Modeling the Effects of Refinery Emissions on Residential Property Values

Authors Robert A. Simons, Youngme Seo, and Paul Rosenfeld

Abstract

We examined the effects of refinery air pollution on house prices near Houston, Texas. The affected area was identified through AERMOD air modeling of past releases of sulfur dioxide, a proxy for respiratory risk. A total of 3,964 residential MLS sales from 2006 to 2011 were used to populate an OLS model, a spatial model, and a spatial model with an additional endogenous variable. The findings indicate that air pollution has a significant negative 6%–8% loss on house prices. For one year, the negative effect is shown to generally diminish with distance up to about two miles from the refinery.

The third largest oil refinery in the United States is located on the southern side of the Houston metropolitan area. It is one of the single largest sources of air pollution in the U.S. It has a troubled operating record, with a deadly explosion in 2005, and 77 reported emissions (of which 12 releases were above those allowed by operating permits) from December 2008 to 2010. The most sustained release involved a 40-day long event beginning in April 6, 2010, when over half a million pounds of petrochemicals, including 17,000 pounds of benzene, were flared off (Cheremisinoff, 2011). The release was caused by technical and maintenance problems with the refinery's Ultracracker unit. Theory and the peer-reviewed empirical literature indicate that real property in south Texas communities near the facility would be negatively affected by these harmful chemical releases.

Few studies that measure air-based environmental impacts on house prices have paid attention to how an affected area is delineated. They have generally defined an affected (subject) area by arbitrarily drawing buffer rings or measuring distances, or by using a zonal approach designating the property as either "in" or "out" of an affected area. The affected area in this research is based on the highest concentration of sulfur dioxide (SO_2) based on 2009 and 2010 emissions, and was scientifically determined by AERMOD.[1] The control areas (unaffected by releases from this particular refinery) include those portions of Texas City and La Marque outside the subject area and demographically similar portions of Pasadena, Baytown, and Deer Park in Harris County. Because this part of Texas is "petroleum positive," it is expected that there is some air pollution in the control areas, and that homebuyers are generally tolerant of petroleum activities because of its positive effect on the economic base.

This research employs hedonic regression to measure the price discount attributable to the airborne chemical releases from the refinery. About 4,000 usable residential property sales from 2006 through mid-October, 2011 were obtained from the local MLS. Two time and space models, as well as three methodological approaches (OLS, SARAR- a spatial model, and SARAR with an additional endogenous variable for mortgage foreclosure) were used to estimate the impact of air pollution from the refinery on residential property. The main finding indicates that significant losses of residential property values are due to pollution from the refinery and the disclosure of airborne chemical releases. There are about 6%–8% losses to home values within the affected area after the major release event, and, based on an analysis of sales in 2011, the negative effect on residential property values fades away with distance from the source.[2]

Literature Review

In this section we discuss the peer-reviewed literature concerning the effect of pollution on residential property values, defining an affected area, and methodological approaches. Properties that are believed to be contaminated may experience substantial diminution in property value, especially before they are remediated and/or officially designated as worthy of no further action. This discount can be substantial. Even after a property has been cleaned, damages are still expected to persist because of the potential for a future reoccurrence of the problem (Simons, Bowen, and Sementelli, 1999; Simons, 2002). The potential for future airborne chemical releases from the still-operating polluting refinery (and thus property owners' associated concern over that potential) remains likely.

Flower and Ragas (1994) studied the effects on residential property values of two petroleum refineries located 1½ miles apart in St. Bernard Parish, Louisiana, just east of New Orleans. They used hedonic regression models to analyze sales of 1,999 homes from 1979 to 1991 near the refineries, based on proximity and air pollution. They found losses of 5% in the area near one refinery and 1.5% for homes within half a mile of the other refinery. Proximity, neighborhood prestige, and the quality of a buffer were found to contribute to differences in the losses experienced by homes near the refineries. The authors used a form of distance rings to determine affected areas.

Simons, Seo, and Robinson (2014) studied the effect of a new hog farm operation on nearby residential property values in a rural area near Benton, Kentucky. Using regression analysis of 240 homes sold from 2005 to 2012, they found that homes within a 1.25 mile zone around the facility had sales prices 25% below comparable homes in the control area. Wind direction was also a major factor, as homes directly downwind within the affected zone experienced significantly higher losses, regardless of distance.

There have also been several general studies of the effects of air pollution on residential property values. Figueroa et al. (1996) surveyed a random sample of households in Santiago, Chile and used hedonic regression to determine the owners' marginal willingness to pay for a 50% improvement in air quality (measured in terms of the concentration of 10-micron particulate matter [PM_{10}]). They found that owners would pay 3.3% of their property's value to live in a neighborhood with 50% cleaner air.

In their hedonic meta-analysis of 37 studies of marginal willingness to pay for improved air quality across several U.S. cities, using ordinary least squares (OLS) regression and an econometric model, Smith and Huang (1995) found that every reduction of 1 $\mu g/m^3$ in PM_{10} resulted in an increase of 0.1% of property value in average willingness to pay.

Anstine (2003) reported an 8% loss in assessed property value for homes located two miles away from a Jonesborough, Tennessee rubber compounding facility that emitted foul odors and air pollution. His study, which employed hedonic regression analysis on data from 171 residential sales in 1996, also found no significant effect on values for homes located near a heavy metals processing plant that did not appear to be polluting the local environment. Homes located between the two plants showed a loss in value of nearly 14%. This study provides evidence that public perception about an industrial facility's environmental performance and potential for risk is a factor in determining house value, beyond the measurable scientific risk that may exist.

Hoen et al. (2011) examined the effect of wind energy facilities on neighboring residential property values through scenic vista and nuisance. The affected area was drawn by multiple buffer rings up to five miles and distance. Time aspects, pre-, post- announcement, and construction were considered in various models. They find that the area's stigma effect generally is not statistically significant and tends to fade away rapidly with time. The nuisance stigma effect has a negative effect on sales, a loss of 4.1% within a half mile and 6.4% within a quarter mile. On the contrary, the scenic vista stigma effect is significant, ranging from 9% to 10% on average and 33% to 35% in areas with water frontage or situated on a cul-de-sac.

Anselin and Gallo (2006) investigated the spatial effects of air quality (ozone) on house prices in Los Angeles, Riverside, San Bernardino, and Orange County, California. They utilized spatial interpolation of point measures of air quality (Thiessen polygons, inverse distance weighting, kriging, and spline) and spatial econometrics (spatial lag and spatial error), along with OLS. They found that there were significant negative effects of ozone on property values. They emphasized that OLS is more likely to be biased, and a properly specified model that considers spatial autocorrelation yields the best results.

Fernandez-Aviles, Minguez, and Montero (2012) studied several air pollutants (including SO_2), individually and in a combined index, and their effect on housing prices in Madrid, Spain, a market that they maintain has a high awareness of air pollution. Their data set had almost 11,800 sales from 2009. They considered a standard OLS hedonic regression model, as well as focused on a careful examination of spatial hedonic models, including a spatial model with the same form used by Osland (2010) in modeling hedonic price models, and in a similar spatial approach as Montero and Larraz (2011), who also modeled sales with a much smaller real estate data set in Toledo, Spain. The authors did not find a significant relationship between air quality and housing prices in the Madrid market, however.

