MYTHBUSTING--DISCREDITING APPRAISAL MYTHS THROUGH PROPERLY APPLIED STATISTICAL REASONING

When properly applied, statistical reasoning methods, such as regression modeling, can support and verify valuation theories, assertions, and conclusions, or debunk them.

Author: JEFFREY B. WOLPIN

JEFFREY B. WOLPIN, MBA, CGREA, CBA, is the president of Accredited Business Appraisals & Brokerage in Santa Clarita, California. He can be reached at 1-800-VALUATE or jbw-appraiser@earthlink.net.

Statistical analysis is a form of mathematical logic applied to data. It is not unlike the narrative logic that attorneys apply to fact patterns. Statistical analysis is an objective measure of probabilities, which has various steps and mathematical rules that must be followed in order for the analysis and conclusion to be valid.

Statistical analysis of data is in all appraisal work, either implicitly in the appraiser's reasoning or explicitly developed by mathematical formula. Consequently, an appraiser's assumptions and conclusions can be tested through statistical methods and modeling. Many appraisers, however, either distort statistical methods implicitly to arrive at a predetermined valuation conclusion, or feel that their own built-in, subjective, inference-making equipment is up to the task of developing an appraisal conclusion.

Experience, however, suggests that most people are incapable of taking large amounts of data and mentally weighing each bit of relevant information to arrive at a good inference. One merely needs to listen to the wide variance of value opinions for both public securities and the stock market to confirm the problems of processing data. In the business of private company valuation, the problem is even more severe. The variance of value opinions is wider, and industry participants are more prone to promote their assertions by standing on intractable valuation myth, unsubstantiated ideology, or accumulation of their credentials.

Proper use of statistical methods can cut through this nonsense. The downside of statistical analysis, however, is that its techniques, when presented to the unwary, can be manipulated and imply unwarranted conclusions. For instance, most are familiar with the phrase, "how to lie with statistics," coined in the title of the book by Darrell Huff, originally printed nearly 50 years ago and still in print today; it identified some common types of statistical manipulations. [1]

The Fallacy of Inductive Reasoning in Business Valuation

Statistical reasoning at its most fundamental level can be broken down into deductive versus inductive reasoning. Deductive reasoning is the process of drawing a conclusion from a set of previously known facts. If the premises or hypotheses are true, the conclusion must be true. This element of statistics seeks only to describe and analyze a given group without drawing any conclusions or inferences about a larger group.

Inductive reasoning, on the other hand, infers from a sample or small set of observations, conclusions about the larger population as a whole. Such reasoning cannot be absolutely certain and must be stated within probabilities.[2] If the premises are probable or inconclusive, the conclusion cannot be certain. Inductive reasoning can be certain only when all possible instances have been examined and tested. Consequently, inductive reasoning can be valid only if care is taken to insure viability and quality—that the sample is random, and unbiased, and closely resembles the population. Improper inductive reasoning can lead to dogma and self-serving ideology that often becomes the exception instead of the rule.

For example, Sherlock Holmes would deduce from a specific set of facts, say claw and tooth marks on a body found in England's moors near the Estate of Baskerville, that a large mastiff killed Sir Charles Baskerville. He would not, however, conclude that all hounds are responsible. Alternatively, in the 16th Century, based on highly limited observations, it was law and dogma, subject to imprisonment by the Church, that the sun and all planets revolved around the earth, until Galileo proved that Copernican theory was correct. Gross generalizations, over-simplifications, and overly broad rules and dogma derived from a small sample of observations and applied to the total population, are the hallmark of improper inductive reasoning and questionable valuation practice.

Another kind of improper inductive reasoning commonly applied in business valuation is "argument from authority," in which the truth of an assertion is based exclusively on attribution to an individual believed to be a high-influence personality, or entirely on the authority of a credential without factual or evidentiary support. For example, it might be asserted that businesses that are within ten times the revenues of each other are acceptable comparables.[3] The proper use of statistical methods and reasoning, like regression analysis and recognition of non-linear patterns, such as those seen in Exhibits 4, 7, and 13, can easily be used to verify or discredit the assertions of such individuals.

A parallel to the flawed inductive "argument from authority" reasoning are the well known conformance studies[4] recognized in the field of behavioral finance. In the 1950s, social scientists and social psychologists found that individuals join groups as a means to form their identities, and that once joined, most of these individuals succumb to a kind of insular "groupthink."[5] They become incapable of independent and objective examination of factual data, or resistance to group peer pressure, despite the fact that a particular group's opinion and acts may be obviously flawed, invalid, and erroneous[6]

Similarly related studies, show the willingness and desire of some individuals to blindly follow others who are perceived to be of high influence and authority, despite the fact that the latter individuals may be unethical, or just incorrect in their opinions.[7] These two core social dynamic principles are frequently the intrinsic driving force behind many persistent, unproven and discreditable business valuation myths. These include the "Rule of 10," just discussed; the issuance of questionable "master" credentials to sell product and ideology; and the frequent abuse of phrases as "generally accepted," "peer reviewed," and "peer accepted" to sustain and promote flawed business valuation databases, myths, and ideologies.

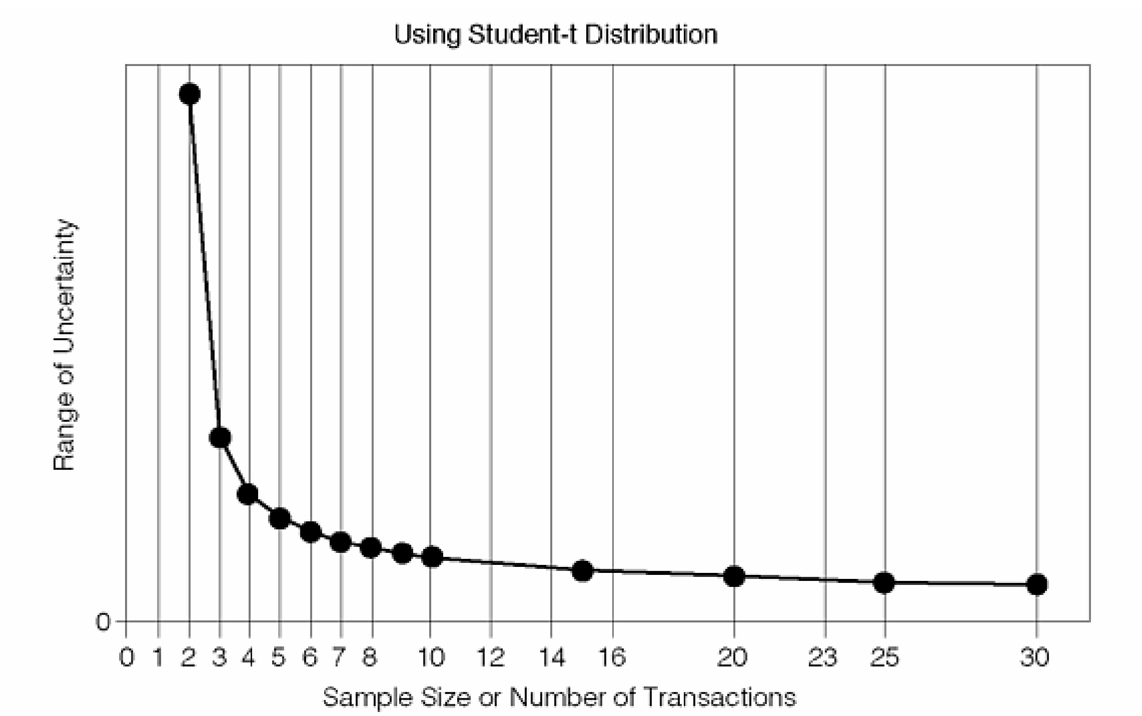
The Direct Market Data Method Myth

One of the more persistently marketed myths based on faulty inductive reasoning is called the "direct market data method." This method's fundamental thesis is that the total market is represented by the Institute of Business Appraiser's (IBA) nonrandomly selected and unverifiable database. [8] However, sound, scientific sampling procedure suggests that the IBA database no more represents the total market than is the total population represented by an opinion poll taken by a political talk show of its listeners. Additionally, the proponents of this method employ the following questionable statistical methods and reasoning:

- Simplistic means and median multiples of sales and earnings, without mentioning necessary measures of reliability for those means and medians, such as confidence intervals and standard deviations.
- Denial of the existence of valuation drivers other than sales and earnings, in other words, acceptance of only those drivers captured by its database.
- Inappropriate outlier elimination, without first verifying whether arbitrarily identified outliers are outliers in fact.
- Small sample sizes of five to seven transactions, which are claimed to be adequate to identify averages and medians.

Regarding the last item, the small sample size argument, Exhibit 1 is a reproduction of a misleading graph produced by the author of the direct market data method, to justify the use of small sample sizes. [9] The case being made is that five to seven transactions are adequate to determine population means because there is only marginal improvement in the range of uncertainty after five to seven transactions. However, this graph violates fundamental graphing principles.

