Statistical String Theory for Courts: If The Data Don't Fit . . . .

David F. Babbel*

Vincent James Strickler**

Ricki Sears***

First Draft: July 1, 2007

Current Draft: July 15, 2008

* The Wharton School of Business, University of Pennsylvania, and CRA International; babbel@wharton.upenn.edu

** Political Science Department, Utah State University; James.Strickler@usu.edu

*** Department of Economics, University of Texas at Austin; Ricki.Sears@gmail.com

The authors express their gratitude to Dr. Kabir Dutta of CRA International, Professor David Hoaglin of Harvard University, Professor James McDonald of Brigham Young University, and the late Professor John Tukey (Princeton University) for their pioneering studies in the statistical theory that underlies this article and helpful suggestions, and Bernard J. Bonn III, J.D., of Dechert for his assistance in preparing the leading legal case. The authors are also grateful to Professor Neil A. Weiss of Arizona State University, Dr. Mark Meyer, Professor Art Rosenbloom, J.D., Dr. Miguel Herce, and Daniel Garrie, J.D., of CRA International, and Elise Hahl and Lisle Updike for their encouragement, insights, and editorial assistance.

Statistical String Theory for Courts: If The Data Don't Fit . . . .

Abstract

The primary purpose of this article is to provide courts with an important new tool for applying the correct probability distribution to a given legal question. In areas as diverse as criminal prosecutions and civil lawsuits alleging securities fraud, courts must assess the relevance and reliability of statistical data and the inferences drawn therefrom. But, courts and expert witnesses often make mistaken assumptions about what probability distributions are appropriate for their analyses. Using the wrong probability distribution can lead to invalid factual conclusions and unjustified legal outcomes. To deal with this problem, we propose the use of a unifying "statistical string theory" – the g-and-h distribution – in legal settings. This parent distribution subsumes many other distributions and spans the widest range of possible skewness-kurtosis combinations. The capacity of the g-and-h distribution to accommodate such a wide variety of data can alleviate judicial fact finders of the difficult task of trying to correctly select among competing distributions. Finally, we report the successful use of this statistical tool in a trial setting for financial data analysis – showing that it can produce more accurate inferences, than those drawn from alternative distributions, and these differences can be judicially decisive.

Part I: Mistaken Distributional Assumptions

Expert witnesses often confront courts with massive amounts of statistical data, leaving judges and juries with a daunting task. The courts must grapple with the reliability and relevance of the data as well as the inferences that experts (and advocates) draw from them. Ultimately, the courts must determine whether the statistical data and the conclusions drawn from them constitute evidence that meets legal standards for certainty.

More than once, the United States Supreme Court has endorsed the use of probability theory as a tool to help decide such cases.[1] But, in considering such statistical evidence, the Court has not been content to passively accept the calculations and conclusions of the contesting parties. In *Daubert v. Merrell Dow Pharmaceuticals*, the Supreme Court specifically counseled judges to engage in an "assessment of whether the reasoning or methodology underlying the testimony is scientifically valid and of whether that reasoning or methodology properly can be applied to the facts in issue."[2] By their examples – such as when Justice Harry Blackmun in the *Castaneda v. Partida* jury selection case,[3] and Justice Potter Stewart in the *Hazelwood School District v. U.S.* employment discrimination case,[4] both calculated standard deviations – the Supreme Court has even encouraged lower court judges to perform their own statistical analyses. Emboldened by such instructions, examples, and the Supreme Court's assurance that "[w]e are confident that federal judges possess the capacity to undertake this review,"[5] lower courts, while admitting "[w]e are not expert statisticians,"[6] have often willingly "embark[ed] upon a journey into the statistical maze."[7]

Statistical methods have most prominently been put to judicial use in employment discrimination cases.[8] Judge Robert L. Carter of United States District Court for the Southern District of New York explained why in the case of *Ste. Marie v. Eastern Railroad Association*:

> Since it is well recognized that specific acts of discrimination may be hard come by in a disparate treatment case, reliance upon inferential proof of purposefulness is appropriate . . . . Plaintiff here has quite properly sought to structure a prima facie showing of sex discrimination on a foundation of statistical proof. Courts have

---

[1] *E.g.*, Castaneda v. Partida, 430 U.S. 482, 497 (1977) and Hazelwood School District v. U.S., 433 U.S. 299, 311 (1977).

[2] Daubert v. Merrell Dow Pharms., 509 U.S. 579, 592-593 (1993). *See also* Presseisen v. Swarthmore College, 442 F. Supp. 593, 600 (E.D. Penn. 1977) (In rejecting a sex discrimination claim against Swarthmore College, noted that a "[c]ourt must carefully analyze the particular statistical methods used and then, assuming their propriety, must determine whether the statistics, either alone or in conjunction with other evidence, establish" what is claimed).

[3] *Castaneda*, 430 U.S. at 496 n. 17.

[4] *Hazelwood*, 433 U.S. at 311 n. 17.

[5] *Daubert*, 509 U.S. at 593.

[6] Palmer v. Schultz, 815 F.2d 84, 93 n. 8 (D.C. Cir. 1987).

[7] *Id.* at 93.

[8] *E.g.*, Equal Employment Opportunity Comm'n v. Akron National Bank & Trust Co., 497 F. Supp. 733 (N.D. Ohio 1980) (wherein a chi-square calculation was accepted as sufficient proof of sex discrimination).

come to regard statistical data as reliable indicia of patterns and practices in Title to VII cases."[9]

In considering such statistical arguments, courts have been required to decide what levels of statistical proof are sufficient. Most commonly, judges have naively borrowed these standards from the social sciences,[10] such as when the United States Court of Appeals for the District of Columbia Circuit held that a "disparity [in group representation within a workforce] measuring two standard deviations (to be more precise, 1.96 standard deviations) correspond[ing] to a 5% probability of randomness under a two-tailed test" was at least necessary to establish an inference of unlawful discrimination.[11] Essentially, a court applying these standards is requiring that there be only a one in twenty chance that observed differences can be explained by random chance, before the court will accept that there is sufficient proof of discrimination.

Such commonly utilized requirements assume that random departures from a nondiscriminatory ideal will conform to a "normal" (bell-shaped) distribution (also known as a Gaussian distribution).[12] The underlying theory of normal distributions can be explained thus:

---

[9] Ste. Marie v. Eastern Railroad Assoc., 458 F. Supp. 1147, 1162 (S.D. N.Y. 1978) (citing Rogers v. International Paper Co., 510 F.2d 1340; Muller v. United States Steel Corp., 509 F.2d 923 (10th Cir.); Brown v. Gaston County Dyeing Machine Co., 457 F.2d 1377; United States v. Jacksonville Terminal Co., 451 F.2d 418 (5th Cir. 1971); and Parham v. Southwestern Bell Telephone Co., 433 F.2d 421 (8th Cir. 1970)).

[10] Castaneda v. Partida, 430 U.S. 482, 497 n. 17 (1977) ("As a general rule for . . . large samples, if the difference between the expected value and the observed number is greater than two or three standard deviations, then the hypothesis that the jury drawing was random would be suspect to a social scientist."). Such methodological borrowing has been criticized by some; see Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U.L. REV. 385, 413 (1985) ("Merely to borrow a standard from the scientific world without examining the values implicit in such a standard is a mistake. This point applies not only to proof of discrimination, but to broader issues in legal proof as well. The [Supreme] Court's lack of analysis of this issue [in *Castaneda*] is an abdication of the responsibility to determine an appropriate allocation of risks.")

[11] *Palmer* 815 F.2d at 92. Not all courts agree that even these, rather exacting, standards are sufficient. A Federal District Court in California rejected statistical evidence that there was only a probability of ".22%" that "random occurrence" could explain the racial patterns in hiring of hotel waiters, because "viewing the data as a whole the Court cannot conclude that they raise an inference of purposeful discrimination." The District Court argued that "The Supreme Court, while noting that disparities 'greater than two or three standard deviations' would be suspect to a social scientist, has never accepted that level as sufficient to raise an inference of intent. In the cases in which it has applied this analysis to determine the presence of purposeful discrimination, it has relied on disparities ranging from five to 29 standard deviations . . . . Statistical disparities considerably more gross and long-lasting than those found here being required to support an inference of purposeful discrimination, the Court finds and concludes that plaintiffs have failed to establish a prima facie case as to their class claims . . . ." Gay v. Waiters' & Dairy Lunchmen's Union, 489 F. Supp. 282, 311 (N.D. Cal. 1980) (citing *Castaneda*, 430 U.S. at 497 and Hazelwood School District v. U.S., 433 U.S. 299, 311 (1977)).

[12] *See* Fred S. McChesney, *Statistics: The Language of Science (Part II)*, 9 KAN. J.L. & PUB. POL'Y 75, 77 (1999) ("This discovery that things take on a bell-shaped curve is what's called, in statistics, the central limit theory. The central limit theorem and distributions are described as 'normal,' or sometimes 'Gaussian' after the German mathematician Gauss who was responsible for developing the central limit theorem. The central limit theorem is the proposition that things are distributed essentially in this bell-shaped way . . . . [W]hy is it that we have this bell-shape emerge? Essentially, what we're talking about is a situation where events themselves, outcomes themselves, are the result of many different influences, all of them acting more or less independently of one another. If we went out and got a distribution of heights or weights of the population, we would surely find something that had this familiar bell-shape. Heights and weights are themselves reflective of any number of different impacts: heredity, genetics, diets, exercise, and what have you . . . .").

> A basic theorem of mathematical statistics links . . . randomness with the famous bell-shaped, normal distribution. More precisely, the central limit theorem states that "the sum of a large number of independent random variables will be approximately normally distributed almost regardless of their individual distributions; any random variable which can be regarded as the sum of a large number of small, independent contributions is thus likely to follow the normal distribution approximately."[13]

The characteristic "bell shape" (in two dimensions) of a normal distribution has a central symmetrical hump or mound and gets wider (in terms of its horizontal reach) yet thinner (in terms of its vertical height) as it extends downward and outward towards its lower edge, or "tails." Sixty-eight percent of the random occurrences represented within the distribution are found within one "standard deviation" of its center or mean; 95% are within two standard deviations of the mean; and about 99.7% are within three standard deviations. Roughly this means, for example, that there is a 95% probability that the value of any particular random item – such as the percentage of black employees in a given nondiscriminatory workforce – will lie within two standard deviations of the central value (the average percentage of black employees in all nondiscriminatory workforces – which is often assumed to be the same as the percentage of blacks in the overall population). Conversely, if the racial composition of a given workforce is more than two standard deviations from the mean percentage, a normal distribution tells us that there is only a five-percent chance that this occurred by random chance – thus creating the inference that it was an intentional act of racial discrimination.

But, assumptions of normality may not always be warranted in employment discrimination cases. As examples, some cases have held that in making hiring and promotion decisions based on exam scores, state employers (to afford more opportunities for minority applicants/employees, and to avoid possible discrimination claims) may treat test scores lying within two standard deviations of each other as equivalent (since there is not as much as a 95% probability that their difference cannot be explained by random variation).[14] And one court even *required* that there be a 95% probability that a difference in exam scores could not be a result of random variation, or the scores could *not* be treated as different by the employer.[15] These decisions were made despite the fact that the relevant test scores did *not* fall into a normal distribution and actually tended to be

---

[13] Daniel A. Farber, *When the Court Has a Party, How Many "Friends" Show Up? A Note on the Statistical Distribution of Amicus Brief Filings*, 24 CONST. COMMENT. 19, 23 (2007) (quoting M.G. BULMER, PRINCIPLES OF STATISTICS 109 (2d ed. 1967, 1979 corrected reprint)). The above statement refers to the "normal distribution." In this study, we will refer to a host of other probability distributions. The focus of our study is not on these other distributions, but on an underlying, unifying distribution which will be discussed at greater length. However, the reader interested in further study of the other distributions mentioned in this study is referred to the sources cited in Notes 113 and 114.

[14] San Francisco Fire Fighters Local 798 v. City and County of San Francisco, 133 P.3d 1028, 1041 (Cal. 2006) (footnote omitted) (The California Supreme Court approved a system in which "the difference between two [examination] scores [must be] large enough so that there is a 95 percent confidence that the difference in scores is a true difference not to be attributed only to random factors. That range of scores then becomes the band within which the promotion applicants are considered to have scored comparably . . . .") and Officers for Justice v. Civil Service Comm'n, 1991 U.S. Dist. LEXIS 12580, *17 (N.D. Cal. Aug. 21, 1991) (Federal district court issued declaratory judgment that the City of San Francisco could use a system of "banding" – treating exam scores within a range of "1.96 multiplied by the 'standard error of measurement'" of each other as equivalent – for purposes of granting promotions within the police department, without violating Title VII's antidiscrimination requirements.).

