# Courtroom Valuation, Rulified Finance, and the Promise (and Perils) of Machine Learning


Andrew Baker
Berkeley Law

Jonah B. Gelbach
Berkeley Law

Eric Talley
Columbia Law


January 30, 2023

**PRELIMINARY AND INCOMPLETE DISCUSSION DRAFT**
**PLEASE DO NOT CIRCULATE, QUOTE OR CITE WITHOUT PERMISSION**

## I. Introduction

About four decades ago, financial valuation methodologies started to play an increasingly pivotal role in the litigated outcomes of high-stakes commercial disputes. While the judicial embrace of such methodologies was initially limited to isolated topics in corporate and securities law,[1] the movement soon mushroomed. Examples now abound where litigation has come to be influenced, and maybe dominated, by valuation disputes that hinge on financial economics—from bankruptcy to tax to family law to fiduciary duties to garden-variety questions in tort, property and contract law.[2] By some accounts, the incursion of modern finance into commercial law has been nothing short of a pioneering "revolution"—a long-overdue hostile takeover of an "ossified, stagnant field".[3] Every top US law school now offers at least one course dedicated to teaching these techniques to law students.

That said, when the revolution's vanguard breached the ossified walls of courtroom valuation practices, it inadvertently smuggled in a Trojan Horse adversary of its own making: As modern finance infused substantive law, it unleashed at least four undesirable collateral consequences. The first stems from the fact that financial economics tends to be a somewhat mathematical enterprise. As quantitative methodologies found their way into litigation settings, they quickly confronted courts with a demanding progression of technical challenges. Most judges are not formally trained asset pricing

---

[1] Particularly notable developments were in Delaware shareholder appraisal (and quasi-appraisal) actions and federal securities fraud claims. See Weinberger v. UOP, 497 A.2nd 792 (Del. 1985) (appraisal and quasi-appraisal); Basic v. Levinson 485 U.S. 224 (1982).

[2] See generally Casey & Simon-Kerr (2015).

[3] See generally Romano (2005) (describing the transformation in greater detail).

specialists, even in business-intensive courts like Delaware's Court of Chancery; yet they are now typically required to admit, exclude, and sometimes weigh technical financial evidence, or even instruct lay juries possessing even less expertise. In most cases, judges are left to pick up key tenets of valuation practice on the fly, often deducing them from the motivated pedagogy of litigants or their hired experts.

Second, despite its seemingly precise technical façade, there is little doubt that financial valuation is as much art as science: the field is awash with free parameters that both require judgment and afford considerable discretion to the expert analyst. Not only can an expert select from a sizeable menu of *general* approaches to render estimates of fair market value, but within each lurks a veritable fractal tree of subsidiary methodologies and assumptions around implementation.[4] In addition, the professional literature that purportedly guides and substantiates financial experts' choices has *itself* grown heterogeneous, offering a variety of self-styled "authoritative manuals" that contain distinct—and surprisingly inconsistent—formulations for best practices,[5] further amplifying the need for judgment (and the concomitant discretion that such judgment entails).

Third, those deploying financial methodologies in the courtroom are typically the litigants' own compensated experts. While most are respected members of academic and/or professional finance circles, they tend to be repeat-player consultants who perform their work largely shielded from public scrutiny, crafting reports that can cherry-pick from assorted best-practice formulations in ways that—(un)remarkably—favor their clients' economic interests.[6] Less self-serving (and more moderate) calculations may be left on the cutting room floor. Expert reports, moreover, are often filed under seal, remaining secluded from public view until long after trial, perhaps indefinitely. With no real threat of professional scrutiny, little stops dueling experts from embracing techniques they would eschew or even deride in academic settings, resulting in valuations that not only may be inconsistent with academic practices, but also might diverge by multifold factors—or even orders of magnitude.[7]

Finally, in groping to cope with an already-flawed valuation ecosystem, courts have perhaps unwittingly distorted it further. Non-expert judges who adjudicate the claims of dueling, disparate financial experts must provide reasons for their judgments. That reasoning frequently adopts one or the other expert's assumptions for each specific issue, though it might also split the difference between

---

[4] By way of example, consider the "Discounted Cash Flow" (or DCF) approach to valuation, which is but one of several popular valuation methodologies. Implementing a DCF analysis requires the analyst to make assumptions around (*inter alia*) (i) projected free cash flow trajectories, (ii) applicable tax rates, (iii) relative debt and equity values; (iv) systemic risk measures (e.g., "Beta"), (v) de- and re-levering procedures, (vi) market equity risk premia, (vii) risk-free returns, (viii) terminal value measurement, (ix) plowback/retention ratios, and (x) the use (if any) of peer firms to triangulate in on the above factors. Many of these subsidiary items also entail additional subsidiary assumptions. Experts deploying DCF may even exercise important discretion regarding the *timing* of their data downloads. See Akey, Robertson & Simutin (2022).

[5] Compare, e.g., Pratt & Niculita (2010), Rosenbaum & Pearl (2014); Damodaran (2006).

[6] Alternatively, there is the so-called file drawer problem: even if experts write reports unhelpful to their clients, litigation rules allow the clients to withhold such reports in discovery and, a fortiori, at trial.

[7] For example, across all published Delaware appraisal opinions between 1985 and 2018, the petitioner's expert posited a median "fair value" opinion that was 225% larger than that of the respondent's expert, and an mean fair-value opinion that was an eye-popping 1,331% larger. (Data on file with co-author Talley.) Perhaps both sides' experts are using fully defensible methods, but if so that strikes us as an awful lot of heterogeneity in valuation.

experts. Either way, the judge's reasoning—especially once memorialized in a written decision—can implicitly entrench and thus amplify the discretionary choices experts make by adding to precedential tapestry that makes up tomorrow's set of approved best practices. Over time, this tapestry becomes less financial economics and more a "rulified" distillation of judicial folk-wisdoms *about* finance—one whose core tenets may wander afield from contemporary social science. Rulified finance thus assumes a life of its own, with each successive opinion contributing to a self-perpetuating echo chamber of accepted conventions.[8] The end result is something resembling a distinct (and ironically nonacademic) *legal* doctrine, which—while sprouting from the finance scholarship of the 1970s and 1980s—has subsequently followed a markedly different path, impelled substantially by interventions from a coterie of motivated litigators, experts, litigation consultants, and best-practice entrepreneurs.[9]

In recent years, various commentators have offered suggestions about how best remedy to the dysfunctional ecosystem described above. For example, at the urging of several academic commentators, certain courts have increasingly embraced the practice of routinely unsealing expert reports for public scrutiny (at least after trial), under the theory that the prospect of professional embarrassment can provide needed discipline. Others have suggested structural reforms, such as having courts retain an independent expert to advise on valuation issues, or experimenting with expert "hot-tubbing", or committing to final-offer (a.k.a. "baseball") arbitration mechanisms to incentivize experts towards moderation.[10] Still others have instead attempted to find ways for courts to skirt altogether the messy enterprise of valuation sausage making, *e.g.,* by simply adverting to the negotiated merger price itself or pre-existing securities market prices.

Although we believe many of these institutional tweaks warrant consideration,[11] this project follows a different path: We advance the thesis that *now is a particularly opportune moment* for courts not to descend into valuation hibernation, but rather to reconcile courtroom practice with recent academic advances, which are increasingly being fueled by a revolution in data science and machine learning. Our argument starts with a simple observation: Notwithstanding their technical trappings and quasi-theoretical origins, every valuation methodology in use is ultimately an exercise in *prediction*. Whether used to estimate equity value, debt value, enterprise value, asset value, synergy value, or some other target, all valuation methodologies and "best practices" are worth deploying only because they are thought to render a credible prediction of that target given the available data. In fact, each of the three leading market valuation methodologies in use today—comparable transactions analysis, comparable companies analysis, and discounted cash flow (DCF) analysis—aspires to deliver a prediction of firm

---

[8] One such questionable convention is the routine judicial acceptance of CAPM "size" premia. See Carney, Bartlett & Geis, ch 3. (2022).

[9] It is, on reflection, ruefully ironic that an early doctrinal adopter of modern finance—shareholder appraisal actions—did so as a means to jettison the (so-called) Delaware Block Method, itself a rulified practice that came to share few discernible ties to then-prevailing social science. See, e.g., Weinberger v. UOP Inc., 457 A.2nd 701, 712-13 (Del. 1983) (declaring the Delaware Block method "clearly outmoded" and holding that thenceforth, appraisal/quasi-appraisal claims "must include proof of value by any techniques or methods which are generally considered acceptable in the financial community").

[10] Each of these suggestions has potential counter-arguments. See Talley (2017) (providing an overview).

[11] One approach we are less taken with is the idea that judges should eschew valuation challenges altogether. Although accurate judicial valuation is no doubt desirable, even inaccurate good-faith valuation is often more desirable then nothing at all, from both efficiency and distributional perspectives. See, e.g., Choi & Talley (2018); Bubb, Catan & Spamann (2022).

value; the three methodologies differ not in this ultimate goal, but rather in the choice and use of data to accomplish the task.[12]

If our claim is correct that prediction may be viewed as the underlying *sine qua non* of courtroom valuation, then it follows that the science of prediction becomes a critical reference point for our collective attention. We now find ourselves at the threshold of a new and powerful era in the field where algorithmic and artificial forms of intelligence are rapidly sharpening our abilities to make more precise, more reliable and more replicable predictions then were ever presumed possible. In this paper, we argue both that such approaches may outperform current practices, and also that they are perfectly consistent with the doctrines and evidentiary rules that govern litigation.

Our project therefore makes three contributions. First, using simulation evidence based on actual firms' value and other financial variables, we show that current practices allow considerable expert discretion—what we refer to as "expert degrees of freedom."[13] Second, we argue that emerging tools from machine learning (ML) may be able to augment, improve, and even supplant prevailing courtroom valuation practices. Third, we argue not only that judges can admit expert evidence based on such approaches, but also that they can and should begin to expect them from litigants and experts.

Our simulation evidence regarding expert degrees of freedom appears in section III has one overarching finding: conventional valuation practices, which are essentially nearest-neighbor algorithms, bring substantial expert degrees of freedom, even when only apparently minor variations in the parameterization of the nearest-neighbor routine are allowed.

Regarding our second contribution, we aim to demonstrate that ML methods can be used to substantially improve the predictive performance of conventional valuation methodologies. We start by observing that each of DCF, comparable companies analysis, and comparable transactions is at its core basically a primitive ML algorithm.[14] Accordingly, each might in principle be improved by disciplining current practice using only rudimentary improvement. Section IV provides a glimpse at a very preliminary initial attempt at such improvement. At present, our results provide reason for both optimism and awareness that ML methods require adequate data for effective fitting of models (also known as model training). The optimism arises because using a simple error-minimizing approach for choosing the pre-valuation-quarter optimal parameterization is enough for a data-driven approach to clearly outperform a random choice from among the 24 parameterizations we consider (more on the parameterizations below). The need for cautious awareness arises because we also find that our rudimentary approach to choosing an optimal parameterization is slightly outperformed by a strategy

---

[12] DCF analysis principally uses cash flow projections and stock returns data. Comparable companies analysis makes use of public equity market data for "peer" companies. Comparable transactions analysis makes use of data for M&A transactions involving "peer" companies.

[13] See Andrew Baker & Jonah B. Gelbach (2020) for an earlier focus on this concept in the 10b-5 securities litigation context.

[14] Comparable companies and comparable transactions analysis can both be thought of as simplistic implementations of "k-nearest-neighbor" (kNN) prediction algorithms, while DCF typically entails a (quasi-Frankensteinian) assortment of kNN and linear regression techniques.

of just reporting the third lowest valuation from among the 24 parameterizations we consider.[15] This isn't too surprising given the weak theoretical basis for current practice and the fact that the optimal parameterization we use is based on only 7 quarters of model training data. These are circumstances under which one might reasonably expect it to be challenging to identify a tight relationship between firm valuation and observable data. We are actively working on improved methods for these reasons.

