Use of Mass Appraisals in Toxic Tort Litigation Involving Loss of Value

Charles D. Cowan, Ph.D.[1]

Abstract

This paper reviews methods used in Toxic Tort Litigation to examine alleged loss of value for properties proximate to buildings suffering environmental disamenities. These methods build on standard mass appraisal methods, use traditional assumptions regarding probability distributions, and introduce new techniques that make efficient use of the appraisal results. The goal of this article is to demonstrate that reasonable procedures with common-sense appeal can prevail even when standard analytic or econometric techniques fail. Care in reducing the problem to an elemental level, buttressed by minimal assumptions, leads to results that can be understood and readily accepted.

Introduction

There has been a release of a contaminant and the issuance of a health warning, perhaps by the EPA. Although we know the extent of the contamination, the problem is not the contamination but the alleged stigma that may (or not!) be associated with the properties near the site. Has there been a reduction in property values near the site, and in which neighborhoods? How far from the site? How many properties may be suffering from an alleged loss of value?

Plaintiffs in a lawsuit may seek to certify as a class a set of properties as affected by their proximity to a site named in a health report; defendants may wish to anticipate a class action by determining whether there is any effect on property values. Is there some area of any shape or size in which proximity to the site had some affect on property values? The answer lies in an examination of the sales and patterns of prices of properties over a lengthy period of time, both before and after the incident that may have stigmatized such properties.

The research method employed is simple to understand and proceeds in discrete steps. As described below, the method applies to sales of residential properties, but similar methods could be applied to commercial properties. We start with the acquisition of all property tax records summarizing residential sales in the county where the incident occurred. Next we develop a mass appraisal model to predict the sales prices for all residences based on sales and housing characteristics. The third step is to calculate for each property the difference between the property's actual sale price as recorded in property tax records and the model's predicted sale price for that property. Finally, we employ search procedures to look for areas where the actual and predicted sales prices are very different. In areas where the differences are large, factors other than standard housing characteristics, such as proximity to a contamination site, could be influencing sales prices.

---

[1] Charles D. Cowan is Chief Executive Officer for Analytic Focus LLC, a firm specializing in litigation support, econometrics and forecasting, and research for financial institutions. He can be reached through the Analytic Focus website, www.analyticfocus.com or toll free at 877-771-7711.

To find these areas, there are several methods of evaluation.

1) A search technique to find groupings based on the difference between actual sales prices and the model predicted sales prices.
2) A technique that finds patterns of values, showing movement up or down in the differences between actual and predicted sales prices.
3) A technique that does not use the model, but rather the inputs to the model (housing characteristics such as age of the property, square footage, etc.), and determines if similar groupings are found where sales prices for houses are high or low relative to individual housing characteristics.
4) A technique that examines whether housing sales prices change over time at a different rate close to the site relative to similar areas far from the site.

This research method does not incorporate the specific location of the site until the last step of the process. We choose first to explore general value patterns, using a method that differentiates locations based on patterns found in the sales values of homes. In this way, an objective search is conducted that will find loss within groups of areas. If there is stigma associated with the properties near the site, this would surface in the analysis.

Information Available

There are several sources of property valuation data that can be examined for the purposes of litigation. The two most promising sources are the Realtors Multiple Listing Service and the County Property Appraiser's database. The choice between the two depends on the area. The MLS database can suffer from shortcomings when used for analysis. One primary shortcoming is that the MLS is restricted to properties listed with certified Realtors. It would not contain properties sold by the owner or which changed hands through other means. Additionally, the MLS may not have the desired historical detail needed. Depending on the area, the MLS lists only the asking price and not the actual sale price. Further, it may have a less complete accounting of descriptors of interest.

The County Property Appraiser's database usually is publicly available for a minimal price. However, the data in this database is retained for tax assessment purposes, not research, and so it may be incomplete with respect to important characteristics needed to predict the value of the property. Further, some counties retain the location of the property according to plat maps, but on electronic records only retain the street address of the owner of the property, who may be in a different state or at least at some other location than the property to be studied. The choice of a database has to be tempered by the difficulty associated with extracting the information needed.