[15] Guardians Ass'n of New York City v. Civil Serv., 630 F.2d 79, 103 (2d Cir. 1980).

bunched at the high end in what can be described as a "fat tail."[16] At least one Federal District Court has recognized this problem and rejected "normal" statistical analysis as proof that rank ordering of exam scores for employment advancement was discriminatory (and that exam scores within two standard deviations should be treated as equivalent) because the "test results did not come out to a bell shaped curve – it was not a normal distribution – rather, it was skewed in that there were many scoring in the upper limit."[17]

Though courts considering employment discrimination cases may not *always* assume the applicability of a normal distribution, they seem inclined in that direction, even when evidence based on other distributions is presented.[18] It is simply the easiest, most convenient, statistical path. As an example, the United States District Court for the Northern District of Texas, in rendering a complicated judgment in a race and sex discrimination in employment case, considered whether its probability analysis should be based on a hypergeometric distribution, a binomial distribution, or a normal distribution, and concluded that the relevant probabilities could just be "accurately approximated" using a simple normal distribution, which would also supply the benefit "that statistical significance levels may [then] be determined using widely published tables."[19]

The tendency of courts to assume normal distributions when conducting probability analyses may be partly explained by a robotic allegiance to precedent in this unfamiliar "statistical maze." In a racial discrimination case before the United States Court of Appeals for the Fourth Circuit, dealing with the composition of grand juries in Colleton County, South Carolina, the court held that "in all cases involving racial discrimination, the courts of this circuit *must* apply a standard deviation analysis such as that approved by the Supreme Court in *Hazelwood* [*School District v. U.S.*] before drawing conclusions from statistical comparisons."[20] In the same decision, the Fourth Circuit also cited the importance of "following guidelines" stated by the Supreme Court in *Castaneda v. Partida* – a case dealing with racial discrimination in jury selection.[21] In both

---

[16] *Id.* at 103.

[17] Nash v. Consolidated City of Jacksonville, 895 F. Supp. 1536, 1551 n. 7 (M.D. Fla. 1995).

[18] For examples of employment discrimination cases not based on a bare normal distribution, *see* McAlester v. United Airlines, Inc., 851 F.2d 1249, 1258 (10th Cir. 1988) (Court accepted "unrefuted" evidence of racial discrimination in employment based upon "statistically significant results with the binomial probability distribution.") and Rich v. Martin Marietta Corp., 467 F. Supp. 587, 603 (Colo. 1979) (Court accepted evidence of racial discrimination in employment which "establish inferences based upon [statistical methods] . . . described as a 'normal approximation to the binomial' and as a 'hypergeometric distribution.'").

[19] Vuyanich v. Republic National Bank of Dallas, 505 F. Supp. 224, 349 (N.D. Tex. 1980).

[20] Moultrie v. Martin, 690 F.2d 1078, 1082 (4th Cir. 1982) (emphasis added) ("The reason for our holding is quite apparent. When a litigant seeks to prove his point exclusively through the use of statistics, he is borrowing the principles of another discipline, mathematics, and applying these principles to the law. In borrowing from another discipline, a litigant cannot be selective in which principles are applied. He must employ a standard mathematical analysis. Any other requirement defies logic to the point of being unjust.") (referring to Hazelwood School District v. U.S., 433 U.S. 299 (1977).

[21] Castaneda v. Partida, 430 U.S. 482, 497 n. 17 (1977). Not all courts interpret the Castaneda precedent so strictly. As an example, the United States Court of Appeals for the Eighth Circuit conducted its own statistical calculations to determine if sex discrimination had taken place in the compensation and promotion of female faculty at St. Cloud State University in Minnesota. Along the way, the court noted that "If a legal rule of analysis can properly be derived from the *Castaneda* footnote, it can only be that standard deviations greater than two or three necessarily exclude chance as a cause of underrepresentation. The converse of this – that standard deviations of not 'more than two or three' necessarily exclude discriminatory design as the cause – is nowhere implied. Statistical evidence showing less marked discrepancies will not alone establish that something other than chance is causing the result, but we shall consider it in

Hazelwood and *Castaneda*, the Supreme Court assumed a normal distribution and looked for differences of two or three standard deviations to determine if there was a legitimate inference of unlawful discrimination.[22]

Yet, not all judges are ready to blindly accept assumptions of normality, even in the area of jury selection (where the Supreme Court clearly has done so). When the United States Court of Appeals for the Fifth Circuit, sitting en banc, made use of a "cumulative binomial probability distribution . . . based on a Bernoulli process"[23] – which assumes that evaluated events are randomly distributed, much like in a normal distribution – to find racial discrimination in the selection of grand jury foremen in Madison Parish, Louisiana,[24] Circuit Judge Thomas Reavley, in dissent, pointed out:

> [T]here is neither anything random about the process of selecting a grand jury foreman in Louisiana, nor any reason for the majority to indulge in the presupposition that the selection of grand jury foremen should, absent discrimination, produce random results. Louisiana has decided to choose the best possible person from the grand jury venire to serve as foreman. That determination alone refutes the assumption of randomness that underlies a Bernoulli process and disallows an inference of racial discrimination by statistically testing the numbers against the cumulative binomial probability distribution.[25]

In a concurring opinion in the case, Circuit Judge John Brown felt compelled to admit that "the analysis of Judge Reavley in his dissent is unanswerable" and that the majority had relied on "inapplicable statistical improbability."[26]

Beyond employment discrimination and jury selection, American courts have dabbled in probability analysis in a number of other topical areas. In a Georgia voting rights case, dealing with possibly racially discriminatory redistricting plans, the United States District Court for the District of Columbia made use of a statistical analysis of the likelihood of a black candidate being elected, as a function of black voting age population – which assumed a normal distribution – to reject one plan.[27] In accepting a settlement agreement in a mass toxic tort case as fair, the United States District Court for the Eastern District of New York considered the strength of the plaintiff's case in terms of a rough estimate of statistical probabilities.[28] The United States Tax Court decided that an insurance company's loss estimates were reasonable, based on the application of certain unspecified probability distributions by expert witnesses.[29] And, to detect contract fraud by a gas

---

conjunction with all the other relevant evidence in determining whether the discrepancies were due to unlawful discrimination." Craik v. Minn. State Univ. Bd., 731 F.2d 465, 475 n. 13 (8th Cir. 1984) (paragraph break omitted).

[22] *Hazelwood*, 433 U.S. at 311 n. 17 and *Castaneda*, 430 U.S. at 497 n. 17.

[23] Guice v. Fortenberry, 661 F.2d 496, 509 (1981) (Reavley, Cir. J., dissenting).

[24] *Id.* (majority opinion).

[25] *Id.* at 511 (Reavley, Cir. J., dissenting).

[26] *Id.* at 514 (Brown, Cir. J., concurring and dissenting in part).

[27] Georgia v. Ashcroft, 195 F. Supp. 2d 25, 81-82 (D.D.C. 2002).

[28] *In re* "Agent Orange" Product Liability Litig., 597 F. Supp. 740, 836-37 (E.D.N.Y. 1984). *See also In re* Hanford Nuclear Reservation Litig., 1998 U.S. Dist. LEXIS 15028 (W.D Wash. August 21, 1998) (court evaluated the reliability of expert testimony about the probability of injury caused by radiation from the Hanford Nuclear Reservation – but did not consider whether the experts were making us of the correct probability distribution in their calculations).

[29] Utah Medical Insurance Association v. Commissioner, 76 T.C.M. (CCH) 1100 (1998).

station against its wholesale supplier, the United States Court of Appeals for the Seventh Circuit made use of "desktop computers and statistical packages" to perform "a quick analysis on the numbers."[30] The court concluded that "[t]he probability that the observed fluctuations [in the volume of gasoline sold] are attributable to random events such as the weather is minuscule."[31]

Probability theory has also found its place in criminal law – for both determining appropriate sentences and for establishing guilt. The United States District Court for the Eastern District of New York relied on statistical analysis (built on normal assumptions of randomness) to estimate how much heroin a drug trafficker had transported, for the purpose of imposing sentence.[32] The Arizona Court of Appeals held that the mere statistical probability that an inmate was HIV-positive prior to being sentenced, without the condition being known or taken into account when his sentence was imposed, entitled him to resentencing.[33] And, the New Jersey Supreme Court applied probability theory in its evaluation of an entertainment game to determine whether it was a game of chance (and, thus, a form of illegal gambling), or a permissible contest of skill.[34]

But, most of the above examples either do not identify the probability distribution upon which they rely, or they assume that the behaviors they are evaluating ideally fall into a normal distribution. Some of the problems with such an assumption are exposed in a remarkable 1988 decision, written by Judge Frank Easterbrook, for the United States Court of Appeals for the Seventh Circuit. In Easterbrook's majority opinion in *Branion v. Gramly*, the court launched a scathing attack on the application of probability theory in this criminal case.[35] While, among other things, publishing, perhaps, "the world's first appellate court opinion to take the partial derivatives of a function,"[36] the statistically-astute court criticized the probability analysis provided by the defense's expert witnesses in detail.

Despite dramatic circumstantial evidence that implicated Dr. Branion in his wife's brutal murder, the defense argued that it would have been impossible for the doctor to have done it, because of the *probable* driving time between his departure from work and the commission of the crime at his home.[37] The court chastised the defense experts for relying on "a series of assumptions generally favorable to Branion (though implausible individually and collectively)."[38] Of particular concern for the court was the defense's "assum[ption] that the distribution of driving and bruise-forming times is Gaussian (that is, characterized by a normal bell-shaped curve centered on the mean of the distribution)," though "[n]othing [in the presented evidence] suggests a Gaussian

---

[30] Baker v. Amoco Oil Co., 956 F.2d 639, 641 (7th Cir. 1992).

[31] *Id.*

[32] U.S. v. Shonubi, 895 F. Supp. 460, 502-06, 518-24 (E.D.N.Y. 1995).

[33] State of Arizona v. Ellevan, 880 P.2d 139, 140-141 (Ariz. Ct. App. 1994).

[34] O'Brien v. Scott, 89 A.2d 280, 284-285 (N.J. Super. Ct. Ch. Div. 1952).

[35] Branion v. Gramly, 855 F.2d 1256 (7th Cir. 1988).

[36] David H. Kaye, *Practice and Procedure: Statistics for Lawyers, Law for Statistics*, 89 Mich. L. Rev. 1520, 1527 (1991) (reviewing MICHAEL O. FINKELSTEIN & BRUCE LEVIN, STATISTICS FOR LAWYERS (1990)), citing *Branion*, 855 F.2d at 1265-66 & n.7.

[37] *Branion*, 855 F.2d at 1257-58.

[38] *Branion*, 855 F.2d at 1266.

distribution or the absence of skewness."[39] This led the appellate court to conclude that "the statistical argument that the jury was *compelled* to find [Branion] innocent collapses."[40]

It is important to note that, although it expressed specific concerns about the defense's application of statistics, the Seventh Circuit seemed generally to endorse the use of statistical methods in judicial settings in the *Branion* case. After citing the importance of probability theory underlying such things as fingerprint testimony and DNA evidence, the court asserted:

> None of these techniques leads to inaccurate verdicts or calls into question the ability of the jury to make an independent decision. Nothing about the nature of litigation in general, or the criminal process in particular, makes anathema of additional information, whether or not that knowledge has numbers attached. After all, even eyewitnesses are testifying only to probabilities (though they obscure the methods by which they generate those probabilities) – often rather lower probabilities than statistical work insists on.[41]

The court concluded that "[s]tatistical methods, properly employed, have substantial value" in criminal prosecutions.[42]

The over-the-counter derivatives market supplies yet another area where statistical inferences have found a home in the courts. The potential of this market was initially unrealized because of an inability to rationally price trade options.[43] These problems were significantly ameliorated when Fischer Black and Myron Scholes (and separately, Robert C. Merton) created a statistically-sophisticated option-pricing model in 1973.[44] With the development of the Black-Scholes model, new businesses were created in the field of derivatives trading for a variety of option-type investments – such as stock options, commodity futures, and currency futures.[45]

The Black-Scholes model derives the price of an option from five factors: (1) the time until the option expires, (2) the strike price, (3) the value of the underlying stock, (4) the volatility of the price of the underlying stock, and (5) the risk free interest rate. The model makes several assumptions, including that price changes of the underlying stock are lognormally distributed. If

---

[39] *Id.* at 1265.

[40] *Id.* at 1266. David Kaye concluded that "[d]espite a lapse or two in its own calculations, the *Branion* court successfully flagged the major problems" found in the statistical analysis at issue in the case. Kaye, *supra* note 36, at 1543. For a more complete defense, and critique, of Branion's defense experts' assumptions and of the Seventh Circuit's analysis, *see Id.* at 1525-44.

[41] *Branion*, 855 F.2d at 1264-1265.

[42] *Id.* at 1263-64.

[43] Henry T.C. Hu, *Misunderstood Derivatives: The Causes of Informational Failure and the Promise of Regulatory Incrementalism*, 102 YALE L.J. 1457, 1469 (1993) (reviewing PETER L. BERNSTEIN, CAPITAL IDEAS: THE IMPROBABLE ORIGINS OF MODERN WALL STREET (1992)) (footnote omitted) (the lack of accepted mechanism for pricing option created "large obstacles to a bank's engaging in a broad range of derivative transactions. A bank seeking to buy or sell an option would not know what price to pay or to charge; nor would it have had a viable hedging strategy").

[44] Fischer Black and Myron Scholes, *The Pricing of Options and Corporate Liabilities*, 81 J. POL. ECON. 637 (1973). *See also* Hu, *supra* note 43, at 1469-1470 ("[T]heir option pricing model generated an exact theoretical price for the market value of options. Whatever the model's accuracy, it finally provided a rational basis for pricing options").