More aggressive ML-assisted valuation approaches might skip the nearest-neighbor framework altogether, or combine it with other approaches. Rather than limiting analysis to quasi-cookbook procedures endorsed by professional valuation manuals, an ambitious ML approach could allow the general structure of financial data—rather than individual experts—to guide choices for model structure and inputs, all with the goal of improving overall prediction quality.[16] Consequently, while each of the traditional "best" practices in financial valuation would remain a *candidate* for the best valuation approach, these practices would have to compete with other data analysis methods in a well-structured horse-race designed to ferret out the best performer. Properly set up, machine learning algorithms can favor whichever approach—or combination of approaches—results in the most reliable set of predictions. This is possible because ML techniques can amalgamate all the prevailing valuation approaches (and their distinct inputs) through an ensemble of models,[17] generating predictions that harness and deploy the predictive power of *all* valuation-relevant data simultaneously—a combined predictive power that current courtroom valuation methodologies only minimally harness. Whereas now, fact-finders frequently attempt to assign "weights" to each alternative valuation methodology presented by experts,[18] an ensemble approach functionally delegates much of this weight-apportionment exercise to the optimized prediction algorithm itself.[19]

We acknowledge that these ideas are not without challenges. Machine learning techniques do best, all else equal, when large quantities of comparable data are available, and that is not always true in valuation disputes. Still, the copious amount of publicly available financial data should make it feasible in many disputes. And insofar as current practices are essentially rudimentary nearest-neighbor algorithms, critiques of ML feasibility are also critiques of current practice.

Our third contribution is to argue that ML-based expert evidence is perfectly consistent with current evidence requirements in effect in state and federal courts. Federal Rule of Evidence 702, for example, conditions expert testimony on establishing the expert's qualifications in "knowledge, skill,

---

[15] The third-lowest valuation was one on which we were focused before seeing these results.

[16] For example, comparable company and comparable transactions analysis both aspire to predict a unitless *valuation ratio* for the valuation target (such as the ratio of total enterprise value to EBITDA), and then multiply that ratio by the relevant denominator for the target (here, EBITDA) to deliver a valuation estimate. While ratio normalizations of this sort are widely assumed to have predictive benefits within traditional valuation analysis, an appropriately formulated ML learner can be trained to utilize such ratios if and only if they improve predictive power (and to discard them otherwise).

[17] Formally, ensemble models are a mechanism for taking several individual predictive models (or "weak learners") and amalgamating them into a model that harnesses the power of all of them in combination. See Dietterich (2000).

[18] See, e.g., Dell, Inc. v. Magnetar, 177 A.3d 1, 5, 16–18 (Del. 2017). 3 DFC Glob. Corp. v. Muirfield Value Partners, L.P., 172 A.3d 346, 360–61 (Del. 2017). See also Korsmo & Myers (2018) (describing this approach in greater detail).

[19] See Dietterich (2000).

experience, training, or education."[20] The Rule requires an expert's testimony to be based on "sufficient facts or data,"[21] and also that the testimony is "the product of reliable principles and methods"[22] that are "reliably applied … to the facts of the case."[23] Because Rule 702's 2000 amendment codified the *Daubert* standard,[24] the Rule embraces *Daubert*'s flexible desiderata concerning the testability of the expert's opinion, its academic pedigree, its error rate, and its acceptance within a relevant scientific community. The majority of states,[25] including Delaware, follow *Daubert*'s principles,[26] and Delaware's Supreme Court has applied them in numerous cases.[27] That said, certain states continue to follow the pre-*Daubert Frye* standard that "the thing from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."[28]

Data science and machine learning, we argue, are not merely promising novelties that may, someday, satisfy these admissibility desiderata. Rather, the field is already there, having in the last two decades spawned numerous peer review journals,[29] academic departments,[30] and scholarly conferences worldwide.[31] In fact, we maintain that the data-driven predictive tasks that ML-based models are designed to solve transparently deliver precisely the kind of reliability diagnostics that would satisfy Rule 702 and *Daubert*. And, because such models are designed to optimize direct statistical measures of prediction accuracy, they often will permit side-by-side comparisons of experts' competing models. By their construction, moreover, well-constructed ML-based methods will typically do no worse than the standard approaches, and, given enough data, can be expected to fare better. Consequently, should courts take up our invitation to embrace ML-based valuation approaches, traditional cookie-cutter

---

[20] Fed. R. Evid. 702.

[21] Fed. R. Evid. 702(b).

[22] Fed. R. Evid. 702(c).

[23] Fed. R. Evid. 702(d).

[24] Daubert v. Merrell Dow Pharmaceuticals Inc., 509 U.S. 579 (1993).

[25] See Funk (2022).

[26] Following its amendment in 2001, D.R.E. 702 has the verbatim text as Fed. R. Evid. 702.

[27] See, e.g., Tumlinson v. Advanced Micro Devices, Inc., 81 A.3d 1264, 1268, 1271 (Del. 2013); Sturgis v. Bayside Health Ass'n Chartered, 942 A.2d 579, 584 (Del. 2007); Spencer v. Wal-Mart Stores East, LP, 930 A.2d 881, 888 (Del. 2007); Bowen v. E.I. du Pont de Nemours & Co., Inc., 906 A.2d 787, 795 (Del. 2006); Tolson v. State, 900 A.2d 639, 645 (Del. 2006); Eskin v. Carder, 842 A.2de 1222, 1227 (Del. 2004); Laugelle v. Bell Helicopter Textron, Inc., C.A. No. N10C-12-054 PRW, slip op. at 4, Wallace, J. (Oct. 6, 2014); Brown v. United Water Delaware, Inc., C.A. No. 07C-07-070-JAP, slip op. at 4-5, Parkins, J. (Del. Super. Oct. 7, 2011).

[28] Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923). Examples include California, Illinois, New York, and Pennsylvania.

[29] See SJRScimago Journal & Country Rank, available at https://www.scimagojr.com/journalrank.php?category=1702).

[30] See Best Artificial Intelligence Programs of 2022 (available at https://www.usnews.com/best-graduate-schools/top-science-schools/artificial-intelligence-rankings).

[31] See Phonexia, Top 9 Machine Learning Conferences in 2023 (available at https://www.phonexia.com/blog/top-machine-learning-conferences-2023/).

valuation methodologies may well begin to face increasingly difficult admissibility challenges of their own.[32]

An important caveat to our analysis deserves mention before proceeding. As noted above, the dysfunction afflicting courtroom valuation at present stems in large part from the significant discretion that extant valuation methodologies afford experts—who then can exploit that discretion to conjure up client-serving projections. Although our proposed ML-based methodology neutralizes most of *those* dimensions of discretion, we should be mindful that it—like any new methodology—will likely spawn a new set of conventions, norms of judgment, and dimensions of discretion for future experts to apply. It is therefore fair to question whether our proposed approach—if embraced by courts—would eventually fall prey to a similar set of forces that have undermined the status quo (and arguably even its predecessors).

While we take this potential criticism seriously, we also submit that it does not render the game unworthy of the candle. First, as our simulation evidence shows, our methodology already does better than, or roughly as well as, the status quo in a prediction horse race—despite the fact that we saddled our optimal parameterization approach with relatively little training data. Second, the approach we advocate would use *all* available data, not just subsets susceptible to careful curation by motivated experts. Third, the approach we advocate ultimately could endogenously select the model and data inputs that optimize with respect to a well-defined objective—prediction accuracy—which allows one to compare among two or more competing models.[33] Current valuation approaches, in contrast, fail even to specify their performance criterion. And finally, even if our approach eventually descends into an expert discretion trap, its successor likely would take the form of improved data analytics technique. In the interim, our approach still offers the promise of a desirable package relative to the alternative of continued reliance on today's ossified, rulified finance.

Our analysis proceeds as follows. In Section II we provide an overview of standard valuation approaches that are prevalent in the literature, focusing on comparable transactions, comparable companies, and DCF analyses. We show that the first two can be interpreted as a nearest-neighbor ML algorithm, with as-practiced DCF sharing that feature at least partially. More sophisticated ML approaches could both loosen those conventions and admit a richer variety of models to predict firm value.

In Section III, we spotlight comparable companies analysis and demonstrate the prevalence of expert degrees of freedom. Section IV then undertakes a rudimentary approach to disciplining the

---

[32] Pending amendments to Rule 702 further underscore this point. Under the proposed changes, the proponent of the expert testimony would have to show the Court that "the expert's opinion reflects a reliable application of the principles and methods to the facts of the case," and also establish the testimony's compliance with Rule 702 to a preponderance, in line with Rule 104(a)'s treatment of other preliminary questions. SUMMARY OF THE REPORT OF THE JUDICIAL CONFERENCE COMMITTEE ON RULES OF PRACTICE AND PROCEDURE 23 (2022) (available at https://www.uscourts.gov/sites/default/files/sept_2022_jcus_rules_report_final_for_website.pdf). These changes have been approved by the Standing Committee, with approval by the Judicial Conference and Supreme Court, and abstention or approval by Congress, still necessary for the changes to take effect.

[33] For the settings we envision, where the objective is to predict fair market value (a continuous variable), data science already suggests a relatively modest menu of familiar criteria (such as mean square error and mean absolute error in out-of-sample prediction). Experts would likely attempt to show that their approach dominates the opposing expert's approach with respect to either criterion.

degree of expert discretion by using two years of pre-valuation data, for each target firm and quarter in our simulation, to select an optimal parameterization—i.e., combination of nearest-neighbor parameters—for the target firm and quarter in question. Our results here are somewhat mixed. The glass is half full because even with only 7 quarters of data, our approach to selecting an optimal parameterization is able to outperform a randomly selected parameterization. The other half of the glass is empty, though, because the optimality of this parameterization doesn't stop it from being slightly outperformed by the third order statistic (see below for more discussion). We suspect the mixed nature of these results points to the need for using more data, likely by pooling data across more than just the target firm and quarter.

Section V, to be added, will discuss ML methods designed to overcome these challenges. Section VI will address procedural and evidentiary dimensions of our enterprise, arguing that ML-based valuation models are consistent with legal rules of admissibility; indeed, if they perform as well as they might, these models might put increasing stress on the admissibility and/or weight accorded conventional approaches.[34]

## II. Conventional Practice

As it is currently practiced in business law courtrooms and boardrooms, modern valuation practice is dominated by three alternative methodologies: Comparable companies (CC), comparable transactions (CT), and discounted cash flow (DCF) approaches. In many cases, a valuation expert will attempt to value a financial asset of interest using two or even all three of these approaches. Particularly in the cases of company valuations associated with merger agreements or bankruptcies, experts may use other valuation methodologies as complements. Such additional approaches include an analysis of historical premiums paid, analyst forecasts, and leveraged buyout/recapitalization analysis. When such alternatives are employed, however, they are typically offered only as "reference" valuations meant to complement CC, CT and DCF approaches.

### *A. Comparable Transactions*

The CT approach may be the most intuitively accessible, as it bears resemblance to the approach that real estate appraisers take when using "comps" to estimate the value of one's home. The basic idea is to find examples of analogous assets that have recently been sold in arm's length-transactions, and then use those sales prices to deliver an estimate of what company in question would deliver. With home appraisals, this process usually begins by selecting neighborhoods in a similar geographic area with similar traits such as walkability, school quality, and income, as well as having similar numbers of bedrooms and square footage. The recently sold properties deemed similar to the

[34] This is an instance of the way that technological advances—here, in statistical method and computational practice—might affect adversarial practice in our legal system. For more on that topic, *see* Engstrom & Gelbach (2021).

property in question along these dimensions are an appraisers' comps. The appraiser then will usually attempt to find a normalized measure of value, such as price per square foot, for each comp transaction. The appraiser then uses an aggregation method such as the mean or median of the comps' price-per-square-foot values, which yields a summary measure of the price per square foot to be applied to the property whose valuation is in question.[35]

The CT approach for companies and other financial assets operates similarly. Much like the property appraiser, a valuation analyst using a CT methodology will first find recent sales of companies deemed comparable. If feasible, the analyst will limit attention to transactions in similar industries, geographic locations, or vintages.[36] Like real estate, companies can vary in size, so finding comps with similar sizes is desirable. In addition they can have unique capital structure traits: for example, corporate debt often can transfer over as part of the sale, in which case the purchase price reflects only equity value. The valuation analyst will thus attempt to produce a measure of firm value that controls for both equity and debt factors. To control for capital structure, the analyst often needs to rescale the purchase price to reflect what is known as the enterprise value of each comparable company, adjusting the sales price of each comp to account for the value of any debt not capitalized into the sales price[37]

Once comps' sales prices are converted to enterprise values, analysts then address the size factor, using an analog of the price per square foot measure used by real estate appraisers. Here, however, the standard normalized metric is typically an *earnings multiple:* That is, re-expressing the enterprise value not in raw dollar terms, but rather as a multiple of some specified measure of earnings. A standard metric that operates as a default is earnings before interest, taxation, depreciation and amortization (EBITDA), which is often a relatively stable proxy for cash flows in mature companies. For less mature companies, however, it is not uncommon to see multiples based on other factors, such as EBIT, sales revenues, or in a pinch, even some other measure of market interest such as "clicks" on the company's website. Non-EBITDA multiples are typically disfavored,[38] however, and tend only to be used when the company generates *negative* EBITDA, which renders any multiples-based approach nonsensical.