Berkman, Hubbard, and Savage (2012) studied the impact of particular matter on residential property values in Ponca City, Oklahoma. Using only hedonic property value econometric models and property-specific particulate matter concentrations, they found that an increase of 10% in particulate matter concentration reduced property values by 1.1%. Pollution sources included a carbon black plant and a nearby oil refinery. A key independent variable was the concentration of particulate matter, which was calculated using an isopleth map generated by AERMOD, although no spatial boundary of emissions was set.

Thus, it appears that most air proximity studies use a zone approach (in or out) or commonly a distance-only concentric rings approach to determining classifying the relationship between air pollution and residential property values. Some studies show that wind direction (a proxy of an affected area driven by odors, in one case) is also important. Although Berkman, Hubbard, and Savage (2012) used an air model (AERMOD), it was used as an independent variable, not to set the edge of the study area. Thus, we are the first to scientific data (AERMOD) to delineate the boundary, and can contribute as a refinement in properly applying scientific data over straight (and convenient) proximity.

To summarize, there are several examples in the literature where researchers measured the impact of various types of air quality problems on house prices. The prevalent methodologies used to define an affected area were either drawing a buffer or measuring the distance from the source. However, it is more likely that an affected area would more accurately be set by environmental factors such as wind direction; thus, a location two miles upwind from a pollution source might be outside the affected area, but a location three miles downwind might be affected. We employ air modeling (AERMOD) to deduce the boundaries of a potentially affected area polluted by emission from an oil refinery, in combination with a concentric rings approach.

Property Data and Models

The study area includes six cities. The affected (subject) area includes parts of Texas City and La Marque, Texas, located in the southern part of the Houston Metropolitan area near Galveston, where the major effects of the air pollution on property values are expected. The potentially affected area was derived using the

AERMOD model (Chen, 2012). The model generates an irregular area (e.g., not concentric rings) that reflects sustained emissions from the main (but not only) source of point-source air pollution in the study area. The properties inside this boundary are considered to be in the subject area.

Once the subject area was determined, control areas in east and south Metro Houston with similar demographics, housing stock, and proximity to petroleum facilities (but not badly polluting refineries) were selected by analyzing secondary data and conducting site visits. The most suitable areas were located in part of the city of Pasadena, and smaller neighborhoods in Deer Park and Baytown. Areas in Texas City and La Marque outside the designated boundary were also included. Exhibit 1 provides a map of the subject and control areas.

Once the control areas were determined, we tested to see if the levels of air pollution were significantly different than the air quality in the case area. The air modelling process is further explained in the Appendix and the results are in Exhibit A1 in the Appendix. They indicate that the air quality (a higher number is undesirable) in the subject area (2.46) is higher than in the main control area (1.35), and that this difference is statistically significant. Both areas have petroleum-related activity, but the control areas do not have any highly-polluting petroleum refineries.

The initial dataset contained 8,246 single-family home sales obtained from MLS sales data at the parcel level that were in or near the subject and control areas. The data covered sales at the parcel level during the period from 2006 through mid-October 2011 in the six cities in the two counties. Approximately 35% of sales were in the subject area, and the average sales price there was $76,973.

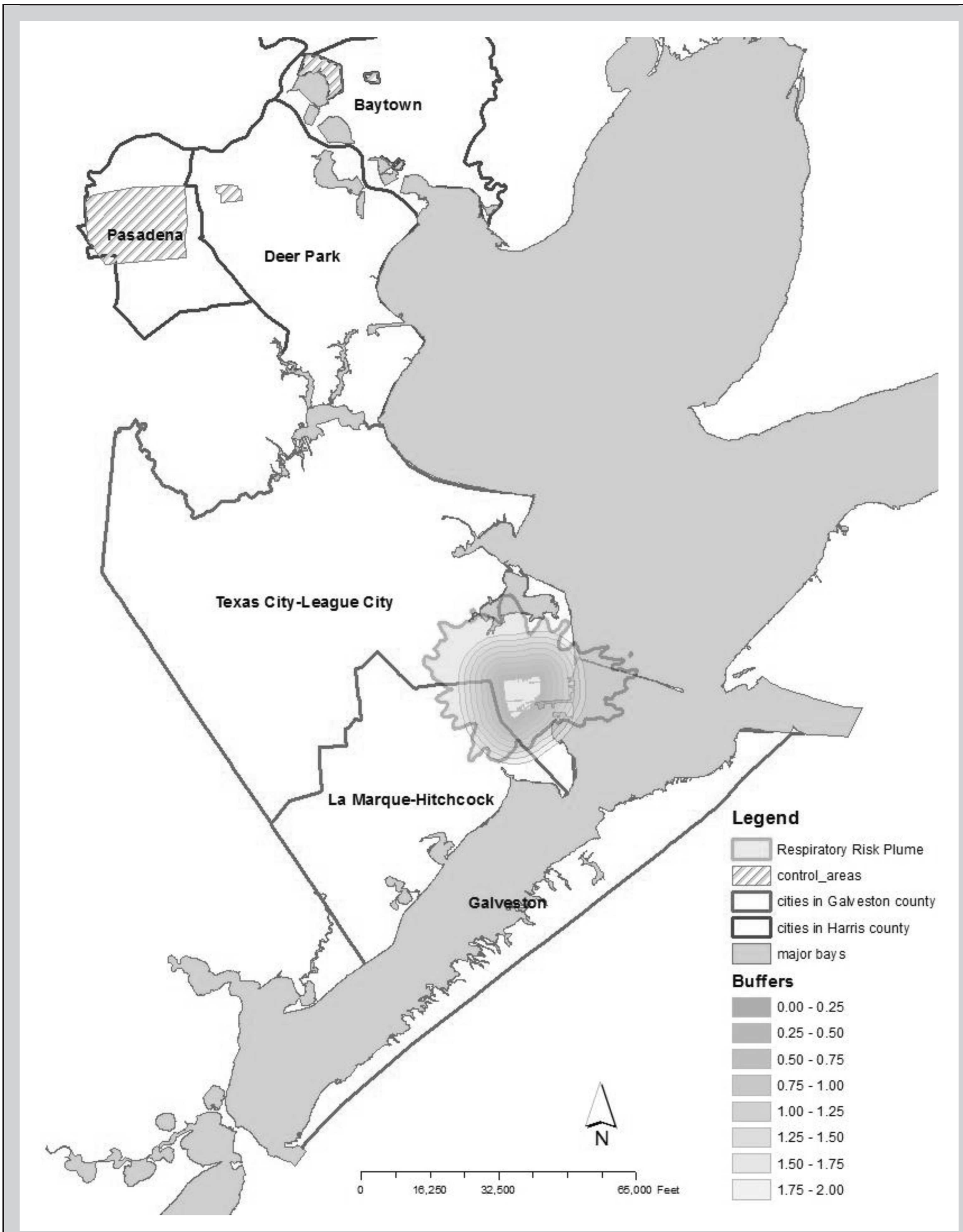
The final dataset contained 3,964 sales. This number was reduced because of missing information, inability to geocode location, sales being outside the prescribed control area boundaries in Harris County, and data outliers.[3]

The dataset contains housing unit, demographic, and location characteristics. Housing unit characteristics include sale amount and year, square footage of the unit, year built, number of bedrooms and bathrooms, lot size, private swimming pool, garage, cooling system, heating system, and foreclosure status.[4] The lot size, living area, and age variables are logged, the others are dummy variables.