Exhibit 1. Ray Miles Graph - Sample Size vs. Range of Uncertainty of Population Average

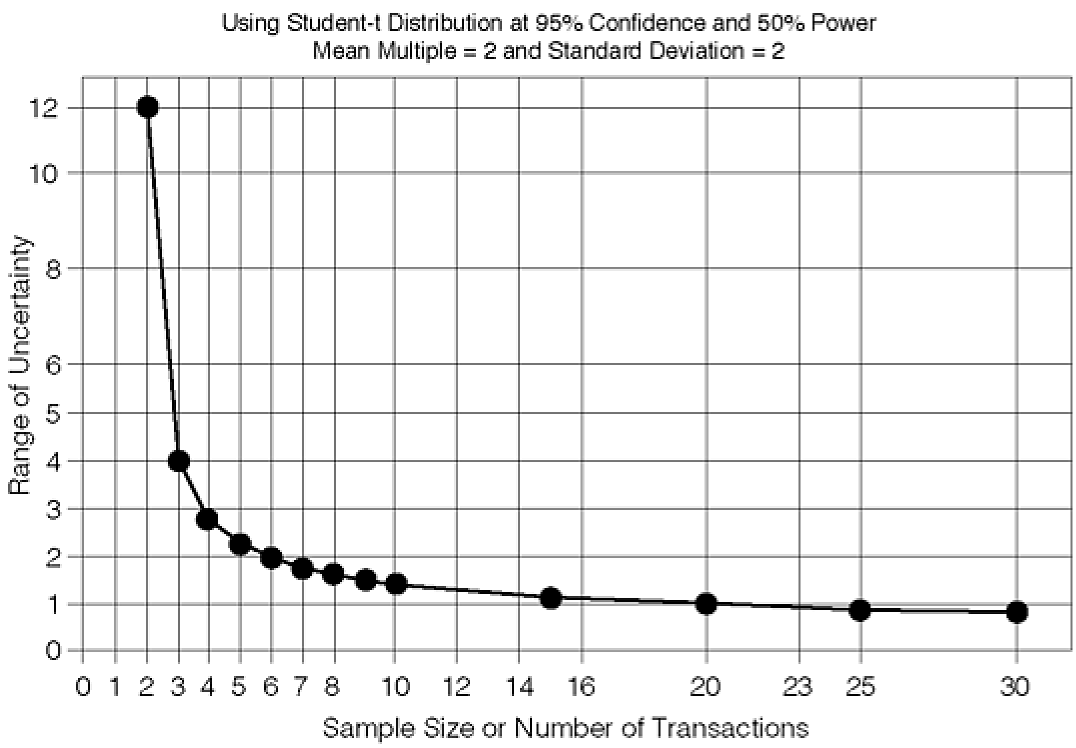


When viewing the graph, one immediately notices two things. First, the vertical scale to the right is ambiguously defined and lacks measurement, [10] and second, this graph is being used to infer its outcome of five to seven transaction sample sizes to all possible populations. This is faulty inductive reasoning.

Regarding the unknown scale on the vertical axis, the real question is what is the percentage margin of error, or the confidence interval, produced by a sample size of five to seven? [11] For example, based on a not uncommon standard deviation of 2 on a mean sales price to seller's discretionary earnings of 2, a sample size of five to seven produces 100% margin of error, where the upper confidence level is 4X and the lower confidence level is 0X. Assuming a properly compiled random sample and sample size calculations, this range is so broad as to provide little guidance on the true multiple. Putting these ranges to actual numbers— say $100,000 in earnings—the average multiple implies an average value of $200,000, whereas the upper end confidence level is $400,000 and a lower end confidence level is $0. Would any client accept this kind of poor accuracy?

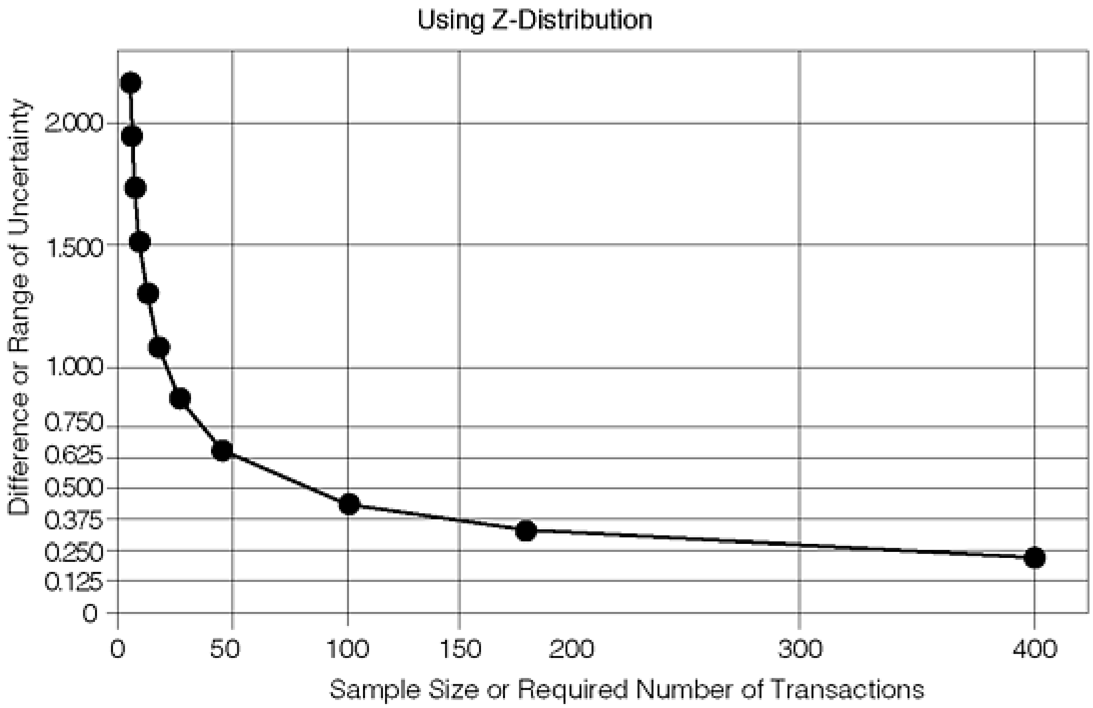
This broad range can be more easily seen in the properly scaled Exhibit 2, in which the range of uncertainty is now represented by the actual difference converted from the percentage margin for error, and in which an actual difference of 2 against a mean of 2X seller's discretionary earnings represents a 100% margin of error.

Exhibit 2. Sample Size vs. Range of Uncertainty to Population Average Including Vertical Scale



Given this graphed relationship in the small business transaction data, the question that should be asked is: What is an acceptable margin of error? If one believes that a 19% margin of error is acceptable, then given a fixed sample or population standard deviation, a considerably larger sample size will be required. A graph reflecting this desired margin of error on the same 2X multiple and 2 standard deviations will look something like Exhibit 3. [12]

Exhibit 3. Sample Size vs. Difference--Using Z-Distribution



Though the basic shape of the graph is the same as the previous graphs, it now indicates that the optimum sample size occurs at around 100; sample sizes larger than 100 produce only minimal improvements in relative uncertainty. The reason this graph indicates a considerably larger sample size requirement is a function of the sample or population standard deviation, which is a fixed variable. At a sample size of 100, the margin of error is approximately 19%, represented by a half-width difference divided by the average or 0.375 ÷ 2 = 19%. [13]

Regression Assumptions and Diagnostic Procedures for Appraisal Analysis

Linear regression analysis is another useful but often misapplied statistical method. Linear regression measures the relationship between one or more predictor variables on a dependent variable. It provides a graphic illustration of this relationship. However, linear regression analysis is often misapplied and misrepresented by appraisers who possess only a limited understanding of the mathematical assumptions required to validate this method. All too often the assumptions required to employ this method properly are unintentionally or intentionally omitted. The assumptions behind regression analysis must be understood and considered in order to evaluate whether a particular regression model properly represents the sample data. The following is a brief overview of some of the assumptions and procedures that must be considered in order to develop a valid regression model.

Basic Regression Statistics—the R-Square, Standard Error, Confidence Intervals, and P-values.

All too often, appraisers focus on one aspect of the regression procedure, the R-Square (R-Sq), to the exclusion of other required elements, as justification for the analysis and conclusions they develop from the regression. Unfortunately, as a sole statistic, R-Sq may be the most misunderstood, misrepresented, and misleading statistic developed by the procedure.

R-Sq is a measure of the strength of the linear relationship between the variables, or alternatively phrased, it is a measure of how much the total variation in the model is explained by the predictor variables (such as revenues) on the dependent variable (such as sales price). R-Sqs will range between 0% and 100%, where 100% represents perfect explanatory power.

However, as a single statistic, it is largely an irrelevant measure without its accompanying measures of standard error, confidence intervals, and p-values. The standard error and confidence intervals are measures of regression model precision. Standard error should be small and confidence intervals narrow. The p-value is the measure of whether a particular variable is a significant factor in the analysis. In a regression, p-values should be preferably less than 5%, and no greater than 10%. For example, R-Sq as a sole statistic is misleading without these other measures, because, for example, it is quite possible to have a high R-Sq with low precision, i.e., a large confidence interval, which indicates a poor regression model; in contrast, a low R-Sq with a small confidence interval can indicate a good regression model.

The Regression Equation.

The regression equation is simply the algebraic equation that computes the average trend line, also known as the line of central tendency. The line of central tendency is another often abused myth, improperly interpreted by many in the business appraisal community as the "most likely" value. However, the line of central tendency has many infirmities, only a few of which are noted herein. Reliance on this trend line and its confidence intervals while ignoring its infirmities is misleading and often results in incorrect value conclusions. Furthermore, there are alternative regression techniques that remedy the many infirmities intrinsic to ordinary least squares trend lines; these are often more appropriate and representative of data trends than are the trends produced by ordinary least squares methodologies.

Margin of Error and the Ratio of the Standard Error Divided by Mean Dependent Variable.

