventional event-study regression estimates how the subject stock ordinarily moves with those broader factors. The difference between the stock's actual return and the model's predicted return is the abnormal return, which is the measured effect of the company's firm-specific disclosure.

The choice of the sector index that represents the subject company's industry, and the methodology for selecting it, can become a disputed issue between forensic-finance experts. One proposed index may yield a statistically significant decline on an event date while another yields a smaller or statistically insignificant abnormal return. With one sector index, more of an event-day decline may be attributed to company-specific information; with another, more may be attributed to industrywide movement. The choice can therefore matter to case-specific price impact, loss causation, or damages analyses, particularly when sector returns are material or the event-day result lies near a significance threshold. It need not matter in every case: closely related controls may yield nearly identical conclusions, and a large company-specific event may remain significant under every reasonable specification.

The ideal regression specification is the one that best predicts the subject company's counterfactual return: the return the stock would have earned if the company-specific news at issue had not occurred. By extension, the best sector index would capture the relevant effect of industrywide information without absorbing the company-specific effect the event study is intended to measure. Although that objective is well understood, true industry effects cannot be directly observed, and therefore, neither can the identity of the ideal sector index. Analysts instead evaluate imperfect proxies.

This article studies one way analysts might choose among proposed sector indices: maximizing the regression model's in-sample R-squared. R-squared measures how well a proposed model fits the observed returns used to estimate it. In a securities case, the estimation period may cover a class period or another period, although the relevant dates and design are case-specific.

Statistical fit can be informative, but it can be a problematic selection criterion if the model-selection process relies on it mechanically or gives it decisive weight without separate economic

[1] See MacKinlay (1997); Bhagat and Romano (2002a).

justification or validation. We do not claim that exclusive reliance on R-squared is common, nor do we estimate how often analysts use this exact rule.

We review methods for identifying potential sector controls and the statistical principles and literature that both motivate the use of R-squared and caution against using it exclusively. We then perform empirical exercises that illustrate those caveats and suggest practical improvements when R-squared is used as a criterion in securities-litigation event-study modeling.

Specifically, we examine the 30 Dow Jones Industrial Average (DJIA) stocks and the ten high-level Dow Jones U.S. Industry Indices constructed by S&P Dow Jones. S&P Dow Jones assigns each firm to one of these indices. We examine how often the sector index selected by the R-squared rule aligns with the S&P Dow Jones designation. We do not assume that the S&P Dow Jones designation reveals the unobservable ideal control, but it provides an external benchmark and allows us to observe when a broad search based on fit selects an index that aligns with a designation that was made independently outside our regression exercise and before we examine returns. The comparisons also allow us to observe how often a search based on fit might choose an index with no obvious connection to the subject company's business.

We organize the empirical work into four exercises. Exercise 1 asks how often R-squared maximization agrees with the S&P Dow Jones designation, first when the choice is made from all ten high-level indices and then when it is limited to the S&P Dow Jones designated sector index and one predetermined plausible alternate based on the firm's line of business. Exercise 2 evaluates intertemporal stability across years, half-years, and rolling windows, both among all ten candidates and within the designated-versus-alternate comparison. Exercise 3 separates coefficient estimation from model ranking through repeated random holdouts. Exercise 4 asks whether an index identified using R-squared maximization in one year continues to fit well in the other year.

These exercises lead to empirical conclusions and practical recommendations for sector-index selection. Broad in-sample R-squared maximization frequently produces an index other than the S&P Dow Jones designation and is sensitive to the sample used. In litigation, an analyst therefore should not present one fit ranking as self-proving, particularly after searching a broad candidate set. At the same time, disagreement with the S&P Dow Jones designation does not itself prove that the maximizing index is economically wrong: several non-designated maximizers continue to fit well in subsequent periods. If R-squared is used as a selection criterion, best practice combines an economically justified candidate set with transparent stability and out-of-sample checks.

2. Identifying Potential Sector-Index Candidates

The first task in selecting an industry sector index for an event-study regression is to identify one or more economically reasonable candidates. R-squared cannot supply the economic rationale for putting an index into the comparison in the first place. At least five sources of candidates are available.

First, a forensic analyst may consider the comparator that the company itself disclosed for a regulated reporting purpose. Item 201(e) of Regulation S-K requires certain registrants' shareholder reports to compare cumulative total shareholder return with both a broad equity-market index and a published industry or line-of-business index, a good-faith peer group, or, in specified circumstances, issuers with similar market capitalizations.[2] The company ordinarily presents this comparator in its annual report to security holders.

This "company's choice" index has several useful features, including that it: (i) may reflect the subject company's knowledge of its own business; (ii) is publicly disclosed; (iii) is selected for a regulated purpose; and (iv) is ordinarily fixed before a later litigation dispute or event-study result. It therefore offers an objective and reproducible starting point. The disclosure requirement was designed for a shareholder-return performance graph, however, not to identify the best event study control.

Examples in which experts have used a company-disclosed comparator include *Monroe County Employees' Retirement System v. Southern Company*; *In re EQT Corporation Securities Litigation*; *Louisiana Sheriffs' Pension & Relief Fund v. Cardinal Health, Inc.*; and *Homyk v. ChemoCentryx, Inc.*[3]

Second, an analyst may use a third-party industry-classification system such as the Global Industry Classification Standard (GICS), S&P Dow Jones Industry Classification System (DJICS), the Industry Classification Benchmark (ICB), the North American Industry Classification System (NAICS), the Standard Industrial Classification system (SIC), or a Fama-French grouping.[4] These systems suggest reproducible candidate selections that are independent

---

[2] 17 C.F.R. § 229.201(e); SEC Division of Corporation Finance, "Item 201 of Regulation S-K—Staff Interpretation," Section 5, https://www.sec.gov/divisions/corpfin/guidance/execcomp201interp.htm. The staff guidance confirms that the graph is required in the annual report to security holders and need not appear separately in Form 10-K unless the filing is used to satisfy the annual-report requirement.

[3] *Monroe County Employees' Retirement System v. Southern Company*, 332 F.R.D. 370 (N.D. Ga. 2019); *In re EQT Corporation Securities Litigation*, No. 2:19-cv-00754 (W.D. Pa. 2022); *Louisiana Sheriffs' Pension & Relief Fund v. Cardinal Health, Inc.*, No. 2:19-cv-03347 (S.D. Ohio 2021); *Homyk v. ChemoCentryx, Inc.*, No. 4:21-cv-03343 (N.D. Cal. 2024). Coauthor Steven Feinstein served as a plaintiffs' expert in the *Southern Company*, *EQT*, and *Cardinal Health* matters; the authors had no involvement in *ChemoCentryx*.

[4] The Global Industry Classification Standard (GICS) was developed jointly by MSCI and S&P Dow Jones Indices; see https://www.spglobal.com/spdji/en/landing/topic/gics/. The Dow Jones Industry Classification System (DJICS) was developed by S&P Dow Jones Indices; for a description of DJICS and a comparison with GICS, see https://www.spglobal.com/spdji/en/documents/index-policies/difference-between-gics-and-djics.pdf. The Industry Classification Benchmark (ICB) was developed by FTSE Russell; see https://www.lseg.com/en/ftse-russell/industry-classification-benchmark-icb. The North American Industry Classification System (NAICS) was developed jointly by the statistical agencies of the United States, Canada, and Mexico; see https://www.census.gov/naics/. The Standard Industrial Classification (SIC) system was developed by the Interdepartmental Committee on Industrial Statistics, established by the U.S. Central Statistical Board; see https://www.osha.gov/data/sic-manual. For the construction of the Fama–French 12-industry portfolios, see Kenneth R. French, "Detail for 12 Industry Portfolios," https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library/det_12_ind_port.html.

of the event-study regression. GICS, for example, divides the market into 11 sectors and then into industry groups, industries, and sub-industries. An example in which an expert used GICS to select a sector index is *Edwards v. McDermott International, Inc.*[5] Companies sharing a classification may reasonably be viewed as occupying the same broad sector, but every taxonomy draws boundaries around firms that may have several lines of business or economic exposures that change over time.

Third, an analyst may construct an index from companies identified as peers or competitors by analysts who cover the subject company. Because different analysts may identify different peers, the construction can use a disclosed screen—for example, including only firms named by a majority or by at least a specified number or proportion of analysts. The reports reviewed, the relevant dates, and the inclusion threshold should be stated so the process can be replicated. Examples in which experts used analyst reports to identify peers include *Pearlstein v. BlackBerry Ltd.* and *In re Walter Energy, Inc. Securities Litigation*.[6]

Fourth, an analyst may construct a sector index through fundamental business analysis, using criteria such as product lines, revenue sources, size, market capitalization, customer base, geography, or shared economic exposures. To guard against results-driven selection, the construction rule should be established, disclosed, and followed before the event-study outcome is examined. Exercise 1B below uses a deliberately simple version of this approach: one alternate sector is selected based on each firm's line of business before the R-squared results are examined. Because those assignments were generated through a general-knowledge prompt rather than a documented expert screen, we use them to study candidate-set restriction, not as validated classifications.