Demographic characteristics include median income and educational attainment from the 2000 U.S. Census, measured at the block group level. Demographic variables are median income and education attainment for the population over age 25 including the percentage of residents who have attained a high school diploma and bachelor's degree.

Location variables are also used. It is assumed that proximity to an airport has a negative effect on residential property values due to noise nuisance. The same principle holds for both primary roads and railroad: 0.1 mile buffers were drawn along these linear nuisances. On the other hand, living close to water has a positive effect due to both an interesting view and for being close to water for recreation.

Exhibit 1 | Subject and Control Areas



A description of the data show that homes had an average sales price of $95,632, sat on a 7,500 square-foot lot (equivalent to a log of just under 9), with 3.06 bedrooms, and 1.58 full bathrooms. Census tract household income in 2000 was $41,102, and 8% of households had a bachelor's degree. A total of 12% of the properties were sold in 2011, 26% experienced mortgage foreclosure, 35% were located within the subject area, and 14% sold in the subject area after 2010. Exhibit 2 contains the descriptive statistics.

Moving on to the hedonic regression models, the basic model has its subject area defined by the irregular polygon determined by the AERMOD procedure, based on releases in 2009 and 2010, the year that included a major release event from the refinery. The basic OLS model is specified as follows:

$$\begin{aligned} Ln\_HP = \beta_0 + \beta_1 HC + \beta_2 N + \beta_3 LOC + \beta_4 TIME \\ + \beta_5 FORE + \beta_6 REFIN\text{-}PLUME\text{-}AFTER + \varepsilon, \quad (1) \end{aligned}$$

where *Ln_HP* is the log form of sales price of each home that sold in this data set; β_0 is the model intercept; *HC* is a vector of physical housing characteristics described above; *N* is a vector of neighborhood characteristics also described above; *LOC* is a vector of location dummy variables for sales within 0.1 mile of a the four factors described above; *TIME* is the date of sales, before or after the sustained release events, or some other date as discussed below; *FORE* represents whether a home has been foreclosed upon; *REFIN-PLUME-AFTER* (an interaction term combining location in the refinery's air plume after the trigger date) is intended to measure the effect of the event(s) on residential sales price in the designated plume area after 2009 or a later date, which can take different forms as discussed below; and ε is the error term. The OLS results are shown in Exhibit 4, Model 1 (left column of results).[5] On the maps shown in Exhibits 1 and 3, the *REFIN-PLUME-AFTER* boundary is equivalent to the AERMOD respiratory risk plume. This is the framework for the basic OLS model.

However, because house prices tend to be affected by nearby sales and may be spatially dependent, spatial autocorrelation is a concern. To detect spatial autocorrelation, Moran's I and the Lagrange multiplier (LM) tests are used after running the OLS model. These tests provide a non-subjective measure by which connectivity between house sales can be determined. Moran's I is a measure of spatial autocorrelation between a dependent variable and the lagged dependent variable. The value of Moran's I in this data set is 21.81, which is highly significant, indicating strong spatial autocorrelation of the residuals. The LM tests for spatial lag and spatial error and robust lag and robust error; a portmanteau test for serial correlation for both are used and suggest the combined spatial autocorrelation model (SARAR) is appropriate.[6] The spatial model is specified as follows:

Exhibit 2 | Descriptive Statistics

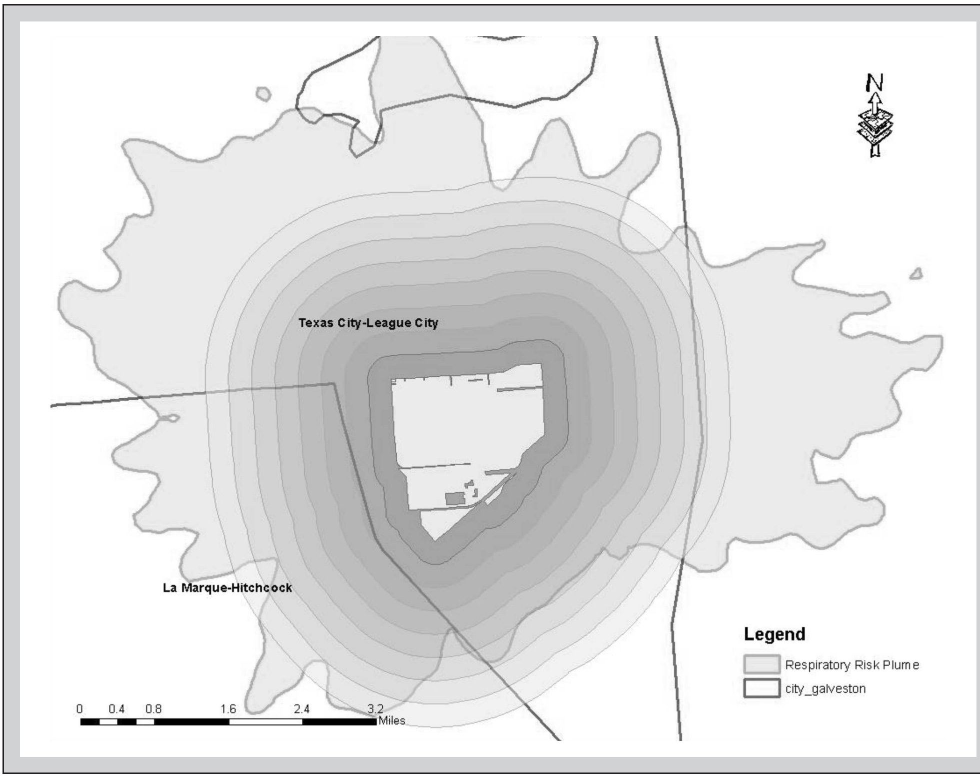
Variables	Label	Mean	Std. Dev.	Min.	Max.
LN_HP	Log of housing price	11.271	0.601	9.306	13.506
SALE_PRICE	Sales price	95632.07	77254.14	11000	734000
W_HP	Weighted log of housing price	11.269	0.495	9.896	13.416
LN_LOT SIZE	Log of lot size	8.983	0.358	7.724	11.513
LN_SQ FOOT	log of square footage	7.277	0.307	6.254	8.382
D_GARAGE	Dummy for garage	0.769		0	1
LN_AGE	Log of age	3.598	0.736	0	4.500
BEDROOMS	# of bedrooms	3.060	0.603	1	7
D_COOL	Dummy for Cooling	0.958		0	1
D_HEAT	Dummy for Heating	0.969		0	1
D_POOL	Dummy for pool	0.039		0	1
BATH_FULL	# of full baths	1.579	0.589	1	5
BATH_HALF	# of half baths	0.209		0	1
D_FORECLOS	Dummy for Foreclosure	0.260		0	1
INCOME	Income	41102	10234	17552	83508
P_HIGH SCH	% high school Diploma	30.852	4.355	16.588	44.574
P_BACHELOR	% bachelor Degree	7.755	4.779	0.569	24.158
DIS_AIRPORT	Distance to the airport (feet)	79830.14	54600.66	16541.57	169404.7

