Statistical Learning Can Help The Judiciary Fulfill Its Gatekeeping Role Over Expert Witnesses

Andrew Baker*

Abstract

1 Introduction

2 Expert Work and Commercial Litigation

Expert testimony has become a critical tool for attorneys to clarify complex issues that arise in disputes between businesses. Expert witnesses provide specialized knowledge that aids the court in understanding intricate technical matters, industry standards, or specialized data that are beyond the common knowledge of judges and juries. Over time, the role of expert testimony has expanded as business transactions have become more sophisticated, and courts increasingly rely on expert witnesses to bridge the gap between legal principles and the detailed factual underpinnings of commercial disputes.

Expert testimony in commercial cases was initially limited to straightforward technical matters, such as accounting practices in cases involving financial fraud. However, as industries have become more specialized and the legal environment more intricate, experts from a wider range of fields, including economics, finance, engineering, and technology, have been called to provide testimony. The rise of global commerce, digital technologies, and complex financial instruments has further driven the need for expert testimony to explain the complexities involved in modern commercial litigation.

One key area where expert testimony has seen significant growth is in damages calculations. In the past, damages were often calculated using basic methods, but as litigation in sectors like antitrust, intellectual property, and securities fraud has increased, courts require more precise models for understanding potential losses or financial harm. Economists and financial experts are now frequently used to create complex models that assess the impact of alleged misconduct, lost

*University of California, Berkeley

profits, or market manipulations. These experts can provide clarity by offering nuanced insights into causation and quantifying harm in ways that were not previously possible.

Another factor that has contributed to the increased use of expert testimony is the expanding scope of regulatory environments. With regulatory bodies like the Securities and Exchange Commission (SEC) and the Federal Trade Commission (FTC) playing an ever-greater role in enforcing business practices, litigation involving regulatory compliance has grown. Experts in fields such as securities law, environmental regulations, or telecommunications standards are often required to explain whether a company's conduct meets or violates established legal standards. Their testimony often becomes pivotal in determining the outcome of these cases, particularly when there is a need to interpret new and evolving regulations that require deep subject matter expertise.

The credibility and reliability of expert testimony have been notionally reinforced by the heightened standards established by legal precedents, such as the *Daubert* standard, which governs the admissibility of expert testimony in federal courts. This standard requires that experts use reliable methods and base their opinions on sufficient data, which has further solidified the role of experts in commercial litigation. Courts have come to expect rigorous, well-reasoned testimony, leading to an increased demand for highly credentialed experts who can withstand judicial scrutiny. As a result, the selection of expert witnesses has become a strategic decision for attorneys, with significant resources being invested in finding and vetting individuals who possess the knowledge and credibility to persuade a judge or jury.

Over time, the role of expert testimony has also been shaped by the increasing complexity of commercial relationships, often involving cross-border disputes or multinational corporations. Experts in international trade, global finance, and cross-jurisdictional regulatory compliance have become invaluable in cases where national legal systems intersect. For instance, in disputes involving international mergers or allegations of anticompetitive behavior, courts often rely on expert testimony to understand how business activities in one jurisdiction affect markets in another. This expansion of expert testimony into international commercial litigation reflects the global nature of modern commerce, where the legal, financial, and economic issues are deeply intertwined across borders.

With advancements in technology, experts are increasingly using sophisticated tools like data analytics, artificial intelligence, and forensic software to support their testimony. These tools enable experts to sift through vast amounts of data and provide insights that were previously unattainable, adding even greater depth to their analyses. The increased reliance on technology in business operations has made such tools indispensable in litigation, especially in cases involving cybersecurity, digital fraud, or data breaches, where expert testimony is needed to explain the nature and extent of the damage.

In conclusion, expert testimony in commercial litigation has grown both in importance and complexity, reflecting the evolving nature of business and legal disputes. From providing clarity on technical issues to offering detailed economic analyses, expert witnesses now play a central role in

shaping the outcomes of high-stakes commercial cases. As industries continue to advance and legal frameworks grow more intricate, the reliance on expert testimony will likely continue to increase, making it a cornerstone of modern commercial litigation.

3 *Daubert*, Judicial Gatekeeping, and Frustration With the Expert Machinations

The *Daubert* Standard is a legal rule that governs the admissibility of expert testimony in U.S. federal courts, particularly in relation to scientific and technical evidence. This standard was established by the U.S. Supreme Court in the 1993 case *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, which set forth the criteria that judges must use to determine whether proffered expert testimony is sufficiently reliable and relevant to be presented to a jury. The overarching goal of *Daubert* is to ensure that expert evidence is grounded in scientific validity rather than speculation or unreliable methodologies. The decision represented a break from the earlier standard, *Frye*, which relied on the general acceptance of the methodology within a relevant scientific community.

Under *Daubert*, judges are required to evaluate several factors to determine the admissibility of expert testimony. These include whether the theory or technique employed by the expert can be (and has been) tested; whether it has been subjected to peer review and publication; the known or potential rate of error; the existence and maintenance of standards governing the methodology's operation; and whether the theory or technique is generally accepted within the relevant scientific community. These considerations guide the court in ensuring that evidence introduced to the court is scientifically grounded.

In applying the *Daubert* Standard, judges serve a crucial role as gatekeepers. It is their responsibility to assess whether the methodology underlying the expert's testimony is not only scientifically valid, but also relevant to the case at hand. This requires judges to move beyond simply evaluating an expert's credentials or field of expertise; they are required to scrutinize the reasoning and processes that lead to the expert's conclusions. Even if a method is reliable as a general principle, it must be shown to have direct relevance to the facts in dispute for it to be admitted. Judges also exercise significant discretion in determining which of the *Daubert* factors are most applicable in each case and how heavily to weigh them. The standard does not require that all factors be met, but it does provide a framework for ensuring that expert testimony is grounded in reliable scientific principles.

