Valuation Strategies

MARCH/APRIL 2009

SMALL BUSINESS TRANSACTION DATABASES

INTELLECTUAL PROPERTY

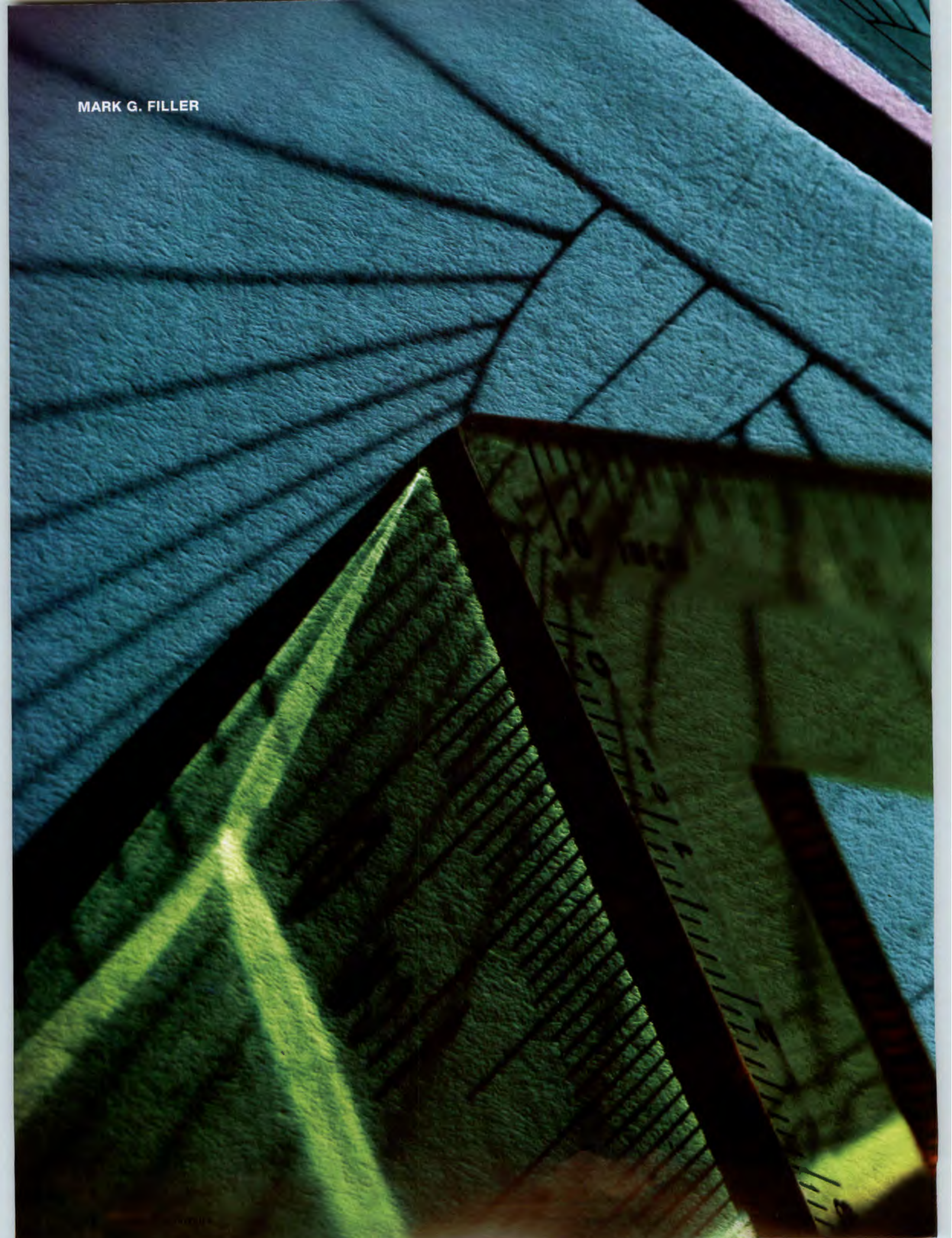
LACK-OF-MARKETABILITY DISCOUNTS

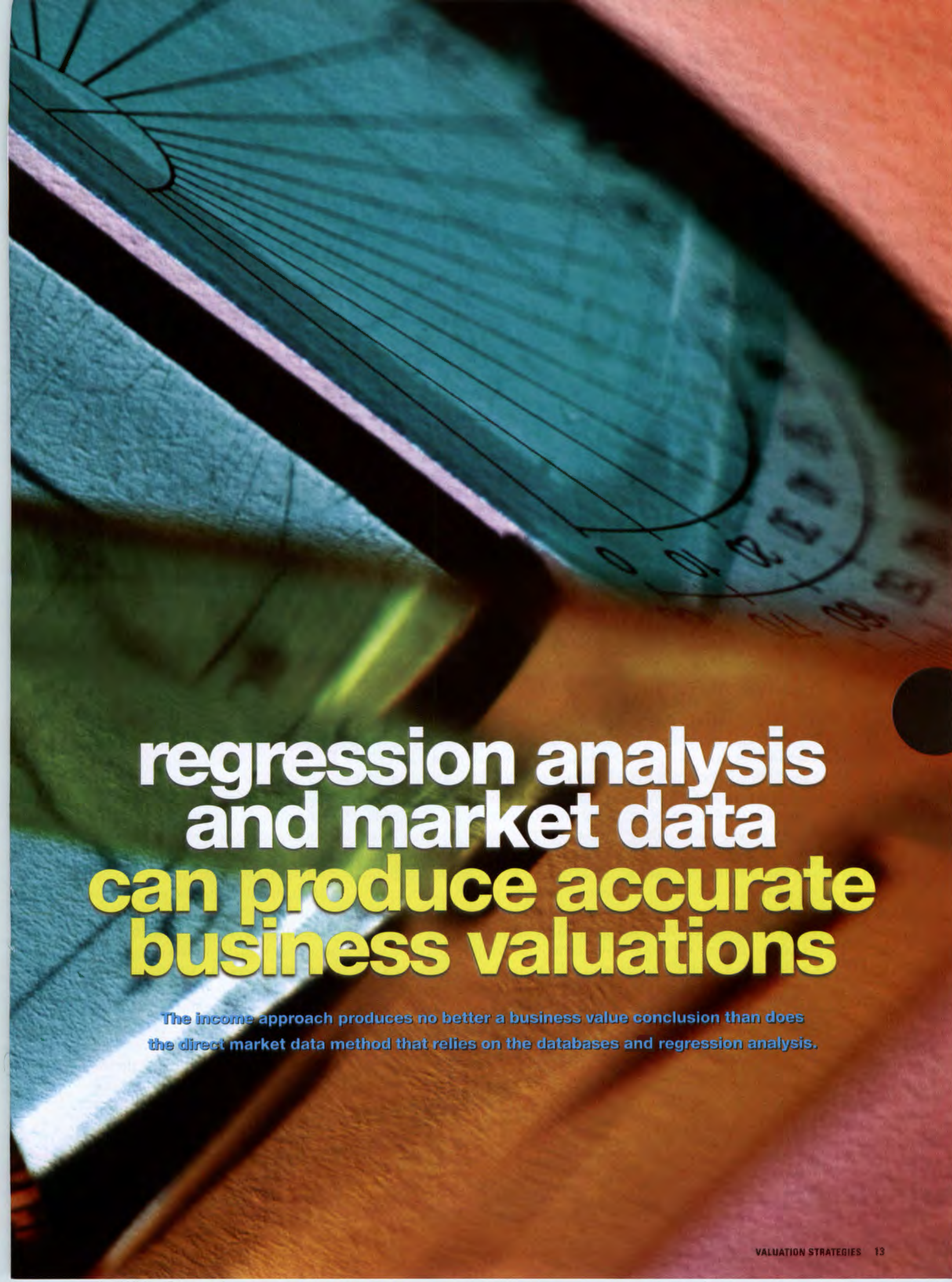
Valuation Strategies

MARCH/APRIL 2009
VOLUME 12, NUMBER 4

WG&L

MARK G. FILLER





regression analysis and market data can produce accurate business valuations

The income approach produces no better a business value conclusion than does the direct market data method that relies on the databases and regression analysis.