Exhibit 2 | (continued)

Descriptive Statistics

Variables	Label	Mean	Std. Dev.	Min.	Max.
D_ROAD	Dummy for major road	0.032		0	1
D_RAILROAD	dummy for railroad	0.018		0	1
D_WATER V	Dummy for water view buffer	0.043		0	1
D_2006	Dummy for year 2006	0.217		0	1
D_2007	Dummy for year 2007	0.190		0	1
D_2008	Dummy for year 2008	0.175		0	1
D_2009	Dummy for year 2009 (Reference)	0.152		0	1
D_2010	Dummy for year 2010	0.146		0	1
D_2011	Dummy for year 2011	0.121		0	1
D_AFTER	Dummy for sales after 2009	0.419		0	1
MAJ_EFF	Dummy for sales within affected area of subject refinery	0.349		0	1
REFIN-PLUME-AFTER	Dummy for sales within affected area of subject refinery after 2009	0.138		0	1
RPA-05_11	Sale in refinery plume area within .05 mile after 2011	0.002		0	1
RPA-075_11	Sale in refinery plume area within .75 mile after 2011	0.005		0	1
RPA-1_11	Sale in refinery plume area within 1 mile after 2011	0.003		0	1
RPA-125_11	Sale in refinery plume area within 1.25 mile after 2011	0.003		0	1
RPA-15_11	Sale in refinery plume area within 1.5 mile after 2011	0.006		0	1
RPA-175_11	Sale in refinery plume area within 1.75 mile after 2011	0.006		0	1
RPA-2_11	Sale in refinery plume area within 2 mile after 2011	0.006		0	1

Exhibit 3 | AERMOD Plume Boundary



$$Ln\_HP = X\beta + \rho WLnHP + \varepsilon, \varepsilon = \lambda M\varepsilon + v, \quad (2)$$

where X + [*HC*, *N*, *LOC*, *AFTER*, *FORE*, *REFIN-PLUME-AFTER*], and W and M are normalized spatial weighting matrices that parameterize the distance between neighborhoods. The coefficients of ρ and λ are scalars that measure the dependence of housing prices on neighboring housing prices and the spatial correlation in the lags and the errors, respectively. The ρ and λ are between -1 and 1. The weight matrix w_{ij} is a contiguity matrix; $w_{ij} = d_{ij}$ if i and j are neighbors $w_{ij} = 1$, otherwise $w_{ij} = 0$. The weight matrices are "row-standardized" by dividing each element in a row by the sum of the elements in the row.[7] The model is a first-order spatial autoregressive process with first-order spatial-autoregressive disturbances (SARAR (1,1)). The v is assumed to be independent and identically distributed (i.i.d.). The spatial model results are shown in Exhibit 4, Model 2 (middle column of results).

In addition to spatial autocorrelation, the model might suffer from endogeneity because housing foreclosure may be determined by other factors in the model,

Exhibit 4 | Regression Models with OLS, Two-Stage Spatial, and Spatial Lag Models

	OLS		Two-Stage SARAR		Two-Stage SARAR with Additional Endogenous Variable	
	Model 1		Model 2		Model 3	
Variables	Coeff.	*t*-Stat.	Coeff.	*z*-Value	Coeff.	*z*-Value
Intercept	7.303	40.37	7.114	35.99	8.008	26.61
LN_LOT SIZE	−0.019	−1.30	0.031	1.97	−0.034	−1.47
LN_SQ FOOT	0.541	23.32	0.513	22.51	0.551	15.31
D_GARAGE	0.082	6.99	0.073	6.60	0.016	0.81
LN_AGE	−0.179	−23.15	−0.169	−20.14	−0.177	−14.56
D_COOL	0.304	10.95	0.256	9.99	0.087	1.71
D_HEAT	0.104	3.27	0.115	3.95	0.037	0.75
D_POOL	0.055	2.26	0.049	2.18	0.067	1.78
BATH_FULL	0.114	9.65	0.096	8.53	0.094	5.12
BATH_HALF	0.064	5.11	0.048	4.08	0.052	2.67
INCOME	0.00001	11.69	0.00001	8.36	0.00001	5.13
P_HIGH SCH	−0.0001	−0.06	−0.001	−0.76	−0.002	−0.88
P_BACHELOR	0.003	1.58	0.002	0.58	0.003	0.73
DIS_AIRORT	−0.000001	−9.57	−0.000001	−5.41	−0.000001	−4.70
D_ROAD	0.103	3.47	0.054	1.56	0.082	1.75
D_RAILROAD	−0.136	−3.52	−0.069	−1.59	−0.139	−2.28
D_WATER V	0.850	29.77	0.807	19.74	0.740	15.28
REFIN-PLUME-AFTER	−0.075	−3.74	−0.072	−3.83	−0.058	−1.85
D_FORECLOS	−0.409	−37.05	−0.384	−37.84	−1.223[a]	−11.10[a]
D_AFTER	−0.098	−9.45	−0.101	−4.27	−0.003	−0.14
Ρ	NA		0.000	−0.02	0.000	1.11
Λ	NA		0.077	26.97	0.011	1.96
l						
R^2	76.06		NA			
Wald (χ^2)	NA		727.46 for error		3.84 for error	
F-values	663.8		NA		NA	

Note: The number of observation in all models is 3,964.
[a]Predicted using two-stage model.

and not be fully exogenous as per the five standard OLS assumptions. Nearby distressed property such as tax-delinquent vacant lots was suspected of suffering from endogenous issues, and was properly modeled using a two-stage least squares approach (Simons, Quercia, and Maric, 1997). Similarly, and more recently, residential mortgage foreclosure status is a complex variable that may also be related to other independent variables in the model. Clauretie and Daneshvary (2009) modeled residential mortgage foreclosure using both a spatial model and a two-stage approach. They considered foreclosure status as a proxy for other variables, including house condition, marketing time, and proximity to other foreclosures.[8] The Durbin-Wu-Hausman test is used to assess endogeneity issues, and the f-value from the Durbin-Wu-Hausman test is 25,319.33, which is highly significant at the 99% confidence interval, indicating there is an endogeneity problem. Thus, a two-stage spatial model with endogenous variables is used. The model is specified as follows:

$$Ln\_HP = Z\beta + \varepsilon, \ \varepsilon = \lambda M\varepsilon + \upsilon, \tag{3}$$

where $Z = [\overline{X}, Wln(HP)]$. The coefficient for foreclosure was created using an instrumental variable using a two-stage spatial (SARAR) model. Brivand and Piras (2013) explain the steps of estimation for SARAR. First, the initial estimator of β is obtained using the regression residuals. The sample moment β and residual obtained from the first step are transformed into a generalized spatial two-stage least squares model. In the second step, the variance-covariance matrix of the sample moment vector is estimated based on the residuals from the generalized least squares model. This model also allows adding additional endogenous variables, and these are included as Model 3 (and later as Model 6). The advantage of this method is simple computation with large samples, and it generates consistent parameters compared to the maximum likelihood method (Kelejian and Prucha, 1998; Arraiz, Drukker, Kelejian, and Prucha, 2010). The two-stage SARAR results are shown in Exhibit 4, Model 3 (right column of results).

