An Analysis of Ratio Studies & Sales Chasing Studies Conducted by the Indiana Department of Local Government Finance and Mr. Robert Denne

LaPorte County, Indiana

May 15, 2008

Prepared by:

Thomas W. Hamilton, PhD, CRE, FRICS
University of St. Thomas-Minnesota
Opus College of Business
Shenehon Center for Real Estate

Executive Summary

At the request of Ms. Carol McDaniel, LaPorte County Assessor, I have conducted a thorough analysis of several ratio studies and sales chasing studies conducted by:

1. Mr. Robert Denne of Almy, Gloudemans, Jacobs and Denne (Denne); and
2. Indiana Department of Local Government Finance (DLGF);
3. Nexus Group (Nexus), on behalf of LaPorte County.

At question is the validity of the March 1, 2006 assessments in LaPorte County, Indiana. Below is a summary of my findings and recommendation of such studies. Supporting material and research follows this executive summary in the form of seven appendices.

A. Ratio Studies

According to the International Association of Assessing Officers (IAAO), a ratio study is defined as:

> *"A study of the relationship between appraised or assessed values and market values. Indicators of market values may be either sales (sales ratio study) or independent "expert" appraisals (appraisal ratio study). Of common interest in ratio studies are the level of assessment and uniformity of the appraisals or assessments." (IAAO 1999 Ratio Study Standard)*

In other words, ratio studies are designed to measure the performance of assessors to a defined standard. In Indiana, the defined assessment standard is market value-in-use.

Denne Ratio Study: I was provided two ratio studies conducted by Denne. The first study was presented to the DLGF in February 2007. I have not reviewed this study as the DLGF rejected the findings and methodology used by Denne in this study. The second study was presented to the DLGF in October 2007. I have reviewed this study in its entirety. The five issues raised by Denne in this study are:

1. the CODs (coefficients of dispersion) for LaPorte County and the Townships contained therein are typically outside the range of acceptable standards;
2. the PRDs (price related differential) for LaPorte County and the Townships contained therein are typically outside the range of acceptable standards;
3. the median ratio of assessments (appraisals) to sales prices—A/S Ratios—are typically outside a band confidence to say that assessments (appraisals) are in-line with market sales prices;
4. each class of property is not within 5 percent of the overall A/S Ratio; and
5. selective appraisal techniques (sales chasing practices) have been employed in LaPorte County.

I will address issues 1-4 as a group in that they all deal with A/S Ratios (CODs and PRDs are functions dependent upon A/S Ratios). Issue #5 will be addressed below under the Sales Chasing Study section.

Given the methodology employed by Denne to establish A/S Ratios for his analysis, it is impossible for his conclusions to be considered valid. Denne used a sales price trending technique that is in direct violation of appraisal theory and acceptable industry practice and runs counter to established IAAO guidelines on the matter. Denne used a single, county-wide price index-based trending process to determine time-adjusted sales prices for a wide-range of properties in all of the classes of properties in his study. This aggregate index process employed by Denne biases the denominator of the A/S Ratios he used in his analysis. As such, all of the arguments Denne makes regarding "bias" and lack of uniformity in the appraisal (assessment) process are equally possible and attributable to his erroneous and biased indexing methodology. As such, items 1, 2, 3 and 4 above cannot be considered true statements based on how Denne constructed his analysis.

DLGF Ratio Study: I was provided two ratio studies conducted by the DLGF. The first study was presented to LaPorte County on Tuesday March 4, 2008. I have not reviewed this study since several errors were identified by LaPorte County. It is my understanding that the DLGF withdrew this study in order to make the necessary corrections. The second study was presented to LaPorte County on April 10, 2008. The DLGF identified several issues with the median ratios, CODs, and PRDs:

1. Springfield (improved residential) outside COD range;
2. Noble (improved residential) outside PRD range;
3. Noble (vacant residential) outside PRD range;
4. Galena (vacant residential) outside PRD range;
5. Hanna (vacant residential) outside PRD range;
6. Hudson (vacant residential) outside median, COD and PRD range;
7. Scipio (vacant residential) outside PRD range;
8. Springfield (vacant residential) outside PRD range;
9. Center (improved commercial) outside PRD range;
10. Michigan (improved commercial) outside PRD range; and
11. LaPorte County as a whole (vacant commercial) outside COD range.

As part of my analysis of the DLGF study, I also reviewed the LaPorte County analysis (see April 16, 2008 letter to Commissioner Cheryl Musgrave) and corrected DLGF ratio study which identified at least eight mathematical and/or formula errors. Upon confirming these errors, I re-ran the DLGF ratio study. These corrections eliminated issues 1, 2, 7, 9, and 10.

Issues 3, 4, 5 & 8 involve PRDs outside of the acceptable range. The DLGF ratio study appears to have made no attempt to remedy the inadequate sample sizes (see 1999 IAAO Standard on Ratio Studies, page 28 at 8.4). The DLGF ratio study also appears to have made no attempt to adjust their study for the existence of outliers (see 1999 IAAO Standard on Ratio Studies, pages 19-21 at 6.6). The DLGF ratio study has made no attempt to examine other statistical measures for vertical inequity, given the small sample sizes (see 1999 IAAO Standard on Ratio Studies, page 26 at 7.6). Finally, using data from LaPorte County, Galena Township vacant residential property represents 1.73% of the total real property assessed value in the township in 2006. Hanna Township vacant residential property represents 4.78% of the total real property assessed value in the township in 2006. Noble Township vacant residential property represents 1.81% of the total real property assessed value in the township in 2006. The DLGF ratio study and Resolution ignores the IAAO recommendation to assume the

statutory level of assessment in such instances (see 1999 IAAO Standard on Ratio Studies, page 28 at 8.4.6)

For vacant residential land in Galena, Hanna, and Springfield townships, the sample size is also a concern. The IAAO recommends two alternative statistical tests to the PRD: regression analysis and Spearman Rank test Correlation. The sample sizes in these townships do not allow for any reasonably reliable regression analysis, so the Spearman test is preferred. Upon running the Spearman test, no bias is found for regressivity in Galena and Springfield townships; however, the test does indicate a potential bias in Hanna Township.

With regard to issue #11, corrections to the DLGF study still yield a COD outside of the range. Any problem with vacant commercial land appears to be concentrated in Center Township. Again, the sample size is relatively small, and again, there appear to be no attempts to increase the sample size, or examine and correct outliers, contrary to the 1999 IAAO Standard on Ratio Studies. In such instances, the IAAO provides several remedies for inadequate sample sizes, including, but not limited to re-stratification, re-examining invalid sales and including independent appraisals. However, two sales were identified which were erroneously included as valid sales. One was an adjoining parcel purchased on a speculative basis to assemble a larger parcel for future development purposes. As such the sale is not typical compared to the other "non-assemblage" parcels which sold and would thus be considered an outlier. The second parcel was identified as a parcel whose assessed value was changed, for some unknown reason, sometime during the 12 month period following the submission of the LaPorte County 2006 ratio study. As such, with the uncertainty of which value should be assigned to this data point, it too would be considered an outlier. By removing these two extreme, or outlier, properties this property class meets Indiana standards for median A/s ratio, COD and PRD.

Once the DLGF ratio study is corrected, appropriate statistical tests are conducted, and inadequate sample size is investigated, the remaining issues involve vacant residential land in Hanna and Hudson townships.

It is difficult, at best, to compare assessments based on 1999 valuations and 2005 valuations (a six-year difference in valuation). Numerous structural changes in the marketplace for real property can take place over six years which can cause—in some cases—drastic shifts in price levels. For example, regular, unleaded gasoline in a city in 1999 was about $1.20 per gallon. By 2005, the price was about double. These are the average prices in the city for each year. The actual prices in the city core were higher and in the city's suburbs it was less. Does this mean that there is a bias or concerted effort to raise the price of gasoline? No. The marketplace bases prices on supply and demand. As demand increases relative to supply, prices rise. The same holds true for real estate. Some markets for real estate, such as lakeshore property, has certain supply constraints (fixed and limited in both the short-run and long-run) and certain demand constraints (not fixed or limited in either the short-run or long-run). As such, with a fixed supply curve and a varying demand curve, the price for lakeshore property will ultimately change over time. As demand increases, so do prices. And, if the demand for lakeshore property increases relative to the demand for other types of residential property, the price for lakeshore property will increase at a rate greater than the other types of residential property.

With these different markets for real estate changing over time, and when the time between observed events lengthens, there will be inevitable differences in prices for various groups (markets) of certain

real estate assets. The greater the difference in time, the more drastic the changes in observed prices for these assets. Therefore, a six-year difference in baseline values of assessments (1999 versus 2005) should not be unexpected. And, given differential supply and demand conditions for differing market segments of properties, the changes in baseline values of assessments should not be expected to be uniform.

As such, one must be careful when comparing property values that are significantly different solely due to the immense time difference in when their baseline values were determined. Markets change, sales prices change, and so should properties' assessed values change to reflect market conditions.

B. Sales Chasing Studies

According to the IAAO, the definition of sales chasing is:

> *"...the practice of using the sale of a property to trigger a reappraisal of that property at or near the selling price. Sales chasing causes invalid uniformity results in a sales ratio study and causes invalid appraisal level results unless similar unsold parcels are reappraised by a method that produces an appraisal level for unsold properties equal to the appraised level of sold properties." (IAAO 1999 Ratio Study Standard)*

Regarding issue 5, the Wilcoxon-Mann-Whitney U-test (WMW U-test) statistic used to claim that "sales chasing" is evident in the data cannot be used in the manner developed and presented in the Denne Study for a number of reasons. First, the WMW U-test requires that the number of items in each of the two samples be approximately the same size. This does not mean that each group must be of exactly equal size, but the two groups must be fairly similar in count to accurately depict differences in the two groups. This is definitely not the case in this dataset as the number of unsold property elements is typically 9 times as large as the number of sold property elements (in number). As the size of two samples in a WMW U-test become less similar, the likelihood of making a Type I error (the two samples are different when in fact the two samples are similar) increases.

Next, the WMW U-test requires that the underlying shapes of the two samples (sold and unsold properties) be similar. This requirement does not mean that the two samples must resemble any specific distribution (standard normal, for example), rather the requirement is that the two samples must have the same shape to be properly tested using a z-statistic. In 10 Townships and in LaPorte County as a whole, the sold and unsold property appraised (assessed) values are not similar to each other and the WMW U-test cannot be used with those data. Lastly, these tests typically assume that the data in the two groups are random events or variables. Sales of property are not considered random events, and the unsold properties are a simply a function of these sales events (either a particular property sells or it doesn't—which is a binary outcome—and the sold and unsold properties are not independent of each other). As such, the WMW-based analysis as employed by Mr. Denne cannot be viewed as a statistically sound methodology.

Therefore, I do not agree with the findings of the Denne report and cannot support the contention that any serious errors exist in the LaPorte County property assessments that would warrant a reassessment of property in the county. The remainder of this report contains issues, comments and justifications for the statements I make in this executive summary.

APPENDIX A

Summary of Pertinent Issues Regarding the Denne Report:

1) Distributions of 2005 (value year 1999) and 2006 (value year 2005) appraised values, by Township and by County, do not have the same variance, skewness, and kurtosis (i.e., distributional shape) for the sold and unsold groupings of properties for the vase majority of property in LaPorte County. As such, the two groups of sold and unsold property cannot be compared using the Mann-Whitney U-test as performed in the Denne Study due to the distributional dependencies required by the Mann-Whitney U-test.

2) Explicit necessary conditions regarding the Mann-Whitney test must be met before using the statistic to test for differences. Two sources support this claim. First, "The Mann-Whitney test requires independent samples from populations with equal variances, but the populations need not be normal." (***Applied Statistics in Business and Economics***, David P. Doane and Lori E. Seward, McGraw-Hill Higher Education, ISBN: 13:978-0-07-296696-1, page 706). Second, "(The Mann-Whitney U-test is a) nonparametric test for detecting differences between two location parameters based on the analysis of two independent samples (from a single population) (***Pocket Dictionary of Statistics***, Hardeo Sahai and Anwer Khurshid, Hill Higher Education, ISBN: 0-07-251693-3). It is from these two sources that we can see that we can know that sample populations in the two groups tested require equal variances and the groups must be independent samples from a single population, where that single population has only one distributional shape (its own). We also know that sold and unsold properties are not independent data because one condition (sold) precludes the other condition (unsold). Parametric comparisons of skewness and kurtosis of Township data show that the sold and unsold data in LaPorte County do differ in distributional shape and therefore cannot be compared using Mann-Whitney.

3) The data trending process developed and used in the Denne Study is inadequate for the purpose of trending individual sales prices to the date of value. If properties in various price-ranges, types, conditions, locations, etc., change value over time at differential rates, then using a single index to adjust prices backward (or forward) will result in a bias in the sales data used in A/S ratio studies. For example, if properties in the lowest 25% of property market values decrease 2% over a year's time, properties in the middle 50% of property market values increase 3.3% over a year's time, and properties in the upper 25% of property market values increase 8.6% over a year's time, the average increase in property values is 3.3%, but only half of the properties (at most) will experience this price increase. Half (most likely more than half) will change in price ranging from a low of about -2% (or lower) to a high of +8.6% (or higher). This creates a significant bias in the denominator of the A-S ratio calculations, the COD calculations, and the PRD calculations. It will also incorrectly allocate "unsold" properties into the "sold" sample (they did not actually sell in the correct study period 2004 to 2005), therefore potentially contaminating the Mann-Whitney test results with outliers. By incorporating error into sales prices used in the study, Mr. Denne biases the A/S ratio study results.