In a series of well-written and comprehensive articles in this journal over the past few years, Jeffrey Wolpin launched an attack on the traditional ways that the IBA, *Pratt's Stats*, and *Bizcomps* transaction databases are considered and used by the majority of the valuation community.[1] With a seemingly firm grasp of statistical concepts and reasoning, he has berated the purveyors and users of the databases for (1) applying incorrect methodologies to the databases, and (2) even attempting to apply correct methodologies to the databases, because, with few exceptions, they are not now, and never can be, statistically valid samples. Continuing this assault, his latest article in *Valuation Strategies*[2] deplores improper reasoning, discredits some valuation "myths," and sums up his opus by demonstrating the use of regression analysis as applied to the databases.

The purpose of the present article is to demonstrate that, except for one major point—that regression analysis is the proper tool of choice to employ with these databases—and a few minor points, Wolpin is incorrect concerning almost all of his claims, assertions, and proofs. The focus of this is on Wolpin's latest article in the order of the ideas presented, while major claims, assertions, and arguments from his prior articles are reexamined when he refers to them in his latest article.

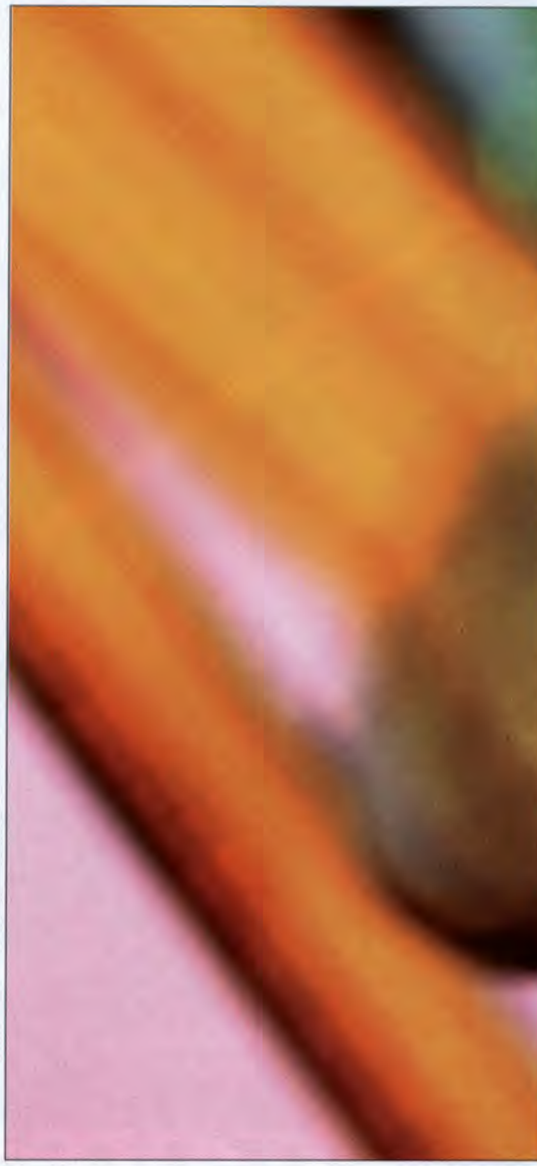
Wolpin begins his latest article by attacking what is known as the fallacy of "appeals to authority," i.e., referencing or citing the writings of better known valuation analysts to buttress an assertion that ought to stand on its own logic or factual evidence. Wolpin is being too harsh, as many valuation experts practice in an arena where "appeals to authority" are mandated. In the courtroom, the old *Frye* rule, and now one of the five *Daubert* prongs, requires that an expert's testimony be "generally accepted" in the community in which he or she practices. Referencing and citing other experts who have published on the topic at hand help establish for the court this general acceptance. Whether these "authorities" are correct is another matter, but appealing to them is not a fault, at least in the litigation ring.

Attacks on Direct Market Data Method

Means and Medians. Wolpin attacks the use of the direct market data method on a number of grounds, the first of which, the "simplistic use of means and median multiples of sales and earnings," is the one major point in which Wolpin and the present author are in complete agreement. While there are some minor reasons for rejecting use of means and medians, the major reason is that when it comes to reducing dispersion and fitting the data, averages and medians are outperformed by regression analysis. If value is truly a range, that range should be as tight as possible. Regression analysis will narrow that range considerably further than any average or median ratio.

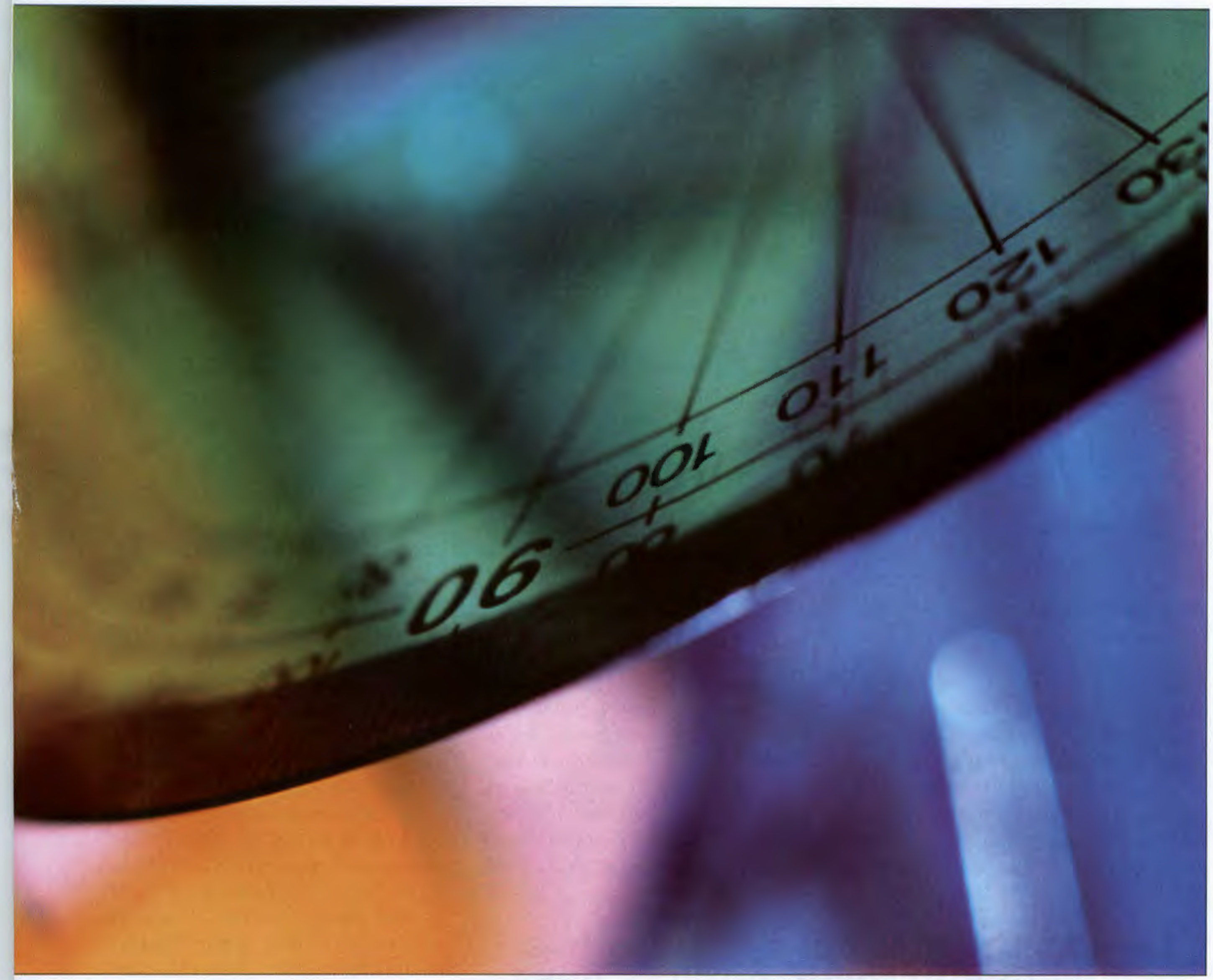
Value Drivers. Wolpin's second point is that the wide degree of dispersion about the mean pricing multiple in the databases is caused by lack of knowledge concerning value drivers. This might be true when annual revenue is the value driver, but when applied to normalized and forecasted seller's discretionary earnings (SDE), it is problematic, as most value drivers are already subsumed in that number. While there might be further adjustments to the value calculated by the base SDE model for such things as a short-term lease, equipment not up to date, volatile earnings, high level of customer concentration, growth potential greater or lesser than normal, and understaffed management, one cannot just assume that these variables definitively account for all the variability in the data—it might just be random error.