[45] Wendt v. Wendt, 1998 Conn. Super. LEXIS 1023, at *100 (Mar. 31, 1998) ("This had a profound effect on economics in general and in the growth of financial markets in particular").

these assumptions hold, Black-Scholes can offer "a precise theoretical value for [an] option."[46] The success – and weaknesses – of Black-Scholes have led to the formulation of both related and competing models (such as the Binomial Model) designed to even better assess option prices.[47]

Though an options-pricing model may "appear[] to a layman to be one of the most complicated formulas ever devised by mankind,"[48] courts have embraced the validity of such models in numerous contexts. As examples, the United States Bankruptcy Court for the Northern District of Illinois ruled that a consultant's valuation of a debtor's common stock using such a model was neither negligent nor unreasonable,[49] and the United States District Court for the Southern District of Iowa rejected the notion that a corporate report using Black-Scholes calculations was entirely speculative, and thus fraudulently failed to disclose actual option values in violation of Rule 10b-5 of the Securities Exchange Act.[50]

Options-pricing models have been of particular judicial use when evaluating tax liabilities.[51] Several cases decided by the United States Tax Court can serve as examples.[52] In one, the Tax Court accepted a Black-Scholes valuation as providing sufficient proof that the defendant's options trades were done at prices to ensure they would result in a loss.[53] In another, the court held that an appraiser's use of Black-Scholes to estimate the value of common stock

---

[46] Hu, *supra* note 43, at 1475. *See also* Charles T. Perry, *Option Pricing Theory and the Economic Incentive Analysis of Nonrecourse Acquisition Liabilities*, 12 AM. J. TAX POL'Y 273, 339 (1995) (The model's output is the "the price an option should cost that will statistically compensate for and therefore eliminate the risk (uncertainty of return) inherent in [an] actual financial position and reproduce a given risk-free rate of return based on a given exercise price for the options. The genius of the formula is its ability . . . to determine how much call and put options should cost within the context of a financial position consisting of a riskless combination of assets, liabilities and options." (footnotes omitted)).

[47] *See, e.g.*, John C. Cox, Stephen A. Ross, and Mark Rubinstein, *Option Pricing: A Simplified Approach*, 7 J. FIN. ECON. 229 (1979).

[48] *Wendt*, 1998 Conn. Super. at *97 (footnote omitted).

[49] *In re* Nanovation Technologies, Inc., Chatz v. Bearingpoint, Inc., 364 B.R. 308 (Bankr. N.D. Ill. 2007).

[50] Martino-Catt v. E.I. DuPont de Nemours and Company, 213 F.R.D. 308 (S.D. Iowa 2002). *See* Securities Exchange Act 15 U.S.C. §§ 78j, 17 C.F.R. § 240.10b-5 (1993).

[51] Several models have achieved judicial acceptance in tax cases. *See, e.g.*, Estate of Harriett R. Mellinger v. Commissioner, 112 T.C. 26 (1999) (U.S. Tax Court noted that expert valuations using Black-Scholes, Noreen-Wolfson, or Shelton options-pricing models can provide a range of values from which a court can determine an appropriate value); Custom Chrome, Inc. v. Commissioner, 217 F.3d 1117, 1124 n. 10 (9th Cir. 2000) (the court advised that "[t]he factfinder may adopt any well established and reliable method of determining the value of options . . . such that the factfinder can assess with reasonable accuracy the value of the options with respect to the evidence in the record"); and Menard, Inc. v. Commissioner, 2004 Tax Ct. Memo LEXIS 215 at *64 (Sept. 16, 2004) (the court expressed a preference for using Black-Scholes, because "generally accepted accounting principles support the use of Black-Scholes for valuing stock options. For example, paragraph 19 of SFAS [Statement of Financial Accounting Standards] No. 123 requires for financial reporting purposes that companies use a fair value method of accounting, such as Black-Scholes, to estimate the companies' stock option expenses").

[52] Other courts, beyond the United States Tax Court, have also accepted the use of options-pricing models when deciding tax liability cases. *See, e.g.,* City of Philadelphia v. Tax Review Board of the City of Philadelphia, 901 A.2d 1113 (Pa. Commw. Ct. 2006) (the court ruled that the City of Philadelphia was not prejudiced by tax payers seeking to use Black-Scholes when calculating their local tax refunds, particularly since the city had used the same formula when negotiating with other tax payers).

[53] Andros v. Commissioner, 71 T.C.M. (CCH) 2472 (1996).

warrants was "reasonable."[54] And, most relevantly, in a third case, the court itself applied an options-pricing model to determine that, because an option was purchased for more than 100 times its actual value, it was not a failed investment (and thus a loss for tax purposes), but was actually a component part of a larger transaction.[55]

Courts have also found options-pricing formulas "relevant because they assist . . . in understanding . . . plaintiffs' damage evidence and determining the amount of . . . plaintiffs' damages."[56] In one case, because "[each of the opposing experts] employed Black-Scholes in measuring damages and testified that it was an appropriate method," the Delaware Court of Chancery "adopt[ed] that method" when determining the compensatory value of employee stock options following a merger.[57] In another case, to determine damages for breach of a stock options agreement, the United States District Court for the Southern District of New York rejected the "inherent value reflected in the profits [a party] actually realized when he exercised the options" and instead insisted upon "the Black-Scholes method of option valuation" which can determine "the actual market value of the options as of the time they were granted."[58] Remarkably, the United States Court of Federal Claims has even argued that a party's failure to apply a Black-Scholes options-pricing analysis, in a breach of contract case, confirmed the "court's suspicion" that the party's options were not worth what the party claimed when seeking damages.[59]

Theoretically, options-pricing models could be judicially useful in many other contexts, from formulating divorce settlements,[60] to pricing patents,[61] to calculating a contract's value,[62] to

[54] Hospital Corporation of America v. Commissioner, 72 T.C.M. (CCH) 1581 (1996).

[55] Berry Petroleum Company v. Commissioner, 104 T.C. 584 (1995).

[56] R.A. Mackie & Co., L.P. v. Petrocorp Incorporated, 329 F. Supp. 2d 477, 514 (S.D. N.Y. 2004) (damages awarded because warrant issuer violated "perpetual" warrant agreements when it merged with another corporation, thus effectively forcing the warrant holders to exchange their warrants for stock before the merger). *See also* Dearlove v. Genzyme Transgenics Corporation, 2004 Phila. Ct. Com. Pl. LEXIS 11 (Dec. 28, 2004) (court approves use of options-pricing formula for assessing damages in class action lawsuit).

[57] Lillis v. AT&T Corp., 2007 Del. Ch. LEXIS 102, at *63 n. 91 (July 20, 2007).

[58] Mathias v. Jacobs, 238 F. Supp. 2d 556, 575 (S.D. N.Y. 2002).

[59] Franklin Federal Savings Bank v. United States, 60 Fed. Cl. 55 (2004).

[60] Ayyad v. Rashid, 2003 Wash. App. LEXIS 383 (Wash. Ct. App. Mar. 10, 2003) (affirming a lower court decision in which the judge requested that a party's expert witness perform a Black-Scholes calculation, conforming to conditions set by the judge, to assess the value of jointly-held stock options). *See also* Fisher v. Fisher, 769 A.2d 1165 (Pa. 2001) (Newman, J., concurring) (after the majority appeared to conclude that the intrinsic value method of assessing the value of stock options "is too speculative" Judge Newman suggested that "There are other more sound and sophisticated ways to value stock options [including Black-Scholes], and if expert testimony is provided to support valuations, it is within the province of the trial court as fact finder to accept or reject such testimony.").

[61] F. Russell Denton and Paul J. Heald, *Random Walks, Non-Cooperative Games, and the Complex Mathematics of Patent Pricing*, 55 RUTGERS L. REV. 1175, 1176-1177 (2003) ("Given that the value of both stock and patented inventions is driven by predictions about future business revenue, the Black-Scholes equation presents a pricing story worthy of further investigation to those interested in pricing patents. Moreover, Black-Scholes may be an especially promising tool because stock options and patents share other important features, such as definite expiration dates and sequentiality of investment moments (the initial purchase of both a stock option and a patent license is followed by later decisions to exercise the option or to develop and market a product).").

[62] Cary Oil Co., Inc. v. MG Refining & Marketing, Inc., 257 F. Supp. 2d 751, 768 (S.D. N.Y. 2003) (expert testimony calculating a contract's value in the same way as a stock option was "not inadmissible for lacking relevancy or not fitting the facts of the case").

valuing employee stock options,[63] and beyond.[64] Legal bargaining itself may even be understandable, and manageable, through the framework of options-pricing analysis.[65]

Yet, despite the many actual and theoretical uses for options-pricing theory in legal settings, some courts have expressed significant concerns about the reasonableness of the

---

[63] Ryan J. Foreman, Comment, *Employee Stock Options in Personal Bankruptcy: Assets or Earnings?* 72 U. CHI. L. REV. 1367, 1388-1389 (2005) ("Assume that financial markets are efficient in the sense that current stock prices . . . incorporate all information about the future value of the stock. This assumption means that it is impossible to systematically 'beat the market': all changes in stock prices are purely random and the expected change in stock prices from one period to the next is zero . . . . [S]o it is impossible for anyone to profit by trading stock on the basis of private (inside) information, even employees of the firm whose stock is at issue. This implies that the debtor and the trustee, when deciding what to do with ESOs, are in the same position as any other market participant. They value ESOs according to the Black-Scholes formula and they have no reason (such as insider information) to depart from the standard 'hold and wait' strategy for optimizing expected option value."). For further discussion of the difficulties and possibility of using option pricing models to value ESOs, *see* Jerome Detemple and Suresh Sundaresan, *Nontraded Asset Valuation with Portfolio Constraints: A Binomial Approach*, 12 REV. FIN. STUD. 835 (1999) (presenting a model indicating that Black-Scholes overvalues ESOs) and Robert C. Merton, *For the Last Time: Stock Options Are an Expense*, HARV. BUS. REV. 62, 65-68 (Mar. 2003) (arguing that the value of ESOs can be accurately estimated using common option-pricing models).

[64] "[T]he original BSOP [Black-Scholes Option-Pricing Theory] insight triggered two further lines of inquiry. One can be called 'options are everywhere;' [because] BSOP rightly understood ought to apply to any contingent claim. A simple classroom exercise is to show students that casualty insurance is a kind of a 'put option,' where the holder has the right to 'put' the asset on the insurance company at a defined price under defined conditions – e.g., if the car is totaled, you get $ 20,000 . . . . More generally, academics and entrepreneurs together have had a merry time exploring hitherto unsuspected 'option aspects' of ordinary deals. This line of inquiry is sometimes shorthanded under the name of 'real options' . . . . Among many other important consequences the study of real options has served to revolutionize the way financial analysts understand ordinary problems of valuation. To take a textbook example: Buying into an (apparently) unprofitable investment may count as buying an 'option to wait and see,' with value in its own right. The other line of inquiry might be called 'everything is an option,' suggesting that if you cannot find an option, you can always make one – constructing a contractual package of rights and liabilities that disaggregate 'the firm' into a series of positive and negative cash flows." John D. Ayer, *Debt Abides: A Prolegomena for Any Future Chapter 11*, 78 AM. BANKR. L.J. 427, 441 (2004) (paragraph breaks omitted).

[65] "At some point . . . it becomes useful to move beyond the construction of abstract models of the litigation process and to apply real option theory to the analysis of actual lawsuits. The challenge in this regard is similar to the challenge observed as real option theory attempts to make the leap from the academic environment to the corporate boardroom. There, practitioners complain that it can be difficult to estimate the variance of the underlying outcomes, that the number of decisions that have to be modeled can be very large, and that the mathematics of solving for options values using formal options methodology, such as the Black-Scholes options-pricing model, can be quite daunting. There are, however, straightforward responses to all of these objections . . . . In particular, litigation is a highly structured process and operates through a well-defined sequence of events. Litigation is, in this respect, better defined than many other investment projects. The likely range of outcomes at each stage of the litigation process is also relatively well defined and is usually bounded in terms of a best and worst possible outcome. Experienced counsel can generally provide reasoned estimates of the distribution of these outcomes at each stage of the process. Indeed, even if counsel lack the experience necessary to generate such estimates, the models can be constructed using the equal ignorance assumption and can be subjected to sensitivity analyses designed to test whether and how various assumptions regarding the model's parameterization influence the lawsuit's potential settlement value." Joseph A. Grundfest & Peter H. Huang, *The Unexpected Value of Litigation: A Real Options Perspective*, 58 STAN. L. REV. 1267, 1326-1327 (2006) (footnotes omitted). *See also* Lucian Arye Bebchuk, *A New Theory Concerning the Credibility and Success of Threats to Sue*, 25 J. LEGAL STUD. 1 (1996); Bradford Cornell, *The Incentive to Sue: An Option-Pricing Approach*, 19 J. LEGAL STUD. 173 (1990); Paul G. Mahoney, *Contract Remedies and Options Pricing*, 24 J. LEGAL STUD. 139 (1995); and Robert J. Rhee, *A Price Theory of Legal Bargaining: An Inquiry into the Selection of Settlement and Litigation under Uncertainty*, 56 Emory L.J. 619, 623 n. 10 (2006).

foundational assumptions of Black-Scholes.[66] Of particular concern is the model's assumption about the volatility of the price of the underlying stock.[67] The Black-Scholes model assumes that a "stock price or value continuously, smoothly, and randomly fluctuates" within a lognormal distribution;[68] but this assumption is often "faulty" and "unrealistic."[69]

> [A]ctual price distributions for assets traded in financial markets typically exhibit "fat tails." The presence of "fat tails" means that higher percentages of actual prices fall within the extreme negative and positive ends of the pricing spectrum than would be predicted under the "normal" distribution assumed in the Black-Scholes model. Indeed, equity markets and other financial markets appear to operate with fatter tails on the left or extreme negative end of the pricing spectrum because "crashes occur more frequently than sudden sharp increases in stock prices."[70]

For these reasons, options traders who rely on Black-Scholes type models are "exposed to the risk of serious losses because catastrophic and near-catastrophic events are likely to occur in financial markets with a significantly higher frequency than the models' assumptions predict."[71]

---

[66] *See, e.g.,* Davis v. Commissioner, 110 T.C. 530 (1998) (court was unwilling to use Black-Scholes when calculating the value of stock gifts, because the model assumes a risk free interest rate) and Wendt v. Wendt, 1998 Conn. Super. LEXIS 1023, (Mar. 31, 1998) ("In markets where [its] assumptions are unreasonable, the Black-Scholes formula may produce misleading results").