Fifth, data-driven tools can construct an index or reconstruct and reweight an index identified through one of the other methods. Baker and Gelbach (2020), for example, show that peer indices and regularized prediction can improve single-firm event-study performance, while also finding that including a related-industry peer index matters more than the particular algorithm used to construct it.[7]

Expert analysis in the *Shoals* litigation illustrates this approach. The plaintiffs' expert began with a collection of potential peer companies drawn from a solar-industry index, a common SIC code, and the S&P 500. The analysis then used a regression method that penalizes complexity and can reduce some candidate firms' weights to zero, leaving a smaller weighted peer set. The technical name for that method is the least absolute shrinkage and selection operator, or LASSO. A related method that blends two forms of shrinkage, known as elastic net, was used to tune the procedure

---

[5] *Edwards v. McDermott International, Inc.*, No. 4:18-cv-04330; The authors had no involvement in the *McDermott* matter.
[6] *Pearlstein v. BlackBerry Ltd.*, No. 1:13-cv-07060; *In re Walter Energy, Inc. Securities Litigation*, No. 2:12-cv-00281. Coauthor Steven Feinstein served as a plaintiffs' expert in *Blackberry* and *Walter Energy*.
[7] Baker and Gelbach (2020).

and selected the LASSO special case. The analysis also separated non-news dates used to select variables from dates used to evaluate the resulting models.[8]

These approaches can be combined. A company-chosen index or third-party classification may provide a starting point. Business and analyst research may identify additional peers or exposures. Fit or predictive performance may help distinguish among and rank the resulting candidate indices. The essential first-stage discipline is to explain why each candidate sector index could capture non-company-specific forces affecting the subject stock. Tabak and Dunbar (1999), for example, recommend considering plausible candidate indices and being prepared to defend a lower-fit choice on substantive grounds.[9]

3. Evaluating Proposed Candidates

Once candidate sector indices have been identified, the question becomes how to evaluate and choose among them. The practical objective is to choose a model that predicts the return the subject stock would have earned if the company-specific information being studied had not occurred. Because that counterfactual return is unobservable, so is the ideal model. A sector label matters only insofar as the associated index helps estimate the counterfactual. Economic plausibility, in-sample fit, stability, and performance on other observations each provide different evidence about the model's usefulness for that purpose.

The practical task of evaluating proposed sector indices arises when the finance expert must adopt an index to execute an event study. However, a court may also need to evaluate competing sector indices when the experts disagree. The *Southern Company* securities litigation illustrates the task. [10] The court described a defense expert testing six regression model specifications over the class period and selecting the one with the highest R-squared. The sector index in that specification was constructed from companies Southern Company had identified in its proxy statements as compensation peers. By contrast, the plaintiffs' expert used the S&P Electric Utility Index that Southern Company had reported as its industry comparator for Regulation S-K purposes. Using their respective indices, the defense expert found that three of seven earnings announcements during the class period had statistically significant stock returns, while the

[8] Expert Report of Jonah B. Gelbach, *In re Shoals Technologies Group, Inc. Securities Litigation*, No. 3:24-cv-00334, ECF No. 127-3 (M.D. Tenn. Jan. 21, 2026), Appendix C. The report describes a candidate universe of more than 100 firms drawn from a solar-industry index, SIC 3674, and the S&P 500; selection of 17 firms using LASSO, with an elastic-net tuning exercise selecting the LASSO special case; and a split between 371 non-news training dates and 201 non-news test dates. It also removes Shoals from the solar-industry index. The authors had no involvement in the *Shoals* matter.

[9] Tabak and Dunbar (1999), p. 10 (discussing objective criteria for index construction, cautioning against blind reliance on adjusted R-squared or t-statistics, and explaining that an expert selecting a lower-fit index should be prepared to defend that choice).

[10] *Monroe County Employees' Retirement System v. Southern Company*, 332 F.R.D. 370, 389–91 (N.D. Ga. 2019). Coauthor Steven Feinstein served as a plaintiffs' expert in that matter.

plaintiffs' expert found that five of the seven were significant.[11] The plaintiffs' expert concluded that the stock reacted consistently to new information; the defense expert disputed that the stock satisfied this market-efficiency test. Establishing market efficiency is a prerequisite for class certification. The disagreement required the court to evaluate competing models and shows how sector-index selection can affect a litigation-relevant statistical conclusion.

R-squared is indeed one useful measure. It reports the fraction of a stock's return variation explained by the model fitted over the estimation period.[12] A higher R-squared means that the model tracks the observed returns more closely in that sample. It can therefore be evidence that one candidate is a closer available proxy for the ideal model. R-squared does not by itself show, however, that the sector control is economically appropriate, that the fitted relationship will remain stable, or that the model better predicts returns not used to estimate it. The measure can be informative, but it can also be misused.

The *EQT* securities litigation provides a contrasting example.[13] In that case, the defendants' expert constructed 479 regression specifications using different combinations of market and sector indices and natural-gas prices.[14] He selected the specification with the highest adjusted R-squared, which he described as having the highest explanatory power. These examples do not establish how common broad specification searches are. They do show why the size and construction of the candidate set matter and why an analyst should disclose how candidate indices were constructed, how the candidate set was compiled, and the selection rule. Comparing a small set of economically plausible indices addresses a different question than searching a large set and accepting whichever specification happens to fit best. The statistical literature reviewed next explains why relying exclusively on R-squared is problematic after a broad search.

4. Literature and Econometric Principles

The statistical literature explains both the strengths and limitations of attempting to maximize R-squared. Kennedy begins with a fact familiar to practitioners: “The coefficient of determination, R^2, is often used in specification searches.”[15] Summarizing Theil, he then writes that “The ‘correct’ set of independent variables produces a higher [adjusted] R^2 on average in repeated samples… .” Kennedy immediately adds: “This procedure is valid only in the sense that the

---

[11] *Monroe County*, 332 F.R.D. at 383–90.

[12] Relative to ordinary R-squared, adjusted R-squared penalizes the inclusion of additional predictive variables because adding predictors cannot mechanically reduce the fraction of variation explained in sample. Because the candidate specifications compared within each firm-period use the same dates and contain the same number of predictors, ordinary and adjusted R-squared yield identical rankings in every in-sample comparison reported here.

[13] *In re EQT Corporation Securities Litigation*, No. 2:19-cv-00754-RJC, ECF No. 256, at 21–22 (W.D. Pa. Aug. 11, 2022), https://www.govinfo.gov/content/pkg/USCOURTS-pawd-2_19-cv-00754/pdf/USCOURTS-pawd-2_19-cv-00754-1.pdf. Coauthor Steven Feinstein served as a plaintiffs’ expert in that matter.

[14] Expert Report of Kenneth M. Lehn, *In re EQT Securities Litigation*, No. 2:19-cv-00754-RJC, ECF No. 156-1, ¶¶ 26–27 & n.61 (W.D. Pa. June 11, 2021).

[15] Kennedy (2008), p. 79.

‘correct set’ of independent variables will produce, on average in repeated samples a higher [adjusted] R^2 than will any ‘incorrect’ set of independent variables.”[16] The words “on average in repeated samples” do critical work. The proposition does not say that the correct specification must produce the highest R-squared in a single realized sample before the analyst or the court.

An observed R-squared is a sample statistic. It varies when the observations change, even if the underlying economic relationships do not. Kennedy therefore cautions: “Any method that selects regressors on the basis of a sample statistic such as R^2 is likely to ‘capitalize on chance’ – to select a regressor because of an accidental feature of the particular sample at hand.”[17] Searching more candidates creates more opportunities for one candidate to benefit from such sample-specific noise.

Other standard authorities state the warning more bluntly. Dhrymes (1970) describes the “game” of maximizing adjusted R-squared.[18] Stock and Watson warn that “heavy reliance on the [adjusted] R^2 (or R^2) can be a trap.” They explain, “In applications, maximizing the R^2 is rarely the answer to any economically or statistically meaningful question.”[19] Cramer writes: “These measures of goodness of fit have a fatal attraction. Although it is generally conceded among insiders that they do not mean a thing, high values are still a source of pride and satisfaction to their author, however hard they may try to conceal these feelings.”[20] Campbell, Lo, and MacKinlay prescribe the corresponding discipline: “[t]he most effective means of reducing the impact of overfitting and data-snooping is to impose some discipline on the specification search by *a priori* theoretical considerations.”[21] These authorities do not make statistical fit irrelevant. They explain why fit should be applied within an economically justified and transparent specification process rather than treated as a self-validating answer.

The problem of rankings driven by random chance becomes more acute when the candidate set expands. Kennedy describes the pitfall directly: “It is worth reiterating that searching for a high R^2 or a high [adjusted] R^2 runs the real danger of finding, through perseverance, an equation that fits the data well but is incorrect because it captures accidental features of the particular data set at hand (called ‘capitalizing on chance’) rather than the true underlying relationship.”[22] Restricting the search to a reasonably small set of candidates with a defensible economic basis mitigates, but does not eliminate, that problem. Stability checks and evaluation on observations not used to fit the coefficients provide discipline and support for a selection.

Model design, prediction, and post-selection inference are distinct pursuits, but model selection affects the validity of inference. Model design addresses which controls are economically

---

[16] Theil (1957), pp. 41–51; Kennedy (2008), pp. 80, 89.
[17] Kennedy (2008), p. 102.
[18] Dhrymes (1970), pp. 177–185.
[19] Stock and Watson (2006), p. 202.
[20] Cramer (1987), p. 253.
[21] Campbell, Lo, and MacKinlay (1997), p. 524.
[22] Theil (1957), pp. 41–51; Kennedy (2008), pp. 80, 89.

plausible before looking at results. Prediction asks how a fitted specification performs on observations not used to estimate it. Post-selection inference uses a selected design to test hypotheses and draw conclusions. Leeb and Pötscher (2005), Berk et al. (2013), and Belloni, Chernozhukov, and Hansen (2014) show why inference after model selection cannot generally proceed as though the chosen model had been fixed in advance.[23] Merely selecting the largest in-sample R-squared does not necessarily produce the model that is best for post-selection hypothesis testing and drawing inferences.