Actually, this is no different from Wolpin berating the Institute of Business Appraisers (IBA) for promoting the idea that sales transactions that produce larger multiples necessarily represent superior businesses, rather than the result of just random variation occurring in a purchase and sales marketplace full of multiple forces pushing and pulling in different directions. No two businesses are exactly the same. Therefore, no two selling prices ought to be exactly the same. An analyst's judgment will come into play, because he or she accounts for, to the extent necessary, the adjustments to base value in any valuation exercise.



Outlier Removal. A third questionable statistical method, according to Wolpin, is that of outlier removal. Wolpin's assertion that certain data points should never be removed unless there is strong evidence that they are indeed outliers is correct, but then he goes on to state that no reasons can be produced for removing these outliers. This last statement is incorrect, because valuing a business does not require performing basic research into what drives value. For example, if the cause of lung cancer is the objective

MARK G. FILLER, CPA/ABV, CBA, AM, CVA, is the founder and president of Filler & Associates, P.A., a CPA firm in Portland, Maine. His personal practice focuses on business valuations, measurement of damages for lost profits, and assistance in adjusting business interruption losses.



of basic research, researchers are quite interested in the outliers—people who smoked one or more packs of cigarettes a day but never got lung cancer, and people who never smoked at all but still came down with lung cancer. These outliers would be important sources of new research.

However, the answer to the question of what drives business value is already known (cash flow), and so outliers are unable to provide any meaningful information. If the subject company is "average" (as calculated by return on sales, return on equity, current ratio, gross margin, debt/equity ratio, etc.), it should be valued using "average" comparison multiples, not including outliers. Factors that might indicate divergence from the "average" could result in an upward or downward adjustment to the penultimate conclusion of value. Unless the subject company is a real winner or loser, outliers have no place in the valuation universe.

When commentators write that when a data set contains influential outliers one should default to the median because the median is not influenced by outliers, they are, in fact, suggesting that outliers be removed, because when they are, the mean will come closer to equaling the median. Others advise elimination of items in the data set that are either too large or too small when compared to the subject company. Following this advice necessarily involves the removal of outliers, albeit two distinct classes of outliers. Removing only data points that are more than 2.0, or 2.5, or 3 standard deviations from the mean just removes another class of items—those that are clearly not members of the population of interest. They are the result of special circumstances that the average pricing multiple is not intended to cover.

What one does when outliers are identified in the data set is not without controversy. If the outlier is a result of a data entry error or is otherwise suspect in terms of its reliability or accuracy, it clearly should be removed from the data set or repaired before any further analysis. What should be done, however, about outliers that are not clearly erroneous, such as those that lie between 2 and 3 standard deviations from the mean of the regression line? Somehow, leaving those observations in the data set has come to be viewed as the "honest" thing to do, and removing them is viewed as "cherry-picking" or "cheating" or "making it work."

Outliers are extreme observations that for one reason or another do not belong with the other observations in data sets. There are two ways that outliers can be introduced into the databases, the first of which, as mentioned above, results from incorrect recording, or especially, data entry errors that can put wild values into the data sets. The second cause of outliers is that data sets are not a homogeneous unit to which a single regression model will apply, but rather a heterogeneous mix of two or more types of transactions, one of which is more frequent. The infrequent observations of the other types will appear as outliers as they represent: (1) foolish buyers who have overpaid, (2) foolish sellers who have accepted less than fair market value, (3) distressed sellers, or (4) synergistic buyers. These items violate the fair market value standard of value, and therefore do not belong in the data set. For that reason, it is necessary to delete them, along with the obvious data input errors. If a data set contains 75 data points, and 65 of them are within 2.5 standard deviations of the mean, why are the other ten needed, and what helpful information do they contain?

If a data set is heterogeneous and contains all types of transactions, why not exclude those that do not fit the fair market value standard of value? By definition, it is true that any transaction outside the mainstream does not conform to that standard, whatever the reason. For example, how can a sale that is 4.5 standard deviations from the mean be at fair market value? Is not the sale at investment value—value to a particular buyer? Even if the assumption is made that a sale at 4.5 standard deviations is truly a fair market value transaction, this question remains: Why did it sell for such a high multiple? Perhaps it has the best location, the best management, superior service, loyal customers, etc. All these things will tend to make its SDE far in excess of the average enterprise in its SIC code number. Therefore, it sold at a premium. Not only was its SDE multiplied by the average multiple, but the buyer paid a premium for its superior performance and the fact that its recipe for success will survive the closing because it has been systematized by management.

Now, the question can be asked: Does a subject company enjoy such profits and have such systems in place? If not, how can the outlier company be similar and relevant to the subject valuation assignment? It cannot be, and thus it should be removed from the data set. Therefore, remove the outliers because they do not represent fair market value transactions, or remove them even if, in the extreme, they do.

In response to a chart created by Ray Miles,[3] Wolpin insists that five to seven transactions are always an inadequate sample size. However, as he has previously pointed out, the less dispersion there is about the mean, the smaller the sample size can be. If all seven data points line up on a trend line, why would they not be an adequate sample? The present author has found bivariate data sets of as few as nine observations, but where r^2 (the coefficient of determination) was in excess of .95. (See Exhibit 1.) Should the method have been dropped due to too small a data set? The present author does not think so.

Unless the subject company is a real winner or loser, outliers have no place in the valuation universe.