Hedonic Model Results

Exhibit 4 presents the results of the effect model for the OLS, the spatial autoregressive with a spatial autoregressive (SARAR) disturbance, and a two-stage SARAR with an additional endogenous variable for mortgage foreclosure. The OLS result in Model 1 indicates that the R^2 is 76%, which is satisfactory. Likewise the F-statistic is 663.8 and is highly significant. The coefficients for housing characteristics and neighborhood characteristics are as expected by theory, with most variables significant at over a 95% level of confidence. Housing square footage, number of bathrooms, and dummies for garage, cooling, heating, and pool have the expected positive signs and are statistically significant. Further, the log of age variable has the expected negative sign and is statistically significant.

The signs and magnitudes of the neighborhood characteristics, median income, and educational attainment variables are also as expected. The median income variable has a positive sign and a correspondingly significant *t*-value. The location variables are also statistically significant, as is the foreclosure dummy variable at −41%. The coefficient of the time dummy variable for sale after 2009 (*d_after*) has a negative sign and is statistically significant. The main result also shows that the parameter estimate for sale in the subject area after 2009 is −8%, and is statistically significant.

After running the OLS, Moran's I and Lagrange Multiplier tests were conducted to detect spatial autocorrelation: results indicate that there is a spatial autocorrelation problem in the lag and the errors. The SARAR results in Model 2 are similar to the OLS results, except for the location-related variables. The coefficients of the dummies for road, railroad, and water have changed, as well as the significance, now measured by z-values. The Ward test was conducted for the lag and the error, and shows that χ^2 for the error is 727.46, which is statistically significant, while the χ^2 for the lag is not significant.[9] The coefficient on *REFIN-PLUME-AFTER* is −7.2%, and statistically significant, very close to the finding with OLS alone.

Assuming that foreclosure is not an exogenous variable, but is rather endogenous, the OLS parameter estimates for foreclosure and *REFIN-PLUME-AFTER* may be biased. After using the predicted foreclosure variable presented in the two-stage Model 3, comparing the OLS and the two-stage spatial models reveals that the coefficients for *REFIN-PLUME-AFTER* are slightly down from −0.075 in the OLS to −0.058 in the two-stage spatial model. The parameter estimate maintains statistical significance at 90%. The difference between the two means is statistically significant. Other variables' parameter estimates are generally similar to those in the basic OLS shown in Model 1. The exceptions are the dummy variables for garage and heat, both of which become statistically insignificant under Model 3. Also, the dummy for time after 2009 has become insignificant. This is likely due to the effects of the (now predicted) mortgage foreclosure, which Clauretie and Daneshvary (2009) consider a proxy variable. Also, foreclosure rates throughout the U.S. were heightened during the post-2009 period. Although it is impossible to directly compare the estimates due to different modeling approaches, the estimates from OLS appear up-biased for *REFIN-PLUME-AFTER* and down-biased for foreclosure because the OLS model suffers from spatial autocorrelation and endogeneity issues.

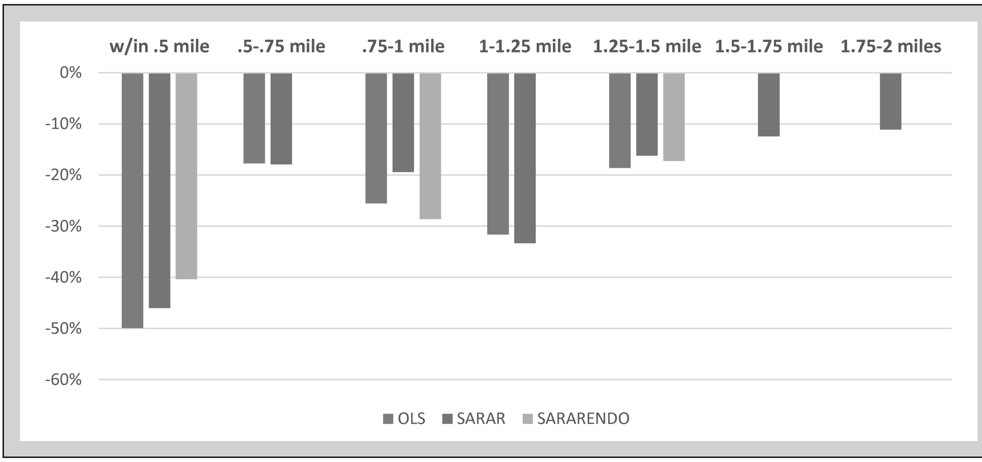
Shifting gears to spatial variation within the subject area, (scientifically determined though air modeling) the commonly used distance-rings approach is used to estimate more precisely the effect of proximity to the refinery.[10] This is demonstrated in Exhibit 5, where *REFIN-PLUME-AFTER* is redefined as only year 2011 sales, to reflect higher losses expected after the market reacts to the major April 2010 release events. House sales (whether within the AERMOD boundary or outside) are differentiated into distance zones based from the refinery: quarter and half-mile bands up to two miles away are modeled. The same three

Exhibit 5 | Interactive Models with One Year and Multi-distance Bands

	OLS		Two-Stage SARAR		Two-Stage SARAR with Additional Endogenous Variable	
	Model 1	Model 2	Model 3		Model 1	Model 2
Variables	Coeff.	*t*-Stat.	Coeff.	*z*-Value	Coeff.	*z*-Value
Intercept	7.319	40.59	7.131	36.24	7.902	28.67
LN_LOT_1	−0.022	−1.48	0.028	1.80	−0.031	−1.42
LN_SQUARE FOOT	0.540	23.36	0.513	22.59	0.547	16.65
D_GARAGE_1	0.092	7.94	0.083	7.63	0.034	1.87
LN_AGE	−0.178	−23.09	−0.168	−20.10	−0.176	−15.67
D_COOL_1	0.308	11.14	0.261	10.24	0.119	2.59
D_HEAT_1	0.099	3.12	0.110	3.80	0.044	0.98
D_POOL_1	0.056	2.32	0.050	2.25	0.066	1.95
BATH_FULL_1	0.112	9.52	0.095	8.47	0.094	5.62
BATH_HALF_1	0.063	5.07	0.047	4.03	0.052	2.95
INCOME	0.00001	11.71	0.00001	8.39	0.00001	5.85
P_HIGH SCH	0.00019	0.14	−0.001	−0.71	−0.002	−0.87
P_BACHELORS	0.003	1.39	0.001	0.46	0.002	0.72
DIS_AIRPORT_1	−0.000001	−9.81	−0.000001	−5.35	−0.000001	−5.28
D_ROAD_1	0.101	3.41	0.051	1.49	0.079	1.81
D_RAILRO_1	−0.138	−3.56	−0.069	−1.60	−0.136	−2.40
D_WATER_V B	0.857	30.24	0.808	19.84	0.757	16.76
D_AFTER_1	−0.101	−10.23	−0.103	−11.33	−0.021	−1.19
D_FORECLOS_1	−0.407	−36.94	−0.381	−37.73	−1.095*	−11.28[a]
RPA-05_11	−0.499	−4.14	−0.460	−4.13	−0.403	−2.38
RPA-075_11	−0.177	−2.51	−0.179	−2.63	0.021	0.21
RPA-1_11	−0.255	−2.73	−0.194	−2.23	−0.286	−2.17
RPA-125_11	−0.316	−3.55	−0.333	−4.07	−0.132	−1.03
RPA-15_11	−0.186	−2.99	−0.162	−2.77	−0.172	−1.96
RPA-175_11	−0.093	−1.47	−0.124	−2.09	−0.061	−0.68
RPA-2_11	−0.067	−1.06	−0.111	−1.83	−0.091	−1.01
P			0.000	−0.09	0.0004	1.06
Λ			0.077	27.19	0.017	2.84
R^2	76.25		NA		NA	
Wald(χ^2)	NA		739.55 for error		8.08 for error	
F-values	509.98		NA		NA	

Note: The number of observation in all models is 3,964.
[a] Predicted using two-stage model.