A proper measure to trend data does not use a single index for all property of all types within a "class" of property (e.g., detached single-family residential improved property), let alone using the single index for all "classes" of property within a jurisdiction (residential, retail, industrial, commercial, etc.). IAAO even states that "(t)he unit of comparison (for tracking price trends) should be appropriate for the type of property, for example, units for apartment buildings, square feet of living area for single-

family residences, gross or net leasable area for retail stores, and acres for rural land. ***Tracking changes in average prices not expressed on a per-unit basis is not advised because prices can vary greatly with size***." Seasonality also needs to be addressed as price changes throughout a year rarely (if ever) are constant throughout time. (Mass Appraisal of Real Property, Robert J. Gloudemans, IAAO, ISBN: 088329-165-5, pages 263-270)

"(M)arket adjustment factors are best applied when appraisal uniformity within strata is acceptable." (***Property Appraisal and Assessment Adminstration***, Joseph K. Eckert, PhD (editor), and Robert J. Gloudemans and Richard R. Almy (senior technical editors), IAAO, ISBN:088329-081-2, page 203.) Regarding value updates, "Updates are annual adjustments applied to properties between appraisals. A mass appraisal system can use ratio studies or other ***market analyses*** to derive trending factors based on ***property type, location, size, age and the like***." In no way can a single county-wide index be construed as a "market analysis" and it would definitely not be based on differentiable market conditions and property-specific issues such as "***property type, location, size, age and the like***". A single index will create an "over-representation" or an "over-simplification" of the actual market conditions present in a sales sample and the population from which the sample arises. If these different markets did not exist, then why do we even classify property and build and develop models/tables/manuals for these differentiable property types? A single model for all property as a single class would suffice.

Furthermore, "A house price index is by definition a summary indicator of spatial and/or inter-temporal house prices. House price indices provide a basis for measuring real estate values and their growth through time. But, all housing is not created equal. The attributes of the home (the square feet, number of baths, quality of materials, etc.) as well as the location of the home add substantial heterogeneity to the value of housing in any location. As a result, any index will measure individual house prices with an error and is best thought of representing overall market conditions. This is even true for house price index estimates at a detailed level of geography such as census tracts or zip codes." (***Aggregation Bias and The Repeat Sales Price Index***, Anthony Pennington-Cross, Office of Federal Housing Enterprise Oversight, 1700 G Street 4th Floor, Washington, D.C. 20052). The OFHEO is the entity which created the baseline index adopted by Mr. Denne in his study and that office recognizes differentiation of properties in the creation of their generalized index and that errors will exist for individual properties.

Therefore to use a single, simple price index to adjust sales prices to a specific day creates a biased, erroneous price (an adjusted property price estimate with an error built in). Add to this fact that spatial and inter-temporal differences exist across property types, value ranges, locations, conditions, etc., the bias in the error term becomes more important to remove to get to a properly estimated sales price for the parcel(s) under appraisal investigation. The data which Mr. Denne used contains an aggregation bias which incorporates systematic errors into individual property's adjusted sales prices.

4) The minimum sample size needed for conducting the Laporte County Sales Ratio Study is also questioned. A typical formula used for determining a minimum sample size is found on page 272 of (Mass Appraisal of Real Property, Robert J. Gloudemans, IAAO, ISBN: 088329-165-5) and a finite population equivalent is found on page 273 of the same. In both cases, the formula uses a Coefficient of Variance (COV) measure as part of the formula and includes the predicating phrase, "If the ratios can be ***assumed*** to be random and are normally distributed (or sample size is large)". COV is based on parametric statistics and the predicating phrase has significant meaning. If the ratios are NOT randomly selected (i.e., sales of properties are not random) and they are NOT normally distributed

(they are not due to the truncated nature of A-S ratios having a lower boundary approaching 0.00 and typically being skewed to the right) then simply being a "large" sample is an insufficient argument to use either sample size formulas on page 272 or 273. To be correct, repeated sampling with replacement must be conducted to obtain a normal distribution from the A/S ratios. Then, and only then, will the skewed data approximate a normal distribution. As such, the data as they are violate underlying parametric statistical requirements. Another proper sample size calculation can be found in the ***Property Tax Journal***, Volume 10, Number 3 (Sept. 1991) on pages 299-311, a publication of IAAO.

APPENDIX B

Aggregation Bias Example using Stocks and the S&P 500

Chart B-1
S&P 500 Index, January '04 to December '06

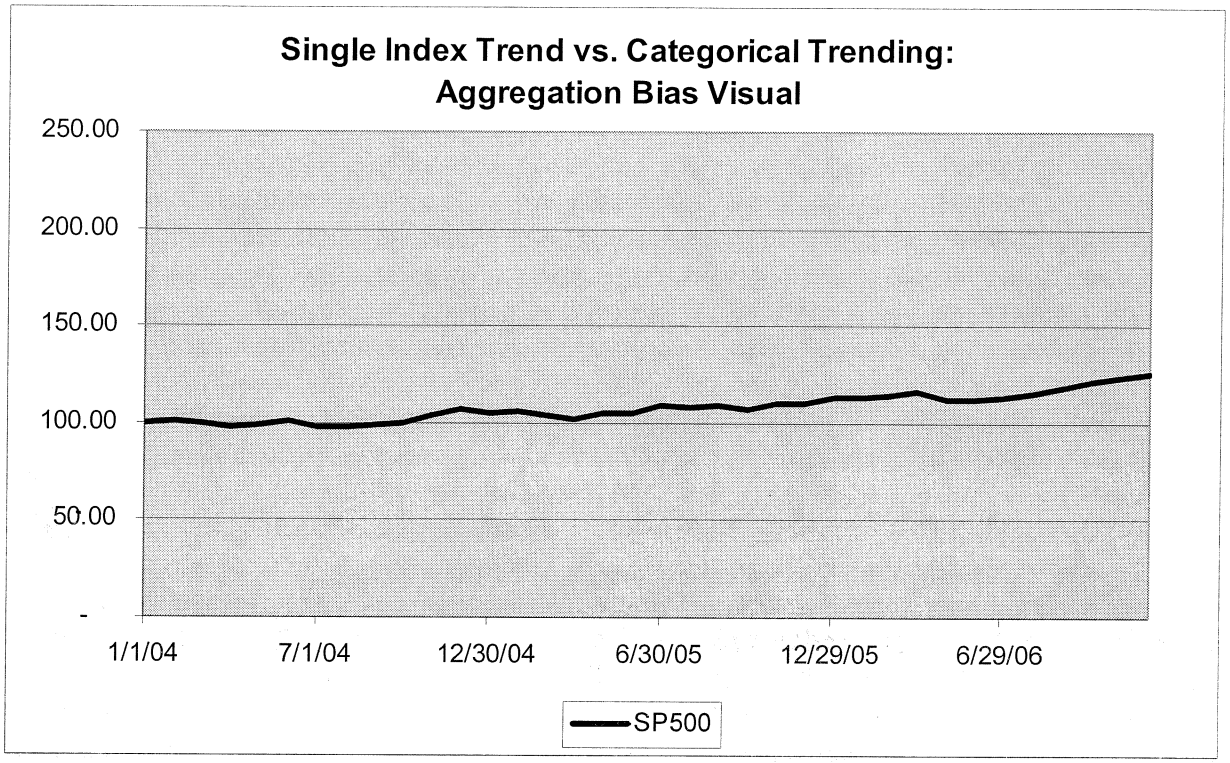
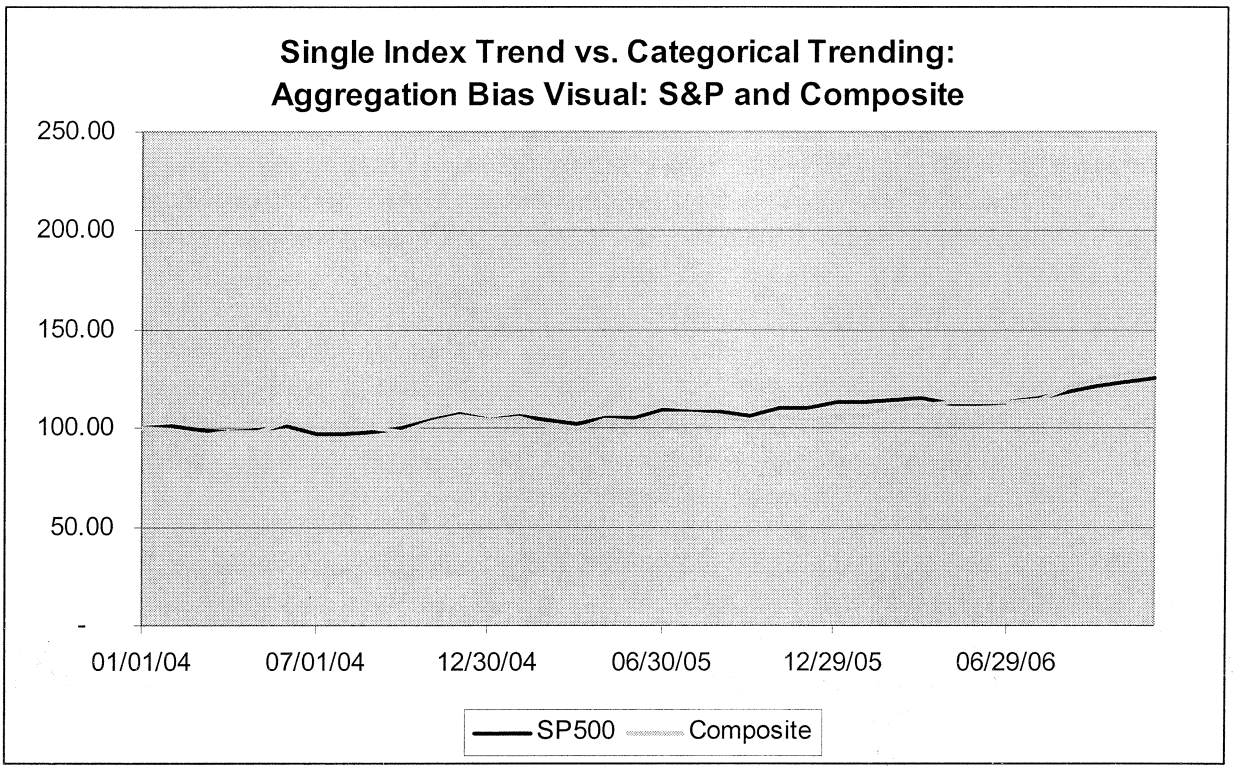


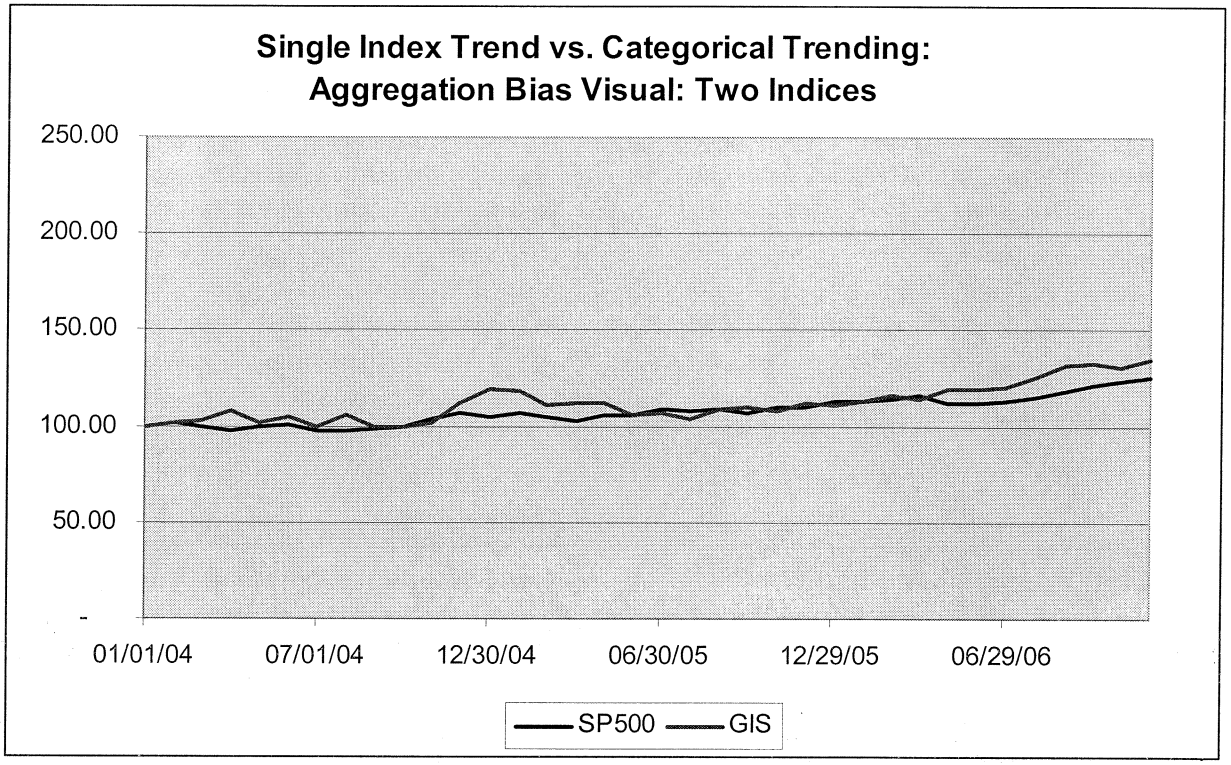
Chart B-1 is a representation of the overall price changes in the S&P 500 Index over the period January 2004 through December 2006. It is an aggregate, value-weighted index of the 500 "Largest" corporations trading on the New York Stock Exchange. It is generally accepted as a proxy for overall market performance of securities in the United States. It is also a common benchmark used to compare the returns of individual firm's stock prices over time. I am using this INDEX as an example as to why over-aggregation of price changes into a single index can distort the actual path that prices follow over time. The next graphs (appearing in Charts B-2 through B-7) show price changes for some of the component stocks that make up the S&P 500 and how they can vary significantly from this single index. It will become obvious from these graphs that this over-aggregation process should not be applied to stock price indexing or to real estate sales price indexing.

