Statistical Reasoning in Family Limited Partnership Appraisals:

New Tax Court Scrutiny

By Jeffrey B. Wolpin

In two recent cases, *Lappo v. Commissioner*, T.C. Memo 2003–258, and *Peracchio v. Commissioner*, T.C. Memo 2003–280, the Tax Court criticized business appraisers for their inadequate statistical reasoning. In *Lappo*, the court noted:

> We are not persuaded that [the taxpayer's appraiser's] guideline group is sufficiently large or made up of companies sufficiently comparable to the partnership. "While we have utilized small samples in other valuations contexts, we have also recognized the basic premise that '[a]s similarity to the company to be valued decreases, the number of required comparables increases.'"

T.C. Memo 2003–258, at 11; see also *McCord v. Commissioner*, 120 T.C. 358 (2003). In this instance, the taxpayer's appraiser selected only seven guideline REITs for determination of the discount, whereas the IRS appraiser selected 52 REIT comparables.

In *Peracchio*, the court noted:

> [The taxpayer's appraiser] calculates the mean (average) discount and the median (midpoint) discount with respect to each of his fund samples. In each instance, the median discount is greater than the mean discount. [The taxpayer's appraiser] opts to use the median, rather than the mean, discount with respect to each sample for purposes of determining a minority interest discount factor for each corresponding asset category of the partnership. At trial, he testified that medians "in my opinion are often more relevant [than means] because it takes outliers out of the equation."

T.C. Memo 2003–280, at 20. The taxpayer's appraiser's written report, however, suggests that he may have accounted for outliers (by excluding them from his samples) before determining sample medians: "After adjusting for outliers and asset homogeneity with the subject assets, we then calculated a weighted average median discount." In any event, the taxpayer's appraiser eventually conceded at trial that he did not have a good reason to use the median, and that either one— the median or the mean—could have been used. The court used the mean discount as the minority interest discount factor, finding that it was "more straightforward" to account for outliers by excluding them.

The courts in *Lappo* and *Peracchio* searched for, and perhaps demanded, better statistical reasoning from business appraisers. This is long overdue. Unfortunately, the business appraisal community has a very poor understanding of statistical reasoning. The business appraisal texts and literature

---

Jeffrey B. Wolpin, CBA, CREA, CCRA, MBA, is the principal in Accredited Business Appraisals in Santa Clarita, California.

are rife with examples of poor statistical logic. Consequently, the business appraisal community, as in the examples of the above two cited cases, frequently asserts valuation theories without the empirical evidence and statistical reasoning to back them up. Even when empirical evidence exists, it is often distorted or used out of context.

It is with good reason that the courts in *Lappo* and *Peracchio* seemed skeptical of business appraisers, their logic, and their conclusions. Courts are insisting upon more than mere speculative arguments and theory about business valuation and are requiring stronger statistical expertise from those who hold themselves out as business valuation experts. This is not to say that the court is itself without error. Although the above statements are on the right track in demanding more concrete statistical reasoning, the courts' own statistical logic also may have been flawed.

For instance, in *Peracchio*, the taxpayer's appraiser apparently admitted that he did not understand the differences between the mean and the median and ended up in essence twice eliminating outliers. Although the court correctly identified this error, it may have been incorrect in eliminating what it defined as "obvious outliers." In so doing, the court may have unwittingly violated a statistical principle of outlier elimination *only* if there is sufficient and substantive evidence to identify apparent outliers as outliers in fact. See Gerald Keller & Brian Warrack, *Statistics for Management and Economics* 670 (5th ed. 1999). In addition, there is some weakness to the *Lappo* court's argument that "as similarity to the company to be valued decreases, the number of required comparables increase[s]." The court correctly determined that small samples of highly similar comparables are statistically acceptable and that a direct relationship exists between the sample size and the strength and quality of the statistical conclusions from the sample. Whether increasing the sample size with additional dissimilar comparables improves the strength and quality of the conclusions is debatable, however.

