2006 V34 1: pp. 77–107
REAL ESTATE
ECONOMICS

The Impact of Environmental Contamination on Condo Prices: A Hybrid Repeat-Sale/Hedonic Approach

Bradford Case,* Peter F. Colwell,** Chris Leishman*** and Craig Watkins****

We extend the literature on the impact of externalities using an approach based on a hybrid of hedonic and repeat-sales methods. The externality in question is groundwater contamination in Scottsdale, Arizona. The use of condominium sales allows us to assume that major physical characteristics remain unchanged, but location parameters may be altered by urban growth and development as well as contamination. We find an economically significant discount for properties located in the contaminated area. Interestingly, it does not appear until several years after the contamination becomes publicly known, and it seems to have disappeared before the end of the study period.

The last three decades have seen the emergence of a voluminous literature examining the interaction between environmental factors and real estate markets, much of it concerning the extent to which negative environmental spillovers are capitalized into real estate values. These interests have given rise to two broad categories of real estate literature reviewed in Jackson (2001). The first of these is dominated by the appraisal profession and has focused on valuation concepts and methods (see, *e.g.*, Chalmers and Jackson 1996, Jackson 1997, Weber 1997, Kinnard and Worzala 1999). The second category reports the results of empirical studies for the effects of environmental spillovers on real estate prices, especially in residential markets. The forms of environmental spillovers considered relate to land, water and air (Segerson 2001) and have included proximity to landfill sites (McClelland, Shulze and Hurd 1990, Kohlhase 1991, Thayer, Albers and Rahmatian 1992) as well as the presence of overhead power lines (Colwell 1990), air pollution (Graves *et al.* 1988), water quality (Leggett and Boekstael 2000), noise pollution (Pennington, Topham and Ward 1990) and groundwater contamination (Kiel 1995, Kiel and McClain 1996).

*Federal Reserve Board, Washington, DC 20551 or Bradford.Case@frb.gov.
**Department of Finance, University of Illinois, Champaign, IL 61820 or pcolwell@uiuc.edu.
***School of Built Environment, Heriot-Watt University, Edinburgh EH14 4AS or c.m.leishman@hw.ac.uk.
****Department of Town and Regional Planning, University of Sheffield, Sheffield S10 2TN or c.a.watkins@sheffield.ac.uk.

There are three general scenarios to describe the response of real estate prices to an environmental spillover. In the most straightforward, the presence (or severity) of an environmental problem is associated simply with the constant presence (or severity) of a price discount on properties affected by that problem. For example, properties affected by environmental contamination could be priced at a discount (perhaps equal to a constant percentage of value) relative to unaffected properties, or the discount could be a linear function of the severity of the contamination affecting that property. As the discount is related only to the presence (or severity) of contamination, it remains constant until the contamination is mitigated (wholly or to some degree).

The presence or severity of contamination, though, can be considered a (negative) hedonic attribute of the property just like any other, and therefore the implicit market price associated with contamination may change over time just as it does for other hedonic attributes. In the second scenario, then, the presence (or severity) of the environmental occurrence is associated with a price discount on affected properties, but the magnitude of the discount evolves in the same way that implicit marginal prices on other hedonic attributes evolve. For example, if the implicit discount associated with environmental contamination is elastic with respect to wealth, then the magnitude of the discount can be expected to grow over time as wealth increases (controlling for any mitigation). Alternatively, if information about environmental contamination dissipates through time then the discount may itself dissipate even if the presence and severity of contamination do not.

Neither of these scenarios, however, takes into account the dynamic response of housing consumers to an occurrence such as environmental contamination. Consumers who attach a relatively high premium to environmental factors can be expected to bid (or ask) relatively high prices for properties with favorable environmental attributes, and to bid (or ask) relatively low prices for those with environmental problems. Therefore, these environmentally sensitive consumers will tend to outbid less sensitive consumers for properties with favorable environmental attributes, and will tend to be outbid by less sensitive consumers for properties with environmental problems. In response to the sudden occurrence (or recognition) of an environmental problem, the outcome of such a market adjustment or “sorting” process is likely to be this: the households that are most sensitive to the occurrence will offer their properties at the greatest discount in order to leave immediately; they will be replaced by households with the least sensitivity to the occurrence; this process will be repeated at successively smaller magnitudes for successively less sensitive sellers and more sensitive buyers; and thus the initial discount will decline steadily in magnitude, even in the absence of any mitigation of the environmental contamination. Of particular interest in this article is the application of house price modeling to assess whether transaction prices reveal the effects of such a sorting process, and if so how long the adjustment process seems to take.

In a recent review article, Boyle and Kiel (2001) examine more than 30 papers that use hedonic approaches to test the effects of environmental externalities on house prices. They note that a common problem is that the tests are often conducted over short periods and as a consequence are generally unable to effectively capture changes in price over time. The main problem that Boyle and Kiel identified is the absence of data from before and after an environmental event, which means that not even the immediate effect of the event on prices can be adequately identified. (A solution to this problem is found in Colwell, Dehring and Lash (2000).) Boyle and Kiel also, however, note a common inability to detect any change in the contamination effects after the initial contamination event—that is, the studies they examined tended to focus on the first and most straightforward scenario without considering any evolution of the price effect over time.

Although interest in the persistence of contamination effects and their longer-term impact on property values is not new (Riechert 1999), these effects have been difficult to determine in hedonic studies, so previous studies provide conflicting evidence of the temporal effects of contamination and remediation (see Kiel 1995 versus Kohlhase 1991 and Kiel and McClain 1996). Some researchers (Mieszkowski and Saper 1978, Hite *et al.* 2001) have tried to use control areas to infer what would have happened had the externality not occurred. The success of this approach is, of course, dependent on the researcher's ability to limit the differences in *inter alia* market conditions and neighborhood quality between the control and study areas or to accommodate the differences by including sufficient spatial variables. Dale *et al.* (1999) attempted a variant of the control area approach by examining various neighborhoods and found that after a cleanup house prices recover more slowly for dwellings in closest proximity to the negative externality.

In contexts other than contamination, real estate analysts have used repeat-sales regression analysis to examine temporal change in property values. Hedonic analysis is superior to repeat-sale analysis when there are few repeat sales, but repeat-sales analysis may be superior to hedonic analysis for this purpose when there are sufficient repeat sales, when data on some of the key property attributes are unavailable and when the analyst can be confident that those attributes will not have changed over time. Changing values for property attributes or model parameters can be modeled within a hedonic price model, but early applications of the repeat-sales model used simple versions that lacked this capability. A notable exception was Palmquist (1982), which incorporated changing values of an indicator of environmental quality (noise pollution), but subsequent studies failed to build on this observation. Shiller (1993), however, showed how changes in measured attributes or in parameter values could be incorporated more generally into the repeat-sales model, thereby removing a key advantage of the hedonic price model.

In this article, we employ the hybrid repeat-sale/hedonic approach suggested by Palmquist (1982) and Shiller (1993) to investigate the effect of environmental contamination in a particularly rewarding situation. In this application the timing of the first perception of contamination is well known while changes in unmeasured attributes are likely to be minimal, which means that observed changes in prices can be reliably decomposed into those attributable to contamination and those attributable to changing marginal prices in the market. Moreover, our data collection period is quite long, especially after the contamination was recognized, which enables us to investigate whether the price effect of the initial contamination persists, grows or dissipates over time.

The remainder of the article is divided into four major sections. The next section introduces the hybridization of hedonic and repeat-sale models. The third section discusses the data and the specific models to be estimated, while the fourth section discusses the empirical results. Our findings are summarized in the concluding section.

Hedonic and Repeat-Sale Models

Hedonic house price studies have been employed for nearly four decades to assess the impact of negative externalities such as air pollution (Ridker and Henning 1968). The hedonic method has subsequently yielded a vast applied literature, the basic premise of which is that by estimating the implicit price of each of the physical and locational attributes associated with a property it is possible to isolate the impact of environmental events on the price surface. In this context, estimates of the price surface from the period before and after the event are essential. For example, if the producers of negative externalities select sites in less desirable neighborhoods to reduce costs, then without before-and-after data lower prices might be interpreted as evidence of negative externalities while they may reflect only preexisting market conditions. As we note above, however, often the paucity of data imposes a constraint on before-and-after analysis.

Analysts tend to use repeat-sale analysis to estimate price indices when they do not have sufficient attribute data to use hedonic analysis for this purpose, but there are several dimensions to the downside of using repeat-sale analysis. First, data on assets that sold only once during the study period are ignored. Second, several empirical studies (Mark and Goldberg 1984 and Case, Pollakowski and Wachter 1991) have concluded that repeat sales understate house price inflation by failing to account for aging and depreciation between sales (Clapp and Giacotto 1992), overrepresentation of starter homes among repeat transactions (Clapp, Giacotto and Tirtiroglu 1991) and overrepresentation of "lemons" (properties that are substandard in some way) (Clapp and Giacotto 1992). Curiously perhaps, others (Crone and Voith 1992, Gatzlaff and Ling 1994) have found evidence that repeat-sales indices can overstate, rather than understate, price appreciation by failing to control for fix-up costs and the high representation of short hold properties (Steele and Goy 1997, Clapp and Giacotto 1999, Costello 2002). Advances in repeat-sales modeling have begun to ameliorate some of these problems: Cannaday, Munneke and Yang (2002), for example, developed a means of disentangling age effects from temporal effects.

The empirical approach employed in this article can be thought of as a hybrid model incorporating hedonic characteristics in a repeat-sales framework. Although a hybrid approach has been used in other studies, the focus of these papers has tended to be on technical aspects of index construction, rather than the assessment of the impact of externalities (Palmquist 1982, Case and Quigley 1991, Case, Pollakowski and Wachter 1991, Shiller 1993, Quigley 1995, Meese and Wallace 1997, Hill, Knight and Sirmans 1997, Englund, Quigley and Redfearn 1999).