Chart B-2
S&P 500 Index and 5 Stock Composite, January '04 to December '06



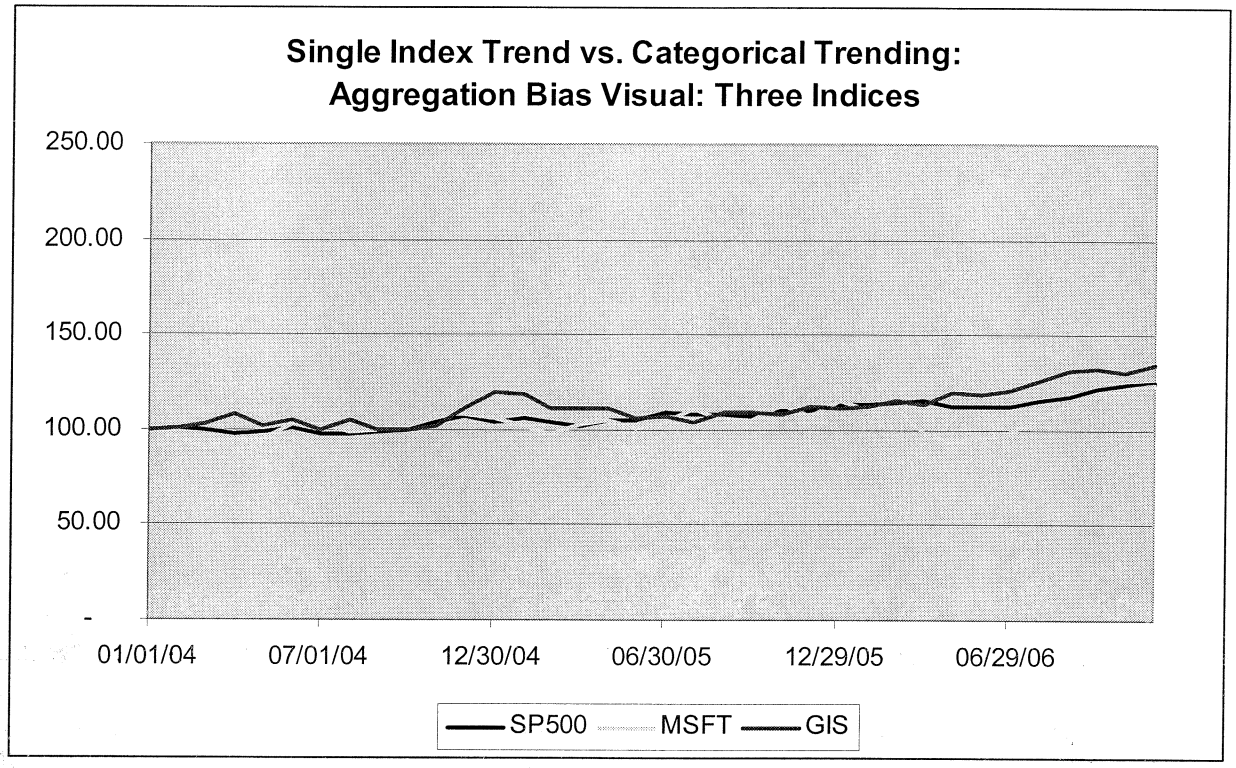
A composite of the five firms' stock prices was calculated and graphed against the S&P 500 Index. As is evident, the 5 firm composite closely follows the S&P 500 Index. This would allow us to potentially use the composite index as a proxy for the S&P 500 Index over this time period with a fairly high degree of confidence because the two lines correlate very well. In other words, the composite and the S&P 500 Index would be reasonable proxies for each other.

Chart B-3
S&P 500 Index and General Mills Stock, January '04 to December '06



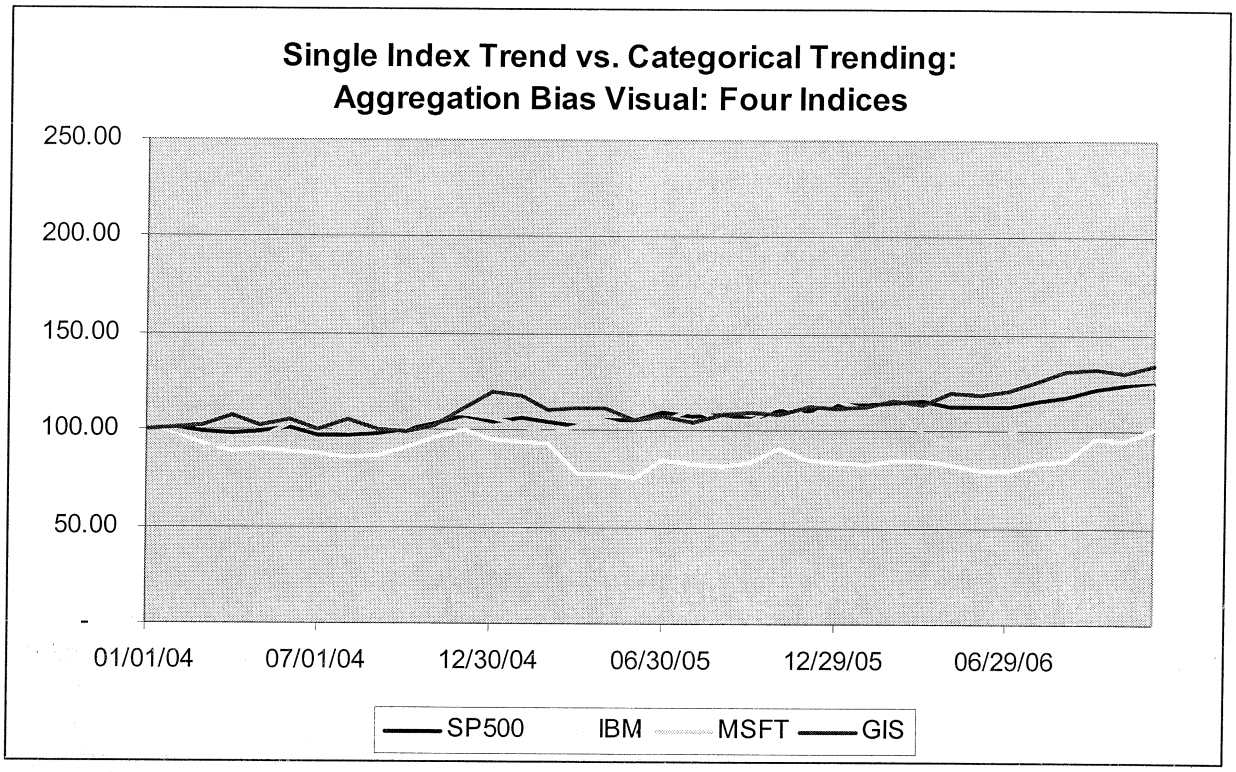
In Chart B-3, the S&P 500 Index is trended along with the stock price of General Mills Corporation. Much like the "Composite" index created for the previous graph, General Mills (GIS) closely follows the S&P 500 Index, but not as closely as does the composite. If one had to use a broad index to estimate the prior selling price of GIS stock, the S&P 500 Index would do an adequate job, but it would be less than perfect (because the two lines do have significant periods where the two index lines differ).

Chart B-4
S&P 500 Index, General Mills & Microsoft Stock, January '04 to December '06



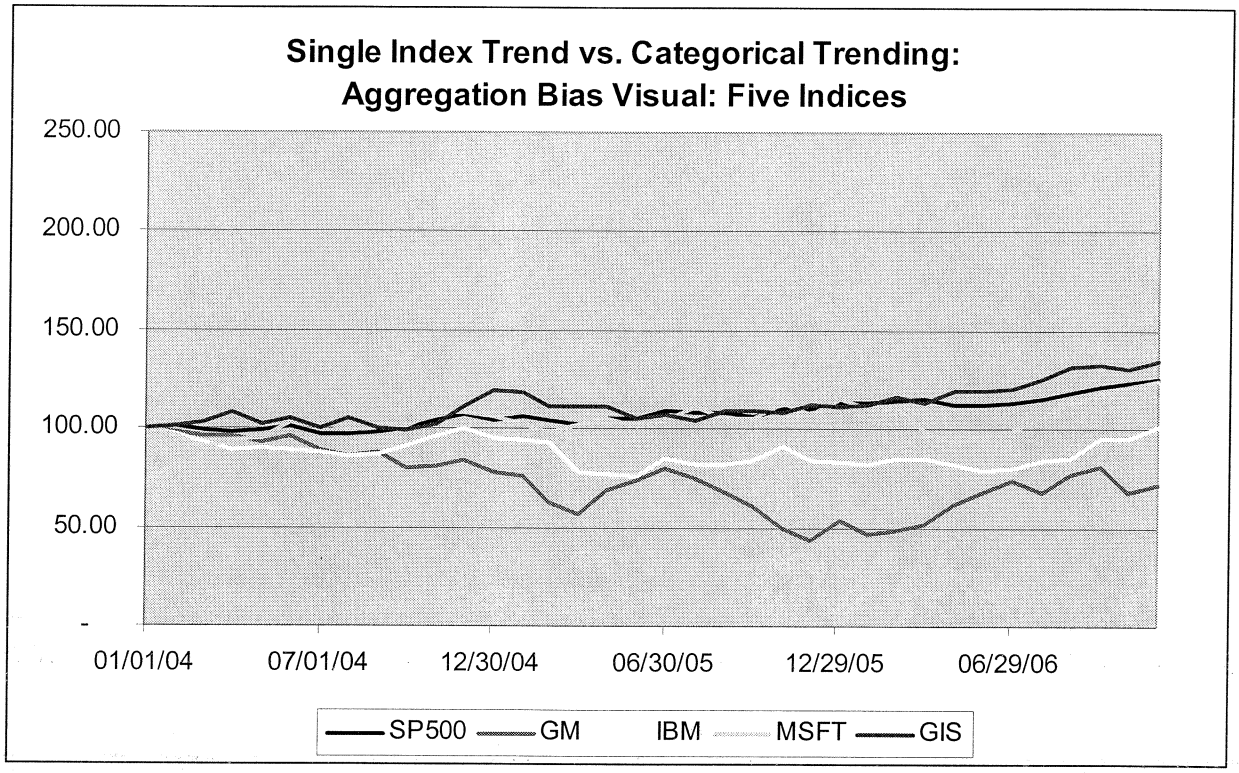
In Chart B-4, I add Microsoft Corporation stock to the S&P 500 Index and to GIS. As you can see, MSFT is not as well correlated to the S&P 500 Index, nor is it extremely well correlated with GIS. The general trend line is in the same direction, but there are substantial periods when MSFT and both the S&P 500 and GIS trend lines differ. This might be similar to what you would see with a general housing price index (S&P 500), single family detached homes in average neighborhoods (GIS) and single family detached homes in expensive neighborhoods (MSFT).

Chart B-5
S&P 500 Index, General Mills, Microsoft & IBM Stock, January '04 to December '06



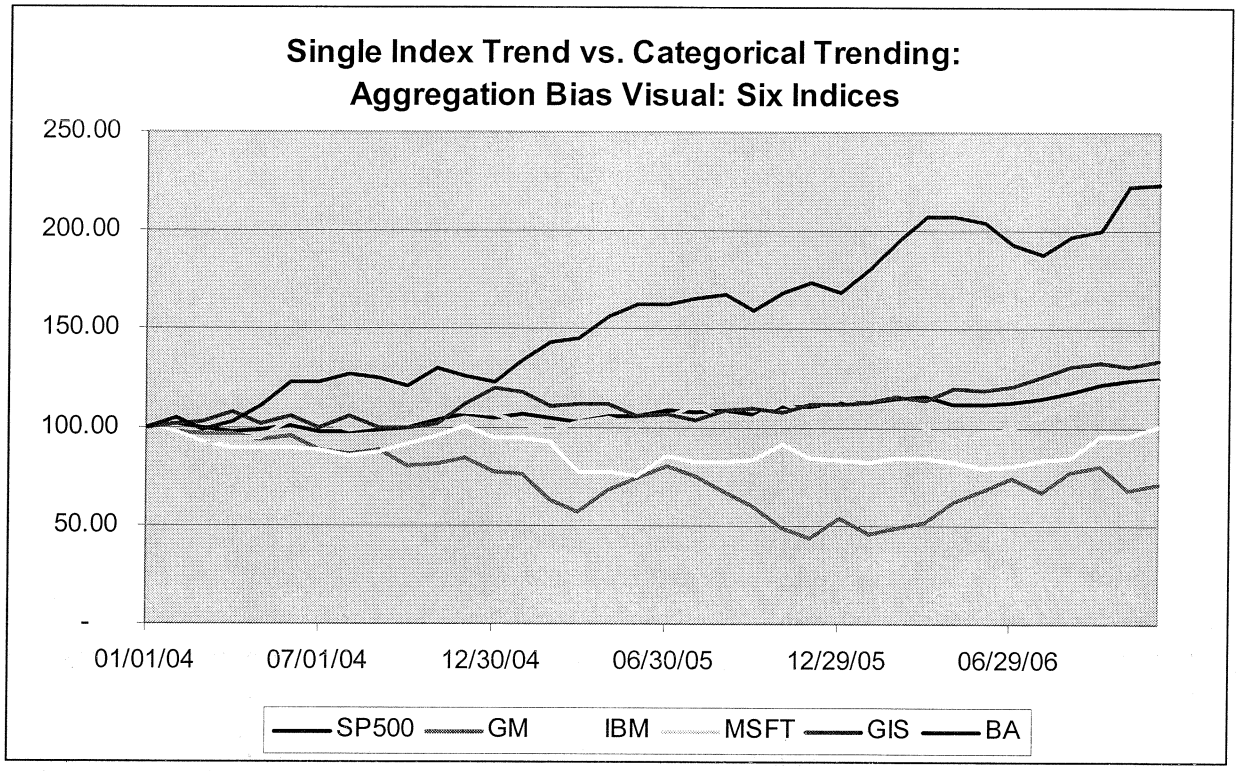
Again, looking at Chart B-5, I add another stock to the mix, this time IBM. IBM's trend line is very poorly aligned with any of the other stocks and with the S&P 500. This situation might be what we would see with multifamily (townhouse or condominium or apartments) price trends versus other single family detached housing.

