An Empirical Evaluation of Central Tendency Measures

Robert J. Gloudemans
Partner - Almy, Gloudemans, Jacobs & Denne

Overview

A common concern in ratio studies is on which measure of central tendency to place primary reliance. The median is generally regarded as the preferred measure for monitoring appraisal performance and application of market adjustment factors. The weighted mean is conceptually preferred for Aindirect@ equalization, in which equalization agencies estimate full market value of local jurisdictions for purposes of apportioning state/provincial aid or applying a state/provincial property tax levy or requisition. The conceptual preference for the weighted mean for such purposes derives from the fact that it weights the ratios based on sale price (or appraised value in appraisal ratio studies) and therefore reflects any price-related assessment biases. However, it is acknowledged that the median may be preferred when samples are small, have wide dispersion, or are unduly affected by outliers (see IAAO 1990a and IAAO 1990b). The mean is generally not favored because, like the weighted mean, it can be more affected by outliers while containing no important offsetting advantages.

Although the above framework is helpful, it does not resolve the question of whether the median or weighted mean is likely to prove better for indirect equalization over the long run in terms of accuracy and stability across a variety of assessment jurisdictions, or whether an alternative or hybrid measure may be preferred. Using actual assessment and sales data, this paper analyzes this issue for small to mid-size jurisdictions, where the issue is of paramount concern, through a series of simulation studies.

As further background, two important gauges of a measure of central tendency are accuracy and stability. Accuracy is the ability of the measure to predict the underlying population parameter. In the studies presented here, the underlying population parameters are known and average errors over many samples for each measure of central tendency are computed and compared.

Stability relates to the variance or standard deviation of a measure. Given a certain population, desirably a measure of central tendency would be consistent from sample to sample. Large errors are especially undesirable. For large, normally distributed populations, it is well known that the mean has minimum variance. The variance of the median is similar to that of the mean for very small samples (for samples of one or two they are equal) and increases to approximately 1.25 times that of the mean for large samples (Mood, Graybill, and Boes 1974, 257 and Birch, Sunderman, and Hamilton 1991). However, these relationships assume a normal distribution. When this assumption is violated, as is typically the case in ratio studies, the comparative stability of the mean is adversely affected (Birch 1994).

For purposes of indirect equalization, the weighted mean is conceptually preferred because the objective of the study is to estimate the full market value of each jurisdiction. It can be shown that, assuming no sampling error in any of the three traditional measures, only the weighted mean will produce the correct estimate as long as there are any price-related biases in assessments. (If there are no price-related biases and all three measures are the same, all three will produce the same estimated market value). However, it is also well known that the weighted mean, like the mean, can be unduly affected by outlier ratios, particularly if they occur for high value properties. In addition to comparing the accuracy and stability of the measures of central tendency themselves, this paper also compares the accuracy and stability of estimated full market values produced by the various measures over all property classes.

The major difference between the weighted mean and median (or mean) is that the weighted mean gives equal weight to each dollar, while the other measures give equal weight to each ratio. Consider two commercial sales: one of a small retail store for $100,000 and one of a large shopping center for $5,000,000. The weighted mean gives 50 times as much weight to the latter as the former. The median and mean weight them equally. Does either of these approaches seem optimal? Two alternatives, also evaluated in the paper, are (1) to give half weight each to the weighted mean and median and (2) to weight the ratios based on the square root of the sale price (or proxy thereof).

The Database

Six hypothetical jurisdictions of various benchmark parcel counts were constructed from actual data. All the properties used in the analysis represent actual sales in which the sale had been initially screened as usable for assessment-sales ratio analyses by assessment authorities. The following minimum sale price parameters were used: $1,000 for vacant land, $10,000 for residential, and $25,000 for commercial. With the exception of vacant land for the two largest hypothetical jurisdictions, the data are from counties in the state of New York, where assessment occurs at the municipality level. In order to obtain adequate sales to serve as market value proxies, assessments from each municipality were trended to a median ratio of 100% and pooled at the county level or combined across counties. Vacant land data for the two largest jurisdictions come from two counties in Arizona where there were adequate sales over a three-year period to simulate populations of the desired size. Again, assessment levels were trended to 100%. Once the assessments were trended and the data pooled in this manner, ratios below .40 or greater than 2.50 were excluded from the analysis to prevent the results from being overly affected by extremes. Finally, the samples were randomly thinned to produce the exact benchmark sizes shown below.