The Foundations of Repeat-Sale Analysis

Repeat-sale analysis is derived from hedonic analysis. The standard repeat-sale analysis has two implicit assumptions: (1) attributes are constant through time and (2) parameters are constant through time. Deriving repeat-sale analysis from its foundations in hedonic analysis reveals that if either of these assumptions is violated, the repeat-sale analysis is suspect. Suppose that a hedonic regression equation has the following form:

$$P_i = \gamma X_i^{\alpha_1} e^{\beta_1 Y_i + \tau_1 T_{i1} + \tau_2 T_{i2} + \cdots + \tau_n T_{in}}, \tag{1}$$

where P_i is the price of property i, X_i is a prototypical, unchanging attribute of the property, Y_i is a second prototypical, unchanging attribute of the property $T_{i\Theta}$ is a dummy time variable (with Θ being a discrete indicator of year) such that

$$T_{i\Theta} = \begin{cases} 1, & \text{if } \Theta = t_i \\ 0, & \text{if } \Theta \neq t_i \end{cases} \quad \text{and } t_i \text{ is the year of sale of the } i\text{th property.}$$

Of course, one of the set of time dummy variables must be omitted from the regression equation. Let us assume that the omitted dummy is T_{i0}. Now suppose that an asset sells twice, once at year t and once at an earlier year $\tilde{t}$ where the tilde denotes the earlier magnitudes of this and the other variables and where the i subscript denotes the sale pair. The ratio of the two predicted prices would be

$$\frac{P_i}{\tilde{P}_i} = \frac{\gamma X_i^{\alpha_1} e^{\beta_1 Y_i + \tau_1 T_{i1} + \tau_2 T_{i2} + \cdots + \tau_n T_{in}}}{\gamma \tilde{X}_i^{\alpha_1} e^{\beta_1 \tilde{Y}_i + \tau_1 \tilde{T}_{i1} + \tau_2 \tilde{T}_{i2} + \cdots + \tau_n \tilde{T}_{in}}} = e^{\tau_1 (T_{i1} - \tilde{T}_{i1}) + \tau_2 (T_{i2} - \tilde{T}_{i2}) + \cdots + \tau_n (T_{in} - \tilde{T}_{in})} \tag{2}$$

or, in logarithmic transformation, $\ln \frac{P_i}{\tilde{P}_i} = \tau_1(T_{i1} - \tilde{T}_{i1}) + \tau_2(T_{i2} - \tilde{T}_{i2}) + \cdots + \tau_n(T_{in} - \tilde{T}_{in})$ where $X_i = \tilde{X}_i$ and $Y_i = \tilde{Y}_i$.

Equation (2) is the well-known repeat-sale analysis equation in which the dependent variable is the ratio of prices, the attributes and the implicit prices of the attributes are gone, and the time variables in parentheses take on the value -1 if the first sale occurs during that period, $+1$ if the second sale occurs during that period and 0 if no sale occurs during that period. Thus, the dummy variable is no longer dichotomous. The equation is estimated by taking the natural logarithm of both sides and using ordinary least squares regression. Although very commonly used for the purpose of price index estimation, by itself this equation will not generate any but the most indirect evidence of a price effect for an environmental spillover.

Changing Attributes

If an attribute changes between the sale times, then the attribute does not cancel out. In keeping with the logarithmic version of the formulation given above, for example, the model would be specified as

$$\begin{aligned} \ln \frac{P_i}{\tilde{P}_i} &= \ln \frac{\gamma X_i^{\alpha_1} e^{\beta_1 Y_i + \tau_1 T_{i1} + \tau_2 T_{i2} + \cdots + \tau_n T_{in}}}{\gamma \tilde{X}_i^{\alpha_1} e^{\beta_1 \tilde{Y}_i + \tau_1 \tilde{T}_{i1} + \tau_2 \tilde{T}_{i2} + \cdots + \tau_n \tilde{T}_{in}}} \\ &= \alpha_1 \ln \frac{X_i}{\tilde{X}_i} + \beta_1 (Y_i - \tilde{Y}_i) + \tau_1 (T_{i1} - \tilde{T}_{i1}) + \tau_2 (T_{i2} - \tilde{T}_{i2}) \\ &\quad + \cdots + \tau_n (T_{in} - \tilde{T}_{in}) \end{aligned} \tag{3}$$

where $X_i \neq \tilde{X}_i$ and $Y_i \neq \tilde{Y}_i$. Of course, the terms are dropped for the attributes that do not change between the sales. A logarithmic transformation yields the traditional repeat-sale analysis plus either the log of the ratio of the attributes (*i.e.*, the Xs) or the difference in the attributes (*i.e.*, the Ys), for just the attributes that change. Unlike the simple version of Equation (2), this equation can be used with before-and-after data to show whether a price change coincided temporally with an environmental event.

Varying Parameters

If the parameters on the attributes change through time, it is important to accommodate such changes in the analysis. Estimating a different hedonic regression for each period allows the implicit prices to change. Slightly overlapping the time periods links the price indices between periods. Alternatively, a simple device to allow parameters to change would be to assume that the parameters follow some particular function of time, such as a linear evolution or a jump at the time of some event. In the formulation

$$\ln\frac{P_i}{\tilde{P}_i} = \ln\frac{\gamma X_i^{\alpha_0+\alpha_1 t_i} e^{(\beta_0+\beta_1\theta_i)Y_i+\tau_1 T_{i1}+\tau_2 T_{i2}+\cdots+\tau_n T_{in}}}{\gamma \tilde{X}_i^{\alpha_0+\alpha_1\tilde{t}_i} e^{(\beta_0+\beta_1\tilde{\theta}_i)\tilde{Y}_i+\tau_1\tilde{T}_{i1}+\tau_2\tilde{T}_{i2}+\cdots+\tau_n\tilde{T}_{in}}}$$
$$= \alpha_1(t_i-\tilde{t})_i \ln X_i + \beta_1(\theta_i-\tilde{\theta}_i)Y_i + \tau_1(T_{i1}-\tilde{T}_{i1}) + \tau_2(T_{i2}-\tilde{T}_{i2}) + \cdots + \tau_n(T_{in}-\tilde{T}_{in}) \quad (4)$$

where $X_i = \tilde{X}_i$ and $Y_i = \tilde{Y}_i$, the elasticity of price with respect to the X attribute is modeled as a linear function of time, while the percentage change in price due to a unit change in the Y attribute is modeled as a discrete jump with θ being a dichotomous dummy variable taking on the value of one at time greater than some critical time. Of course, these devices can be combined with both the intercept and the slope of the function being allowed to change at a critical point in time.

Changing Attributes and Varying Parameters

Shiller (1993) recognized that a generalization of the standard repeat-sale formulation enabled the model to be estimated in the presence of both changing attributes and varying parameters. In terms of the formulation presented above, Shiller's hybrid model is

$$\ln\frac{P_i}{\tilde{P}_i} = \ln\frac{\gamma}{\tilde{\gamma}} + \alpha_1\ln X_i - \tilde{\alpha}_1\ln\tilde{X}_i + \beta_1 Y_i - \tilde{\beta}_1\tilde{Y}_i + \tau_1(T_{i1}-\tilde{T}_{i1}) + \tau_2(T_{i2}-\tilde{T}_{i2}) + \cdots + \tau_n(T_{in}-\tilde{T}_{in}) \quad (5)$$

where, as before, the tilde indicates the values (of property attributes or of model parameters) that were in effect at the time of the initial sale in each transaction-pair. This can be expressed equivalently as

$$\ln\frac{P_i}{\tilde{P}_i} = \ln\frac{\gamma}{\tilde{\gamma}} + \alpha_1\sum_{\Theta=1}^{n}\ln\widehat{X}_{i\Theta}^{1\Theta} + \beta_{11}\widehat{Y}_{i1} + \beta_{12}\widehat{Y}_{i2} + \cdots + \beta_{1n}\widehat{Y}_{in} + \tau_1\widehat{T}_{i1} + \tau_2\widehat{T}_{i2} + \cdots + \tau_n\widehat{T}_{in} \quad (6)$$

where $\widehat{X}_{i\Theta}$ takes on the value $-\tilde{X}_i$ if the first sale occurs during that period, $+X_i$ if the second sale occurs during that period and zero if no sale occurs during that period. $\widehat{Y}_{i\Theta}$ is defined analogously; so is $\widehat{T}_{i\Theta}$, which thereby takes on the familiar value of -1, $+1$ or 0.

Given appropriate data, this formulation can be used to investigate (1) the discount associated with an environmental event (a discrete change in attributes), (2) the evolution of house prices in response to mitigation of the environmental problem (a continuous change in attributes) and (3) the evolution over time in the environmental discount even in the absence of mitigation (a parameter variation).

Unmeasured Physical Changes

One curious point in the previous formulation is that the gammas do not necessarily cancel out, which leaves the repeat-sale model with the intercept $\ln(\frac{\gamma}{\tilde{\gamma}})$. Although the gross physical characteristics of a dwelling are not likely to change over time, some authors highlight the existence of a nontemporal component to house price appreciation (Goetzmann and Spiegel 1995, 1997). This implies that there might be changes associated with the physical attributes of the property through refurbishment or modifications associated with a purchase or sale. It is likely that households will undertake routine fix-ups, modifications, refurbishments or improvements prior to sale or just after purchase. The absence of detailed attribute data means that these changes are not generally captured by our parameter estimates. As Goetzmann and Spiegel (1995) explain, this can be remedied by the introduction of an intercept term in the repeat-sale model. The inclusion of an intercept captures the nontemporal components of house price appreciation and removes them from the estimated index values. Thus, the intercept can be interpreted as measuring the price change normally associated with these physical improvements.

Data and Modeling

We employ a data set that is particularly appropriate for assessing changes over time in the discount associated with the sudden occurrence of an environmental problem. Our data set was developed as part of a case known as *Baker vs. Motorola*, in which a group of homeowners sued a manufacturer for damages sustained as a result of the discovery of long-term environmental contamination. Until the 1970s, industrial solvents containing volatile organic compounds (VOCs) were typically disposed of directly onto the ground or in dry wells. In Scottsdale and Tempe, Arizona, these practices resulted in contamination of groundwater over an area of approximately 13 square miles by VOCs including trichloroethylene (TCE), tetrachloroethylene (PCE) and chloroform.

The groundwater contamination was discovered only in late 1981 in several drinking water wells. Local water providers immediately stopped using those wells for drinking water, and in 1983 the entire site was placed on the U.S. Environmental Protection Agency's National Priorities (Superfund) List. Remediation efforts have since been ongoing; specifically, the contaminated groundwater is pumped to the surface where exposure to air removes the volatile compounds, after which the groundwater is pumped back underground. It is anticipated that this treatment will be required for the next 40 years, although the effectiveness of the treatment is uncertain. The length of time that mitigation is expected to require, as well as the uncertainty of ultimate success, suggest that any decline in the discount associated with groundwater contamination is likely to signal a decline in the market's assessment of the value of clean groundwater, rather than a decline in the physical severity of the contamination.

The defense in the lawsuit acquired data on transactions of properties in areas both affected and unaffected by the contamination, covering time periods both before and after the contamination was discovered. The data were accepted by the plaintiff as being reasonably accurate and were entered as evidence in the case, making it public information according to the plaintiffs' attorneys.

The area covered by the data set encompasses a large number of condominium properties, which makes it particularly useful for our purposes because structural characteristics of condominium properties are quite unlikely to have changed over time. Importantly, there was no public information regarding possible contamination prior to the announcement very late in 1981 that contamination had occurred. (If there was any leakage of information prior to the public announcement—which we do not believe occurred—then that fact would affect our estimate of the magnitude of the initial discount associated with environmental contamination, but not our analysis of its persistence after the public announcement.) The study period encompassed by the data set extends from the beginning of 1980 through 1998, which means that we have essentially two years of property transactions that were unaffected by contamination followed by 17 years of transactions after the event. (Of course it would have been preferable for our purposes to have more transactions prior to the event, but the defense selected the time period and the plaintiff acquiesced, each for interests that did not coincide with ours.)