Chart B-6
S&P 500 Index, General Mills, Microsoft, IBM & GM Stock, January '04 to December '06



Adding GM to the mix of stocks and the S&P 500 Index in Chart B-6, we see very divergent price changes compared to the other stocks and the S&P 500 Index. This might be a situation analogous to adding industrial/warehouse properties (GM) to the mix of residential properties (IBM, MSFT, GIS). One would have a very difficult time using the S&P 500 Index trend to explain price changes for GM stock because they do not correlate well at all. The same might be said for using a housing index (like Mr. Denne's) for non-residential property.

Chart B-7
S&P 500 Index, General Mills, Microsoft, IBM, GM & Boeing Stock, January '04 to December '06



In Chart B-7, I have added Boeing (BA) to the mix to see how it compares to the balance of the other stocks used to develop the "composite index" shown in the first graph in Chart B-1. BA might be analogous to apartments in a marketplace. It is housing, it is real property, but its price trend line is quite different from the other types of residential property and also different from warehouse/industrial property. Regardless, it would not make sense to index apartments using the housing index (S&P 500) to adjust prices of apartments (BA).

The purpose of the preceding graphs shown in Charts B-1 through B-7 is to show that when constructing an index for heterogeneous, differentiable goods (such as stocks or housing), one must be careful not to aggregate information into a single index. For example, if we used the S&P 500 index as our basis for adjusting (trending) stock prices for the five companies' stocks shown in the graphs, the index would work fairly well for Microsoft (MSFT) and General Mills (GIS), but the index would be a terrible indicator for International Business Machines (IBM), Boeing (BA) and General Motors (GM). The S&P 500 Index would systematically overstate (bias) the prices for GM and IBM, but it would systematically understate (bias) the prices for Boeing. A "composite portfolio" of the five stocks did correlate very well with the S&P500 Index, but it too would not perform well as a proxy for IBM, Boeing or GM. The "composite portfolio" is similar to the index developed by Mr. Denne for time trending LaPorte County sales. Individual Townships would be analogous to the individual stocks in this example. Many Townships will not proxy well with the "composite" and their time-trended values will differ significantly from their true, expected prices had they sold in January 2005 or had the index been properly calibrated for the 2006 sale's particular characteristics. The same systematic bias from the over-aggregation of data into a single index for housing prices will be explained in the next section.

Chart B-8

Aggregation Bias Example using Housing Indexes (Chicago)

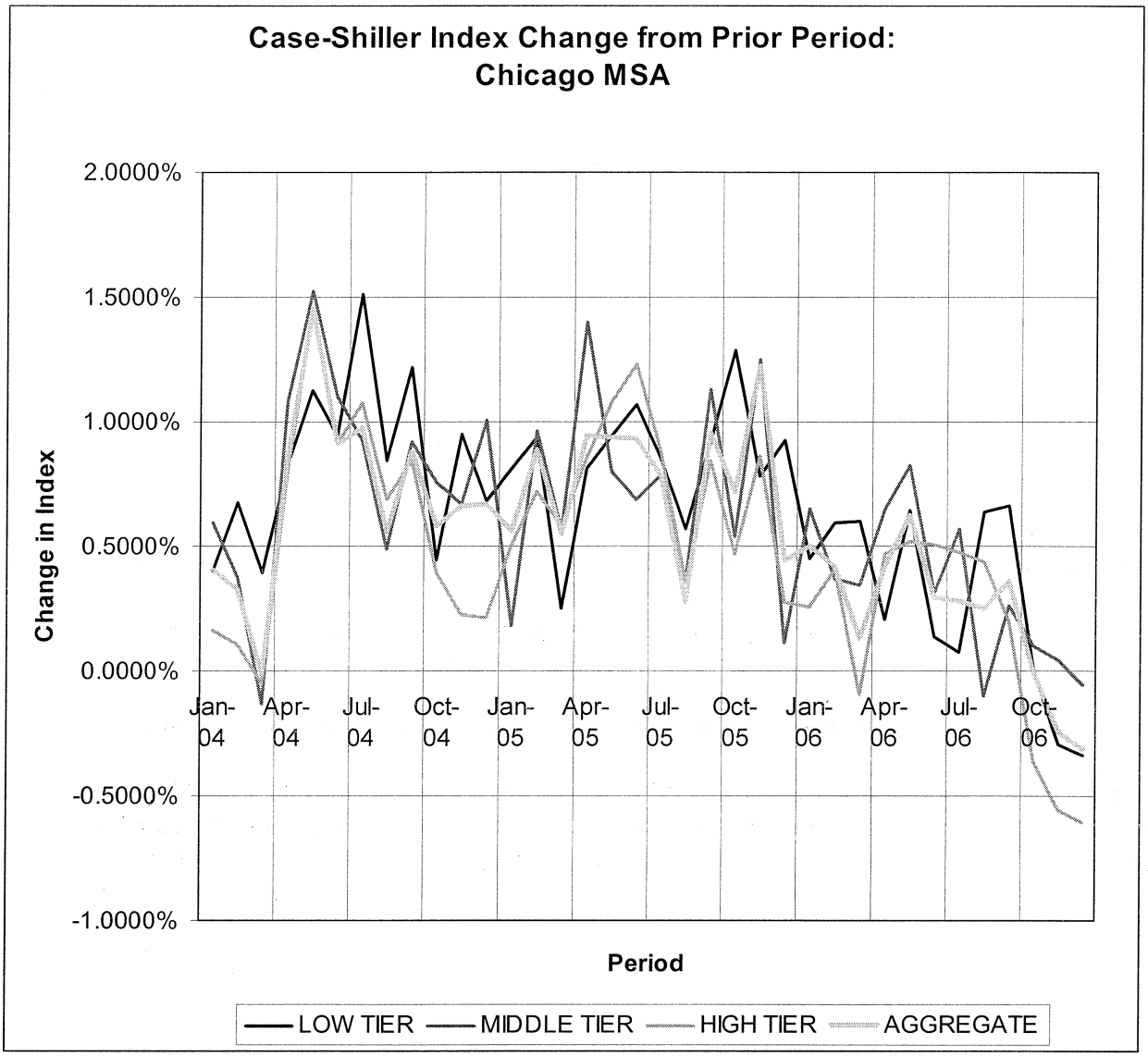


Chart B-8 shows the S&P500 Case-Shiller housing price index referenced by Mr. Denne as the other well-known index. As can be seen above for Chicago, properties in different price regimes exhibited varying price changes over the period January 2004 to December 2006. The "aggregate" index is similar in methodology to that which Mr. Denne created for LaPorte County. What is intriguing is that in some periods (for example, December 2004), prices changes fell for some groups (high tier properties) while prices changes rose for other groups (middle tier properties), but the composite price change was basically flat. In a general sense, there is correlation between price ranges over time, but depending on the particular month of sale, the aggregate, composite index compared to the actual tier index shows severe deviations which will result in a biased estimate of a property's trended sale price. IAAO Standards on Ratio Studies does not support a single-index time trend when differentiable classes and markets for property exist (more on this later). The charts below show how sales price changes vary for different priced properties over time.