[67] *See, e.g.,* Litman v. United States, 78 Fed Cl. 90 (2007) (though the court "generally approve[d]" of the Black-Scholes approach, it acknowledged "weaknesses . . . that give the court pause" (particularly concerning assumptions about normal volatility) and applied certain discounts to the figures generated by the model); Spicer v. Chicago Board Options Exchange, Inc., 1990 U.S. Dist. LEXIS 1478, at *38 n. 19 (N.D. Ill. E.D. 1990) ("for Black-Scholes to be of use in resolving damage questions, we must decide what the proper volatility measure is in some other manner"); and Hutchens Non-Marital Trust v. Commissioner, 1993 Tax Ct. Memo LEXIS 613, *69 (Dec. 16, 1993) (after expressing concern about the volatility variable used for computing Black-Scholes (among other things), the court concluded that "[b]efore we accept the option pricing formula as a means for valuing corporate stock, its proponents must make a better case for its validity").

[68] Perry, *supra* note 46, at 337-338.

[69] Arthur E. Wilmarth, Jr., The Transformation of the U.S. Financial Services Industry, 1975-2000: Competition, Consolidation, and Increased Risks, 2002 U. ILL. L. REV. 215, 344 (2002) ("In addition to the faulty assumptions regarding constant volatility and '[log]normal' distribution of prices, the Black-Scholes option pricing theory depends on [other] unrealistic assumptions regarding market conditions") and Hu, *supra* note 43, at 1478 (citing Black & Scholes, *supra* note 44, at 640) ("As an empirical matter, uncertainty in estimating volatility creates 'wide margins of error' in pricing and hedging . . . . [T]heoretical models, including Black-Scholes, all depend on unrealistic assumptions").

[70] Wilmarth, *supra* note 69, at 343-344 (footnotes omitted) (quoting Katerina Simons, *Model Error*, NEW ENG. ECON. REV., Nov.-Dec. 1997, at 23).

[71] Wilmarth, *supra* note 69, at 344. "Several modifications of the Black-Scholes model have been proposed to deal with the 'fat tails' problem. However, these alternative models involve complexities that make them very difficult, and often impractical, for dealers to apply. As a result, due to the 'simplicity' of the Black-Scholes assumption regarding lognormal distribution, it 'continues to be used despite its frequently proclaimed and well-documented flaws.'" *Id.* at 344 n. 537 (quoting Simons, *supra* note 70, at 23-26). *See also* Hu, *supra* note 43, at 1478-79 ("Even now, most option models are based on an arbitrage assumption that simply does not hold in the real world. Currency options illustrate the problems generated by unrealistic assumptions. The leading model assumes that prices of the underlying currencies follow a 'lognormal' probability distribution, but studies have shown that the historical distribution is different, especially for minor currencies. They have "fat tails." Also, the models assume constant volatility, even though currency volatility changes over time. The fall 1992 European currency crisis further demonstrates that unrealistic assumptions undermine the quality of currency options pricing models. Shortly before the French referen-

And judicial fact finders who use the wrong options-pricing model, built upon the wrong assumptions, are exposed to the possibility of making serious errors in judgment.

The potential for such errors can be illustrated by considering Rule 10b-5 securities fraud class action cases. Because the vast majority of such 10b-5 cases are settled out of court, it is uncommon for these cases to go to trial and very rare for them to reach the appellate level.[72] It is in the initial certifying of the class for such a case that a court exercises what can be described as "lethal force."[73] The standard approach in the class certification exercise is for the plaintiffs to allege some misbehavior on the part of a firm or its auditors and demonstrate that a loss in equity value has been occasioned by the revelation or discovery of that misbehavior to the markets. An "event study" is often performed to identify the stock price change associated with the market receiving notice of the misbehavior and to determine whether the loss in value is "abnormal,"[74] or beyond the bounds of typical daily fluctuations.[75]

---

dum on the Maastricht treaty, none of the usual assumptions about price distribution, transactions costs, and the ability to hedge continuously held true. As a consequence, the models were 'totally inappropriate' for valuing options on European currencies.").

[72] *See* James Bohn & Steven Choi, *Fraud in the New-Issues Market: Empirical Evidence on Securities Class Actions*, 144 U. PA. L. REV. 903, 930-31 (1996) ("The vast majority of the IPO class-action suits eventually settled (ninety-six out of the 123 suits [studied]). Seven class-action suits resulted in a pretrial resolution in favor of the IPO defendants (dismissal, denial of class certification, or summary judgment for the defendant). Three class-action IPOs ended in a trial verdict for the defendant; only one class-action IPO ended in a trial verdict for the plaintiff. Five of the class-action IPOs resulted in bankruptcy. The study failed to ascertain (due to unavailability of information) the resolution of the remaining eleven IPOs.").

[73] Oscar Private Equity Investments v. Allegiance Telecom, Inc., 487 F.3d 261, 262 (2007) (In rejecting a class certification, the court ruled that "Given the lethal force of certifying a class of purchasers of securities . . . we now in fairness insist that such a certification be supported by a showing of loss causation . . . .").

[74] "Abnormal returns are the difference between the expected return on a security, taking into account its volatility, and the actual return. A statistically significant abnormal return may indicate that something fishy is causing the value of the security to deviate from its expected value. For example, abnormal positive returns could be the result of insider trading. Abnormal negative returns could be the result of fraud." Schleicher v. Wendt, 529 F. Supp. 2d 959, 976 n. 12 (S.D. Ind. 2007). *See also In re* Credit Suisse First Boston Corp. (Lantronix, Inc.) Analyst Sec. Litig., 2008 U.S. Dist. LEXIS 14198 at *23-24, *24 n. 14 (Feb. 26, 2008) (quoting the transcript of earlier proceedings before a different judge of the same court) (In decertifying a class in a securities fraud class action, the court observed that "[b]oth parties agree, and this Court finds, that there were no statistically significant abnormal market returns . . . and the Court finds that these returns likely represent 'just random fluctuations in the stock price.'" The court noted that "To prove that a stock was responding to a specific piece of information on a Specific day under the generally accepted event study approach (1) the return Must be abnormal; (2) the abnormal return must be [statistically] significant; and (3) there must not be confounding news.").

[75] "One method increasingly recognized by courts and mandated by some of them is an event study, a statistical method of measuring the effect of a particular event such as a press release, a Form 10-K, or a prospectus, on the price of a company's stock: 'An event study is a statistical regression analysis that examines the effect of an event on a dependent variable, such as a corporation's stock price. This approach assumes that the price and value of the security move together except during days when disclosures of company-specific information influence the price of the stock. The analyst then looks at the days when the stock moves differently than anticipated solely based upon market and industry factors-so-called days of "abnormal returns." The analyst then determines whether those abnormal returns are due to fraud or non-fraud related factors . . . . [E]vent study methodology has been used by financial economists as a tool to measure the effect on market prices from all types of new information relevant to a company's equity valuation.'" *In re* Enron Corp. Securities Litig., 529 F. Supp. 2d 644, 720 (S.D. Tex. 2006) (quoting Jay W. Eisenhoffer, Geoffrey C. Jarvis, and James R. Banko, *Securities Fraud, Stock Price Valuation, and Loss Causation: Toward A Corporate Finance-Based Theory of Loss Causation*, 59 BUS. LAW. 1419, 1425 - 26 (2004)) (citing *In re* Imperial Credit Industries, Inc. Sec. Litig., 252 F. Supp. 2d 1005, 1014-15 (C.D. Cal. 2003); *In re* Northern Telecom Sec. Litig.,

It is at this point that mistakes can occur.[76] Economic or accounting experts often examine the daily returns on the stock and adjust them for effects associated with general stock market or industry movements. The residual returns on the company stock are then examined for statistical outliers, where a Gaussian, or normal, distribution of residuals is typically assumed. Then, if the adjusted "event return" is, say, two or three standard deviations away from what is expected, it is regarded as evidence that the particular misbehavior alleged is responsible for the abnormal return.[77] The problem arises because what is considered abnormal depends upon the underlying distribution of residual returns assumed.

> [W]hether a residual is deemed to be statistically abnormal depends on the probability distribution assumed. For example, a normal distribution assigns little likelihood to the occurrence of extreme values. However, there is substantial evidence that stock prices are not normally distributed and extremes occur quite often, a condition known as "leptokurtosis," as in the case of a Pareto-Levy or Paretian distribution. In fact, the arrival of information may not occur randomly, reaction to information may not occur in a consistent time frame, and old information may have a feed-back effect on new information.[78]

Most standard regression software programs report statistics under the assumption that residual returns are normally, identically, and independently distributed. They also provide statistics that can be used to test whether those assumptions are justified. If the true distribution is not Gaussian (normal), then deviations identified as outliers and their probabilities of occurrence are incorrect.

116 F. Supp. 2d 446, 460, 468 (S.D.N.Y. 2000); *In re* Executive Telecard Ltd. Sec. Litig., 979 F. Supp. 1021, 1025-26 (S.D.N.Y. 1997); and *In re* Oracle Sec. Litig., 829 F. Supp. 1176, 1181 (N.D. Cal. 1993). *See also In re* Williams Securities Litig. 496 F. Supp. 2d 1195, 1272-1273 (quoting *In re* Imperial Credit Industries, Inc. Securities Litig., 252 F. Supp. 2d 1005, 1015 (C.D.Cal. 2003)) ("An event study can play a pivotal role in a securities fraud case: 'Because of the need to distinguish between the fraud-related and non-fraud related influences of [sic] the stock's price behavior, a number of courts have rejected or refused to admit into evidence damages reports or testimony by damages experts in securities cases which fail to include event studies or something similar.'"). For an excellent synopsis of the event study approach and methodology, *see* A. Craig MacKinlay, Event Studies in Economics and Finance, 35 Journal of Economic Literature 13, 14-16 (1997).

[76] It appears that not all courts agree that such mistakes should be pointed out. *See, e.g.,* Fogarazzo v. Lehman Brothers, Inc., 232 F.R.D. 176, 190 (S.D.N.Y. 2005) (citing *In re* Initial Public Offering Sec. Litig., 227 F.R.D. 65, 111 (S.D.N.Y. 2004), wherein the court asserted that "class certification is not the time to wage a battle of the experts; nor is it a time to engage in the weighing of contradictory facts."

[77] *But, see In re* Dura Pharmaceuticals, Inc., 2001 U.S. Dist. LEXIS 25907, at *23 n. 1 (S.D. Cal. Nov. 2, 2001), *rev'd on other grounds*, Brouda v. Dura Pharmaceuticals, Inc., 339 F.3d 933 (9th Cir. 2003) (the court, in a Rule 10b-5 case, refused to apply the Black-Scholes model (even when urged by the Securities and Exchange commission to do so), because, it claimed, even if the model can identify suspicious trades, its capacity to "establish scienter is questionable because it does not consider . . . the individual circumstances of the insider." *See also In re* Apple Computer, Inc., 243 F. Supp. 2d 1012, 1028 (N.D. Cal., 2002) ("this Court is not aware of any other court that has used the Black-Scholes model [alone] to establish scienter in a securities fraud case").

[78] Andrew R. Simmons, Kenneth A. Sagat, & Joshua Ronen, *Dealing with Anomalies, Confusion and Contradiction in Fraud on the Market Securities Class Actions*, 81 KY. L.J. 123, 146-147 (1993) (paragraph break and footnotes omitted) (citing David M. Cutler et al., *What Moves Stock Prices?,* 15 J. PORTFOLIO MGMT., Spring 1989, at 4, 9; Richard Roll, R^2, 43 J. FIN. 541, 561 (1988)).

A careful econometric analysis will not make this mistake, but unfortunately it is often overlooked.[79]

As illustrated by the many examples discussed above, the consideration of non normal distributions can also have dispositive importance in a host of other areas beyond options pricing – from employment discrimination cases to criminal prosecutions, etc.[80] But, asking courts to make correct decisions about what probability distribution should be applied in a given case – much less performing the correct statistical calculations – is easier advised than done. As David Kaye has explained: "Probability theory remains an arcane and dangerous tool for lawyers and courts . . . . The challenge for the legal system is to develop the knowledge and rules that will tend to produce technically sound and conceptually appropriate mathematical demonstrations, at trial or on appeal."[81]

---

[79] *See, e.g.,* Jonathan Klick & Robert H. Sitkoff, *Agency Costs, Charitable Trusts, and Corporate Control: Evidence from Hershey's Kiss-Off*, 108 COLUM. L. REV. 749, 810 (2008) (In the process of using "standard event study econometric methodology" to isolate abnormal returns in the stock of the Hershey Company, the authors became concerned that the variance in the firm's returns might not be well approximated by a normal distribution, so they used a statistical method called the "Chebyshev inequality" to determine that returns on certain identified dates were "so large that as a mathematical matter they are statistically significant in all possible abnormal return distributions.").