Data

The initial data set describing condominium sales in Scottsdale and Tempe comprised 30,199 transactions. A small number (333) were duplicates, and a few (38) were eliminated because they were incomplete (lacking data on price and/or year) or because we judged them not to be valid repeat transactions. (For

Table 1 ■ Summary of usable and unusable property transaction records.

No. of Times Transacted	Properties	Transactions	Transaction-Pairs
1 (Unusable)[a]	7,347	7,347	0
2 (Unusable)[b]	185	370	185
2 (Usable)	4,926	9,852	4,926
3 (Unusable)[b]	6	18	12
3 (Usable)	2,361	7,083	4,722
4	890	3,560	2,670
5	240	1,200	960
6	47	282	235
7	14	98	84
8	1	8	7
9	1	9	8
Unusable total	7,538	7,735	197
Usable total	8,480	22,092	13,612
Total	16,018	29,827	13,809

[a]Unusable in a repeat-sales modeling framework; would be usable in hedonic price model or Case and Quigley hybrid model.
[b]Unusable because all transactions occurred in the same time interval (calendar year).

example, several of them appeared to be transactions not of property but only of partial interests, such as those resulting from a divorce; others appeared to involve an initial transaction of raw land along with a subsequent transaction of the developed property.)

The remaining 29,827 transactions are summarized in Table 1. As the table shows, 7,347 transactions described properties which sold only once during the study period. Although Case and Quigley (1991) and others have shown how data on once-transacting properties could be incorporated into a hybrid repeat-sale/hedonic model, for this application we focus only on repeat-transacting properties and therefore removed the once-transacting observations from the analysis.

It is common practice to eliminate any pair of consecutive transactions that occurred during the available time interval (the calendar year for this data set) because the observations on the time dummies in a repeat-sales model would all equal zero. As the Appendix points out, however, this approach is at best inefficient and in practice may introduce bias. The use of consecutive (rather than nonconsecutive) transaction pairs is simply a matter of convenience and is not required. Thus, the alternative approach suggested in the Appendix is to replace pairs of consecutive transactions occurring in the same time interval with pairs of nonconsecutive transactions occurring in different time intervals to preserve the information contained in all independent price-relatives. A property

Table 2 ■ Descriptive statistics for properties included in the analysis.

Variable	Mean	Standard Deviation	Minimum	Maximum
Transaction year	1989.7	5.75 years	1980	1998
Age (second transaction)	12.52 years	8.17 years	1 year	40 years
Price (second transaction)	$64,428	$22,703	$10,000	$494,200
In contaminated area	44.2%	49.7%	0	1

that transacted only in one time interval, however, cannot be incorporated into a repeat-sales analysis no matter how many transactions were involved in that time interval; in our data set this meant that 191 properties that sold more than once but never in more than one year had to be deleted.

Table 2 presents basic descriptive statistics for the 13,612 transaction-pairs that are available for use in this analysis. As noted, the properties in our data set are all condominium units, which implies that changes in structural characteristics were rare if indeed they ever occurred—an important advantage when a repeat-sales modeling approach is employed. The average condominium was 12.5 years old at the time of its second transaction and sold for $64,428. Most importantly, 44% of the transaction-pairs involved properties that were located within the contaminated area.

Modeling

As a baseline, we first estimate the repeat-sale model commonly used to construct price indices, with an intercept:

$$\ln\left(\frac{P_i}{\tilde{P}_i}\right) = \ln\left(\frac{\gamma}{\tilde{\gamma}}\right) + \tau_1(T_{i1} - \tilde{T}_{i1}) + \tau_2(T_{i2} - \tilde{T}_{i2}) + \cdots + \tau_n(T_{in} - \tilde{T}_{in}) \quad (7)$$

Of course, we assume that there is an error term of the usual sort in this and all the other estimation models.

Our analysis of the price effects of environmental contamination begins with Model 2, where we add a single hedonic variable indicating whether (1) the property is located in the contaminated area and (2) the second sale occurred after the announcement while the first sale occurred before:

$$\ln\left(\frac{P_i}{\tilde{P}_i}\right) = \ln\left(\frac{\gamma}{\tilde{\gamma}}\right) + \tau_1(T_{i1} - \tilde{T}_{i1}) + \tau_2(T_{i2} - \tilde{T}_{i2})$$
$$+ \cdots + \tau_n(T_{in} - \tilde{T}_{in}) + \beta_0 \mathrm{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i}) \quad (8)$$

where

$$\text{CNTM}_i = \begin{cases} 1, & \text{if located in plaintiff class area} \\ 0, & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\phi_{82,i} = \begin{cases} 1, & \text{if second sale occurs in 82 or thereafter} \\ 0, & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\tilde{\phi}_{82,i} = \begin{cases} 1, & \text{if first sale occurs in 82 or thereafter} \\ 0, & \text{otherwise.} \end{cases} \quad \text{and}$$

$$\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i}) = \begin{cases} 1, & \text{contaminated area and sales before and after} \\ 0, & \text{otherwise.} \end{cases} \quad \text{and}$$

The announcement date was quite late in 1981, so we assume that 1982 was the first year in which the contamination was made public and the information was used in negotiating transaction prices. (Note that the difference variable $\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$ can never take on the value -1 because if the first sale occurred after 1981 then the second surely occurred after 1981.) The estimated coefficient on the contamination dummy variable indicates whether the contamination had any price effect.

Next we extend our analysis by estimating two models enabling us to consider whether the contamination discount grew or dissipated following the initial event. In Model 2a we restrict the evolution of the contamination discount to a simple linear time trend:

$$\ln\left(\frac{P_i}{\tilde{P}_i}\right) = \ln\left(\frac{\gamma}{\tilde{\gamma}}\right) + \tau_1(T_{i1} - \tilde{T}_{i1}) + \tau_2(T_{i2} - \tilde{T}_{i2}) + \cdots + \tau_n(T_{in} - \tilde{T}_{in}) + \beta_0\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i}) + \beta_1\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})(\Delta t_{82,i}) \quad (9)$$

where $\Delta t_{82,i} = \phi_{82,i}(1 - \tilde{\phi}_{82,i})(t_i - 82) + \tilde{\phi}_{82,i}(t_i - \tilde{t}_i) =$ minimum of the number of years between sales and the number of years from 1982 to the year of the second sale. In Model 2b we generalize further by estimating a separate price index for the effect of contamination, using the estimation procedure suggested by Shiller (1993), without restricting it to a straight-line time trend:

$$\ln\left(\frac{P_i}{\tilde{P}_i}\right) = \ln\left(\frac{\gamma}{\tilde{\gamma}}\right) + \tau_1(T_{i1} - \tilde{T}_{i1}) + \tau_2(T_{i2} - \tilde{T}_{i2}) + \cdots + \tau_n(T_{in} - \tilde{T}_{in}) + \beta_0\text{CNTM}_i(T_{82,i} - \tilde{T}_{82,i}) + \beta_1\text{CNTM}_i(T_{83,i} - \tilde{T}_{83,i}) + \cdots + \beta_n\text{CNTM}_i(T_{98,i} - \tilde{T}_{98,i}) \quad (10)$$

Next we introduce two additional location effects. The Salt River approximately divides Scottsdale and Tempe, and a gradient describes house price differences as a function of distance either north or south of the river. In Model 3, then, we allow for evolution through time in each of the location-price gradients:

$$\begin{aligned}\ln\left(\frac{P_i}{\tilde{P}_i}\right) &= \ln\left(\frac{\gamma}{\tilde{\gamma}}\right) + \tau_1(T_{i1} - \tilde{T}_{i1}) + \tau_2(T_{i2} - \tilde{T}_{i2}) + \cdots + \tau_n(T_{in} - \tilde{T}_{in}) \\ &\quad + \beta_0\, \mathrm{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i}) + \delta_2 \mathrm{NSR}_i \mathrm{DSR}_i(\Delta t_{81,i}) \\ &\quad + \delta_3(1 - \mathrm{NSR}_i)\mathrm{DSR}_i(\Delta t_{81,i}) \qquad (11)\end{aligned}$$

where

$$\Delta t_{81,i} = \phi_{81,i}(1 - \tilde{\phi}_{81,i})(t_i - 80) + \tilde{\phi}_{81,i}(t_i - \tilde{t}_i) = \text{number of years between sales},$$

$$\mathrm{NSR}_i = \begin{cases} 1, & \text{if located north of Salt River} \\ 0, & \text{otherwise,} \end{cases} \quad \text{and}$$

$$\mathrm{DSR}_i = \text{distance from Salt River in thousands of feet.}$$

Model 3a again incorporates the possibility of straight-line growth or dissipation over time in the contamination discount:

$$\begin{aligned}\ln\left(\frac{P_i}{\tilde{P}_i}\right) &= \ln\left(\frac{\gamma}{\tilde{\gamma}}\right) + \tau_1(T_{i1} - \tilde{T}_{i1}) + \tau_2(T_{i2} - \tilde{T}_{i2}) + \cdots + \tau_n(T_{in} - \tilde{T}_{in}) \\ &\quad + \beta_0 \mathrm{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i}) + \beta_1 \mathrm{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})(\Delta t_{82,i}) \\ &\quad + \delta_2 \mathrm{NSR}_i \mathrm{DSR}_i(\Delta t_{81,i}) + \delta_3(1 - \mathrm{NSR}_i)\, \mathrm{DSR}_i(\Delta t_{81,i}) \qquad (12)\end{aligned}$$

while Model 3b generalizes to estimate a separate price index for the effect of contamination without restricting it to a straight-line trend:

$$\begin{aligned}\ln\left(\frac{P_i}{\tilde{P}_i}\right) &= \ln\left(\frac{\gamma}{\tilde{\gamma}}\right) + \sum_{t=81}^{98} \tau_t(T_{t,i} - \tilde{T}_{t,i}) + \sum_{t=82}^{98} \beta_t \mathrm{CNTM}_i(T_{t,i} - \tilde{T}_{t,i}) \\ &\quad + \sum_{t=81}^{98} \delta_t \mathrm{NSR}_i \mathrm{DSR}_i(T_{t,i} - \tilde{T}_{t,i}) \\ &\quad + \sum_{t=81}^{98} \delta_t(1 - \mathrm{NSR}_i)\, \mathrm{DSR}_i(T_{t,i} - \tilde{T}_{t,i}) \qquad (13)\end{aligned}$$

Case and Shiller (1989) suggested that the drift of individual property values through time would likely be reflected in heteroskedastic disturbance terms in any repeat-sales model, and that this source of heteroskedasticity could be eliminated by regressing the squared residuals from a first-stage repeat-sales regression on the time interval between sales. We employ this correction (using a quadratic functional form) in all of our empirical estimates.

