Evaluating Environmental Stigma with Multiple Regression Analysis

by Thomas O. Jackson, PhD, MAI

This edition of "Environment and the Appraiser" continues our series on generally accepted methods and techniques for evaluating the effects of environmental contamination on the value of real property. The series began with an overview of these techniques and methods and an introduction to paired sales analysis, environmental case studies, multiple regression analysis, market interviews, and income capitalization analysis.[1] Subsequently, environmental case studies and market interviews were addressed in more detail.[2] In each of these articles, specific examples of the application of the techniques were provided. In keeping with this approach, this column will discuss the technique of multiple regression analysis generally and then present an example of its use in analyzing potential proximity stigma effects.

As with the other techniques that have been discussed, the primary focus will be on analysis of the effects of increased environmental risk or stigma. As previously explained, the market value of real property can be affected by three potential effects: *cost effects*, or deductions for costs to remediate a contaminated property to appropriate regulatory standards; *use effects*, or limitations on the highest and best use of the subject property due to environmental contamination and its remediation; and *risk effects*, or the effects on value due to the market's perception of increased environmental risk and uncertainty.[3] The effect of increased environmental risk on property value is referred to as *environmental stigma*.[4] Further, and as noted in Advisory Opinion 9, the "analysis of the increased environmental risk and uncertainty on property value (environmental stigma) must be based on market data, rather than unsupported opinion or judgment."[5] The "market data" referred to in the advisory opinion is sales data. Thus, multiple regression analysis in this context analyzes sales data to estimate the risk-related effects on property value known as environmental stigma.

Appraisers contemplating the use of multiple regression analysis in evaluating environmental stigma, or for any appraisal-related purpose, should remember the requirements in the COMPETENCY RULE of the *Uniform Standards of Professional Appraisal Practice* (USPAP).[6] To review, appraisers who do not have sufficient training and experience "must: (1) disclose the lack of knowledge and/or experience to the client before accepting the assignment; (2) take all steps necessary or appropriate to complete the assignment competently; and (3) describe the lack of knowledge and/or experience and the steps taken to complete the assignment competently in the report."[7] These requirements would apply to appraisers who had limited experience and/or training in the use of multiple regression analysis generally and as applied to properties that may be impacted by environmental contamination.

1. Thomas O. Jackson, "Methods and Techniques for Contaminated Property Valuation," The Appraisal Journal (October 2003): 311–320.
2. Thomas O. Jackson, "Case Studies Analysis: Environmental Stigma and Monitored Natural Attenuation," *The Appraisal Journal* (Spring 2004): 111–118; and Thomas O. Jackson, "Surveys, Market Interviews and Environmental Stigma," *The Appraisal Journal* (Fall 2004): 300–310.
3. Appraisal Standards Board, Advisory Opinion 9, "The Appraisal of Real Property That May Be Impacted by Environmental Contamination," *Uniform Standards of Professional Appraisal Practice*, 2005 ed., Lines 171–180 (Washington, DC: The Appraisal Foundation, 2005), 143–147.
4. Ibid., Lines 92–93.
5. Ibid., Lines 180–182.
6. Appraisal Standards Board, COMPETENCY RULE, Lines 356–397.
7. Ibid., Lines 359–364.

Multiple Regression Analysis and Real Estate Appraisal

Multiple regression analysis is an increasingly used technique in the real estate appraisal field. The Appraisal Institute has two books on this topic as it relates to automated valuation modeling.[8] In addition, the Appraisal Institute offers a seminar entitled *Regression Analysis in Appraisal Practice: Concepts and Applications*. Most universities offer statistics courses that cover regression analysis through either their statistics or economics departments. There is an advisory opinion that addresses the use of automated valuation models (AVMs).[9] This advisory opinion lists regression analysis as a method that may underlie an AVM, but notes that "The output of an AVM is not, by itself, an appraisal. An AVM's output may become a basis for appraisal, appraisal review, or appraisal consulting opinions and conclusions if the appraiser believes the output to be credible and reliable for use in a specific assignment."[10]

In other words, the estimates produced by a regression analysis of sales data for properties that may be impacted by environmental contamination are not an appraisal, but can provide the basis for an opinion as to these impacts. The appraiser must interpret the results of the modeling process and apply those results to the facts of the situation being analyzed. The appraiser must also have sufficient basis for concluding that the results are credible and reliable. This could involve statistical testing of the results to ensure that they are not biased in a statistical sense and that the model is adequately specified. Complex multiple regression analysis models can produce results that do not reflect the market's reaction to a specific environmental condition or influence, but may reflect other influences masquerading as adverse environmental effects on property values.