Jurisdiction	Total	Residential	Vacant	Commercial
1	500	300	150	50
2	1,000	500	400	100
3	2,500	1,500	800	200
4	5,000	3,500	1,200	300
5	10,000	7,500	2,000	500
6	15,000	10,000	4,000	1,000

The assessment jurisdictions from which the data were drawn were chosen so as to represent typical assessment performance. Exhibit 1 shows sales ratio statistics for the six resulting benchmark jurisdictions after the edits described above. Residential CODs are between 10.0 and 15.0. Vacant land and commercial CODs are between 20.0 and 30.0.

Appendices 1 through 6 show histograms of the sales prices by property type and a scatter graph of the ratios plotted against value, defined as one-half of the assessment value (trended to 100%) plus one-half of the sale price, in order to minimize the Aerrors in variables@ problem associated with using only one or the other as the comparison variable. The graphs will appear typical to ratio study analysts. The residential sales prices depart somewhat from a normal distribution, being skewed more or less to the right. The vacant and commercial sales are highly skewed, being characterized by many modest or mid-range sales and several high value sales. In all, the data appear reasonably representative of many small to mid-sized jurisdictions.

Simulation Studies - Measures of Central Tendency

To simulate real-world ratio studies, it was assumed that 5% of the properties would generate sales usable for ratio study purposes and that, if this yielded less than five sales in any stratum, supplemental sales or appraisals would be added. This would imply, for example, 28 samples in the smallest jurisdiction: 15 residential (.05 x 300), 8 vacant land (.05 x 150 rounded up to 8), and 5 commercial (minimum sample size). Two thousand random samples of the indicated size were then drawn from each property class in each jurisdiction, yielding a total of 36,000 samples (6 x 3 x 2000) of five to 500 sales each.

Each of the three measures of central tendency (median, mean, and weighted mean) were calculated for each sample and evaluated based on two criteria. First, average absolute percent errors over the 2,000 samples were calculated. This gauges the accuracy of the central tendency measures in estimating their corresponding population parameters. Second, the standard deviations of 2,000 central tendency measures were calculated. The lower the standard deviation, the more consistent or stable is the measure of central tendency. Of course, the two measures -- average percent error and standard deviation -- tended to move together, in general, the smaller the average error, the more consistent the estimates. The maximum error for the 2,000 samples was also calculated.

Exhibit 2 shows the results of these calculations. In the large majority of cases, the median ratio produced the lowest average error and also had the lowest standard deviation of the three measures. Interestingly, the median=s superior performance was not related to sample size. In fact, the mean or weighted mean performed better for the smallest jurisdiction in the residential and commercial categories. However, in the other five jurisdictions the median clearly performed best. Exhibits 3, 4, and 5 graph the average errors and standard deviations of the three measures by property type.

Overall, as shown at the bottom of exhibit 2 and as can be confirmed by an inspection of the graphs, there was little difference in the three measures for residential property. However, the median is clearly better behaved for vacant land and commercial properties, which are characterized by several high-value sales and greater dispersion in the ratios.

Simulation Studies - Total Value Estimates

The fact that the median performed better than the weighted mean in terms of estimating its corresponding population parameter does *not* necessarily mean that it will do so in terms of estimating total market value, which in our simulation studies represents the sum of the known sales prices for all properties in each jurisdiction, e.g., the sum of all 500 sales prices used as market value proxies in jurisdiction 1. On theoretical grounds, the weighted mean is preferred for this purpose because it weights each ratio in terms of dollars and thus, unlike the median and mean (which give equal weight to each ratio), is capable of capturing any price-related bias in the ratios. Further, a review of the price-related differentials (PRDs) in exhibit 1 and the scatter graphs in appendix 1 suggest that such biases are sometimes considerable in the present database. On the other hand, as already seen, the weighted mean has more sampling variability than the median in the present database. Which of these two forces will dominate will be evaluated empirically (the unweighted mean was dropped from further consideration). .

In addition, two other estimators of total market value were evaluated: (1) the average of the median and weighted mean and (2) the mean of the ratios weighted by the square root of sale price. We will term these two measures the *median-weighted mean* and the *square root weighted mean*. The rationale of the two measures is to forge a compromise between the conceptual advantages of the weighted mean and practical stability of the median. In effect, first measure, the median-weighted mean, gives equal weight to parcels and dollars.

The rationale of the square root weighted mean (SQRWM) can best be appreciated by contrasting it with the weighted mean (WM), which can be calculated by the formula:

WM = 3 (weight □ ratio)

where weight = SP) 3(SP). Thus, the weighted mean weights each ratio relative to its sale price, e.g., a $500,000 sale has twice the weight of a $250,000 sale. In the square root weighted mean, each ratio is weighted relative to the square root of its sale price:

SQRWM = 3 (weight □ ratio)

where weight = sqrt(SP)) 3(sqrt(SP)). Thus, higher value sales are given extra weight, but not to the extent they are in calculation of the weighted mean. For example, a $500,000 sale will receive 41% more weight than a $250,000 sale:

sqrt(500,000) / sqrt(250000) = 707 / 500 = 1.41.