[80] Beyond the many topics discussed above, when mistaken distributional assumptions may wrongly influence the outcome of judicial proceedings, legal analysts should consider similar possibilities in the realms of legislation and regulation. Lawrence Cunningham has pointed out the inapplicability of Gaussian statistical assumptions to a number of policy areas – from stock market regulation, to airline security, to space exploration, to levy construction, to asteroid impact defense, to environmental protection, to earthquake detection, to even the possibility of "large audit firms exiting the industry" – where catastrophic risks lurk among many small, relatively inconsequential, events: "While analysts [typically] use standard tools such as the normal curve in statistical probability distributions, an increasingly large number of systems are known to behave in accordance with other probability distributions that require redefining the nature of the unexpected." Lawrence A. Cunningham, *Too Big to Fail: Moral Hazard in Auditing and the Need to Restructure the Industry Before it Unravels*, 106 COLUM. L. REV. 1698, 1724-1725 (2006). *See also* Eric L. Talley, *Cataclysmic Liability Risk among Big Four Auditors*, 106 COLUM. L. REV. 1641, 1645 (2006) (discussing the significance of "'fat-tail' distributions" for insurance market risks). Legal scholars have also found nonnormal distributions useful for studying topics such as judicial opinion citation patterns, amicus brief citations patterns, and patent citation networks: Daniel A. Farber, *Earthquakes and Tremors in Statutory Interpretation: An Empirical Study of the Dynamics of Interpretation*, 89 MINN. L. REV. 848, 875-876 (2005) ("citation frequencies deviate greatly from the normal distribution. The data reflect leptokurtosis . . . . Viewing appellate opinions as seismic events – large or small legal shifts that resolve stresses between conflicting legal forces while sometimes creating new stresses – may prove to be a fruitful perspective"); Farber, *supra* note 13, at 24 ("We could not expect an exact correspondence between citation data and the normal distribution, if only because the normal distribution requires an infinite domain in both directions while the number of citations to an opinion cannot be a negative number . . . . Leptokurtosis in data has an important implication for decisionmaking. Change data from human institutions have, in comparison to the Gaussian (normal) distribution, an excess of cases in the central peak, an excess of cases in the tails of the distribution, but a paucity of cases in the 'shoulders,' the area between the central peak and the tails. In terms of amicus briefs, the idea would be that most briefs get some average amount of attention from interested groups that results in filings, but there may be a tendency for attention to snowball once a case begins getting attention. The snowballing effect can lead to [large] distribution tails . . . ."); Katherine J. Strandburg, Gabor Csardi, Jan Tobochnik, Peter Erdi, & Laszlo Zalanyi, *Law and the Science of Networks: An Overview and an Application to the "Patent Explosion"*, 21 BERKELEY TECH. L.J. 1293, 1303 (2006) (While using "network science" to analyze patent citation patterns, Strandburg, et al noted that "that many networks encountered in the real world are not homogeneous like the random networks [described by a normal distribution]. Instead, many real world degree distributions are highly skewed, and very broad, with what are called "fat tails.'").

[81] *Id.* at 1544.

In Part II of this article, we provide courts with an important new tool for applying the correct probability distribution to a given legal question.

Part II: String Theory for Courts

In the process of drawing inferences, opinions or conclusions from data, analysts will often assume that the actual data behave according to a particular theoretical construct – a "distributional assumption." Assuming that the data conform to a particular distribution allows analysts to employ powerful mathematical and statistical methods to make clear statements on the likelihood that a particular event (as exhibited in the data) is explained by identified factors or the product of mere chance. To the extent, however, that experts rely on underlying distributional assumptions that differ from the statistical distribution actually exhibited by the data, they will present probability assessments that have little relationship with true likelihoods. This is because probability assessments are inextricably linked to the underlying distribution assumed.

An example of this is the widespread use of the normal distribution (and its derivative distributions). If the data (or an appropriate transformation) follow a normally distributed process, an analyst or expert can make strong statistical statements based solely on estimates of the sample mean (average) and variance (measure of dispersion from the average). They can do this because the mean and variance parameters are sufficient to completely describe the normal distribution. Underlying this, however, is the specific and restrictive nature of the normal distribution. The mean locates the distribution and the variance describes how far the distribution extends around the mean – both are crucial to any statistical evaluation. The assumption of a normal distribution, however, imposes additional structure on the statistical distribution. In particular, the assumption of normality requires that the distribution be symmetric – equal amounts of the distribution are on either side of the mean. Also, the normality assumption embeds a requirement on the proportion of the observations that are contained in the shoulder area (i.e., the shoulder and waist portions of the characteristic bell shape) relative to the proportion contained in the area far away from the mean – the so-called "tails" of the distribution (or the lip and mouth of the bell shape). There is no reason to assume that all real-world data conform to the normal distribution or its derivatives.[82] To the extent, therefore, that an expert inappropriately assumes normality, or any other unsupported distribution for that matter, and derives conclusions based on that assumption, those conclusions are suspect and should receive little, if any, weight in the courts.

We propose a simple probability distribution that subsumes many other distributions through its ability to map the entire space of possible distribution outcomes for their first four moments – mean, variance, skewness (lopsidedness or asymmetry), and kurtosis (fat tails and peakedness).[83] It is, as it were, a "string theory" distribution for courts.[84] By restricting the parameters of

[82] Many real-world data are described well by the normal distribution. In addition, there are transformations of the data that will result in distributions that are sufficiently normal in shape. There is an extensive literature on "Central Limit Theorems" and asymptotic distribution theory that characterize the conditions under which one can use the normal distribution to make valid statistical statements. An interesting historical perspective is provided in ANDERS HALD, A HISTORY OF MATHEMATICAL STATISTICS FROM 1750 TO 1930 303-350 (1998). A simplified modern treatment is found in HENK TIJMS, UNDERSTANDING PROBABILITY: CHANCE RULES IN EVERYDAY LIFE 149-211 (2004).

[83] A probability distribution can be mapped if one knows all of its "statistical moments." Technically, a moment is the expected value of a positive integral power of a random variable – e.g., the i^{th} moment, $M_i = E[X^i]$, where X is a random variable and E is an expected value operator. The first moment is the mean (or average) value of a distribution.

this parent, or "umbrella" distribution to certain values, many distributions are seen to be its proper subset. Other distributions find their full ranges of skewness-kurtosis combinations subsumed by this same parent distribution.

Expert witnesses can minimize invalid inferences about probabilities by relying on the parent distribution, which fits possible combinations of the first four moments of a wide class of probability distributions. Unfortunately, they often derive probability estimates from an assumed distribution that is often unlikely to closely fit the data being examined. The courts need not be relegated to electing between competing distributions, because the parent distribution likely includes all of the competing assumptions under a single, unified set of assumptions. Thereby, the court can draw conclusions based on the best possible statistical fit to the evidence at hand, and commensurately better probability estimates. Even if the underlying data truly fit a normal (i.e., bell-shaped), lognormal, uniform, or a plethora of other continuous distributions, the parent distribution will provide exactly the same probability estimates as its subset distributions. But if there are any deviations in the actual data from a hypothesized distribution, the parent distribution will fit it better and provide more valid probability estimates. Accordingly, in most cases, the court need rely upon only one parent distribution in order to sort through a number of conflicting testimonies.

The Underlying String Theory Distribution

The parent distribution to which we refer, known as the g-and-h distribution, is a functional transformation of the standard normal distribution, and spans the entire range of possible skew-

---

Higher moments are centered about the mean of a distribution. The second moment (about the mean) is its variance, or a measure of its dispersion about the mean. The third moment (about the mean) is a measure of the distribution's skewness, asymmetry or lopsidedness. For example, a distribution with a positive skew is one whose mass is concentrated on the left, but whose tail extends further toward the right than toward the left. A distribution with a negative skew has a longer tail on its left than on its right, but whose mass is concentrated on the right. The fourth moment (about the mean) is a measure of the distribution's kurtosis, or sharpness of its peak and shortness of its tails – i.e., how much of a distribution's mass is located near its mean as opposed to extreme locations. A distribution with high kurtosis has a greater concentration of observations in its tails and near its center, relative to a normal distribution, whereas a distribution with low kurtosis has skinny tails and its mass concentrated in the shoulders surrounding its mean. The first four moments of a continuous probability distribution are generally considered to be its most important defining characteristics. For a more precise mathematical treatment of these statistical moments, *see* CYRUS DERMAN, ET AL., A GUIDE TO PROBABILITY THEORY AND APPLICATION 217-23 (1973).

[84] We borrow the phrase "string theory" from fundamental physics. String theory views nature as a set of one-dimensional extended objects called "strings," rather than the zero-dimensional point particles that form the basic model of particle physics. It is used to "unify" the known natural forces (gravitational, electromagnetic, weak nuclear and strong nuclear) by describing them with the same set of equations. *See* BRIAN GREENE, THE ELEGANT UNIVERSE: SUPERSTRINGS, HIDDEN DIMENSIONS, AND THE QUEST FOR THE ULTIMATE THEORY 15-16 (1999) and BARTON ZWIEBACH, A FIRST COURSE IN STRING THEORY 6-7 (2004). We use the "string theory" moniker here in a limited sense to convey that by use of a more general probability distribution, a unifying paradigm can be constructed that accommodates a large set of seemingly disparate distributions. Very recently, a new family of distributions has been developed, known as the g-and-k distributions, where skewness and kurtosis are estimated independently. However, it is not clear that it adds any more flexibility than g-and-h. As will be shown later (Figure 5), for a g-and-h distribution any combination of skewness and kurtosis is possible. *See* Michele A. Haynes, H. L. MacGillivray and K. L. Mengersen, *Robustness of Ranking and Selection Rules Using Generalized g-and-k Distributions,* 65 J. STAT. PLAN. & INFERENCE 45 (1997). A helpful comparison between the g-and-h and the g-and-k families of distributions is provided in Michele A. Haynes and H. L. MacGillivray, *Numerical Maximum Likelihood Estimation for the g-and-k and Generalized g-and-h Distributions,* 12 STAT. & COMPUTING 57 (2002).

ness-kurtosis combinations – a much wider area in the skewness-kurtosis plane than all other well-known skewed and leptokurtic (i.e., “fat-tailed”) continuous distributions. Introduced in 1977 by the preeminent statistician, John W. Tukey to study asymmetry in income distribution, the g-and-h is a four-parameter distribution, and many well-known distributions can be derived as special cases of it.[85] It performs significantly better than simpler distributions such as normal, lognormal, Burr III, Weibull, and most members of the Pearsonian family of distributions.[86] While the g-and-h distribution has been around since 1977,[87] and has been used in applications ranging from weather to earthquakes to molecular motion, it has only recently been applied to financial markets and operational risk assessments for financial institutions,[88] with significant findings and implications.[89] It also recently met its first judicial test in a financial valuation trial.[90]

---

[85] JOHN W. TUKEY, EXPLORATORY DATA ANALYSIS (1977). The late John Tukey, formerly a Princeton professor of statistics, is considered to be one of the most important contributors to modern statistics, and to have made more original contributions to statistics than anybody else since World War II. Tukey received his Ph.D. at Princeton in mathematics, and later received 7 honorary degrees, including another Ph.D. from Princeton in 1998. He served four consecutive four-year terms as scientific adviser to the president of the United States, received the National Medal of Science in 1973 from President Nixon, and was a member of the National Academy of Sciences and the Royal Society of England. One of Tukey’s major contributions to statistical practice was his clear articulation of the important distinction between exploratory data analysis and confirmatory data analysis, believing that much statistical methodology placed too great an emphasis on the latter. He almost single-handedly developed systematic ways of analyzing data. These ways have been widely accepted by the scientific community. This distinction has particular relevance to the courtroom, where more effort should be made to discover which underlying probability distribution is consistent with the observed data, rather than assume a particular distribution and require a high threshold to reject the assumption. Only by following this approach are any probabilities derived from the distribution likely to have sufficient reliability to inform judicial decisions. For retrospective analyses of Tukey’s life and his contributions to statistical science*, See* George Casella, et al., *Tribute to John W. Tukey*, 18 STAT. SCI. 283 (2003); David R. Brillinger, *John W. Tukey: His Life and Professional Contributions*, 30 ANNALS STAT. 1535 (2002); and THE PRACTICE OF DATA ANALYSIS: ESSAYS IN HONOR OF JOHN W. TUKEY (David R. Brillinger, Luisa Turrin Fernholz, & Stephan Morgenthaler eds., 1997).

[86] Descriptions of each of these distributions are given in the references cited in footnotes 87, 88 and 95, and in Pandu R. Tadikamalla, *A Look at the Burr and Related Distributions*, 48 INT’L STAT. REV. 337 (1980). A compendium of common probability distributions is available online: MICHAEL P. MCLAUGHLIN, REGRESS +: APPENDIX A: A COMPENDIUM OF COMMON PROBABILITY DISTRIBUTIONS (2001), http://www.causascientia.org/math_stat/Dists/Compendium.pdf.

[87] For studies of the properties of the g-andh distribution, *see* David C. Hoaglin, *Summarizing Shape Numerically: The g-and-h Distributions, in* EXPLORING DATA TABLES, TRENDS, AND SHAPES, 461 (Eds. David C. Hoaglin, Frederick Mosteller, & John W. Tukey, 1985); Jorge Martinez & Boris Iglewicz, *Some Properties of the Tukey g and h Family Distributions*, 13 COMM. STAT. – THEORY & METHODS 353 (1984).

[88] *See* Swaminathan G. Badrinath & Sangit Chatterjee, *On Measuring Skewness and Elongation in Common Stock Return Distributions: The Case of the Market Index*, 61 J. OF BUS. 451 (1988); Swaminathan G. Badrinath & Sangit Chatterjee, *A Data-Analytic Look at Skewness and Elongation in Common-Stock-Return Distributions*, 9 J. BUS. ECON. STAT. 223 (1991); Terence C. Mills, *Modelling Skewness and Kurtosis in the London Stock Exchange FT-SE Index Return Distributions*, 44 STATISTICIAN 323 (1995).