Exhibit 6 | Distance Effects for 2011



types of models OLS Model 4, SARAR Model 5, and two-stage endogenous Model 6 are presented.

The R^2 of the interactive hedonic regression Model 4 is slightly higher than OLS Model 1, but the F-statistic is somewhat lower, so the models are roughly equivalent. As shown in Model 4 of Exhibit 5, the effects generally fade according to distance from the refinery. In 2011, and within 0.5 miles, the variable *RPA-5_11* (refinery plume after, within 0.5 miles of the refinery, for 2011) is statistically significant and negative, indicating a price reduction of almost 50% attributable to pollution from the refinery. The negative effects generally decrease further from the refinery (although not monotonically) until about 1.5 miles, where the last significant loss is 19%. Some of these distance bands have less than ten sales however, so small sample size is a concern. Since some sales at two miles are outside the designated boundary, it is not surprising that the results for this distance band are negative, but not statistically significant.[11] As for all models shown in Exhibit 5, the figures are higher than the 6%–8% shown in Exhibit 4 because of the one year selected (2011) is right after the main prolonged release event associated with this refinery in April 2010. Losses within the overall plume area for that year averaged over 10%, when market knowledge was fresh. For the most part, the results for Models 4–6 were consistent with the base OLS runs shown in Model 4. Among the variables not related to refinery discounts, only heat and the post 2009 time dummy became insignificant, likely for the same reasons discussed previously.

The two-stage SARAR Model 5 results are generally the same as those of the OLS except the location-related variables have a modestly smaller price reduction in each band, implying that the coefficients in the OLS model may be biased. The effect of the proximity to the refinery starts at 46% close in, then decreases (but not monotonically) until it fades away after a 12% loss at 1.75 miles. At up to two miles away, the loss is 11%, and significant at a 90% level of confidence.

The two-stage SARAR model with endogenous variables (Model 6) also has generally similar results to Models 4 and 5. The effects for each distance band start at 40% close in, then drop monotonically to 29% at a mile and 17% at 1.5 miles, although there are distance gaps where parameter estimates are not statistically significant. This might imply that the model captures the concentration of low house prices and foreclosure in the subject area.

Exhibit 6 shows the results of the three models over space. Values not significant at a confidence level of least 90% are shown as zero. All three models show a generally decreasing trend: the SARAR alone (Model 5) seems to have the most stable results. However, some zones contain a relatively small number of sales (as small as 8, but typically 12–24 per distance ring), so the results for individual zones should be viewed with caution.

Conclusion

The subject refinery is one of the worst polluting refineries in the U.S. According to the TCEQ Point Source Emission Inventory, Texas City and La Marque are heavily polluted, and their air quality is characterized as unhealthy. The information has widespread awareness among potential homebuyers and the air pollution has a statistically significant negative effect on residential property values. Scientific air modeling (AERMOD), based on the refinery's reported baseline releases for an increase in SO_2 concentrations in ambient air, predicted and identified the potentially affected area. This research is among the first to utilize a scientifically-determined influence area, as opposed to pure proximity, to estimate property value diminution. Berkman, Hubbard, and Savage (2012), in a related article, did not use AERMOD to provide a cutoff of the affected area, and, further, relied solely upon OLS for their conclusions. In addition, the articles cited above generally used straight OLS regressions. Berkman, Hubbard, and Savage employed a spatial model and the spatial model with additional endogenous variables to estimate an unbiased parameter, to complement statistical analysis of using just OLS alone, where this research uses all three and allows limited comparison between model results.

Statistical analysis of air monitoring data demonstrates that SO_2 levels measured in the area affected by the refinery were elevated above levels from the control areas, and that this difference was statistical significant. Thus, residential properties within the subject area are subjected to additional degraded air quality and any associated risks, due to their proximity to the refinery, and this is capitalized into property value reductions.

Looking at all sales after 2009, our first three model results support and show property value losses of 6%–8% (all sales, all years). OLS model results tended to have slightly higher parameter estimates than models that adjusted for spatial autocorrelation (SARAR) or SARAR plus a two-stage model to control for endogeneity of the housing foreclosure variable. These findings are consistent with the peer-reviewed literature cited above (Flower and Ragas, 1994; Anstine, 2003; Figueroa et al., 1996; and Berkman, Hubbard, and Savage, 2012) that addressed refinery emissions or point source air plumes. While Fernandez-Aviles, Minguez, and Montero (2012) did not find a statistically significant relationship between pollution and house prices, their OLS and spatial models generated similar results, as did ours.

Moving to a more precise look at the distribution of losses over space, losses after information about the most serious release in April 2010 was fully incorporated into market knowledge in 2011 indicates losses of up to 40%–50% closest to the refinery, declining to about 20%–30% a mile away, and 16%–19% 1.5 miles away, depending which modeling approach is relied upon. The SARAR model results indicate double digit or greater losses up to two miles away.

Based on sales data for the 2006–2011 period, residential properties in the area affected by the frequency and severity of airborne chemical releases from the refinery have suffered a reduction in property value of 6%–8%, with many areas within the plume area closer to the refinery showing higher losses for the last year data were available.

Appendix

Air Quality

Using the Texas Commission on Environmental Quality's (TCEQ) Geographical Texas Air Monitoring (GeoTAM) Viewer application,[12] three air monitoring sites were identified in the control region that provided parallel continuous SO_2 measurements for 2009–2010 to those in Texas City: Clinton (AQS 482011035), Park Place (AQS 482010416), and Houston Monroe (AQS 482010062). These monitors are the closest active sites to the control area and provide the best representation of local air quality for comparison to the Texas City data.