Even after a multiple has been selected, there may be multiple ways to quantify the denominator of the multiple. For example, one might base a multiple on the last fiscal year's numbers, or the last twelve months, or projections for the next twelve months or fiscal year. In each case, the multiple of the comparable company may change somewhat, and it is not uncommon for analysts to assess CT multiples using several measures, thereby cobbling together a range of valuations based on the aggregated outcomes of such approaches, as well as a variety of

---

[35] Either a point estimate or a range of estimates might be provided, to be clear.

[36] These and other traits are specified in a small set of valuation manuals that have become accepted over time in the profession.

[37] Other adjustments include netting off cash (and cash equivalents), as well as making sometimes-controversial changes to working capital.

[38] Our simulation analysis focused on an EBIT-based measure for data availability reasons.

permutations of the mean, median or inter-quartile range for each measure. In formulating the comps, the multiple formulations, and the ranges, the analyst typically retains significant discretion—an issue to which we return in our analysis below.

Two potential constraints on the CT approach often affect its ability to deliver valuation projections. The first is a lack of data. Because bona fide arms-length sales of companies within a given industry are generally rare, the set of "comparable transactions" might itself be relatively sparse, which may force the analyst to make a projection from a very small group (perhaps even as small as one). One potential fix to this problem is to lengthen the time horizon or broaden the criteria by which comparatives are drawn (e.g., by expanding the industries considered). The second potential limitation on the CT approach is that it is predicated against sales of comparable firms through negotiated transactions. Such sales often come with a premium baked into the sales price, reflecting the value of control. This baked-in control premium may sometimes be inappropriate if (for example) one is interested in gauging only the cash flow valuation of the company. In such settings, CT requires a (sometimes hazardous) attempt to shave off control premia from precedent sales.

### *B. Comparable Companies*

The CC approach is a close cousin to the CT approach, differing only in the source of the data used to assess comparable firms. While CT uses sales price data from acquisitions of comparable firms, CC uses the full spectrum of data from large and thick securities markets. In thick markets, stock prices are thought to be a good proxy for the economic value of a fractional share of the company, at least on average and as viewed by the marginal investor. Thus, rather than using the sales price to predict value, the total market capitalization of comparable firms can be based on public trading data. Beyond that, however much of the CC valuation process is identical to that in CT, including the conversion from an enterprise value multiple to enterprise value, the specification of the multiple itself, the use of judgment about how to measure such multiples, such as last twelve months, next twelve months, etc., and a summary measure (mean or median) for aggregating comp multiples. In other words, beyond the different source of valuation metrics for the comparable firms, nearly every other part of a CC analysis tracks the CT analysis almost directly.

CC methodologies have one obvious advantage over CT approaches: data. It can be difficult to find precedent M&A transactions to use in developing CT comps, but CC is facilitated because thousands of public companies trade continuously and have observable prices each day. Consequently, CC allows one to build sizeable groups of comparable firms. On the other hand, CC also tethers value of companies to the trading value of their stocks, which in turn tends to reflect the value the market ascribes to a publicly held valuation target. Thus, the CC approach might neglect the value of the control premium. Another potential limitation of the CC approach is that it depends critically on the on-average value efficiency of trading markets. This may often be appropriate, but CC fits less well when the target firm is traded in a thin, volatile, poorly-developed market where market valuations and fundamental valuations can diverge.

*C. Discounted Cash Flow*

The third major form of valuation is discounted cash flow (DCF) analysis. Compared to CC and CT approaches, the DCF has significantly more moving parts, and is generally viewed as more technically demanding. Rather than looking for comparable firms (at least directly), the DCF approach conceives of the value of a financial asset as the equivalent to the present discounted value of the free cash flows the asset is projected to produce. Borrowing from the well-known formula in finance for present values, the DCF approach can be captured as follows:

$$\text{FMV} = \text{PV(Cash Flows)} = \frac{FCF_1}{(1+WACC)} + \frac{FCF_2}{(1+WACC)^2} + \cdots + \frac{FCF_T}{(1+WACC)^T} + \frac{S_T}{(1+WACC)^T},$$

where projected future time periods are divided up into discrete units through some "terminal" projection period $(t = 1,2,3 \ldots T)$; $FCF_t$ denotes Expected Future "Free Cash Flows" in each of the periods projected (typically 3-10 years into the future[39]); $S_t$ denotes a terminal (or "salvage") valuation of the asset as of the terminal projection year; and $WACC$ represents a risk-adjusted discount rate known as the "Weighted Average Cost of Capital." If all of these ingredients are known (or can be reliably estimated under different scenarios), then a cash-flow prediction (or predicted range) of asset valuation is possible.

Like an onion, each of the ingredients of the DCF approach has its own layers of complexity (and concomitant expert discretion). Free cash flows are sometimes generated from analyst forecasts, or through management forecasts done in the ordinary course, or through investment banker forecasts for the purposes of a (disputed) transaction, or by some other source. They are typically, though not always, unlevered (and thus do not carve off interest payable to capital creditors, so as to summarize the entire pool of earnings available to satisfy both debt holders and equity holders). And, in some cases cash flow projections of comparable firms (if available) can be used to form composite projections that more closely track the industry at large.

The $WACC$ discount rate is typically a blend of expected return estimates for debt and equity (adjusted for leverage ratios and tax deductions), with equity return estimates the product of an underlying asset pricing model (such as the still-dominant market "beta" from the capital asset pricing model). In many cases, peer company betas are also blended in with company-specific data to create more of a composite measure (usually after an elaborate process adjusting peers' betas for differences in peer leverage ratios)

Finally, the terminal value measure ($S_T$) represents something of a capitulation to our inability to make projections indefinitely into the future. Because company projections are typically no longer than 10 years (and are more frequently 5-7 years), an analyst using DCF must make an assumption of what the asset will be worth in its terminal period (when no more forward looking projections are available). The assessment of terminal value is perhaps one of the most freighted with expert degrees of freedom since there is little to tether it to firm-level data. One approach for terminal values is to

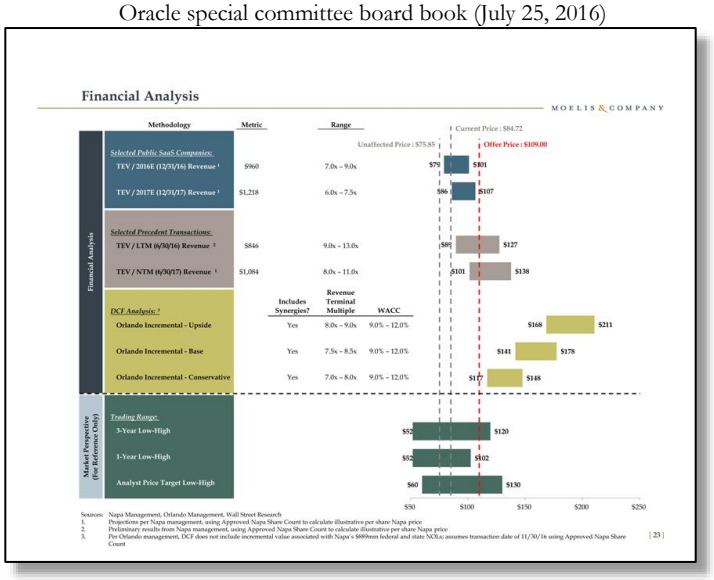
[39] Free cash flows are typically computed by starting with projected EBITDA for the relevant period, and then deducting (a) projected taxes, (b) expected capital expenditures, and (c) expected changes in working capital.

extrapolate the final period's free cash flow indefinitely into the future as "growing perpetuity" at some posited growth rate, backing out a present valuation (as of period T) from a well-known formula for the present value of growing perpetuities.[40] Another frequently used approach, however, is simply to revert (once again) to peer company multiples using a recycled CC or CT approach applied as of the terminal period.

*D. Presentation*

Once a financial analyst has performed the valuation analyses above (possibly adding in a few others as a "check" for good measure), the results are conventionally encapsulated in a report/presentation delivered to the relevant decision maker (such as a board of directors or committee thereof). The report will often summarize each of the approaches used in a moderate level of detail, and then present the totality of results on a summary page sometimes known as a "football field" that posits ranges of valuations under a variety of different scenarios. Figure 1 reproduces the football field summary of the proposed acquisition of NetSuite that was presented to the special committee of the Oracle board of directors in 2016.

Figure 1: NetSuite valuation "Football Field" from
Oracle special committee board book (July 25, 2016)




In the proposed Oracle/NetSuite deal, NetSuite shareholders were to be paid $109/share (represented by the red dotted line in the chart). The committee's financial advisor (Moelis & Co.) performed each of a Comparable Companies analysis ("Selected Public SaaS Companies"), a Comparable Transactions analysis ("Select Precedent Transactions"), and a DCF analysis. In

[40] For a posited perpetuity growth rate $g$, the formula is given by: $S_T = \frac{FCF_{T+1}}{(WACC-g)} = \frac{FCF_T \times (1+g)}{(WACC-g)}$.

addition (and placed below the dotted black line for comparison purposes) were a few other informal analyses based on analyst projections and historical trading ranges. Note that for each of the CC, CT, and DCF analyses, the report makes use of several "scenarios" related to data inputs that jostle around the valuation ranges. It should be noted that the report excerpt above was presented to the board of a company that was dominated by a controller (Larry Ellison) who had a substantial financial interest in NetSuite, and thus stood to make private financial benefits if the sale was consummated. Consequently, it should not be surprising that the data inputs into the valuation report excerpted above have faced heavy scrutiny by plaintiff attorneys in ongoing litigation.[41]

### *D. Valuation practice is a rudimentary form machine learning*

Although there are clearly several moving parts to modern valuation practice as applied in judicial and regulatory settings, several attributes of the description above stand out. First, experts have substantial discretion in how they go about delivering a valuation—discretion that ranges from what approach they adopt to how the measure the ingredients of that approach to subsidiary decisions about how to model "inputs to the inputs" of their valuation metric. Second, peer-group comparisons are pervasive – indeed such comparisons are the very backbone of CC and CT analysis, and even in DCF analysis peer group comparisons sneak in in myriad ways (free cash flow projections, terminal values, beta estimates, etc.).

But more noteworthy than any of these factors in isolation is how closely valuation practice resembles the nearest-neighbor machine learning algorithm, both generally and in specific application. From a general perspective, conventional valuation measures reflect the same philosophy as machine learning generally, and the nearest-neighbor approach in particular, along three dimensions: (1) Their core aspirations are to deliver core predictions (rather than to test causal theories); (2) They use a significant amount of unstructured data to deliver those predictions; and (3) They habitually make use of observable/claimed similarities of instances (via "comps") to buttress, support, and even drive their predictive enterprise.

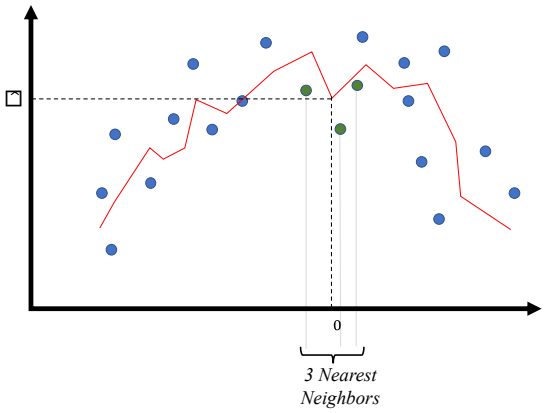
The kNN learner was first developed over a half century ago.[42] It is a non-parametric method for learning often used to classify and/or make predictions about an outcome label of interest by reference to the attributes of a given number, $k$, of the target's closest neighbors in some feature space, using a specific proximity measure. The method can be used for both classification problems, where the object is to predict which of a discrete set of categories the item of interest belongs, or regression problems, where the objective is to predict the value of a potentially continuously measured outcome variable's value. Valuation practices are an instance of the latter, with the outcome variable being enterprise value.

---

[41] See Jeff Montgomery, "Del. Chancery Winds Up Trial On $9.3B Oracle-NetSuite Deal", Law360 (Aug. 16, 2022).