[89] For the first use of a g-and-h distribution for estimating financial option values, *see* Kabir K. Dutta & David F. Babbel, *Extracting Probabilistic Information from the Price of Interest Rate Options: Tests of Distributional Assumptions*, 78 J. BUS. 841 (2005). For the first application of a g-and-h distribution to assessing bank operational risk, *see* Kabir K. Dutta, K. and J. Perry, *A Tale of Tails: An Empirical Analysis of Loss Distribution Models for Estimating Operational Risk Capital*, FRB of Boston Working Paper No. 06-13.

[90] During a January 2007 trial by jury in Cambridge, Massachusetts, Professor David F. Babbel became the first to apply the g-and-h parent distribution for the valuation of a complex financial instrument. Wellesley Leasing LLC v. Northern Light Technology, LLC, Middlesex Superior Court, CA No. 02-1507.

Martinez and Iglewicz showed that the g-and-h distribution covers most of the Pearsonian family of distributions up to an adequate approximation, and can also produce a variety of other types of distributions.[91] By choosing the values of *A, B, g,* and *h* (parameters used to transform the standard normal distribution into a g-and-h distribution, which relate to the mean, variance, skewness, and kurtosis of the transformed distribution), they demonstrated how the g-and-h distribution can generate more than twelve distributions.[92] The normal and lognormal distributions are two very important special cases of the g-and-h distribution. This property of the g-and-h distribution makes it a very general leptokurtic distribution (i.e., one capable of having a higher concentration of data around the mean and fatter tails than a normal distribution) for modeling asset returns and for the valuation of options and derivative assets, the precision of which depends on accurate modeling of the underlying asset returns.

In order to study the skewness and kurtosis of a distribution we need to evaluate its moments. Martinez and Iglewicz,[93] and Hoaglin,[94] showed that calculations for the moments of the g-and-h distribution are straightforward. Its ability to accommodate various combinations of moments exceeds even that of the very attractive and flexible Generalized Beta of the Second Kind (GB2) family of distributions.[95]

The g-and-h distribution can accommodate a wide variety of tail behavior – fat, thin, skewed, and lumpy tails. This allows great flexibility in fitting the actual data. Martinez and Iglewicz,[96] and Hoaglin[97] also compared the skewness and kurtosis of the g-and-h distribution with the distributions belonging to the Pearsonian family. The Pearsonian family of distributions can be broadly classified into seven different categories, and all have finite first four moments. Martinez and Iglewicz observed that the g-and-h distribution covered a much wider area in the skewness-kurtosis plane than the distributions in the Pearsonian family.[98]

---

[91] Martinez & Iglewicz, *supra* note 87, at 361-68.

[92] *A, B, g,* and *h* are parameters used to transform the standard normal into a g-and-h distribution and relate to the mean, variance, skewness, and kurtosis of the transformed distribution. For example, when $g = 0$ and $h = 0$, the g-and-h distribution collapses to a normal distribution. When $h = 0$, it collapses to a lognormal distribution. In addition to normal and lognormal distributions, Martinez & Iglewicz showed how to generate uniform, student-t, exponential, double exponential, Cauchy, beta, gamma, Weibull, chi-square, and logistic distributions. *Id.,* at 362-63. We can also derive Burr III as a combination of the beta and gamma distributions.

[93] Martinez & Iglewicz, *supra* note 87, at 354-56.

[94] Hoaglin, *supra* note 87, at 485-87.

[95] James B. McDonald & Yexiao J. Xu, *A Generalization of the Beta Distribution with Applications*, 66 J. ECONOMETRICS 133, 134-38 (1995).

[96] Martinez & Iglewicz, *supra* note 87, at 362.

[97] Hoaglin, *supra* note 87, at 504-07.

[98] Martinez & Iglewicz, *supra* note 87, at 361-62. As discussed above, by constraining the parameters of the g-and-h distribution, some popular continuous distributions can be derived as its proper subset. However, g-and-h can numerically approximate any other continuous distribution. We have not encountered a continuous distribution that could not be approximated by the g-and-h distribution at a very high level of accuracy (say at 99.999%). In that sense we say that every continuous distribution is a subset of the g-and-h distribution. In other words, we can have a g-and-h equivalent representation of any continuous distribution at a very high degree of accuracy. Invoking the vocabulary of measure theory used in statistics, any continuous distribution can be approximated by the g-and-h almost everywhere – i.e., the measure of that space where g-and-h and another continuous distribution do not agree is zero.

In Figures 1, 2, 3, 4 and 5, we depict the coverage of the g-and-h versus better known distributions in the skewness-kurtosis plane.[99] This plane shows the standardized values of skewness and kurtosis for different distributions.[100] Figure 1 identifies all combinations of skewness and kurtosis that fall within the feasible range.[101] Figure 2 identifies the skewness-kurtosis combinations that are featured in several important distributions used in financial economics. For example, a normal distribution is symmetric and hence exhibits no skewness; however, it exhibits a kurtosis of 3.0. A uniform distribution is also symmetric, but has a lower kurtosis of 1.8, while a logistic distribution has a higher kurtosis of 4.0. The exponential distribution exhibits a skewness index of 4 and a kurtosis of 9. The triangular distribution features a kurtosis index of 2.4 but has skewness that ranges from 0 to 0.32, depending on how the triangle is shaped. Hence, its combinations are mapped by a short horizontal line segment rather than a single point. Note that if the true underlying distribution of a set of data does not match these particular combinations of skewness and kurtosis, the probability estimates derived from these distributions will be incorrect and invalid for statistical inference. This understanding is extremely important for a judicial proceeding in which probabilities of occurrence, and financial values related to probabilities of occurrence, are germane to the outcome of a trial.

---

[99] These charts are an adaptation from Dutta & Perry *supra* note 89, at 20, updated by Ricki Sears. For a technical explanation of the process for calculating and comparing the moments of distributions, *see* Appendix: Skewness and Kurtosis Calculations, *infra* pp. 28-31.

[100] Technically speaking, the value scale for the horizontal axis is the square of the standardized values of skewness, given as M_3^2 / M_2^3, and for the vertical scale, the standardized values of kurtosis. M_4 / M_2^2, where M_i is the i^{th} central moment of the distribution. The g-and-h distribution approaches, but does not reach the impossible region of the lower side unless h is rendered as a polynomial.

[101] Mathematicians have demonstrated that certain combinations of skewness and kurtosis cannot obtain. For example, it is impossible to have skewness (a lopsided tail) if there are no tails at all (kurtosis). All possible combinations of skewness and kurtosis are within the non-shaded area of Figure 1.

FIGURE 1: Skewness – Kurtosis Plane

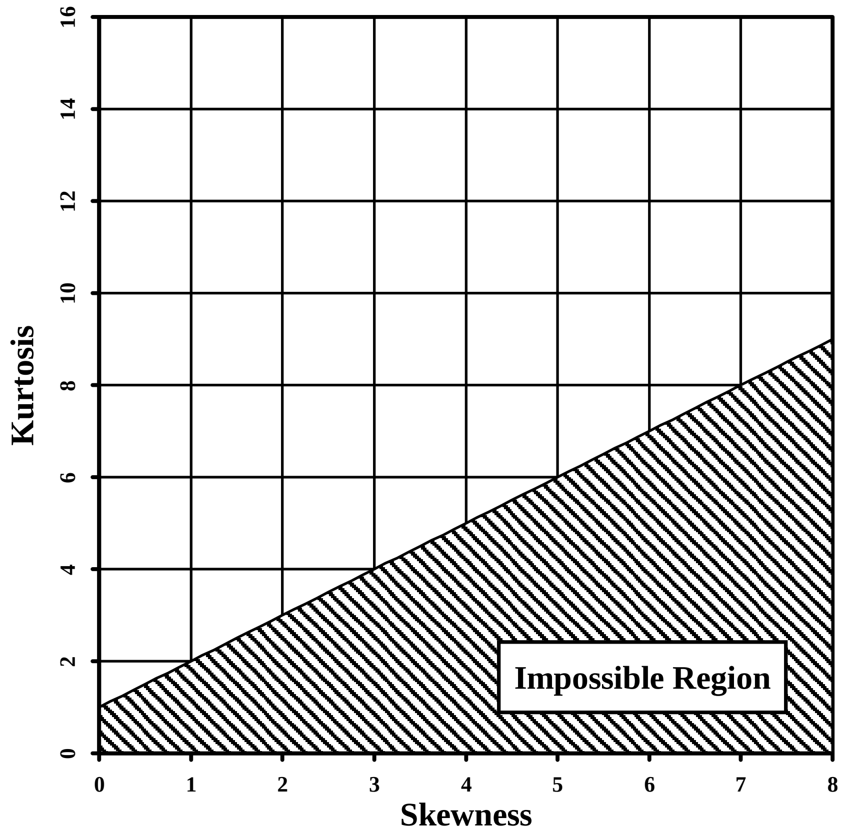
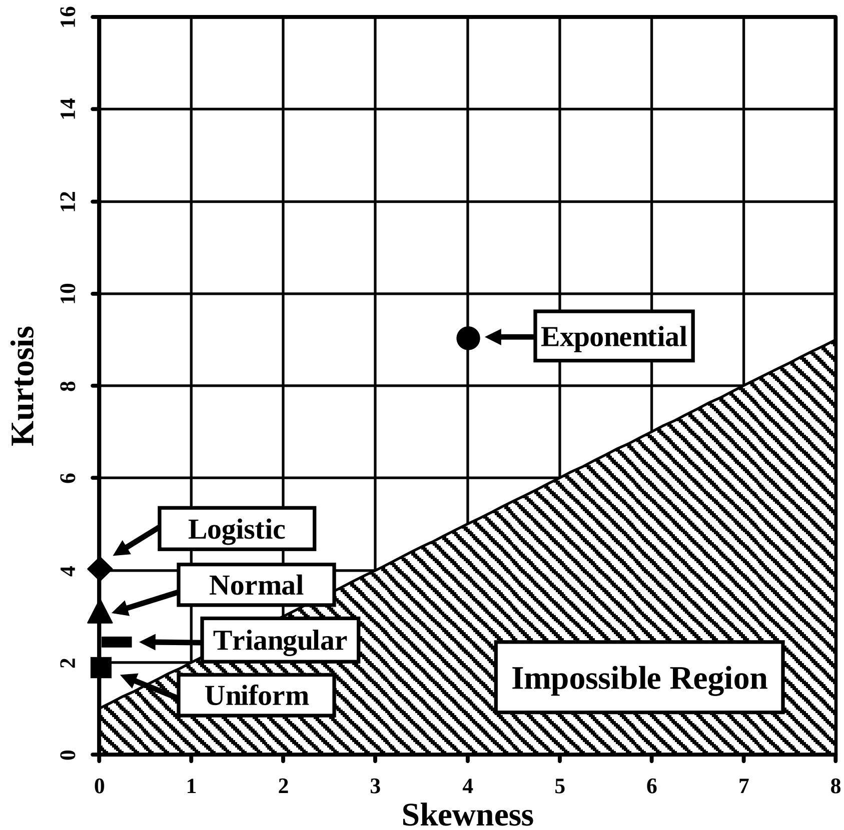


FIGURE 2: Common Distributions on Skewness –Kurtosis Plane



In Figure 3, we present several other important probability distributions that are commonly used in financial economics. Each line or curve in Figure 3 maps the set of skewness-kurtosis combinations that are possible in its respective distribution. For example, the lognormal distribution includes at its initial point (0.0, 3.0) the same combination of skewness (0.0) and kurtosis (3.0) exhibited by the normal distribution; however, it also allows for a set of other skewness-kurtosis combinations, as denoted by the designated curve shown that extends in a northeasterly direction from its initial point. The Generalized Pareto distribution includes, as special cases, the uniform and the exponential distributions. Note that its skewness-kurtosis combinations are mapped by two separate lines.

FIGURE 3: Other Common Distributions on Skewness – Kurtosis Plane

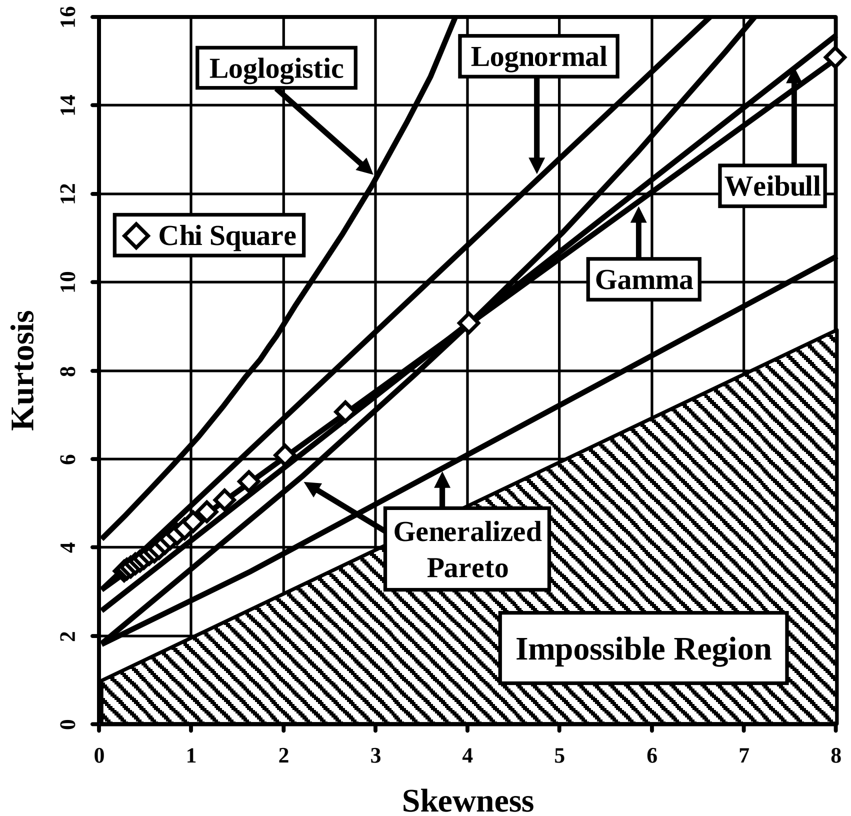


Figure 4 shows how one distribution, known as Generalized Beta of the Second Kind (GB2), covers a much wider range of possibilities than any of the distributions shown in Figures 2 and 3, thereby affording the statistician a good chance for a better fit to the data. Indeed, many prominent statistical distributions are nested within GB2.[102] Accordingly, it is widely used in insurance applications, and has begun to be used in financial economics applications as well.[103]

[102] GB2 is a four parameter distribution that denotes the parameters by a, b, p and q. It collapses to a Burr distribution when p=1, a loglogistic distribution when a=p=1, a generalized Pareto distribution when b=1, and a Pareto distribution when b=p=1. For additional special cases, *see* McDonald & Xu, *supra* note 95 at 134-40. g-and-h can approximate the GB2 distribution at a very high level of accuracy. However, we do not think the converse is true. That is, we can construct a g-and-h distribution for which there is no equivalent GB2 distribution. In that sense, g-and-h is more general than GB2, even though both of them are four parameter distributions. In our experience, we have found that in some situations the tail approximation for the extreme values in the data is better modeled by g-and-h than by GB2.