Careful event-study design can also help support valid inference. When the purpose is to predict the return that would have occurred absent an event, the event date should not be used to estimate the expected-return relationship. Bhagat and Romano (2002a) and MacKinlay (1997) describe the standard sequence of defining the event, estimating expected returns outside the event window, and then evaluating abnormal performance; Bhagat and Romano's companion article surveys applications in corporate law and regulation.[24] Additionally, single-firm, single-event studies have statistical characteristics that distinguish them from cross-sectional studies examining many firms or many events collectively. Gelbach, Helland, and Klick (2013) propose inference procedures designed specifically for these settings.[25] Our exercises address sector choice in single-firm event studies rather than considering event-date significance test designs.

Another concern that may affect the reliability of R-squared for identifying a sector index can arise in the presence of simultaneity or reverse causation. In the context of an event study, simultaneity can occur when the returns of the subject stock influence the returns of the sector index, rather than the other way around. The clearest case is mechanical own-stock inclusion. If the subject company is a constituent of a candidate sector index, the stock's returns contribute to the sector index returns, thereby mechanically increasing their measured association and potentially inflating R-squared. A high R-squared is then no longer clean evidence that the index predicts the stock's counterfactual return when company-specific news is absent. The standard response, which we follow for the S&P Dow Jones-designated index in our empirical exercises below, is to remove the subject stock from its own index.

Mechanical own-stock inclusion should be distinguished from common shocks and information spillovers, which can also produce contemporaneous co-movement but operate through different mechanisms. A marketwide or industrywide shock may move the subject company and its peers together, but capturing that common movement is precisely the purpose of a market or sector control. Separately, a company disclosure, such as an earnings announcement, may convey information about the company's peers or broader industry. For example, a large retailer's positive or negative earnings announcement may change investors' expectations about consumer demand and the earnings of other retailers. Patton and Verardo (2012) address such cross-firm

[23] See Leeb and Pötscher (2005); Berk et al. (2013); Belloni, Chernozhukov, and Hansen (2014).
[24] See MacKinlay (1997), pp. 14–19; Bhagat and Romano (2002a, 2002b).
[25] Gelbach, Helland, and Klick (2013).

information spillover effects.[26] Information spillover can increase the measured association between the subject company and a candidate sector index, and the effect can differ across candidate indices.

Whether spillover is a problem that should be addressed depends on the information at issue and the purpose of the event study. A spillover caused by the company disclosure would not have occurred in the counterfactual scenario without that disclosure, which may suggest that a sector control is absorbing part of the event being studied. But if the disclosure reveals information that properly belongs to the sector – for example, new information about industry demand – then the sector index may be doing its job of controlling for industrywide information. On the other hand, if alleged misconduct inflated both the subject company's stock and peer-company stocks, and a corrective disclosure dissipated that inflation in both, using the peer movement as a control could attribute too much of the subject company's loss to the sector. Thus, whether and how to adjust for information spillover depends on the facts and circumstances of the case, the counterfactual the event study is intended to estimate, and the ultimate purpose of the analysis. When using R-squared for sector index selection, analysts should explain why movements in the chosen sector index provide an appropriate control for the stock's counterfactual return.

Basic principles and the statistics literature therefore support a central point: R-squared may be informative for model selection, but it has limitations and can be misused. The practical lesson is not to ignore fit. Rather, when fit is used as a criterion, the entire model-selection process should be transparent, well reasoned, and auditable: explain the economic rationale for every candidate; disclose the candidates and fit statistics considered; remove mechanical own-stock contamination; compare models on the same observations; account for differences in model dimension; test the stability of the ranking; and evaluate predictions on observations not used to estimate the coefficients.

5. Empirical Design

Our empirical design applies the preceding principles through four exercises, each aimed at a different question about R-squared-based selection: (1) Does the index selected by maximizing R-squared agree with an external designation? (2) Does the R-squared ranking remain stable when the sample period changes? (3) Does it remain stable when coefficients are estimated on different observations from those used to rank the candidates? (4) Does a winner identified in one year continue to fit well in the other year?

Each exercise uses the same fixed cohort of 30 firms corresponding to the DJIA roster effective at the opening of trading on November 8, 2024. We apply that roster retrospectively to both 2023 and 2024, recognizing that it is not the point-in-time DJIA roster throughout either year. Amazon joined the DJIA in February 2024, and NVIDIA and Sherwin-Williams joined in November

[26] Patton and Verardo (2012), pp. 2789–90.

2024.[27] The fixed cohort provides a transparent, balanced sample of large firms, but it is of course not representative of all public companies.

Selection based on a future-date roster could introduce selection or survivorship concerns. Although we are not predicting future returns using future information, inclusion in the November 2024 roster may be associated with firm characteristics that also affect fit or stability. Interpretation of our results should therefore consider possible systematic influence from the roster choice, and we treat the results as descriptive rather than representative. Expanding the analysis to a wider, point-in-time universe of firms would address this limitation in future work. Table 1 identifies the firms, the sector index designated for each by S&P Dow Jones, and the alternate index used in Exercise 1B.

[PLACE TABLE 1 HERE]

For each firm and period, we estimate ten versions of the same return model, substituting each of the ten high-level Dow Jones U.S. Industry Indices ("DJI Indices") as the sector factor:[28]

$$Subject\ Stock\ Return = \alpha + \beta_1 Market\ Return + \beta_2 Sector\ Return + \varepsilon.$$

All returns are daily logarithmic returns. Focal-stock daily total returns and daily constituent weights come from S&P Dow Jones constituent extracts. The ten industry-factor returns are calculated from DJI gross-dividend total-return index levels. Market returns are computed from the value-weighted CRSP NYSE/AMEX/NASDAQ/ARCA Market Index, which includes dividends. S&P Dow Jones assigns each focal firm to one of the ten DJI Indices using its classification system.[29] For each respective stock, we refer to this as the "S&P Dow Jones-designated index" or the "designated index." The S&P Dow Jones designation is external to our regression comparison and known before the regression analysis. We do not assume that the designated index is the uniquely correct event-study control; rather, it provides a reference against which to compare other choices.

[27] S&P Dow Jones Indices announced that Amazon would replace Walgreens Boots Alliance effective before trading on February 26, 2024, and that NVIDIA and Sherwin-Williams would replace Intel and Dow effective before trading on November 8, 2024. See S&P Dow Jones Indices, "Amazon.com Set to Join Dow Jones Industrial Average" (Feb. 20, 2024), https://www.spglobal.com/spdji/en/documents/indexnews/announcements/20240220-1470711/1470711_djiadjtawbajblu-feb2024.pdf; S&P Global, "NVIDIA and Sherwin-Williams Set to Join Dow Jones Industrial Average" (Nov. 1, 2024), https://press.spglobal.com/2024-11-01-NVIDIA-and-Sherwin-Williams-Set-to-Join-Dow-Jones-Industrial-Average-Vistra-to-Join-Dow-Jones-Utility-Average.

[28] S&P Dow Jones Indices, *Dow Jones Global Indices Methodology* (May 2025), p. 15, https://www.spglobal.com/spdji/en/methodology/article/dow-jones-global-indices-methodology/. S&P Dow Jones calls these highest-level groupings "industry indices" and calls finer divisions "sector indices." We call the highest-level groupings "sectors" throughout.

[29] S&P Dow Jones Indices, *Dow Jones Global Indices Methodology* (May 2025), pp. 23–24, https://www.spglobal.com/spdji/en/methodology/article/dow-jones-global-indices-methodology/.

The designated index contains the focal firm, so we reconstruct the index to remove it.[30] This adjustment removes one source of mechanical simultaneity by preventing the stock's own return from increasing the fit of its designated index. The other nine DJI Indices do not contain the focal firm under the S&P Dow Jones classification, so no reconstitution is necessary.

For each firm-period, all ten regressions use exactly the same observation dates and contain the same number of estimated parameters. The full-year samples contain 249 observations in 2023 and 251 in 2024.

Exercise 1: Alignment with the S&P Dow Jones designation. Exercise 1A estimates each model separately over calendar years 2023 and 2024 and asks which of the ten DJI Indices maximizes R-squared. It then asks whether that index matches the designated index. Exercise 1B restricts the comparison to two indices: the designated index and one reasonable alternate DJI Index for each firm. Comparing two economically plausible indices answers different questions than does an unconstrained comparison across all ten candidates.

To generate an alternate index for each stock, we prompted ChatGPT 5.6 to use general business knowledge to choose from the remaining DJI Indices one sector that appeared plausible given the focal company's line of business, without examining any R-squared results.[31] The resulting outcome-blind assignments are reported in Table 1. They appear plausible but are not expert or externally validated classifications; we use them to study candidate-set restriction, not as validated classifications.

If disagreement falls sharply when the candidate set is restricted, that would show that broad searching contributes to the ten-index result and that the number of candidates can affect the role of sample-specific chance. If plausible alternates still frequently have higher R-squared, that would show that third-party designation and statistical fit provide different information and that neither can reliably proxy for the other. Neither finding would establish that a particular candidate is the correct counterfactual control.

Exercise 2: Stability across estimation periods. Exercise 2A compares the ten-candidate maximizers in 2023 and 2024 and then repeats the maximization in four non-overlapping half-year samples and fourteen rolling 120-trading-day windows. Frequent changes after modest shifts in sample dates would make the original ranking less persuasive as evidence of a stable relationship. Exercise 2B constrains the candidate set to the designated index and the reasonable alternate. Greater stability in the restricted pair exercise would suggest that identifying a small set of plausible candidates improves reproducibility, although stability alone would not identify the better counterfactual model.

---

[30] The constituent data include daily index weights. We remove the focal firm using its lagged weight so that the current focal-stock return is not mechanically embedded in the S&P Dow Jones-designated factor.