In addition to this now-standard correction of the basic repeat-sales model, Palmquist (1982) pointed out that the error covariance matrix of a repeat-sales model is not diagonal if any of the properties transacted more than twice during the study period. The exact structure of the covariance matrix is known, however, so it is straightforward to apply the estimator suggested by Aitken (1935), as we describe in the Appendix.[1]

Results

The regression results are very much what we anticipated. The results suggest that the shape of the value surface has undergone both discrete and continuous changes over time, indicating the value of including some hedonic characteristics in the empirical estimation. Most importantly, we discover a contamination discount that varied over the post-contamination period.

As a baseline we estimated the standard repeat-sales formulation commonly used in constructing price indices. (As in all the models, 1980 was the excluded dummy year.) The adjusted R^2 is 29.0%, and F tests indicate that we can reject the null hypotheses that all annual price changes are zero ($F_{18,13593} = 295$) and that all annual price changes are identical ($F_{17,13593} = 312$). The results are shown in Table 3. The most obvious feature of the results is that after a general rise in condo prices from 1980 through 1983, prices declined steadily through 1991, reaching a trough more than 20% below the 1980 base year ($1 - e^{-0.22520} = 20.16\%$) before rising steadily through 1998 until they stood

[1] Surprisingly, however, we are not aware of any published studies employing repeat-sales models that have employed the Aitken estimator. The Appendix considers the importance of this correction and finds that it may be empirically more important than the Case–Shiller heteroskedasticity correction. In both cases the uncorrected estimates are unbiased, but applying the correction reduces the standard error of the estimates, so the corrected estimates should be closer to the true parameters being estimated. In our application the heteroskedasticity correction reduced the magnitude of the estimated effect of contamination on property values by between 2.6% and 12.1% (Models 2, 2a, 3 and 3a below), while applying the Aitken estimator reduced the estimated effect additionally by between 8.5% and 24.2%. This suggests that failing to apply the heteroskedasticity correction would have led, in this application, to an overestimate of the contamination effect—but that failing to apply the Aitken estimator would have led to a larger mistake in the same direction.

Table 3 ■ Model 1: Traditional repeat sale.

Variable	Parameter Estimate	Standard Error	T Statistic
Intercept	0.03563	0.00379	9.39
$T_{81} - \tilde{T}_{81}$	0.03894	0.00847	4.60
$T_{82} - \tilde{T}_{82}$	0.03768	0.00931	4.05
$T_{83} - \tilde{T}_{83}$	0.06036	0.00894	6.75
$T_{84} - \tilde{T}_{84}$	0.05384	0.00876	6.14
$T_{85} - \tilde{T}_{85}$	0.02078	0.00888	2.34
$T_{86} - \tilde{T}_{86}$	−0.01737	0.00896	−1.94
$T_{87} - \tilde{T}_{87}$	−0.04725	0.00960	−4.92
$T_{88} - \tilde{T}_{88}$	−0.06547	0.00941	−6.96
$T_{89} - \tilde{T}_{89}$	−0.13229	0.00991	−13.36
$T_{90} - \tilde{T}_{90}$	−0.18318	0.00981	−18.66
$T_{91} - \tilde{T}_{91}$	−0.22520	0.00981	−22.95
$T_{92} - \tilde{T}_{92}$	−0.20376	0.01043	−19.54
$T_{93} - \tilde{T}_{93}$	−0.15098	0.00965	−15.65
$T_{94} - \tilde{T}_{94}$	−0.11122	0.00930	−11.96
$T_{95} - \tilde{T}_{95}$	−0.03799	0.00968	−3.93
$T_{96} - \tilde{T}_{96}$	0.02764	0.00979	2.82
$T_{97} - \tilde{T}_{97}$	0.09573	0.01005	9.52
$T_{98} - \tilde{T}_{98}$	0.15621	0.01011	15.45
	$N = 13{,}612$	$R^2 = 0.2907$	

almost 17% above their base level ($e^{0.15621} - 1 = 16.91\%$). Another clear result is that the intercept is positive and significant, indicating nontemporal improvements on the order of approximately 3.6%.

The next step is to introduce a constant contamination discount through the variable $\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$. As shown in Table 4, this change improves the fit of the model only slightly, increasing the adjusted R^2 to 29.1%. F tests again indicate that we can reject the hypotheses that all annual price changes are zero ($F_{18,13592} = 291$) and that all annual price changes are identical ($F_{17,13592} = 308$). The dip in condo prices is largely unaltered by the inclusion of the contamination variable. Perhaps the only notable change to the annual price indices is that prices in the 1998 were more than 19% above the 1980 prices as compared to slightly less than 17% in the baseline model.

The environmental story is that the coefficient on the contamination variable is negative and significant, indicating that condos located in the contaminated area suffered a discount of 4.65% relative to condos in noncontaminated areas. This is interesting from two perspectives. First, it strongly suggests that the effects of groundwater contamination were capitalized into lower property price: unless

Table 4 ■ Model 2: Repeat sale with contamination.

Variable	Parameter Estimate	Standard Error	T Statistic
Intercept	0.03556	0.00379	9.37
$T_{81} - \tilde{T}_{81}$	0.03944	0.00846	4.66
$T_{82} - \tilde{T}_{82}$	0.05623	0.00999	5.63
$T_{83} - \tilde{T}_{83}$	0.07821	0.00960	8.15
$T_{84} - \tilde{T}_{84}$	0.07220	0.00947	7.62
$T_{85} - \tilde{T}_{85}$	0.03939	0.00960	4.10
$T_{86} - \tilde{T}_{86}$	0.00099	0.00966	0.10
$T_{87} - \tilde{T}_{87}$	−0.02887	0.01025	−2.82
$T_{88} - \tilde{T}_{88}$	−0.04681	0.01009	−4.64
$T_{89} - \tilde{T}_{89}$	−0.11418	0.01052	−10.86
$T_{90} - \tilde{T}_{90}$	−0.16484	0.01045	−15.78
$T_{91} - \tilde{T}_{91}$	−0.20696	0.01044	−19.82
$T_{92} - \tilde{T}_{92}$	−0.18536	0.01103	−16.80
$T_{93} - \tilde{T}_{93}$	−0.13232	0.01032	−12.83
$T_{94} - \tilde{T}_{94}$	−0.09275	0.00997	−9.30
$T_{95} - \tilde{T}_{95}$	−0.01952	0.01033	−1.89
$T_{96} - \tilde{T}_{96}$	0.04589	0.01042	4.41
$T_{97} - \tilde{T}_{97}$	0.11417	0.01068	10.69
$T_{98} - \tilde{T}_{98}$	0.17480	0.01074	16.28
$\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$	−0.04761	0.00936	−5.09
	$N = 13{,}612$	$R^2 = 0.2920$	

there is an excluded variable that is correlated with the contamination variable and that offers an alternative explanation, it is reasonable to conclude that groundwater contamination led to price discounts. Second, the results provide our first glimpse at the usefulness of including hedonic characteristics in a repeat-sales model.

Although Model 2 suggests the existence of a contamination discount, we are more interested in whether the contamination discount grows or dissipates through time, and to this end Table 5 shows the results of Model 2a permitting the contamination effect to follow a linear trend. The contamination discount for the year 1982 is the negative of the coefficient on the $\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$ variable, that is, the immediate effect of the contamination announcement was to lower prices in the area of contamination by approximately 2.43% (slightly less than in Model 2). The negative of the coefficient on the $\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})(\Delta t_{82,i})$ variable is the constant annual increase in the contamination discount. If the contamination discount dissipated, the coefficient on this variable would be positive; in contrast, we estimate it to be negative and significantly different from zero at almost the 95% level of confidence. The results provide no support

Table 5 ■ Model 2a: Repeat sale with straight-line contamination dissipation.

Variable	Parameter Estimate	Standard Error	T Statistic
Intercept	0.03333	0.00397	8.39
$T_{81} - \tilde{T}_{81}$	0.04055	0.00848	4.78
$T_{82} - \tilde{T}_{82}$	0.05542	0.01000	5.54
$T_{83} - \tilde{T}_{83}$	0.07847	0.00960	8.17
$T_{84} - \tilde{T}_{84}$	0.07254	0.00947	7.66
$T_{85} - \tilde{T}_{85}$	0.04005	0.00960	4.17
$T_{86} - \tilde{T}_{86}$	0.00216	0.00968	0.22
$T_{87} - \tilde{T}_{87}$	−0.02724	0.01029	−2.65
$T_{88} - \tilde{T}_{88}$	−0.04494	0.01014	−4.43
$T_{89} - \tilde{T}_{89}$	−0.11185	0.01059	−10.56
$T_{90} - \tilde{T}_{90}$	−0.16227	0.01053	−15.40
$T_{91} - \tilde{T}_{91}$	−0.20405	0.01055	−19.34
$T_{92} - \tilde{T}_{92}$	−0.18219	0.01116	−16.33
$T_{93} - \tilde{T}_{93}$	−0.12876	0.01048	−12.28
$T_{94} - \tilde{T}_{94}$	−0.08887	0.01018	−8.73
$T_{95} - \tilde{T}_{95}$	−0.01523	0.01057	−1.44
$T_{96} - \tilde{T}_{96}$	0.05049	0.01070	4.72
$T_{97} - \tilde{T}_{97}$	0.11908	0.01099	10.84
$T_{98} - \tilde{T}_{98}$	0.18022	0.01111	16.22
$\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$	−0.02460	0.01531	−1.61
$\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})(\Delta t_{82,i})$	−0.00227	0.00119	−1.90
$N = 13{,}612$		$R^2 = 0.2922$	

for the hypothesis that the contamination discount dissipated through time—at least along a steady trend—and only weak evidence that it grew.

While Model 2a restricts the contamination discount to follow a linear time trend, in Model 2b we use the formulation suggested by Shiller (1993) to generalize this time-varying effect by estimating a separate price index for the effect of contamination. The empirical estimates are shown in Table 6. F tests indicate that we can reject the hypotheses that all annual price changes are zero and that all annual price changes are identical for both the overall price index ($F_{18,13577} = 148$ and $F_{17,13577} = 157$) and the contamination price index ($F_{16,13577} = 12.9$ and $F_{15,13577} = 11.8$).

The pattern of the contamination coefficients, however, tells a very different story than was suggested by the more restrictive models: they suggest that the effect of contamination was not capitalized into the prices of properties in the contaminated area until about 1986, well after the contamination became public knowledge. During 1986–1996, however, the prices of properties in the contaminated area dropped below those in unaffected areas by up to 13.55%

Table 6 ■ Model 2b: Repeat sale with price index on contamination effect.