Multiple Regression Analysis and the Sales Comparison Approach

Ultimately, all generally accepted appraisal techniques must be related to one of the three approaches to value: sales comparison approach, income capitalization approach, or cost approach. The use of multiple regression analysis has been likened to a form of the sales comparison approach. In a sale price regression analysis, the sale price (referred to as the dependent variable) is modeled as a function of a number of variables reflecting the property's physical and market characteristics (referred to as independent or predictor variables). The analysis estimates coefficients for each of the independent variables. Linné, Kane, and Dell note that the "adjustments in the sales comparison approach are analogous to the coefficients in a regression analysis model, but the two sets of values are not usually equivalent."[11] The lack of equivalence could be due to correlations between the independent variables (multicollinearity) or omitted variables (specification error), as will be discussed.

The use of regression analysis in relation to the sales comparison approach, and in particular the technique of analyzing sales through an adjustment grid (the "grid method"), has been discussed in the academic literature for some time. In 1983, Colwell, Cannaday, and Wu noted that "the grid method is shown to be less biased than the pure regression method" due to the omitted variable problem, but that regression-based "hedonic price functions (price as a function of property and other characteristics, added) underlie both grid and regression approaches to appraisal."[12] Later, Kang and Reichert analyzed the use of regression model coefficients in making adjustments within the appraisal-grid framework and compared this hybrid technique to standard regression models.[13] They found that where markets are in equilibrium and have homogenous housing and neighborhood characteristics, the grid method with regression-based ad-

8. Mark R. Linné, M. Steven Kane, and George Dell, *A Guide to Appraisal Valuation Modeling* (Chicago: Appraisal Institute, 2000); and M. Steven Kane, Mark R. Linné, and Jeffrey A. Johnson, *Practical Applications in Appraisal Valuation Modeling* (Chicago: Appraisal Institute, 2004).
9. Appraisal Standards Board, Advisory Opinion 18, "Use of Automated Valuation Model (AVM)," 178–185.
10. Ibid., Lines 18–19.
11. Linné, Kane, and Dell, 49.
12. Peter F. Colwell, Roger E. Cannaday, and Chunchi Wu, "The Analytic Foundations of Adjustment Grid Methods," *Journal of the American Real Estate and Urban Economics Association* 11, no.1 (1983): 27–28.
13. Han-Bin Kang and Alan K. Reichert, "An Empirical Analysis of Hedonic Regression and Grid-Adjustment Techniques in Real Estate Appraisal," *Journal of the American Real Estate and Urban Economics Association* 19, no. 1 (Spring 1991): 70–91.

justments is the preferred approach, while in "less homogenous markets with significant price variation" a "straight regression" using a log-linear form[14] is "more appropriate."[15] Likewise, Pace and Gilley found that "neither OLS[16] nor the grid estimator excel in all circumstances."[17] Lai and Wang compare the adjustment grid method, but with weighting of the comparables (minimum-variance grid method[18]) to multiple regression; they conclude that while both are unbiased, the minimum-variance grid method is preferred because of a lower variance in its prediction error.[19] Lastly, Lipscomb and Gray compare various methods of estimating sale price adjustments, including regression-based adjustments and adjustments derived through various types of matched pair analysis.[20] They find that the standard matched pair analysis approach (in which the average price difference between matched pairs is used to estimate adjustments)[21] and matched pair regression (with sale price as the dependent variable, physical characteristics as independent variables and a dummy variable for the feature of interest)[22] "produce the same adjustment estimate," but that the matched pair regression is preferred because it is simpler to perform and provides a standard error of its estimates.[23] Lai and Wang also noted the advantage of regression analysis in being able to estimate a confidence interval (based on standard errors) and perform a hypothesis test of the "true property value."[24]