[42] See, e.g., Evelyn Fox & Joseph Hodges, (1951). Discriminatory Analysis. Nonparametric Discrimination: Consistency Properties Report). USAF School of Aviation Medicine, Randolph Field, Texas; Thomas M. Cover & Peter E. Hart (1967). "Nearest neighbor pattern classification" (PDF). IEEE Transactions on Information Theory. 13 (1): 21–27

Figure 2: kNN Heuristic Representation (with $k$=3)




By way of illustration, Figure 2 shows an example set of training data in two dimensional space. The figure depicts variables $x$ and $y$, whose labels are observed for all members of the training data. Our objective is to use these data to predict an outcome label $\hat{y}$ for a new observation that has observed $x$-value $\boldsymbol{x_0}$. The Figure utilizes a 3-NN approach, where the three closest points to $\boldsymbol{x_0}$ in $x$-space (measuring closeness here by simple Euclidian distance) are treated as comps. We then predict $y$ for the target of interest using the mean of the three nearest neighbors in $x$-space, which serves as the prediction $\hat{y}$. Using the mean of a sample of $y$ values among comps having close values of $x$ can be thought of as an instance of $k$-NN mean regression.

The kNN algorithm can be generalized by adding large numbers of $x$-variables, as well as vector-valued outcomes. Its key ingredients are (a) an appropriate outcome variable $y$ (often called "label" in machine learning communities) to predict (b) a specified set of characteristics other than the outcome label by which to assess similarity/proximity (the $x$s); (c) a distance metric for measuring proximity (such as Euclidean distance); (d) a specified inclusion/weighting system for selecting comparators into the prediction set (such as picking the closest 3 neighbors); and (e) a metric for generating predictions from the included training variables (such as voting; median values; mean values; inter-quartile range, etc.).

With respect to Comparable Companies and Comparable Transactions analysis, this discussion reveals that the two valuation methodologies are not simply "like" the kNN algorithm—they *are* the kNN algorithm, if implemented in a somewhat casual form. Both approaches involve (a) taking earnings multiples of either trading values (CC) or acquisition values (CT) as the relevant outcome variable; (b) a specified set of data for identifying comparable companies/transactions; (c) a comparability assessment (although in practice this is rarely specified precisely); (d) rules of thumb for determining the number of comps to use (also involving expert discretion); and (e) a means for predicting the earnings multiple of the company in question (often the mean or median of the comparable firms, again at the expert's discretion). Indeed, while neither the CC method nor the CT

method has to our knowledge previously been directly identified with kNN algorithm, their resemblance is evident.

Although DCF valuation is not quite as clearly the splitting image of the kNN algorithm, it, too, shares some features with kNN. For example, it is common to use comparables to generate estimates of terminal value, earnings projections, and asset betas – all of which are a core ingredients of the DCF model. Thus, the kNN approach often enters DCF analysis at several junctures in material ways.

But if standard valuation methodologies are functionally kNN learners, it is important to understand the approach's potential limitations. The kNN approach can be attractive for a variety of reasons. First, kNN is a form of instance-based learning, which means that unlike other supervised learning approaches, it does not require a training stage for use. This makes kNN simpler and faster to use than other algorithms that require training. Finally, when the data set grows large, kNN regressions are known to be Bayes optimal.[43]

On the other hand, kNN classification has several limitations that can hamper its performance. First, as the amount of data grows, the cost of calculating distances explodes (this is an example of the so-called curse of dimensionality that many non-parametric methods confront). kNN use is also complicated by the presence of high-dimensional data (many *x*s), although this is not a problem with the numbers of variables typically used in valuation. Third, kNN can be highly sensitive to scale and how distance is calculated, so good applications of kNN must involve normalization before applying the algorithm—which adds a layer of expert discretion, because multiple scalers exist. Finally, kNN can be sensitive to noisy data, missing values, and outliers.

The fact that valuation practice has not previously been identified with machine learning might help explain why it hasn't already become a *very good* type of machine learning. It appears by our lights that courtroom valuation techniques have been locked in a form of cryogenic suspension since the 1980s, lying dormant in the historical sediment somewhere between a long-forgotten Led Zeppelin belt buckle and a Devo flower pot hat. Nevertheless, the similarities between valuation theory and machine learning invite us to explore how current valuation practice can be recast in the parlance of machine learning in order to systematically asses how expert degrees of freedom can, if left unchecked, importantly affect valuation. We think a more systematic embrace of tools, norms, and data transparency practices in machine learning can help address the problems of rulified finance.

## III. Illustrating the Role of Expert Degrees of Freedom Using Comparable Companies

Ultimately we aim to show that machine learning techniques can be useful across accepted valuation approaches, and can be useful for aggregating various approaches. But to fix ideas, this section illustrates our basic points using the comparable companies (CC) approach.

---

[43] See Cover & Hart (1967).

The first part of our analysis in this section uses a randomly chosen target firm and target quarter. Together, we refer to these as the "target valuation case." For any target valuation case, the goal of CC analysis can be viewed as using information on the value of peer firms to predict the target firm's value in the target quarter. Results for a randomly chosen target valuation case show that seemingly minor variations in the parameters used to choose peer companies can generate strikingly substantial variation in the target valuation. We show in detail how this variation is connected to the fact that different parameterizations select different comps from among the available peers. These findings aptly illustrate the substantial extent of expert degrees of freedom in CC valuation.

The second part of our analysis extends our analysis from the single target valuation case to a sample of 10,000 randomly chosen target valuation cases. Statistics based on that sample demonstrate that the core finding from the single case extend generally. Expert degrees of freedom pervade our simulation sample.

### *A. Data*

To conduct our simulations, we use data from two familiar financial databases. The first is S&P's Compustat, which provides information on quarter financial information such as revenue, profit, and assets. The second, used in later sections of the paper [still to be added] is the Center for Research in Security Prices (CRSP), which provides information on daily stock returns, which we will use to measure market risk, aka beta, and excess returns. We use quarterly data from Q1 2000 to Q4 2020, merged to CRSP through the linking file provided by the Wharton Research Data Services (WRDS).

For a firm to be eligible to be drawn as a simulated target firm, we required that it have non-missing values of covariates in the union of the Rosenbaum & Pearl and Pratt & Niculita sets, as well as additional variables necessary to identify the quarter, the firm, firm enterprise value, firm EBITDA, and firm market capitalization. For analysis of quarter $t$, we required that the firm have non-missing values of all these variables in both quarter $t$ and quarter $t$-1. In addition, to focus attention on firm-quarters for which it is possible to predict a target firm's multiple with $k$=9, we required that there be at least 9 available peers with available data on the variables just described. After applying these data restrictions, we have 83,337 distinct firm-quarter combinations available to serve as simulated target valuation firm-quarters. Our data contain 3,576 distinct firms, and 80 distinct quarters.[44]

### *B. An example using a specific firm*

We begin this section with an illustrative example. We first picked a target valuation firm and target valuation quarter at random from our data. Table 1 reports some basic financial facts about

[44] The number and pattern of quarterly observations varies across firms for familiar reasons—firms are newly created, or go public, or are acquired, or go out of business, etc.

the target valuation firm and quarter. The randomly drawn target valuation company is Universal Technical Institute (ticker UTI),[45] and the randomly drawn target valuation quarter is the quarter ending September 30, 2011. The company's SIC code of 8200 indicates that it is involved in the "SERVICES-EDUCATIONAL SERVICES" industry. Its closing share price at the end of the quarter was $13.59, and it had nearly 25 million shares outstanding, so that its total market capitalization was $336 million.

Table 1: Basic Information About Target Valuation Firm-Quarter

| Variable | Value |
|---|---|
| Randomly drawn target firm name | Universal Technical Institute |
| Randomly drawn quarter | Ending 9-30-2011 |
| SIC code | 8200 |
| Closing price | $13.59 |
| Shares outstanding | 24.69 million |
| Market cap | $336 million |

Table 2 lists the covariates that are included in either or both of two well-known practitioner guides for valuation experts: Rosenbaum & Pearl and Pratt & Niculita. There are 21 covariates in all, of which 10 belong to both sets. Of the remaining 11, six appear in the Pratt & Niculita set but not the Rosenbaum & Pearl set, and the reverse is true for the remaining 5 covariates. For reference, the table also reports UTI's values of the 21 covariates for the target quarter.

[45] The company's ticker is UTI; further information may be found from Yahoo! Finance at https://finance.yahoo.com/quote/UTI/.

Table 2: Values of Covariates Suggested by Rosenbaum & Pearl and by Pratt & Niculita

| Variable | RP | PN | Value |
|---|---|---|---|
| Age | No | Yes | 40 |
| Assets | No | Yes | 266 |
| capitalization_ratio | No | Yes | 0.00 |
| Debt | No | Yes | 0.00 |
| debt_equity | No | Yes | 336 |
| ebit_growth | **Yes** | **Yes** | -0.18 |
| ebit_margin | Yes | No | 0.09 |
| ebitda_growth | **Yes** | **Yes** | -0 |
| ebitda_margin | Yes | No | 0 |
| eps_growth | **Yes** | **Yes** | -0.21 |
| Ev | **Yes** | **Yes** | 232 |
| gross_profit_margin | Yes | No | 0.56 |
| implied_dividend_yield | **Yes** | **Yes** | 0.00 |
| leverage_ratio | No | Yes | 0.00 |
| mkt_cap | **Yes** | **Yes** | 336 |
| ni_growth | **Yes** | **Yes** | -0.22 |
| ni_margin | Yes | No | 0.05 |
| Roa | **Yes** | **Yes** | 0.02 |
| Roe | **Yes** | **Yes** | 0.01 |
| Roic | **Yes** | **Yes** | 0.02 |
| Sales | Yes | No | 111 |

[NEED TO ADD FURTHER DISCUSSION OF COVARIATES]

There are 17 firms that both shared UTI's two-digit SIC code value of 82 and, in the target quarter, passed the other data selection criteria we required of peer firms.[46] From these 17 firms, we selected the $k$ nearest matches to our target firm in the target quarter, where $k$ is allowed to be either 7, 8, or 9. To determine the nearest matches for any value of $k$, we alternately used one of two distance metrics.

- The scaled Euclidean metric.[47]

[46] See discussion below.

[47] The scaled Euclidean metric is a modified version of the simple Euclidean metric, which is just the square-root of the sum of coordinate-wise differences between two vectors. For example, if $x_1=(1,2)$ and $x_2=(3,7)$, then the difference between the two vectors is $d=x_2-x_1=(2,5)$, and the sum of the squared values of the two coordinates is $2^2+5^2=29$. The simple Euclidean distance between $x_1$ and $x_2$ is then the square-root of this sum, i.e., the square-root of 29. A problem with the simple Euclidean distance metric is that it is very sensitive to scale. For example, suppose the first coordinate involves a variable measured in dollars. By changing units to cents for the first coordinate, we would change its value in $x_1$ and $x_2$ from 1 and 3 to 100 and 300, which will lead to a drastic change in the numeric Euclidean distance—now the square-root of 40,025. The scaled Euclidean metric avoids this problem. It does so by first calculating the standard deviation of each coordinate among points in the data of interest and then dividing each coordinate by its standard deviation. This solves the arbitrary-units problem with the simple Euclidean metric because if we change units for a

- The Mahalanobis metric.[48]

Given a metric, we calculate a measure of the distance between UTI's covariates and those of each of its 17 peer firms. We deem the $k$ peer firms with the smallest distance for each metric and covariate set to be the comps for the given parameterization—i.e., combination of the distance metric, covariate set, and $k$. Finally, we calculate the summary measure—mean or median—of the comparable companies' enterprise value multiple in the target quarter. This value is the comparable companies-based predicted value of the target firm's enterprise value multiple for the given parameterization.

Table 3 reports the target firm prediction for each of the 24 combinations of parameters: distance metric (either scaled Euclidean or Mahalanobis); covariate set (either Rosenbaum & Pearl or Pratt & Niculita); number of comparable companies ($k$ = 7, 8, or 9); and summary measure used to predict target valuation case multiple (either mean or median of peer firms' enterprise value multiples). We report the predicted values from the 24 parameterizations in ascending order.

The mean of the predictions in Table 3 is 22.5, relatively close to the actual enterprise value multiple, which was 24.2; thus the prediction error was -1.7 (=22.5-24.2), which is about 7% of the actual enterprise value multiple. The standard deviation across the 24 predictions is 2.0, roughly equal to the prediction error. These facts—relatively low average prediction error and similarly low standard deviation of predicted multiples—might seem to be cause for confidence.

However, these summary measures of the distribution of predicted multiple values mask a substantial degree of variation. The min-to-max range is 6.6 units, which is more than 25% of the multiple's actual value. This range isn't the result of extreme outliers, either: the spread between the 3rd and 22nd order statistics is 5.1, still above 20% of the true value.

---

coordinate from dollars to cents, the standard deviation rises proportionately; the ratio of coordinate values to their standard deviation is thus unaffected by changes in units.