Variable	Parameter Estimate	Standard Error	T Statistic
Intercept	0.03582	0.00377	9.49
$T_{81} - \tilde{T}_{81}$	0.04053	0.00842	4.81
$T_{82} - \tilde{T}_{82}$	0.03938	0.00925	4.26
$T_{83} - \tilde{T}_{83}$	0.06787	0.01044	6.50
$T_{84} - \tilde{T}_{84}$	0.05070	0.01032	4.91
$T_{85} - \tilde{T}_{85}$	0.02944	0.01045	2.82
$T_{86} - \tilde{T}_{86}$	−0.00371	0.01044	−0.36
$T_{87} - \tilde{T}_{87}$	−0.02923	0.01138	−2.57
$T_{88} - \tilde{T}_{88}$	−0.06442	0.01116	−5.77
$T_{89} - \tilde{T}_{89}$	−0.10582	0.01125	−9.40
$T_{90} - \tilde{T}_{90}$	−0.14400	0.01146	−12.57
$T_{91} - \tilde{T}_{91}$	−0.17026	0.01147	−14.85
$T_{92} - \tilde{T}_{92}$	−0.16057	0.01219	−13.17
$T_{93} - \tilde{T}_{93}$	−0.11893	0.01110	−10.71
$T_{94} - \tilde{T}_{94}$	−0.08624	0.01065	−8.10
$T_{95} - \tilde{T}_{95}$	−0.01721	0.01088	−1.58
$T_{96} - \tilde{T}_{96}$	0.04140	0.01091	3.79
$T_{97} - \tilde{T}_{97}$	0.10165	0.01123	9.05
$T_{98} - \tilde{T}_{98}$	0.15778	0.01127	13.99
$\text{CNTM}_i(T_{82,i} - \tilde{T}_{82,i})$	−0.01539	0.01590	−0.97
$\text{CNTM}_i(T_{83,i} - \tilde{T}_{83,i})$	–	–	–
$\text{CNTM}_i(T_{84,i} - \tilde{T}_{84,i})$	0.01432	0.01524	0.94
$\text{CNTM}_i(T_{85,i} - \tilde{T}_{85,i})$	−0.01938	0.01519	−1.28
$\text{CNTM}_i(T_{86,i} - \tilde{T}_{86,i})$	−0.03616	0.01509	−2.40
$\text{CNTM}_i(T_{87,i} - \tilde{T}_{87,i})$	−0.05127	0.01618	−3.17
$\text{CNTM}_i(T_{88,i} - \tilde{T}_{88,i})$	−0.00435	0.01572	−0.28
$\text{CNTM}_i(T_{89,i} - \tilde{T}_{89,i})$	−0.07532	0.01720	−4.38
$\text{CNTM}_i(T_{90,i} - \tilde{T}_{90,i})$	−0.10580	0.01640	−6.45
$\text{CNTM}_i(T_{91,i} - \tilde{T}_{91,i})$	−0.14559	0.01588	−9.17
$\text{CNTM}_i(T_{92,i} - \tilde{T}_{92,i})$	−0.11745	0.01720	−6.83
$\text{CNTM}_i(T_{93,i} - \tilde{T}_{93,i})$	−0.08661	0.01490	−5.81
$\text{CNTM}_i(T_{94,i} - \tilde{T}_{94,i})$	−0.06472	0.01352	−4.79
$\text{CNTM}_i(T_{95,i} - \tilde{T}_{95,i})$	−0.05541	0.01393	−3.98
$\text{CNTM}_i(T_{96,i} - \tilde{T}_{96,i})$	−0.03555	0.01377	−2.58
$\text{CNTM}_i(T_{97,i} - \tilde{T}_{97,i})$	−0.01373	0.01385	−0.99
$\text{CNTM}_i(T_{98,i} - \tilde{T}_{98,i})$	0.00152	0.01363	0.11
$N = 13{,}612$		$R^2 = 0.3014$	

(in 1991) before recovering in the last two years of the study period. It is interesting to speculate why the market response to the contamination event should apparently have been delayed by several years. One possibility is that information about the contamination may have disseminated relatively slowly to buyers—or perhaps even to sellers, although the news coverage of this event makes it difficult to credit that explanation.

Another possibility is that the relatively large transaction costs associated with moving made it difficult for owners to respond quickly. If transaction costs are large, then the earliest transactors following the announcement of the contamination may not have been those owners who were most sensitive to environmental problems, but rather those owners who were already close to overcoming the transaction costs before the announcement was made; the maximum discount, then, would occur only at a later time when the most environmentally sensitive owners overcame their transaction costs to become environmentally sensitive sellers.

It is also interesting that the most severe discount (in percentage terms) on properties in contaminated areas occurred in 1991, coinciding with the trough of the overall condominium market in the Scottsdale/Tempe area. As our data set covers only one market cycle we cannot use it to assess whether environmental contamination generally depresses the value of affected properties most severely during market downturns, but it is a possibility worth exploring in other contexts.

The third substantial step in our estimation experiment is to introduce varying location-price gradients as well as the contamination discount. The results are reported in Table 7. The adjusted R^2 is 30.2%, only slightly higher than the traditional repeat-sale and the simpler hybrid model, and F tests again indicate that we can reject the hypotheses that all annual price changes are zero ($F_{18,13590} = 144$) and that all annual price changes are identical ($F_{17,13590} = 151$). Once again we observe the profound dip in condo prices. The contamination discount is estimated at 4.88%, slightly higher than in Model 2 and significant at a high level of confidence.

The results for the time trend in both of the location-price gradients are positive and significant, suggesting that relative prices are rising in the far north and the far south of the area from which the data are drawn, with the gradient for locations north of the Salt River growing more rapidly than the one for locations south of the Salt River.

The value of including hedonic variables is even more dramatic in Model 3 than in Model 2: all three hedonic variables are significant. As suggested by Case and Quigley (1991), we should expect that the coefficients on location variables are likely not to be stationary, so it may generally be important to include hedonic characteristics with a functional form enabling implicit prices as well as property attributes to vary.

Table 7 ■ Model 3: Repeat sale with gradients and contamination.

Variable	Parameter Estimate	Standard Error	T Statistic
Intercept	0.03531	0.00377	9.36
$T_{81} - \tilde{T}_{81}$	0.04182	0.00839	4.98
$T_{82} - \tilde{T}_{82}$	0.05831	0.00991	5.88
$T_{83} - \tilde{T}_{83}$	0.07927	0.00952	8.33
$T_{84} - \tilde{T}_{84}$	0.07284	0.00939	7.76
$T_{85} - \tilde{T}_{85}$	0.03968	0.00952	4.17
$T_{86} - \tilde{T}_{86}$	0.00280	0.00958	0.29
$T_{87} - \tilde{T}_{87}$	−0.02728	0.01017	−2.68
$T_{88} - \tilde{T}_{88}$	−0.04659	0.01001	−4.65
$T_{89} - \tilde{T}_{89}$	−0.11481	0.01043	−11.01
$T_{90} - \tilde{T}_{90}$	−0.16352	0.01036	−15.78
$T_{91} - \tilde{T}_{91}$	−0.20520	0.01036	−19.81
$T_{92} - \tilde{T}_{92}$	−0.18142	0.01095	−16.57
$T_{93} - \tilde{T}_{93}$	−0.12913	0.01023	−12.62
$T_{94} - \tilde{T}_{94}$	−0.11368	0.01005	−11.31
$T_{95} - \tilde{T}_{95}$	−0.06385	0.01087	−5.87
$T_{96} - \tilde{T}_{96}$	−0.01946	0.01164	−1.67
$T_{97} - \tilde{T}_{97}$	0.02505	0.01282	1.95
$T_{98} - \tilde{T}_{98}$	0.06567	0.01395	4.71
$\mathrm{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$	−0.04998	0.00931	−5.37
$\mathrm{NSR}_i \mathrm{DSR}_i(\Delta t_{81,i})$	1.27×10^{-6}	9.02×10^{-8}	14.08
$(1 - \mathrm{NSR}_i)\mathrm{DSR}_i(\Delta t_{81,i})$	7.57×10^{-7}	1.09×10^{-7}	6.97
	$N = 13{,}612$	$R^2 = 0.3027$	

Model 3a incorporates a straight-line trend into the hybrid model to investigate whether the effect of contamination grows or diminishes over time. F tests again indicate that we can reject the hypotheses that all annual price changes are zero ($F_{18,13589} = 136$) and that all annual price changes are identical ($F_{17,13589} = 143$). Table 8 shows that the coefficient on the contamination term is somewhat smaller in Model 3a (2.05%) than in Model 2a (2.43%) and not significantly different from zero, but the coefficient on the contamination/time interaction term is also negative and significant at the 98% level of confidence. This can be interpreted as evidence that there was very little contamination discount immediately after the contamination became public knowledge, but that the discount grew over time and amounted to about 6.5% by 1998.

Finally, in Model 3b we estimate a price index for the effect of contamination in the manner suggested by Shiller (1993), relaxing the constraint that the effect of contamination grow or diminish at a constant rate.[2] The results are shown

[2] We also estimate full price indices for the two hedonic characteristics, distance north and south of the Salt River; coefficient estimates for these variables are not shown but are available on request.

Table 8 ■ Model 3a: Repeat sale with gradients and straight-line contamination dissipation.

Variable	Parameter Estimate	Standard Error	T Statistic
Intercept	0.03247	0.00395	8.22
$T_{81} - \tilde{T}_{81}$	0.04324	0.00841	5.14
$T_{82} - \tilde{T}_{82}$	0.05736	0.00992	5.78
$T_{83} - \tilde{T}_{83}$	0.07972	0.00952	8.37
$T_{84} - \tilde{T}_{84}$	0.07334	0.00939	7.81
$T_{85} - \tilde{T}_{85}$	0.04061	0.00953	4.26
$T_{86} - \tilde{T}_{86}$	0.00434	0.00960	0.45
$T_{87} - \tilde{T}_{87}$	−0.02513	0.01021	−2.46
$T_{88} - \tilde{T}_{88}$	−0.04413	0.01006	−4.39
$T_{89} - \tilde{T}_{89}$	−0.11184	0.01050	−10.65
$T_{90} - \tilde{T}_{90}$	−0.16018	0.01045	−15.32
$T_{91} - \tilde{T}_{91}$	−0.20142	0.01047	−19.23
$T_{92} - \tilde{T}_{92}$	−0.17735	0.01107	−16.02
$T_{93} - \tilde{T}_{93}$	−0.12453	0.01041	−11.97
$T_{94} - \tilde{T}_{94}$	−0.10848	0.01028	−10.55
$T_{95} - \tilde{T}_{95}$	−0.05796	0.01114	−5.20
$T_{96} - \tilde{T}_{96}$	−0.01298	0.01195	−1.09
$T_{97} - \tilde{T}_{97}$	0.03210	0.01315	2.44
$T_{98} - \tilde{T}_{98}$	0.07362	0.01433	5.14
$\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$	−0.02073	0.01523	−1.36
$\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})(\Delta t_{82,i})$	−0.00290	0.00120	−2.43
$\text{NSR}_i\text{DSR}_i(\Delta t_{81,i})$	1.27×10^{-6}	9.02^{-8}	14.08
$(1 - \text{NSR}_i)\text{DSR}_i(\Delta t_{81,i})$	7.30×10^{-7}	1.09^{-7}	6.69
$N = 13{,}612$		$R^2 = 0.3030$	

in Table 9. F tests indicate that we can reject the hypotheses that all annual price changes are zero and that all annual price changes are identical for all four price indices: the overall price index ($F_{18,13541} = 48.8$ and $F_{17,13541} = 49.8$), the contamination price index ($F_{16,13541} = 4.63$ and $F_{15,13541} = 4.94$), the gradient for distance north of the Salt River ($F_{18,13541} = 22.9$ and $F_{17,13541} = 21.2$), and the gradient for distance south of the Salt River ($F_{18,13541} = 18.8$ and $F_{17,13541} = 15.9$).