[31] The exact prompt was: "I'd also like you to add a separate 'plausible index' column (ticker and sector) that uses your best judgement about each company to pick an alternate sector index that the company might best fit into given its line of business." The resulting assignments are reported in Table 1.

Exercise 3: Random holdout. For every firm-year, we first remove the four earnings-announcement event dates, which are likely to contain firm-specific information.[32] We then conduct 500 deterministic random splits of the remaining trading days. Each split holds out 50 days, fits all ten models on the other 195 days in 2023 or 197 days in 2024, and ranks the models' by R-squared computed on only the held-out days.[33] This exercise separates the data used for model estimation from data used for model evaluation. The held-out returns are not used to estimate the regression coefficients. They do determine the maximizer in each split, so they do not constitute an independent, out-of-sample evaluation. If the same index repeatedly wins across splits, that would support the stability of its predictive relationship. If the winner varies widely, it would show that separating estimation from ranking does not eliminate sample sensitivity.

Exercise 4: Later-year evaluation. For firms whose 2023 full-year maximizer differs from the designated index, we carry the identity of that maximizer into 2024. We then re-estimate both the R-squared-maximizing model and the designated-index model on 2024 data and compare their 2024 fit. We also report the reverse comparison, using the 2024 maximizer as the identified index and 2023 as the evaluation year. If the identified index continues to outperform the designation, its original win is harder to dismiss as a one-sample accident; if it does not the original ranking is not durable, which suggests the original result may have been driven by sample-specific effects. Because the coefficients are re-estimated in the evaluation year, this exercise tests the persistence of the index relationship, not the predictive accuracy of the earlier year's fitted coefficient vector or permanent structural superiority.

6. Results

The tables are designed to let the reader see each comparison directly. Table 1 reports the S&P Dow Jones-designated and predetermined alternate indices for every firm. Tables 2 and 3 show the designated index's R-squared and rank, the alternate index's R-squared, and the sector that maximizes R-squared across all ten candidates. Each displayed fit cell gives ordinary R-squared followed by adjusted R-squared in parentheses; the rankings are identical under the two measures. Panel A contains the firms for which the maximizing and designated sectors match; Panel B contains the firms for which they differ. Table 4 compares full-year rankings across 2023 and 2024, both among all ten candidates and within the designated-versus-alternate pair. Table 5 summarizes the shorter-window, random-holdout, and later-year exercises.

[32] See Ball and Brown (1968); Beaver (1968); Patell and Wolfson (1984); Beaver, McNichols, and Wang (2018, 2020).

[33] For a held-out set, we calculate $R^2_{test} = 1 - \sum_t (y_t - \hat{y}_t)^2 / \sum_t (y_t - \bar{y}_{test})^2$, where $\hat{y}_t$ denotes the predicted return calculated using regression coefficients estimated only on the training dates. Test-set R-squared can be negative; candidates are ranked by the resulting value.

6.1 Alignment with S&P Dow Jones-designated industries

Exercise 1A: Ten candidates.

In 2023, the R-squared-maximizing index differs from the S&P Dow Jones-designated index for 15 of 30 firms. In 2024, it differs for 18 of 30 firms. Sector index choice based on R-squared, when selecting from among all ten candidate indices, therefore aligns with the external designation in 45% of the 60 firm-years.

[PLACE TABLE 2 HERE]

[PLACE TABLE 3 HERE]

Some disagreements are economically counterintuitive. The Healthcare Index maximizes full-year R-squared for McDonald's in both years, while the Utilities Index does so for Coca-Cola in 2023. A reader should of course not infer from these labels alone that Healthcare is the “right” control for McDonald's or Utilities the “right” control for Coca-Cola. The result instead shows that a broad fit search can produce a choice that requires a substantial economic explanation. Nor does the disagreement prove that the S&P Dow Jones designation is better: common risk exposures, correlated sector returns, and broad classification boundaries can generate economically meaningful non-designated associations.

The result is not unique to the S&P Dow Jones Industry Classification System that we used for this exercise. We repeated the broad-sector comparisons using all eleven S&P 500 GICS high-level sector indices. We excluded focal stocks from their respective indices using an approximated weight computed from the stock’s market capitalization and the market capitalization of the entire index. The R-squared-maximizing sector index matched the GICS assignment in 30 of 60 firm-years, compared with 27 of 60 under the primary S&P Dow Jones analysis. This similar alignment rate corroborates the original finding that index selection based on R-squared maximization frequently departs from the designations of an external source.[34]

For practical applications, Exercise 1A shows why the candidate list cannot be treated as an incidental input. If the analyst searches every broad sector or an expansive universe of candidates

[34] Because the GICS weights are reconstructed rather than observed official index weights, we treat this result as an approximate robustness diagnostic rather than an exact replication. The comparison uses all eleven S&P 500 GICS sector total-return indices, although none of the 30 focal firms is assigned to Utilities or Real Estate. Because official point-in-time constituent weights were unavailable, we approximate each firm’s prior-day weight in its assigned GICS sector as its market capitalization multiplied by S&P’s investable weight factor, divided by Bloomberg’s reported market capitalization for the sector index, and use that reconstructed weight to subtract the firm’s return. The assigned sector maximizes R-squared for 17 firms in 2023 and 13 in 2024. As a diagnostic of mechanical own-stock inclusion, we also conduct comparisons using the as-published index returns. In those comparisons, the maximizing sector matches the GICS assignment for 23 firms in 2023 and 19 in 2024, while the maximizing DJI Index matches its designation for 22 firms in 2023 and 20 in 2024—42 of 60 firm-years under either classification system. The higher alignment in both as-published comparisons indicates that retaining the focal stock can materially increase the assigned index’s measured fit. Because the GICS weights are reconstructed rather than official published constituent weights, the stock-excluded GICS results remain approximate.

and then relies only on the winner's R-squared, the choice may ultimately be driven by things other than a true structural relationship. The process may generate a sector with no obvious connection to the company's business. The analyst should disclose the search and explain the economic basis for the result rather than presenting the fit statistic as self-validating.

Exercise 1B: Two candidates.

Candidate-set construction materially affects the result. In the restricted comparison, the designated index has higher R-squared than the single predetermined alternate for 23 of 30 firms in 2023 and 20 of 30 in 2024. Thus, the designation-disagreement rate falls from 50% to 23% in 2023 and from 60% to 33% in 2024 when the comparison is limited to two candidates. The decline shows that restricting the comparison to a small, predetermined set reduces the opportunities for a more remote sector to win because of sample-specific fit.

The remaining disagreements also have a practical point. Even against only one outcome-blind alternate chosen as plausible from business fundamentals, the S&P Dow Jones-designated index does not maximize R-squared for 7 firms in 2023 and 10 firms in 2024. The restricted comparison shows why an external classification is useful as an outcome-blind benchmark, but not necessarily as a conclusive choice. The results identify cases in which a predetermined, economically plausible alternative fits better, but they do not establish that the external classification is inaccurate or that the higher-fit alternative is the better counterfactual control. Exercise 1B therefore supports identifying plausible candidates without reference to fit and then choosing among them based on both economic rationale and statistical performance.

6.2 Stability across estimation periods

Exercise 2A: Ten-candidate stability. Table 4 shows that for 18 of 30 firms, the index that maximizes R-squared in 2023 does not do so in 2024. Panel A of Table 5 reports the shorter-window comparisons. Across four non-overlapping half-years, the R-squared-maximizing index differs from the S&P Dow Jones designation in 70 of 120 firm-periods. Considering the four possible half-year periods for each of the 30 firms, 25 firms have more than one R-squared-maximizing industry, and the average firm has 2.4 distinct maximizers.

[PLACE TABLE 4 HERE]

The rolling analysis examines seven overlapping 120-trading-day windows in each calendar year, advanced successively by 20 trading days. The R-squared-maximizing index differs from the designation in 243 of 420 firm-windows. Twenty-five firms have more than one maximizer, and the average firm has 3.2 distinct maximizers. Because adjacent windows share 100 observations, they are not independent replications. Their value is diagnostic: modest changes in the start and end dates often change the winner.

In practice, a model should not appear persuasive merely because it wins over one conveniently bounded calendar or estimation period. If reasonable changes to the start or end date repeatedly change the R-squared-maximizing index, the analyst should report that instability and rely more heavily on economic justification and case-specific robustness.

Exercise 2B: Two-candidate full-year stability. The restricted comparison is substantially more stable. The same candidate has higher R-squared in both full years for 25 of 30 firms. The designated index wins in both years for 19 firms, the alternate wins in both years for 6, and the winner switches for only 5. This result reinforces the value of narrowing the candidate set on economic grounds: a designated-versus-plausible-alternate comparison is much more stable across these two full-year samples than a ten-sector horse race. But the six persistent alternate wins also reinforce the other side of the conclusion. A third-party designation can provide an objective starting point while remaining contestable on R-squared grounds. Persistence still does not establish that the alternate is economically appropriate or that it predicts the unobservable event-date counterfactual better.

6.3 Random-holdout results

Exercise 3 asks a simple question: if models' regression coefficients are estimated on one set of dates, which model best predicts returns on a *different* set of dates? Panel B of Table 5 reports 30,000 random-holdout splits (500 for each of the 60 firm-years). The designated index maximizes held-out R-squared in 44.7% of 2023 splits and 34.7% of 2024 splits, or 39.7% overall. It is the modal maximizer for 16 of 30 firms in 2023 and 14 of 30 firms in 2024. Nonetheless, even when model estimation is separated from evaluation, selecting based on R-squared aligns with the external designation less than 50% of the time. Moreover, repeated splits produce an average of 7.3 distinct maximizers per firm in 2023 and 7.7 in 2024. Separating estimation from ranking does not make the identity of the maximizer stable.