[48] The Mahalanobis distance metric is similar to the scaled Euclidean, except that it accounts for covariances in the data between the variables represented by each coordinate. For example, if firm age and firm assets were the only covariates, and older firms had greater assets, that association would be unaccounted for by the scaled Euclidean metric's coordinate-wise approach. Using the Mahalanobis metric to measure the difference, d, between two vectors can be viewed as using the following steps: (i) calculate V, the full covariance matrix for all variables used in the distance calculation (i.e., all covariates in the particular covariate set; (ii) calculate $d^*=V^{-½}d$, where $V^{-½}$ is the matrix square-root of V, and (iii) calculate the simple Euclidean distance between d* and the zero vector that has the same dimension as d*. Step (ii) standardizes the vector d, eliminating both coordinate-specific variance and cross-coordinate covariance scale differences.

Table 3: Predicted Target Firm Enterprise Value Multiple Based on Comparable Companies

| Metric | Covariate Set | Summary Measure | *k* | Prediction |
|---|---|---|---|---|
| Scaled Euclidean | Rosenbaum & Pearl | Mean | 7 | 19.0 |
| Scaled Euclidean | Rosenbaum & Pearl | Median | 7 | 19.5 |
| Scaled Euclidean | Rosenbaum & Pearl | Mean | 9 | 20.4 |
| Scaled Euclidean | Rosenbaum & Pearl | Mean | 8 | 20.4 |
| Scaled Euclidean | Rosenbaum & Pearl | Median | 9 | 20.5 |
| Mahalanobis | Pratt & Niculita | Median | 9 | 20.5 |
| Mahalanobis | Pratt & Niculita | Median | 7 | 20.5 |
| Mahalanobis | Pratt & Niculita | Mean | 7 | 21.7 |
| Mahalanobis | Rosenbaum & Pearl | Mean | 7 | 21.7 |
| Scaled Euclidean | Pratt & Niculita | Mean | 7 | 21.8 |
| Scaled Euclidean | Rosenbaum & Pearl | Median | 8 | 22.0 |
| Mahalanobis | Rosenbaum & Pearl | Mean | 8 | 22.4 |
| Mahalanobis | Pratt & Niculita | Median | 8 | 22.6 |
| Mahalanobis | Pratt & Niculita | Mean | 9 | 23.2 |
| Scaled Euclidean | Pratt & Niculita | Mean | 8 | 23.2 |
| Scaled Euclidean | Pratt & Niculita | Mean | 9 | 23.4 |
| Mahalanobis | Rosenbaum & Pearl | Mean | 9 | 23.5 |
| Mahalanobis | Pratt & Niculita | Mean | 8 | 23.6 |
| Mahalanobis | Rosenbaum & Pearl | Median | 7 | 24.6 |
| Scaled Euclidean | Pratt & Niculita | Median | 7 | 24.6 |
| Mahalanobis | Rosenbaum & Pearl | Median | 8 | 25.1 |
| Mahalanobis | Rosenbaum & Pearl | Median | 9 | 25.5 |
| Scaled Euclidean | Pratt & Niculita | Median | 9 | 25.5 |
| Scaled Euclidean | Pratt & Niculita | Median | 8 | 25.6 |

This is a substantial amount of variation in predicted values given the innocuous-seeming variation in the parameterizations. After all, authoritative sources recommend each covariate set; each distance metric has been used extensively in financial economics research [VERIFY]; the values of *k* are not exotic; and both the mean and median summary measures are often used in financial research. In the remainder of this subsection, we demonstrate that for our randomly drawn target valuation case, the parameterizations' variation is not innocuous.

Begin with a look at Table 4, which lists the 17 peer firms for our randomly drawn target valuation example and reports the scaled Euclidean distance between the target firm and each of the 17 peers, computed using the Rosenbaum & Pearl covariate set. Peer firm names and their distances from the target firm are presented in Table 3 in ascending order of distance. The peer firm with the lowest distance was Learning Tree Intl Inc (scaled Euclidean distance = 1.23), and the peer with the greatest distance was Apollo Education Group Inc (scaled Euclidean distance = 7.60). The peers that would be included as comparable companies with *k*=7 are presented in bold, the additional peer

that would be included with $k$=8 is presented in italics, and the additional peer that would be included with $k$=9 is underlined.

Table 4: Peer Firm Distances From Target Firm in Target Quarter,
Using Rosenbaum & Pearl Covariates

| | Distance From Target Firm | |
|---|---|---|
| Peer firm | Scaled Euclidean Distance | Rank |
| **Learning Tree Intl Inc** | **1.23** | **1** |
| **Capella Education Co** | **2.29** | **2** |
| **Perdoceo Education Corp** | **2.29** | **3** |
| **National Amern Univ Hldg Inc** | **2.99** | **4** |
| **Strategic Education Inc** | **3.08** | **5** |
| **Corinthian Colleges Inc** | **3.20** | **6** |
| **Grand Canyon Education Inc** | **3.28** | **7** |
| *GP Strategies Corp* | *3.62* | *8* |
| Graham Holdings Co | 3.71 | 9 |
| Education Management Corp | 4.07 | 10 |
| Lincoln Educational Services | 4.98 | 11 |
| American Public Education | 5.08 | 12 |
| Adtalem Global Education Inc | 5.46 | 13 |
| Zovio Inc | 5.92 | 14 |
| ITT Educational Services Inc | 5.92 | 15 |
| Archipelago Learning Inc | 6.26 | 16 |
| Apollo Education Group Inc | 7.60 | 17 |

The first column of Table 5 reports each peer firm's distance rank using the scaled Euclidean metric and the Rosenbaum & Pearl covariate set, again sorted in ascending order. The remaining columns report the corresponding rank of each of the 17 peer firms' distance from the target firm when we replace the Rosenbaum & Pearl covariate set with the Pratt & Niculita one, and when we replace the scaled Euclidean distance metric with the Mahalanobis distance metric.[49] The table shows both that there are a number of peers with consistently low distances across the four distance parameterizations (e.g., Capella Education Co and Strategic Education Inc), and that there are a number with consistently high distances (e.g., Education Management Corp and ITT Educational Services Inc).

[49] We don't report the distance values themselves because the scaled Euclidean metric is sensitive to the joint distribution of the covariates used to calculate distance, even though this metric normalizes by the variance of each covariate, it does not adjust for covariance. Consequently, the scaled Euclidean distance will have a different scale for different covariate sets. The Mahalanobis distance metric adjusts for both variance and covariance values, so the Mahalanobis distance values share a common scale for the two covariate sets. However, the Mahalanobis distance values will not be on the same scale as either scaled Euclidean distance metric if the covariate sets have differential covariance patterns.

Table 5: Peer Firm Ranks in Distance From Target Firm in Target Quarter, By Distance Metric & Covariate Set

| | Metric | | | |
|---|---|---|---|---|
| | Scaled Euclidean | | Mahalanobis | |
| Peer firm | RP | PN | RP | PN |
| Learning Tree Intl Inc | 1 | 1 | 13 | 9 |
| Capella Education Co | 2 | 2 | 1 | 1 |
| Perdoceo Education Corp | 3 | 3 | 7 | 17 |
| National Amern Univ Hldg Inc | 4 | 5 | 5 | 13 |
| Strategic Education Inc | 5 | 6 | 2 | 3 |
| Corinthian Colleges Inc | 6 | 11 | 15 | 7 |
| Grand Canyon Education Inc | 7 | 4 | 8 | 2 |
| GP Strategies Corp | 8 | 7 | 3 | 6 |
| Graham Holdings Co | 9 | 14 | 6 | 4 |
| Education Management Corp | 10 | 17 | 10 | 10 |
| Lincoln Educational Services | 11 | 12 | 14 | 11 |
| American Public Education | 12 | 8 | 16 | 15 |
| Adtalem Global Education Inc | 13 | 13 | 9 | 12 |
| Zovio Inc | 14 | 9 | 4 | 16 |
| ITT Educational Services Inc | 15 | 10 | 17 | 14 |
| Archipelago Learning Inc | 16 | 16 | 11 | 8 |
| Apollo Education Group Inc | 17 | 15 | 12 | 5 |

Each column of Table 6 shows, in ascending value of distance, the peer firms that have the 9 lowest distance values for each distance metric and covariate set. The peers that would be included as comparable companies with $k$=7 are presented in bold, the additional peer that would be included with $k$=8 is presented in italics, and the additional peer that would be included with $k$=9 is underlined. Table 6 shows that the set of comparable firms can be made to vary substantially, simply by varying the parameters used to choose them, even when we consider only seemingly innocuous combinations of parameters.

Table 6: *k* Nearest Neighbor Sets for Target Firm in Target Quarter,
By Distance Metric & Covariate Set

| Metric | | | |
|---|---|---|---|
| Scaled Euclidean | | Mahalanobis | |
| RP | PN | RP | PN |
| **Learning Tree Intl Inc** | **Learning Tree Intl Inc** | **Capella Education Co** | **Capella Education Co** |
| **Capella Education Co** | **Capella Education Co** | **Strategic Education Inc** | **Grand Canyon Education Inc** |
| **Perdoceo Education Corp** | **Perdoceo Education Corp** | **GP Strategies Corp** | **Strategic Education Inc** |
| **National Amern Univ Hldg Inc** | **Grand Canyon Education Inc** | **Zovio Inc** | **Graham Holdings Co** |
| **Strategic Education Inc** | **National Amern Univ Hldg Inc** | **National Amern Univ Hldg Inc** | **Apollo Education Group Inc** |
| **Corinthian Colleges Inc** | **Strategic Education Inc** | **Graham Holdings Co** | **GP Strategies Corp** |
| **Grand Canyon Education Inc** | **GP Strategies Corp** | **Perdoceo Education Corp** | **Corinthian Colleges Inc** |
| *GP Strategies Corp* | *American Public Education* | *Grand Canyon Education Inc* | *Archipelago Learning Inc* |
| Graham Holdings Co | Zovio Inc | Adtalem Global Education Inc | Learning Tree Intl Inc |

To make sense of the jumble of firms in the table, it will help to define the "sometimes-comparable set"—the set of companies that are (i) found to be comparable, i.e., are among the *k* nearest neighbors, using at least one set of parameters, and are also (ii) found to be noncomparable using at least one other set of parameters. Also define the "always-comparable set" as the set of companies found to be within the *k* nearest neighbors of the target firm for *every* one of the four combinations of distance-measuring parameters. Table 7 lists the always-comparable and sometimes-comparable companies that would be selected for our target valuation case for each choice of *k*. For *k*=7 and 9, there are nine sometimes-comparable companies; for *k*=8, there are eight sometimes-comparable companies.

Table 7: Always-Comparable and Sometimes-comparable Sets, By *k*

| | Value of *k* | | |
|---|---|---|---|
| Peer Firm Companies | 7 | 8 | 9 |
| **Always-Comparable** | Capella Education Co<br>Strategic Education Inc | Capella Education Co<br>GP Strategies Corp<br>Grand Canyon Education Inc<br>Strategic Education Inc | Capella Education Co<br>GP Strategies Corp<br>Grand Canyon Education Inc<br>Strategic Education Inc |
| **Sometimes-Comparable** | Apollo Education Group Inc<br>Corinthian Colleges Inc<br>GP Strategies Corp<br>Graham Holdings Co<br>Grand Canyon Education Inc<br>Learning Tree Intl Inc<br>National Amern Univ Hldg Inc<br>Perdoceo Education Corp<br>Zovio Inc | American Public Education<br>Apollo Education Group Inc<br>Archipelago Learning Inc<br>Corinthian Colleges Inc<br>Graham Holdings Co<br>Learning Tree Intl Inc<br>National Amern Univ Hldg Inc<br>Perdoceo Education Corp<br>Zovio Inc | Adtalem Global Education Inc<br>American Public Education<br>Apollo Education Group Inc<br>Archipelago Learning Inc<br>Corinthian Colleges Inc<br>Graham Holdings Co<br>Learning Tree Intl Inc<br>National Amern Univ Hldg Inc<br>Perdoceo Education Corp<br>Zovio Inc |

The comparatively large numbers of sometimes-comparable companies might not matter much if these companies had similar enterprise value multiples. In that event, the predicted enterprise value multiple would not be very sensitive to the particular roster of comparable companies drawn from the sometimes-comparable companies. Table 8 reports summary statistics for the predicted enterprise value multiple for always- and sometimes-comparable companies, for each value of *k*. The first two columns report the mean and median, and these columns show generally slight differences in measures of central tendency for always- and sometimes-comparable companies' multiples.[50]

But these similarities mask the fact that the distributions of enterprise value multiples among the sometimes-comparable companies create the potential for substantial variation. For *k*=7, we saw in Table 7 that there were 2 always-comparable and 9 sometimes-comparable companies. This means that 5 companies must be drawn from the 9 sometimes-comparable companies to round out the comparable companies set. As the third column of Table 8 reports, there are 126 ways to choose 5 of 9 distinct objects. The final column of Table 8 reports the standard deviation of the enterprise value multiple across these 126 combinations of sometimes-comparable companies. This standard deviation is 8.0 units for the *k*=7 case, indicating that taking random draws of 5 peer companies from the sometimes-comparable companies roster would generate substantial variability in the resultant mean enterprise value multiple. For the other values of *k*, the standard deviation is even larger, at 10.0 units. Although we do not mean to suggest that actual comparable companies *are*

[50] The exception is for the median when *k*=8; the always-/sometimes-comparable companies difference is 5.4 units in this case.

randomly drawn, the large standard deviation helps explain why the predicted enterprise value multiples reported in Table 3 can have such a substantial range of variation.