Chart B-9

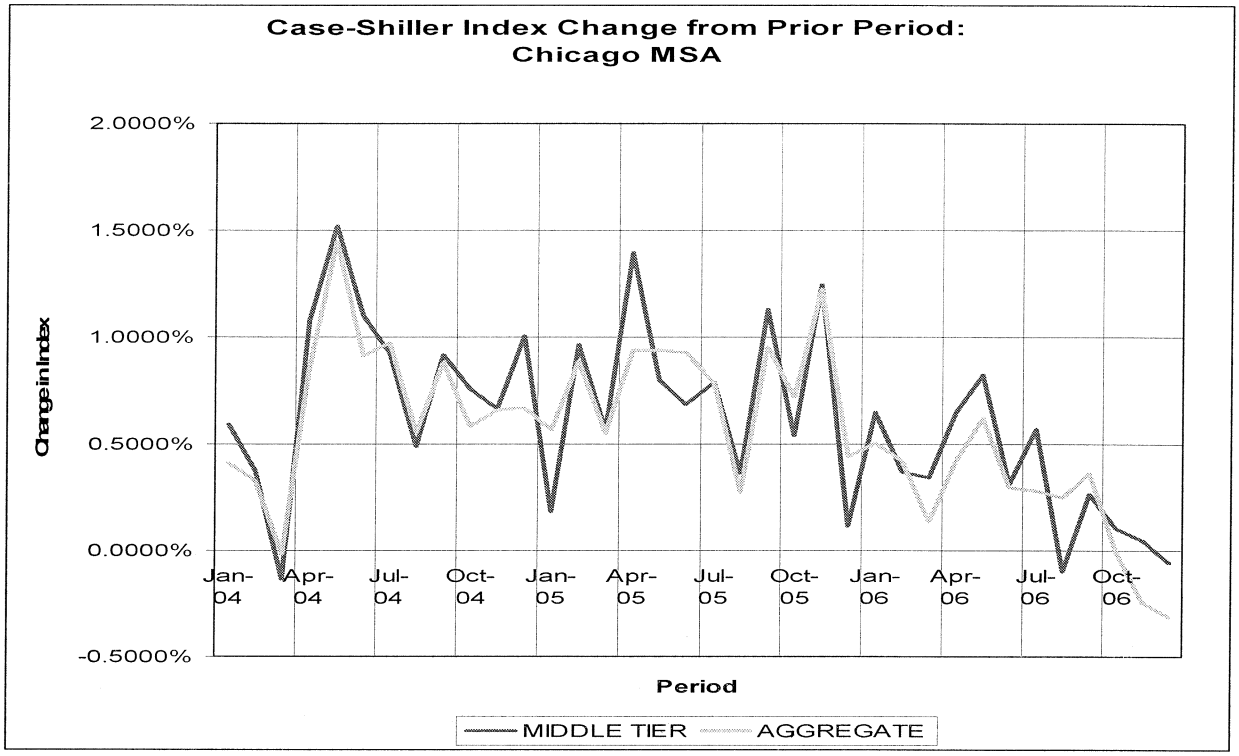


Chart B-10

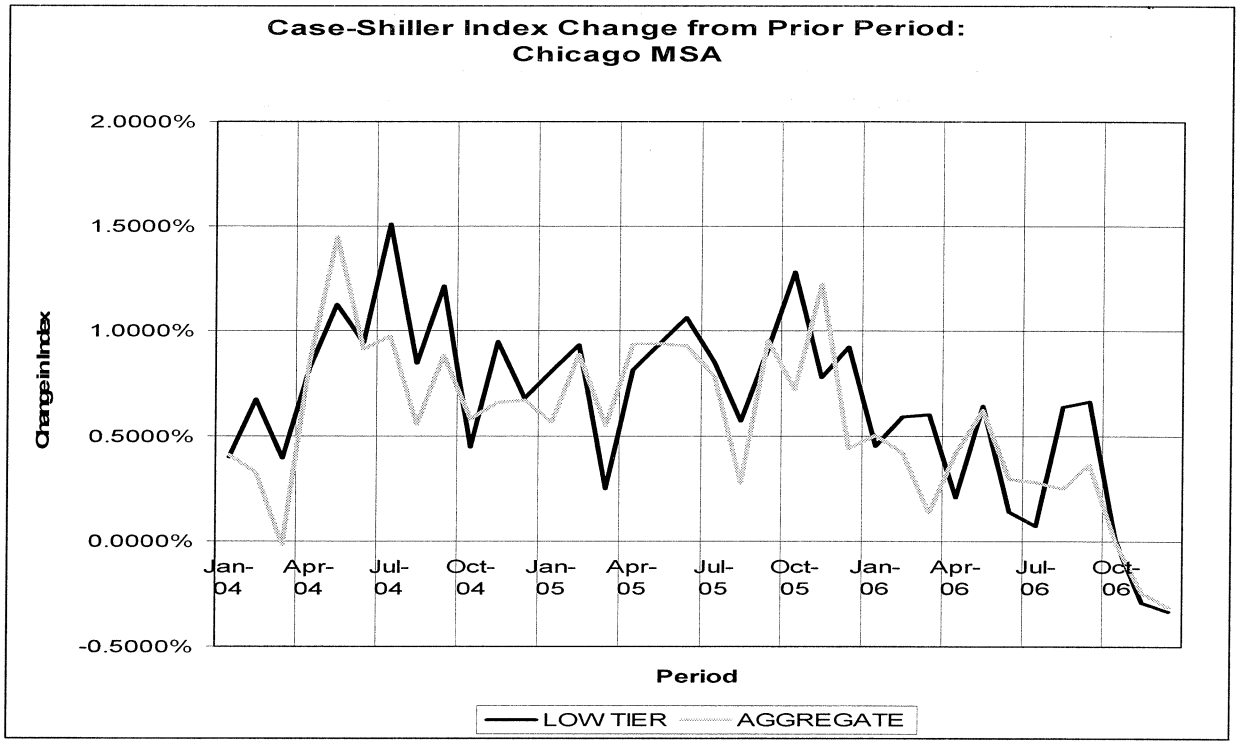


Chart B-11

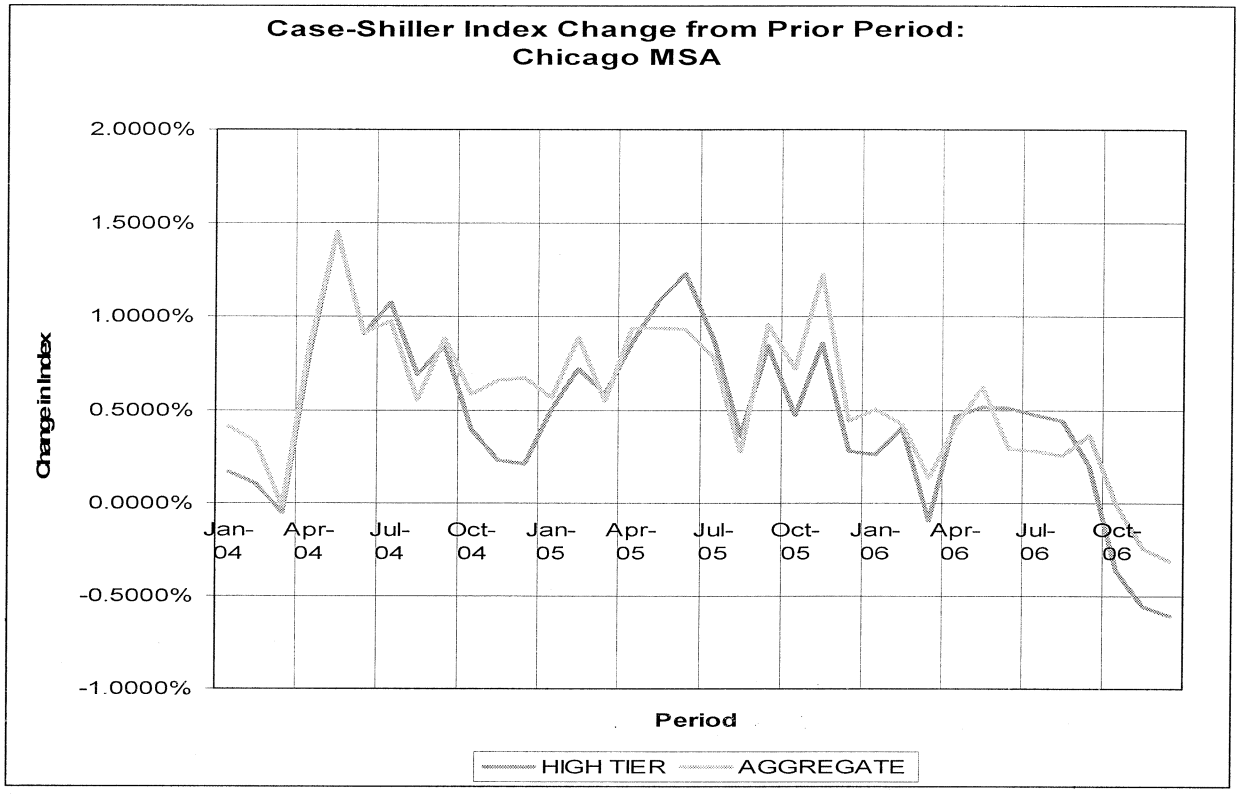


Chart B-12

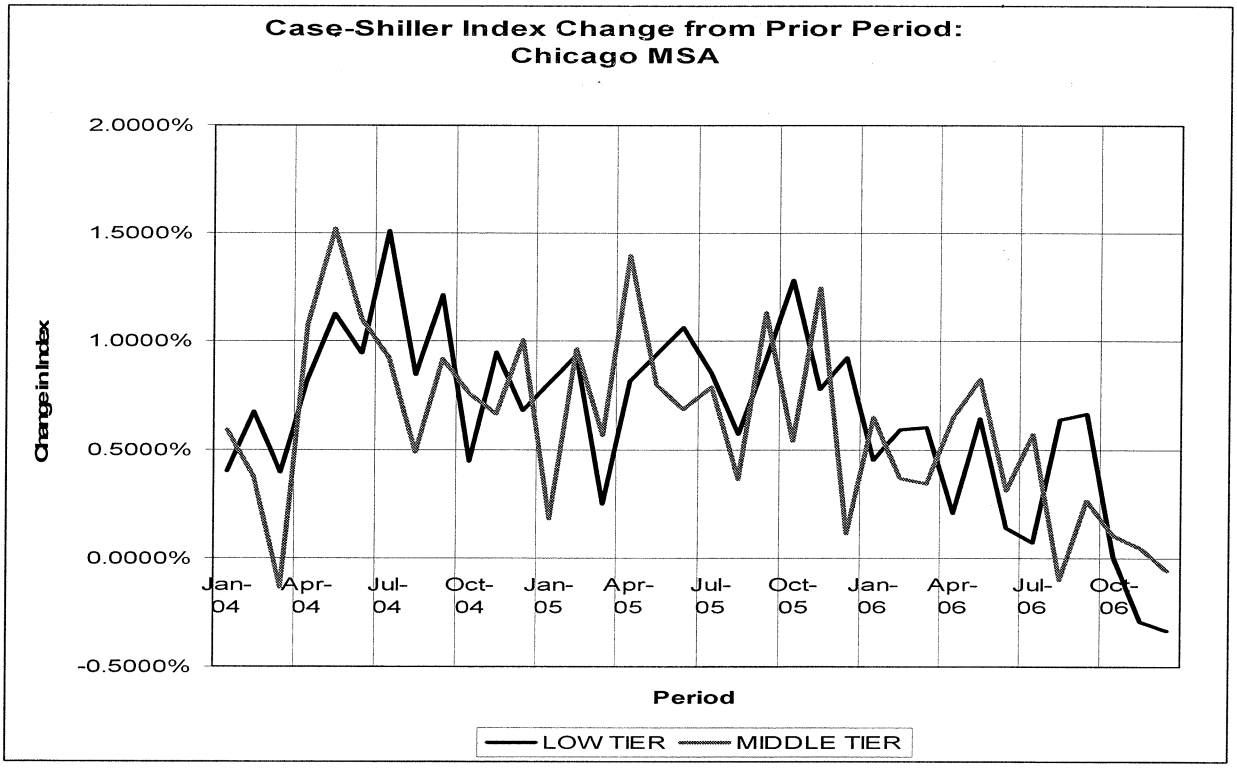


Chart B-13

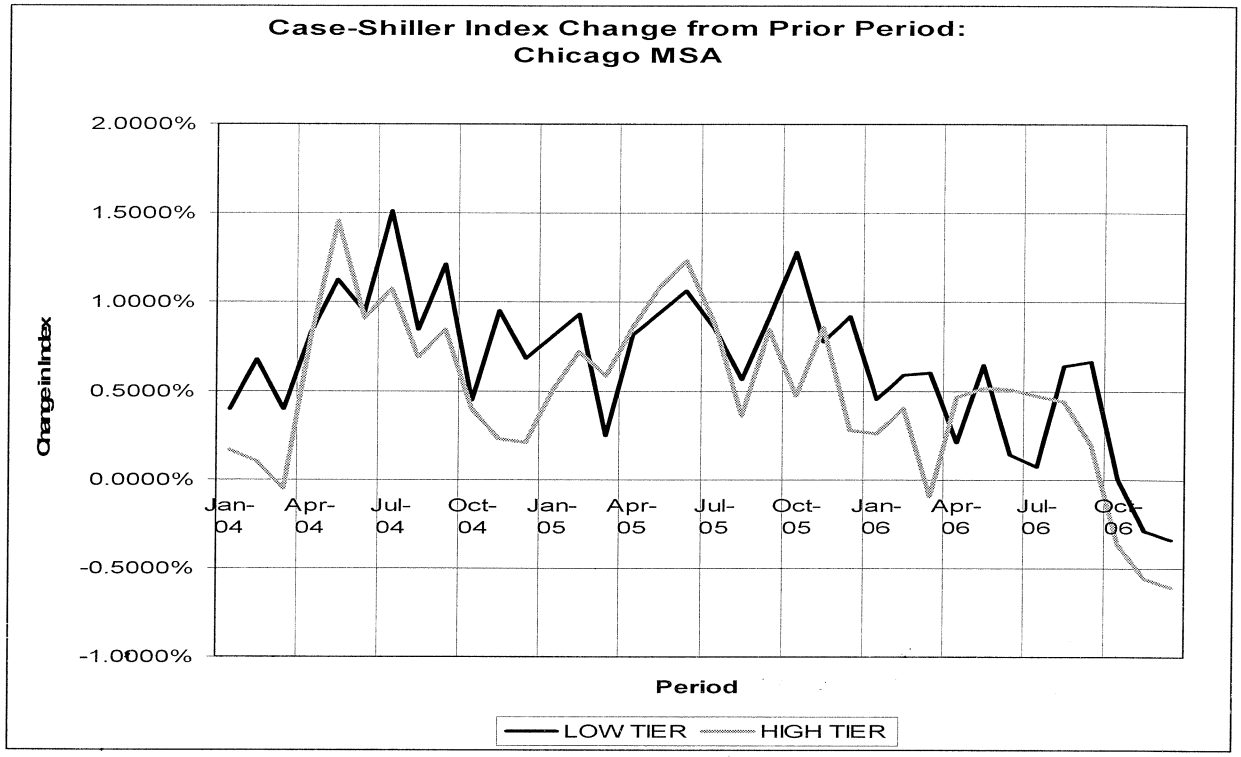


Chart B-14

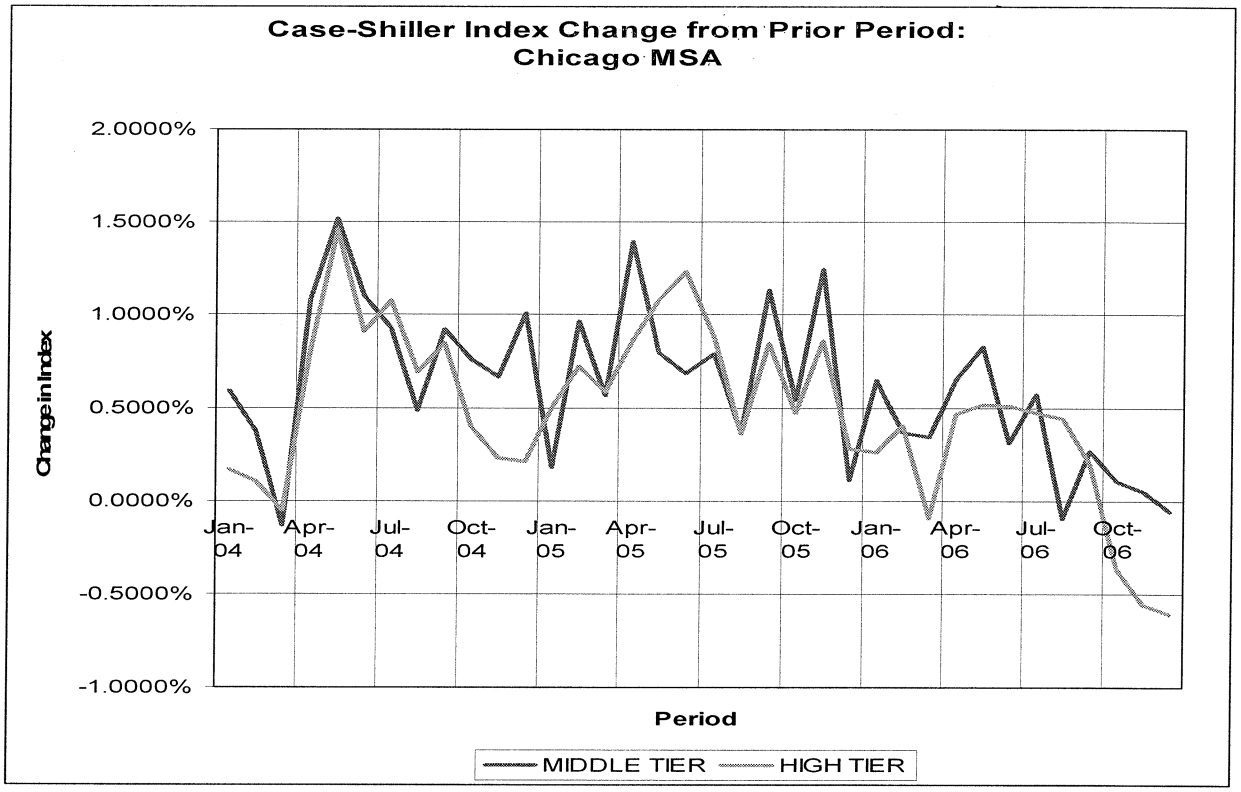


Table B-1

Correlation Matrices between Chicago Price Tier Indices and the Aggregate Index

	AGGREGATE	*LOW TIER*
AGGREGATE	1	
LOW TIER	0.796508781	1
	63.44%	

	AGGREGATE	*MIDDLE TIER*
AGGREGATE	1	
MIDDLE TIER	0.884174512	1
	78.18%	

	AGGREGATE	*HIGH TIER*
AGGREGATE	1	
HIGH TIER	0.912234065	1
	83.22%	

	LOW TIER	*MIDDLE TIER*
LOW TIER	1	
MIDDLE TIER	0.546470953	1
	29.86%	

	LOW TIER	*HIGH TIER*
LOW TIER	1	
HIGH TIER	0.740703469	1
	54.86%	

	MIDDLE TIER	*HIGH TIER*
MIDDLE TIER	1	
HIGH TIER	0.744732298	1
	55.46%	

What Charts B-8 to B-14 and correlation matrix in Table B-1 above show is that a single price index is insufficient to use as a means to adjust properties' sales prices to some point in time (i.e., create a universal time trend adjustment). Economically it makes absolutely no sense to do so. There is too much variation at different ranges of property value to indicate that a single index will properly adjust sales prices for single family properties. On average you may be correct, but the highest and lowest valued properties sales prices will be systematically over or under adjusted by using the single, mid-point index (like the one introduced by Mr. Denne in his study). I strongly believe that appraisers could very well lose their state certification if they used an index like the one proposed by Mr. Denne, hence that is why we don't see these types of time adjustment methods used by practicing appraisers.

APPENDIX C

Aggregation Bias and its effects on Sales Ratio Studies

By oversimplifying the sales price adjustment process, Mr. Denne has introduced a bias into the sales prices used in the sales ratio study. A sales ratio is typically:

Assessment divided by Sales Price, or (A/S).

A sales ratio study is typically used to gauge assessor performance through the calculation of a median A/S ratio, and CODs and PRDs. These performance measures have standards which can be used to tell us "how well" an assessor is doing her job—appraising and assessing properties. The problem with the Denne Study is that the sales prices are not based on actual market transactions. They are a "function" of the actual market transactions. The function is a "systematic" transformation of actual market sales prices, and that "systematic" transformation applies a "bias" to the denominator of the A/S ratio, which I will denote S` (S with the character ` immediately following it to differentiate it from my nomenclature for unbiased sales prices S).