In upholding their gatekeeping role, judges often hold *Daubert* hearings as part of the pretrial process to assess the admissibility of expert testimony. These hearings provide both sides an opportunity to argue for or against the use of a particular expert, and they offer judges a venue to explore the scientific foundations of the proposed evidence. The rulings made during these hearings often significantly shape the course of a trial, as the exclusion of expert testimony can

weaken a party's case or change the dynamics of the evidence presented to the jury.

The *Daubert* Standard has, arguably, had a profound impact on the use of expert witnesses in courtrooms, placing a greater burden on those experts to demonstrate not only their expertise but also the scientific rigor of their methodologies. By emphasizing factors such as testability, peer review, and error rates, the standard seeks to filter out so-called "junk science" from influencing court decisions. At the same time, it has increased the responsibility placed on judges, who must now have a degree of understanding in scientific and technical matters to effectively evaluate expert evidence. In some cases, courts may even appoint their own neutral experts to assist with this evaluation, especially when the subject matter is highly specialized.

The *Daubert* standard is the test currently used in the federal courts, and state courts are free to follow their own rules regarding expert evidence. Currently, only five states continue to use the Frye or Frye-plus standard: Illinois, Minnesota, New York, Pennsylvania and Washington. While states who have adopted *Daubert* are in the majority, some 16 states have adopted modified versions of *Daubert*: Alaska, California, Colorado, Connecticut, Hawaii, Idaho, Indiana, Iowa, Maine, Montana, New Mexico, Tennessee, Texas, Utah, and West Virginia. Regardless of the precise legal standard governing expert evidence, in every state judges are required to play some role in gatekeeping evidence provided by experts to juries.

4 Model-Driven vs. Data-Driven Estimation

[Insert discussion here about the difference between standard model-based analysis (where, e.g., an expert selects control variables or closest values in a k-NN type analysis) with the data-driven approach from statistical learning. See discussion from Breiman - Two Cultures. Discuss how statistical learning works well in settings where you care about prediction rather than model parameters, and why it would be particularly useful in cases where judges need to adjudicate between motivated experts.]

5 Empirical Examples

In this section I document, through a series of empirical examples, how statistical learning can cabin expert discretion in a judicially reliable manner.

5.1 Event Studies and Securities Litigation: The Case of *Halliburton*

5.1.1 Event Studies and Securities Litigation

Securities class action lawsuits are governed by the SEC's Rule 10b-5, promulgated under Section 10(b) of the Securities and Exchange Act of 1934. Rule 10b-5 makes it unlawful for any person

to make an untrue statement of a material fact, or to omit to state a material fact necessary to make other statements not misleading, in connection with the purchase or sale of security.[1] In a securities fraud suit brought under Rule 10b-5, plaintiffs are required to prove the existence of a material misrepresentation or omission that is made with scienter (or a mindset embracing an intent to deceive). In addition, plaintiffs bear the burden of proving reliance, which, building upon the common law of deceit, requires the plaintiff to have actually and justifiably relied on the misrepresentation in causing them to transact in the security in question. Under the Supreme Court's test in *Basic v. Levinson*,[2] there is a presumption of reliance where the defendant makes a material representation in an informationally efficient market. Finally, plaintiffs must prove loss causation—that the defendant's wrongful act was the proximate cause of the plaintiff's loss—and be able to prove class-wide damages in justiciable manner.

Antifraud cases brought under our securities laws, particularly those brought pursuant to Rule 10b-5, represent an area of commercial litigation where expert-provided evidence is often outcome determinative. Occasionally, experts are asked to opine on business and industry facts that can help a court determine the materiality of a specific piece of information. In addition, experts sometimes instruct the court on what a reasonable investor would intuit from a given disclosure. In nearly every case,[3] however, experts are hired to conduct an "event study" analysis linking specific misstatements and disclosures to the firm's stock price.

An event study is an empirical technique used to identify the effect of an event on the value of a firm's security (typically, though not always, the value of its common equity). Event study evidence is used, and often de-facto required, to support multiple of the elements of a plaintiffs cause of action; including reliance,[4] materiality,[5] loss causation,[6] and damages.[7] Each of these elements is *critically dependent* on the provision of a reliable event study by a qualified expert.[8] As some

[1] 17 C.F.R. § 240.10b-5(b) (2024)

[2] 485 U.S. 224 (1988).

[3] At least those cases that make it past a motion for summary judgment.

[4] Event studies are often used to determine whether the market for a firm's stock is informationally efficient, by analyzing whether the stock responds in a consistent and statistically significant manner to news regarding the firm's prospects.

[5] Following the Supreme Court's holding in *Halliburton II*, defendants are entitled to an opportunity to rebut "price impact" at the class certification stage. Although materiality does not need to be established for a class to be certified, defendants are now allowed to present evidence rebutting the materiality of alleged misrepresentations based on a price impact analysis. *Halliburton II*, 134 S. Ct. 2398, 2417 (2014).

[6] Loss causation is the plaintiff's burden to establish a direct connection between the alleged fraud and the economic harm to the shareholders. This harm is measured at two points in time–when the security was purchased and when the fraud was disclosed. *Dura Pharmaceuticals Inc. v. Broudo*, 544 U.S. 336, 345-46. Evidence of price distortion nearly always requires a formal event study analysis to disentangle the return on the security from other contemporaneous market changes.

[7] Damages in securities class actions follow the "out-of-pocket" damages established in *Affiliated Ute Citizens v. United States*, 406 U.S. 128, 155 (1972). Under *Ute*, defrauded purchasers of a security are entitled to the difference between the price paid for the security and the price it would have traded at had the material misrepresentation or omission not occurred. This determination also nearly always requires an event study to disentangle the effect of normal variation in returns from the fraud-induced change.