[103] *See* Richard M. Bookstaber & James B. McDonald, *A General Distribution Describing Security Price Returns*, 60 J. BUS. 401 (1987); McDonald and Xu, *supra* note 95; and J. David Cummins, James B. McDonald, and Craig Merrill, *Risky Loss Distributions and Modeling the Loss Reserve Pay-out Tail*, 3 REV. APPLIED ECON. 1 (2007).

FIGURE 4: GB2 Distribution on Skewness – Kurtosis Plane

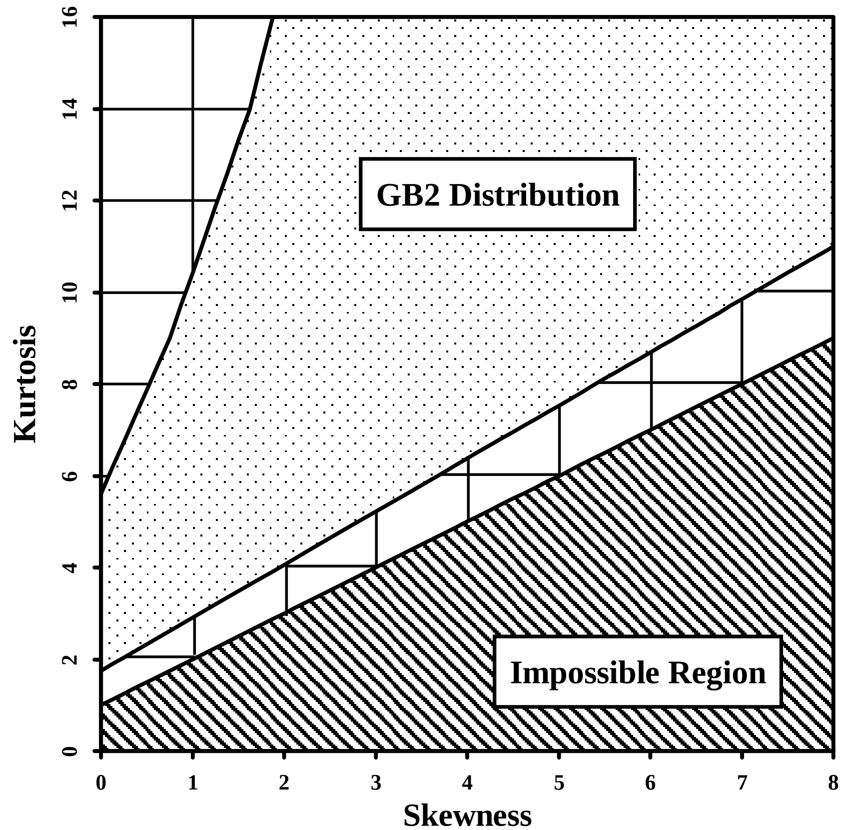
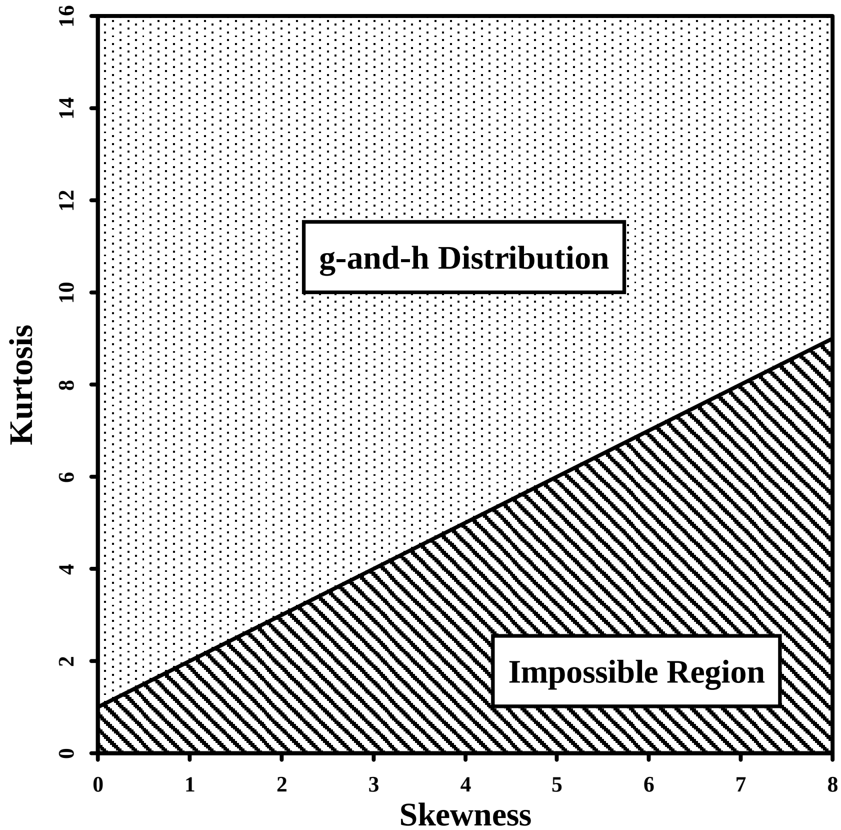


Figure 5 shows that g-and-h covers *all* of the possible combinations of skewness and kurtosis shown earlier in Figure 1; hence, any other continuous distributions can fit the data no better, and often worse, than the fit possible through g-and-h. The parent distribution (g-and-h) not only covers more possibilities than others, but it is easy to work with, provides well-studied, defined probabilities, and is amenable to combining data from disparate distributions without violating the underlying assumptions of each contributing distribution. Therefore, it is a superior general distribution to utilize when making probabilistic inferences about data, and often more reliable than many of the standard distributions used by financial and economic experts in court testimony.

FIGURE 5: g-and-h Distribution on Skewness – Kurtosis Plane



Test Case

In a jury trial last year, a securities case was presented before the Middlesex Superior Court in Cambridge, Massachusetts.[104] The case revolved around the theoretical value of a non-traded stock "put option."[105] In this instance, defendant had acquired some restricted stock whose value was at risk prior to the removal of trading restrictions. Plaintiff, who had a residual interest in the stock, contended that the owner of the restricted stock should have protected his value by purchasing a put option that would allow him to "lock in" a guaranteed minimum price for the restricted stock once it came off restriction. Because there were no traded put options on the particular stock, plaintiff maintained that it was possible to synthetically create the equivalent of a put option by dynamically hedging the instrument, and that the cost of that dynamic hedge could be determined by the hedge ratios implicit in an option pricing formula.[106]

---

[104] Wellesley Leasing LLC v. Northern Light Technology, LLC, et al., Middlesex Superior Court, CA No. 02-1507.

[105] A put option is a financial contract between two parties that obligates the seller of the option to buy the underlying instrument (e.g., stock or commodity) for a certain price if the buyer of the put option exercises the option to buy at that price within the pre-specified time period.

[106] This contention is related to the field of "financial engineering," in which the behavior of certain options is replicated by taking positions in the underlying securities along with a position in debt, and dynamically rebalancing those positions over time. For a primer on this technique, *see* JOHN COX & MARK RUBINSTEIN, OPTIONS MARKETS (1985).

Plaintiff's expert used the standard Black-Scholes put option valuation formula,[107] showing that the option would be worth 28 percent of the value of the underlying traded stock. The Black-Scholes put option pricing formula stands on the assumption that rates of return on the underlying traded stock are distributed normally.

As illustrated in Figure 6, the defense's expert examined the underlying rates of return on the underlying stock and showed that they were not distributed normally. Figure 7 then shows the results of applying a Jarque-Bera (JB) test for normality. The defense's expert showed that there was not one chance in a million that the stock rates of return could be adequately characterized by a normal distribution – indeed, less than one chance in a billion![108] He then demonstrated how a g-and-h distribution fit the data quite well, and applied a put option pricing formula based on a fitted g-and-h distribution.[109] The fitted value for the standardized skewness was only 0.0216 but for standardized kurtosis was 5.2206. To the eye trained in statistics, the point denoting this skewness-kurtosis combination lies well above the loglogistic combinations depicted in Figure 3, indicating that none of the standard distributions shown in Figures 2 and 3 are capable of reproducing the behavior observed in the underlying stock. This skewness-kurtosis combination generated an implied value of the put option equal to 63% of the value of the underlying traded stock.[110] In the case before the jury, if the estimated value of the put option were in excess of 40% of the value of the underlying traded stock, no damages would arise based on plaintiff's theories. The jury found for the defense.[111] The more robust option pricing formula was the key contribution of the case, because it properly accounted for the observed deviations from normality, which were shown to be substantial.[112]

---

[107] Black & Scholes, *supra* note 44.

[108] The data were tested for bell curve fit using the Jarque-Bera (JB) Test. The JB test statistic was 692.61, indicating that the data do not fit a bell curve (i.e., normal distribution). The null hypothesis that this is normally distributed was rejected at the 0.000001 probability level. *See* Carlos M. Jarque and Anil K. Bera, *Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals*, 6 ECON. LETTERS 255 (1980).

[109] *See* Dutta & Babbel, *supra* note 89. (showing evidence that actual traded option prices were much better estimated using a g-and-h option pricing formula than any other of the popular option pricing formulas. In fact, the best alternative formula produced estimated option values that were 700% further removed from actual traded option values than those generated through the application of the g-and-h option pricing model).

[110] To the eye untrained in statistics, the fitted value of 5.2206, which lies near the vertical axis of Figure 2 and between the 4 and 6 vertical markers, may not appear that far out of range of the kurtosis level given by some of the distributions shown there, including the normal distribution, notwithstanding the Jarque-Bera test statistic that demonstrates the distribution of interest is decidedly non-normal. However, from a pricing perspective, this deviation is crucial: the associated price of 63% is 2.25 times higher than the calculated price of 28% that is based on normally distributed returns!

[111] "Modern finance theory accepts several models for determining the net present value of an option grant, Black-Scholes and the binomial model are just two of them. While the differences in total valuation may, on these facts, result in only small differences of value, [the parties to a case are] *entitled to present those differences as a question of fact to a jury*." Lucente v. International Business Machines Corporation, 117 F. Supp. 2d 336, 360 (S.D. N.Y. 2000) (emphasis added).

[112] It is an understatement to say that "one of the tests of a model is how well the theoretical price approximates market prices." Hu, *supra* note 43, at 1501 (citing Kenneth S. Leong, *Model Choice*, RISK, Dec. 1992, at 60-61). Models built on the g-and-h distribution excel in this regard.