Attorneys hiring business appraisers should require sound expertise in both statistical reasoning and market insight. The two go hand in hand. One cannot apply statistical techniques absent market insight, and vice versa. Similarly, one cannot perform a valuation based solely on parroting text book theories.

Statistics 101

Some statistical definitions are in order at this point.

The *mean* is the sum of the observations divided by the number of observations. It is the arithmetic average of the measurements in a data set. A data set can have only one mean and its value is influenced by extreme outliers or measurements. Trimming outliers can help reduce the degree of influence outliers have on the mean. By trimming outliers, however, one may end up falsely eliminating data or introducing bias. Statisticians combine means of subsets to determine the mean of the complete data set. See R. Lyman Ott & Michael T. Longnecker, *An Introduction to Statistical Methods and Data Analysis* 77 (5th ed. 2001).

The *median* is the value that falls in the middle of a set of observations when the observations are arranged in order of magnitude. It is the central value, where 50% of the measurements lie above it and 50% lie below it. There is only one median for a data set and it is not influenced by extreme outliers or measurements. One cannot combine medians of subsets to determine the median of the complete data set. The median is more robust than the mean as well as more resistant to erratic or extreme observations, although it is inferior to the mean with symmetrical distributions not far from normal.

Many appraisers regularly overlook the fact that one should express both the mean and the median in ranges or confidence intervals, particularly when dealing, as in business and real estate appraisals, with uncertain data and probabilities. The range or confidence interval indicates the level of uncertainty contained within the data. For example, if a survey of average starting attorney salaries reveals, with 95%

confidence, a range of $50,000 to $150,000, the interval is so wide that little information can be derived from the data. On the other hand, if the range is narrower, say between $70,000 and $80,000, one has more precise information about average starting salaries. The *Peracchio* court intrinsically realized this when it discussed values in terms of ranges, "because valuation necessarily involves an approximation, the figure at which we arrive need not be directly traceable to specific testimony if it is within the *range* of values that may be properly derived from consideration of all the evidence." The range (more properly called the interval for the mean and the median) is expressed in terms of a statistical formula and shown graphically.

Calculating the Confidence Interval for the Mean

Most of us are familiar with the formula for the mean's interval during an election season, when the media talk of polls and their respective predictive ability within a plus or minus margin of error. In the business appraisal context, as in the election context, the actual statistical formula for the mean interval is

$$E = z_{\alpha/2}(\sigma \div \sqrt{n})$$

where E is the margin of error, $z_{\alpha/2}$ is the desired probability or confidence level, σ is the standard deviation or variation of the data set, and n is the sample size. Z is the notation for the normal probability distribution, or "bell curve," which most of us are familiar with from high school and college grading. The normal probability distribution is most appropriate for sample sizes over 30. When one is faced with a sample size under 30, however, a different probability distribution (referred to as the student-*t* or *t* probability distribution) is substituted, which recognizes the greater variability intrinsic to smaller sample sizes.

For example, to compute the margin of error with a 95% confidence level, on a data set with 11 items, where the average or mean is 7.88%, and the variation or standard devia-

tion is 4.49%, the above formula would indicate the following:

Data Set	Yield
1	6.44%
2	4.44%
3	10.17%
4	10.87%
5	10.40%
6	10.00%
7	10.98%
8	0.00%
9	11.57%
10	0.00%
11	11.76%
Average	7.88%
Standard Deviation	4.49%
Mean Margin of Error or	
Half Width (+ or -)	3.01%

In other words, excluding all other qualitative information, given the data set size and its standard deviation, we can be 95% confident that the mean of 7.88% is accurate to within ±3.01%. In this example, ±3.01% represents a 38.2% margin for error. This is a very wide range and very low level of accuracy for the data set. The margin of error (also known as the half width) can be reduced in one of two ways: by reducing the confidence level or by increasing the sample size, or both. This is expressed in the algebraic relationship of the mean interval formula by solving for sample size *n:*

$$ n = (z_{\alpha/2} \cdot \sigma \div E)^2 $$

The above formula indicates that the required sample size (*n*) is a function of the sample variability (σ) and the required margin for error (*E*), given a selected confidence level.