Regression Models for Evaluating Environmental Stigma[25]

General Model Specification

The basic specification of a multiple regression model for analyzing the risk-related effects of environmental contamination on sale price is as follows:

$$Price = \beta_O + \beta_1 X_1 + \ldots + \beta_n X_n + \beta_{n+1} ENV_1 + \ldots + \beta_{n+1+p} ENV_p + \varepsilon,$$

where:

$Price$ = the sale price of the property; adjusted for the estimated remediation costs to be paid by the buyer, which are known or can be estimated as of the date of sale. Uncertainty concerning these costs would be a risk factor included in the effects captured by the ENV variable(s). Could be transformed through logarithmic transformation (in which case interpretation of independent variable coefficients would be different)

β_O = a constant term

$X_1 \ldots X_n$ = a vector of nonenvironmental property characteristics such as building size, age, lot size, etc. Could also include property location and date of sale for property-level model type (discussed next)

$ENV_1 \ldots ENV_p$ = a vector of discrete and/or continuous variables indicating the environmental condition of the property at the time of sale and corresponding to the risk-related (stigma) effects of the contamination. Could include location for control and proximity models and/or remediation status for property-level model (discussed next)

ε_i = a random error term

14. Logarithm of sale price as the dependent variable.
15. Kang and Reichert, 89.
16. OLS stands for ordinary least squares.
17. R. Kelley Pace and Otis W. Gilley, "Generalizing the OLS and Grid Estimators," *Real Estate Economics* 26, no. 2 (Summer 1998): 331–347.
18. The minimum-variance grid method was originally proposed by Kerry D. Vandell, "Optimal Comparable Selection and Weighting in Real Property Valuation," *Journal of the American Real Estate and Urban Economics Association* 19, no. 2 (Summer 1991): 213–239.
19. Tsong-Yue Lai and Ko Wang, "Comparing the Accuracy of the Minimum-Variance Grid Method to Multiple Regression in Appraised Value Estimates," *Real Estate Economics* 24, no. 4 (Winter 1996): 531–549.
20. Joseph B. Lipscomb and J. Brian Gray, "An Empirical Investigation of Four Market-Derived Adjustment Methods," *Journal of Real Estate Research* 5, no. 1 (Spring 1990): 53–66.
21. Ibid., 56.
22. Ibid., 57.
23. Ibid., 65.
24. Lai and Wang, 548.
25. Portions of this section are from Jackson, "Methods and Techniques for Contaminated Property Valuation," 316.

In this model specification, the nonenvironmental characteristics that influence sale price are independent, or predictor, variables in the equation. In this way, the variation in sale price "explained" by the nonenvironmental variables (size, age, etc.) would not be incorrectly attributed to the environmental condition variables being tested in the model (distance from contamination source, remediation status, location in contaminated neighborhood, etc.). An analysis of the statistical significance of the environmental condition variables would indicate whether there was adequate statistical evidence to conclude that there were significant environmental impacts on value. In areas with multiple adverse influences and/or diverse submarkets and property types, it may not be possible to reliably estimate the effect of a single contamination source through regression analysis or a common effect on values across a single area.

The omitted variable problem mentioned earlier refers to important predictor variables that are not included in the mode specification. For analyzing environmental impacts, this would present a problem if the omitted variables were somehow correlated with the environmental condition variable of interest (*ENV*), and the effects on property value were incorrectly interpreted as being due to environmental condition rather than the omitted variable. Otherwise, if a variable that could be a significant predictor was omitted, then its influence could be captured by one of the other variables in the model. For example, lot size is frequently missing from multiple listing service (MLS) information and other property records. The influence of lot size on price could be picked up and indirectly accounted for by house size since they tend to be correlated. The model's overall prediction of sale price would still be unbiased; only if lot size correlated with environmental condition would omitting that variable bias the estimated coefficient for environmental condition.[26]

As the following discusses, the general model specification can be used in three types of multiple regression analyses to estimate the impacts of risk-related effects of environmental contamination on properties in an area around or near a contamination source.