The available air monitoring data was organized to set up two populations, the "Area of Concern/case area" (Texas City/La Marque), and the "Control Area" or "Background" (Pasadena). All air monitoring data was obtained through the TCEQ Texas Air Monitoring Information System (TAMIS)[13] web interface. The following air monitoring sites provided relevant data for Texas City: Texas City Ball Park (AQS 481670005): Hourly SO_2 data for all of 2009–2010; Subject 31st Street (AQS 481670615): Hourly SO_2 data beginning on 1/21/2010; Subject Logan Street (AQS 481670621): Hourly SO_2 data beginning on 5/21/2010; Subject Onsite (AQS 481670616): Hourly SO_2 data beginning on 3/23/2010. A total of 37,869 measurements were compiled from these four monitors to comprise the "Subject Area" dataset; there were 306 missing data points. Deviations in accuracy of equipment calibration resulted in 294 negative concentrations values

Exhibit A1 | ProUCL Wilcoxon-Mann-Whitney Test Output

Wilcoxon-Mann-Whitney Site vs. Background Comparison Test for Full Data Sets without NDs

Panel A: User selected options

From File	Augmented. wst
Full Precision	OFF
Confidence Coefficient	95%
Substantial Difference	0
Selected Null Hypothesis	Site or AOC Mean / Median Less Than or Equal to Background Mean / Median (Form 1)
Alternative Hypothesis	Site or AOC Mean / Median Greater Than Background Mean / Median

Panel B: Subject area of concern data: all TC

Background Data: All Pasadena

Raw Statistics	Site	Background
Number of Valid Observations	37869	50956
Number of Missing Values	306	0
Number of Distinct Observations	18483	26701
Min.	0	0
Max.	141.9	70.45
Mean	2.456	1.347
Median	1.122	0.514
Std. Dev.	4.628	2.797
SE of Mean	0.0238	0.0124

Panel C: Wilcoxon-Mann-Whitney (WMW) test

H0: Mean / Median of Site or AOC ≤ Mean / Median of Background

Site Rank Sum W-stat	1.93E+09
WMW Test U-stat	2802
WMW Critical Value (0.050)	1.645
P-value	0
Conclusion with Alpha	0.05
Reject H0, Conclude Site > Background	
P-value < alpha	(0.05)

from the three monitors over the two-year period, and each of these values was treated as a zero measurement because concentrations below zero are not physically possible.

The Control Area dataset was comprised of hourly SO_2 concentrations from the three aforementioned monitors bordering the identified Control Area neighborhoods (Clinton, Park Place, and Houston Monroe): a total of 50,956 hourly SO_2 measurements.

The EPA-promulgated ProUCL[14] (V4.1) statistical analysis software was then utilized to examine the concentrations of SO_2 in ambient air in the Texas City/La Marque and Control Area neighborhoods. A Wilcoxon-Mann-Whitney test was performed because the datasets from the respective Area of Concern and Control Area locations were not expected to have equal variances, or distributions. The test treats the two collections of values independently of each other and assesses the probability that one set is significantly elevated above the other. The output of the test is presented as Exhibit A1.[15]

Endnotes

1 AERMOD is an air-dispersion model used to simulate air contaminant concentrations in ambient air. For the purposes of modeling site-wide emissions from the refinery, data reported to the TCEQ (Texas Commission On Environmental Quality) Annual Emissions Inventory were compiled for 2009 and 2010 for SO_2. The subject area was delineated based on the 10^{th} highest ground level one-hour concentration of SO_2 during 2009 through the end of 2010. This level was selected to approximate a moderate (not the highest) sustained level of respiratory risk. Based on peer-reviewed literature (e.g., Crain et al. 1994; Segala et al., 1998; Delfino, Gong, Linn, and Pellizzari, 2003; Lewis et al., 2004; Joseph et al., 2005; Perry et al., 2005; Wilson, Zeng, Zhan, and Zou, 2010; Bhati, Marrapu, Mohan, and Sreeniva, 2011), this level is consistent with a statistically significant elevated level of respiratory risk (Chen, et al., 2012).

2 The senior authors have been retained by plaintiffs' counsel in litigation against the sources of pollution in this legal case.

3 Outliers included sales less than \$10,000; more than \$750,000; age (less than one year, greater than 90 years old); square footage (less than 500 SF or more than 5,000 SF); lot size (less than 1,000 SF, more than 100,000 SF); bedrooms (less than one and more than 7); and bathrooms (less than one and not more than 5).

4 The foreclosure status includes pre-foreclosure.

5 As an extension, an interactive model was created that considers both distance from the refinery (quarter or half mile bands up through two miles away) for only one year, in this case 2011. The interactive model is designed to measure the sales price losses in a more precise location and time frame, such as a "potentially affected" period subsequent to a major release event. The demarcation between those two periods is the point in time when knowledge of the subject event(s) becomes widespread in the local community, such that it can be reasonably inferred that potential buyers of class area properties would be aware of the existence of the conditions at issue, and thus could reasonably be expected to capitalize those effects into their purchase price for the affected properties. These models, numbered 4–6, are discussed later.

[6] The LM tests for LM lag, LM error, Robust lag, Robust error, and SARMA are 768.41, 473.63, 319.73, 24.95, 793.36, respectively. The test results are statistically significant, which means we cannot reject the null hypotheses of no spatial autocorrelation in lag and error.

[7] $W = \overline{w_{ij}}/\Sigma_i \overline{w_{ij}}$. Assuming $W = M$.

[8] Foreclosure status as a proxy for other variables helps explain a conundrum with the pre-temporal condition of causality, since foreclosure is an ex post factor for some of the sales in the model. It is acknowledged that since foreclosure may have affected properties differently during the study period due to a change in national economic conditions, a single foreclosure variable may be simplistic. Still, as a cumulative proxy variable intended to hold constant the effects of foreclosure-related conditions during 2009–2011 at the end of the study period, foreclosure as modeled is adequate as a control variable.

[9] Although the LM test result for the lag indicates there is spatial autocorrelation, the Wald test result after running the two-stage spatial model indicates spatial autocorrelation in the lag is not statistically significant.

[10] Simons and Seo (2011) found a positive externality of a religious facility campus on neighboring housing sale prices. They used hedonic regression analysis using 2,500 sales in Ohio, and identified sales within quarter-mile distance buffers. A similar distance-ring approach was taken by Reichert, Small, and Mohanty (1992), Smolen, Moore, and Conway (1992), and Ready (2010) in their analyses of the negative amenity from proximity to landfills, and by Ding, Simons, and Baku (2000) in modeling housing investment.

[11] There are 29 sales that are outside of the SO_2 plume boundary but inside of the 1.75 and 2 mile rings.

[12] Geographical Texas Air Monitoring (GeoTAM) Viewer, Texas Commission on Environmental Quality. Accessed at http://gis3.tceq.state.tx.us/geotam/index.html.

[13] Raw Data Reports, TAMISWeb v4.0.5. Texas Commission on Environmental Quality. Accessed at http://www5.tceq.state.tx.us/tamis/index.cfm.

[14] ProUCL Software, Site Characterization and Monitoring Technical Support Center, United States Environmental Protection Agency. Accessed at http://www.epa.gov/osp/hstl/tsc/software.htm.

[15] The test results confirmed that concentrations of SO_2 in Texas City are elevated above levels typically measured in the Pasadena region with statistical significance, and concluded that the two datasets represented different populations. To evaluate whether the two datasets could have possibly come from the same population, an additional *t*-test was run under the presumption that the two distributions were equal. Results of the *t*-test were consistent with the Wilcoxon-Mann-Whitney test, and it was confirmed that the Texas City data were significantly different from the Pasadena values.

References

Anselin, L. and J. Le Gallo. Interpolation of Air Quality Measures in Hedonic House Price Models: Spatial Aspects. *Spatial Economic Analysis*, 2006, 1:1, 31–52.