Regardless of the source of the data, a county listing is too voluminous to realistically relate to any stigma associated with the events in a particular city or town, so some filtering is needed to identify properties to be chosen within the county. This can be done by drawing the smallest possible rectangle that could be fit around city boundaries containing the site in question and selecting all the properties encompassed by the rectangle. This process has the benefit of including non-incorporated areas associated with the city and of refining the data to just the area where a potential buyer may be searching for a home.

Predicting Sales Prices

Having gathered the information, we turn to exploring the geographical patterns of housing sales prices. In this analysis, we want to:

1) Predict sales prices of housing based on common characteristics,
2) Compare actual sales prices to predicted sales prices for each housing unit,
3) Find areas where actual sales prices were higher or lower than predicted,
4) Determine where these geographical patterns were relative to the site.

To predict sales prices for housing based on the characteristics of the housing, we used a technique called **multiple linear regression analysis**. Multiple regression is used by most appraisers and recommended by the International Association of Assessing Officers as a generally accepted method to evaluate trends in sales prices and to determine mass appraisal values. Using multiple regression, we can calculate a predicted sales price for each property in the city containing the site using information about the characteristics for the housing units, such as the age, the square footage, and the number of bedrooms for each unit.

Regression methods involve fitting a line through a set of points – the line that best explains the variability on the up and down axis, as in the chart to the right. The line can then be used to predict the value of the house, based on the values on the left to right axis. Points above the line have values higher than predicted; points below the line have values less than predicted. For example, if the slope of this line is two, a house with twice the square footage of another house would be expected to have twice the sales price.

Multiple regression methods do the same operation. In the example above, there was only one predictor of sales price: square footage. Multiple regression allows several predictors to be used simultaneously. The regression-predicted sales price is a useful way of determining what the sales price of a home is expected to be based on how characteristics of a house contribute to sales prices in the market overall. Using this method, for each housing unit included in our data set, we calculate the difference between the actual last sales price and the model predicted sales price. This difference indicates which properties sold for less than would be expected based on housing characteristics for each housing unit.

We examine which variables on the property tax database best explain the variability in the sale price of housing units. In a multiple regression to last sale price, we may include the following characteristics:

Square footage	Number of bedrooms
Date of sale	Number of bathrooms
Age of the residence	Whether the property was on a lakefront, river, or other waterfront
Lot size	Whether the housing unit was built in the last year

Note that the variability explained is based only on the indicators used in the analysis. In some situations involving mass appraisals, as many as 100 housing unit characteristics might be used, including indicators for special features of homes such as whether a swimming pool, a screened-in porch, or a garage are present, materials from which the house is made, and whether the property is at the intersection of two streets, among others. These items in other analyses can raise the amount of variability explained, which is our goal in this first portion of the analysis. We want the remaining variability in sales price to essentially be random. If there are components left that are not random, and if these components are related to proximity to the site in question, then this should show up when we compare actual and predicted sales prices.

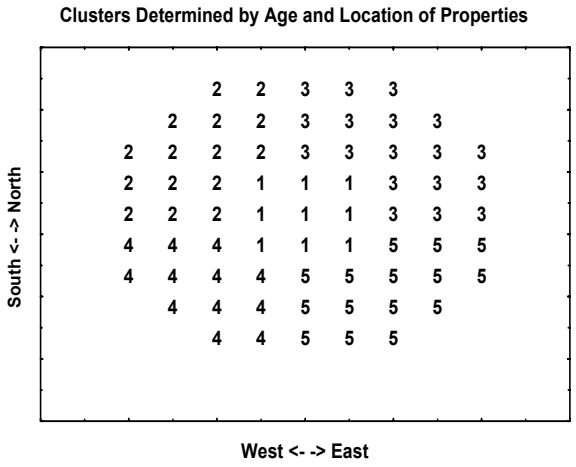
The results of this regression analysis are used to estimate a predicted sale price for properties based on the characteristics of each housing unit. The difference between the actual sale price and the predicted sale price is called a "residual". We calculate a residual for each property in the data set. Using the residuals, we identify the individual cases where the regression line was furthest from the true value of sales price. This is where proximity to the site would have its effect and would be shown by otherwise unexplained groupings of properties.