Property-Level Model

The first approach and model will be referred to as the *property-level model*. Examples of this model approach and specification in the analysis of contaminated industrial properties are provided by Jackson.[27] In Jackson's models, the risk effects of the properties' environmental condition on sale price were analyzed before, during, and after remediation (ENV_1, ENV_2, and ENV_3). Included in the sales analyzed were unimpaired comparable properties, so that the impacts on sale price due to environmental condition were analyzed relative to otherwise similar but uncontaminated properties. Also included are categorical (dummy) variables for the location of the property and the year of sale. The focus of these analyses was on changes in stigma or risk-related effects due to remediation status, rather than on allegations of impacts on properties in an area that may have become contaminated from a common source (non-source properties).[28]

Proximity Analysis

In a *proximity analysis*, the regression model is usually specified so that one of the independent variables or a set (vector) of independent variables reflects the distance of each of the sale properties analyzed from the source of the environmental contamination. These variables can be specified as continuous distance from the contamination source or as discrete distance bands, or concentric bands, around the source.

Before drawing conclusions from such an analysis, the appraiser should consider the possibility that multiple adverse influences on sale price might exist in areas with a number of contamination sources or other disamenities. In such situations, it may be difficult or impossible to sort out the relative influence of any one source as distinct from the others. Another limitation on this type of analysis involves the general tendency for residential properties closer to older industrial facilities and landfills to sell for less than otherwise similar properties located farther away, regardless of whether the facilities have released any environmental contamination. In this situation, lower sale prices closer to an industrial facility or landfill might not be due to hazardous environmental contaminants, but could be incorrectly interpreted as such.

26. A priori, it is hard to imagine lot size and environmental condition being somehow related, but the appraiser must be aware of this issue when data are missing.

27. Thomas O. Jackson, "The Effect of Previous Environmental Contamination on Industrial Real Estate Prices," *The Appraisal Journal* (April 2001): 200–210; and Thomas O. Jackson, "Environmental Contamination and Industrial Real Estate Prices," *Journal of Real Estate Research* 23, no.1/2 (2002): 179–199.

28. See definitions of *source*, *non-source*, *adjacent*, and *proximate properties* in Advisory Opinion 9, Lines 104–108.

Control Area Analysis

A multiple regression *control area analysis* can be used to analyze the effects of contamination on properties in a neighborhood area where it is claimed that property values have been diminished because of environmental stigma. In this type of analysis, sale prices of properties in the potentially impacted area, referred to as the "subject area," are compared to prices of similar properties in a comparable neighborhood having the same characteristics as the subject area but without the adverse environmental condition under study. This comparable neighborhood is referred to as the "control area." In many such analyses, the locational influences of the subject and control areas are compared before and after a contamination event. Such events could be the actual release of the contamination or a public announcement of the release. Typically, such events are publicized in the media.

Issues in developing a reliable control area analysis involve potential time and area interactions and the influence of confounding nonenvironmental factors. In comparing two or more areas, even well-matched areas can be influenced by differing market and locational conditions over time, and these differing influences may be incorrectly attributed to the adverse environmental condition under study.[29] However, the subject and control areas do not need to be identical, but should be influenced by the same general market conditions over time so that changes in relative pricing could be appropriately attributed. Thus, the initial selection of the control areas is a critical step in this type of analysis. In general, the subject and control areas for the analysis should represent definable neighborhoods or market areas for the subject's property type.[30]

Regression Analysis Application: Alleged Odor Impacts on Residential Properties

To illustrate the application of multiple regression analysis in evaluating environmental impacts, analyses from an actual situation are presented below. The contamination source was a sewer lift station that had allegedly emitted odors, which could potentially diminish property values due to a stigma effect in the surrounding neighborhood. The neighborhood is located in a Dallas suburb, and the properties were all owner-occupied, single-family residences. As of the date of value in 2004, there were no reported odors and there were no expenditures to remediate the residential properties or to repair the lift station. The odors had occurred previous to this time, and the lift station had been repaired at the operators' expense. Thus, any impact would be risk or stigma related. Two types of analyses were performed, a control area analysis and a proximity analysis; the following discusses the results of these analyses.