[8] Andrew Baker, *Single-Firm Event Studies, Securities Fraud, and Financial Crisis*, 68 Stan. L. Rev. 1207

have argued, "the law governing event studies has become inseparable from the substantive law governing securities fraud litigation" because "[c]ourts have effectively collapsed securities fraud actions into a single question: Whether the defendant's misrepresentation or omission created a disparity between the transaction price of a security and its true value measured by the precise reaction of the market price to the disclosure of the concealed information."[9] It is thus unsurprising how much litigant and judicial effort (and expense) has been dedicated to conducting and attacking event studies in securities disputes.

In order to conduct an event study, an expert first specifies an econometric model that relates the return of a security to the corresponding return on market factors. The expert defines an "estimation period", over which the model will be estimated, and an "event window", which is the period over which the effect of the event on the security will be analyzed. After estimating the model, an expert is able to determine the potential magnitude and significance of an event by comparing the actual return of the security over the event window to the return predicted by the model's estimated parameters. If this difference is sufficiently large in magnitude in comparison to the model's typical estimation error[10] then the expert will testify that the returns cannot be explained by normal patterns in the data. Note that, in addition to the quantitative results provided by the model, the expert still has a qualitative role to play in convincing the finder of fact that there were not other firm-specific events that occurred at the same time as the alleged misstatement or correct disclosure that could explain the residual portion of the firm's return.

A key ingredient in constructing an event study analysis is specifying the model that links the expected returns to other contemporaneous returns in the market. In support of their analysis experts will frequently cite to academic work, although the event study was created by financial economists as an empirical technique to the assess the impact of a general *type of event*—such as mergers or dividend announcements—on the value of a *set* of securities. In such a setting, modeling errors, if uncorrelated with treatment timing, can be expected to average out in the aggregation process. However, in a litigation setting we are almost always dealing with an event that only impacted one firm, without any ability to margin out prediction errors.[11]

Initial academic work used simple expected return models—including the constant mean return model (where the predicted return is equal to the average firm return over the estimation window)

(2016).

[9]Michael J. Kaufman & John M. Wunderlich, *Regressing: The Troubling Dispositive Role of Event Studies in Securities Fraud Litigation*, 15 Stan. J.L. Bus. & Fin. 183 (2009).

[10]Model's will never perfectly predict the security's return over the estimation period, so we compare the unexplained portion of the return in the event window to the *typical* unexplained component of the return over the estimation period.

[11]This core limitation to the single-firm event study has now been extensively addressed in the literature, and presents difficulties for both generating adequate predictions, and conducting valid statistical inference, especially in the presence of changes in time-varying volatility. *See, e.g.*, Andrew Baker, Single-Firm Event Studies, Securities Fraud, and Financial Crisis, 68 Stan. L. Rev. 1207 (2016), Jonah B. Gelbach et al., Valid Inference in Single-Firm, Single-Event Studies, 15 Am. L. & Econ. Rev. 495, 499 (2013), Edward G. Fox et al., *Economic Crisis and the Integration of Law and Finance: The Impact of Volatility Spikes*, Colum. L. Rev. 325 (2016).

and the market adjusted model (where the predicted return is simply the contemporaneous return on a market index). Under one particularly influential theory in academic finance, the Capital Asset Pricing Model (CAPM), the return on a stock is solely a function of its systematic risk, measured as the covariance between its return and the return on a market portfolio.[12] This led naturally to the "market model" event study approach, which estimates predicted returns using a linear regression of the firm's returns on the market index over the estimation period. The market model identifies two parameters, α, which is the expected return on the stock when the market return is zero, and β, which measures the firm's systematic risk. Unfortunately, the CAPM assumptions don't hold: β is not the only risk that explains returns. Later models, including that of Fama and French and its extensions, supplement the market risk factor with portfolio return risk factors meant to capture the effect of size, valuation, and momentum.[13] An important takeaway from this shift in the literature is that the field of empirical asset pricing largely moved away from structural, *a priori* model-based estimation to a predictive exercise in finding risk factors, or anomolies, that can predict firm returns. The state of the art methods in the literature now include using non-linear, machine learning methods to forecast predicted returns.[14]

5.1.2 *Halliburton*

Erica P. John Fund, Inc. v. Halliburton Co. was a long-running securities class action brought under Rule 10b-5 that twice made its way to the Supreme Court. Plaintiffs alleged that Halliburton and its executives issued material misrepresentations regarding the company's potential liability in asbestos litigation, its expected revenue from a series of construction projects, and the benefits of a merger.[15] The defendants initially argued that plaintiffs had not met their burden to invoke *Basic's* reliance presumption because they could not adequately plead loss causation. After winning on that theory at the district and circuit courts, the Supreme Court overturned, holding that proving loss causation is a separate inquiry from reliance, and not a requirement at the class certification stage.[16] On remand, Halliburton argued that class certification was inappropriate because the event study evidence provided by their expert to disprove loss causation also demonstrated a lack of "price impact"—i.e. proof that "the alleged misrepresentations affected the market price in the first place."[17]. This, arguably, "sever[ed] the link between the alleged misrepresentation and ... the price received (or paid) by the plaintiff,"[18] rendering the presumption of reliance from *Basic* inapplicable. Halliburton lost on this secondary argument at the lower courts, with the Supreme

[12] Eugene F. Fama and Kenneth R. French, *The CAPM is Wanted, Dead or Alive*, 51(5) J. Fin. 1947, 1948 (1996).

[13] *See* Eugene F. Fama and Kenneth R. French, *Common risk factors in the returns on stocks and bonds*, J. of Fin. Econ. 33 (1993) and Mark M. Carhart *On persistence in mutual fund performance*, 52 J. Fin. 57 (1997)

[14] Shihao Gu et al., *Empirical asset pricing via machine learning*, 33(5) Rev. Fin. Stud. 2223 (2020)

[15] *Halliburton Co. v. Erica P. John Fund, Inc.*, 573 U.S. 258, 264 (2014).