Table 8: Enterprise Multiple Values for Firms in the Always-Comparable & Sometimes-Comparable Sets

| | Predicted Enterprise Multiple | | | |
|---|---|---|---|---|
| | Mean | Median | Number of Combinations of Sometimes-Comparable Peers | Sometimes-Comparable Peers' Std. Dev. |
| *Panel A: k=7* | | | | |
| Always-Comparable Peers | 21.6 | 21.6 | — | — |
| Sometimes-Comparable Peers | 20.6 | 20.5 | 126 | 8.0 |
| *Panel B: k=8* | | | | |
| Always-Comparable Peers | 25.1 | 25.9 | | |
| Sometimes-Comparable Peers | 22.1 | 20.5 | 126 | 10.0 |
| *Panel C: k=9* | | | | |
| Always-Comparable Peers | 25.1 | 25.9 | | |
| Sometimes-Comparable Peers | 23.2 | 23.0 | 252 | 10.0 |

In sum, analyzing the randomly chosen target firm Universal Technical Institute for the quarter ending September 30, 2011 illustrates the substantial leeway experts may have to choose results by choosing parameters. Of course, a randomly chosen example could just be an outlier. We thus turn to a more systematic consideration based on a large random sample of firms and quarters.

### *C. Comprehensive analysis of a large random sample of target firm-quarters*

We turn now to our broader simulation exercise, which involves a random sample of size 10,000 drawn without replacement from our data. The unit of analysis is a firm-quarter, which we call a "case." Note that because we have a firm-quarter panel of data and select at the firm-quarter level, some firms are drawn multiple times, and likewise for quarters (though any particular firm-quarter appears in our random sample either once or not at all).

Our key findings for the single randomly chosen target valuation case considered just above were:

(i) varying the matching parameters substantially affected the set of peers considered comparable companies, and

(ii) there is substantial variation in the predicted enterprise value multiple for the target firm and quarter.

Using our 10,000-size random sample, we show that these findings characterize comparable companies valuation generally.

To assess whether finding (i) persists, for each randomly chosen target valuation case we calculate the number of peer firms that are in the sometimes-comparable set (i.e., because they are

selected at least once but fewer than 24 times). Table 9 reports quartile cutoffs for the distribution of this statistic, considered across the 10,000 randomly chosen target valuation cases.[51] Panel A of the table shows that there are typically large numbers of possible combinations of sometimes-comparable companies, with median values of 792 for $k$=7 and in the thousands for the other two $k$ values. As a result, there is scope for wide variation in the enterprise value multiple, depending on which combinations of available comparable companies would be selected to fill out the $k$ comparable companies: Panel B shows that among sometimes-comparable companies, the standard deviation of the enterprise value multiple is above 10 even at the 25th percentile, with a median above 20 for each value of $k$. In sum, the results in Table 9 confirm that the combination of the underlying data structure and the 24 parameterizations we consider here can be expected to generate substantial variation in comps' enterprise value multiples.

Table 9: Summary Statistics Regarding Sometimes-Comparable Companies

| | $k$=7 | $k$=8 | $k$=9 |
|---|---|---|---|
| *Panel A: Number of Combinations* | | | |
| 25th Percentile | 210 | 252 | 462 |
| Median | 792 | 1716 | 3432 |
| 75th percentile | 5005 | 11440 | 31824 |
| *Panel B: Std. Dev. of Multiple Among Sometimes-comparable companies* | | | |
| 25th Percentile | 14 | 14 | 14 |
| Median | 20 | 21 | 21 |
| 75th percentile | 33 | 34 | 35 |

To assess question (ii), whether this allows experts considerable leeway to obtain results that favor a client's position relative to others, we calculate several statistics for each of the 24 parameterizations for each target valuation case. These include the predicted enterprise value multiple and its associated prediction error (the prediction minus the target's enterprise value multiplier for the target quarter). We also assess variability in predicted enterprise value, which equals the product of the predicted enterprise value multiple and target quarter EBITDA, as well as the prediction error involving this variable. We then calculate a variety of summary statistics across the 24 parameterizations for each target valuation case.

Table 10's first column provides average values, across the 10,000 randomly chosen target valuation cases, of various distributional statistics for the predicted enterprise value multiple, which was our focus in the previous subsection. For each target valuation case, we calculate the predicted enterprise value for each of the 24 parameterizations. Of these predicted multiples, the one with the smallest value is labeled the "min" predicted enterprise value multiple (in our example in the

[51] We do not report the mean, because it is clear there are some extreme outlier cases that dominate the rest of the distribution.

previous subsection, this value was 19.0). The average value, across the 10,000 target valuation cases, of the case-specific minimum predicted multiple is what we report in the first row of Table 10. Thus, across our 10,000 target valuation cases, the minimum predicted enterprise value multiple averaged 38.

Table 10: Average Value of Summary Statistics Regarding Enterprise Value Multiples
(unit of analysis is target valuation case)

| | Average Value of | | | |
|---|---|---|---|---|
| | Enterprise Value Multiple | | Enterprise Value | |
| | Prediction | Prediction Error | Prediction | Prediction Error |
| Target valuation case's: | | | | |
| Min | 38 | -25 | 7408 | -1398 |
| 3rd order statistic | 40 | -23 | 7857 | -949 |
| Median | 49 | -14 | 9331 | 525 |
| 22nd order statistic | 69 | 6 | 12338 | 3532 |
| Max | 76 | 13 | 13427 | 4621 |
| *Note: Actual means are: enterprise value multiple=63, enterprise value=8806.* | | | | |

The remaining rows of Table 10 are calculated analogously. For a given target valuation case, the 3rd order statistic is the third-smallest value resulting from the 24 predictions for that case, the median is the midpoint between the 12th and 13th order statistics, the 22nd order statistic is the third-largest value, and the max is the largest of the 24 values. The statistics reported in the table are the averages of each of these summary statistics, calculated across the 10,000 randomly chosen target valuation cases. On average, then, the max value of the predicted enterprise value multiple averaged 76 across target valuation cases—twice the min value. As with our single example in the previous subsection, the 3rd and 22nd (of 24) order statistics show that this large range of variation is not an artifact of looking only at the min and the max; the range between them is 29, nearly as great as the range of 38 between the min and max.

The second column of Table 10 reports the corresponding statistics for enterprise value multiple prediction error. For each target valuation case, we create the 24 enterprise value multiple prediction error values by subtracting the true enterprise value multiple (this was 24.2 in our single firm example in the previous subsection) from each of the 24 predicted enterprise value multiples. Because the true enterprise value multiple is a constant for each target valuation case, each of the 24 summary statistics in Table 10's second column is just the first-column value shifted left by the mean true enterprise value multiple across all target valuation cases, which was 63. It is interesting to note that across target valuation cases, the median predicted enterprise value multiple averaged 49, so that using the case-specific median as the prediction value would be associated with an average prediction error across cases of -14 (=49-63). In fact, the 22nd order statistic is more accurate on average, with an average enterprise value multiple prediction error of only 6; this is about 10% of the cross-case average value of the true enterprise value multiple.

The third column of Table 10 reports average values, across our 10,000 target valuation cases, of case-specific summary statistics for predicted enterprise value. For each of the 24 parameterizations for a case, we calculate implied enterprise value by multiplying together that parameterization's predicted enterprise value multiple and the target valuation case's actual EBIT value. As an example, a target valuation case with an actual EBIT value of 10 and a minimum predicted multiple across the 24 parameterizations of 20 would have a minimum predicted enterprise value of 200 (=10x20). For reference, the average of the *actual* enterprise value in our data was 8,806.

Table 10's third column shows that on average, the minimum predicted enterprise value across our 10,000 target valuation cases was 7408. On average, the maximum predicted enterprise value was nearly twice this value, at 13427.[52] Turning to the case-level median predicted enterprise value, it averaged 9,331.

The final column of Table 10 reports the average values of case-level summary statistics related to prediction error for enterprise value. The prediction error associated with the case-level medians averaged only 525, or about 6% of the true average EV among the 10,000 target valuation cases of 8,806. This prediction error is noticeably lower than the prediction error at the other distributional points reported in Table 10. Thus, the case-level median predicted enterprise value is considerably more accurate than the other reported case-level distributional points.[53]

Summing up findings from this subsection:

- The 24 parameterizations yield substantial variability in the ranges of both predicted enterprise value multiples and the implied predicted enterprise value itself. By picking among them, opposing testifying experts could easily come up with very different valuations, using the same data and the same general *k*-nearest neighbors approach to peer companies—varying only the particular choices in how that approach is operationalized.
- The differences in valuation that result from these parameterization choices are economically significant. For example, using the 22nd order statistic of the predicted enterprise value multiple yields an average enterprise value prediction of 12338—nearly 60% greater than the average prediction of 7857 based on the 3rd order statistic.
- Because actual EBIT among our target cases is negatively associated with the predicted enterprise value multiple, the practice of predicting multiples might be less accurate than would be a practice of predicting enterprise value directly. This posy is one we plan to take up in future versions of this paper.

---

[52] The fact that the average value of the max predicted enterprise value multiple was exactly twice the average value of the min predicted enterprise value multiple indicates that actual EBIT and the ratio of max-to-min predicted enterprise value multiple are negatively associated across cases; if the max predicted enterprise value multiple were exactly twice the min predicted enterprise value multiple for all firms, then the predicted enterprise value max-to-min ratio also would be exactly 2.

[53] As discussed above in relation to the second column of Table 10, the prediction error of the enterprise value multiple is -14; relative to the actual mean enterprise value multiple of 63, this is a prediction error of more than 20%.

## IV. Choosing the Optimal Valuation Using Nearest-Neighbors on Pre-Target-Quarter Neighbor

This section considers one example of how a more systematic approach to valuation could work. For a given target valuation case—target firm and quarter—we determine which peer firms could serve as potential comps. We then use data on the target firm and these available peers to determine which of the 24 parameterizations discussed in Section III would have yielded the lowest enterprise value multiple prediction error using data from the 8 quarters preceding the target quarter. We call that parameterization the "optimal parameterization." Then we calculate the predicted enterprise value multiple using each of the 24 parameterizations for the target quarter (i.e., we do what we did in Section III). Finally, we assess how well the optimal parameterization does by comparison to a randomly chosen parameterization, as well as to the 3rd and 22nd order statistics of the set of predicted enterprise value multiples that we discussed in Section III.

### *A. Performance of the optimal k-nearest neighbors parameterization: Simulation evidence*

This part of our analysis again involves randomly choosing a sample of 10,000 target valuation cases. Because we use lagged data for the target firm and its peers, we need to add data restrictions to ensure that each included firm has all data required for our analysis of each target valuation case.[54]

For each target valuation case, we determined which peer firms were available for use as comps in both the target quarter and in the 7 quarters immediately preceding it. Because the nearest neighbors approach we described in Section III uses peer- and target-firm data from the quarter before the target quarter, this means available peers are those with non-missing data on the ratio of enterprise value to EBIT for the 8 quarters ending in the target quarter, as well as non-missing data on the RP and PN covariates for the 8 quarters ending in the quarter before the target quarter. For simplicity, we required that included firms have non-missing data on all the variables just described for the target quarter and the 8 quarters preceding it.

After imposing these conditions, we were left with 47,667 firm-quarters available to serve as simulated target valuation firm-quarters. The data contain 2,393 distinct firms arrayed across 76 distinct quarters. We note that the additional data requirements for this section led to a large reduction in the number of firm-quarters available—more than 40% of available firm-quarters in our first simulation are unavailable for use in this one due to missing data.[55] Thus, the firm-quarters analyzed in this section are not necessarily representative of those analyzed in Section III.

For each target valuation case, we repeated our Section III simulation calculations for the target quarter—i.e., we determined the peers that would be selected as comps under each of the 24 parameterizations described in Section III, and then we used the selected peers from each parameterization to calculate the target-quarter prediction of the target firm's valuation.