Control Area Analysis

The subject area for this analysis was delineated on the basis of the locations of those residents claiming their properties had been adversely affected, and then working outward to approximate a neighborhood. A smaller area drawn more compactly around the houses reporting odors was also analyzed, with results similar to those for the larger neighborhood area. The control areas for this analysis were delineated on the basis of field observation of houses in the vicinity of the subject areas, following accepted guidance for such determinations. While no two areas are identical, these areas appeared to be generally similar in housing type, age, and other characteristics. This was confirmed by statistical summaries of housing data for the subject and control areas.

Table 1 presents a basic control area analysis of differences in sale price between the subject and control areas. The parameter estimates correspond to relative price differences due to each variable. For example, the analysis indicates that on average, each square foot of house size increases sale price by $36.04 and for each additional year a house ages, its sale price declines by $1,067. A swimming pool is shown to add $18,095 to sale price, although this variable may reflect additional amenities that could cluster with pools. The model explains more than 80% of the variation in sale price and overall is statistically significant at acceptable levels.[31] The analysis is based on 1,426

29. Warren Rogers, "Errors in Hedonic Modeling Regressions: Compound Indicator Variables and Omitted Variables," *The Appraisal Journal* (April 2000): 208–213.

30. Neighborhoods and market areas are discussed in Chapter 8 of *The Appraisal of Real Estate*, 12th ed. and in other academic and professional literature. Appraisal Institute, "Market Areas, Neighborhoods, and Districts" in *The Appraisal of Real Estate*, 12th ed. (Chicago: Appraisal Institute, 2001), 163–187.

31. The base models were tested for residual normality, multicollinearity, and heteroscedasticity, and no problems were indicated.

sales. These were all the sales available from the MLS system for the January 2000 to November 2004 period that had complete data on the variables in the analysis.[32]

If the market value of properties in the subject area had been reduced by stigma related to the odors, this would be reflected in reductions in prices in the subject areas relative to prices of otherwise similar housing in the control areas. The results of the statistical analyses and tests in Table 1 show just the opposite. The analysis in Table 1, comparing sale prices for the subject area to the control areas indicates that prices of houses in the subject area are $2,129 greater on average than houses in the control areas, holding constant or controlling for the effects of the other variables in the model such as house size and age. This premium is small compared to the average sale price in the subject area of $168,497. However, given the large number of sales in the analysis, this premium is statistically significant.[33] Accordingly, the results of comparing the prices of sales in the subject and control areas over the 2000 to 2004 period do not show any measurable or discernable reduction in prices in the area around the sewer lift station and in the broader neighborhood surrounding it. Indeed, there is even a slight subject area sale price premium. This is the opposite pattern from what might be found if property values in the subject area had been adversely affected.[34]

Proximity Analysis

The second test of potential adverse impacts on sale price due to the lift station involves an analysis of any sale price differences due to proximity to the lift station site. To accomplish this analysis, the distance of each property sold in the subject area from the lift station was measured and entered as a variable in the multiple regression analysis. The results are presented in Table 2. As shown there, the coefficient for distance (in meters) from the lift station for the 311 sales in the area is -0.79, indicating that prices decrease by $0.79 for each meter of distance from the lift station. With a corresponding *t*-statistic of -0.412 and a *p*-value of 0.6810, this is a statistically insignificant relationship. In other words, there is no relationship between distance from the lift station and sale price.[35] Prices in the subject area did not vary with distance from the lift station, indicating no adverse pattern of sale price discounts or value diminution that could be attributable to proximity to the lift station. If there had been adverse impacts due to any gasses released from the lift station, the prices would show a significant decline with distance.

Stated Preferences, Revealed Preferences, and Market Knowledge

In a previous edition of "Environment and the Appraiser," we discussed the difference between stated and revealed preferences.[36] In general, sale prices

Table 1 Multiple Regression Analysis of Housing Prices, Subject Area Relative to Control Areas

Variable	Parameter Estimate	*t*-Statistic	*p*-Value
Intercept	$74,915.65	55.744	0.0001
House size (square feet)	36.04	66.756	0.0001
Age of house (in years)	-1,067.32	-16.964	0.0001
Year of sale*	3,349.06	18.505	0.0001
Swimming pool	18,095.18	24.354	0.0001
Subject area 1	2,128.93	3.320	0.0010
Adjusted R^2	0.830		
F-value	1389.979		
p-value	0.0001		

* 0 = 2000; 1 = 2001; 3 = 2003; 4 = 2004
Note: Analysis based on 1,426 sales of single-family residential properties from January 2000 to November 2004.