As we saw in Model 2b, the results for Model 3b suggest that the effect of contamination was not capitalized into property values immediately after the information became public, but rather several years later in 1990. The contamination discount peaked in 1991 at 6.66%, and by 1995 it had disappeared. (Indeed, the results suggest that in 1998 properties located in the contaminated area were actually selling at a premium relative to properties located in otherwise identical but unaffected areas.)

Table 9 ■ Model 3b: Repeat sale with price index on gradients and contamination effect.

Variable	Parameter Estimate	Standard Error	T Statistic
Intercept	0.03392	0.00374	9.07
$T_{81} - \tilde{T}_{81}$	0.03439	0.02306	1.49
$T_{82} - \tilde{T}_{82}$	0.01566	0.02470	0.63
$T_{83} - \tilde{T}_{83}$	0.08554	0.02412	3.55
$T_{84} - \tilde{T}_{84}$	−0.02727	0.02320	−1.18
$T_{85} - \tilde{T}_{85}$	−0.08407	0.02394	−3.51
$T_{86} - \tilde{T}_{86}$	−0.11780	0.02415	−4.88
$T_{87} - \tilde{T}_{87}$	−0.13230	0.02573	−5.14
$T_{88} - \tilde{T}_{88}$	−0.20671	0.02480	−8.34
$T_{89} - \tilde{T}_{89}$	−0.21900	0.02448	−8.95
$T_{90} - \tilde{T}_{90}$	−0.28985	0.02459	−11.79
$T_{91} - \tilde{T}_{91}$	−0.34730	0.02500	−13.89
$T_{92} - \tilde{T}_{92}$	−0.32525	0.02822	−11.53
$T_{93} - \tilde{T}_{93}$	−0.29233	0.02412	−12.12
$T_{94} - \tilde{T}_{94}$	−0.28218	0.02233	−12.64
$T_{95} - \tilde{T}_{95}$	−0.24164	0.02228	−10.84
$T_{96} - \tilde{T}_{96}$	−0.16528	0.02204	−7.50
$T_{97} - \tilde{T}_{97}$	−0.09004	0.02194	−4.10
$T_{98} - \tilde{T}_{98}$	−0.04545	0.02216	−2.05
$\mathrm{CNTM}_i(T_{82,i} - \tilde{T}_{82,i})$	0.00005	0.01790	0.00
$\mathrm{CNTM}_i(T_{83,i} - \tilde{T}_{83,i})$	–	–	–
$\mathrm{CNTM}_i(T_{84,i} - \tilde{T}_{84,i})$	0.03666	0.01733	2.12
$\mathrm{CNTM}_i(T_{85,i} - \tilde{T}_{85,i})$	0.01113	0.01738	0.64
$\mathrm{CNTM}_i(T_{86,i} - \tilde{T}_{86,i})$	0.00181	0.01744	0.10
$\mathrm{CNTM}_i(T_{87,i} - \tilde{T}_{87,i})$	−0.01346	0.01866	−0.72
$\mathrm{CNTM}_i(T_{88,i} - \tilde{T}_{88,i})$	0.05328	0.01804	2.95
$\mathrm{CNTM}_i(T_{89,i} - \tilde{T}_{89,i})$	−0.02492	0.01932	−1.29
$\mathrm{CNTM}_i(T_{90,i} - \tilde{T}_{90,i})$	−0.03137	0.01840	−1.70
$\mathrm{CNTM}_i(T_{91,i} - \tilde{T}_{91,i})$	−0.06890	0.01786	−3.86
$\mathrm{CNTM}_i(T_{92,i} - \tilde{T}_{92,i})$	−0.04191	0.01965	−2.13
$\mathrm{CNTM}_i(T_{93,i} - \tilde{T}_{93,i})$	−0.03001	0.01688	−1.78
$\mathrm{CNTM}_i(T_{94,i} - \tilde{T}_{94,i})$	−0.01305	0.01538	−0.85
$\mathrm{CNTM}_i(T_{95,i} - \tilde{T}_{95,i})$	−0.00077	0.01565	−0.05
$\mathrm{CNTM}_i(T_{96,i} - \tilde{T}_{96,i})$	0.00838	0.01553	0.54
$\mathrm{CNTM}_i(T_{97,i} - \tilde{T}_{97,i})$	0.02279	0.01555	1.47
$\mathrm{CNTM}_i(T_{98,i} - \tilde{T}_{98,i})$	0.04646	0.01551	3.00
	$N = 13{,}612$	$R^2 = 0.3282$	

Figure 1 shows the price indices estimated for contaminated and unaffected areas using versions of each of the three models, while Figure 2 shows the estimated difference in prices between contaminated and unaffected areas. The general pattern of the results is quite striking: the results suggest that properties

Figure 1 ■ Estimated price indices in unaffected and contaminated areas.

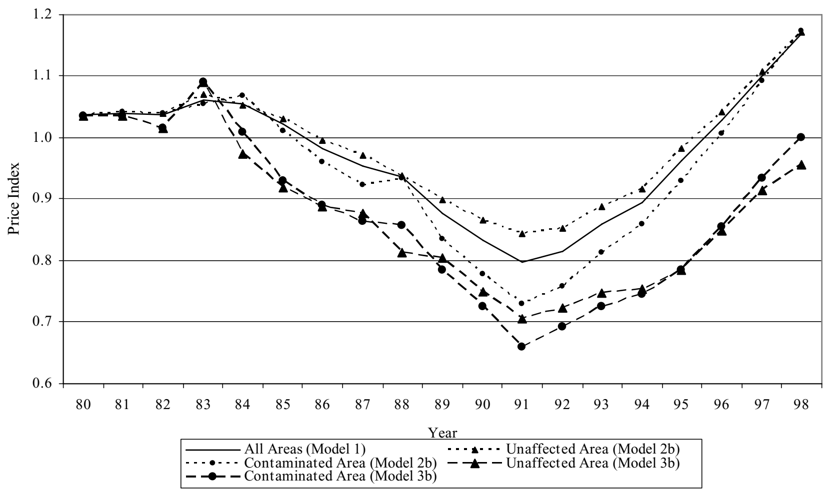
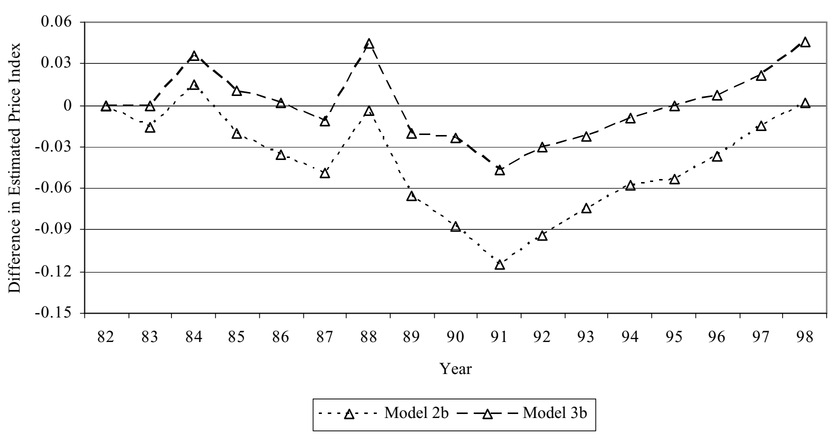


Figure 2 ■ Estimated difference in prices between contaminated and unaffected areas.



in the contaminated area suffered more severely than did properties in unaffected areas during the trough of the market, from 1989 to 1994. According to Model 3b, for example, there was essentially no difference in prices between contaminated and unaffected areas during 1982 and 1983, the first years after the contamination became public knowledge. Aside from an odd uptick estimated for 1984 in contaminated areas relative to unaffected areas, over the next eight years the prices of properties in the unaffected areas declined by slightly less than 30% while those in the contaminated area sank by more than 34%, meaning that properties in the contaminated area lost about 16% more of their value than did properties in unaffected areas.

Both Models 2b and 3b suggest that the effect of contamination was not capitalized into market values immediately after the contamination became public knowledge; rather, a contamination discount does not appear in transaction price data until 1985 (Model 2b) or even 1987 (Model 3b). Moreover, both models suggest that the contamination discount had been eliminated by 1998 (Model 2b) if not earlier (Model 3b). We believe this suggests the value of using a flexible specification of the type proposed by Shiller (1993) rather than a restrictive specification of the types embodied in Models 2, 2a, 3 and 3a. It also raises the question why there should be such a delay in realizing the price discount associated with such a widely recognized disamenity as groundwater contamination. We speculate that the answer to that question lies in the protracted spread of information to buyers (if not sellers as well), the large transaction costs associated with moving residence, or both.

Conclusions

Every empirical model that we estimated suggested that condominiums in the contaminated area transacted at a discount that was both statistically and economically significant during at least part of the study period. When the contamination discount was restricted to remain constant over the study period (Model 2) it was estimated at 4.65%. When the contamination discount was allowed to vary over time it grew steadily from 2.43% to 5.91% (Model 2a) or varied as high as 13.55% in 1991 (Model 2b). When the model was expanded to take into account changes in the price gradient associated with locations further north or south of the Salt River (Models 3, 3a and 3b), the contamination discount declined slightly in magnitude but remained significantly different from zero.

Perhaps the most interesting finding is that the effect of groundwater contamination does not appear to have been capitalized into condominium transaction prices until several years after the contamination became public knowledge. It is also interesting that the contamination discount was estimated to be growing rather than dissipating through the study period, according to the restrictive straight-line specification of Models 2a and 3a. When this was generalized to permit the estimation of a separate price index for the contamination discount (Models 2b and 3b), however, the results suggested that the contamination discount operated primarily during the trough of the property market cycle (1990–1993) and had disappeared by the end of the study period.