Anstine, J. Property Values in a Low Populated Area When Dual Noxious Facilities are Present. *Growth and Change*, 2003, 34:3, 345–58.

Arraiz, I., D.M. Drukker, H.H. Kelejian, and I.R. Prucha. A Spatial Cliff-Ord-Type Model with Heteroskedastic Innovations: Small And Large Sample Results. *Journal of Regional Science*, 2010, 50:2, 592–614.

Berkman, M.P., K.J. Hubbard, and T.H. Savage. The Adverse Impact of Particulate Matter on Property Values. *International Real Estate Review*, 2012, 15:2, 215–30.

Bhati, S., P. Marrapu, M. Mohan, and A. Sreenivas. Performance Evaluation of AERMOD and ADMS-Urban for Total Suspended Particulate Matter Concentrations in Megacity Delhi. *Aerosol and Air Quality Research*, 2011, 11, 883–94.

Brivand, R. and G. Piras. Comparing Implementations of Estimation Methods for Spatial Econometrics. Working Paper #2013-01, West Virginia University.

Chen, J.A., A.R. Zapata, A.J. Sutherland, D.R. Molmen, B.S. Chow, L.E. Wu, P.E. Rosenfeld, and R.C. Hesse. Sulfur Dioxide and Volatile Organic Compound Exposure to a Community in Texas City, Texas Evaluated Using AERMOD and Empirical Monitoring Data. *American Journal of Environmental Science*, 2012, 8:6, 622–32.

Cheremisinoff, N.P. Trends in BP Texas City Refinery Flaring Practices. August 2011.

Clauretie, T.M. and N. Daneshvary. Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time. *Real Estate Economics*, 2009, 37, 43–67.

Crain, E.F. et al. An Estimate of the Prevalence of Asthma and Wheezing Among Inner-City Children. *Pediatrics*, 1994, 94, 356–62.

Delfino, R.J., H. Gong, W.S.H.Y. Linn, and E.D. Pellizzari. Respiratory Symptoms and Peak Expiratory Flow in Children with Asthma in Relation to Volatile Organic Compounds in Exhaled Breath and Ambient Air. *Journal of Exposure Analysis and Environmental Epidemiology*, 2003, 13:348–63.

Ding, C., R.A. Simons, and E. Baku. The Effect of Residential Investment on Nearby Property Values: Evidence from Cleveland, Ohio. *Journal of Real Estate Research*, 2000, 19:1/2, 23–48.

Fernandez-Aviles, G., R. Minguez, and J.M. Montero. Geostatistical Air Pollution Indexes in Spatial Hedonic Models: The Case of Madrid, Spain. *Journal of Real Estate Research*, 2012, 34:2, 243–74.

Figueroa, E., G. Donoso, G. Lagos, R. Alvarez, and J. Muñoz. Sustentabilidad Ambiental del Sector Exportador Chileno. In: *Sustentabilidad Ambiental del Crecimiento Económico Chileno*, O. Sunkel (ed.). Programa de Desarrollo Sustentable, CAP, Universidad de Chile, 1996, 47–86.

Flower, P. and W. Ragas. The Effects of Refineries on Neighborhood Property Values. *Journal of Real Estate Research*, 1994, 9:3, 319–38.

Hoen, B., R. Wiser, P. Cappers, M. Thayer, and G. Sethi. Wind Energy Facilities and Residential Properties: The Effect of Proximity and View on Sales Prices. *Journal of Real Estate Research*, 2011, 33:3, 279–316.

Joseph, C.L.M., et al. Effect of Asthma Intervention on Children with Undiagnosed Asthma. *The Journal of Pediatrics*, 2005, 146:1, 96–104.

Kelejian, H.H. and I.R. Prucha, A Generalized Spatial Two-stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 1998, 17:1, 99–121.

Lewis, T.C. et al. Identification of Gaps in the Diagnosis and Treatment of Childhood Asthma Using a Community-based Participatory Research Approach. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 2004, 81:3, 472–88.

Montero, J.M. and B. Larraz. Interpolation Methods for Geographical Data: Housing and Commercial Establishment Markets. *Journal of Real Estate Research*, 2011, 33:21, 233–44.

Osland, L. An Application of Spatial Econometrics in Relation to Hedonic House Price Modeling. *Journal of Real Estate Research*, 2010, 32:3, 289–320.

Perry, R.J., A.J. Cimorelli, R.F. Lee, R.W. Paine, S.G. Brode, W.D. Peters, A. Venkatram, J.C. Weil, and R.B. Wilson. AERMOD: A Dispersion Model for Industrial Source Applications. Part II: Model Performance against 17 Field Study Databases. *Journal of Applied Meteorology*, 2005, 44, 694–708.

Ready, R.C. Do Landfills Always Depress Nearby Property Values? *Journal of Real Estate Research*, 2010, 32:3, 233–44.

Reichert, A., M. Small, and S. Mohanty. The Impact of Landfills on Residential Property Values. *Journal of Real Estate Research*, 1992, 7:2: 297–314.

Segala, C., B. Fauroux, J. Just, L. Pascual, A. Grimfeld, and F. Neukirch. Short-term Effect of Winter Air Pollution on Respiratory Health of Asthmatic Children in Paris. *European Respiratory Journal*, 1998, 11:3, 677–85.

Simons, R.A. Estimating Proximate Property Damage from PCB Contamination in a Rural Market: A Multiple Techniques Approach. *The Appraisal Journal*, 2002, 388–400.

Simons, R.A., W. Bowen, and A. Sementelli. The Price and Liquidity Effects of UST leaks from Gas Stations on Residential and Commercial Property Values. *The Appraisal Journal*, 1999, 14(1/2), 29–42.

Simons, R.A., R. Quercia, and I. Maric. The Value Impact of Neighborhood Transition on Residential Sales Price. *Journal of Real Estate Research*, 1997, 15:2, 147–61.

Simons, R.A. and Y. Seo. Religious Value Halos: The Effect of a Jewish Orthodox Campus on Residential Property Values. *International Real Estate Review*, 2011, 14:3, 330–53.

Simons, R.A., Y. Seo, and S. Robinson. The Effect of a Large Hog Barn Operation on Residential Sales Prices in Marshall County, KY. *Journal of Sustainable Real Estate*, 2014, 6:1, 93–111.

Smith, V.K. and J. Huang. Can Market Value Air Quality? A Meta-analysis of Hedonic Property Value Models. *Journal of Political Economy*, 1995, 103, 209–27.

Smolen, G.E., G. Moore, and L.V. Conway. Economic Effects of Hazardous Chemical and Proposed Radioactive Waste Landfills on Surrounding Real Estate Values. *Journal of Real Estate Research*, 1992, 7:3, 293–96.

Wilson, J.G., Y. Zeng, F.B. Zhan, and B. Zou. Performance of AERMOD at Different Time Scales. *Simulation Modelling Practice and Theory*, 2010, 18, 612–23.

Robert A. Simons, Cleveland State University, Cleveland, OH 44122 or r.simons@csuohio.edu.

Youngme Seo, Iowa State University, Ames, IA 50011 or yseo@iastate.edu.

Paul Rosenfeld, Soil Water Air Protection Enterprise (SWAPE), Santa Monica, CA 90405 or prosenfeld@swape.com.