---

[54] Future versions of this paper may include the use of multiple imputation to fill in the blanks of missing variables.

[55] [Something on imputing data rather than dropping missing values.]

Our goal was to use a data-driven approach to determining which of the 24 parameterizations is best in predicting value in the target quarter. We devise a method for using pre-target quarter data to measure each parameterization's errors in predicting the enterprise value multiple. Then we use these pre-target quarter prediction errors to determine which parameterization has the least aggregate prediction error over the pre-target quarter period; we dub this the optimal parameterization. Then we use this optimal parameterization to predict the target quarter enterprise value multiple. Because the designation of optimality was made using pre-target quarter data, there is no guarantee that the designated-optimal parameterization is actually any better than a randomly chosen parameterization would be.[56] Finally, we compare the target-quarter performance of the optimal parameterization to the performance of the other parameterizations, as described below.[57]

We emphasize that the optimality involved here is limited to optimizing over 7 quarters of pre-target quarter data *for each target firm and quarter considered in isolation*. This is not a lot of pre-target quarter data, and the number of available peers can be small. Accordingly, it is important not to overstate our claim of optimality here. We return to this point below.

To implement our procedure, we repeated the steps described in Section III for the target firm and its available peers for each of the 7 most recent quarters before the target quarter. That yields 7 quarterly predicted valuations for each parameterization, as well as the associated prediction errors based on the target firm's actual enterprise value multiple for the same 7 quarters. The optimal parameterization is the one that has least aggregate prediction error based on these 7 quarters.

We measure pre-target quarter aggregate prediction error in two ways. First, we compute each parameterization's sum of squared prediction errors over the 7 pre-target quarters. Second, we compute the sum of the absolute values of prediction errors over the 7 pre-target quarters. The parameterizations with the lowest values of aggregated prediction error for each target valuation case are then, respectively, the optimal squared-error parameterization and the optimal absolute-value parameterization. The two aggregated prediction error methods might yield either the same or different optimal parameterizations in any target valuation case.

We next suppose that a testifying expert could be expected to report which parameterization is optimal for each of the squared-error and absolute value approaches. The predicted valuation of the target case could then be based on the optimal parameterization(s) for the target quarter. That is, the expert could be expected to report the one (or two) valuation prediction(s) that result from using the optimal parameterization(s) for the target quarter.

---

[56] For example, if each firm's relationship between covariates and the enterprise value multiple were independent across quarters, then even though the designated-optimal parameterization's pre-target quarter performance was best by construction, its *target-quarter* performance would, on average, be indistinguishable from the other parameterizations' performance.

[57] In standard machine learning terms, one can view our target quarter data as being the test set, with the 7 quarters of pre-target quarter data being the training set. [Something about cross-validation here?]

### 1. An example with one firm

To illustrate this procedure, we walk through another one-firm example. This time our target firm is Balchem Corp,[58] which is in two-digit SIC group 28 (Chemicals and Allied Products), and our target quarter is the one ending September 30, 2002. In that quarter, its enterprise value multiple was 29.8, with its actual enterprise value being 101.6, and its EBIT being 3.4.[59] Balchem had 38 useable peers for use in our analysis.

The optimal parameterization for predicting Balchem's valuation in the target quarter when we used the squared-error approach to measuring pre-target quarter prediction error was the one that used Rosenbaum & Pearl covariates, used the Mahalanobis distance metric, used the median summary measure (i.e., use the median of comps' enterprise value multiples to predict Balchem's target quarter enterprise value multiple), and used the 9 nearest neighbors as comps. Using the absolute value approach to measuring pre-target quarter prediction error, the optimal parameterization also involved the Rosenbaum & Pearl covariates, the median summary measure, and used the 9 nearest neighbors, but the optimal parameterization for the absolute value approach used the standardized Euclidean metric rather than the Mahalanobis.

Table 11 reports the simulation results for Balchem in the target quarter. Its top set of rows shows that the actual enterprise value multiple was 29.8. For the squared-error approach, the optimal parameterization yielded a prediction of 40.8, for a prediction error of 11.0; for the absolute value approach, the optimal parameterization's prediction was 32.0, for a prediction error of just 2.1. The top panel also reports the prediction error for the 3rd order statistic and the 22nd order statistic in the target quarter. These were 8.5 and 22.6 (by construction, they do not vary across the approach used to aggregate prediction errors from the pre-target quarter period). Thus in the Balchem case for target quarter ending 9/30/2002, the 3rd order statistic slightly outperformed the optimal parameterization using the squared approach. Using the absolute value approach, though, the optimal parameterization actually achieved the lowest target quarter prediction error.

The table's next set of rows report transformations of the prediction errors in the target quarter using the respective aggregation approaches to measure prediction error, i.e., using the squared prediction error in the first column and the absolute value in the second column.[60] Row [a] reports the minimum value of the aggregated prediction error in the target quarter among all 24 parameterizations, which is not necessarily the optimal parameterization's prediction error given that optimality is determined using only pre-target quarter data. The minimum squared prediction error in the target quarter was 4.5, and the minimum absolute value prediction error was 2.1. Row [b]

---

[58] The firm we used to illustrate our procedure in Section III does not have enough useable peers for the pre-target quarters to be included in the analysis in this section; otherwise we would have used it to illustrate the procedure here as well.

[59] This is the actual enterprise value multiple, rounded to one digit; because the enterprise value and EBIT statistics in the text are rounded, the ratio of the statistics reported in the text is slightly different.

[60] Because all the prediction errors involved were positive, the absolute value approach yields the same values as raw prediction errors.

reports the aggregated prediction error for the optimal parameterization, which was 120.1 for the squared approach[61] and 2.1 for the absolute value approach.

Rows [c] and [d] report the aggregated prediction errors for the parameterizations that were the 3rd and 22nd order statistics of the enterprise value multiple prediction series. The 3rd order statistic had a squared prediction error of 71.9, and its absolute value prediction error was 8.5. For the 22nd order statistic, the squared prediction error was 509.5 and the absolute value was 22.6.

Table 11: Optimal Parameterization Results for Balchem, Quarter Ending 9/30/2002

| Variable | Error-Aggregation Approach | |
|---|---|---|
| | Squared-Error | Absolute Value |
| *Enterprise value multiple in target quarter* | | |
| Actual | 29.8 | 29.8 |
| Prediction, optimal parameterization | 40.8 | 32.0 |
| Prediction error, optimal parameterization | 11.0 | 2.1 |
| Prediction error, 3rd order statistic | 8.5 | 8.5 |
| Prediction error, 22nd order statistic | 22.6 | 22.6 |
| *Prediction errors in target quarter, using aggregation approach†* | | |
| [a] Minimum among 24 parameterizations | 4.5 | 2.1 |
| [b] Optimal parameterization | 120.1 | 2.1 |
| [c] 3rd order statistic | 71.9 | 8.5 |
| [d] 22nd order statistic | 509.5 | 22.6 |
| *Relative error in target quarter††* | | |
| Minimum among 24 parameterizations ([a]÷[a]) | 1 | 1 |
| Optimal parameterization ([b]÷[a]) | 27 | 1 |
| 3rd order statistic ([c]÷[a]) | 16 | 4 |
| 22nd order statistic ([d]÷[a]) | 113 | 11 |
| Mean, across all 24 parameterizations, of relative error | 60 | 7 |
| Optimal parameterization relative error < mean relative error? | Yes | Yes |

*Notes:*

Optimal parameterizations:
For squared-error approach: Rosenbaum & Pearl, Mahalanobis, Median, $k$=9
For absolute value approach: Rosenbaum & Pearl, Standardized Euclidean, Median, $k$=9

† Reported values are squared, in the Squared-Error column, and absolute value, in the Absolute Value column.

†† Reported values are ratios involving squared prediction error, in the Squared-Error column, and absolute value, in the Absolute Value column.

[61] This figure differs slightly from the square of 11.0, which is the value of the prediction error for the squared approach reported in the top set of rows, as already discussed, only because of rounding.

The table's bottom panel reports relative errors for the target quarter. These are computed by dividing each aggregated prediction error in the second set of rows by the minimum aggregated prediction error. By construction, the relative error for the minimum aggregated prediction error is 1. The optimal parameterization's relative error is 27 for the squared-error approach—the ratio of 120.1 to the minimum squared prediction error of 4.5. A relative error of 27 sounds like quite poor performance for the optimal parameterization in the squared-error approach. However, the final row of Table 11 shows that for the squared-error approach, the mean of the relative errors across all parameterizations was a whopping 60.[62]

Thus, it seems the CC approach to predicting Balchem's valuation for the quarter ending 9/30/2002 is just a hard problem, at least using nearest-neighbor comps. As poorly as it does, the optimal parameterization nevertheless substantially outperforms the mean—which is a natural measure of the performance of a randomly chosen parameterization. In this target valuation case using squared-error aggregation, the 3rd order statistic actually does better in the target quarter than the optimal parameterization. However, the relative error of the 22nd order statistic exceeded 100.

As for the absolute value approach, because the optimal parameterization achieves the minimum prediction error for that approach, its relative error is 1. By its nature, the absolute value approach puts less weight on extreme outlier errors, given that they are not squared. Still, the relative error for the 3rd order statistic using the absolute value approach is 4 times the minimum absolute prediction error, and the corresponding 22nd order statistic has relative error of 11, an order of magnitude worse than the optimal parameterization; in between, the mean relative error for the absolute value approach was 7.

The Balchem 9/30/2002 example illustrates several points. First, the optimal parameterization based on pre-target quarter data might or might not yield the least prediction error for the target quarter itself. Second, it is possible for the optimal parameterization to miss by a lot, with apparently high relative error compared to the minimum prediction error for the target quarter—and it is also possible that even in such a case, the optimal parameterization might substantially outperform the expected value of the relative error from randomly choosing the parameterization. Third, the party-biased order statistics we considered in Section III might miss by more or less than the optimal parameterization—and when they miss, they might miss by a lot.

### 2. Larger-scale simulation results

We now turn to our second set of simulation results for target valuation cases. As in Section III, we randomly drew 10,000 target valuation cases—i.e., target firm-quarter combinations. For each case, we followed the procedure just described for the quarter ending 9/30/2002 for Balchem, which was one of the target valuation cases we randomly drew. Because our code did not finish successfully for four of our 10,000 target valuation cases, we actually have 9,996 cases.

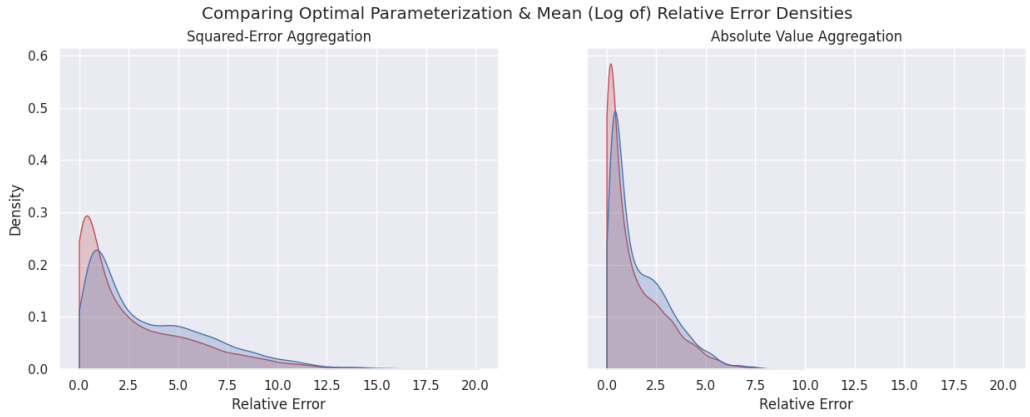
---

[62] This statistic is calculated as the average, across all 24 parameterizations, of the relative error for the squared error approach.

Our primary focus is on the target quarter relative error performance of the optimal parameterization based on pre-target quarter data. In particular, we are interested in whether using this optimal parameterization outperforms other feasible ways to choose the parameterization. One simple alternative to using the optimal parameterization would be to randomly choose which of the 24 parameterizations to use. The expected relative error from such an approach is given by the mean relative error across all 24 parameterizations.

Figure 3 displays density plots over our 10,000 simulations for the optimal parameterization's relative error (displayed in red) and the mean relative error (displayed in blue). We used the log of the relative error because there were extreme outliers that distort the density plot without this transformation; by construction, a log minimum relative error of 0 is the lowest possible value. The figure's left panel is for the squared-error approach to error aggregation, and the right panel is for the absolute value approach. Both plots show that the optimal parameterization's density (in red) is pushed up closer to zero than is the density for the mean relative error (in blue). These plots thus provide evidence that the optimal parameterization outperforms a randomly chosen parameterization. Additional evidence comes from the fact that the optimal parameterization had lower relative error than the mean relative error in roughly two-thirds of our 10,000 simulations reps (for the squared-error approach this share was 68%, and for the absolute value approach it was 62%).