Wolpin misconstrues Miles' chart. If the chart is recast into the schedule in Exhibit 2, perhaps some light can be shed on the subject. The critical value of t for a sample size of 7 is 2.45, while the critical value for a sample size of 16 is 2.13, and for one of 31 it is 2.04. While these differences are not insignificant, they are not substantial. Miles' point is that a large sample size is not all that it is cracked up to be, and that meaningful information can be derived from small samples. It should be noted that William Gossett invented the t-distribution to compensate for the fact that his employer, Guinness Brewery, constrained his sample sizes, so that he often had to work with only four or five observations.[4] Wolpin has fallen into the trap of thinking that large sample sizes correct for all sorts of errors in the data, and that all small samples should be eschewed.

Regression Analysis

Wolpin next enters the realm of regression analysis, his and the present author's preferred method for handling the databases. However, he gets off on the wrong foot immediately with his definition of r^2, the coefficient of determination, by claiming that it measures the strength of the linear relationship between the x and y variables. That is what r, the coefficient of correlation, does. Wolpin's alternative phrase does capture the correct definition of r^2.

Wolpin goes on to suggest, and the present author agrees, that r^2 alone is insufficient as a regression metric, and that it must always be used in conjunction with the standard error of the estimate (SEE). However, he claims that one can have both a low r^2 and a small SEE. While a large influential point can produce a result that demonstrates a high r^2 and a larger SEE, it is mathematically impossible to have a result that demonstrates a low r^2 and simultaneously, a small SEE. The SEE and r^2 are produced from the same source—the sum of squared residuals. Therefore, if one is large, the other is small, and vice-versa. See Exhibit 3 for the calculations that demonstrate this.

The only thing that can disturb this equanimity is a large influential point that will increase total sum of squares without changing residual sum of squares, thereby driving up r^2 and leaving the SEE unchanged. Removing the influential point will lower r^2, but leave the SEE mostly unchanged. Mathematically, the reverse cannot happen, as there is no opposite of an influential point. Models that produce a low r^2 and a high t-statistic, or narrow confidence interval, are useless in a valuation setting, as they have no explanatory power. The low r^2 indicates that the model is misspecified, i.e., other variables are missing and need to be included in the model for it to be able to say something other than "there is a relationship between x and y." That relationship is necessary but not sufficient—the real questions are, how good is the model's fit and how much explanatory power does it have?

Wolpin misnames regression terms, such as the standard error of the estimate (SEE) and the coefficient of variation (COV). The correct name for the standard deviation of the regression line is the "standard error of the estimate," not the standard error, which in regression parlance is also a term that relates to sampling procedures, and is associated with the x coefficient(s). The SEE is not computed using the square root of n, while the standard error of the coefficient is, using the sum of squared deviations of the x variable as a substitute for n. This makes the SEE a standard deviation, which has to do with the data set as a whole, and the standard error of the x coefficient a true standard error, which has to do with confidence intervals for sample means. This is an important distinction. Either unknown to or unstated by Wolpin, the result of dividing the SEE by the average of the dependent variable has a name—it is called the coefficient of variation of the trend line.

However, Wolpin fails to focus on the SEE, and instead emphasizes standard errors, confidence intervals, and p-values, as if these alone were enough to indicate goodness of fit and model accuracy. This results from both an over-reliance on survey techniques that are not strictly on point in a valuation setting, and a failure to distinguish

EXHIBIT 1
Example: Small Sample Size/Small Dispersion About Mean

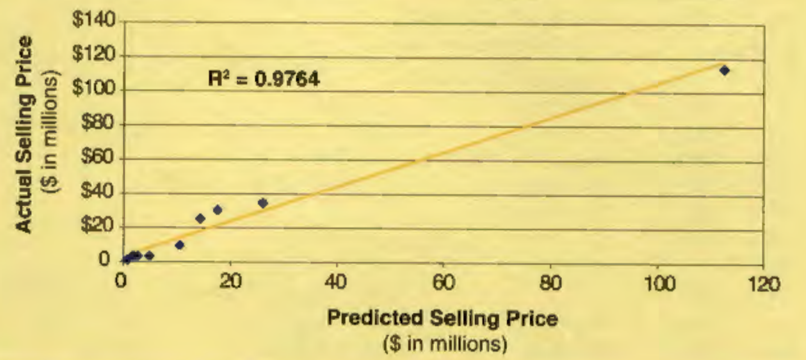


EXHIBIT 2
Recast Schedule of Miles's Chart

Sample Size	df	t-critical 95.0% Conf. Level	% Change	SDE	Multiple	Multiple Std Dev	Lower Interval	Mean Value	Upper Interval
3	2	4.30		100	2	0.55	(37)	200	437
4	3	3.18	26.0%	100	2	0.55	25	200	375
5	4	2.78	12.8%	100	2	0.55	47	200	353
6	5	2.57	7.4%	100	2	0.55	59	200	341
7	6	2.45	4.8%	100	2	0.55	65	200	335
8	7	2.36	3.4%	100	2	0.55	70	200	330
9	8	2.31	2.5%	100	2	0.55	73	200	327
10	9	2.26	1.9%	100	2	0.55	76	200	324
11	10	2.23	1.5%	100	2	0.55	77	200	323
16	15	2.13	4.3%	100	2	0.55	83	200	317
21	20	2.09	2.1%	100	2	0.55	85	200	315
26	25	2.06	1.3%	100	2	0.55	87	200	313
31	30	2.04	0.8%	100	2	0.55	88	200	312

1 Wolpin, "Should Appraisers Rely on the Small Business Transaction Databases to Determine Fair Market Value?" 5 Val. Strat. 4 (July/August 2002); Wolpin, "Examining the Reliability of Small Business Transaction Databases," 7 Val. Strat. 4 (November/December 2003); Wolpin, "Unreliability of the Small Business Transaction Databases and the Guideline Company Method," 8 Val. Strat. 12 (September/October 2004).

2 Wolpin, "Myth-busting—Discrediting Appraisal Myths Through Properly Applied Statistical Reasoning," 11 Val. Strat. 20 (January/February 2008).

3 Miles, "How to Use the IBA Market Data Base," Part VIII, "Choosing the Method—How Many Guideline Transactions are Needed?" Ray Miles has been authoring, on an irregular monthly basis since 1993, four-page monographs pertaining to the IBA transaction database and the direct market data method that are part of the IBA newsletter mailed to its membership. To date, the total number of monographs is 48.

4 Ziliak and McCloskey, *The Cult of Statistical Significance* (University of Michigan Press, 2008), p. 3.

5 Wolpin, note 1, *supra* (2002, p.8).

6 Tatum, *Transaction Patterns* (Toby Tatum, 2000), p. 96; Miles, note 3, *supra*, Part XIV, "More About the IBA Market Data Base," p. 4.

7 Tatum, note 6 *supra*, pp. 35 - 40. Tatum plotted individual and average Price/SDE ratios over time from 1989 through 1999 and found that there has been no increase or decrease whatsoever in the Price/SDE ratios of small businesses over that period. Annual fluctuations could not be correlated with changes in the prime lending rate.

8 Wolpin, note 2, *supra*, p.27.

9 *Id.*, p.28.

10 *Id.*, p.29.

11 A spline fit is a data analysis technique for estimating (via the least squares criterion) the parameters in a spline polynomial model. It is commonly used to fit curves that have different shapes in different areas of the x axis variable. Loess smoothing is a curve-fitting technique based on local regression.