The spatial gradients appear to have been evolving during the study period. Ordinarily, a gradient is the proportionate rate of decline in price as distance increases away from the downtown. In this article, the gradients are the proportionate rate of increase in price as distance from the Salt River increases. The Salt River cuts from east to west across the approximate middle of the area from which data are drawn. The price gradient north of the Salt River grows by an annual amount that is slightly more than the growth of the gradient south of the Salt River.

The methodological results are nearly as interesting as the environmental or economic results. We have demonstrated the value of incorporating hedonic characteristics such as location into repeat-sale models. It may be generally important to do this because the parameters on location variables can change profoundly over time. We found that including gradient variables caused the path of the price index to change dramatically. The distinctions allowed by this approach reveal relative changes in condo prices within the region.

References

Aitken, A.C. 1935. On Least Squares and Linear Combinations of Observations. *Proceedings of the Royal Society of Edinburgh* 55: 42–48.

Bailey, M.J., R.F. Muth and H.O. Nourse. 1963. A Regression Method for Real Estate Price Index Construction. *Journal of the American Statistical Association* 58: 933–942.

Boyle, M.A. and K.A. Kiel. 2001. A Survey of Hedonic Studies of the Impact of Environmental Externalities. *Journal of Real Estate Literature* 9: 117–144.

Cannaday, R.E., H.J. Munneke and T.T. Yang. 2002. A Repeat Sales Index for Durable Assets: Disentangling the Price Effects of Age and Time. Working Paper. University of Georgia: Athens, GA.

Case, B., H.O. Pollakowski and S.M. Wachter. 1991. On Choosing Among House Price Index Methodologies. *Journal of the American Real Estate and Urban Economics Association* 19: 286–307.

Case, B. and J.M. Quigley. 1991. The Dynamics of Real Estate Prices. *Review of Economics and Statistics* 83: 50–58.

Case, K. and R. Shiller. 1989. The Efficiency of the Market for Single Family Homes. *American Economic Review* 79: 45–56.

Chalmers, J.A. and T.O. Jackson. 1996. Risk Factors in the Appraisal of Contaminated Property. *The Appraisal Journal* 64: 44–59.

Clapp, J.M. and C. Giacotto. 1992. Estimating Price Indices for Residential Property: A Comparison of Repeat Sales and Assessed Value Methods. *Journal of the American Statistical Association* 87: 300–306.

——. 1999. Revisions in Repeat-Sales Price Indexes: Here Today, Gone Tomorrow? *Real Estate Economics* 27: 79–104.

Clapp, J.M., C. Giacotto and D. Tirtiroglu. 1991. Housing Price Indices Based on All Transactions Compared to Repeat Subsamples. *Journal of the American Real Estate and Urban Economics Association* 19: 270–285.

Colwell, P.F. 1990. Power Lines and Land Values. *Journal of Real Estate Research* 5: 117–127.

Colwell, P.F., C. Dehring and N.A. Lash. 2000. The Effect of Group Homes on Neighborhood Property Values. *Land Economics* 76: 615–636.

Costello, G. 2002. Pricing Size Effects in Housing Markets. *Journal of Property Research* 17: 203–220.

Crone, T.M. and R.P. Voith. 1992. Estimating House Price Appreciation: A Comparison of Methods. *Journal of Housing Economics* 2: 324–338.

Dale, L., J.C. Murdoch, M.A. Thayer and P.A. Waddell. 1999. Do Property Values Rebound from Environmental Stigmas? Evidence from Dallas. *Land Economics* 75: 311–315.

Englund, P., J.M. Quigley and C.L. Redfearn. 1999. The Choice of Methodology for Computing Housing Price Indexes: Comparisons of Temporal Aggregation and Sample Definition. *The Journal of Real Estate Finance and Economics* 19: 91–112.

Gao, A.H. and G.H.K. Wang. 2004. Multi Transactions Model for Constructing Housing Market Index. Paper from the 2005 Fannie Mae Association Annual Meeting. Available at http://www.fma.org/Chicago/Papers/q041230.000MultiTransactionsModel-All.Zpdf.

Gatzlaff, D.H. and D.C. Ling. 1994. Measuring Changes in Local House Prices: An Empirical Investigation of Alternative Methodologies. *Journal of Urban Economics* 35: 221–244.

Goetzmann, W.N. and M. Spiegel. 1995. Non-Temporal Components of Residential Real Estate Appreciation. *Review of Economics and Statistics* 77: 199–206.

——. 1997. A Spatial Model of Housing Returns and Neighbourhood Substitutability. *The Journal of Real Estate Finance and Economics* 14: 11–31.

Graves, P., J.C. Murdoch, M.A. Thayer and D. Waldman. 1988. The Robustness of Hedonic Price Estimation: Urban Air Quality. *Land Economics* 64: 220–233.

Hill, R.C., J.R. Knight and C.F. Sirmans. 1997. Estimating Capital Asset Prices. *Review of Economics and Statistics* 79: 226–233.

Hite, D., W. Chern, F. Hitzhusen and A. Randall. 2001. Property Value Impacts of an Environmental Disamenity: The Case of Landfills. *The Journal of Real Estate Finance and Economics* 22: 185–202.

Jackson, T.O. 1997. Investing in Contaminated Real Estate. *Real Estate Review* 26: 38–43.

Jackson, T.O. 2001. The Effects of Environmental Contamination on Real Estate: A Literature Review. *Journal of Real Estate Literature* 9: 93–116.

Kiel, K.A. 1995. Measuring the Impact of the Discovery and Cleaning of Hazardous Waste Sites on House Values. *Land Economics* 71: 428–435.

Kiel, K.A. and K.T. McClain. 1996. House Price Recovery and Stigma After a Failed Siting. *Applied Economics* 28: 1351–1358.

Kinnard, W.N. and E.M. Worzala. 1999. How North American Appraisers Value Contaminated Property and Associated Stigma. *Appraisal Journal* 67: 269–279.

Kohlhase, J.E. 1991. The Impact of Toxic Waste Sites on Housing Values. *Journal of Urban Economics* 30: 1–26.

Leggett, C.G. and N.E. Boekstael. 2000. Evidence of the Effects of Water Quality on Residential Land Prices. *Journal of Environmental Economics and Management* 39: 121–144.

Mark, J.H. and M.A. Goldberg. 1984. Alternative House Price Indices: An Evaluation. *Journal of the American Real Estate and Urban Economics Association* 12: 30–49.

McClelland, G.H., D.H. Shulze and B. Hurd. 1990. The Effects of Risk Beliefs on Property Values: A Case Study of a Hazardous Waste Site. *Risk Analysis* 10: 485–497.

Meese, R.A. and N.E. Wallace. 1997. The Construction of Residential Housing Price Indices: A Comparison of Repeat-Sales, Hedonic Regression and Hybrid Approaches. *The Journal of Real Estate Finance and Economics* 14: 51–73.

Mieszkowski, P. and A.M. Saper. 1978. An Estimate of the Effects of Airport Noise on Property Values. *Journal of Urban Economics* 5: 425–440.

Palmquist, R.B. 1982. Measuring Environmental Effects on Property Values Without Hedonic Regressions. *Journal of Urban Economics* 11: 333–347.

Pennington, G., N. Topham and R. Ward. 1990. Aircraft Noise and Residential Property Values Adjacent to Manchester Airport. *Journal of Transport Economics and Policy* 24: 49–59.

Quigley, J.M. 1995. A Simple Hybrid Model for Estimating Real-Estate Price Indexes. *Journal of Housing Economics* 4: 1–12.

Ridker, R.G. and J.A. Henning. 1968. The Determination of Residential Property Value with Special Reference to Air Pollution. *Review of Economics and Statistics* 49: 246–257.

Riechert, A. 1999. The Persistence of Contamination Effects: A Superfund Site Revisited. *Appraisal Journal* 67: 126–135.

Segerson, K. 2001. Real Estate and the Environment: An Introduction. *The Journal of Real Estate Finance and Economics* 22: 135–139.

Shiller, R.J. 1993. Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures. *The Journal of Finance* 48: 911–931.

Steele, M. and R. Goy. 1997. Short Holds, the Distribution of First and Second Sales and Bias in the Repeat-Sales Price Index. *The Journal of Real Estate Finance and Economics* 14: 133–154.

Thayer, M., H. Albers and M. Rahmatian. 1992. The Benefits of Reducing Exposure to Waste Disposal Sites: A Hedonic Housing Value Approach. *Journal of Real Estate Research* 7: 265–282.

Weber, B.R. 1997. The Valuation of Contaminated Land. *Journal of Real Estate Research* 14: 379–398.

Appendix

If any properties have transacted three or more times during the study period, then more than one price-relative may be formed and the estimation must take into account the fact that the disturbance terms on these price-relatives are not independent of each other. This issue was first pointed out in the seminal article on repeat-sales price index estimation by Bailey, Muth and Nourse (1963) and was explicated by Palmquist (1982), but a casual review of the repeat-sales literature suggests that most if not all researchers fail to address it.

Consider a property that has sold exactly three times, once in each year of a three-year study period. Then three price-relatives may be formed:

$$\frac{P_i}{\tilde{P}_i} = \frac{P_2}{P_1} = \frac{\gamma X_i^{\alpha_1} e^{\beta_1 Y_i + \tau_1 T_{i1} + \tau_2 T_{i2} + \tau_3 T_{i3} + \varepsilon_i}}{\gamma \tilde{X}_i^{\alpha_1} e^{\beta_1 \tilde{Y}_i + \tau_1 \tilde{T}_{i1} + \tau_2 \tilde{T}_{i2} + \tau_3 \tilde{T}_{i3} + \tilde{\varepsilon}_i}}$$

$$= e^{\tau_1 (T_{i1} - \tilde{T}_{i1}) + \tau_2 (T_{i2} - \tilde{T}_{i2}) + \tau_3 (T_{i3} - \tilde{T}_{i3}) + (\varepsilon_i - \tilde{\varepsilon}_i)}$$

$$= e^{-\tau_1 + \tau_2 + (\varepsilon_{i2} - \varepsilon_{i1})}, \tag{A.1}$$

$$\frac{P_3}{P_2} = e^{-\tau_2 + \tau_3 + (\varepsilon_{i3} - \varepsilon_{i2})} \tag{A.2}$$

and

$$\frac{P_3}{P_1} = e^{-\tau_1 + \tau_3 + (\varepsilon_{i3} - \varepsilon_{i1})}. \tag{A.3}$$