The 50 held-out dates determine the maximizer in each split, so the maximum held-out R-squared is itself the largest of ten noisy estimates. Performing well in the holdout period is therefore not an independent validation of the winning model. A further evaluation set would be needed to estimate the winner's predictive performance without that selection effect, which is the spirit of our Exercise 4. However, the narrower result here is still useful: broad-set rankings remain highly sample-dependent even when coefficients are estimated on different observations from those used to rank the candidates.

In practice, the exercise shows the value of reserving dates not used to estimate the coefficients. It also shows why a single holdout winner should not automatically replace a single in-sample winner. Repeated splits and a separate final evaluation set can provide useful evidence, but economic rationale remains important.

6.4 Later-year evaluation

Exercise 4 evaluates an index identified in one year using data from the other year. Panel C of Table 5 reports these comparisons. Among the 15 firms with a non-designated 2023 maximizer, that index has higher R-squared than the S&P Dow Jones-designated index in 2024 for 12 firms.

That is, a non-designated winner in 2023 usually beat the designated index again in 2024.[35] In 2024, the mean difference between the 2023 maximizer's R-squared and the designated-index's R-squared was positive 0.023. The result indicates that the association between many 2023 maximizers and the focal stocks remains strong in the later year. It does not evaluate a 2023-fitted coefficient vector on 2024 returns because both models' coefficients are re-estimated in 2024.

The temporally reversed comparison is weaker. Among the 18 non-designated 2024 maximizers, 9 had higher R-squared than the designated index when evaluated on 2023 data, with a mean difference in 2023 R-squareds near zero. The asymmetric results (going forward versus backward in time) show why disagreement with S&P Dow Jones cannot simply be attributed to chance and why analysts may want to consider both stability and predictive performance rather than alignment with a classification alone.

For litigation practice, later-period and out-of-sample evidence can strengthen the case for a non-designated index when the business relationship is stable and the evaluation period is free of the events at issue. It can also reveal that a seemingly strong in-sample choice does not persist. The evaluation period and any re-estimation of coefficients should be described clearly.

[PLACE TABLE 5 HERE]

7. Practical Implications and Limitations

The choice of sector index affects two parts of an event study that can matter in litigation. It affects predicted returns and therefore the size of abnormal returns. It can also affect the model's residual variance and therefore the test of statistical significance. A different sector index will not necessarily change the statistical conclusion, but it can do so when sector returns are material or the event-day result is near a threshold.

An untabulated significance-overlap diagnostic illustrates the point. In the 33 firm-years for which the designated and full-year maximizing indices differ, the designated-index model identifies 359 daily returns as statistically significant at the 5% level and the R-squared-maximizing model identifies 353; 271 dates overlap. Thus, the models agree on most dates, but each fails to flag roughly one-quarter of the dates flagged by the other. These are in-sample trading-day classifications, not tests of alleged misrepresentations or corrective disclosures, and there is no objective yardstick for deciding whether the disagreement rate is economically large. The exercise nevertheless shows how sector choice can matter when a litigation-relevant return

[35] This later-year-evaluation exercise is different from the stability tests in Exercise 2, which examine whether a winner in one period won again in the next. In those tests, the focus was on the selection methodology, assessing whether it is stable and picks the same winner in successive periods. The later-year-evaluation assessed in Exercise 4 focuses on the later performance of a prior winner, relative to the designated index.

lies near a significance threshold. An analyst should evaluate the actual event date under each economically reasonable specification.

The same caution applies to damages. Changing the sector control can move a particular abnormal return in either direction. A model that fits the estimation sample more closely has a smaller in-sample sum of squared residuals, but that mechanical fact does not establish that it will produce systematically higher or lower abnormal returns on properly separated litigation events. The effect in a particular case depends on the event date, the sign and size of the sector movement, the estimated coefficients, confounding information, and the applicable legal and damages framework.

The simultaneity discussion has a similar implication. A value-weighted sector index that includes the subject stock can appear to fit well partly because company-specific news moves both the stock and the index containing it. Unless the focal stock is removed, the fit statistic does not cleanly answer how well the index predicts the stock's return when that company-specific news is absent. After removal, the analyst must still decide whether movements among the remaining peers reflect an appropriate common shock, an informative industry spillover, or movement that should not be used to absorb the event at issue. Simultaneity can produce a persistently high but misleading R-squared ranking.

The paper has important limits. It does not estimate how often broad R-squared searches occur in litigation, determine an acceptable designation-disagreement rate, or establish that S&P Dow Jones identifies the true event-study model. The GICS robustness comparison uses reconstructed rather than official point-in-time constituent weights, so its focal-stock-excluded returns are approximate. It also does not test Fama-French portfolios, multifactor specifications, or data-driven peer indices, any of which may be preferable in a particular application. Exercise 1B uses only one alternate sector per firm. Although the alternate assignments were made before the R-squared results were examined, firm-specific rationales were not elicited and the assignments have not been externally validated.

The sample is also narrow. It consists of large, visible firms in a roster fixed as of November 2024 and applies that roster retrospectively to two years. Because inclusion in that future-dated roster may be associated with firm characteristics that also affect fit or stability, the results are descriptive rather than representative. Multi-line firms may have exposures poorly represented by any single high-level sector index, and relationships observed in 2023–2024 may not persist in other regimes. The rolling windows overlap. The random-holdout exercise uses its holdout to identify the maximizer rather than to evaluate a previously specified index, and the later-year exercise covers only one transition.

Finally, we do not use these exercises to infer whether any actual alleged misrepresentation or corrective disclosure produced a significant abnormal return. We also do not attempt to assess whether or how any particular model or selection criterion affects loss causation determinations or damages. These questions require case-specific event dates, confounding-information analysis, a properly separated estimation period, and appropriate single-event inference. Brav and Heaton (2015) explain why low power, confounding information, and selection on statistically detected effects can complicate price-impact and damages inferences in single-firm litigation event studies.[36] Our contribution is narrower: to help assess what economic classification, in-sample fit, stability, and simple validation exercises can and cannot establish about sector-control selection.

8. Conclusion

Selecting a sector index requires economic judgment and can benefit from statistical evidence. The ideal control would predict the stock's return in the absence of the company-specific news being studied, but that counterfactual cannot be observed directly. A company disclosure or third-party designation offers an objective, pre-specified starting point. R-squared provides evidence about fit and can be useful in the sector index selection process. However, it should not, by itself, end the inquiry.

Our four exercises show why. When the choice is made from ten broad indices, the R-squared maximizer agrees with the S&P Dow Jones designation in only 45% of firm-years. When the comparison is limited to the designation and one predetermined alternate, disagreement becomes less frequent and the same candidate wins in both years for 25 of 30 firms. Even then, the alternate wins in both years for six firms. Across the broader search, the R-squared-maximizing index changes for many firms across years, half-years, rolling windows, and random holdouts. Those results show that both the candidate set and sample dates can strongly influence the winner, while a limited, economically grounded comparison can be more stable.

Economic plausibility, external classification, in-sample fit, stability, and later-period performance answer different questions informing sector index selection. When relying on R-squared in a litigation event study, analysts should also: identify and disclose economically plausible candidates before drawing conclusions; remove the subject stock from any index that contains it; compare candidates on common observations and account for differences in model dimension; and test whether the results survive reasonable changes in the estimation period and predict observations not used to estimate the coefficients. Relying exclusively on fit is ill-advised.

[36] Brav and Heaton (2015).

References

Baker, A. C., and Gelbach, J. B. (2020). Machine Learning and Predicted Returns for Event Studies in Securities Litigation. *Journal of Law, Finance, and Accounting*, 5(2), 231–272.

Ball, R., and Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, 6(2), 159–178.

Beaver, W. H. (1968). The Information Content of Annual Earnings Announcements. *Journal of Accounting Research*, 6 (Supplement), 67–92.

Beaver, W. H., McNichols, M. F., and Wang, Z. Z. (2018). The Information Content of Earnings Announcements: New Insights from Intertemporal and Cross-Sectional Behavior. *Review of Accounting Studies*, 23(1), 95–135.

Beaver, W. H., McNichols, M. F., and Wang, Z. Z. (2020). Increased Market Response to Earnings Announcements in the 21st Century: An Empirical Investigation. *Journal of Accounting and Economics*, 69(1), 101244.

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives*, 28(2), 29–50.

Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013). Valid Post-Selection Inference. *Annals of Statistics*, 41(2), 802–837.

Bhagat, S., and Romano, R. (2002a). Event Studies and the Law: Part I: Technique and Corporate Litigation. *American Law and Economics Review*, 4(1), 141–168.

Bhagat, S., and Romano, R. (2002b). Event Studies and the Law: Part II: Empirical Studies of Corporate Law. *American Law and Economics Review*, 4(2), 380–423.

Brav, A., and Heaton, J. B. (2015). Event Studies in Securities Litigation: Low Power, Confounding Effects, and Bias. *Washington University Law Review*, 93, 583–619.

Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Cramer, J. S. (1987). Mean and Variance of R-squared in Small and Moderate Samples. *Journal of Econometrics*, 35(2–3), 253–266.

Dhrymes, P. J. (1970). On the Game of Maximizing R-bar-squared. *Australian Economic Papers*, 9(15), 177–185.

Gelbach, J. B., Helland, E., and Klick, J. (2013). Valid Inference in Single-Firm, Single-Event Studies. *American Law and Economics Review*, 15(2), 495–541.

Gelbach, J. B. (2026). Expert Report of Jonah B. Gelbach. *In re Shoals Technologies Group, Inc. Securities Litigation*, No. 3:24-cv-00334, ECF No. 127-3 (M.D. Tenn. Jan. 21, 2026).