[16] *Erica P. John Fund, Inc. v. Halliburton Co.*, 563 U.S. 804, 813 (2011).

[17] *Halliburton I*, 563 U.S. at 814(2011)

[18] *Basic*, at 248.

Court again stepping in, this time in (partial) support of the company to find that "defendants must be afforded an opportunity before class certification to defeat the presumption through evidence that an alleged misrepresentation did not actually affect the market price of the stock."[19]

After the Supreme Court vacated the lower court judgements and remanded the case for further class certification proceedings, the district court ordered additional briefing on price impact and its relation to class certification. Both parties submitted additional expert reports centered around their event study evidence,[20] and, as expected, a dispute among the two experts (Chad Coffman for the funds and Lucy Allen for the company) arose. As the court noted, "[t]he determination of whether lack of price impact ha[d] been shown largely turns on the competing methodologies of the parties' experts."[21] The two experts disagreed on a number of methodological issues—from the relevant dates to analyze, the correct estimation period based on the testing dates, the use of one-day or two-day event windows, and whether, and how, to adjust for multiple testing. However, here I will focus on a more fundamental difference between the two experts that has arisen in a number of securities suits[22]—how to risk-adjust within the market model.

Under the CAPM, β completely captures the explainable portion of a stock's return. A common event study specification, frequently used in litigation, that builds upon this model is:

$$r_t = \alpha + \beta M_t + \epsilon_t$$

where r_t is the return on the company's stock on date t, α is the model intercept that captures the expected return when the market return is zero, M_t is the return on a broad market index, like the S&P 500, and β is the measure of the firm's systematic risk. ϵ_t is the model error, and reflects the fact that we can never perfectly capture the expected return on a security.

It is common to supplement this specification in litigation, where the event study is estimated for only a single security, with the inclusion of a second index designed to capture industry-specific trends in returns. Consistent with the CAPM not fully explaining the cross-section of expected returns, Baker and Gelbach (2020)[23] shows through a simulation analysis that the inclusion of a simple industry index based on two-digit SIC codes increases the out-of-sample predictive power of the market model. The experts in the *Halliburton* litigation disagreed about the proper way to adjust for the industry component of Halliburton's return prediction.

Lucy Allen, the expert retained by the defense, estimated an event study that controlled for the company's two primary lines of business: energy services and engineering and construction (E&C). She used the S&P 500 Energy Index to control for the former, and a bespoke equally-weighted index of composed of firms in the Fortune 1000 that are classified as being in the E&C

[19] *Halliburton II*, at 284.

[20] *Erica P. John Fund, Inc. v. Halliburton Co.*, 309 F.R.D. 251, 256 (N.D. Tex. 2015).

[21] *Id.* at 262.

[22] Add examples

[23] Andrew Baker & Jonah B. Gelbach, *Machine Learning and Predicted Returns for Event Studies in Securities Litigation*, 5 J.L., Fin., Acct. 231 (2020).

industry for the latter.[24] Chad Coffman, the expert retained by the funds, argued that the Allen model incorrectly controlled for Halliburton's primary business, because the S&P 500 energy index was driven in large measure by energy *producers* rather than energy *servicers*. Coffman created a separate index based off of the listed peers in Halliburton's analyst reports, another common way to generate industry indices.[25]

Table 1: Expert Model Results and Predictions

	Replication		Report Values	
	Allen (D)	Coffman (P)	Allen (D)	Coffman (P)
Model Results				
Intercept	0.00	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
	-1.18	-1.70	-0.09	-0.97
	0.24	0.09	0.93	0.33
S&P Energy Index	1.40	0.28	1.38	0.28
	(0.06)	(0.07)	(0.06)	(0.06)
	23.99	4.31	23.30	4.57
	0.00	0.00	0.00	0.00
Fortune E&C Index	0.13	0.08	0.16	0.12
	(0.06)	(0.04)	(0.06)	(0.04)
	2.40	1.85	2.78	2.87
	0.02	0.06	0.01	0.00
Industry Peer Index		0.90		0.85
		(0.04)		(0.04)
		22.70		24.04
		0.00		0.00
Predictions for December 4, 2001				
Excess Return	-0.027	-0.035	-0.029	-0.037
t-Statistic	-1.156	-2.007	-1.290	-2.270
p-value	0.248	0.045	0.198	0.024

This table reports the event study model estimates from the Lucy Allen and Chad Coffman Reports in the Halliburton Securities Lawsuit. The first two columns represent my best attempt at replication, while the second two columns present the results are showin in their reports. I also report the the event study results for December 4, 2001, including the excess return and associated t-statistic and p-values from the alternative approaches.

I report the results of both event study models in Table 1. The first two columns present my best attempt at replicating the models as described in the reports,[26] and the second two columns provide the reported values. For each estimated coefficient I report the estimate, standard error (in

[24] Allen report, at Para. 20.

[25] Coffman Report, Para 30.

[26] One reason why I do not perfectly match the estimated model coefficients in the reports is because they remove Halliburton from the S&P 500 Energy Index. Unfortunately, this requires the industry weights, which is a proprietary dataset that I currently don't have access to.

parenthesis), t-Statistic, and corresponding p-value (in that order). As can be seen from the model results, I closely, though not exactly, match the results submitted to the court. The difference in industry controls generate a dispute between the experts over one disclosure date in particular—December 4, 2001—when Halliburton announced an adverse judgment in a Texas case regarding its asbestos liability. The as-reported Allen model generates an excess return estimate of -2.9%, with a p-value of 0.20, while the Coffman model leads to a -3.7% excess return with a p-value of 0.02, below the conventional cutoff for statistical significance.