Only $N - 1 = 2$ of these price-relatives are independent, so one must be dropped. Importantly, it does not matter which,[3] so let us drop (A.3). The covariance matrix is no longer equal to $\sigma^2 I$; rather, if we define $\nu_2 = \varepsilon_2 - \varepsilon_1$ and $\nu_3 = \varepsilon_3 - \varepsilon_2$, then

$$E(\nu\nu') = E\left(\begin{bmatrix} \varepsilon_2 - \varepsilon_1 \\ \varepsilon_3 - \varepsilon_2 \end{bmatrix} [\varepsilon_2 - \varepsilon_1 \quad \varepsilon_3 - \varepsilon_2]\right)$$

$$= E\begin{bmatrix} \varepsilon_2^2 - 2\varepsilon_1\varepsilon_2 + \varepsilon_1^2 & \varepsilon_2\varepsilon_3 - \varepsilon_1\varepsilon_3 - \varepsilon_2^2 + \varepsilon_1\varepsilon_2 \\ \varepsilon_2\varepsilon_3 - \varepsilon_1\varepsilon_3 - \varepsilon_2^2 + \varepsilon_1\varepsilon_2 & \varepsilon_3^2 - 2\varepsilon_2\varepsilon_3 + \varepsilon_2^2 \end{bmatrix} \tag{A.4}$$

and, as $E(\varepsilon_i^2) = \sigma^2$ and $E(\varepsilon_i \varepsilon_j) = 0$ when $i \neq j$,

$$E(\nu\nu') = \sigma^2 \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = \sigma^2 \Omega \tag{A.5}$$

where $\Omega \neq I$. Because we know the exact structure of the covariance matrix, we can apply the estimator suggested by Aitken (1935). Specifically, we premultiply both sides of the regression equation by the root matrix **P** such that $P'P = \Omega^{-1}$; then $E(P\nu\nu'P') = \sigma^2 P\Omega P' = \sigma^2$.

[3] Gao and Wang (2004), however, argue that the transactions selected to create each price-relative can affect both the estimated rate of appreciation and the confidence interval around the estimated appreciation rate.

Now consider a property that sold exactly three times, but for which the second and third transactions both occurred during Year 2. Now the price-relative shown in Equation (A.2) cannot be formed, but Equations (A.1) and (A.3) can still be used. The premultiplication matrix **P**, however, is different, because

$$E(\nu\nu') = E\left(\begin{bmatrix} \varepsilon_2 - \varepsilon_1 \\ \varepsilon_3 - \varepsilon_1 \end{bmatrix} [\,\varepsilon_2 - \varepsilon_1 \;\; \varepsilon_3 - \varepsilon_1\,]\right)$$

$$= E\begin{bmatrix} \varepsilon_2^2 - 2\varepsilon_1\varepsilon_2 + \varepsilon_1^2 & \varepsilon_2\varepsilon_3 - \varepsilon_1\varepsilon_2 - \varepsilon_1\varepsilon_3 + \varepsilon_1^2 \\ \varepsilon_2\varepsilon_3 - \varepsilon_1\varepsilon_2 - \varepsilon_1\varepsilon_3 + \varepsilon_1^2 & \varepsilon_3^2 - 2\varepsilon_1\varepsilon_3 + \varepsilon_1^2 \end{bmatrix} \text{so that}$$

$$E(\nu\nu') = \sigma^2 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. \tag{A.6}$$

Finally, consider a property that sold exactly four times, with the first two occurring in year 1 and the last two occurring during year 2. We could in principle use any $N - 1 = 3$ independent price-relatives of the $N(N-1)/2 = 6$ available; again, it would not matter which we choose, except that in this case we cannot form $\frac{P_2}{P_1}$ or $\frac{P_4}{P_3}$. Let us, then, use $\frac{P_3}{P_1} = e^{-\tau_1+\tau_3+(\varepsilon_{i3}-\varepsilon_{i1})}$, $\frac{P_3}{P_2} = e^{-\tau_2+\tau_3+(\varepsilon_{i3}-\varepsilon_{i2})}$ and $\frac{P_4}{P_1} = e^{-\tau_1+\tau_4+(\varepsilon_{i4}-\varepsilon_{i1})}$. Now

$$E(\nu\nu')$$

$$= E\left(\begin{bmatrix} \varepsilon_3 - \varepsilon_1 \\ \varepsilon_3 - \varepsilon_2 \\ \varepsilon_4 - \varepsilon_1 \end{bmatrix} [\,\varepsilon_3 - \varepsilon_1 \;\; \varepsilon_3 - \varepsilon_2 \;\; \varepsilon_4 - \varepsilon_1\,]\right)$$

$$= E\begin{bmatrix} \varepsilon_1^2 - 2\varepsilon_1\varepsilon_3 + \varepsilon_3^2 & \varepsilon_3^2 + \varepsilon_1\varepsilon_2 - \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_3 & \varepsilon_1^2 - \varepsilon_1\varepsilon_3 - \varepsilon_1\varepsilon_4 + \varepsilon_3\varepsilon_4 \\ \varepsilon_3^2 + \varepsilon_1\varepsilon_2 - \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_3 & \varepsilon_2^2 - 2\varepsilon_2\varepsilon_3 + \varepsilon_3^2 & \varepsilon_1\varepsilon_2 + \varepsilon_3\varepsilon_4 - \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4 \\ \varepsilon_1^2 - \varepsilon_1\varepsilon_3 - \varepsilon_1\varepsilon_4 + \varepsilon_3\varepsilon_4 & \varepsilon_1\varepsilon_2 + \varepsilon_3\varepsilon_4 - \varepsilon_1\varepsilon_3 - \varepsilon_2\varepsilon_4 & \varepsilon_4^2 - 2\varepsilon_1\varepsilon_4 + \varepsilon_1^2 \end{bmatrix}$$

$$= \sigma^2 \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{bmatrix}. \tag{A.7}$$

For this application we have data on 188 properties that sold exactly three times, and each of them can be represented as a set of price-relatives with the covariance matrix shown in (A.5). A covariance matrix of the same form (but appropriately larger) can be used to represent 157 properties that sold exactly four times, 85 properties that sold exactly five times, 19 properties that sold

Table A.1 ■ Parameter estimates for Model 3 under alternative treatments of error covariance.

Variable	No Correction	Case-Shiller Heteroskedasticity Correction	Case-Shiller Correction with Aitken Estimator
Intercept	0.03882	0.03973	0.03531
$T_{81} - \tilde{T}_{81}$	0.04742	0.04876	0.04182
$T_{82} - \tilde{T}_{82}$	0.06453	0.06586	0.05831
$T_{83} - \tilde{T}_{83}$	0.08319	0.08391	0.07927
$T_{84} - \tilde{T}_{84}$	0.06635	0.06549	0.07284
$T_{85} - \tilde{T}_{85}$	0.05242	0.05256	0.03968
$T_{86} - \tilde{T}_{86}$	0.01478	0.01494	0.00280
$T_{87} - \tilde{T}_{87}$	−0.02162	−0.02146	−0.02728
$T_{88} - \tilde{T}_{88}$	−0.03390	−0.03349	−0.04659
$T_{89} - \tilde{T}_{89}$	−0.10258	−0.10290	−0.11481
$T_{90} - \tilde{T}_{90}$	−0.15533	−0.15595	−0.16352
$T_{91} - \tilde{T}_{91}$	−0.20889	−0.20944	−0.20520
$T_{92} - \tilde{T}_{92}$	−0.17632	−0.17494	−0.18142
$T_{93} - \tilde{T}_{93}$	−0.12631	−0.12651	−0.12913
$T_{94} - \tilde{T}_{94}$	−0.11182	−0.11235	−0.11368
$T_{95} - \tilde{T}_{95}$	−0.06564	−0.06603	−0.06385
$T_{96} - \tilde{T}_{96}$	−0.03029	−0.03014	−0.01946
$T_{97} - \tilde{T}_{97}$	0.01411	0.01483	0.02505
$T_{98} - \tilde{T}_{98}$	0.04773	0.04895	0.06567
$\text{CNTM}_i(\phi_{82,i} - \tilde{\phi}_{82,i})$	−0.05638	−0.05489	−0.04998
$\text{NSR}_i\text{DSR}_i(\Delta t_{94,i})$	1.44×10^{-6}	1.42×10^{-6}	1.27×10^{-6}
$(1 - \text{NSR}_i)\text{DSR}_i(\Delta t_{94,i})$	1.04×10^{-6}	9.95×10^{-7}	7.57×10^{-7}

exactly six times, 10 properties that sold exactly seven times, one that sold exactly eight times and one that sold exactly nine times. There are also two properties that sold exactly four times, two that sold exactly five times and one that sold exactly six times and that cannot be represented with a covariance matrix of the form shown in Equation (A.5), but that can be represented with a covariance matrix of the form shown in Equation (A.6).

As noted, the left-hand-side and right-hand-side variables for properties that sold more than twice during the study period were premultiplied by the root matrix **P** such that $P'P = \Omega^{-1}$ before performing the regression analysis. Properties that sold exactly two times, of course, need no premultiplication because there is only one available price-relative. If the two transactions occurred during the same year—or if all transactions occurred during the same year for a property that sold more than twice—then the property could not be used at all in the analysis.

The effect of applying the weights suggested by Bailey, Muth and Nourse (1963) and Palmquist (1982) to the data at hand is shown in Table A.1, which compares the coefficients estimated for Model 3 without any correction, with just the Case-Shiller correction for heteroskedasticity, and with the Aitken estimator suggested by Palmquist as well as the Case-Shiller heteroskedasticity correction. Comparing the estimates of the (constant) contamination discount, failing to correct at all for nonspherical disturbance terms, results in an estimate of 5.48% ($= 1 - e^{-0.05638}$), whereas applying the Case-Shiller correction for heteroskedasticity reduces that estimate by 2.6%, to 5.34%. Employing the Aitken estimator, however, reduces it by another 11.1%, to 4.88%. This means that the Case-Shiller heteroskedasticity correction accounted for just 23% of the total change in the estimated contamination discount from taking into account the nonsphericality of the disturbance terms, while the Aitken estimator resulted in more than three times as much additional change in the empirical estimate.

Finally, it is worthwhile emphasizing that in many cases data on property transactions can be used in a repeat-sales model even if two transactions occur during the same increment of time (*e.g.*, year). In general, they should not be dropped if it is possible to include them for two reasons. First, of course, because additional data should improve the efficiency of the empirical estimates, and second because the manner in which observations might be selected for exclusion may bias the empirical results. In particular, for example, one can imagine a researcher sorting transactions by property and date, and then deleting any transaction of a given property other than the first for that property during that time increment. If this selection rule tends to result in the exclusion of transactions at higher prices within each time increment, then the resulting price index will tend to be biased to suggest lower index levels in earlier time periods and higher levels in later time periods. Of course, a strategy of randomly selecting the transaction to be excluded would eliminate this bias, but the empirical estimates would still be inefficient.