The Effect of Proximity to the Site on Housing Sales Prices

To explore whether the presence of the site has an effect on housing sale prices, we proceed with spatial analyses. These analyses include a clustering of the residuals, a contour plot of the residuals, and a cluster analysis of the sales price and housing characteristics. Each of these approaches is addressed in turn below.

Cluster Analysis

Cluster analysis is a classification method that seeks to form groups of individual units that are similar on two or more characteristics of interest. Units are placed in clusters so as to minimize differences between the units within a cluster and maximize differences measured on the same variables between clusters. In other words, the clusters that result are as homogeneous as possible.



As an example of clustering, consider the history of any small or medium sized city. Decades ago, a town was formed around a crossroads or at a landing point on a river. The town will have a few homes and some small buildings as a central point. Over time, from this core there will be new homes built. Building will progress in the direction of economic development. When the spaces around the core are fully developed, building will continue in other directions or further out.

A clustering program will use the location of the properties and the ages to find groupings of properties that were built at approximately the same time. In the example to the right, we have five clusters determined by the age of the properties and the location – a core of properties at the center, denoted with the number 1. After that, four more clusters are shown, numbered 2, 3, 4,

and 5. Note that clusters 2 and 3 may have been built at around the same time, but are denoted as two separate clusters because of the split between East and West locations. In this process, the separation of the clusters by age is as important as the separation of the clusters by location. For our analysis, instead of examining clusters of houses by their ages, we want to cluster properties by their relative value.

An important criterion in forming clusters is that no clusters of only one or two housing units results from the clustering operation. Clusters with only a few units indicate something unique about those housing units, not about the area in the city being analyzed. We find that the number of clusters overall is the number needed to adequately summarize patterns in the data, while producing the greatest possible differentiation between the clusters.

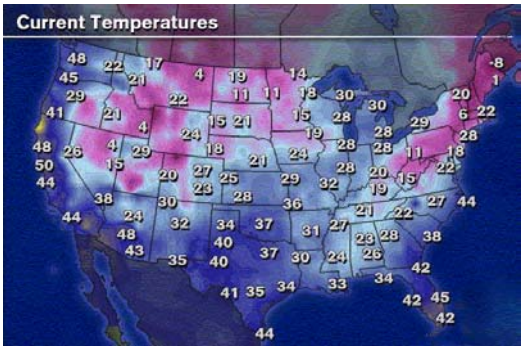
In the cluster analysis, we use only the:

- Latitude – North/South position of each property,
- Longitude – East/West position of each property, and
- Residual – the difference in actual and predicted price as calculated by the multiple regression model.

This model forces grouping of housing units by area, based solely on the differences between actual and predicted prices. The clusters of housing units differ only on location and on areas where differences are reflected in large positive or large negative residuals, or uniformly low residuals. We can examine each cluster and determine whether it shows that the actual and predicted prices are nearly equal, or if the differences for housing units included in the cluster are mostly negative or mostly positive.

Contour Plotting

Another way we can consider the changes in value for housing units in a city is to construct contour plots of the differences estimated from the pricing model. Again, the residuals represent the difference between the actual and the predicted housing unit prices. A contour plot works somewhat differently than a cluster analysis. The cluster analysis attempts to find discrete groupings of housing units that are as similar to one another as possible. Contour plots instead find patterns. The map uses two variables, latitude and longitude, to predict a pattern in a third variable, the size of the residual.



Everyone has seen contour maps – they are the charts that come in map books describing the height of a mountain or the depth of a valley. They are used for navigational purposes to show channels for ingress and egress to a port or a marina. A contour map connects areas that have similar values of a quantity being measured. The map of the United States at the right depicts variations in temperature. It is a contour plot showing variation in color bands for the U.S.

In this example, researchers did not measure the temperature of each square foot of the U.S.. Rather, temperatures were measured at weather stations, then a mathematical algorithm

computed the temperatures at the points in-between. In the map, the dark red areas represent the lowest temperatures while the blue sections of land areas represent the highest temperatures. Contour plotting of housing sales prices works similarly. We use the actual sales price data for housing, then an algorithm determines what the sales prices of the in-between areas are likely to be. The procedure fills in predicated sales prices of the in-between areas based on the known sales prices, assuming that the areas contain single family residential properties. Colors then indicate similar levels of sales prices and how they are dispersed geographically.