This is the type of methodological dispute that a generalist judge is ill-suited to adjudicate. Both experts provide a plausible story for the inclusion (or lack of inclusion) of different industry controls, without a clear way to resolve the dispute. Moreover, the difference matters for the resolution of the case, as the inclusion or exclusion of different disclosure dates changes the effective class period and estimates of class-wide damages.[27] Moreover, while Coffman argues for his model based on the superior adjusted-R^2 (a measure of a model's explanatory power), it is not clear that is the right way to do model selection in this setting.

In Baker and Gelbach (2020) we argue for a re-framing of the question in the litigation context. Rather than view an event study as a way to determine the best model of expected return—surely a fool's errand—we can instead view an event study as an example of a prediction problem. The opposing experts are attempting to generate a prediction of the expected return, in light of contemporaneous returns in the market, on a set of pre-determined dates. The experts and the court are not concerned with the model's parameters—namely, the weights placed on different market and industry indices—but instead are solely concerned with generating accurate predictions of the counterfactual return over the event window. Viewed from that angle, a natural way to adjust for the return on a firm's industry is to use a data-driven procedure to select peer firms based on a constructive notion for their use—the extent to which a given peer firm's returns assist in generating a valid prediction of the target firm's returns. This will avoid getting into the thicket of whether a given firm is truly a valid industry peer,[28] which is not the probitive question for the prediction exercise at hand.

In Baker and Gelbach (2020) we present one, intuitive and interpretable, manner for doing such a prediction exercise. Rather than create bespoke indices of peer firms, we use a penalized regression model to predict the return on a given target stock based on the returns on the market index, and the returns of each individual peer firm. There are multiple ways to penalize the inclusion of additional factors in the model—from lasso to ridge, and the elastic net which is a combination of the two—and the penalization parameters can be optimized using cross-validation or leave-one-out prediction error. The advantage of this approach in the context of securities litigation is that it transforms the debate from a relatively subjective one (what is the correct industry and set of peers based on the business attributes of the company) to a comparatively

[27] Cite to more cases with this dispute.
[28] Cite examples

objective one (which combination of firms and weights seems to best predict the return of the stock during the estimation period using out-of-sample prediction methods).

Table 2: Penalized Regression Weights on Peer Firms

Index/Company	Lasso	Ridge	Elastic Net
(Intercept)	0.00	0.00	0.00
S&P Energy Index	0.24	0.24	0.24
Baker Hughes	0.15	0.13	0.14
Beazer Homes	0.00	-0.02	0.00
BJ Services	0.12	0.11	0.11
Centex	0.00	0.01	0.00
Champion Enterprises	0.00	0.00	0.00
Clayton Homes	0.00	0.01	0.00
Comfort Systems	0.03	0.04	0.04
Cooper Cameron	0.07	0.10	0.09
DR Horton	0.00	-0.01	0.00
Emcor Group	0.00	0.00	0.00
Fluor	0.00	0.01	0.00
Foster Wheeler	0.00	0.01	0.00
Granite Construction	0.00	-0.01	0.00
IT Group	0.00	0.01	0.00
Jacobs	0.00	-0.03	0.00
Lennar	0.00	-0.03	0.00
McDermott Intl	0.00	0.02	0.01
MDC Holdings	0.01	0.03	0.02
NVR Inc	0.00	0.01	0.00
Oakwood Homes	0.00	0.01	0.00
Oceaneering	0.04	0.06	0.05
Pulte	0.00	0.03	0.01
Ryland Group	0.00	0.01	0.00
Schlumberger Ltd	0.23	0.18	0.20
Smith Intl	0.07	0.10	0.09
Standard Pacific	0.00	0.00	0.00
Toll Brothers	0.00	-0.01	0.00
URS Corp	0.00	0.02	0.01
Weatherford Intl	0.13	0.11	0.12

This table reports the coefficient values on the peer firms and energy index using different forms of penalized regression. The outcome variable is the log return for Halliburton and the features that enter the regression are the log returns on the index and the peer firms. We use daily data over the class period, omitting the misstatement and disclosure dates, and optimize the tuning parameter using leave-one-out cross validation.

Table 2 shows the model coefficients from different forms of penalization for the event study in the Halliburton case. Similar to the Allen and Coffman models, I use the class period as the estimation period, omitting the days where the plaintiffs allege an affirmative misstatement or a

corrective disclosure was made, and include the returns on the S&P Energy Index and all of the firms with a full trading data over the period that enter either of the two indices used by Coffman in his report. Leave-one-out cross validation determines the penalty value λ that minimizes the root mean squared prediction error over this period. The first column reports the value from using the L_1 norm as a penalty parameter (or penalizing the absolute value of size of the coefficient on each index or firm return); this type of model is typically used for model selection and will "shrink" the estimated coefficients towards zero. As shown in the table, the lasso model drops the returns of twenty of the twenty-nine potential peer firms. The second column reports the estimated coefficients from the ridge model, which uses the L_2 norm, penalizing the square of each coefficient; ridge models tend to shrink each of the estimated coefficients towards each other rather than towards zero, and we see far fewer firms dropping entirely out of the model. The elastic net model finds the optimal combination of each form of penalization; in this case the optimal combination value, called α, is equal to 0.1, so the elastic net and lasso models generate very similar estimated coefficients.

Table 3: Expert and Statistical Learning Predictions

	Expert Predictions		Statistical Learning Approach		
	Allen (D)	Coffman (P)	Lasso	Ridge	Elastic Net
Excess Return	-0.029	-0.037	-0.033	-0.032	-0.033
t-Statistic	-1.290	-2.270	-1.900	-1.835	-1.922
p-value	0.198	0.024	0.058	0.067	0.055

This table presents the excess return calculations using the values from the expert reports, as well as the data-driven predictions from the statistical learning models. I report the excess return, as well as the associated t-Statistic and p-value. The optimal value of α for the elastic net model is 0.1.