There are two measures of precision of the model. The margin of error is the width of one side of the confidence interval. The standard error measures the displacement of the data observations relative to the trend line. When the standard error is divided by the average of the dependent variable (sales price), this ratio provides a useful measure by which to compare different models. The lower the margin of error and the lower the ratio of the standard error to average dependent variable, the more precise the model.

Regression Coefficients— Slope and Intercept.

The slope is the angle of ascent or descent of the trend line. The intercept is the point on the graph where the trend line begins. Each of these coefficients, and the individual trend line, has probable ranges or confidence intervals (margins of error) around them. These confidence intervals also reflect the precision of the model. Omission of either confidence intervals around the coefficients or the trend line is consequently misleading. By focusing exclusively on whether an R-Sq is a sufficiently high percentage, as many appraisers do, appraisers may be implying a precision in their regression models and conclusions that does not exist.

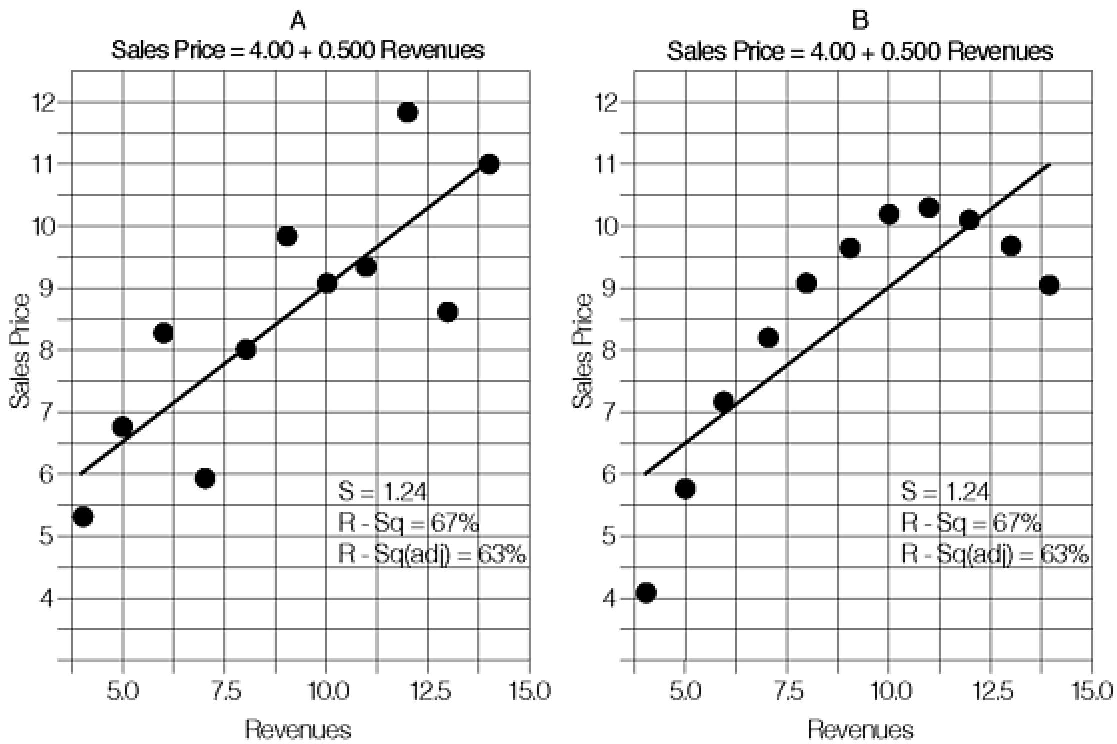
Required Linear Regression Model Assumptions.

One of the most important assumptions required for a linear regression to be valid is the requirement that the variables be linearly related. This is measured by the symmetry of the data points around the trend line. Alternatively stated, the variables must be normally distributed around the trend line (as in the classic bell curve). If there is a normal distribution, the relationship between variables is linear. If there is not a normal distribution, the relationship is not linear. The relationship may be upwardly or downwardly curved, exponentially curved, quadratically curved, cubically curved, etc. If the relationship is curved, the application of a linear regression model or the fitting of a straight trend line to a curved relationship is inappropriate.

This assumption may be tested by certain statistical diagnostic and pattern recognition techniques, such as probability plots and residual vs. fits plots, and the less traditional spline and lowess procedures. [14] The more traditional diagnostic procedures of probability plots and residual vs. fits plots are essentially another way of representing the normal distribution bell curve around the trend line. Interpretation of the probability plot requires that the distribution of the points in the graph fall approximately on a straight line diagonal for the normal distribution assumption to be satisfied. Interpretation of the residuals vs. fits plot requires that the distribution of points in the graph fall evenly and randomly around a center line with no discernable pattern, for the normal distribution assumption or linear relationship to be satisfied. As noted, many appraisers omit this analysis when performing and showing regression graphs, and consequently their linear relationship conclusions may be misleading or incorrect. In other words, one cannot apply a straight line to a curving pattern in the data and expect the straight line to predict that pattern accurately.

An example of how seriously incorrect and misleading conclusions can be reached by simplistically assuming a linear relationship can be seen in Exhibit 4. [15] Though the regression equation, R-Sq, and standard error are all the same, the relationship of the data in each graph is acutely different, as should be the respective conclusions.

Exhibit 4. Fitted Line Plots



Proper Treatment of Outliers, High Influence Points, and Leverage.

Lastly, outliers and high influence points, which frequently occur in sample data and regression models, require careful and proper treatment. Unfortunately, all too frequently they are improperly treated. Outliers are defined as aberrant points relative to the other data points in the sample. High influence points are defined as points that, by their removal, singly or in combination with two or three other points, will cause substantial changes in the regression model's equation. They cause a change in the regression model because they do not have corroborating points or observations in the graph. High influence points are said to have large leverage on the direction and slope of the trend line.

One entrenched self-serving myth in the business valuation industry is that outliers should be eliminated. However, well-grounded statistical procedure calls for the elimination of outliers *only* if they can be verified as outliers in fact, not merely by their apparent aberrant presence on a graph in relation to the other points. [16] Unfortunately, the myth of outlier elimination is often used by database vendors, their proxies, and appraisers to sell their data and opinions in the face of the apparent, frequent, irrational presence of outliers and high influence points, which conflict with their preconceived biases and opinions of value. [17] Statisticians have explained that, "[o]utliers and influential observations should not routinely be deleted or automatically down-weighted because they are not necessarily bad observations. On the contrary, if they are correct, they may be the most informative points in the data." [18]

Improper Regression Model Building, Analysis, and Interpretation

Perhaps an even better way to illustrate proper applications of regression modeling is to observe specific examples of improper applications. Recently, an article was published that serves that purpose. [19] This article contained many of the common myths, misrepresentations, manipulations, and mistakes regarding the application of regression modeling and various other valuation principles.

To reiterate, linear regression analysis is a statistical method that measures the relationship between one or more predictor variables on a dependent variable. In the case of the referenced article, the author used the predictor variable of gross revenues to predict the dependent variable, the sales price of the company, from a commonly available database of small business transactions. [20] The article erroneously justified a conclusion that Segment 1 and Segment 13 (see Exhibit 5) were the best models representing the data. As noted in Exhibit 5, the author based this conclusion exclusively on the high R-Sq for each of these models.

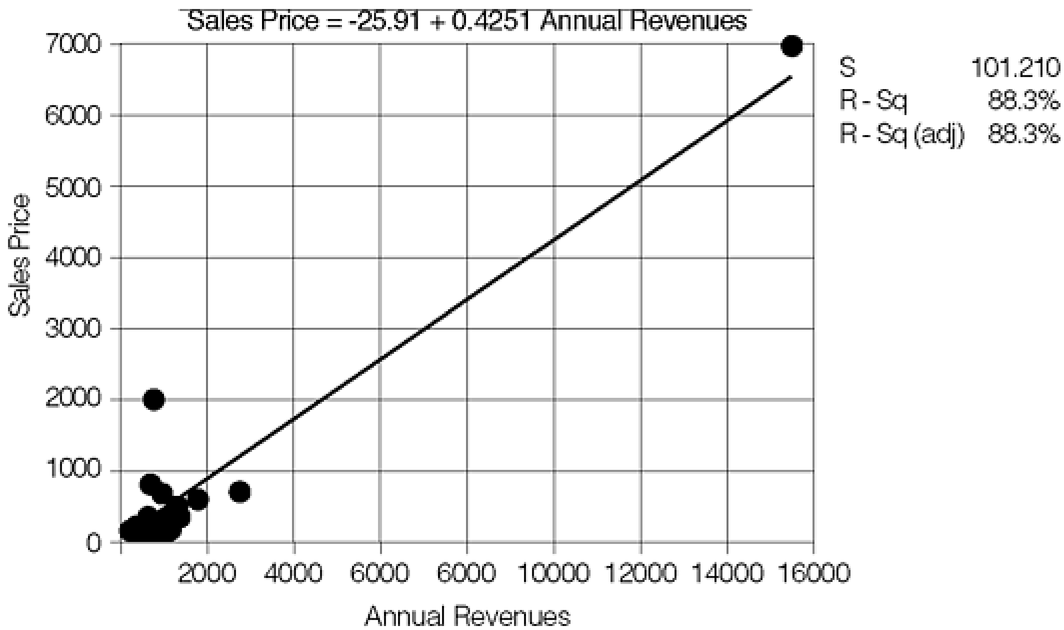
Exhibit 5. SIC Code #7231 Beauty Shops--Sales Price to Gross