Kennedy, P. (2008). *A Guide to Econometrics*, 6th edition. Blackwell Publishing.

Leeb, H., and Pötscher, B. M. (2005). Model Selection and Inference: Facts and Fiction. *Econometric Theory*, 21(1), 21–59.

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature*, 35(1), 13–39.

Patell, J. M., and Wolfson, M. A. (1984). The Intraday Speed of Adjustment of Stock Prices to Earnings and Dividend Announcements. *Journal of Financial Economics*, 13(2), 223–252.

Patton, A. J., and Verardo, M. (2012). Does Beta Move with News? Firm-Specific Information Flows and Learning about Profitability. *Review of Financial Studies*, 25(9), 2789–2839.

S&P Dow Jones Indices. (2025). *Dow Jones Global Indices Methodology*. https://www.spglobal.com/spdji/en/methodology/article/dow-jones-global-indices-methodology/.

Stock, J. H., and Watson, M. W. (2006). *Introduction to Econometrics*, 2nd edition. Pearson.

Tabak, D., and Dunbar, F. (1999). Materiality and Magnitude: Event Studies in the Courtroom. NERA Working Paper No. 34.

Theil, H. (1957). Specification Errors and the Estimation of Economic Relationships. *Review of the International Statistical Institute*, 25, 41–51.

Table 1. Cohort and sector assignments

Ticker	Firm	S&P Dow Jones-designated sector (index)	Alternate sector (index)
AAPL	Apple Inc	Technology (DJUSTC)	Consumer Goods (DJUSNC)
AMGN	Amgen Inc	Healthcare (DJUSHC)	Technology (DJUSTC)
AMZN	Amazon.com Inc	Consumer Services (DJUSCY)	Technology (DJUSTC)
AXP	American Express Co	Financials (DJUSFN)	Consumer Services (DJUSCY)
BA	Boeing Co	Industrials (DJUSIN)	Technology (DJUSTC)
CAT	Caterpillar Inc	Industrials (DJUSIN)	Basic Materials (DJUSBM)
CRM	Salesforce Inc	Technology (DJUSTC)	Consumer Services (DJUSCY)
CSCO	Cisco Systems Inc	Technology (DJUSTC)	Telecommunications (DJUSTL)
CVX	Chevron Corp	Oil & Gas (DJUSEN)	Basic Materials (DJUSBM)
DIS	Walt Disney Co	Consumer Services (DJUSCY)	Telecommunications (DJUSTL)
GS	Goldman Sachs Group Inc	Financials (DJUSFN)	Technology (DJUSTC)
HD	Home Depot Inc	Consumer Services (DJUSCY)	Consumer Goods (DJUSNC)
HON	Honeywell International Inc	Industrials (DJUSIN)	Technology (DJUSTC)
IBM	International Business Machines Corp	Technology (DJUSTC)	Consumer Services (DJUSCY)
JNJ	Johnson & Johnson	Healthcare (DJUSHC)	Consumer Goods (DJUSNC)
JPM	JPMorgan Chase & Co	Financials (DJUSFN)	Technology (DJUSTC)
KO	Coca-Cola Co	Consumer Goods (DJUSNC)	Consumer Services (DJUSCY)
MCD	McDonald's Corp	Consumer Services (DJUSCY)	Consumer Goods (DJUSNC)
MMM	3M Co	Industrials (DJUSIN)	Consumer Goods (DJUSNC)
MRK	Merck & Co Inc	Healthcare (DJUSHC)	Technology (DJUSTC)
MSFT	Microsoft Corp	Technology (DJUSTC)	Consumer Services (DJUSCY)
NKE	Nike Inc	Consumer Goods (DJUSNC)	Consumer Services (DJUSCY)
NVDA	NVIDIA Corp	Technology (DJUSTC)	Industrials (DJUSIN)
PG	Procter & Gamble Co	Consumer Goods (DJUSNC)	Basic Materials (DJUSBM)
SHW	Sherwin-Williams Co	Industrials (DJUSIN)	Basic Materials (DJUSBM)
TRV	Travelers Companies Inc	Financials (DJUSFN)	Consumer Services (DJUSCY)
UNH	UnitedHealth Group Inc	Healthcare (DJUSHC)	Financials (DJUSFN)
V	Visa Inc	Financials (DJUSFN)	Technology (DJUSTC)
VZ	Verizon Communications Inc	Telecommunications (DJUSTL)	Technology (DJUSTC)
WMT	Walmart Inc	Consumer Services (DJUSCY)	Consumer Goods (DJUSNC)

Notes

- The cohort is the 30-firm DJIA roster effective at the opening of trading on November 8, 2024, applied retrospectively to 2023 and 2024.
- Designated denotes the S&P Dow Jones industry assignment used in the analysis; it is not assumed to be the uniquely correct event-study control.
- Before the R-squared results were examined, ChatGPT 5.6 used general business knowledge to assign one alternate sector to each firm. The alternate assignments are used only for the restricted comparison and are not expert or externally validated classifications.

Table 2. Full-year sector-index comparisons, 2023

Panel A. R-squared-maximizing sector matches the designated sector

Ticker	Firm	S&P-designated sector (index)	Designated R^2 (Adj. R^2)	Designated rank	Alternate sector (index)	Alternate R^2 (Adj. R^2)	R^2-maximizing sector (index)	Maximum R^2 (Adj. R^2)
AMGN	Amgen Inc	Healthcare (DJUSHC)	0.237 (0.231)	1	Technology (DJUSTC)	0.155 (0.148)	Healthcare (DJUSHC)	0.237 (0.231)
AXP	American Express Co	Financials (DJUSFN)	0.490 (0.486)	1	Consumer Services (DJUSCY)	0.414 (0.409)	Financials (DJUSFN)	0.490 (0.486)
CAT	Caterpillar Inc	Industrials (DJUSIN)	0.463 (0.459)	1	Basic Materials (DJUSBM)	0.420 (0.416)	Industrials (DJUSIN)	0.463 (0.459)
CRM	Salesforce Inc	Technology (DJUSTC)	0.314 (0.308)	1	Consumer Services (DJUSCY)	0.303 (0.297)	Technology (DJUSTC)	0.314 (0.308)
CVX	Chevron Corp	Oil & Gas (DJUSEN)	0.694 (0.691)	1	Basic Materials (DJUSBM)	0.210 (0.204)	Oil & Gas (DJUSEN)	0.694 (0.691)
GS	Goldman Sachs Group Inc	Financials (DJUSFN)	0.520 (0.516)	1	Technology (DJUSTC)	0.473 (0.469)	Financials (DJUSFN)	0.520 (0.516)
HON	Honeywell International Inc	Industrials (DJUSIN)	0.503 (0.499)	1	Technology (DJUSTC)	0.480 (0.476)	Industrials (DJUSIN)	0.503 (0.499)
JNJ	Johnson & Johnson	Healthcare (DJUSHC)	0.338 (0.332)	1	Consumer Goods (DJUSNC)	0.072 (0.065)	Healthcare (DJUSHC)	0.338 (0.332)
JPM	JPMorgan Chase & Co	Financials (DJUSFN)	0.486 (0.482)	1	Technology (DJUSTC)	0.416 (0.412)	Financials (DJUSFN)	0.486 (0.482)
MMM	3M Co	Industrials (DJUSIN)	0.434 (0.429)	1	Consumer Goods (DJUSNC)	0.366 (0.361)	Industrials (DJUSIN)	0.434 (0.429)
MRK	Merck & Co Inc	Healthcare (DJUSHC)	0.335 (0.330)	1	Technology (DJUSTC)	0.126 (0.119)	Healthcare (DJUSHC)	0.335 (0.330)
NVDA	NVIDIA Corp	Technology (DJUSTC)	0.437 (0.432)	1	Industrials (DJUSIN)	0.347 (0.342)	Technology (DJUSTC)	0.437 (0.432)
TRV	Travelers Companies Inc	Financials (DJUSFN)	0.294 (0.289)	1	Consumer Services (DJUSCY)	0.131 (0.124)	Financials (DJUSFN)	0.294 (0.289)
UNH	UnitedHealth Group Inc	Healthcare (DJUSHC)	0.143 (0.136)	1	Financials (DJUSFN)	0.024 (0.016)	Healthcare (DJUSHC)	0.143 (0.136)
VZ	Verizon Communications Inc	Telecommunications (DJUSTL)	0.546 (0.542)	1	Technology (DJUSTC)	0.144 (0.137)	Telecommunications (DJUSTL)	0.546 (0.542)