This prevents high-value sales from overly dominating the analysis. Further, there is an inherent balance in the measure since, as is well know, the reliability of central tendency measures increase approximately in proportion to the square root of sample size. Thus, the leverage afforded several high-value sales will be balanced by the stability of the more numerous low to mid-value sales. Neither will control the result as completely as sole reliance on the median or weighted mean.

Exhibit 6 shows the mean (absolute) percent error and standard deviation of the four measures by jurisdiction and property type. The average errors were computed by the formula:

Ave % Error = 100 x □EMV - TMV□/TMV

where TMV = total (known) market value and EMV = estimated market value based on dividing total assessed value for the property class by the central tendency measure. Standard deviations were computed by the formula:

Standard Deviation = 100 x StdDev (EMV/TMV).

Interestingly, one of the two hybrid measures achieved the smallest average error in 11 of the

18 strata. The median produced the lowest average error in five strata and the weighted mean in only two. The median obtained the lowest standard deviation in 13 of the 18 strata. To be clear, unlike the previous simulations, a low standard deviation by itself is not necessarily desirable, as a measure could produce estimated market values that are consistently too high or too low. However, the median rather consistently outperformed the weighted mean in terms of *both* accuracy and stability. Overall, the hybrid measures proved comparable to the median in both respects. Exhibits 7 through 9 show the average errors and standard deviations of the various measures graphically.

Of course, indirect equalization does not stop with computing estimated market values by class. The class estimates must be summed to a *total* market value estimate. This was done for the six hypothetical jurisdictions and compared against known total values (calculated as the sum of all the sales prices used to construct the hypothetical jurisdictions). Exhibit 10 shows the average percent error and standard deviation of the estimates (calculated as shown above) for each jurisdiction for each of the four central tendency measures. Note that the average error for all six jurisdictions is under 5.00% based on all four measures, although the standard deviations indicate that much larger errors will occur from time to time, especially in jurisdictions 1 and 2. In general, the relatively small average errors (versus exhibit 6) are explained by the relative importance of residential property with its relatively low average errors and the tendency of positive and negative errors for the various classes to somewhat balance each other. Notice that an error of more than 5% would be quite rare in any of the three largest jurisdictions.

Of the four measures, the median again performs most consistently. It produces a lower average error than the weighted mean in all but the smallest jurisdiction and provides a lower standard deviation in all six jurisdictions. The two hybrid measures perform similar to the median.

Summary and Conclusions

Ratio studies and equalization will always be difficult in smaller jurisdictions. On the positive side, the results reported here show clearly how accuracy and stability improve with larger samples. This underscores the need to consolidate jurisdictions for analysis where possible and to increase samples sizes by using larger time frames.

Despite its theoretical advantages in terms of dollar-weighting and reflecting price-related biases, the results strongly suggest that the weighted mean is not necessarily the best measure to use for indirect equalization, at least in small and mid-sized jurisdictions. In fact, the median demonstrated better performance for all except the smallest jurisdiction (where price-related biases for vacant land and commercial properties were severe). It is well accepted that the weighted mean can be skewed by outliers. An analysis of the present results also suggests that the measure can be skewed by Aoverly clean@ samples caused by insufficient samples to reflect the full range of data and any price-related biases inherent in the population. Further, the presence of some such observations in some samples but not in others breeds instability in the measure. Even in larger samples, the dollar-weighting associated with the measure can cause problems when the highest value properties, whose market values can be problematic, are included in the sample.

The paper introduced two alternatives to the median and weighted mean for indirect equalization: the median-weighted mean, which is the simple average of the traditional measures, and the square root weighted mean, which weights the ratios based on the square root of their sales prices. Both strike a compromise between the equal parcel weighting implicit in the median (and mean) and the strict dollar-weighting implicit in the weighed mean. Both performed similar to the median in the present studies and merit further consideration. Of the two measures, the square root weighted mean has a sounder statistical basis and is more amenable to further mathematical manipulations, namely confidence interval calculations.

Although the jurisdictions used in the study are hypothetical, the data represent actual arm=s-length transactions and appear representative of many small to medium-size jurisdictions. Of course, additional stratification could be undertaken in larger jurisdictions with the expectation that results would improve in all respects.

In any case, selecting the proper measure of central tendency for indirect equalization is not a cut-and-dry issue. Although the weighted mean is conceptually preferred, the median and hybrid measures tested here produced more accurate and consistent estimates.