The coefficient values from Table 2 can be used to predict the returns for the target firm during the event window. Table 3 reports the predictions and confidence levels from the expert reports, along with the corresponding predictions and confidence from the statistical learning models. One noteworthy feature of regularization-based estimates is that they are very close to each other, generating predicted excess returns of -3.2 to -3.3%, regardless of how you shrink the estimates. This is an advantage of a data-driven approach to conducting an event study: the models will typically pick up some low-dimensional set of factors that predict returns, rather than over-fitting based on the subjective design choices made by experts. From the perspective of a fact finder this is particularly appealing—they can focus their attention on how the question is framed and the answer will be driven by the data, not by expert discretion. In this case, the statistical learning models produce estimates that are close to, but smaller (in magnitude) to the Coffman model. The statistical significance for the estimate is slightly above 5% which is higher than the threshold

set by many, but not all courts.[29]

5.2 Valuation Disputes

Litigation over firm valuation—or, adjudicating disputes about the fundamental value of a publicly-traded firm—is another area of common disagreement among experts that frequently calls for judicial oversight. Initially founded in corporate and securities litigation, financial valuation now plays an increasingly pivotal role in nearly all areas of high-stakes commercial litigation. As a result, much of the judicial burden in commercial litigation has become, in some areas, dominated by valuation disputes that hinge on complex financial economics—spanning a range of practice areas from bankruptcy, to tax, to family law, to fiduciary duties, to garden-variety questions in tort, property and contract law.

5.2.1 The Use of Valuation in Commercial Litigation

In both courtrooms and boardrooms, financial valuation is primarily driven by three competing methodologies: Comparable Companies (CC), Comparable Transactions (CT), and Discounted Cash Flow (DCF) analyses. These approaches are commonly used, often in conjunction, to assess the value of a company or financial asset, especially in complex merger litigation and bankruptcies. Experts occasionally also employ other techniques—such as historical premium analysis, analyst forecasts, or leveraged buyout evaluations—but these are typically supplemental to the three core methods.

Comparable Companies (CC)

The Comparable Companies method does largely what it says: it uses publicly available financial data from actively traded companies to generate a counterfactual valuation for a target firm. This method compares a company's financials to those of similar firms that are publicly listed and traded, offering an advantage in terms of data availability when comparing to other approaches (like Comparable Transactions). Stock prices are often considered a proxy for a company's economic value, providing a robust dataset for generating comparable firm valuations.

The CC process begins by identifying comparable firms in the same industry, of similar size, and with similar capital structures. Analysts then convert the firm's valuation to enterprise value and apply valuation multiples—most commonly the ratio of enterprise value (EV) to earnings before interest, tax, depreciation, and amortization (EBITDA). The key analytical difficulty, from the perspective of a neutral fact-finder, is that CC engenders a substantial amount of discretion, even in applying the selection of the appropriate multiple, with options including the last fiscal year's earnings, the last twelve months, or projections of future earnings.

[29] Cite to Jonah's paper showing that there is no need to use a 95% confidence level at this stage of the case.

An advantage of the CC approach is the volume of available data. Since stock prices for publicly traded companies are observable daily, large sets of comparable companies can be built, unlike the more limited transaction data available in the CT approach. However, the CC method anchors a company's valuation to stock market prices, which may not reflect intrinsic value, particularly in illiquid or volatile markets. Additionally, when CC is used to value private companies, the liquidity premium that comes with publicly traded firms must be considered. Most importantly for our purposes, there is little guidance on how to identify potential peer firms, how many peer firms to consider, and how to calculate the target ratio from the identified peers.

Comparable Transactions (CT)

The Comparable Transactions approach mirrors how real estate appraisers use recent home sales to estimate property value. The idea is to identify analogous assets that were recently sold under similar conditions and use those sale prices to estimate the value of the company in question. For companies, this means looking at sales of firms in similar industries, regions, or with similar capital structures.

The first step in CT is to identify the appropriate comparable firms. Ideally, these are companies of similar size, industry, and capital structure. Analysts then adjust the purchase prices to reflect both the equity and debt structure of the firms, converting the sales price into an "enterprise value" (EV) to standardize comparisons. EV represents the total value of a firm's equity and debt, accounting for differences in financing structures. Analysts will then typically apply valuation multiples to normalize the data after calculating the enterprise value. The most common metric is again the EV/EBITDA multiple, which provides a proxy for cash flow. For less mature companies, other metrics, such as revenue multiples, may be used. However, EBITDA-based multiples are favored in most cases, as they are considered more reliable for mature firms. Normalizing the value of the firm by a measure of profits allows an expert to generate comparisons for a target firm even among comparables of different scale.

The CT approach faces two notable constraints. First, finding sufficient data can be challenging, as genuine arm's-length sales within a particular industry may be rare, forcing analysts to work with a small pool of comparables. Second, transaction prices often include a control premium—the added value paid for acquiring a controlling interest in the company. This control premium can distort the pure cash flow value of the company, and analysts must adjust for it if the valuation's purpose is to exclude such a premium (as in an appraisal action).

Discounted Cash Flow (DCF)

The DCF approach diverges from the comparative nature of CC and CT by focusing on the company's expected future cash flows. Instead of looking for similar firms, the DCF model esti-

mates the intrinsic value of a company by calculating the present value of its future free cash flows, discounted at an appropriate rate to account for risk.

The DCF formula can be expressed as follows:

$$FMV = PV(\text{Cash Flows}) = \sum_{t=1}^{T} \frac{FCF_t}{(1+WACC)^t} + \frac{S_T}{(1+WACC)^T}$$

Here, FCF_t represents projected free cash flows, S_T is the terminal value at the end of the forecast horizon, and $WACC$ is the Weighted Average Cost of Capital, a risk-adjusted discount rate.