The assessor's performance usually is "measured" against a "market" standard (sales price, or S). As such, A/S ratio studies are viewed as a way to see how well the assessor does her job. The median A/S ratio (the 50^{th} percentile of a distribution of data) is a statistic that is used in other performance objectives of an A/S ratio study. The COD (coefficient of dispersion) uses the median A/S ratio to determine the "typical spread" or variation in A/S ratios around the median A/S ratio. This is similar to a standard deviation which would be compared to the mean of data to some performance target or standard.

When Mr. Denne introduced a biased, systematic factoring process to adjust sales prices for all sold properties in LaPorte County, he chose to use a single factoring vector to adjust all sold properties' sales prices to a specific point in time. The factoring vector is from the OFHEO and is an index of SINGLE FAMILY HOUSING price changes in LaPorte County. It is a single indicator of how ALL single family residential properties in LaPorte County change in value over time. It is NOT a matrix of differentiable vector indices (plural) for the numerous classes and types of properties in LaPorte County. On average (for single family residential property), the index is probably statistically valid and would work for the very middle segment of properties in LaPorte County (those with sales prices at or near the "Median" house price), but it will become less and less valid to use as houses become less similar to the "Median" house. It is unequivocally INVALID to use for anything other than single family housing (and VALID ONLY for the limited number of homes which could be considered "typical" for the County). So, Mr. Denne chose to use a single index to modify sales prices of sold properties of all types and classes, and as such introduced a bias into the A/S ratio process. This is why I refer to these statistics as A/S` ratio study indicators.

In the A/S` ratio study produced by Mr. Denne, he is concerned about the level (median A/S` ratio) of assessments, the COD of the A/S` ratios, and the PRD of A/S` ratios. He claims that in many instances performance standards—as determined by the construction of the A/S` ratio study—are outside normal boundaries for acceptable assessment performance. He claims that the problem is with the ASSESSMENT function of the A/S` ratio study, and that all errors are the result of improper

ASSESSMENT techniques. Since Mr. Denne biased the A/S` ratio study with the introduction of an inexcusable, single time trending factor process, the issues he raises are equally possible to be the fault of his indexing of sales prices and could very well have nothing to do with the numerator (assessment component) of the A/S` ratios he constructed and used in his analysis.

Given that the Nexus had conducted their own study with UNBIASED sales prices and found that the assessment process fell within the guidelines of IAAO and Indiana standards, I submit that the issues presented by Mr. Denne are of his own construction in his development of an inexcusable, unprofessional indexing procedure that no appraiser of any level of expertise would endorse for mass production purposes.

APPENDIX D

IAAO Standard on Ratio Studies (July 2007)

Appendix D of the IAAO Standard on Ratio Studies (pages 56-57) describes Sales Chasing Detection Techniques. I will examine each in the following paragraphs. These standards are no different than the standards which appear in the July 1999 Standard on Ratio Studies in Section 10 on pages 30 and 31.

Technique D.1 (Comparison of Average Value Changes) evaluates property assessments from one period to the next and looks at the average percent change in value for a specified group or stratum and compares changes in the overall population of properties. What is necessary to use Technique D.1 are appraised values from period to period where the same appraisal methodology is employed in both periods. For example, in the first appraisal period land is measured on a front-footage basis and in the second period land is measured on a square footage basis. This was done to improve the overall valuation process. There could be a "statistically significant" rate of appraised value changes in either the sold or unsold group simply if one of the groups had a larger valuation change in land causing an overall valuation change to be confined primarily to that group. The overall appraisal process is better by using square footage (the premise of this example, only), but Technique D.1 might imply sales chasing. For LaPorte County, appraisals in 2005 and 2006 are based on extremely different processes, so Technique D.1 is not a reliable technique. Besides, the baseline year for appraised values for the 2005 appraisals are 1999 appraisals (indexed and trended to 2005), and the baseline year for appraised values for the 2006 appraisals is a new appraisal as of 2005. There is a six year difference in baseline measures (techniques) plus a radical change in methodology. **CONCLUSION: CANNOT USE D.1 RELIABLY.**

Technique D.2 (Comparison of Average Unit Values) looks at common measures for sold and unsold appraised properties (for example, value per square foot) and then uses a Mann-Whitney U-test or a "standard" t-test to determine if significant differences exist. Parametric tests assume many things that typically do not exist in small samples or do not exist in appraisal data altogether and typically are not used (and for the most part, correctly so). Non-parametric tests, like Mann-Whitney, do not make extreme assumptions about the underlying data distributions (such as "normality", continuous data, etc.), but Mann-Whitney does require that the two samples' data distributions are of the same shape, must be drawn randomly from a population, and must be approximately equal in size. Since property

sales are not a "random" event, large enough sales samples must exist in large enough property groups (populations of properties) such that a true, representative and unbiased random selection of sold and unsold properties can be drawn from the overall population and tested with Mann-Whitney. LaPorte County has many smaller data sets (neighborhoods and Townships) that most likely cannot produce reliable random samples to use this technique. The LaPorte County sold and unsold data sets also do not typically have the necessary similar distributional shapes needed for a proper Mann-Whiney U-test. **CONCLUSION: CANNOT USE D.2 RELIABLY.**

Technique D.3 (Split Sample Technique) looks at sales prior to the appraisal date and sales following the appraisal date, both adjusted for date of sale ***as appropriate***. Except for random sampling error and any error in time adjustments, the before and after studies should be similar to each other. If the before study is consistently better than the after study, then sales chasing is indicated. For this technique to work reliably, the time adjustment process must accurately reflect market price change trends for each class and type of property in the study. For example, if there are 5 classes of properties and 9 types of properties in each class in a study, then there must be 45 accurate time adjustment trends applied to the 45 different property groups. Without doing so, such as using a single time trending process for all 45 different property groups/markets, the time adjustment process will introduce an appraiser bias into the adjusted sales price. Additionally, the ratio studies conducted on such data will embed an undesirable combination of appraisal variation across properties and sales price variation across time. There is no realistic way to tell which component (appraised value or adjusted sales price) is causing one study to be "better" than the other. In the Denne Study, the date of sale adjustment process is NOT appropriate and violates the intention of this technique. **CONCLUSION: CAN USE D.3, BUT ITS USE REQUIRES NUMEROUS COMPLEX TIME ADJUSTMENT METRICS AND POSSIBLY REPEAT SAMPLING WITH REPLACEMENT TO GENERATE SUFFICIENT AND RELIABLE, EQUAL SAMPLE SIZES OF SOLD AND UNSOLD PROPERTIES.**

Technique D.4 (Comparison of Observed versus Expected Distribution of Ratios) also assumes an appropriate time adjustment technique (see Technique D.3 above) along with other proper adjustments (whatever those might be). This technique is based on parametric statistics and then "loosened" to be "conservative" (so as not to falsely claim sales chasing when it does not exist, a Type I error). To end the description of this technique (on page 57 of the 2007 Standard) IAAO says, "Even when critical proportions of ratios shown in table D-1 are exceeded, further investigation should be conducted

before concluding that sales chasing has occurred." Again, a proper time adjustment process is necessary for this technique to be valid, and the time adjustment methodology in the Denne Study violates the intention of this technique. Additionally, concentrations of ratios near the measure of central tendency (or near some other nodes in the distribution of ratios) can be a function of not only the numerator of the A/S` ratio (the appraised value), but it can also be a function of the denominator (S`, or time adjusted sales price). **CONCLUSION: CAN USE D.4, BUT ITS USE REQUIRES NUMEROUS COMPLEX TIME ADJUSTMENT METRICS AND CAREFUL ANALYSIS OF THE OVERALL VALUATION PROCESS (e.g., ADDITIVE MODELS VS MULTIPLICATIVE MODELS).**

Technique D.5 (Mass Appraisal Techniques) is a method of comparing two independent appraisals to each other. The "First Appraisal" is the original appraisal from the assessor's office and the "Second Appraisal" is an independent valuation conducted by another agency or entity using a model which is independently derived and not the same as the one which develop the "First Appraisal" values. In this technique, an Appraisal-Appraisal ratio study can be conducted and differences in appraisal values can be compared to see if systematic differences exist in the unsold portions of each method. This process assumes—should differences exist between the unsold portions of each method's results—that the "Second Appraisal" is unbiased and more robust at establishing values and that the "First Appraisal" has a systematic component which might be sales chasing-driven. The presumption of sales chasing is a big assumption to make and the prior assumption that the "Second Appraisal" would be "better" is a significant issue for anyone using Technique D.5. **CONCLUSION: CAN USE D.5, BUT ITS USE REQUIRES THE DEVELOPMENT OF AN EXPENSIVE, TIME-INTENSIVE PARALLEL APPRAISAL PROCESS WHICH MAY RESULT IN NO EFFICIENCY GAINS AND IT TOO MIGHT BE SUBJECT TO APPRAISER BIAS IN THE MODEL'S DEVELOPMENT.**

A Note on "Adjustments for Time":

From the 2007 IAAO Standard on Ratio Studies (pages 51 and 52) and the 1999 IAAO Standard on Ratio Studies (pages 18 and 19), valid time-adjustment techniques "compare per-unit values over time in homogeneous strata, such as a subdivision or condominium complex", and "(t)ime adjustments must be based on market analysis and supported with appropriate documentation". Additionally, "(c)hanges in price levels should be monitored and time adjustments made by geographic area and type of property, because different segments of the market tend to change in value at different rates." The Denne study fails to account for differential price level changes and bias his results.

APPENDIX E

Proper and Improper Non-parametric Testing: Distributional Shape Literature Review

This appendix contains several articles and other works which explain the proper use and limits of tests of uniformity, including—specifically—the Mann-Whitney U-test.

A Note on Homogeneity of Variance of Scores and Ranks (1996). *Journal of Experimental Education, 4,* 351-362.

When any two or more sets of scores with unequal variances are combined and ranked together as one set, the corresponding sets of ranks inherit the unequal variances. This fact is well known in the theory of nonparametric statistics, but in practice researchers and applied statisticians frequently overlook its implications. Because of this property, familiar nonparametric rank tests cannot overcome effects of heterogeneous variances of treatment groups in statistical significance testing. A simulation study demonstrates explicitly that transformation of scores to ranks reduces variance heterogeneity, although not enough to prevent gross distortion of the probabilities of Type I and Type II errors of statistical significance tests, including the *t* test, the Wilcoxon-Mann-Whitney test, and the van der Waerden, or normal scores test. The present note also focuses attention on an aspect of the problem that is neglected in the literature: The equivalence of various nonparametric tests and their parametric counterparts performed on ranks, or the *rank transformation* concept, provides a rationale for the influence of unequal variances on test statistics calculated from ranks.

Invalidation of Parametric and Nonparametric Statistical Tests by Concurrent Violation of Two Assumptions (1998). *Journal of Experimental Education, 67,* 55-68.

To provide counterexamples to some commonly held generalizations about the benefits of nonparametric tests, the author concurrently violated in a simulation study two assumptions of parametric statistical significance tests—normality and homogeneity of variance. For various combinations of non-normal distribution shapes and degrees of variance heterogeneity, the Type I error probability of a nonparametric rank test, the Wilcoxon-Mann-Whitney test, was found to be biased to a far greater extent than that of its parametric counterpart, the Student *t* test. The Welch-Satterthwaite separate-variances version of the *t* test, together with a preliminary outlier detection and downweighting procedure, protected the significance level more consistently than the nonparametric test did. Those findings reveal that nonparametric methods are not always acceptable substitutes for parametric methods such as the *t* test and *F* test in research studies when parametric assumptions are not satisfied. They also indicate that multiple violations of assumptions can produce anomalous effects not observed in separate violations.

Some Properties of Preliminary Tests of Equality of Variances in the Two-Sample Location Problem (1996). *Journal of General Psychology, 123,* 217-231.

A simulation study was conducted to examine probabilities of Type I errors of the two-sample Student *t* test, the Wilcoxon-Mann-Whitney test, and the Welch separate-variances *t* test under violation of homogeneity of variance. Two-stage procedures in which the choice of a significance test in the second stage is determined by the outcome of a preliminary test of equality of variances in the first stage were

also examined. Type I error rates of both the *t* test and the Wilcoxon test were severely biased by unequal population variances combined with unequal sample sizes. The two-stage procedures were not only ineffective, they actually distorted the significance level of the test of location. Furthermore, the distortion was greatest when the discrepancy between variances was slight rather than extreme. Unconditional substitution of the Welch separate-variances *t* test for the Student *t* test whenever sample sizes were unequal was the most effective way to counteract modification of the significance level. Conditional substitution of the Welch test, depending on the outcome of a preliminary test, was far less effective.

A Warning about the Large-Sample Wilcoxon-Mann-Whitney Test (2003). *Understanding Statistics, 2,* 267-280.

It is known that the Wilcoxon-Mann-Whitney test is strongly influenced by unequal variances of treatment groups combined with unequal sample sizes. The present simulation study indicates that, for various continuous and discrete distributions, the discrepancy between the empirical Type I error rate and the nominal significance level is large even when sample sizes are equal. In some cases, it exceeds the similar discrepancy characteristic of the Student *t* test. Furthermore, for some distributions, the discrepancy becomes increasingly more extreme as sample sizes increase. When sample sizes are relatively large, so that the normal-approximation form of the Wilcoxon-Mann-Whitney statistic is appropriate, minor and usually undetected differences in variability of treatment groups can substantially inflate the Type I error rate. For several distributions, including some that occur frequently in psychological research, ratios of population standard deviations as small as 1.1 or 1.2 have sizeable effects.

Two Separate Effects of Variance Heterogeneity on the Validity and Power of Significance Tests of Location. 2006. *Statistical Methodology, 3,* 341-394.