Segment	Segment Definition	R-Square	R-Sq. Rank	Sample Size
1	All Transactions	88.0%	2	320
2	Revenues From $0-$100K	25.0%	-	86
3	Revenues From $100-$200K	9.4%	-	102
4	Revenues From $201-$300K	1.5%	-	48
5	Revenues From $0-$300K	29.7%	-	236
6	Revenues From $300-$600K	9.4%	-	64
7	Revenues From $600-$900K	0.5%	-	9
8	Revenues From $900-$1200K	1.2%	-	9
9	Revenues From $0-$600K	40.7%	-	300
10	Revenues From $600-$1200K	5.2%	-	18
11	East Coast Location	46.1%	-	144
12	Not East Coast	70.6%	3	115
13	1983-1989 Sales Date	99.9%	1	50
14	1990-2004 Sales Date	20.6%	-	261
15	2000-2004 Sales Date	18.1%	-	52

This table and graph are reproductions of the referenced article's table and graph. It was reproduced using the same database used by the author except that the database was more recent and had some additional data points. Consequently, the Segment 1 regression in this table produced a slightly different R-Sq.--88% as compared to the author's 91% R-Sq. For purposes of this article, however, these differences are immaterial. Alternatively, the differences may be due to the fact that the author of the referenced article performed his regression analysis in Excel. Excel's statistical program is known to produce numerous calculation errors, and is consequently avoided as a legitimate statistical program by anyone seriously applying statistical methodologies. See Cryer, "Problems Using Microsoft Excel for Statistics," Proceedings of the 2001 Joint Statistical Meetings (2002) or Knusel, "On the Reliability of Microsoft Excel XP for Statistical Purposes, 39 Computational Statistics and Data Analysis 109 (2002).

Exhibit 6 illustrates the questionable statistical analysis and regression model building many appraisers employ in their analysis. [21]

Exhibit 6. Beauty Shops--Revenues Predicting Sale Price--Segment 1--All Transactions



Before getting into the principal discussion of regression analysis and Exhibit 6, however, some fundamental points about the data and inappropriate data manipulation, apparent in Exhibit 5, must be emphasized:

- Because it is usually impossible to measure a total population, samples are taken to reflect the total population. For that sample to be a valid representation of the population, the sample used must be verifiable and randomly collected with a high degree of quality control, free from sampling and nonsampling errors. [22] For the most part, the sample business transaction databases fail to meet these rigorous sampling requirements and, but for a small handful of SIC Codes, are consequently invalid for determining fair market value, from the get-go, despite claims to the contrary by the database owners and promoters. [23] For example, would you trust a TV opinion poll of the viewers of, say, *The O'Reilly Factor* to accurately reflect the opinion of the total population, regardless of how large the poll sample, or how loudly the TV show pounds the table and trumpets its result? [24] Real estate data, on the other hand, tends to be a considerably better sample and truer to the population than small business transaction data. It is highly localized, is verifiable, and frequently represents the entire population within a selected area.
- As a rule of thumb, a minimum sample size of about 50 is required for a regression analysis to be valid. [25] The smaller the sample size, the higher the probability that the sample does not represent the population. Consequently, each time data is segmented and the sample size reduced, the imprecision of one's analysis and conclusions increases. As noted previously, contrary to database vendor claims that samples as small as five or seven are adequate to determine market multiples, [26] required sample sizes are directly related to an estimate of the population standard deviation and the desired precision. [27]
- Many appraisers, either intentionally or unintentionally, group or segment data in an attempt to make sense of the data's complexity, or to find some rationalization for a preconceived conclusion. Such segmenting is often subjective, biased, or

arbitrary, and lacking in factual evidentiary foundation. Therefore, it is meaningless. This is what is occurring in Exhibit 5, where following Segment 1, which contains all of the transactions, the author segments the data into groups from which he justifies various conclusion about those groups. The author, however, provided no explanation for the data segmentation. A reader of this type of segmentation should be skeptical. For example, the data is segmented first by company size as determined by revenues, second by location (east coast vs. not east coast), and third by various time periods. Some very brief comments on each of these segmentations follows:

(1.) *Segmenting business transactions by company size.* Segmenting companies by revenue size may indeed be appropriate. Small companies do not sell in the same market as large companies and multiples are different for each. However, the author's approach was arbitrary. There are, in fact, statistical pattern recognition techniques that can provide evidence of how to segment the data when used in conjunction with some basic common sense. The statistical application for identifying potentially relevant groups can be accomplished through the use of spline fitting or lowess curves. [28] A graphic example of lowess curves applied to a different business and year 2000 data set is provided in Exhibit 7. As can be seen, the lowess curve identifies three bends and three clusters of observations, each of which have different regression equations with different slopes (or angles of ascent). Depending on external factual evidence, this graph may be more representative of the market than one with a single trend line incorporating all of the observations or one in which groups are arbitrarily identified.
(2.) *Segmenting small business transactions by location.* Segmenting by location, east vs. not east coast, simply lacks common sense and is likewise arbitrary and meaningless. The conclusion drawn from the segmentation is that location is not a factor on value. However, the east coast vs. not east coast differentiation is so broad a distinction that it tells one absolutely nothing about the effects of location on a company. Small businesses, like real estate, compete and sell locally, sometimes within a radius of a mile or less. Gas stations and car wash businesses, for example, are known to get a premium price depending on, among other things, the quality of the traffic count, traffic capture, and even what corner of an intersection they are located on. [29] Consequently, without knowing the micro-location dynamics, one cannot know how much of a business' success or failure is due to location vs. management. Additionally, the low coefficients of determination (R^2), evidenced in properly performed regression analysis, and typically produced by these databases, further confirm that factors other than mere sales and earnings contribute to or detract from value as represented by these databases. [30] Therefore, contrary to the claims of the database vendors, it does not follow that location is irrelevant to business value simply because the databases are of insufficient quality to identify these local contributory factors.
(3.) *Segmenting small business transactions by time.* Exhibit 5 indicates two time factor conclusions, (1) that time was a factor in value during the period 1983 to 1989 and (2) that it was not a factor in the other periods. However, here again, these groupings are arbitrary, and any conclusion from them is consequently meaningless. In fact, the conclusion that time is a factor, which the author derives for the 1983-1989 time frame, is based exclusively on one high-influence transaction. (The flaw in basing conclusions on a single, high influence point will be discussed in more detail shortly.) A more relevant grouping would be based on interest rate and financing cycles, which logically would affect the cost of a business purchase. Alternatively, if the data were analyzed monthly or annually, a trend in multiples may become apparent, such as that evidenced in Exhibit 8, based on a different, year 2000, data set. Statistical identification of such a trend

would warrant further investigation into its causes for confirmation. Unfortunately, as with location, database vendors generally promote the proposition that time is not a factor, because such overly broad dogmas promote the sale of their databases, which do not recognize potential time factors.

Exhibit 7. Hardware Stores Scatterplot of Revenue Predicting Sales Price

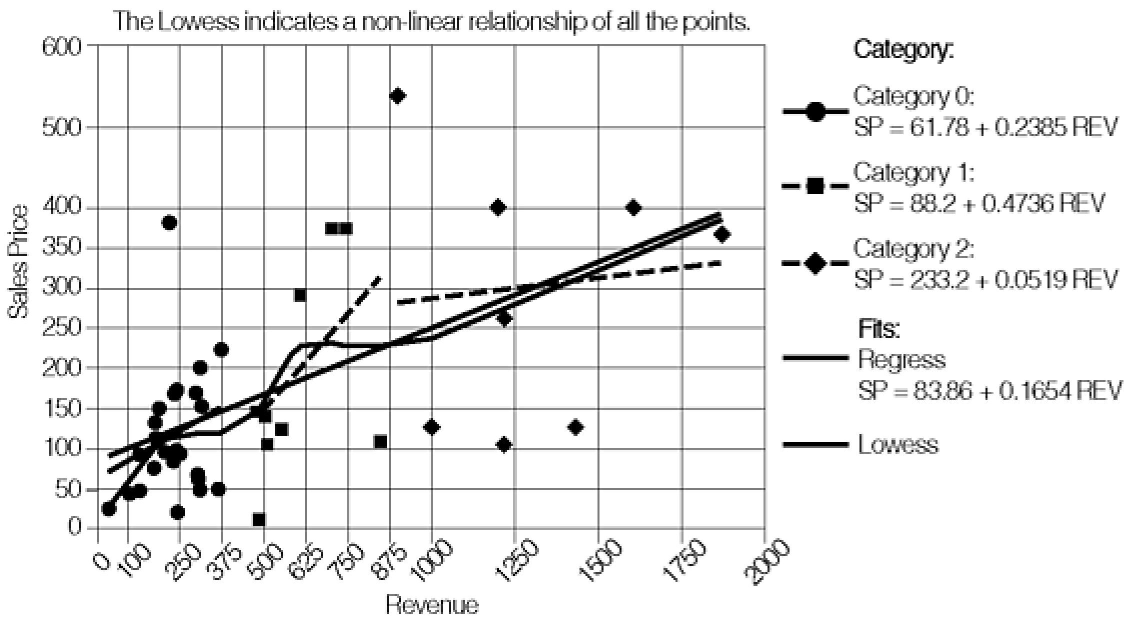
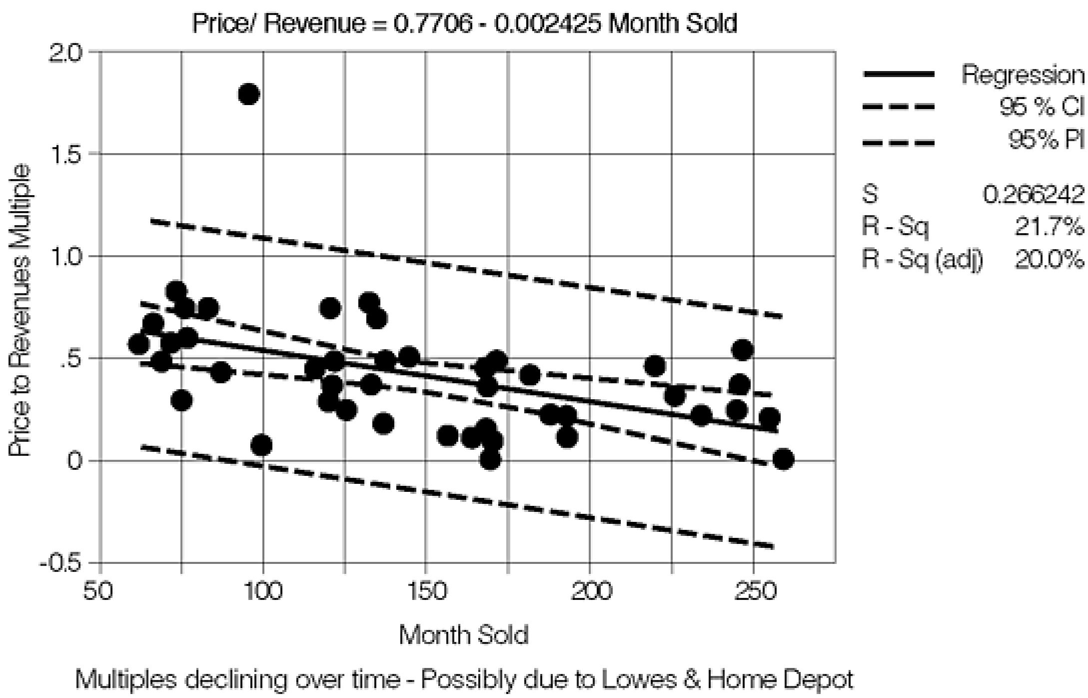