DCF models require careful forecasting of cash flows, often based on internal company projections, management estimates, or external financial forecasts. These projections typically cover a period of 5-10 years, after which a terminal value is calculated to represent the company's remaining value. The terminal value can be determined by assuming the firm will grow indefinitely at a constant rate (using the growing perpetuity formula) or by reverting to a valuation multiple based on comparable companies, effectively blending CC and DCF methods.

The DCF approach offers a more fundamental analysis of a company's value but is also more technically demanding and sensitive to the assumptions used for cash flow projections, discount rates, and terminal values. Each component of the DCF model introduces its own complexities. For example, determining the appropriate discount rate requires careful estimation of the company's cost of equity and debt, often derived from asset pricing models such as the Capital Asset Pricing Model (CAPM). Similarly, cash flow projections can be influenced by broader market trends or company-specific factors.

Summary

Each of the three valuation methodologies—CT, CC, and DCF—provides different insights and comes with its own set of challenges. CT and CC offer market-based valuations, but can be constrained by data availability and the need for careful adjustments, such as removing control premiums. DCF, while offering a more granular and intrinsic valuation, requires complex forecasting and careful discretion in applying assumptions. While DCF is often viewed as the "gold standard" in valuation practice for litigation, it is not at all clear that this is a warranted presumption. One academic has argued that DCF "is a speculative exercise disguised in the trappings of mathematical rigor but squarely within the domain of pseudoscience."[30] Moreover, there is substantial evidence that the actual valuation of firms in the market is done through comparing multiples—essentially the Comparable Companies analysis—rather than discounting cash flows.[31]

[30] J.B. Heaton, *Why does Pseudoscience Still Thrive Under Daubert? The Case of Discounted Cash Flow Valuation.*

[31] Itzhak Ben-David and Alex Chinco, *Expected EPS× Trailing P/E.* No. w32942. National Bureau of Economic

In practice, analysts often use a combination of these approaches to create a more comprehensive valuation, as each method compensates for the limitations of the others. By triangulating between CC, CT, and DCF, a more balanced and robust valuation can ostensibly be achieved.

5.2.2 *In re. Mirant Corp.*

Mirant Group was a company that produced and marketed electric power, and their revenue was largely derived from long-term contract sales of power to utilities and from sales of power and capacity in the wholesale energy market. Most of the company's facilities were put in operation while Mirant Group was controlled by its parent-firm TSC. Unfortunately, the company overbuilt its generation facilities, and, following a downturn in the energy market in 2001 and 2002, found itself in financial straits. Its debtors sought relief under Chapter 11 of the bankruptcy code after the company failed to accomplish an out-of-court settlement with their creditors.

The debtors proposed a restructuring plan based on the assumption that unsecured creditors would not receive full satisfaction from the enterprise value of Mirant Group. The equity holders of Mirant were to receive only the *potential* right to receive distributions after paying off the Mirant creditors and other beneficiaries of subordinated debt. The equity committee for the shareholders filed a complaint contending that debtors had undervalued the firm in their plan, directly to the harm of existing shareholders. Given the latent dispute over valuation, the court called a valuation hearing with interested parties.

The valuation hearing lasted for 27 days over 11 weeks, and included numerous expert reports, with the parties placing into evidence a total of 454 exhibits. Following the hearing, the court adjudicated the merits of the competing reports, and ordered the value of Mirant Group "be recalculated to effect changes required by the court." In ordering the re-valuation of the firm, the court registered its not-so-mild exasperation with the practice of valuation in litigation that is worth mentioning. The court felt the need "at this juncture [to] comment briefly on the questionable reliability of these valuation methods." It cited approvingly prior claims about the limitations of the valuation exercise:

> "[Deciding] going concern value is hardly elementary. It involves consideration of what Shelley in 'A Defense of Poetry' called 'the gigantic shadows which futurity casts upon the present.' Those who would prepare future cash flow analyses and discount them to present values are not oracles. The opinion evidence they present ... should be taken as a set of assumptions that are factored into a model and critical analysis then employed to test those assumptions. The evidence in the exercise is hardly clear, is highly judgmental and consists largely of inferences."[32]

Research, 2024.

[32] In re Beker Indus. Corp, 58 B.R. at 739.

According to the court, "[a]t best, the valuation of an enterprise like Mirant Group is an exercise in educated guesswork. At worst it is not much more than crystal ball gazing. There are too many variables, too many moving pieces in the calculation of value of Mirant Group for the court to have great confidence that the result of the process will prove accurate in the future. Moreover, the court is constrained by the need to defer to experts and, in proper circumstances, to Debtors' management." While the court admittedly had "misgivings about the accuracy" of the valuations in the case, "let alone a valuation subject to inherent methodological weaknesses and assumptions unsupported by history", they felt constrained by the law and their comparative disadvantage at the task. "These are the tools available to the court in its task. The necessary result, however, from dealing with evidence best described as 'soft,' is that the court must exercise caution in establishing a range of values for Mirant Group that will necessarily include or exclude equity participation under the Plan and effectively limit its role in the confirmation process."

The court refused to "simply weight the opinions of the various experts to arrive at a melded value" because the range of values was simply too large, from $7.2 billion (Houlihan for the debtors) to $13.6 billion (PJSC for the equity committee). "[F]or the court to simply average these numbers—derived based on varying assumptions and data—would make a mockery of the valuation process and would be terribly unfair to parties whose rights are thereby disposed of." Consequently, the court concluded that the parties had to recalculate the value of Mirant Group based on stipulated changes in data and assumptions. In order for the court to find any equity participation going forward, the range of values for Mirant Group had to $11 billion.