Selected References

Birch. J. W., M.A. Sunderman, and T.W. Hamilton. 1991. Determining the minimum sample size for assessment and reappraisal work. *Property Tax Journal* 10: 299-312.

Birch, John W. 1994. *Minimum Sample Sizes for Assessment and Reappraisal: An Extension*. 1994. *Assessment Journal* 2(1): 54-59.

Denne, Robert C. 1993. The Mean, Weighted Mean, Median, and Other Robust Estimators of the Assessment Ratio: An Independent Exploration of Issues Raised by Nielson. *Property Tax Journal* 12(3): 261-72.

International Association of Assessing Officers(IAAO). 1990. *Property Appraisal and Assessment Administration*. Chicago: IAAO.

International Association of Assessing Officers(IAAO). 1990. Standard on Ratio Studies. Chicago: IAAO.

Mood, A.M., F.A. Graybill, and D.C. Boes. 1974. Introduction to the Theory of Statistics. 3d ed. New York: McGraw-Hill.

Exhibit 1
Sales Ratio Statistics

Jurisdiction 1:

PTYPE	PARCELS	MIN	MAX	MED	MEAN	WTDMN	PRD	COD
RES	300	.643	1.505	.997	.999	.988	101.1	12.9
VAC	150	.439	2.115	1.000	1.053	.971	108.5	26.2
COM	50	.495	2.307	1.000	1.063	.826	128.6	25.0

Jurisdiction 2:

PTYPE	PARCELS	MIN	MAX	MED	MEAN	WTDMN	PRD	COD
RES	500	.585	2.389	1.000	1.039	1.004	103.4	14.1
VAC	400	.413	2.400	1.000	1.090	.997	109.2	29.8
COM	100	.480	2.419	1.000	1.035	1.053	98.3	24.7

Jurisdiction 3:

PTYPE	PARCELS	MIN	MAX	MED	MEAN	WTDMN	PRD	COD
RES	1500	.552	2.362	1.000	1.019	.998	102.1	13.3
VAC	800	.400	2.500	1.000	1.067	.966	110.4	29.9
COM	200	.423	2.388	1.000	1.046	1.123	93.2	24.9

Jurisdiction 4:

PTYPE	PARCELS	MIN	MAX	MED	MEAN	WTDMN	PRD	COD
RES	3500	.484	2.143	1.000	1.022	.991	103.1	12.7
VAC	1200	.400	2.475	1.000	1.069	1.007	106.2	28.8
COM	300	.403	2.384	1.000	1.073	1.063	100.9	24.7

Jurisdiction 5:

PTYPE	PARCELS	MIN	MAX	MED	MEAN	WTDMN	PRD	COD
RES	7500	.503	2.486	.999	1.017	.992	102.5	13.3
VAC	2000	.400	2.500	1.000	1.046	.931	112.4	28.8
COM	500	.403	2.476	.998	1.064	1.080	98.5	23.7

Jurisdiction 6:

PTYPE	PARCELS	MIN	MAX	MED	MEAN	WTDMN	PRD	COD
RES	10000	.582	2.477	.999	1.031	1.002	102.9	13.8
VAC	4000	.401	2.463	1.000	1.021	.939	108.7	24.9
COM	1000	.411	2.499	1.001	1.052	1.038	101.3	26.4