Exhibit 8. Hardware Store--Month Sold Predicting Price to Revenue Multiple



Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	0.88370	0.883701	12.47	0.001
Error	45	3.18981	0.070885		
Total	46	4.07351			

Proper Regression Model Building, Analysis, And Interpretation

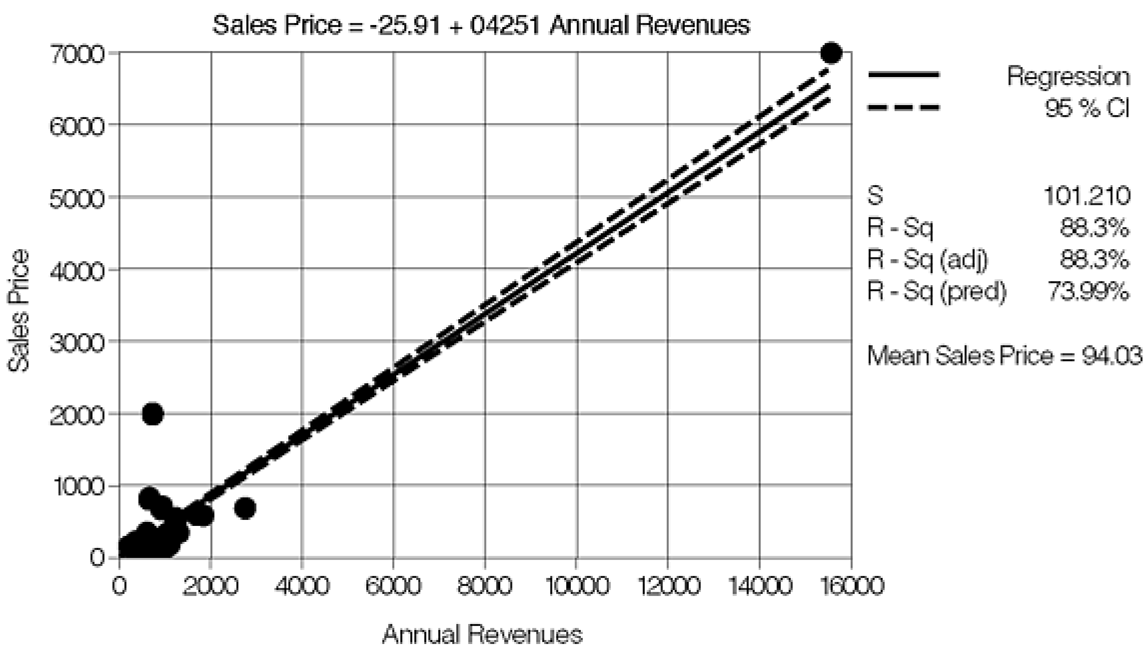
Having established various regression definitions, and having made preliminary observations and a critique of the data and data manipulation, this article next examines Model 1 from the referenced article (Exhibit 6) against the way a regression model should be properly developed.

Linear Regression Model 1

Model 1 is an improper graph and a model frequently presented in appraisal work. As can be seen, a lot of important information is missing. The only information shown is the apparently high R-Sq, the trend line and the model's equation. Of particular note in this graph is the lone, uncorroborated, high-influence point to the far, upper right on the graph. [31]

More of the important and relevant information about this model is presented in the second presentation of Model 1, Exhibits 9 and 10. Immediately noticeable in Exhibit 9, in addition to the high R-Sq, implying a high correlation between variables, is the inclusion of the average percentage trend line margin for error, which is 21%, and the standard error to the mean ratio, which is 107%. Exhibit 10 is the regression model's probability plot and residuals vs. fits graphs, which test whether the model meets the requirement for linearity. As can be seen by these two graphs, the points severely violate the diagonal in the probability plot, and in the residuals vs. fits graph, the points are not randomly and evenly distributed around the center line.[32] (The two relevant statistics in the probability plot are the AD or Anderson Darling statistic and the p-value. If the distribution linearly fits the data, the AD statistic will be small, and the associated p-value will be larger than the commonly chosen α-level of 0.05 - 0.10.[33])

Exhibit 9. Beauty Shops--Revenues Predicting Sale Price--Regression Model 1



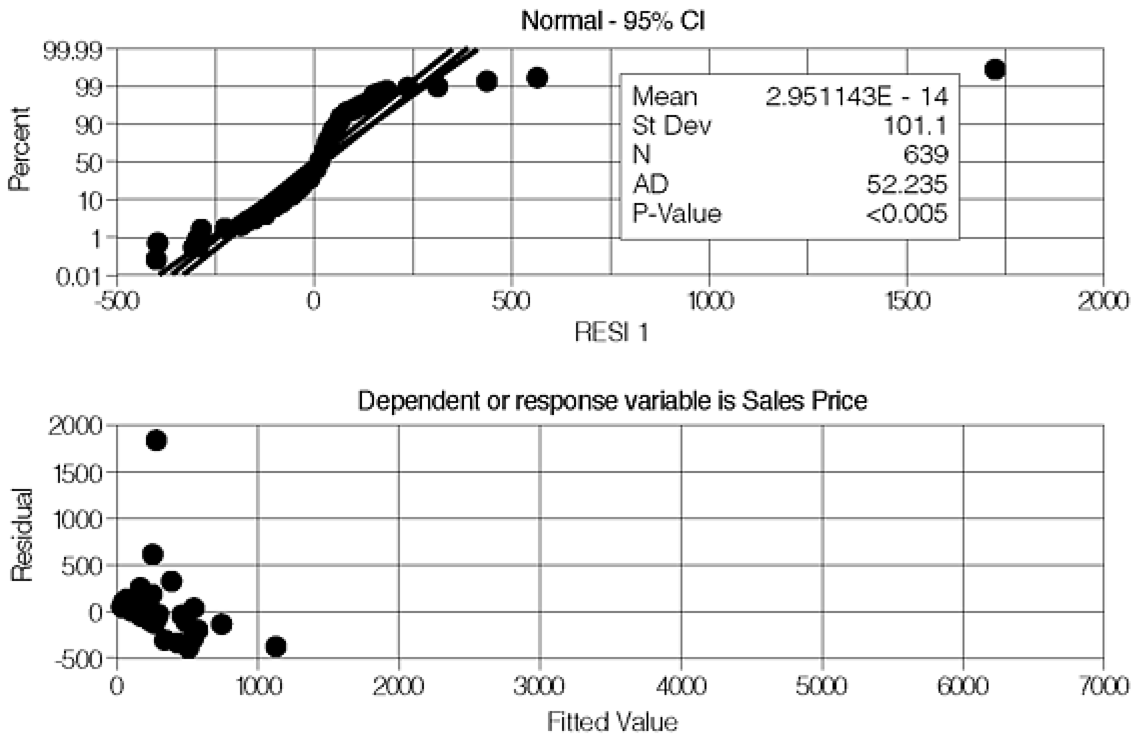
Standard Error / Mean Sales Price = 101.2/94 = 1.077

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	49479700	49479700	4830.36	0.00
Error	637	6525098	10243		
Total	638	56004798			

At 95% Confidence Intervals
Mean Trend Line % Margin for Error = 21.2%
Median Trend Line % Margin for Error = 5.2%
1st and 3rd Quartile Trend Line % Margin for Error = 5.2% to 29.7%