If, in the above data set example, one reduced the margin of error from 38.2% to 10%, solving for sample size, the minimum required sample size must increase from 11 to 125.

Minimal Sample Size =	125.0
Desired C.I. =	95%
Estimated Std. Dev. =	0.0449
10% Margin for Error (Half Width) =	0.788%

Alternatively, if the data set has a smaller standard deviation, the margin of error likewise would be smaller. The data and its standard deviation, however, speaks for itself. The appraiser can only adjust the confidence level or increase the sample size.

These formulas do not address how qualitative facts about the subject to be valued affects or is affected by the comparable data. The qualitative characteristics of the subject may determine where in (or even outside) the range the subject may belong relative to the comparable data. For instance, the qualitative aspects of a family limited partnership to the comparables may include such items as the leverage, management experience, performance, quality of assets, and partnership agreement restrictions on control and marketability. Many family limited partnership agreements contain more transfer restrictions than the agreements of the public limited partnerships or REITs against which they are often compared, or they have inferior assets and management performance. Sometimes, the reverse may be true. Consequently, the comparable data's range, mean, or median may not reflect the range appropriate to the family limited partnership. Family partnerships may deserve discounts or yields that are at the high or low end or possibly even outside the range seen in the comparable data because of these qualitative characteristics. Qualitative characteristics are, by nature, subjective, permitting room for differing opinions. As a result, the strength of these qualitative arguments may determine where inside or outside the range the subject belongs. Lastly, the wider the comparable data's range, the more room there is for divergent appraisal opinion. In these cases, the strength of the qualitative arguments in determining where inside or possibly even where outside the range the subject should fall becomes all the more critical.

Rule of Thumb Method for Calculating the Confidence Interval for the Median

As with the mean, the median also has a confidence interval or a probability range. There are two procedures for calculating the median's confidence interval: a simple rule of thumb method and a more elaborate method called the Sign test. The Sign test is a nonparametric test for small sample sizes—that is, the test is "distribution free" in that the statistician does not assume or require a normal or symmetrical distribution assumption to be valid. Because the Sign test is complex, it is best performed through readily available computer programs.

The formula for the rule of thumb method for calculating a conservative confidence interval for the median is as follows:

$$ (n + 1)/2 \pm z \sqrt{n}/2 $$

where *n* is the sample size and *z* is the normal standard deviation corresponding to the desired confidence probability. Taking *z* = 2 for 95% limits, these numbers simplify to

$$ (n + 1)/2 \pm \sqrt{n} $$

See A.M. Mood & F. A. Graybill, *Introduction to the Theory of Statistics* 480 (2d ed. 1963). When applying this formula, the procedure calls for the lower limit to be rounded down and the upper limit rounded up to the nearest integers. These limits, however, are conservative; the exact confidence probability varies with the sample size.

The term "conservative limits" means that the confidence level or probability is generally greater than 95%, which in turn means the actual interval range is wider than it would ordinarily have to be for the more accepted 95% confidence level. Generally, for purposes of appraisal, confidence levels between 90% and 95% are acceptable. As a rule of thumb, when the sample size is less than 15 with 95% limits, the confidence probabilities actually average 97.5%; for sample sizes between 15 and 50, they average 96.8%. See George W. Snedecor & William G. Cochran, *Statistical Methods* 137 (8th ed. 1989).

Another reason why this formula is considered a rule of thumb method is that the *z* = 2 value is used as a rounded substitute for the critical value $z_{\alpha/2}$, which is usually 1.96 at the 95% confidence level on the normal curve where 1.96 represents a two standard deviation distance from the mean, containing 95% of all values (the normal stan-

dard deviation corresponding to the desired confidence probability). The rounded value makes the calculation and the simplification of the algebraic formula easy. This rule of thumb works best for sample sizes of 50 or 100 or more, as the rounding has less effect on the formula result for large sample sizes, and such large sample sizes tend toward symmetry.