The key difference between contour plots and clustering is that the contour plots fill in expected values for areas between observations in the plot. In the cluster plot, we show only the observations according to the cluster with which they are associated. With thousands of housing units in a city, we have a fairly complete picture of housing and values. In the contour plot, we show how the differences between actual and predicted sales prices are increasing or decreasing.

Based on all the observations – units – in the city and the strength of the prediction, this gives us another way of viewing these differences in detail and identifying areas where the sale prices are lower than one would expect based on the characteristics of the housing.

An Alternative Clustering Method

In the first cluster analysis, the latitude and longitude of each property is included with the residual from the model. In this alternative cluster analysis, latitude and longitude are included in the clustering, as well as the housing unit characteristics, such as sale price, the age of the property, the number of bedrooms, the number of bathrooms, the square footage, the average price for properties sold the previous year, the year of the sale, and the land value as determined by the Property Assessor. The residuals used in the first cluster analysis are not included.

This second type of clustering is used to provide an alternative way to view the sale prices of housing. In this method, using the listed characteristics, if the pricing relationships are not described by a straight line, as required by regression analysis, this method would still locate clusters of housing units where anomalies existed between the sale price of the unit and the characteristics of the property. This is a strong advantage to this method since it eliminates the need for any modeling effort on price. If the pricing model were inaccurate or ineffective, this method could still find differences between pricing and associated characteristics.

The disadvantage of this method is that it will also find clusters or differences that have nothing to do with the pricing. For example, two clusters could be defined on age of the housing unit and the number of bedrooms, where in both clusters the houses were over 40 years old, but one cluster averages 2.5 bedrooms per housing unit and the other cluster averages 5.0 bedrooms per housing unit. A regression model might perfectly predict the sale price differences between these houses on the basis of square footage and the other characteristics, but for this example, cluster analysis would still find two clusters that differ only on size relative to age as the distinguishing characteristics.

The Outcome

The result is three different counts of properties:

1) the number of properties that exist in geographically close proximity to one another that share a single characteristic, namely that the price of these units is consistently above or consistently below what would be predicted for these units, based on their characteristics
2) the number of properties that exist in geographically close proximity to one another that share multiple characteristics, including the level of price when one considers the characteristics in common for these units
3) the number of properties contained in a circle, ellipse, or torus that is defined by the waxing and waning of relative property values when one compares actual sales prices to predicted sales prices.

The counts defined in these ways should be somewhat similar, but reflect different ways of summarizing the information common to the properties. Most important, all three of these methods use no information regarding the site under study. In this way, there is an objective search criterion employed to find whether housing prices are depressed near a site, and if so, where they are depressed and **how many** are lower than expected.

A Reality Check

A final, less formal method is available to search for areas impacted by proximity to an environmental disamenity. From the regression described above, the residuals are coded to two values: greater than zero and less than zero. These are plotted on a map of the area under study, with the greater than zero values indicated with a blue marker and the less than zero values indicated with a red marker.

A plot of these markers overlaid on a map would readily show patterns of loss or gain in the area under study. Of greatest interest is whether there is clustering of the red or the blue values in any part of the map, whether such clusters can be found near the environmental site, and whether such clusters are sufficiently large to warrant concerns about a class of properties. There will be some natural clustering in any scattering of a large number of properties – the question is whether the clusters observed in this way are larger than would be expected under normal entropy conditions and whether such clusters relate to groundwater flows or air plumes.

The advantage of this method is that it is very easily understood by any reviewer or interested party, such as a judge or a jury listening to an explanation of where loss might exist in a community. The same is true for the clustering and plotting techniques presented above. The key in any such analysis is to make sure that the methods and presentation is completely transparent. Anyone should be able to understand immediately what was done and what the results are; any competent researcher should be able to replicate the analysis. Simplicity and transparency are the two essentials for a presentation in any litigation setting.