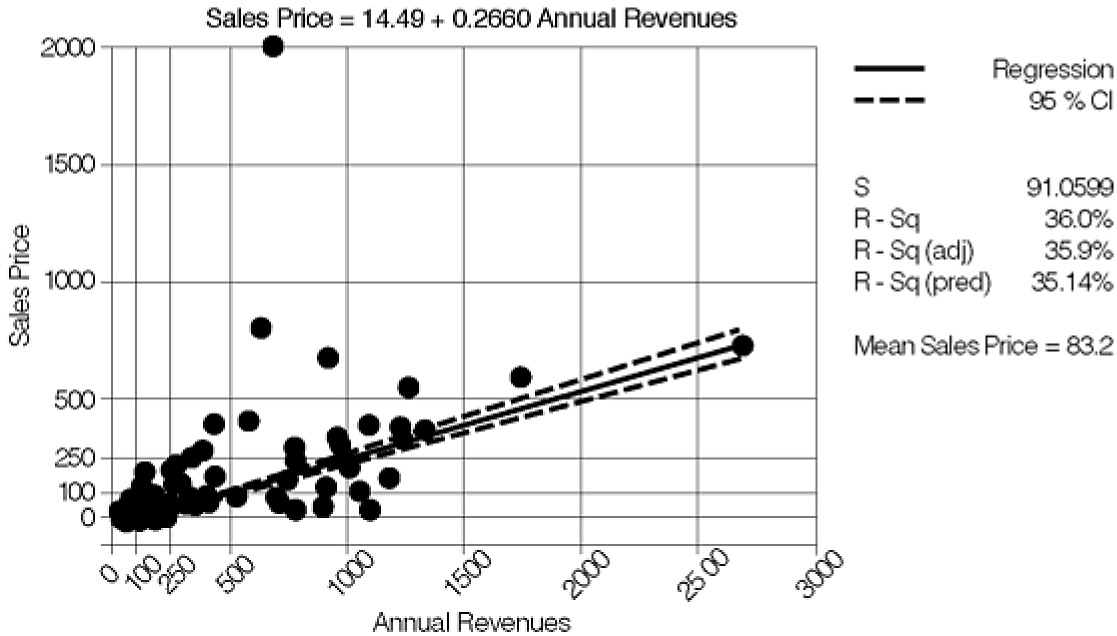
Exhibit 10. Probability Plot of RESI1/Residuals vs. Fits -- Regression Model 1



Linear Regression Model 2

Model 2 (Exhibit 11) uses the same data as the Model 1 graph, but it eliminates the one high-influence point. One can immediately see how dramatically the graph and the regression model equation changes. The points or observations become more apparent, the slope or angle of ascent declines from 42.5% to 26.6%, and the intercept is now positive at 14.49; it was previously negative at -25.91. (This intercept change from negative to positive is quite important and will be discussed later in the article.) The R-Sq, or how well the predictor variable (revenues) explains the dependent variable (sales price), dramatically declines from 88% to 35%, indicating a considerably weaker relationship. The ratio of the standard error to the mean is about the same, but the average trend line percentage margins of error is somewhat improved.

Exhibit 11. Beauty Shops--Revenues Predicting Sale Price--Regression Model 2



Standard Error / Mean Sales Price = 91.06/83.2 = 1.094

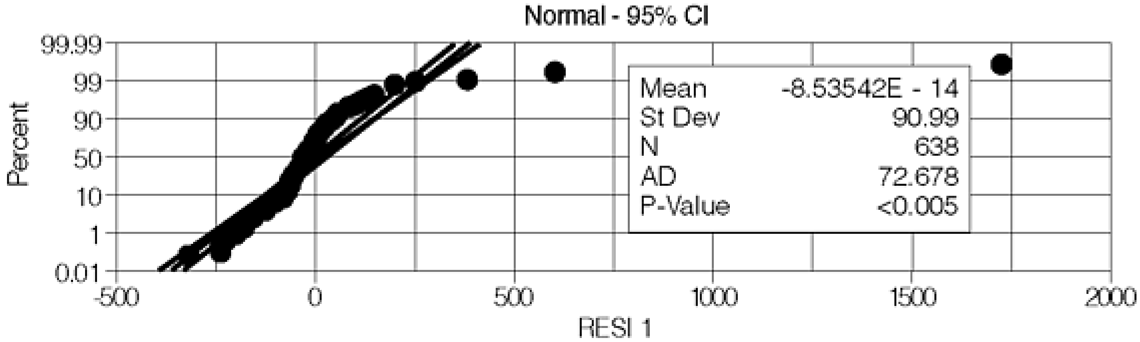
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	2963969	2963969	357.45	0.00
Error	636	5273651	8292		
Total	637	8237620			

At 95% Confidence Intervals
Mean Trend Line % Margin for Error = 15%
Median Trend Line % Margin for Error = 11.8%
1st and 3rd Quartile Trend Line % Margin for Error = 7.6% to 29.2%

However, the diagnostic graphs in Exhibit 12 of the probability plot and the residuals vs. fits continue to indicate severe violations of the normal distribution and the linear relationship. Overall, however, for the points or observations that do not exceed annual revenues of $1.25 million, Model 2 is a considerably better expression of the relationship of the variables than is Model 1.

Exhibit 12. Probability Plot of RESI1/Residuals vs. Fits--Regression Model 2



NonLinear Regression Model 3

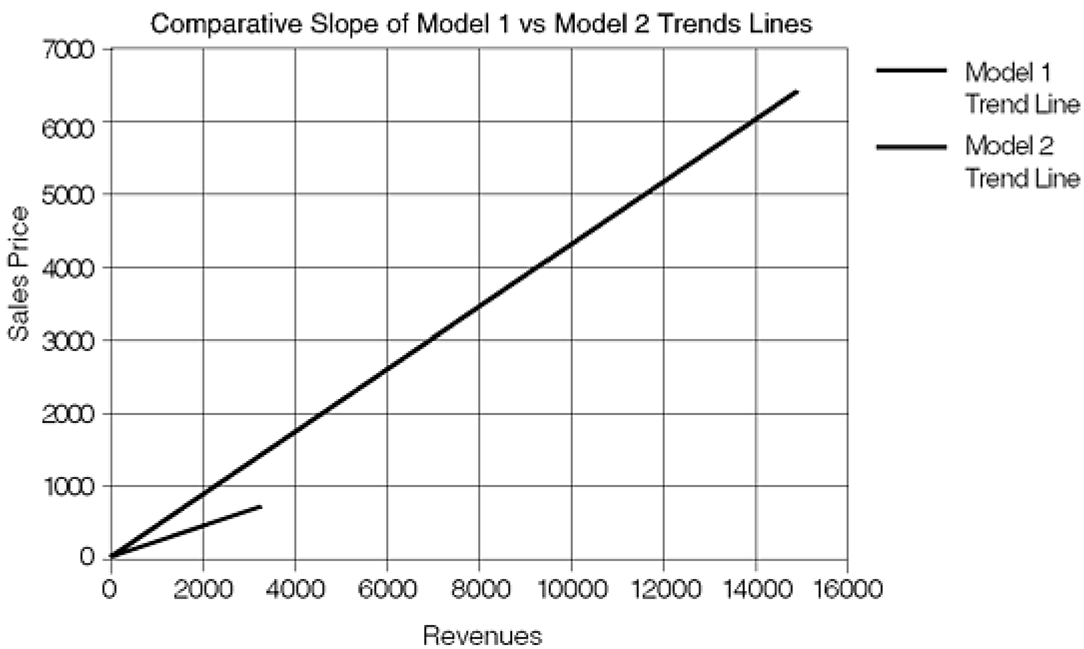
Experience indicates that repeated iterations of data segmentation by companies with smaller and smaller revenue size will not necessarily solve the problem of the nonlinear relationship in the data. Admittedly, that is something of a judgment call that comes from experience working with numerous statistical models. The application of a lowess line to the data (not shown here) indicates that perhaps the data should be segmented into two groups, those with revenues above $3,000 and those with revenues below $3,000. This may produce a superior model to Model 1, but there are three problems with segmentation:

(1) Additional segmentation often brings with it different lowess line patterns that indicate further segmentation.
(2) As the amount of data or the sample size is reduced, the R-Sq declines and the imprecision of the model (i.e., the margins of error) increases.
(3) Implicit in segmentation is the conflict of improperly omitting outliers and data that may be important to the analysis and relationship of the variables.

Experience with the data suggests a curved pattern that can be accommodated by a transformation procedure.