Heterogeneity of variances of treatment groups influences the validity and power of significance tests of location in two distinct ways. First, if sample sizes are unequal, the Type I error rate and power are depressed if a larger variance is associated with a larger sample size, and elevated if a larger variance is associated with a smaller sample size. This well-established effect, which occurs in *t* and *F* tests, and to a lesser degree in nonparametric rank tests, results from unequal contributions of pooled estimates of error variance in the computation of test statistics. It is observed in samples from normal distributions, as well as non-normal distributions of various shapes. Second, transformation of scores from skewed distributions with unequal variances to ranks produces differences in the means of the ranks assigned to the respective groups, even if the means of the initial groups are equal, and a subsequent inflation of Type I error rates and power. This effect occurs for all sample sizes, equal and unequal. For the *t* test, the discrepancy diminishes, and for the Wilcoxon-Mann-Whitney test, it becomes larger, as sample size increases. The Welch separate-variances *t* test overcomes the first effect but not the second. Because of interaction of these separate effects, the validity and power of both parametric and nonparametric tests performed on samples of any size from unknown distributions with possibly unequal variances can be distorted in unpredictable ways.

From: http://www2.chass.ncsu.edu/garson/pa765/mann.htm

Tests for Two Independent Samples: Mann-Whitney U, Wald-Wolfowitz Runs, Kolmogorov-Smirnov Z, & Moses Extreme Reactions Tests

Quantitative Research in Public Administration

Prof. G. David Garson
NCSU Box 8102
Raleigh, NC 27695-8102
Tel. 919-515-3067
Fax: 919-515-7333 E-mail: David_Garson@ncsu.edu
Office location: Winston 022
Mailbox location: 212 Caldwell

Assumptions for all non-parametric tests:

- Random sampling is assumed, as in all significance tests. **(Property sales are not a random event and their sampling representation of a population's true parameters cannot be assumed.)**

- Independent samples are assumed. The two samples should not be correlated (ex., not before-after studies, panel studies, or matched-pairs studies). **(Properties in a Sales Chasing study are not independent: a property sells (1) or it does not sell (o)—this is a binary, conditional relationship and creates a dependency between the two samples.)**

- Data distribution. The tests in this section are non-parametric, not assuming the normal distribution. The Mann-Whitney U test, but not the Wald-Wolfowitz or Kolmogorov-Smirnov tests, also assumes that the distribution in each sample is similar in shape. If the researcher can assume a normal distribution, t-tests are preferable since they can detect true differences between groups using a lower sample size than nonparametric tests in this section. Put another way, t-tests have greater power. Use independent, random samples. The Mann-Whitney U test requires that the two tested samples be similar in shape. **(Non-parametric does not mean "distribution free". Twice the author emphasizes that the two samples be "similar in shape". This can be done by looking at the data's variance, skewness and kurtosis. If two distributions have similar shapes (their variance, skewness and kurtosis are similar) then, and only then, can Mann-Whitney be used as a proper test statistic. If sales were truly a random event, then the variance, skewness and kurtosis of the two samples would most likely be similar.)**

- Data level. All the tests in this section assume ordinal data or higher.

- Data pairs. When the proportion of pairs which are tied is high, none of the tests in this section should be used.

- Sample size. For the Mann-Whitney, Wald-Wolfowitz, and Kolmogorov-Smirnov tests, sample size must be the same in the two samples so that each has the same range of rank values, from 1 to n. Small deviations from this requirement usually do not affect substantive conclusions. Populations, of course, need not be of equal size. **(Rarely would it be the case where half (or even nearly half) of the data in any jurisdiction would sell in any given year. To claim that a county's assessment data for a given year is a "Population" is a weak claim in that it is a single year's data in a stream of years' data sets, and market forces for real estate transactions extend beyond the bounds of any one county to adjoining counties. Therefore, the sample sizes of sold and unsold should be similar. This is simply not the case in any Property Class, Neighborhood, Township or grouping thereof.)**

Necessary Steps for Conducting a Proper "Sales-Chasing" Statistical Analysis

- Ascertain the appropriate strata. This may include: property type; particular delineation; neighborhood; grouping of similar neighborhoods; tax district; township or county.

- Exclude parcels, neighborhoods, etc. that have experienced significant changes since the last assessment period. This would include parcels that have undergone reassessment, new construction, demolition, etc.

- Conduct initial statistical analysis to verify that that the distributional shapes and measures of the parcels' distribution are similar either between analysis periods and/or between the sold and unsold parcel groups (or samples). As the time period lengthens between valuation periods, this requirement is more likely not to be met, hence the IAAO suggestion to conduct such analysis on a year-to-year basis.

- If conducting a Mann-Whitney "U" test, the sample sizes of sold and unsold parcels should be approximately equal. Several random samples of approximately equal size should be selected from the entire population and the various random samples should be selected with replacement (i.e., any parcel, even if previously selected for a prior test, has the same probability.)

- All statistical tests should be considered as *possible* indications, or lack thereof, regarding sales chasing. No test is definitive. For example, even if only the sold parcels change in value from one year to the next, sales chasing may not exist if there is evidence that sold parcels differ substantially from unsold parcels. Less complex comparative measures also provide insight, such as a comparison of average valuation changes between years. The oversight agency should develop a tolerance level that is acceptable, regardless of the statistical evidence. Given a sufficiently large sample, this might include both absolute and relative measures, such as allowing up to a 3% difference (ex. 5% vs. 2%) between the change in sold versus unsold parcels, and/or relatively no more than a 25% difference between the two groups (ie. 40% vs. 30%).

APPENDIX F

Sales Chasing, Distributional Issues: Township-Level Analysis

Ignoring the randomness of sales transactions and ignoring the massive size variation in the sold and unsold sample sets (both are conditions—when not met—that preclude the use of a non-parametric test such as Mann-Whitney), I tested whether the sold and unsold property groups' distributions are similar in shape. Using sold and unsold data from LaPorte County, we get the following relationships between the sold and unsold samples' appraised values for the 2005 and 2006 appraisal years. Any difference in the distributional shapes of the sold and unsold properties appraised values will preclude the use of a Mann-Whitney U-Test to determine if differences exist between the Sold and Unsold properties' appraised values within either year's data or between the two years' data.

Table F-1

Descriptive Statistics: LaPorte County

	N	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Sold 05AV	3974	79205.488	6.836	.039	127.391	.078
UnSold 05AV	32316	85411.440	15.214	.014	857.115	.027
Sold 06AV	3974	127840.103	7.230	.039	113.712	.078
UnSold 06AV	32316	140345.714	12.557	.014	497.213	.027
Valid N (listwise)	0					

As can be seen in Table F-1, the Skewness and Kurtosis for the 2005 Appraised value subgroups (Sold 05AVand Unsold 05AV) are vastly different from each other and the difference is statistically significant at the 99% level of confidence. The same holds true for the 2006 Appraised Value subgroups of sold and unsold properties. Therefore, the Mann-Whitney test cannot be applied to test for treatment differences in appraised values for either year at the LaPorte County level of analysis. This difference in distributions cannot be solely caused by purported selective appraisal techniques (sales chasing) because the sold and unsold appraisal data are not random draws from the overall county data set due to the fact that sales of property are not random events. Additionally, Mann-Whitney U-tests cannot be applied to data with distributional shapes that differ. What follows is a township-level analysis to determine whether the distributions of sold and unsold properties are of the same shape.

Summary of Township Distributional Shapes

The following Townships' assessed value distributions of sold and unsold properties are not sufficiently similar to conduct a Wilcoxon-Mann-Whitney (WMW) U-test. Although ALL 21 Townships fail the presumption that both groups (samples) are approximately equal in size (unsold properties outnumber sold properties by an average of 9 to 1) and also fail the presumption that sales of

property are randomly drawn from the population—and those two conditions alone are sufficient grounds to reject the potential use of, and results of, WMW U-tests—the list of Townships shown below in Table F-2 have significant distributional differences in skewness and/or kurtosis which indicate that the data of sold and unsold properties for the township are not similar in shape and they cannot be compared using a WMW U-test. The WMW U-test is a nonparametric test used by assessors to check for selective appraisal, it is a test recognized by IAAO to check for selective appraisal and is written into property assessment standards and by reference in Indiana Administrative Code, but the underlying statistical and mathematical premises which derive (and limit) the U-test's usefulness as a measure to test for selective appraisal cannot be ignored by promulgated use, standard or administrative code. As such, the WMW U-test results as presented by Mr. Denne in Exhibit A of Mr. Atherton's letter to Mr. Shaw dated January 25, 2008 are invalid on the basis that they violate the necessary and irrefutable statistical properties necessary to properly implement the WMW U-test. It is a common misnomer that nonparametric tests are "independent" of any distributional form, or they do not require a normal distribution and therefore are more robust measures to use when the underlying distribution is unknown. These conditions are "overgeneralizations" of what may be thought of as generally true conditions; however, the two test groups in a WMW U-test ***MUST*** be drawn at random from the same distribution of data and the data ***MUST*** be identical in ***SHAPE***. The statistical parameters which determine a population's (or a sample's) distributional shape are the:

- mean or median (central tendency);
- variance (typical spread around the central tendency);
- skewness (one tail of the distribution extends right or left); and
- kurtosis (an exaggeration (over or under) of the "thickness" of the distribution's tails).

When two data sets (the control and the experiment) are randomly drawn from a population the data sets will exhibit similar distributional shapes and the WMW U-test can be properly applied and hypotheses tested based on the WMW U-test results. Since property sales in any neighborhood, township, county, state, country or the world are not "random" events, the sold property set and the unsold property set may not pass the similar distributional shape requirement of the WMW U-test. It also causes us to commit a Type I error (rejecting similarity between groups when we can't reject the null hypothesis that the groups are similar).

Table F-2

Townships with Sold and Unsold Properties Having Different Distributions	
Based on Distributional Differences in Sold/Unsold Samples' Skewness and Kurtosis	Based on Size Differences of Sold/Unsold Samples, Distributional Means Differences in Sold/Unsold Samples; Not Random Sampling
Center	ALL TOWNSHIPS IN LAPORTE COUNTY
Coolspring	
Galena	
Hanna	
Kankakee	
Michigan	
New Durham	
Pleasant	
Springfield	
Union	
LaPorte County (as a whole)	

Tables F-3 through F-5 look at the distributions of sold and unsold property data by year and by township. These tables are the basis for the conclusions represented in Table F-2 above.

Table F-3

Distributional Shape Parameters: Mean, Standard Deviation, Kurtosis and Skewness

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Sold 05AV * Twp Name	3974	11.0%	32316	89.0%	36290	100.0%
UnSold 05AV * Twp Name	32316	89.0%	3974	11.0%	36290	100.0%
Sold 06AV * Twp Name	3974	11.0%	32316	89.0%	36290	100.0%
UnSold 06AV * Twp Name	32316	89.0%	3974	11.0%	36290	100.0%

Table F-4

Township W-M-W Distributional Shape Comparison

Twp Name		Sold 05AV	UnSold 05AV	Sold 06AV	UnSold 06AV
CASS	N	47	525	47	525
	Mean	85442.55	100164.57	93648.94	106252.76
	Std. Deviation	40096.637	41688.744	44540.899	44976.883
	Kurtosis	.917	1.758	.140	1.632
	Std. Error of Kurtosis	.681	.213	.681	.213
	Skewness	.342	.838	.341	.868
	Std. Error of Skewness	.347	.107	.347	.107
CENTER	N	1177	7391	1177	7391
	Mean	101859.98	99440.91	112767.03	108955.54
	Std. Deviation	65460.311	59387.361	73130.809	65893.211
	Kurtosis	26.844	16.621	25.782	11.101
	Std. Error of Kurtosis	.142	.057	.142	.057
	Skewness	3.827	2.729	3.533	2.370
	Std. Error of Skewness	.071	.028	.071	.028

Table F-4, continued

CLINTON	N	48	370	48	370
	Mean	102000.00	91695.95	115504.17	107442.16
	Std. Deviation	48063.048	45881.266	48278.170	49584.150
	Kurtosis	-.263	.913	.824	1.909
	Std. Error of Kurtosis	.674	.253	.674	.253
	Skewness	.489	.637	.704	.827
	Std. Error of Skewness	.343	.127	.343	.127
COOLSPRING	N	422	3797	422	3797
	Mean	99555.21	104152.88	117473.22	121983.07
	Std. Deviation	49789.079	50686.882	57342.418	58425.915
	Kurtosis	6.511	2.626	4.565	2.601
	Std. Error of Kurtosis	.237	.079	.237	.079
	Skewness	1.707	1.132	1.485	1.165
	Std. Error of Skewness	.119	.040	.119	.040
DEWEY	N	36	324	36	324
	Mean	66725.00	67596.91	73411.11	75167.28
	Std. Deviation	32415.300	34081.691	35823.343	35245.281
	Kurtosis	-.076	2.413	2.945	1.699
	Std. Error of Kurtosis	.768	.270	.768	.270
	Skewness	.318	.807	1.253	.792
	Std. Error of Skewness	.393	.135	.393	.135
GALENA	N	62	552	62	552
	Mean	91988.71	95964.49	114767.74	119846.01
	Std. Deviation	52475.087	88355.381	59746.878	96640.314
	Kurtosis	-1.084	183.257	-.992	157.899
	Std. Error of Kurtosis	.599	.208	.599	.208
	Skewness	.247	10.489	.241	9.415
	Std. Error of Skewness	.304	.104	.304	.104