Applying the rule of thumb formula to the example results in the following:

$$(11 + 1)/2 \pm \sqrt{11} =$$
$$6 \pm \sqrt{11} =$$
$$6 \pm 3.32 = (2.68, 9.31)$$

Lower Confidence Interval (rounded down) = 2, the second number in the ranking.
Upper Confidence Interval (rounded up) = 10, the tenth number in the ranking.

Ranked Data Set	Yield
1	0.00%
2	0.00%
3	4.44%
4	6.44%
5	10.00%
6	10.17%
7	10.40%
8	10.87%
9	10.98%
10	11.57%
11	11.76%
Median	10.17%
Lower CI	0.00%
Upper CI	11.57%

In this data set example, the mean's margin of error at ± 3.01 is quite similar to the median's margin of error at ± 3.32, underscoring the important influence the standard deviation has on the margin of error for each.
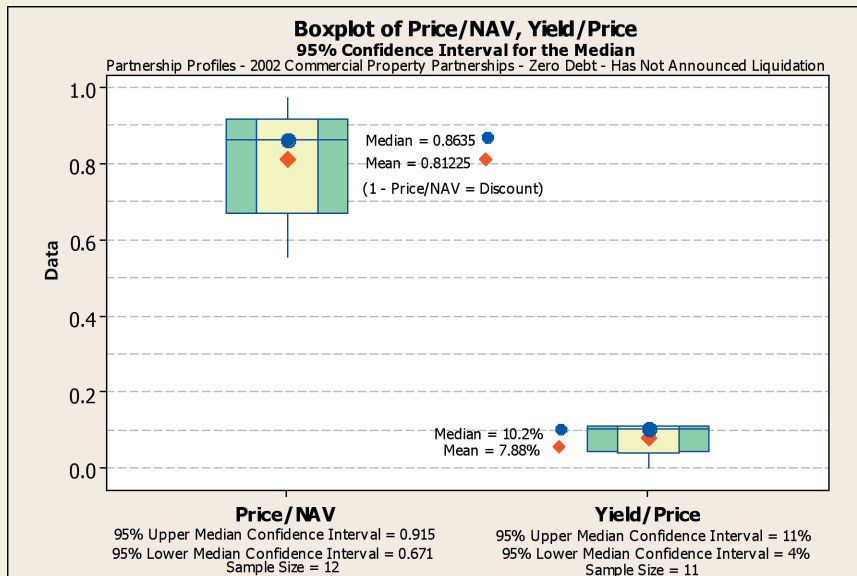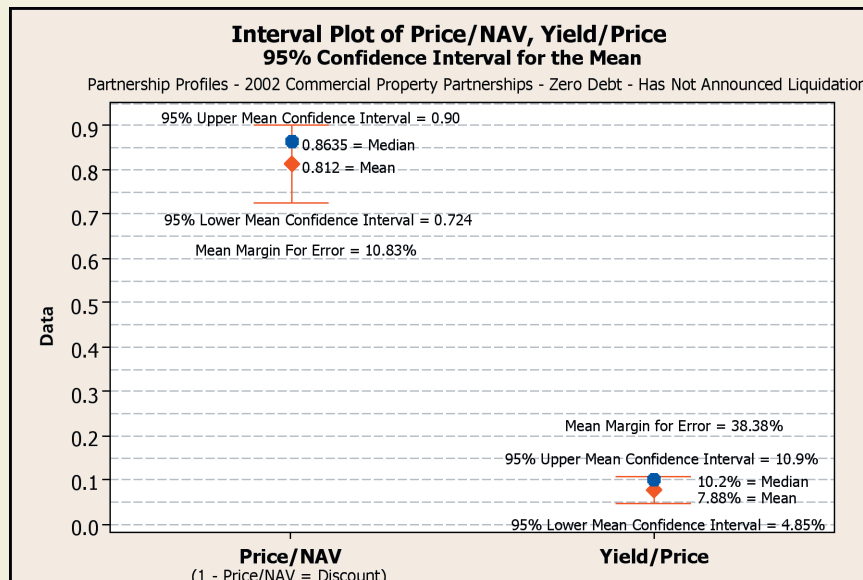
The Sign Test Method for Calculating the Confidence Interval for the Median

The Sign test is the more exacting and preferred method for calculating the median confidence level. This is a non-parametric ranking method for testing a median from small sample sizes. A similar procedure called the Wilcoxon signed rank test can also be used.