Exhibit 2
Average Errors and Stability of Central Tendency Measures

				Average % Error			Standard Deviation			Maximum % Error		
JUR	PROP	PARCELS	SALES	MED	MEAN	WM	MED	MEAN	WM	MED	MEAN	WM
1	RES	300	15	3.84	3.37	3.47	5.04	4.27	4.34	19.83	16.69	15.39
1	VAC	150	8	9.36	9.10	9.43	12.33	12.09	11.02	66.01	38.79	42.61
1	COM	50	5	9.89	12.44	23.59	15.60	16.71	19.29	130.67	76.72	104.75
2	RES	500	25	2.53	3.27	3.24	3.29	4.23	4.10	13.58	15.95	16.10
2	VAC	400	20	5.35	6.78	7.62	7.38	9.20	9.70	36.34	31.39	38.75
2	COM	100	5	11.38	11.83	17.88	16.05	15.60	25.88	91.12	66.60	91.32
3	RES	1500	75	1.78	1.70	1.50	2.25	2.15	1.87	8.72	7.48	6.26
3	VAC	800	40	3.30	4.87	7.13	5.11	6.50	8.83	21.57	25.19	61.61
3	COM	200	10	7.38	8.61	17.47	9.78	11.37	24.22	55.07	50.36	87.79
4	RES	3500	175	.96	1.03	1.25	1.25	1.33	1.58	5.11	4.64	5.76
4	VAC	1200	60	1.86	3.86	5.47	3.24	5.20	6.99	17.53	17.09	27.32
4	COM	300	15	5.90	6.57	10.36	7.87	8.79	14.02	37.71	28.85	56.78
5	RES	7500	375	.70	.76	1.18	.87	.97	1.47	3.42	3.50	5.18
5	VAC	2000	100	2.95	2.89	5.22	3.99	3.79	6.12	14.92	12.97	22.43
5	COM	500	25	3.91	4.93	11.33	5.31	6.60	15.43	25.37	23.04	47.53
6	RES	10000	500	.59	.73	.63	.76	.95	.78	2.60	3.66	3.04
6	VAC	4000	200	1.79	1.82	2.86	2.36	2.32	3.46	8.46	8.73	14.89
6	COM	1000	50	3.62	3.99	7.37	4.76	5.19	9.73	20.50	20.46	35.09
RES - Totals				1.73	1.81	1.88	2.72	3.00	2.79	19.83	16.69	16.10
VAC - Totals				4.10	4.88	6.29	6.70	7.60	8.69	66.01	38.79	61.61
COM - Totals				7.01	8.06	14.66	10.89	11.62	19.26	130.67	76.72	104.75
Grand Totals				4.28	4.92	7.61	7.55	8.37	12.56	130.67	76.72	104.75

Exhibit 6
Average Errors and Stability of Total Value Estimates

				Average % Error				Standard Deviation (*)			
JUR	PROP	PARCELS	SALES	MED	WM	MED/WM	SQRWM	MED	WM	MED/WM	SQRWM
1	RES	300	15	3.95	3.48	3.50	3.35	5.16	4.41	4.52	4.25
	VAC	150	8	9.55	9.08	8.58	8.84	12.14	11.22	10.52	9.99
	COM	50	5	18.09	19.45	17.17	18.74	12.21	20.22	13.55	14.58
2	RES	500	25	2.57	3.23	2.60	3.08	3.25	4.06	3.28	3.55
	VAC	400	20	5.16	7.62	5.70	6.80	7.00	9.84	7.31	7.83
	COM	100	5	14.09	18.37	14.95	14.95	17.61	23.09	18.30	18.32
3	RES	1500	75	1.79	1.50	1.54	1.64	2.24	1.87	1.93	1.87
	VAC	800	40	4.87	7.15	4.89	6.01	5.14	9.02	6.13	6.37
	COM	200	10	14.70	19.28	15.71	14.50	11.30	20.63	14.64	14.54
4	RES	3500	175	1.17	1.25	1.06	1.67	1.24	1.60	1.28	1.29
	VAC	1200	60	2.05	5.49	3.42	4.35	3.15	7.04	4.41	4.88
	COM	300	15	7.89	10.48	7.87	7.36	8.01	13.06	9.27	9.15
5	RES	7500	375	.89	1.18	.87	1.36	.86	1.48	1.02	1.01
	VAC	2000	100	7.28	5.26	4.63	6.08	3.67	6.67	4.31	3.92
	COM	500	25	8.59	11.30	8.29	6.73	5.68	13.45	8.67	8.05
6	RES	10000	500	.64	.63	.57	1.11	.76	.78	.70	.73
	VAC	4000	200	5.76	2.85	3.30	3.66	2.24	3.65	2.48	2.36
	COM	1000	50	4.70	7.40	5.11	4.63	4.93	9.36	6.22	5.75
Totals				6.32	7.50	6.10	6.38	9.86	11.96	9.67	9.64

* Standard Deviation of 100 * Estimated/Actual Value

Exhibit 10
Average Errors and Stability of Total Value Estimates

	Average % Error				Standard Deviation (*)			
JUR	MED	WM	MED/WM	SQRWM	MED	WM	MED/WM	SQRWM
1	4.99	4.69	4.47	4.89	4.69	5.19	4.36	4.31
2	3.38	4.46	3.53	3.55	4.14	5.47	4.30	4.46
3	2.53	3.71	2.87	2.44	2.56	4.04	3.00	2.99
4	1.17	1.81	1.32	1.70	1.47	2.27	1.66	1.68
5	.73	1.40	.93	1.32	.89	1.73	1.15	1.12
6	.75	1.19	.85	1.39	.93	1.49	1.06	1.01
Totals	2.26	2.88	2.33	2.55	3.40	3.98	3.35	3.27

* Standard Deviation of 100 * Estimated/Actual Value