Panel B. R-squared-maximizing sector differs from the designated sector

Ticker	Firm	S&P-designated sector (index)	Designated R^2 (Adj. R^2)	Designated rank	Alternate sector (index)	Alternate R^2 (Adj. R^2)	R^2-maximizing sector (index)	Maximum R^2 (Adj. R^2)
AAPL	Apple Inc	Technology (DJUSTC)	0.513 (0.509)	4	Consumer Goods (DJUSNC)	0.471 (0.467)	Basic Materials (DJUSBM)	0.542 (0.538)
AMZN	Amazon.com Inc	Consumer Services (DJUSCY)	0.351 (0.346)	9	Technology (DJUSTC)	0.467 (0.463)	Financials (DJUSFN)	0.481 (0.477)
BA	Boeing Co	Industrials (DJUSIN)	0.284 (0.278)	3	Technology (DJUSTC)	0.295 (0.289)	Technology (DJUSTC)	0.295 (0.289)
CSCO	Cisco Systems Inc	Technology (DJUSTC)	0.226 (0.220)	10	Telecommunications (DJUSTL)	0.230 (0.224)	Oil & Gas (DJUSEN)	0.234 (0.228)
DIS	Walt Disney Co	Consumer Services (DJUSCY)	0.327 (0.322)	3	Telecommunications (DJUSTL)	0.335 (0.329)	Telecommunications (DJUSTL)	0.335 (0.329)
HD	Home Depot Inc	Consumer Services (DJUSCY)	0.439 (0.434)	5	Consumer Goods (DJUSNC)	0.430 (0.426)	Industrials (DJUSIN)	0.489 (0.485)
IBM	International Business Machines Corp	Technology (DJUSTC)	0.234 (0.228)	3	Consumer Services (DJUSCY)	0.189 (0.182)	Financials (DJUSFN)	0.254 (0.248)
KO	Coca-Cola Co	Consumer Goods (DJUSNC)	0.182 (0.176)	3	Consumer Services (DJUSCY)	0.121 (0.114)	Utilities (DJUSUT)	0.291 (0.286)
MCD	McDonald’s Corp	Consumer Services (DJUSCY)	0.199 (0.193)	9	Consumer Goods (DJUSNC)	0.213 (0.207)	Healthcare (DJUSHC)	0.291 (0.285)
MSFT	Microsoft Corp	Technology (DJUSTC)	0.469 (0.465)	3	Consumer Services (DJUSCY)	0.318 (0.313)	Industrials (DJUSIN)	0.489 (0.485)
NKE	Nike Inc	Consumer Goods (DJUSNC)	0.229 (0.223)	6	Consumer Services (DJUSCY)	0.247 (0.241)	Consumer Services (DJUSCY)	0.247 (0.241)
PG	Procter & Gamble Co	Consumer Goods (DJUSNC)	0.111 (0.104)	5	Basic Materials (DJUSBM)	0.092 (0.085)	Healthcare (DJUSHC)	0.242 (0.236)
SHW	Sherwin-Williams Co	Industrials (DJUSIN)	0.392 (0.388)	2	Basic Materials (DJUSBM)	0.367 (0.362)	Oil & Gas (DJUSEN)	0.394 (0.389)
V	Visa Inc	Financials (DJUSFN)	0.408 (0.403)	2	Technology (DJUSTC)	0.405 (0.400)	Healthcare (DJUSHC)	0.410 (0.405)
WMT	Walmart Inc	Consumer Services (DJUSCY)	0.113 (0.106)	5	Consumer Goods (DJUSNC)	0.146 (0.139)	Consumer Goods (DJUSNC)	0.146 (0.139)

Notes

- Each firm is compared across all ten high-level DJII indices on an identical common-date sample. The focal stock is removed from its S&P-designated-index return.
- The alternate sector is the fixed assignment reported in Table 1. The designated-index R-squared exceeds the alternate R-squared for 23 of 30 firms.
- The maximizing sector and designated-index rank are based on all ten candidates; rank 1 indicates the highest R-squared. In the column headings, 'Adj. R^2' denotes adjusted R-squared; each fit cell reports ordinary R-squared followed by adjusted R-squared in parentheses. Ordinary- and adjusted-R-squared rankings are identical because every candidate uses the same observations and number of predictors.

Table 3. Full-year sector-index comparisons, 2024

Panel A. R-squared-maximizing sector matches the designated sector

Ticker	Firm	S&P-designated sector (index)	Designated R^2 (Adj. R^2)	Designated rank	Alternate sector (index)	Alternate R^2 (Adj. R^2)	R^2-maximizing sector (index)	Maximum R^2 (Adj. R^2)
AMGN	Amgen Inc	Healthcare (DJUSHC)	0.226 (0.219)	1	Technology (DJUSTC)	0.184 (0.177)	Healthcare (DJUSHC)	0.226 (0.219)
AXP	American Express Co	Financials (DJUSFN)	0.503 (0.499)	1	Consumer Services (DJUSCY)	0.318 (0.313)	Financials (DJUSFN)	0.503 (0.499)
CAT	Caterpillar Inc	Industrials (DJUSIN)	0.491 (0.487)	1	Basic Materials (DJUSBM)	0.459 (0.455)	Industrials (DJUSIN)	0.491 (0.487)
CSCO	Cisco Systems Inc	Technology (DJUSTC)	0.288 (0.282)	1	Telecommunications (DJUSTL)	0.284 (0.278)	Technology (DJUSTC)	0.288 (0.282)
CVX	Chevron Corp	Oil & Gas (DJUSEN)	0.638 (0.635)	1	Basic Materials (DJUSBM)	0.161 (0.155)	Oil & Gas (DJUSEN)	0.638 (0.635)
GS	Goldman Sachs Group Inc	Financials (DJUSFN)	0.626 (0.623)	1	Technology (DJUSTC)	0.500 (0.496)	Financials (DJUSFN)	0.626 (0.623)
JPM	JPMorgan Chase & Co	Financials (DJUSFN)	0.561 (0.557)	1	Technology (DJUSTC)	0.392 (0.387)	Financials (DJUSFN)	0.561 (0.557)
MRK	Merck & Co Inc	Healthcare (DJUSHC)	0.061 (0.054)	1	Technology (DJUSTC)	0.027 (0.019)	Healthcare (DJUSHC)	0.061 (0.054)
SHW	Sherwin-Williams Co	Industrials (DJUSIN)	0.360 (0.355)	1	Basic Materials (DJUSBM)	0.331 (0.326)	Industrials (DJUSIN)	0.360 (0.355)
TRV	Travelers Companies Inc	Financials (DJUSFN)	0.273 (0.267)	1	Consumer Services (DJUSCY)	0.080 (0.072)	Financials (DJUSFN)	0.273 (0.267)
V	Visa Inc	Financials (DJUSFN)	0.288 (0.283)	1	Technology (DJUSTC)	0.247 (0.240)	Financials (DJUSFN)	0.288 (0.283)
VZ	Verizon Communications Inc	Telecommunications (DJUSTL)	0.425 (0.421)	1	Technology (DJUSTC)	0.125 (0.118)	Telecommunications (DJUSTL)	0.425 (0.421)

Panel B. R-squared-maximizing sector differs from the designated sector

Ticker	Firm	S&P-designated sector (index)	Designated R^2 (Adj. R^2)	Designated rank	Alternate sector (index)	Alternate R^2 (Adj. R^2)	R^2-maximizing sector (index)	Maximum R^2 (Adj. R^2)
AAPL	Apple Inc	Technology (DJUSTC)	0.261 (0.255)	10	Consumer Goods (DJUSNC)	0.263 (0.257)	Industrials (DJUSIN)	0.338 (0.332)
AMZN	Amazon.com Inc	Consumer Services (DJUSCY)	0.473 (0.468)	8	Technology (DJUSTC)	0.492 (0.488)	Industrials (DJUSIN)	0.501 (0.496)
BA	Boeing Co	Industrials (DJUSIN)	0.075 (0.068)	8	Technology (DJUSTC)	0.096 (0.088)	Basic Materials (DJUSBM)	0.099 (0.092)
CRM	Salesforce Inc	Technology (DJUSTC)	0.234 (0.228)	5	Consumer Services (DJUSCY)	0.225 (0.219)	Telecommunications (DJUSTL)	0.252 (0.246)
DIS	Walt Disney Co	Consumer Services (DJUSCY)	0.087 (0.080)	6	Telecommunications (DJUSTL)	0.088 (0.081)	Technology (DJUSTC)	0.108 (0.101)
HD	Home Depot Inc	Consumer Services (DJUSCY)	0.266 (0.260)	10	Consumer Goods (DJUSNC)	0.283 (0.277)	Technology (DJUSTC)	0.361 (0.355)
HON	Honeywell International Inc	Industrials (DJUSIN)	0.225 (0.219)	4	Technology (DJUSTC)	0.250 (0.244)	Financials (DJUSFN)	0.262 (0.256)
IBM	International Business Machines Corp	Technology (DJUSTC)	0.215 (0.209)	5	Consumer Services (DJUSCY)	0.195 (0.188)	Telecommunications (DJUSTL)	0.237 (0.231)
JNJ	Johnson & Johnson	Healthcare (DJUSHC)	0.209 (0.203)	2	Consumer Goods (DJUSNC)	0.026 (0.018)	Technology (DJUSTC)	0.249 (0.243)
KO	Coca-Cola Co	Consumer Goods (DJUSNC)	0.046 (0.038)	7	Consumer Services (DJUSCY)	0.004 (-0.004)	Healthcare (DJUSHC)	0.223 (0.216)
MCD	McDonald's Corp	Consumer Services (DJUSCY)	0.070 (0.063)	10	Consumer Goods (DJUSNC)	0.134 (0.127)	Healthcare (DJUSHC)	0.159 (0.152)
MMM	3M Co	Industrials (DJUSIN)	0.127 (0.120)	3	Consumer Goods (DJUSNC)	0.089 (0.082)	Technology (DJUSTC)	0.138 (0.131)
MSFT	Microsoft Corp	Technology (DJUSTC)	0.553 (0.549)	4	Consumer Services (DJUSCY)	0.514 (0.510)	Industrials (DJUSIN)	0.611 (0.608)
NKE	Nike Inc	Consumer Goods (DJUSNC)	0.090 (0.083)	3	Consumer Services (DJUSCY)	0.097 (0.090)	Utilities (DJUSUT)	0.099 (0.091)
NVDA	NVIDIA Corp	Technology (DJUSTC)	0.393 (0.389)	8	Industrials (DJUSIN)	0.454 (0.450)	Financials (DJUSFN)	0.556 (0.553)
PG	Procter & Gamble Co	Consumer Goods (DJUSNC)	0.051 (0.043)	5	Basic Materials (DJUSBM)	0.024 (0.016)	Utilities (DJUSUT)	0.158 (0.151)
UNH	UnitedHealth Group Inc	Healthcare (DJUSHC)	0.066 (0.059)	2	Financials (DJUSFN)	0.040 (0.032)	Technology (DJUSTC)	0.080 (0.073)
WMT	Walmart Inc	Consumer Services (DJUSCY)	0.086 (0.078)	2	Consumer Goods (DJUSNC)	0.087 (0.079)	Consumer Goods (DJUSNC)	0.087 (0.079)

Notes

- Each firm is compared across all ten high-level DJII indices on an identical common-date sample. The focal stock is removed from its S&P-designated-index return.
- The alternate sector is the fixed assignment reported in Table 1. The designated-index R-squared exceeds the alternate R-squared for 20 of 30 firms.
- The maximizing sector and designated-index rank are based on all ten candidates; rank 1 indicates the highest R-squared. In the column headings, 'Adj. R^2' denotes adjusted R-squared; each fit cell reports ordinary R-squared followed by adjusted R-squared in parentheses. Ordinary- and adjusted-R-squared rankings are identical because every candidate uses the same observations and number of predictors.