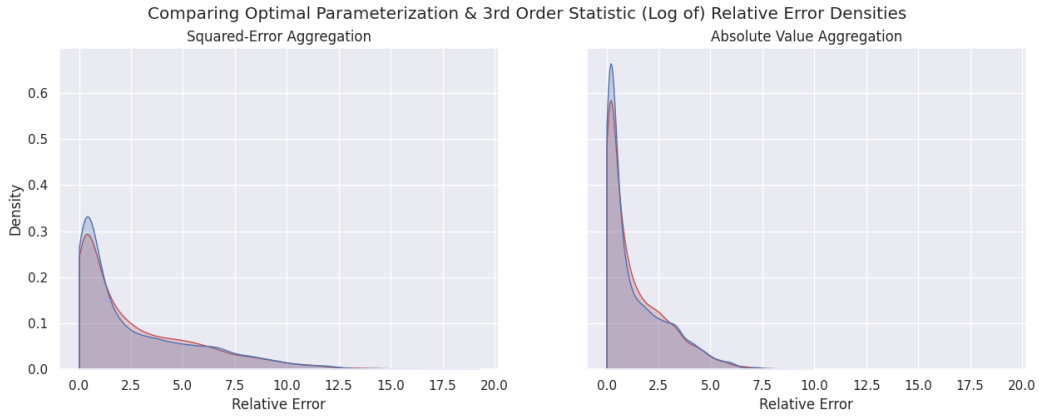
Figure 3




The optimal parameterization similarly outperforms the 22nd order statistic. The pair of plots in panel (a) of Figure 4 show densities for the (log) relative error for the optimal parameterization (in red) and the 22nd order statistic (in blue); these are as clearly in favor of the optimal parameterization as for the mean relative error comparison in Figure 3. The optimal parameterization relative error actually did slightly less well against the 22nd order statistic's relative error; relative error is lower for the optimal parameterization in 61% of all simulations.

Figure 4


(a)




(b)

Perhaps surprisingly, though, the optimal parameterization is slightly worse than the 3rd order statistic of the target quarter. This is evident in the pair of density plots in panel (b) of Figure 4. In addition, the optimal parameterization has lower relative error than the 3rd order statistic in only 48.2% of simulations (this figure is the same for the squared-error and absolute value approaches).

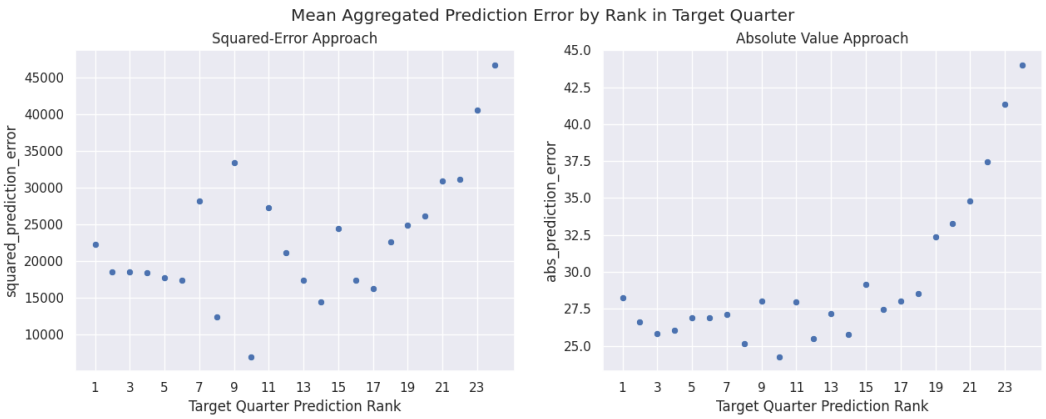
It is natural to wonder why the optimal parameterization based on pre-target quarter data does not do better against a parameterization, the 3rd order statistic, that one might think would likely be chosen by an expert exercising discretion to look for defendant-favoring results (assuming lower valuation benefits defendants in valuation disputes). We can't give a full explanation why this is so at this stage. But we can provide some basic statistics that help show, descriptively, why this is happening.

Figure 5 plots the mean aggregated prediction error (squared-error on the left, absolute value on the right) by target quarter rank of the parameterization.[63] Thus, the point on the x-axis labeled "1"

[63] Unlike the density plots above, these plots use the mean level, rather than log, of the aggregated prediction errors.

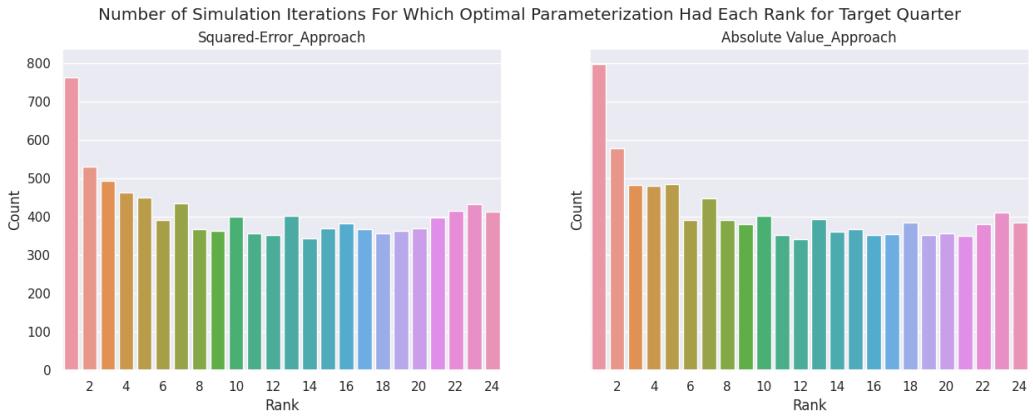
corresponds to the average squared prediction error, over all 10,000 simulation replications, for the parameterization that had the smallest enterprise value multiple in a given simulation replication; the point labeled "3" corresponds to the 3rd order statistic, and so on. The left-side figure shows that the lowest order statistics tend to have quite low squared-error by comparison to many of the other order statistics. The one on the right side tells an even clearer story, with the absolute value of prediction error being relatively similar, on average, among most statistics below the 19th or so—and with the 3rd order statistic being among those with the lowest average absolute value prediction error.

Figure 5




Why do the results in Figure 5 matter? Figure 6 shows why, by plotting the number of simulation replications for which the optimal parameterization takes on each target quarter rank in terms of aggregated prediction error. The plot on the left side shows that the optimal parameterization was the lowest-rank target quarter parameterization in roughly 750 of all simulation replications—by far the most common rank for the optimal parameterization. The first order statistic does worse than the 3rd order statistic in both sides of Figure 5. Further, over the remainder of the 10,000 simulation replications, the optimal parameterization is spread across the other order statistic ranks, with a substantial fraction of replications turning out to be among the higher-ranking order statistics—those which Figure 5 shows have very high aggregated prediction error.

Figure 6




Together, then, Figure 5 and Figure 6 show that there are limits to the value of the optimizing process that underlies the optimal parameterization—finding the parameterization with least aggregated prediction error across the 7 quarters preceding the target quarter. Our results are extremely preliminary. But this finding perhaps suggests that there is limited capacity to improve nearest neighbor-based comparable companies analysis using optimization with pre-target quarter data. One reason that could be true is that there might be a relatively limited—or perhaps a rather complex—relationship between the target quarter enterprise value multiple and the lagged values of the covariates on which comparable companies analysis is conventionally based. Roughly speaking, to put it in conventional statistical terms, if the relationship between Y and X is not simple, then choosing peers that have similar X might have limited value in predicting Y.

## V. Other ML Methods

Our approach in Sections III and IV has the virtue of being rooted in the basic framework commonly used in CC analysis. We used the same overall set of potential peer firms to select comps, and we used the same basic computational methods to determine the performance of each parameterization. Our only substantive methodological innovation was to rely on a data-driven approach to selecting which parameterization to use to do the actual target quarter prediction.

But there is no reason why we, or courts, should restrict attention to the nearest neighbors framework. Many machine learning algorithms exist, and others might well perform better than nearest neighbors. Nor must attention be restricted to predicting enterprise value multiples; we could also consider the possibility of predicting enterprise value directly. That is, rather than predicting the enterprise value to EBIT multiple, and then multiplying the predicted multiple by actual target quarter EBIT, we could simply predict enterprise value directly. A concern with these suggestions is that we might be introducing additional expert degrees of freedom—before related to which parameterization of the nearest neighbors framework is used, and now related to which algorithm is used, and which target variable (enterprise value multiple or enterprise value itself).

Machine learning methods allow optimization over all these dimensions at once by using ensemble methods. These methods use information on the prediction error from different individual algorithms in order to determine a combination of those algorithms that has lower prediction error than any of the individual algorithms. The actual predicted valuation would therefore be based on the results of this combined algorithm—the ensemble. Because the ensemble's particular characteristics are data-driven, using this approach eliminates expert discretion in model selection, reducing expert degrees of freedom.

Future versions of this paper will engage with additional approaches.

## VI. Valuation, Machine Learning, and Evidence Law

[To be added.]

## Conclusion

[To be added.]

**References**


1. Pat Akey, Adriana Robertson & Mikhail Simutin, *Noisy Factors* (working paper, 2022)
2. Andrew Baker & Jonah B. Gelbach, *Machine Learning and Predicted Returns for Event Studies in Securities Litigation*, Journal of Law, Finance, and Accounting, v5, pp. 231–272 (2020)
3. Ryan Bubb, Emiliano Catan & Holger Spamann, *A Functional Analysis of Shareholder Rights in Mergers* (working paper 2022)
4. William Carney, Robert Bartlett & George Geis, CORPORATE FINANCE: PRINCIPLES AND PRACTICE (2022)
5. Anthony J. Casey & Julia Ann Simon-Kerr, *A Simple Theory of Complex Valuation* 113 MICH. L. REV. 1175 (2015)
6. Albert H. Choi & Eric L. Talley, *Appraising the "Merger Price" Appraisal Rule*, 34 J. L. ECON. & ORG. 543 (2018)
7. Aswath Damodaran, DAMODARAN ON VALUATION (2nd Ed. 2006).
8. Thomas G. Dietterich, *Ensemble Methods in Machine Learning*, in MULTIPLE CLASSIFIER SYSTEMS (Josef Kittler & Fabio Roli eds. 2000)
9. Gregory Eaton, Feng Guo, Tingting Liu & Micah Officer, *Peer selection and valuation in mergers and acquisitions*, 146 J. FIN. ECON. 230 (2021)
10. David Freeman Engstrom & Jonah B. Gelbach *Legal Tech, Civil Procedure, and the Future of Adversarialism*, 169 U. Pa. L. Rev. 1001 (2021)
11. Christine Funk, *Daubert Versus Frye: A National Look at Expert Evidentiary Standards* (April 11, 2022) (available at https://www.expertinstitute.com/resources/insights/daubert-versus-frye-a-national-look-at-expert-evidentiary-standards/).
12. Trevor Hastie, Robert Tibshirani & Jerome Friedman, ELEMENTS OF STATISTICAL LEARNING (Springer 2nd Ed. 2017).
13. Charles Korsmo & Minor Myers, *The Flawed Corporate Finance of Dell and DFC Global*, 68 EMORY L. J. 221 (2018).
14. Gareth James, Daniela Witten, Trevor Hastie & Robert Tibshirani, AN INTRODUCTION TO STATISTICAL LEARNING (WITH APPLICATIONS IN R) (Springer 2nd Ed. 2021).
15. Shannon Pratt & Alina V. Niculita, THE LAWYER'S BUSINESS VALUATION HANDBOOK (American Bar Association, 2nd Ed. 2010).
16. Joseph Rocca, *Ensemble methods: bagging, boosting and stacking* (Apr 22, 2019), Towards Data Science Blog (available at https://towardsdatascience.com/ensemble-methods-bagging-boosting-and-stacking-c9214a10a205)
17. Roberta Romano, *After the Revolution in Corporate Law*, 55 JOURNAL OF LEGAL EDUCATION 342 (September 2005).
18. Joshua Rosenbaum & Joshua Pearl, INVESTMENT BANKING (Wiley 2013).
19. Andreas Schueler, *Valuation with Multiples: A Conceptual Analysis,* J. BUS. VALUATION & ECONOMIC LOSS ANALYSIS (2020).
20. Eric Talley, *Finance in the Courtroom: Appraising Its Growing Pains*, DEL. LAWYER (2017).