12 Filler, "Is There a Buy-a-Job Phenomenon in Business Valuations?" 7 Val. Strat. 20 (July/August 2004). The data sets are rarely distributed in such a manner that the application of simple linear regression will provide a relevant and reliable answer. This is because the individual databases are (1) hardly ever linear, (2) infrequently homogeneous as to variance (the larger the x variable, the greater, or smaller, the dispersion about the regression line), and (3) not often normal, or even symmetrical. If the data is linear, simple linear regression can be used without having to resort to more complex models, i.e., one can stick with the tools that Excel provides. The reasons that homogeneity and normality, or at least, symmetry are good things are beyond the scope of this article, but suffice it to say that without these qualities, standard statistical tests and confidence intervals will not be reliable, nor will variation in data be able to be explained away as noise, or ordinary and expected random error. Simple tests will make it apparent that a model is deficient. Fortunately, only one procedure (transformation of either or both the x and y variables) is needed to fix these three problems. This is so because data that is not normally distributed is also often neither linear nor homogeneous. Thus, transformation provides a simple way both to fix statistical problems (nonsymmetrical and heterogeneous distributions) and to fit curves to data (curvilinear regression). Dummy, or categorical variables, allow data to be segregated by size, location, etc., thereby filtering the data to control for various pricing interventions.

13 Wolpin, note 2, *supra*, pp. 30 - 31.

14 Log-log, or power, transformations are where both the x and y variables are transformed by taking them to either their base 10 logarithms or their natural logarithms.

15 *Ibbotson SBBI 2008 Valuation Yearbook* (Morningstar, 2008), p. 134.

EXHIBIT 3
Calculations of SEE and r^2

x SDE	y Price
19	61
24	31
47	85
50	150
60	60
70	82
71	55
75	50
87	219
95	94
100	150
105	215
115	129
115	249
125	240
129	83
133	78
153	382
180	519
210	434
212	503
230	630
253	516
270	750
393	810
466	1263
847	2258

SUMMARY OUTPUT
With Influential Point

Regression Statistics	
Multiple R	0.982
R Square	0.965 = 1 - (207,581/5,980,131)
Adjusted R Square	0.964
Standard Error	91.122 = SQRT(8,303)
Observations	27

ANOVA

	df	Sum of Squares	Mean Squares
Regression	1	5,772,450	5,772,450
Residual	25	207,581*	8,303
Total	26	5,980,031	

* 8,303 = 207,581/25

Without Influential Point

Regression Statistics	
Multiple R	0.954
R Square	0.911 = 1 - (204,980/2,295,931)
Adjusted R Square	0.907
Standard Error	92.417 = SQRT(8,541)
Observations	26

ANOVA	df	Sum of Squares	Mean Squares
Regression	1	2,090,951	2,090,951
Residual	24	204,980*	8,541
Total	25	2,295,931	

* 8,541 = 204,980/24

between two quite different, but similar sounding, statistical ideas: standard deviation and standard error. As mentioned above, the sample standard deviation for a single set of data is computed by summing the squared distance each data point is from the mean, averaging this sum by dividing it by the number of data points in the series minus 1, and then taking the square root of the resulting quotient. The standard deviation represents the average distance each data point is from the mean of the series, and is the best measure of variability when the data series is near bell shaped, or normal.

To compare the variability of two data series, compute and compare the coefficient of variation (COV), a dimensionless ratio, of each data series, which is derived by dividing the standard deviation by the mean. The standard deviation is used when one is interested only in describing a particular, specific sample, or in the case of business valuation, determining where the value of the subject company might lie on the normal curve in comparison to the values of the companies in the data set.

Samples. The standard error for a single set of data is computed by dividing the standard deviation by the square root of the number of points in the data series. It comes into play, through the use of hypothesis testing or confidence interval constructing, when one wants to draw conclusions about the reliability and accuracy of the sampling techniques one is using to estimate a target population's mean. In that case, one selects a representative sample, computes the descriptive statistics, and then uses them to estimate a target population's mean. The beauty of the standard error is that it allows the statistician, researcher or scientist to draw conclusions about the population while only having to draw one sufficiently large (about 30 data points if the population is near bell shaped) sample.

For example, assume a known population of 1,000 price/revenue multiples that average .638 with a standard deviation of .327. Draw 1,000 samples (with re-sampling) of 30 data points each from these 1,000 data points in the population, then compute the average of each sample (total the 30 data points in each sample and divide by 30), and then average the 1,000 sample averages just computed, and compute the standard deviation of the 1,000 sample averages. The results of .639 for the mean of the 1,000 samples and .060 for the standard deviation of the 1,000 samples will approximate the known population average (.638) and the standard error of the known population (.327 / SQRT(30) = .060). Most amazingly, the same results will be obtained from just one sample of 30 observations, i.e., the mean and standard error of that one sample will closely approximate the mean and standard error of the population. This result is known as the control limit theorem. See Exhibit 4 for a demonstration.

Therefore, if all that is known is the sample size, the sample average, and the sample standard deviation, inferences can be drawn, hypotheses tested, and confidence intervals built, without having to draw multiple samples. For example, because of the empirical rule, 95.44% of the sample means will lie within approximately two estimated standard errors, .114 (.057*2), of the unknown population mean. Does this mean that there is a 95.44% probability that the interval .623 ± .114 contains the true average? It does not, as the interval computed from a given sample either contains the true average or it does not. Instead, the level of confidence is associated with the method of calculating the interval.

The confidence coefficient, 95.44% in this case, is simply the proportion of samples of a given size that may be expected to contain the true average. That is, for a 95.44% confidence interval, if many samples are collected and the interval cutoff computed, in the long run about 95.44% of those samples will fall within the intervals. Using this same information, the minimum sample size can be calculated that will be needed, depending on our required confidence level and the standard error of the sample mean.

However, the question facing business appraisers is not what sample size is needed in order to determine:

EXHIBIT 4
Demonstration of Control Limit Theorem

Item Count	Population Price/Revenue	Sample Average	1,000 Samples of Size 30 1	2	3	4	5	6	7		28	29	30
1	0.100	0.650	0.584	0.285	1.091	0.821	0.322	0.842	0.305	..	0.130	0.857	0.174
2	0.100	0.636	1.164	0.191	0.265	0.305	0.629	0.874	0.371	..	0.702	0.788	1.201
3	0.100	0.663	0.186	0.685	0.907	0.870	0.911	0.874	1.005	..	0.943	0.587	0.670
4	0.100	0.616	0.219	0.572	1.243	0.781	0.117	0.474	0.818	..	0.150	0.457	0.383
5	0.105	0.652	0.874	0.653	0.956	0.670	0.474	0.764	1.054	..	0.747	0.825	0.599
6	0.105	0.630	0.395	0.960	0.265	0.415	0.452	0.375	0.182	..	1.022	1.145	0.682
7	0.105	0.665	0.850	1.145	0.572	1.095	0.243	0.788	0.535	..	0.641	1.066	0.231
8	**0.105**	**0.623**	**0.732**	**0.968**	**0.162**	**0.768**	**0.105**	**0.162**	**0.191**	**..**	**0.338**	**0.510**	**1.071**
9	0.108	0.605	0.341	0.870	0.719	0.498	0.764	1.169	0.199	..	0.813	0.985	0.830
10	0.108	0.668	0.523	0.747	0.793	0.560	0.665	0.452	0.452	..	0.575	0.154	0.756
11	0.108	0.610	0.612	0.203	1.054	0.408	0.665	0.125	0.206	..	0.919	0.117	0.547
12	0.108	0.616	1.108	1.083	0.641	0.854	1.120	0.560	0.395	..	0.130	0.400	1.005
13	0.113	0.679	1.206	0.383	0.474	0.572	0.956	1.029	0.830	..	0.474	0.265	1.238
14	0.113	0.682	0.100	1.120	0.203	0.919	0.228	0.513	0.329	..	0.469	0.985	0.629
15	0.113	0.776	1.169	0.167	0.747	0.646	1.034	0.830	1.034	..	0.850	1.132	0.157
16	0.113	0.454	0.771	0.302	0.427	0.665	0.395	0.326	0.732	..	0.727	0.167	0.636
17	0.117	0.400	0.424	0.555	0.830	0.580	0.334	0.830	0.277	..	0.887	1.206	0.305
.		.	.		.	.	.	.	.	..		.	.
.		.	.		.	.	.	.	.	..		.	.
990	1.238	0.692	0.722	0.526	0.805	0.842	0.457	0.764	1.059	..	0.722	0.560	0.690
991	1.238	0.705	1.120	0.854	1.226	0.596	1.017	0.243	0.403	..	0.567	0.506	0.690
992	1.238	0.645	0.338	0.228	0.538	0.658	0.498	0.854	0.857	..	0.461	0.305	1.226
993	1.243	0.680	0.461	0.145	0.850	1.054	0.821	0.481	1.221	..	0.105	0.776	0.907
994	1.243	0.755	0.616	0.609	0.751	1.017	0.673	0.943	0.735	..	0.383	0.489	0.469
995	1.243	0.557	0.341	0.162	0.486	0.624	0.747	0.833	0.506	..	0.621	0.481	0.882
996	1.243	0.604	1.022	0.449	0.292	0.661	1.234	0.690	0.874	..	1.181	0.302	0.322
997	1.246	0.628	0.572	0.948	1.017	1.226	0.518	0.580	0.616	..	0.461	0.854	0.862
998	1.246	0.515	0.661	0.130	0.621	0.297	0.108	0.427	0.948	..	0.223	0.366	1.189
999	1.246	0.674	0.375	0.194	0.535	0.587	1.152	0.857	0.130	..	0.592	0.518	0.378
1000	1.246	0.687	1.078	1.140	1.194	0.415	0.100	0.167	0.216	..	0.801	1.206	0.661

Population	
Minimum	0.100
Maximum	1.250
Mean	0.638
Standard Deviation	0.327
Standard Error (sample size 30)	0.060

Samples	
Average of 1,000 samples	0.639
Average of 1 sample	0.623
Standard Deviation of 1,000 samples	0.060
Standard Error of 1 sample	0.057

- Who will win the next election based on a sample of voters selected today?
- What is the average expected tread life of the latest model steel-belted radial tire based on a selected sample that is driven 40,000-plus miles?
- Does the average cereal box being filled contain at least 15.7 ounces and no more than 16.3 ounces based on a sample selected from the production line?

What do these survey-type questions, or their answers, have to do with valuing a company using pricing multiples from the transaction databases? Not much, as there is little if any interest in the population as a whole—the true interest lies with the data set at hand and the value of a particular company. The sample size determination is not really germane to the issue at hand, which is: How can a value be put on the subject company using the available database information?

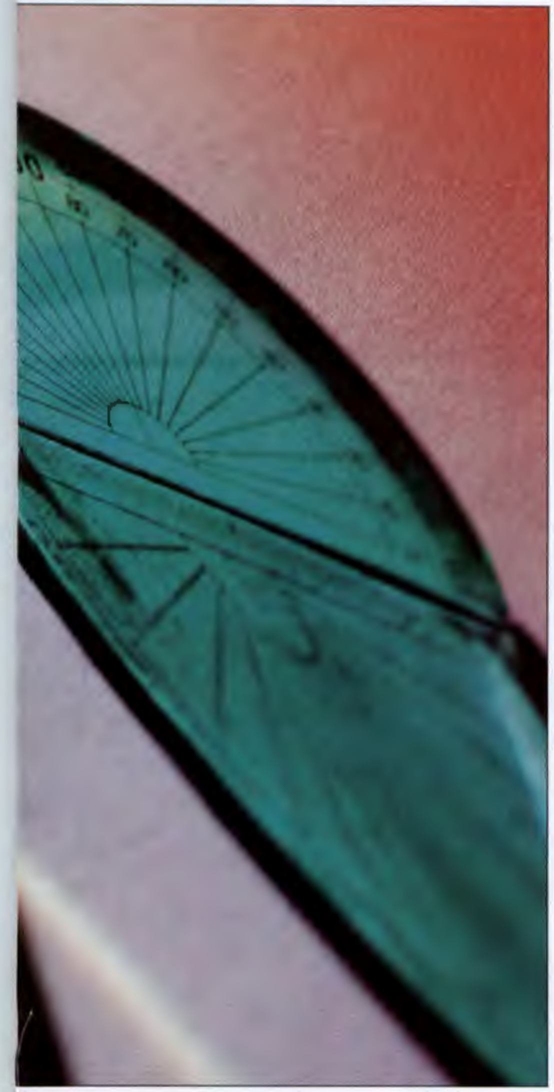
This problem is more akin to that of the restaurant owner who owns pizza parlors located on the outskirts of campuses in 20 different college towns located in the Midwest. Based on the data collected from those locations, how can the owner compute the revenue expected from the 21st store? A more simple question can be answered: How much does a particular type of advertising (radio, television, flyers, etc.) drive sales, and what level of combined advertising is needed to reach a certain sales goal? In sum, the strength of the relationship between two or more variables is what is of ultimate interest, and what that relationship can tell the owner about the expected revenue from the 21st store, the expected revenue produced by any combination of advertising, and for valuators, the estimated selling price of the subject company based on its SDE.

Further, the standard error of the x coefficient does not pertain to the accuracy of the value prediction, as Wolpin claims, but to the accuracy of the sampling process. Invoking margins of error as a measure of unreliability regarding the accuracy of a valuation point estimate misunderstands the use of the standard error. In regression analysis, a low standard error for the s coefficient does not compensate for a large variance about the trend line—there is still a great deal of dispersion, and therefore the concluded value will still be subject to the empirical rule as driven by the standard deviation of the trend line, the SEE, not the standard error of the x coefficient. This occurs because the SEE does not change (except for sampling error) no matter the size of the sample, while the standard error of the x coefficient must vary for purely mathematical reasons as the sample size changes—the larger the sample the smaller the standard error, and vice versa. This is the fundamental error Wolpin makes again and again—he confuses large sample sizes with the absence of sample bias. He also argues for its converse, that low standard errors are evidence of reduced dispersion. He is incorrect on both points—if there is sample bias, it exists in both large and small samples, and it is the SEE that drives the dispersion metric, not the standard error of the x coefficient.

Normal Distribution. In his explication of regression model assumptions, Wolpin correctly states that the x and y variables must be linearly related for a regression to be valid. He then goes on to claim that the only way for the x and y variables to be linearly related is for the variables to be normally distributed. Otherwise, he says, the relationship is not linear, and some other tool besides simple linear regression must be used to describe the relationship between x and y. While it is true that other tools must be substituted if the relationship is nonlinear, it is not true that the sample variable distributions themselves must be normal for them to be linearly related. In fact, only the residuals need be normally distributed. The test for the x and y variables is not one of normality, but of co-variance. If x and y co-vary to a high degree, they can each be nonnormally distributed, but they will be linearly related, and linear regression will work just fine, as can be demonstrated by any of the standard residual analysis techniques.