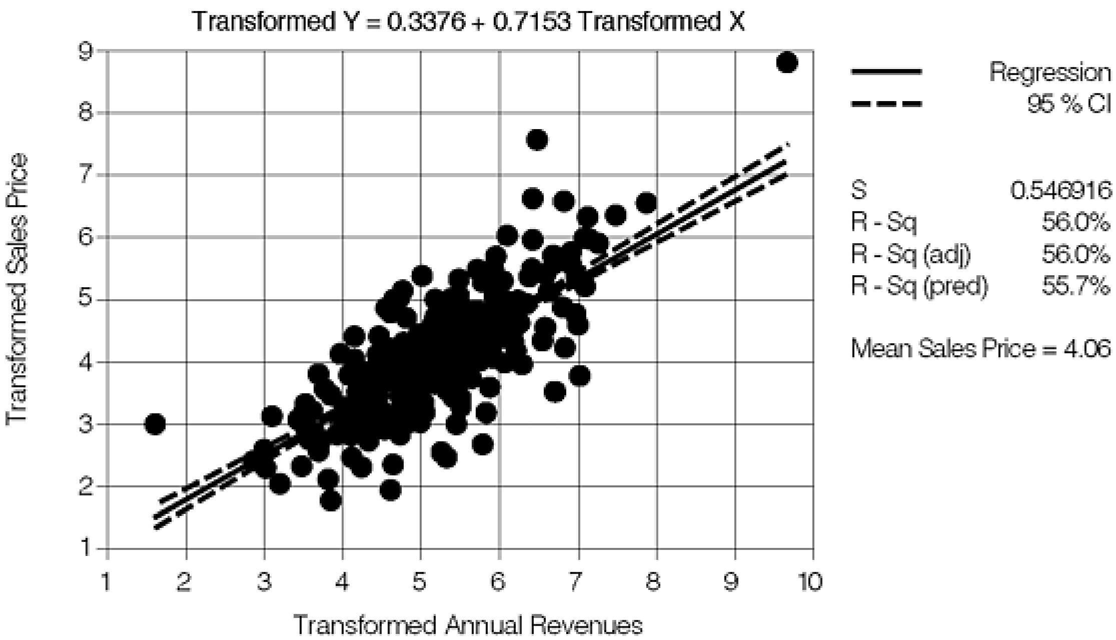
Various patterns lend themselves to various mathematical transformations of the variables to fit the trend line to the pattern.[34] Transformation of variables is a process by which the data is changed to a different scale such that a curved pattern in the original scale becomes linear in the new scale of measurement. It is beyond the scope of this article to describe the variety of ways in which data can be transformed. However in this case, as seen by comparing Models 1 and 2, the trend line starts out at a shallow angle and then steeply curves upward at an increasing rate to the high influence point. This can be better seen in Exhibit 13, which overlays the Model 1 and Model 2 trend lines on one graph.

Exhibit 13. Comparative Slope of Model 1 vs. Model 2 Trend Lines



This is a commonly occurring pattern in both appraisal work and in other scientific applications for which transformation equations have been developed to fit the trend line to the curved pattern. Knowing this, one can apply the transformation equation to develop Model 3 in Exhibits 14 and 15. [35]

Exhibit 14. Beauty Shops--Revenues Predicting Sale Price--Transformed X and Transformed Y Regression Model 3



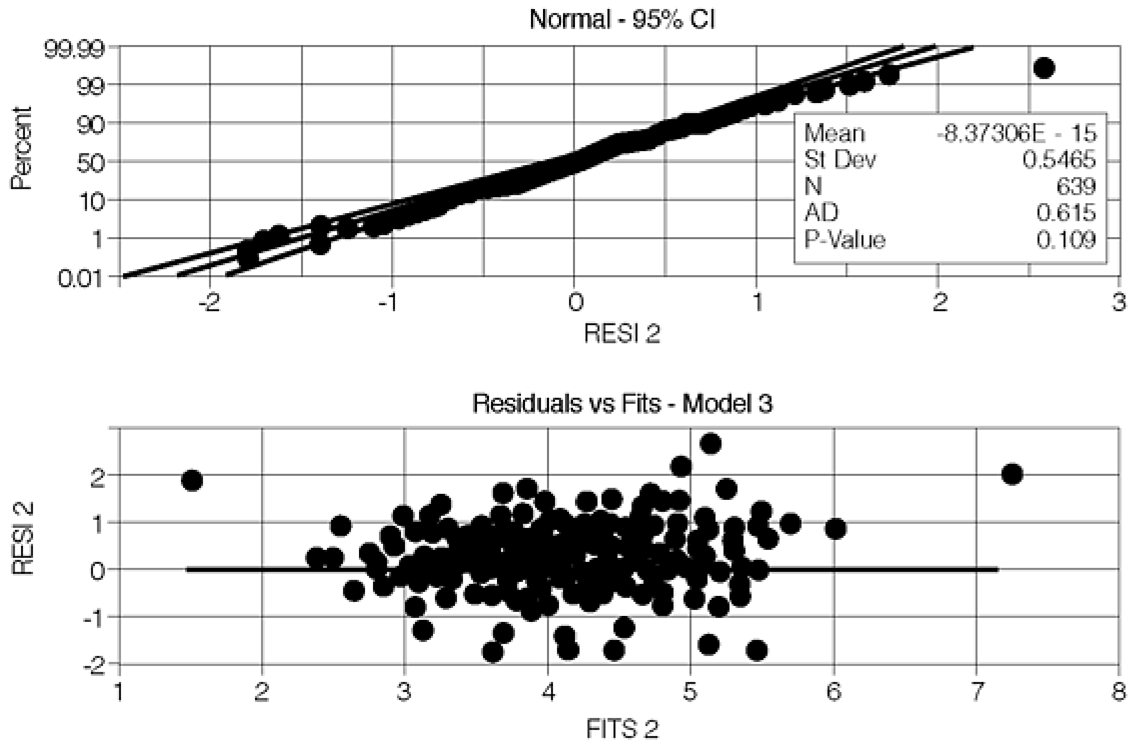
Standard Error / Mean Transformed Sale Price = 0.569/4.06 = 0.14

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	1	242.79	242.79	811.68	0.00
Error	637	190.54	0.30		
Total	638	433.32			

At 95% Confidence Intervals
Mean Trend Line % Margin for Error = 1.46%
Median Trend Line % Margin for Error = 1.35%
1st and 3rd Quartile Trend Line % Margin for Error = 1.16% to 1.66%

Exhibit 15. Probability Plot of Residuals / Residuals vs. Fits--Model 3



What one immediately sees in this Model 3 is that the R-Sq is now 55%, which is improved over the 35% R-Sq in Model 2, but less than the 88% in Model 1. In other words, revenues now explain about 55% of the sales price. This makes considerably more sense than the 88% explanation in Model 1, as it is common sense that considerably more goes into determining the value of a company than its revenues.

More importantly, the margins of error have declined dramatically to less than 2%, from 21% in Model 1 and 15% in Model 2. The graphs in Exhibit 15, the probability plot and residuals vs. fits, indicate that the transformation now satisfies normal distribution requirements for the regression. Overall, Model 3 is a better model than either Models 1 or 2.

Regression Modeling and Diagnostic Conclusions

Because the scale of the Model 3 is transformed, its graph is not directly comparable to the graphs of Models 1 and 2. To make the comparison, the trend line results have to be first converted back to the original scale. Exhibit 16 shows the comparative results for Models 1, 2, and the back converted Model 3.

Exhibit 16. Beauty Shops--Revenues Predicting Sale Price--Transformed X and Transformed Y Regression Model 3

Comparative Results For 3 Regression Models

	Revenues (000) 25	50	100	250	500	1000	R Square	Model's Trend Line Average % Margin for Error
Model 1 Sale Price:	-15.28	-4.66	16.6	80.37	186.64	399.19	88%	21.2%
Model 2 Sale Price:	21.14	27.79	41.09	80.99	147.49	280.49	36%	15.0%
Model 3 Sale Price:	13.75	23.01	37.77	72.74	119.43	196.5	56%	1.45%

What should be immediately noticeable is that Model 1 produces a sales price consistent with Models 2 and 3 only within the very narrow company revenue size corridor of $250,000, and then becomes quite inconsistent with the other two models, producing comparatively extreme and unrealistic values. Additionally, Model 1 actually produces improbable negative values. This is the result of Model 1's negative intercept. Common sense dictates that a company is unlikely to have a negative value, except in those rare cases in which liquidation costs exceed its asset values. A negative intercept such as that exhibited by Model 1 is a sure sign that the regression model is incorrect and inappropriate. In other words, such rarity indicates that the regression model is not representative of the market. Yet business appraisers fail to recognize this and repeatedly produce these types of nonsensical models to prove their valuation assertions. [36]

Model 2 has a somewhat broader corridor of consistency with Model 3, tending to confirm values within the company revenue range of $50,000-$250,000 in revenues. However, it still shows wide departures in relative consistency, because its linear trend line does not reflect the curvature pattern in the data.

Model 3 has the best potential for accurately representing how trends in revenue predict the sales prices of companies, given the limitations in the database. Its 1.45% percentage margin of error is far more reliable than the margin of error in the other two models, and the corridor of reliable values is considerably broader than in either Models 1 or 2. Finally, Model 3 fits the curved pattern in the data.

Conclusion

The statistical reasoning methods used in this article can be applied to most data used in business and real estate appraisal. When properly applied, these methods can support and verify valuation theories, assertions, and conclusions, or debunk them. These procedures, through well crafted *Daubert* challenges, [37] can also discredit appraisers who improperly apply, implicitly or explicitly, these methods, and expose situations in which valuation assertions are rendered without appropriate supporting evidence.

Unfortunately, even proper application of statistical methods to the existing databases—which but for a small handful of SIC Codes are poorly compiled, unverifiable, non-random samples, yet still marketed by the database vendors—does little to prove value. Improper application of these statistical methods to these poor-quality databases does even less to prove value. In these instances, the GIGO principal applies: garbage in equals garbage out.

It should come as no surprise that privately held database vendors and their proxies engage in hyperbole with regard to the necessity of their data, and the associated ideologies that they are selling, in which they have woven threads of improperly applied statistical methods, self-serving myth, specious “master” credentials, and in some cases voodoo science. However, by pulling on any of these threads, with the proper application of statistical reasoning, one finds that the entire fabric of their ideology often unravels.