FIGURE 6: Comparison of Actual Daily Stock Returns with Bell Curve Data Assumed by Plaintiff's Expert   July 19, 2000 – July 8, 2002

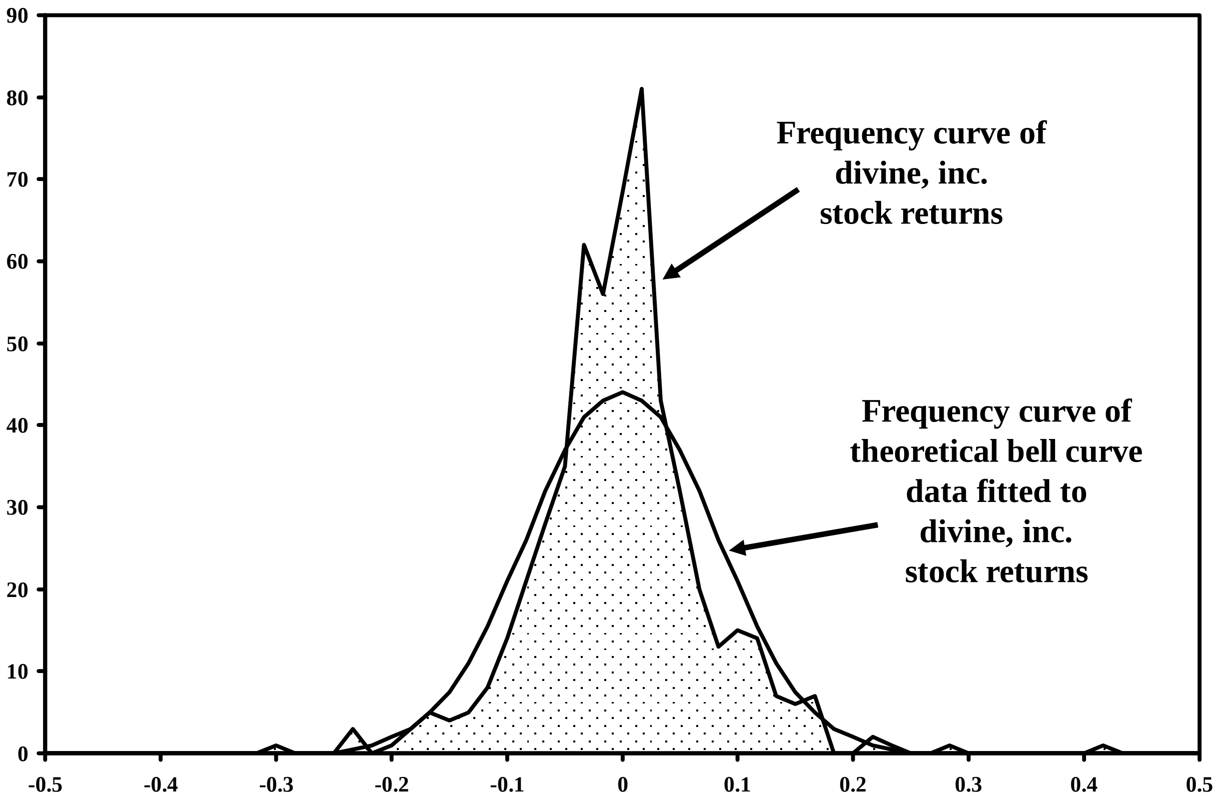
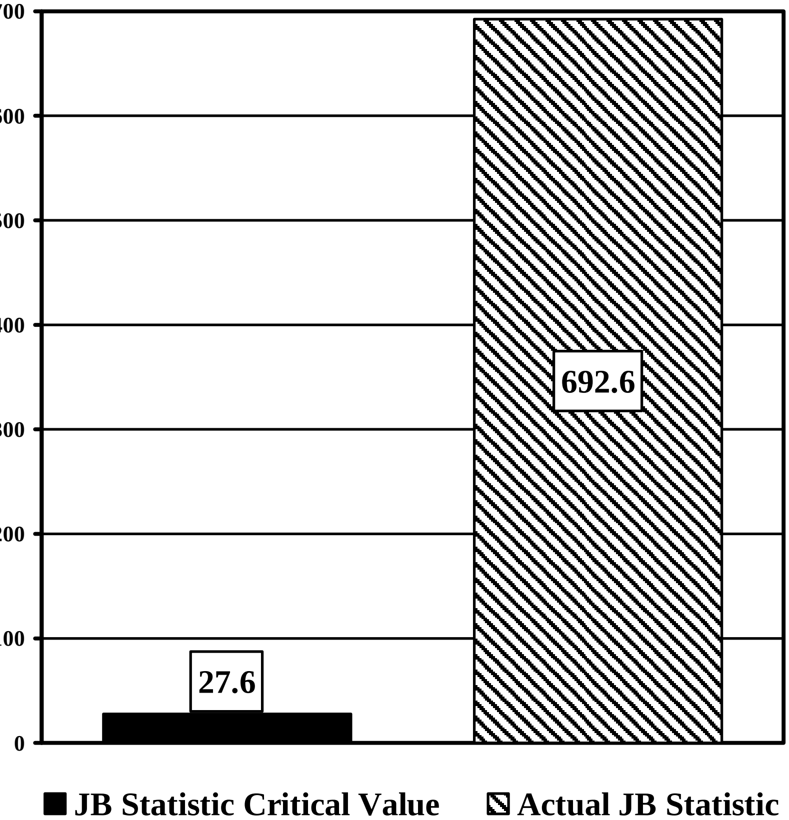


FIGURE 7: Result of the Jarques-Bera Test for Bell Curve Shape



It is noteworthy that the option pricing formula based on the g-and-h distribution would have generated the same value as the Black-Scholes formula if the rates of return on the underlying stock had behaved normally. Accordingly, this case illustrates well how utilizing a general distribution which subsumes more restrictive distributions can clarify the issues, probabilities, and values of concern to a court. It is surprising that the efficacy of g-and-h in financial analysis has not been recognized until relatively recently. Because a g-and-h stochastic (random) process characterizes stock returns and interest rates much better than more restrictive processes, in future legal cases, courts will likely see more valuations, probabilities, and tests based on the g-and-h distribution.

Appendix: Skewness and Kurtosis Calculations

The first four moments of a distribution can be used to define the mean, variance, skewness and kurtosis of that distribution. The k^{th} moment of a random variable X about a fixed point a is the expected value of $(X-a)^k$. We use the notation M'_k to represent the k^{th} moment about the origin, where $\mathrm{M}'_k = \mathrm{E}[X^k]$.

We can find a distribution's moments either through explicit k^{th} moment formulas or through a moment-generating function.[113] The moment-generating function, $\Psi(t)$, is defined as:
$\Psi(t) = \mathrm{E}(e^{tx})$.

If the moment-generating function is differentiable at zero, then its k^{th} derivative at zero, $\Psi^{(k)}(0)$, is equal to $\mathrm{E}(X^k)$. Thus, $\Psi^{(k)}(0) = \mathrm{E}[X^k] = \mathrm{M}'_k$. The mean, μ, is defined as the expected value of X, which equals the k^{th} *moment about the origin* when k equals 1:

$$\mu = \mathrm{M}'_1 = \mathrm{E}[X].$$

We use *central moments*, defined as the k^{th} *moment about the mean*, in the definition for the variance, skewness and kurtosis of a distribution. We use the notation M_k to represent k^{th} central moment.

Central moments can also be defined in terms of moments about the origin:

$$\mathrm{M}_k = \sum_{i=0}^{k} (-1)^k \binom{k}{i} \mathrm{M}'_{k-i}\, \mu^i.$$

The variance of a function is defined as the second central moment, M_2:

$$\sigma^2 = \mathrm{M}_2 = \mathrm{E}\left[(\mathrm{X}-\mu)^2\right].$$

The standard deviation, σ, is the square root of the variance.

The k^{th} *standardized moment* of a probability distribution is M_k/σ^k. Skewness, β_1, is the standardized third moment and Kurtosis, β_2, is the standardized fourth moment.

$$\beta_1 = \mathrm{M}_3/\sigma^3 = \mathrm{M}_3/\mathrm{M}_2^{3/2}$$

$$\beta_2 = \mathrm{M}_4/\sigma^4 = \mathrm{M}_4/\mathrm{M}_2^2$$

Skewness and kurtosis are denoted by β_1 and β_2 because they are also known as the Beta Coefficients. Such standardized moments are often used to compare distributions directly because they are scale invariant – the dimensions cancel, so that they are dimensionless numbers. For example, if one distribution is measured in inches while another one is measured in pounds, they can

[113] For more information on moments and moment-generating functions, *see, e.g.*, NEIL A. WEISS, A COURSE IN PROBABILITY 352-353, 583, 630-638 (2006).

nonetheless be compared with each other in terms of their relative skewness and kurtosis by referring to their standardized moments.

TABLE 1: Moment Formulae for Mean, Variance, Skewness and Kurtosis

MEASUREMENT	FORMULA
Mean	$\mu = \mathrm{M}_1'$
Variance	$\sigma^2 = \mathrm{M}_2$
Standardized Skewness	$\beta_1 = \mathrm{M}_3 / \mathrm{M}_2^{3/2}$
Standardized Kurtosis	$\beta_2 = \mathrm{M}_4 / \mathrm{M}_2^2$

For brevity's sake, we occasionally abbreviate the "gamma function" as $\Gamma(.)$ below. Note that for positive integers, *n*, the gamma function is merely a factorial function, $\Gamma(n) = (n-1)!$, but for a complex number *z* with positive real part, it is defined by $\Gamma(z) = \int_0^\infty t^{z-1}e^{-t}dt$.

TABLE 2: Central Moments

DISTRIBUTION	k^{th} CENTRAL MOMENT (M_k)
Chi-Square	$2^k \dfrac{\Gamma\left(k + \nu/2\right)}{\Gamma\left(\nu/2\right)}$
Exponential	$\dfrac{k!}{\lambda^k}$
g-and-h	$\sum_{i=0}^{k}\binom{k}{i}A^{k-i}B^i \dfrac{\sum_{r=0}^{i}(-1)^r \exp\left[(i-r)^2 g^2 / 2(1-ih)\right]}{(1-ih)^{1/2} g^i}$
Gamma	$\dfrac{b^k\Gamma(c+k)}{\Gamma(c)}$
Generalized Beta of the 2nd Kind	$\dfrac{b^k B\left(p + k/a, q - k/a\right)}{B(p,q)}$
Generalized Pareto	—
Logistic	—
Log-Logistic	$\alpha^k \dfrac{k\pi/\beta}{\sin(k\pi/\beta)}$
Log-Normal	$e^{k\mu + k^2\sigma^2/2}$

Normal	0 (for k is odd) $\frac{\sigma^k k!}{2^{k/2}(k/2)!}$ (for k is even)
Triangular	—
Uniform	$\frac{1}{k+1}\sum_{i=0}^{k} a^i b^{k-i}$
Weibull	see below

Note: A dash indicates that a simplistic formula does not exist.

While no simple explicit formula exists for the moment about the mean of the Weibull distribution, the moments about the origin can be calculated as: $M_k = \eta^k \Gamma\left(1 + k/\beta\right)$.[114]

As confirmed in Table 3 below, the exponential, logistic, normal and uniform distributions are single points on the SK-plot. The chi-square distribution is a series of points on the SK-plot, corresponding to different values of the parameter *v*, which is an integer. As a special case, when *v* = 3 the skewness and kurtosis values of the chi-square distribution are equal to those of the exponential distribution. The chi-square distribution matches points along the gamma distribution and approaches the normal distribution as *v* becomes large.

The gamma, generalized Pareto, log-logistic, log-normal, triangular and Weibull distributions appear as lines or curves within the SK-plot. The generalized Pareto on the SK-plot gives a pair of lines which both run through the uniform distribution (0, 1.8).[115] The upper line of the pair also intersects the Exponential distribution at (4, 9). The log-logistic distribution coincides with the logistic distribution at (0, 4.2). The log-normal and Weibull distributions both approach the normal distribution as skewness approaches zero. The triangular distribution has a fixed kurtosis of 2.4 and a squared-skewness value which ranges from 0 to 0.32.

Two distributions appear as areas on the SK-plot – the generalized beta of the second kind (GB2) and the g-and-h. The GB2 covers a smaller area than the g-and-h, having no skewness values in the upper left area of the SK-plot starting just below a kurtosis value of 6. The g-and-h distribution approaches the impossible region and covers the skewness-kurtosis combinations of all other distributions on the plot.[116]

---

[114] MERRAN EVANS, NICHOLAS HASTINGS & BRIAN PEACOCK, STATISTICAL DISTRIBUTIONS 192-203 (3d ed. 2000).

[115] *See* Jonathan Richard Morley Hosking & James R. Wallis, *Parameter and Quantile Estimation for the Generalized Pareto Distribution*, 29 TECHNOMETRICS 3 (1987).

[116] *See* Dutta & Perry, *supra* note 89, and Kabir K. Dutta & David F. Babbel, *On Measuring Skewness and Kurtosis in Short Rate Distributions: The Case of the US Dollar London Inter Bank Offer Rates*," (Wharton Financial Institutions Center, Working Paper No. 02-25, 2002).

TABLE 3: Selected Values for Skewness and Kurtosis

DISTRIBUTION	TYPE	β_1 (SKEWNESS)*	β_2 (KURTOSIS)
Chi-Square[1]	Series of Points	$\sqrt{8/v}$	$3 + 12/v$
Exponential[4]	Point	2	9
g-and-h	Area	Lengthy	Lengthy
Gamma[1]	Line	$2/\sqrt{c}$	$3 + 6/c$
Generalized Beta of the 2nd Kind	Area	Lengthy	Lengthy
Generalized Pareto	Two Lines	$\frac{2(1-k)\sqrt{(1+2k)}}{(1+3k)}$	$\frac{3(1+2k)(3-k+2k^2)}{(1+3k)(1+4k)}$
Logistic[1]	Point	0	4.2
Log-Logistic	Line	Lengthy	Lengthy
Log-Normal[1]	Line	$(\omega+2)\sqrt{(\omega-1)}$	$(\omega^4 + 2\omega^3 + 3\omega^2 - 3)$
Normal[1]	Point	0	3
Triangular[4]	Line	$\frac{\sqrt{2}(a+b-2c)(2a-b-c)(a-2b+c)}{5(a^2+b^2+c^2-ab-ac-bc)^{3/2}}$	2.4
Uniform[4]	Point	0	1.8
Weibull	Line	$\frac{\Gamma\left(1+3/k\right)\lambda^3 - 3\mu\sigma^2 - \mu^3}{\sigma^3}$	Lengthy

* Note: β_1^2 is presented as our standardized measure of skewness in the skewness-kurtosis plot rather than β_1. Squaring the number eliminates the issue of negativity, and it also enables us to present the combinations of skewness-kurtosis in a more linear fashion, than in the curves that would otherwise clutter the chart. A negative skewness can be achieved by modeling the negative of any distribution with a positive skewness.