Parametric tests are used to test parameters that conform to the strict condition of a normal probability distribution. Means are just such parameters. "Parametric" implies that a symmetrical distribution is assumed for

the population. Under the Central Limit Theorem, a well-established statistical rule, when sample sizes increase, the distribution tends toward normality or symmetry. Briefly, the Central Limit Theorem holds that when sample sizes exceed 30, the distribution tends to be normal or symmetrical. Thus, as sample sizes increase, the need to discriminate for

normality decreases, and testing for a mean tends to be the more appropriate test. See R. Dennis Cook & Sanford Weisberg, *Applied Regression Including Computing and Graphics* 130 (1999). Small sample sizes frequently violate this assumption, particularly if they are skewed or exhibit extreme outliers. In addition, large sample sizes are not, in and of themselves, a guarantee that

Interval Plot of Price/NAV, Yield/Price
95% Confidence Interval for the Mean
Partnership Profiles - 2002 Commercial Property Partnerships - Zero Debt - Has Not Announced Liquidation

- 95% Upper Mean Confidence Interval = 0.90
- 0.8635 = Median
- 0.812 = Mean
- 95% Lower Mean Confidence Interval = 0.724
- Mean Margin For Error = 10.83%
- Mean Margin for Error = 38.38%
- 95% Upper Mean Confidence Interval = 10.9%
- 10.2% = Median
- 7.88% = Mean
- 95% Lower Mean Confidence Interval = 4.85%

Price/NAV (1 - Price/NAV = Discount)   Yield/Price

Boxplot of Price/NAV, Yield/Price
95% Confidence Interval for the Median
Partnership Profiles - 2002 Commercial Property Partnerships - Zero Debt - Has Not Announced Liquidation

- Median = 0.8635
- Mean = 0.81225
- (1 - Price/NAV = Discount)
- Median = 10.2%
- Mean = 7.88%

Price/NAV
95% Upper Median Confidence Interval = 0.915
95% Lower Median Confidence Interval = 0.671
Sample Size = 12

Yield/Price
95% Upper Median Confidence Interval = 11%
95% Lower Median Confidence Interval = 4%
Sample Size = 11

Computer Output: Sign CI: Price/NAV, Yield/Price Sign confidence interval for median:

	N	N*	Median	Confidence Achieved Confidence	Interval Lower	Upper	Position
Price/NAV	12	0	0.8635	0.8540	0.7860	0.9090	4
				0.9500	0.6710	0.9156	NLI
				0.9614	0.6300	0.9180	3
Yield/Price	11	1	0.1020	0.9346	0.0440	0.1100	3
				0.9500	0.0404	0.1105	NLI
				0.9883	0.0000	0.1160	2
NLI = Non-linear interpolation							

the population distribution is normal or symmetrical. Heavy tailed and heavily skewed samples, regardless of their large size, may indeed be representative of nonnormal population distributions in which parametric and testing for a mean is inappropriate.

Therefore, if the sample is small, heavily tailed, or heavily skewed with extreme outliers—that is, asymmetrical—the data set may not be normally or symmetrically distributed. In these cases, the median is the more reliable measure of central tendency. This is what one frequently finds in business valuation data. Because parametric tests assume a normal distribution, when parametric tests are used against small samples or extreme nonsymmetrical distributions, they may lead to misleading conclusions. The benefit of employing a nonparametric test is that it does not imply an assumption of symmetry for the population. Consequently, nonparametric tests tend to be more robust against violations of the normal or symmetrical population distribution assumption and are more robust and reliable for testing medians.

The Sign test for the median's margin for error is more complicated than the rule of thumb method discussed above. Although its explanation is beyond the scope of this article, its principles and methodology are well established and well accepted. What is important to remember is that medians, much like means, have margins of error and confidence intervals associated with them. And it is these corre-

sponding margins for error and confidence intervals that tell the reliability of both the mean and the median. Therefore, a good appraisal does not omit the margin of error or confidence interval associated with its respective mean and median.