Once more turning to outliers, one type of outlier that does not necessarily have to be removed is what is called an influential outlier. Wolpin conflates an influential outlier with a leverage point, but they are not the same thing. An outlier that is consistent with the main body of x values is neither influential nor a leverage point, as the main body of data points will offset its dominating power. An outlier that is outside the main body of x variables, and whose y variable does not lay on the plane of the trend line, is a leverage point. An outlier that is far removed from the main body of x variables, but whose y variable does lie on the plane of the trend line, is influential. The key test for these outliers that survive the 2.5 standard deviation culling process is to see how much the x coefficient changes when they are removed. Most leverage points will fail this test and have to be removed, but not those influential outliers that lie on the plane of the trend line.

Wolpin attempts to set the minimum size of a data set for a valid regression analysis at 50. However, there are as many "minimum" sample sizes as there are statistics textbook authors. Once more, context is everything—it all depends on what one is trying to



Invoking margins of error as a measure of unreliability regarding the accuracy of a valuation point estimate misunderstands the use of the standard error.

accomplish, how much data there is to work with, how that data is distributed, and to what degree x and y co-vary. Wolpin assumes that valuators are concerned with "precision" and therefore, with making inferences about the population mean, when all they are trying to do is value a subject company with some degree of accuracy. There is much imprecision in business valuation, whether it is valuation of publicly traded or privately held companies. As will be demonstrated later, it does not matter whether a market approach or an income approach is employed—imprecision is omnipresent.

Another sign that Wolpin has either willfully misconstrued accepted valuation methodology or is completely confused about the topic is where, in one article,[5] he took the need and desire of valuation analysts to reconcile the value conclusions derived from two or more different approaches and suggested that matching up the result obtained by way of the databases with an income approach is a form of circular reasoning. However, he is confused, and his confusion results from a failure to distinguish between two quite different logical constructions: equivalence and tautology. Equivalence is a pure $A = B$ construction. Under Wolpin's statement, $6 = 2 \times 3$ would also be a tautology. It is not, however; it is a statement of equivalence.

In contrast, a tautology is essentially a logical argument or proposition of causation (or, if you prefer, premise and conclusion) with two parts: (1) an if...then...relationship, such as "If A, then B," coupled with (2) a definition of A (the cause or the premise) by reference to B (the effect or the conclusion). If that were the case, all value reconciliations would be tautologies, or circular references. However, they are not tautologies; they are equivalents, as the concluded values under each approach do not depend on each other, a point Wolpin misses.

As to the location of the selling business, it has been demonstrated by Toby Tatum and Ray Miles that the macro location of a business (east coast vs. midwest) has no (or minimal) impact on its pricing multiples.[6] Therefore, there is nothing to be gained by segmenting a database by geographical location. Wolpin agrees with this, but instead he argues that the micro location of a business (what corner it is on) is a significant value driver that the databases do not capture as distinct from its management, or any other positive attribute it may possess. However, SDE, by definition, will capture all of those value drivers in total, and so a particular business' superior (or inferior) location, for instance, will help determine its selling price. The fact that sometimes a regression model will produce low r^2s and large SEEs is not necessarily the result of the databases' failure to capture all the value drivers of all the transactions. Again, it could be just random error. However, valuation analysts are advised to avoid using a data set with these characteristics.

Wolpin claims that small business market pricing multiples fluctuate with changes in interest rates and the financial cycle and over time as well. However, Tatum shot down both of these ideas quite effectively.[7] Furthermore, Wolpin's explanation of his Exhibit 8 created doubt about whether he really understands what the databases are all about and what information they convey.[8] If the x-axis of his Exhibit 8 was in years, and the time scale began before Lowes and Home Depot opened, his observation about declining multiples over time would be correct. However, the scale is in months, and it just reflects what happens in a typical market, both for residential and commercial real estate as well as for businesses—the price drops over time. The longer the asset is on the market, the lower its price must decline until it hits the market-clearing price.

Wolpin argues that the high influence data point in his Model 1 ought to be removed.[9] Well, which is it? Outliers should not be removed or they should be removed. After the removal of that one leverage point, he notes that the coefficient for the intercept, or constant, of the model changed from negative to positive, something he claims is an indication of a model's usefulness. Once more, Wolpin is confused. The intercept is the value the model produces when the x coefficient is zero. While the intercepts in both Model 1 and Model 2[10] are the best, unbiased estimates of the predictions based on those data, it is clearly a value in both cases that cannot be interpreted, because no business in the data set had annual revenue of zero. The intercept is that amount necessary to insure that the slope of the trend line goes through the point where the average of y meets the average of x. Therefore, it usually has no meaning.

For example, assume a data set that has five years of sales as its y variable, and years 1 through 5 as its x variable. If the years are coded as 2001, 2002, 2003, 2004, and 2005, the regression model will produce a negative intercept. However, if the years are coded 1, 2, 3, 4, and 5, the model produces a positive intercept. In either case, the x coefficient will be the same, and choosing either x variable (e.g., 5 or 2005) will give the same predicted value. Therefore, a negative intercept does not have a negative connotation—in fact, it has no connotation at all.

Size Segmentation

In his latest article, Wolpin urges valuators to avoid size segmentation of the data set in order to solve the problem of nonlinearity, a suggestion with which the present author agrees. However, there is no need to use complicated techniques like spline fitting and loess curves[11] to account for the quantitative and qualitative differences between a $1.7 million beauty shop and a $170,000 beauty shop. These quantitative and qualitative differences between small and large businesses, even those in the same industry, can easily be accounted for in a standard regression model. While it has been suggested that a size factor be used when limiting the range of the x variable, this is an unnecessary elimination of observations. Because regression analysis is all about relationships, a too large range of the x variable should not matter as long its relationship to the y variable can be maintained regardless of size. As the present author has demonstrated elsewhere,[12] the proper use of dummy variables or quadratic models and transformation of the x or y variable will allow for the use of all of the nonoutlier data points in a transaction data set regardless of the range of x values.

Wolpin recognizes this, at least to the extent of the use of transformations. He sets out to demonstrate this technique by using SIC Code no. 7231—Beauty Shops, with annual revenue as the x variable. Unfortunately, even with more than 700 data points, neither the x nor the y variable is normal, which fact contradicts one of Wolpin's continuing assertions—that SIC Code numbers with large samples, being by definition actively traded markets, must be normally distributed. In this case, however, neither the x nor the y variable is normally distributed, even after removing the $7.0MM transaction.

Then again, as previously mentioned, the x and y variables of a data set do not have to be normally distributed to produce a large r^2 and a small SEE, as long as they co-vary with each other, i.e., as long as their standardized variables match up closely. This means that a data set will be perfectly correlated if the distance from the mean (measured in standard deviations) of one variable (x) corresponds exactly to that distance (again measured in standard deviations) from the mean of the other variable (y).

Valuators should avoid size segmentation of the data set in order to solve the problem of nonlinearity.

Distinctly nonnormal distributions can have this co-varying characteristic, but the chance that x and y will co-vary goes up if they are both normally distributed.