32. Five sales were eliminated due to missing data on one or more of the variables in the analysis.
33. When the model is reestimated in log-linear form, the premium becomes statistically insignificant.
34. This result was also tested over time for each year from 2000 to 2004 through an analysis of covariance procedure and estimated marginal means. The findings on a year-by-year basis indicated that price differences were generally not statistically significant. In addition, each individual control area (there were three) was tested against the subject area. No patterns of reduced sale prices were evident.
35. In the distance model, the coefficients for house age and year of sale are less significant than they were in the previous models. The houses in these areas were built about the same time and generally sold about the same time so there is less variation in price due to these factors.
36. Jackson, "Surveys, Market Interviews, and Environmental Stigma."

Table 2 Multiple Regression Analysis of Housing Prices, Proximity to Sewer Lift Station in Subject Area

Variable	Parameter Estimate	*t*-Statistic	*p*-Value
Intercept	$77,914.79	23.739	0.0001
House size (square feet)	37.53	29.892	0.0001
Age of house (in years)	-373.26	-0.864	0.3880
Year of sale	631.78	1.120	0.2630
Swimming pool	22,042.91	7.583	0.0001
Distance to lift station	-0.79	-0.412	0.6810
Adjusted R^2	0.766		
F-value	203.817		
p-value	0.0001		

* 0 = 2000; 1 = 2001; 3 = 2003; 4 = 2004

Note: Analysis based on 311 sales of single-family residential properties in subject area from January 2000 to November 2004. Distance from lift station is measured in meters.

are considered indicators of the preferences of the market as "revealed" in an actual transaction. These transactions ultimately establish market value. Stated preferences, on the other hand, are the preferences of the market that may be elicited through various forms of survey research. These preferences may or may not be consistent with revealed preferences (i.e., transactions). When stated and revealed data differ, preference should be given to the transactional data. Even in situations where there may be adverse perceptions, these perceptions must be acted upon and become revealed for market values to be affected. Appraisers and others who base their conclusions entirely on stated preference data rather than sales data are speculating as to impacts that have not occurred, and in some cases substituting their judgment for that of the market. As we know, the market establishes value, not the appraiser.

In the odor analysis example presented above, a survey of a nearby neighborhood was conducted to determine if there would be any preference for buying a house in the subject area given its history of odors from the sewer lift station. None of those interviewed indicated that they would buy a house there, so arguably property value impacts should have been quite severe. This finding, of course, is directly contradicted by the hundreds of actual transactions that occurred after the release of the odor from the lift station and after the lift station was repaired. Indeed, subsequent to the odor reports, builders in the subject area disclosed the odor issues and even required potential homebuyers to sign statements as to their knowledge of the situation. The divergence between this revealed and stated preference data is consistent with the observations of the late William Kinnard that "The results from survey analyses must be tempered with the knowledge that the expectation of events is almost invariably more negative and more sharply delineated, at least when [the events] are expected to affect oneself negatively, than is realized when the event occurs."[37]

Thomas O. Jackson, PhD, MAI, CRE, is the director of real estate programs and a clinical associate professor in the Department of Finance of the Mays Business School at Texas A&M University, where he teaches real property valuation in the Land Economics and Real Estate Program. In addition, he is the president of Real Property Analytics, Inc., based in College Station, Texas, where he specializes in analyzing the effects of environmental contamination on real property. **Contact: T 979-690-1755; E-mail: tomjackson@real-analytics.com; Web site: www.real-analytics.com**

37. William Kinnard, Jr., "Measuring the Effects of Contamination on Property Values: The Focus of the Symposium in the Context of Current Knowledge," in *Measuring the Effects of Hazardous Materials on Real Estate Values: Techniques and Applications* (Chicago: Appraisal Institute, 1992), 5.