Fortunately, reliable statistical computer programs can quite easily compute and graphically illustrate both the median and mean confidence intervals. For example, putting this article into perspective, the data set used in the examples contains actual data frequently used in family limited partnership analysis. These data represent the 2002 yields seen in public limited partnerships containing commercial properties, as reported by Partnership Profiles. The confidence intervals for both the mean and median from this data were calculated by a statistical computer program. In addition to the yields, the corresponding discounts in the public limited partnerships are graphically exhibited on page 49.

As can be seen by comparing the rule of thumb method to the Sign test method, the rule of thumb method produced confidence intervals of 0.00%–11.57%. These match the 0.00%–11.6% confidence intervals at the conservative 98.8% confidence level produced by the Sign test method. As noted earlier, the rule of thumb method tends to produce overly conservative confidence intervals. The Sign test method, calculated by computer, will also interpolate the confidence intervals down to the 95%

confidence level, which is the more commonly accepted level.

Interpretation

In the above instance, the small sample sizes and the relatively large standard deviation produce confidence intervals for both the mean and the median with great breadth. Consequently, the quality and reliability of any valuation conclusion based solely on the data are poor. Merely stating a mean or a median, absent confidence intervals, is misleading. These numbers, in and of themselves, do not properly indicate the probability of accuracy of the numbers, which is best determined by the confidence interval. In other words, opposing appraisers, given a reasonable qualitative rationale, can identify a value anywhere within the confidence interval and each be statistically correct. Thus, when confidence intervals are wide, the analysis is dependent upon the strength of the qualitative arguments that appraisers bring to the valuation. Alternatively, when confidence intervals narrow, the qualitative arguments become less critical, as the data speak for themselves.

Further illustrating these points, the chart at the bottom left shows the July 2, 1996, REIT data used by the taxpayer's appraiser in the *Lappo* case, about which the court was critical. The chart shows the calculations for the means, medians, standard deviations, and their respective confidence intervals and margins for error. As can be seen, these data also indicated wide variation and margin for error, which undermined the taxpayer's conclusions of value.

Conclusion

In its opinions in both *Lappo* and *Peracchio*, the court asks for better statistical information and knowledge from appraisers appearing before it. This is long overdue. The above statistical procedures described provide the court with the information for which it asks. Given the court's request for better statistical reasoning, it is important for attorneys to query their appraisers about whether they have the requisite statistical knowledge that meets the court's demands. ∎

LAPPO - TAXPAYER REIT DATA AS OF JULY 2, 1996

Name of REIT or LP	Price to NAV	Yield to NAV	Yield to Common	Yield to Total Market Value
Arbor Prop. Trust	58.50%	5.60%	9.66%	9.66%
BRE Properties	110.60%	7.20%	6.48%	6.48%
Catellus Dev. Corp.	79.70%	1.90%	3.55%	2.36%
Cedar Inc. Fund.	69.10%	7.10%	10.32%	10.32%
Meridian Industrial Trust	100.50%	6.60%	8.06%	3.54%
Rouse Co.	84.70%	2.90%	4.35%	3.43%
Shopco Laurel Centre, L.P.	47.20%	0.00%	0.00%	0.00%
Average	78.61%	4.47%	6.06%	5.11%
Median	79.70%	5.60%	6.48%	3.54%
Standard Deviation	22.44%	2.86%	3.68%	3.84%
Mean Upper Confidence Interval	99.37%	7.12%	9.46%	8.67%
Mean Lower Confidence Interval	57.86%	1.82%	1.56%	1.56%
% Mean Margin for Error	26.40%	59.25%	56.15%	69.54%
Median Upper Confidence Interval	103.20%	7.13%	9.84%	9.84%
Median Lower Confidence Interval	55.50%	1.39%	2.60%	1.73%
% Median Margin for Error from Upper CI	29.49%	27.27%	51.85%	177.97%
% Median Margin for Error from Lower CI	30.36%	75.13%	59.88%	51.13%