5.2.3 A Better Way to Value Firms in Litigation

In Baker, Gelbach, and Talley (2024)[33] we propose an alternative appraoch to valuation, building off of the Comparable Companeis approach, that uses statistical learning to automate the subjective portion of the valuation process. Rather than have experts disagree about which of a group of peers is truly a "comparable firm" for the target, we use a data-driven procedure to select peers based on the objective ability of the comparable set to predict the target firm's valuation in a clean period. Similar to the event study context, we use the weights from this exercise to create a counterfactual value as of the valuation date. I note here that a valuation approach based on penalized regression is precisely the type of weighted estimate that the *Mirant* court suggested would be appealing:

> In this regard the court is compelled to note that weighting of comparable companies based on their similarity to the subject being valued would seem to have some appeal. The experts whom the court questioned about this rejected the idea, and the court

[33] *Validating Valuation: How Statistical Learning Can Cabin Expert Discretion in Valuation Disputes*, Working Paper.

therefore will not adopt such an approach; it may be that raising the question here will prove useful in future valuations.

Both sides in *Mirant* issued expert reports that used Comparable Companies analysis to value the firm. Blackstone issued a report for the debtors, and selected four comparables: AES, Reliant, NRG and Dynegy. The equity committee urged the court to also consider Calpine as a peer, which the court ultimately declined to do because of its "precarious financial condition" that made its stock price more of an option than "a true reflection of equity value." In addition, multiple witnesses testified that operational differences between Calpine and Mirant Group "were sufficient, when considered together with Calpine's relatively weak financial condition, to disqualify its use as a comparable." The court also noted that AES was substantially larger than Mirant and had more of an international focus, and that the countries where AES operated were generally more stable. The court entertained using NRG, clearly the closest competitor, as the sole comparable, but ultimately held that it did not believe it appropriate to rely solely on one company in formulating a value by the Comparable Method.

This discussion reflects the limitations to gatekeeping a dispute over Comparable Companies analysis. There is very little here to guide a judge or jury in how to consider which expert has selected a more appropriate peer set of firms. In part, this is due to the fact that the objective is not clear. In Baker, Gelbach, Talley (2024) we argue that the similarity between firms is relevant only insofar as it assists in predicting the valuation, or valuation multiple, of the target firm. Lengthy investigations into the similarity of business lines, geographic regions, and financial position are at best a questionable use of scarce court time, and at worst a hopeless diversion from the true underlying question.

Similar to the event study exercise above, we estimate the valuation of Mirant from regularized regressions with its peers, using a data-driven procedure to select the peers and their weights. Instead of using firm returns as the outcome variable, we instead use the firm's market capitalization (the product of equity price and shares outstanding).[34] An predicate decision under this approach is to determine an appropriate estimation window for the model. This would be an appropriate exercise for the court to decide after relevant testimony from the experts, as it involves selecting a period where the valuation is untainted by the allegations in the complaint, but which is close enough in time to the valuation date for the weights to remain accurate. Given data limitations in this case with peer firms also entering bankruptcy themselves, I use the period from May 1, 2001 to December 31, 2001 to get the estimated weights.[35]

The results of this exercise are reported in Table 4. The model intercept captures the expected valuation of Mirant if the peer firms went to zero. The other values reflect the marginal increase

[34] In Baker, Gelbach, Talley (2024) we show how one can use returns and an event study framework to calculate equity market value. However, in this example the length of time between the estimation window and valuation date is long enough that we stick with market capitalization as the outcome variable.

[35] Mirant went into bankruptcy protection in mid-2003.

(in thousands of dollars) for Mirant's equity that arise from a thousand dollar increase in the peer. As mentioned earlier, the lasso regression model is typically used for model selection, as it will tend to "drop" predictors that don't sufficiently explain the outcome. Given the dispute regarding the inclusion (or exclusion) of Calpine as a peer, it is noteworthy that the lasso model *does not* drop the firm, suggesting that it does help in explaining Mirant's valuation. However, AES, perhaps for the reasons explained by the court, is given zero weight and thus is arguably not a useful peer for valuation purposes.

Table 4: Penalized Regression Weights on Peer Firms (Thousands)

Company	Lasso	Ridge	Elastic Net
(Intercept)	\$757,561.82	\$554,096.59	\$345,549.94
AES	\$0.00	\$0.04	\$0.03
Calpine	\$0.38	\$0.19	\$0.25
NRG	\$1.01	\$1.76	\$1.61
Reliant	\$0.12	\$0.19	\$0.14
Dynegy	\$0.38	\$0.40	\$0.43

This table reports the coefficient values on the peer firms using different forms of penalized regression for the valuation of Mirant. The outcome variable is the market capitalization for Mirant and the features that enter the regression are the market capitalization values for the peer firms. I use daily data for Mirant and the peer firms from May 1, 2001 to December 31, 2001, and optimize the tuning parameter using leave-one-out cross validation. The units are in thousands of USD.

Table 5 reports the predicted equity valuation for Mirant using the regularized models. In this example, given that the outcome variable (valuation) is in levels rather than returns, and the long period of time between model estimation and valuation, the valuation range is substantially larger than in the event study example, with a lower bound of \$5.7 billion and an upper bound of \$8.3 billion. After adding back in Mirant's last reported debt levels before bankruptcy of \$3.7 billion, this suggests a total enterprise value range of \$9.4 to \$12.0 billion. This range is between the two values provided by each respective side, and could potentially support a (small) recovery for the plaintiffs.

Table 5: Statistical Learning Predictions for Equity Value on June 27, 2005 (Millions)

	Lasso	Ridge	Elastic Net
2005-06-27	\$5,703.05	\$8,281.79	\$7,420.19

This table presents the predicted market capitalization using the data-driven predictions from the statistical learning models. The units are in millions of USD.

5.3 Employment Discrimination Litigation

[Include example here.]

5.4 Antitrust

[Include example here.]

6 Conclusion