The major reason, overlooked by Wolpin, as to why the beauty shop data set is not normally distributed, despite its size, is that beauty shop gross revenue is earned in two different and mutually exclusive ways—either as total sales from clients, with a subsequent commission split with the technician, or as booth rentals from the technicians coupled with the direct sales of the owner. Unfortunately, the IBA database does not allow one to distinguish which method is applicable to each of the reported sales transactions, resulting in a lack of consistency among those that are based on annual revenue. Hence, there is no discernible relationship between all the x variables and the y variable, a problem that transformation cannot fix. This does not prevent Wolpin from trying, and his regression output metrics look pretty good. They do not tell the whole story, however. After removing the $7 million transaction as it is (1) such an anomaly and (2) overstates r^2 for the reason given above, Model 3[13] was run using the same log-log transformation that Wolpin used. What became apparent was that Wolpin was comparing apples to oranges.

The metrics in his Model 3 are expressed in transformed space, which makes them appear better than they actually are, because the transformation process reduces the spaces between the data points. For this reason, the only things that should be done in transformed space are the residuals tests, which Wolpin has correctly presented in that manner. These tests, including those for normality, are done on a regression

model's residuals, not on the y or x variables that make up the data set. If the residuals are normal, it can be assumed that the y variable, if not both variables, is from a random sample. However, r^2, SEE, and COV should be calculated from back-transformed predicted values in natural space. When done so, all three metrics of Model 3 are worse than Models 1 and 2, demonstrating that transformation is not the panacea that Wolpin makes it out to be.

Log-log transformations,[14] as used by Wolpin, rarely work when applied to the transaction databases. They are useful when the relationship between the x and y variables can be expressed as: When x increases by 1%, y increases by b%. This type of relationship can be seen in the value of raw land, where the price per acre drops as the lot size increases. In business valuation, the relationship between SDE or revenue and selling price usually cannot be made linear by a transformation of this type. Therefore, Wolpin's heroic attempt to rescue the beauty shop data set from its intractable state ultimately fails.

Removing outliers greater than 2.5 standard deviations does improve the performance of the revenue model somewhat, as r^2 is now .80, but the wide dispersion in the data set still puts the COV at 60%. Actually, there is no regression model that can handle the database—it is just too obdurate in its current state. The degree of variability and lack of constant variance is insurmountable. Using SDE as the x variable gives much better results, but only after using a cut-off of 2 standard deviations and hence removing 35% of the data points. This is one of those situations in which the analyst might perhaps abstain from the market approach and default to the income approach.

Alas, while this might be the solution when confronted with beauty shop-type SIC Code numbers, it will not solve the problem signaled by the relatively wide spread of the SEE and the resulting large COV (approximately 25%) in almost all the SIC Code numbers represented in the databases if the cut-off metric is 2.5 standard deviations. (Using a lower cut-off metric results in lesser COVs.) This is because even the income approach, whether the single-period capitalization method or the multi-period discounted cash flow method, is afflicted with this large variance problem.

To demonstrate, assume a company has $100,000 of cash flow to equity, that its long-term expected growth rate is 5% and that its capitalization rate is 22%, with standard deviations of $20,000, 1%, and 4.4%, respectively. Simulating the Gordon growth model 10,000 times produces an average value of $449,426 with a standard deviation of $158,816. This gives us a COV of 31.8%, a percentage that exceeds the COVs typically produced by the databases. This is not surprising, given that the average growtb in GDP since 1929 is 6.73% with a standard deviation of 7.12%, and the value-weighted COV for the New York Stock Exchange for the period 1926-2007 is 172% as shown in the Ibbotson SBBI yearbook.[15] These results make the variability in the databases appear modest.

In the conclusion to his latest article, Wolpin returns to the idea that the transaction databases contain SIC Code Segment data sets that are unverifiable, poorly compiled, and nonrandom, and therefore do little to prove value. It is important to point out that one can hardly expect a normal distribution if the data are not independent and random. However, one is allowed to reason backwards peofitably. If independence and randomness are necessary conditions for a linear regression, one can assume that the resulting residuals, that distribute themselves into a bell curve, come from observations that are independent of one another and random in nature. This point is sufficient to reject Wolpin's notion that the data sets in the transaction databases are nonrandom samples, *(Continued on page 48)*

Transaction Databases

(Continued from page 23) or that they cannot easily be made normal just by the removal of outliers, among other devices.

Sample Bias

Along the same lines, Wolpin claims that the databases are statistically unreliable because they suffer from nonsampling error, or sample bias. Sample bias is claimed to exist since sales of small businesses fall into three categories:

1. Those sold by business brokers and reported to the database purveyors.
2. Those sold by business brokers and not reported.
3. Those not sold by business brokers and therefore not reported.

Therefore, Wolpin claims that eliminating sample bias would require the valuation analyst to use transactions drawn from all three sources. As previously discussed, if the data collected is in fact essentially random, the sample bias issue will go away. For now, however, assume that Wolpin's assertion that the databases suffer from sample bias is valid. Technically, they do suffer from sample bias, because they all consist of convenience samples, which are samples collected in a particular manner because it is convenient to do so.

For example, a restaurateur who devises a new lobster roll for the summer season asks the opinion of the first 50 customers who try the sandwich on a given day. Unless he has some reason to believe that these 50 customers are somehow radically different from the entire population of all customers, it is probably safe to treat them as a random sample. Or if you are asking about evening television viewing habits, people who are not at home in the evenings are likely to be quite different from people found at home at that time. However, if you are interested in which of two spaghetti sauces they prefer, it may be more reasonable to believe that at-homes can be treated as nearly, if not exactly, representative of not-at-homes.

If the databases are convenience samples, is there any reason to believe that they are not representative of the total population of sales of small businesses at arm's length to third parties? Wolpin presents no evidence that sales transactions that are not reported to the database owners are any different from those that are. In fact, just the opposite is more likely to be true—the private sales that the present author has been involved in were priced at the same multiples as those derived from the databases. Could this be because multiples from the databases produce selling prices that coincide with other valuation approaches and methods? Can all 80,800 sales of private businesses a year really be transacted at incorrect prices?

Wolpin goes on to state that nonsampling error cannot be corrected by increasing the sample size, i.e., bias is bias. But he then claims that the most actively traded SIC Code No. segments of the databases do not suffer from nonsampling error, and at the same time, he defines actively traded as sufficiently large. Aside from being a self-contradictory argument, the beauty-shop data set shows that being a large SIC Code segment is no guarantee of being an unbiased sample. This goes along with all the good things that flow from such a state, such as being random, symmetrical, or normal, with a high degree of co-variance between selling price and revenue or SDE. Therefore, it must be assumed that the transactions included in the databases are just samples available to be statistically investigated as to whether they enjoy those qualities.

Conclusion

In his articles, Wolpin engaged in *ad hominem* attacks against the purveyors of the databases, in which he practically accuses them of being snake oil salesmen, selling something to the business valuation community that they know is not helpful. He charges them with pursuing this nefarious end out of pure greed, as they should know full well that the databases are unreliable for their intended purpose.

Even if Wolpin was right about the databases, it could be that no one was aware of this information until he came along, and that the whole business valuation community was laboring under a delusion. Perhaps the database purveyors really thought they had something useful to sell, and it was only out of ignorance, not greed, that they sold their "damaged" goods. Wolpin should be a little more charitable in his personal assessments of people's motives, and he should confine his attacks to the valuation and statistical theories at hand.

So where does this leave valuation professionals? Where they were before Wolpin wrote his series of articles, but a little wiser; in that they have now been informed that the proper tool to apply to the transaction databases is regression analysis; that outliers may be removed, and that the income approach produces no better a value conclusion than does the direct market data method that relies on the databases. ●