Table F-4, continued

HANNA	N	32	298	32	298
	Mean	68903.12	67901.01	89725.00	88124.50
	Std. Deviation	39779.931	43179.352	49706.526	51230.399
	Kurtosis	-.730	3.269	-.813	3.331
	Std. Error of Kurtosis	.809	.281	.809	.281
	Skewness	.480	1.320	.509	1.257
	Std. Error of Skewness	.414	.141	.414	.141
HUDSON	N	89	892	89	892
	Mean	67373.03	67768.27	78629.21	76973.65
	Std. Deviation	38723.509	42930.687	44516.609	46337.572
	Kurtosis	-.222	1.277	.599	1.060
	Std. Error of Kurtosis	.506	.164	.506	.164
	Skewness	.492	.920	.747	.843
	Std. Error of Skewness	.255	.082	.255	.082
JOHNSON	N	3	36	3	36
	Mean	69266.67	91019.44	97066.67	102275.00
	Std. Deviation	19150.544	45689.629	21936.803	48301.863
	Kurtosis	.	3.423	.	3.106
	Std. Error of Kurtosis	.	.768	.	.768
	Skewness	.039	1.381	.300	1.282
	Std. Error of Skewness	1.225	.393	1.225	.393
KANKAKEE	N	139	1113	139	1113
	Mean	113864.03	112708.89	122308.63	120838.54
	Std. Deviation	66848.105	62317.233	71424.416	66012.952
	Kurtosis	.293	2.351	1.434	3.501
	Std. Error of Kurtosis	.408	.147	.408	.147
	Skewness	.725	1.091	1.009	1.273
	Std. Error of Skewness	.206	.073	.206	.073

Table F-4, continued

LINCOLN	N	89	873	89	873
	Mean	62211.24	63937.00	75825.84	74970.68
	Std. Deviation	37743.764	44678.426	43075.345	46419.784
	Kurtosis	.459	.731	.192	.469
	Std. Error of Kurtosis	.506	.165	.506	.165
	Skewness	.797	1.002	.731	.791
	Std. Error of Skewness	.255	.083	.255	.083
MICHIGAN	N	1092	9737	1092	9737
	Mean	115012.91	108331.08	179827.20	165334.24
	Std. Deviation	115868.426	129378.961	210135.942	231760.710
	Kurtosis	94.385	532.654	49.777	214.000
	Std. Error of Kurtosis	.148	.050	.148	.050
	Skewness	6.767	13.867	5.049	8.667
	Std. Error of Skewness	.074	.025	.074	.025
NEW DURHAM	N	124	989	124	989
	Mean	111019.76	103350.00	128850.00	121332.05
	Std. Deviation	52697.109	58195.338	57250.175	65853.400
	Kurtosis	-.736	1.127	-.622	.929
	Std. Error of Kurtosis	.431	.155	.431	.155
	Skewness	.061	.628	.008	.559
	Std. Error of Skewness	.217	.078	.217	.078
NOBLE	N	65	446	65	446
	Mean	87981.54	80285.65	111716.92	103249.78
	Std. Deviation	47269.605	45980.454	53054.255	53004.348
	Kurtosis	.216	.634	.037	.660
	Std. Error of Kurtosis	.586	.231	.586	.231
	Skewness	.705	.773	.508	.689
	Std. Error of Skewness	.297	.116	.297	.116

Table F-4, continued

PLEASANT	N	137	973	137	973
	Mean	95375.18	92601.13	110596.35	108196.30
	Std. Deviation	36061.830	44503.900	44093.559	54693.497
	Kurtosis	.160	20.864	.278	21.042
	Std. Error of Kurtosis	.411	.157	.411	.157
	Skewness	.180	2.381	.260	2.454
	Std. Error of Skewness	.207	.078	.207	.078
PRAIRIE	N	4	28	4	28
	Mean	92900.00	109378.57	112800.00	131457.14
	Std. Deviation	69147.041	47066.890	77505.871	51323.198
	Kurtosis	2.655	-.081	2.756	-.060
	Std. Error of Kurtosis	2.619	.858	2.619	.858
	Skewness	1.426	.360	1.508	.168
	Std. Error of Skewness	1.014	.441	1.014	.441
SCIPIO	N	147	1237	147	1237
	Mean	124285.03	125995.55	137870.75	137854.08
	Std. Deviation	65516.931	54879.076	67167.268	56900.613
	Kurtosis	3.010	2.459	3.218	3.117
	Std. Error of Kurtosis	.397	.139	.397	.139
	Skewness	1.546	1.122	1.537	1.143
	Std. Error of Skewness	.200	.070	.200	.070
SPRINGFIELD	N	120	1177	120	1177
	Mean	83322.50	84111.34	123065.83	118683.86
	Std. Deviation	52545.240	48829.930	101201.686	82270.461
	Kurtosis	8.587	4.536	24.337	7.697
	Std. Error of Kurtosis	.438	.142	.438	.142
	Skewness	1.759	1.106	3.818	2.109
	Std. Error of Skewness	.221	.071	.221	.071

Table F-4, continued

UNION	N	71	719	71	719
	Mean	50749.30	53234.08	58322.54	61263.00
	Std. Deviation	25321.024	28013.768	28215.731	45134.476
	Kurtosis	.869	3.069	.684	211.126
	Std. Error of Kurtosis	.563	.182	.563	.182
	Skewness	.962	1.297	.950	11.077
	Std. Error of Skewness	.285	.091	.285	.091
WASHINGTON	N	37	385	37	385
	Mean	97440.54	92937.40	111243.24	102737.40
	Std. Deviation	52015.262	47772.021	58853.063	49202.262
	Kurtosis	-.153	.561	.408	.601
	Std. Error of Kurtosis	.759	.248	.759	.248
	Skewness	.697	.652	.871	.594
	Std. Error of Skewness	.388	.124	.388	.124
WILLS	N	33	454	33	454
	Mean	136060.61	119422.03	151518.18	131960.13
	Std. Deviation	75956.842	51893.664	80397.456	52489.630
	Kurtosis	1.253	4.166	1.415	2.079
	Std. Error of Kurtosis	.798	.229	.798	.229
	Skewness	.720	1.119	.748	.694
	Std. Error of Skewness	.409	.115	.409	.115
Total	N	3974	32316	3974	32316
	Mean	102459.98	99832.27	130704.45	126613.51
	Std. Deviation	79205.488	85411.440	127840.103	140345.714
	Kurtosis	127.391	857.115	113.712	497.213
	Std. Error of Kurtosis	.078	.027	.078	.027
	Skewness	6.836	15.214	7.230	12.557
	Std. Error of Skewness	.039	.014	.039	.014

Table F-5

ANOVA Table

			Sum of Squares	df	Mean Square	F	Sig.
Sold 05AV * Twp Name	Between Groups	(Combined)	9.256E11	20	4.628E10	7.623	.000
		Within Groups	2.400E13	3953	6.071E9		
		Total	2.492E13	3973			
UnSold 05AV * Twp Name	Between Groups	(Combined)	6.805E12	20	3.403E11	48.000	.000
		Within Groups	2.289E14	32295	7.089E9		
		Total	2.357E14	32315			
Sold 06AV * Twp Name	Between Groups	(Combined)	4.369E12	20	2.184E11	14.258	.000
		Within Groups	6.056E13	3953	1.532E10		
		Total	6.493E13	3973			
UnSold 06AV * Twp Name	Between Groups	(Combined)	2.738E13	20	1.369E12	72.589	.000
		Within Groups	6.091E14	32295	1.886E10		
		Total	6.365E14	32315			

APPENDIX G

Random Sampling and Size Differences in the WMW Test

EXAMPLE 1

In this example, I assume that 1001 properties exist in a "Township". The properties range in appraised value from $50,000 to $150,000 in $100 increments and those appraisals are perfect representations of the marketplace. In other words, the data are perfect. I also assume in this first example that 101 properties sell (roughly 10%). They also happen to be the first 101 data (the 101 lowest valued properties). The results for the WMW test are shown below:

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Appraised Value	1001	100000.00	28910.811	50000	150000
Sold100	1001	.10	.301	0	1

Mann-Whitney Test

Ranks

	Sold100	N	Mean Rank	Sum of Ranks
Appraised Value	0	900	551.50	496350.00
	1	101	51.00	5151.00
	Total	1001		

Test Statistics[a]

	Appraised Value
Mann-Whitney U	.000
Wilcoxon W	5151.000
Z	-16.497
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Sold100

According to this Mann-Whitney test, "sales chasing" has occurred, even though in the construction of the example the data are "perfectly appraised" and no sales chasing exists.

EXAMPLE 2

In this example, the 100 properties which sell alternate (first, eleventh, twenty-first, thirty-first, etc. in ascending value) through the final property in the "Township". Again all the appraised values are correct, and in this example, the WMW test says there is no evidence of sales chasing.

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Appraised Value	1001	100000.00	28910.811	50000	150000
Sold 100 Alternate	1001	.10	.300	0	1

Mann-Whitney Test

Ranks

	Sold 100 Alternate	N	Mean Rank	Sum of Ranks
Appraised Value	0	901	500.56	451001.00
	1	100	505.00	50500.00
	Total	1001		

Test Statistics[a]

	Appraised Value
Mann-Whitney U	44650.000
Wilcoxon W	451001.000
Z	-.146
Asymp. Sig. (2-tailed)	.884

a. Grouping Variable: Sold 100 Alternate

This example of a small data set selling in a non-random manner causes us to question why if all the properties are properly valued that in one case "sales chasing" exists (Example 1) and in the other case "sales chasing" does not exist (Example 2)? The answer is not in the construction of the nonparametric statistic, but rather the fact that a non-random event caused the results. Sales of property are not random events and the distributions of the sold properties in the first case differ substantially from the distribution of the sold properties in the second case.

EXAMPLE 3

In this example, 501 properties sell (approximately 50%), but it is the first 501 which sell. The final 500 properties do not sell. The results from the Mann-Whitney test would indicate that "sales chasing" has occurred.

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Appraised Value	1001	100000.00	28910.811	50000	150000
Sold500	1001	.50	.500	0	1

Mann-Whitney Test

Ranks

	Sold500	N	Mean Rank	Sum of Ranks
Appraised Value	0	500	751.50	375750.00
	1	501	251.00	125751.00
	Total	1001		

Test Statistics[a]

	Appraised Value
Mann-Whitney U	.000
Wilcoxon W	125751.000
Z	-27.386
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Sold500

The Z-statistic is very large in this case. Sales chasing must have occurred even though the data are constructed as "perfect".

EXAMPLE 4

In this example, the 501 properties which sell alternate (first, third, fifth, seventh, etc.) through the final property in the "Township". Again all the appraised values are correct, and in this example, the WMW test says there is no evidence of sales chasing.

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Appraised Value	1001	100000.00	28910.811	50000	150000
Sold 500Alternate	1001	.50	.500	0	1

Mann-Whitney Test

Ranks

	Sold 500Alternate	N	Mean Rank	Sum of Ranks
Appraised Value	0	501	501.00	251001.00
	1	500	501.00	250500.00
	Total	1001		

Test Statistics[a]

	Appraised Value
Mann-Whitney U	125250.000
Wilcoxon W	251001.000
Z	.000
Asymp. Sig. (2-tailed)	1.000

a. Grouping Variable: Sold 500Alternate

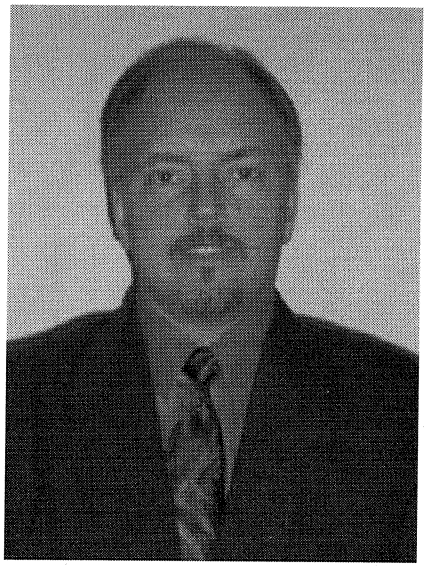
As we can see here, an equal-sized data set (approximately 50% sell and 50% don't) with properties selling in a non-random manner causes us to question why if all the properties are properly valued that in once case "sales chasing" exists (Example 3) and in the other case "sales chasing" does not exist (Example 4)? Again, the answer is not in the construction of the nonparametric statistic, but rather the fact that a non-random event caused the results. Again, sales of property are not random events and the distributions of the sold properties in the first case differ substantially from the distribution of the sold properties in the second case. This is why we must first be certain that the distributions of the data are sufficiently similar before conducting the Mann-Whitney test. An alternative, to ensure distributional

similarity, is to randomly select sold and unsold properties—with replacement—and simulating data to ensure distributional symmetry exists.

Thomas W. Hamilton, PhD, CRE, FRICS, is Associate Professor of Real Estate at the University of St. Thomas Opus College of Business' Shenehon Center for Real Estate; founder and president of Property Tax Evaluation and Consulting; and co-developer of VHAAS© ad valorem tax software. He holds an M.S. and a Ph.D. in Real Estate and Urban Land Economics from the University of Wisconsin-Madison, an M.S. in Finance from the University of Wyoming, and a B.S. in Natural Resources from the University of Wisconsin-Madison. He has been a consultant to the telecommunications industry and the utility industry in the United States for over 15 years, verifying and evaluating ad valorem tax strategies and problems and has been a consultant to numerous local and state agencies in the analysis of property tax equity and incidence since 1989. He is a Counselor of Real Estate (CRE), Chairman of the Minnesota Chapter of CRE, and a Fellow of the Royal Institution of Chartered Surveyors (FRICS). He is a former Fellow and Adjunct Faculty member of the Lincoln Land Institute in Cambridge, MA and former Research Associate with the International Association of Assessing Officers in Chicago, IL. Dr. Hamilton has researched and published numerous articles on valuation techniques and ad valorem tax equity issues and concerns in national and international journals and has been a nationally syndicated columnist with Knight-Ridder/Tribune news service. His research currently focuses on income-based valuation methodologies for complex real estate systems, and he has successfully developed methods to identify and correct systematic, inequitable mass appraisal valuation processes.

Contact Info:

Tom Hamilton, PhD, CRE, FRICS
Associate Professor of Real Estate
University of St. Thomas Opus College of Business
2115 Summit Avenue, Suite MCH316
St. Paul, MN 55105

651-962-5551 (UST)
651-962-5093 (FAX)
651-335-3560 (CELL)

twhamilton@stthomas.edu
drtomhamilton@msn.com