If business appraisal is part science and part art, proper application of statistical methods and reasoning represents the science component. When statistical methods are omitted from appraisal work or misapplied, unsubstantiated biased opinion and discreditable dogma and ideology inevitably result. Such appraisals are more akin to the abstract art of Jackson Pollock than the realism of da Vinci or Michelangelo.

1

Huff, *How to Lie With Statistics* (W. W. Norton & Co., 1993).

2

See Spiegel and Stephens, *Theory and Problems of Statistics, 3rd Ed*. (McGraw-Hill, 1999), p. 1.

3

See, Curtiss, Lurie, and McMorrow, “The Rule of 10: Size Comparability and the Market Approach,” Business Appraisal Practice 54-56 (Winter 2004-2005), with additional attribution for the premise given by the authors to Gary Trugman.

4

See Stanford Prison Study, Asch Conformity Line Studies and Milgram Experiment.

5

Groupthink occurs when a group makes faulty decisions because group pressures lead to a deterioration of mental efficiency, reality testing, and moral judgment. It is a type of thought exhibited by group members who try to minimize conflict and reach consensus without critically testing, analyzing, and evaluating ideas. Groups manifesting “groupthink” often rationalize away warnings that challenge the group's assumptions; stereotype individuals who oppose them as weak, stupid, or disloyal; engage in self-censorship of ideas that conflict with their own; pressure dissenters not to express views and opinions which differ from the group; harbor illusions of unanimity among group or industry members; appoint leaders and directors who agree only with them and shield the group from dissenting ideas; and believe in the rightness and superiority of their cause and therefore ignore the ethical or moral consequences of their decisions. Examples of groupthink disasters are the space shuttle explosion and accounting frauds such as Enron, WorldCom, Cendant. See, Janis, *Victims of Groupthink* (Houghton Mifflin, 1972); *Groupthink: Psychological Studies of Policy Decisions and Fiascoes, 2d Ed.* (Houghton Mifflin, 1982).

6

See Asch, “Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority,” Psychological Monographs, 70 (Whole No. 416) (1956); Asch, “Effects of Group Pressure Upon the Modification and Distortion of Judgment,” Guetzkow (ed.), *Groups, Leadership and Men* (Carnegie Press, 1951); Asch, “Opinions and Social Pressure,” 193 Scientific American 31 (1955).

7

See, Milgram, "Behavioral Study of Obedience," 67 Journal of Abnormal and Social Psychology 371 (1963); Milgram, *Obedience to Authority; An Experimental View* (HarperCollins, 1974).

8

Miles, *Introduction to The Direct Market Data Method* (Institute of Business Appraisers, 1998), page I-1, in which the data and the method are described to represent the “overall market." See also, Fannon, "Uses and Abuses of Market Data: An In-Depth Look at the Tools of Our Trade," Business Valuation Review, (Summer 2006), p. 74. In which this author, parrots the spurious and discreditable "total market" ideology given to her by the database vendors.

9

Id., p. I-10.

10

See note 1, *supra*, Chapter 5, pp. 60-65.

11

The formula for solving for a minimal, random sample size required to estimate the population mean μ, is: $n = [(z_{\alpha/2} * \sigma) \div E]^2$ where n = the sample size, $z_{\alpha/2}$ = critical z score based on the desired degree of confidence, E = desired margin of error, and σ = population standard deviation. See Wolpin "Examining the Reliability of Small Business Transaction Databases," 7 Val. Strat. 4 (November/December 2003), for a more thorough discussion of proper sample size calculations.

12

This graph shifts to the Z-distribution, which is used to accommodate sample sizes over 30. The first graph used a student t distribution, which accommodates sample sizes below 30. Use of the student t distribution instead of the Z-distribution accounts only for a relatively small, immaterial difference to this broad sample size range.

13

Regardless of the optimum point of the curve, if a data set produces a margin of error as large as 19%, there will always be the nagging question of whether the dataset is sufficiently precise, informative, and useful.

14

Spline and lowess procedures are statistical methods that identify two phenomena: (1) bending, breaking, and changing in the trend of the overall data where segmentation may be warranted and (2) non-linear relationships in the data. For a good explanation of these advance techniques, see, Ott and Longnecker, *An Introduction to Statistical Methods and Data Analysis, 5th Ed.* (Thomson Learning, 2000), pp. 535-538. See also Nuter, Kutner, Nachsheim, and Wasserman, *Applied Statistical Models, 4th Ed.* (McGraw Hill, 1996), pp. 136-138.

15

Graphs based on: Anscombe, "Graphs in Statistical Analysis," 27 The American Statistician 17 (1973); and Tufte, *The Visual Display of Quantitative Information* (Graphics Press, 1983).

16

See Wolpin, "Statistical Reasoning in Family Limited Partnership Appraisals: New Tax Court Scrutiny," 18 Probate & Property 46 (May/June 2004), reprinted in 72 Appraisal Journal 363 (Fall 2004).

17

See Curtiss, "Statistical Confidence," Business Appraisal Practice (Spring 2003). This author, contrary to sound statistical practice and reasoning, advocates outlier elimination and erroneously assumes that the IBA database is randomly collected and that the large size of the IBA database sample overcomes and renders irrelevant (1) the serious infirmity of a biased sample, (2) the incorrect procedure of unconfirmed outlier elimination, and (3) the necessity of identifying the conditions of a business sale. "Some abusers of statistics are simply ignorant and careless, but others have personal objectives and are willing to supress unfavorable data while emphasizing supporting data." See Triola, *Elementary Statistics, 8th Ed.* (Addison Wesley Longman, 2001), p.11.

18

Chatterjee, Hadi, and Price, *Regression Analysis by Example, 3rd Ed.* (John Wiley & Sons, 2000), p.108.

19

Rudich, "Analyzing the IBA Database Transaction Results," Business Appraisal Practice 4 (Winter 2006-07); and 13 Business Valuation Update 8 (August 2007), pp. 18-23. This article was selected for publication in a so-called peer reviewed IBA journal by an editorial staff composed of IBA "master appraisers".

20

Although the referenced statistical analysis was performed on a database of small business transactions, the methods and analysis are applicable to all databases of transactions, large and small.

21

Both Segments 1 and 13 produced similar-looking graphs, therefore only the Segment 1 graph is shown and discussed in this article. Additionally, the sample size count was different in this article's reproduced graph. Though larger, for purposes of illustrating proper statistical technique, this difference is immaterial

22

See, Thompson, *Sampling, 2nd Ed.,* (John Wiley & Sons, 2002), pp. 1-5.

23

There are two small business databases that are heavily promoted by their owners to business appraisers, the Institute of Business Appraisers (IBA) database and the BIZCOMP database. Neither of these databases are properly collected random samples, and both suffer from poor quality control and verification problems. Consequently, conclusions from either are questionable.

24

See references in basic statistics books to the 1936 Literary Digest Poll, which predicted that Alfred Landon would defeat Franklin D. Roosevelt by a 3 to 2 margin.

25

See, Green, "How Many Subjects Does it Take to Do a Regression Analysis?" 26 Multivariate Behavioral Research 499, in which the actual rule of thumb formula = 50 +8k, where k = the number of predictors. There are more exacting procedures than rules of thumb, such as the use of *a priori* power analysis on the regression coefficients, but they are beyond the scope of this article.

26

See note 9, *supra*.

27

See Snedecor and Cochran, *Statistical Methods, 8th Ed.* (Iowa State University Press, 1989), pp. 438-439), See also Wolpin, "Examining the Reliability of Small Business Transaction Databases,"7 Val. Strat. 4 (November/December 2003).

28

See note 14, *supra*.

29

See "The Truth About Traffic Counts," 23 Professional Carwashing & Detailing (November 1999).

30

See Wolpin, “Should Appraisers Rely on the Small Business Transaction Databases to Determine Fair Market Value?” 5 Val. Strat. 4 (July/August 2002), for a short list of variables affecting value not captured by the databases and a discussion of the 3Ds or 4Ds of motivation (death, divorce, disability, distress) reflected in business brokerage sales. These motivations known as “conditions of sale” in real estate parlance, are omitted from the databases; they tend to skew database values below fair market value for those companies that otherwise do not suffer from these motivations. For further discussion of “condition of sale adjustments,” see *The Appraisal of Real Estate, 12th Ed.* (Appraisal Institute, 2001), pp. 433, 453-454. See also Fisher and Martin, *Income Property Valuation* (Dearborn Financial Publishing, 1994), p. 187.

31

Unfortunately, the IBA database provides no way to determine whether this loner point is a point of high influence to be preserved and considered, or a miscoded, illegitimate point that should be eliminated. The difference is highly significant as the effects of each are quite different. For the illustrative purposes of this article, the assumption is made that it is a legitimate point to be considered.

32

The distinct widening or funnel shape of points around the center line in the residuals vs. fits graph is called heteroscedasticity in statistical parlance, and is an indicator of a nonlinear relationship.

33

This is opposite of the required p-value interpretation when presented in a linear regression, which requires a value smaller than 0.05-0.10 as proof of variable significance.

34

For instance, radioactive (half-life) decay is not linear; it is a downwardly accelerating curve, reflecting a negative compounding over time. A special nonlinear regression transformation technique is available to identify and correctly determine this relationship between variables.

35

All of these procedures can be performed almost instantaneously with any one of the modern statistical computer packages, except Excel, once one has had proper training in statistical methods.

36

See Ross, “A Graphic Appraisal Method: How a Picture Can Be Worth 10,000 Words,” Business Valuation Update, (June 2007), in which the author used a model with a negative intercept and constructed incorrect confidence intervals.

37

See Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 US 579 (1993).