Table 4. Stability of full-year R-squared rankings

Panel A. 2023 and 2024 R-squared-maximizing sectors match

Ticker	Firm	S&P-designated sector (index)	2023 R-squared-maximizing sector (index)	2024 R-squared-maximizing sector (index)
AMGN	Amgen Inc	Healthcare (DJUSHC)	Healthcare (DJUSHC)	Healthcare (DJUSHC)
AXP	American Express Co	Financials (DJUSFN)	Financials (DJUSFN)	Financials (DJUSFN)
CAT	Caterpillar Inc	Industrials (DJUSIN)	Industrials (DJUSIN)	Industrials (DJUSIN)
CVX	Chevron Corp	Oil & Gas (DJUSEN)	Oil & Gas (DJUSEN)	Oil & Gas (DJUSEN)
GS	Goldman Sachs Group Inc	Financials (DJUSFN)	Financials (DJUSFN)	Financials (DJUSFN)
JPM	JPMorgan Chase & Co	Financials (DJUSFN)	Financials (DJUSFN)	Financials (DJUSFN)
MCD	McDonald's Corp	Consumer Services (DJUSCY)	Healthcare (DJUSHC)	Healthcare (DJUSHC)
MRK	Merck & Co Inc	Healthcare (DJUSHC)	Healthcare (DJUSHC)	Healthcare (DJUSHC)
MSFT	Microsoft Corp	Technology (DJUSTC)	Industrials (DJUSIN)	Industrials (DJUSIN)
TRV	Travelers Companies Inc	Financials (DJUSFN)	Financials (DJUSFN)	Financials (DJUSFN)
VZ	Verizon Communications Inc	Telecommunications (DJUSTL)	Telecommunications (DJUSTL)	Telecommunications (DJUSTL)
WMT	Walmart Inc	Consumer Services (DJUSCY)	Consumer Goods (DJUSNC)	Consumer Goods (DJUSNC)

Panel B. 2023 and 2024 R-squared-maximizing sectors differ

Ticker	Firm	S&P-designated sector (index)	2023 R-squared-maximizing sector (index)	2024 R-squared-maximizing sector (index)
AAPL	Apple Inc	Technology (DJUSTC)	Basic Materials (DJUSBM)	Industrials (DJUSIN)
AMZN	Amazon.com Inc	Consumer Services (DJUSCY)	Financials (DJUSFN)	Industrials (DJUSIN)
BA	Boeing Co	Industrials (DJUSIN)	Technology (DJUSTC)	Basic Materials (DJUSBM)
CRM	Salesforce Inc	Technology (DJUSTC)	Technology (DJUSTC)	Telecommunications (DJUSTL)
CSCO	Cisco Systems Inc	Technology (DJUSTC)	Oil & Gas (DJUSEN)	Technology (DJUSTC)
DIS	Walt Disney Co	Consumer Services (DJUSCY)	Telecommunications (DJUSTL)	Technology (DJUSTC)
HD	Home Depot Inc	Consumer Services (DJUSCY)	Industrials (DJUSIN)	Technology (DJUSTC)
HON	Honeywell International Inc	Industrials (DJUSIN)	Industrials (DJUSIN)	Financials (DJUSFN)
IBM	International Business Machines Corp	Technology (DJUSTC)	Financials (DJUSFN)	Telecommunications (DJUSTL)
JNJ	Johnson & Johnson	Healthcare (DJUSHC)	Healthcare (DJUSHC)	Technology (DJUSTC)
KO	Coca-Cola Co	Consumer Goods (DJUSNC)	Utilities (DJUSUT)	Healthcare (DJUSHC)
MMM	3M Co	Industrials (DJUSIN)	Industrials (DJUSIN)	Technology (DJUSTC)
NKE	Nike Inc	Consumer Goods (DJUSNC)	Consumer Services (DJUSCY)	Utilities (DJUSUT)
NVDA	NVIDIA Corp	Technology (DJUSTC)	Technology (DJUSTC)	Financials (DJUSFN)
PG	Procter & Gamble Co	Consumer Goods (DJUSNC)	Healthcare (DJUSHC)	Utilities (DJUSUT)
SHW	Sherwin-Williams Co	Industrials (DJUSIN)	Oil & Gas (DJUSEN)	Industrials (DJUSIN)
UNH	UnitedHealth Group Inc	Healthcare (DJUSHC)	Healthcare (DJUSHC)	Technology (DJUSTC)
V	Visa Inc	Financials (DJUSFN)	Healthcare (DJUSHC)	Financials (DJUSFN)

Panel C. Designated-versus-alternate winner across full years

Higher R-squared in 2023	Higher R-squared in 2024	Firms	Share
Designated index	Designated index	19	63.3%
Designated index	Alternate index	4	13.3%
Alternate index	Designated index	1	3.3%
Alternate index	Alternate index	6	20.0%

Notes

- The maximizing sector changes between the two full-year samples for 18 of 30 firms when all ten candidates are compared.
- In Panel C, the winner is whichever of the designated and predetermined alternate indices has the higher full-year R-squared; there are no ties. The same candidate wins in both years for 25 of 30 firms.
- These are stability diagnostics, not independent tests of which sector is economically appropriate.

Table 5. Stability and validation checks

Panel A. Designated-index maximization across estimation windows

Analysis	Period	Firm-periods	Designated index maximizes R^2	Firms with multiple maximizers	Mean distinct maximizers
Full-year comparison	2023	30	15/30 (50.0%)		
Full-year comparison	2024	30	12/30 (40.0%)		
Non-overlapping half-year	2023 H1	30	15/30 (50.0%)		
Non-overlapping half-year	2023 H2	30	13/30 (43.3%)		
Non-overlapping half-year	2024 H1	30	9/30 (30.0%)		
Non-overlapping half-year	2024 H2	30	13/30 (43.3%)		
Non-overlapping half-years	2023-2024 pooled	120	50/120 (41.7%)	25/30 (83.3%)	2.37
Rolling 120-day windows	2023-2024 pooled	420	177/420 (42.1%)	25/30 (83.3%)	3.20

Panel B. Random-holdout maximizers

Sample	Firm-years	Splits per firm-year	Designated index maximizes R^2	Designated index is modal maximizer	Designated index maximizes in majority	Mean distinct maximizers	Mean maximum R^2	Mean designated-index R^2	Mean maximum-minus-designated R^2
2023	30	500	6,709/15,000 (44.7%)	16/30	14/30	7.30	0.424	0.381	0.0430
2024	30	500	5,208/15,000 (34.7%)	14/30	11/30	7.67	0.340	0.276	0.0640
2023-2024 pooled	60	500	11,917/30,000 (39.7%)	30/60	25/60	7.48	0.382	0.328	0.0535

Panel C. Later-year evaluation

Identification year	Evaluation year	Non-designated maximizers	Maximizer exceeds designated index	Mean difference R^2 (Adj. R^2)	Median difference R^2 (Adj. R^2)	Mean maximizer rank	Mean designated-index rank
2023	2024	15	12/15 (80.0%)	0.0227 (0.0229)	0.0148 (0.0149)	3.27	5.40
2024	2023	18	9/18 (50.0%)	0.0002 (0.0002)	0.0002 (0.0002)	3.17	3.56

Notes

- Panel A identifies the R-squared-maximizing sector from ten candidates. The four half-years do not overlap. The fourteen 120-trading-day windows advance by 20 trading days, so adjacent windows overlap and are not independent. The multiple-maximizer columns summarize variation within firms across all periods of the indicated design.
- Panel B reports 500 deterministic splits per firm-year after four effective earnings-announcement dates are removed. Each split estimates coefficients on training dates and identifies the sector with the highest R-squared on 50 held-out dates. Modal means the designated index maximizes held-out R-squared more often than any other sector; majority means it does so in more than half of the splits.
- Because the same held-out dates determine the maximizer, Panel B's mean maximum-minus-designated R-squared is descriptive and is not an independently validated performance estimate. Panel B reports held-out test R-squared, not adjusted R-squared; the conventional adjusted-R-squared degrees-of-freedom correction is an in-sample statistic and is not applied to these held-out observations.
- Panel C carries the non-designated full-year maximizer from the identification year into the evaluation year, then re-estimates both that model and the designated-index model in the evaluation year. In Panel C, each difference is the maximizer's fit minus the designated index's fit; each cell reports the R-squared difference followed by the adjusted-R-squared difference in parentheses. Rank 1 indicates the highest R-squared in the evaluation year. The firms share an evaluation year, so the counts are reported descriptively.