A Hedonic Price Analysis of Traffic Noise in the Twin Cities Housing Market

August 28, 2013

Abstract

This paper implements the hedonic price method to estimate the impact of traffic noise on properties within the St. Paul, Minnesota metropolitan real-estate market over the time period from 2005 - 2010. We use Locally Weighted Regression (LWR) estimation techniques in order to allow the hedonic function to vary over time and space. Significant evidence suggests spatial and temporal non-stationarity in the hedonic price function within our study area. In contrast to previous research, however, we find little evidence that the impact of traffic noise exhibits a significant non-linear relationship with the decibel level of the traffic noise, nor does the impact seem to vary over space. We find significant differences in the impact of traffic noise from the beginning to the end of our sample. Our results suggest that an increase in traffic noise of one decibel decreases house prices by an average of 0.19 percent before September, 2008, and an average of 0.37 percent after September, 2008. While significantly different, both of these estimates are still within the typical range of previous estimates in the literature.

1 Introduction and Past Research

Prolonged exposure to traffic noise affects people in a number of ways, ranging from simple annoyance (Miedema and Oudshoorn 2001; Ouis 2001), to sleep disturbance (Netherlands 2004), to increasing risk for stroke (Sorensen, Hvidberg et al. 2011), hypertension(Jarup, Babishch et al. 2008; Bodin, Albin et al. 2009), and myocardial infarction (Babisch, Beule et al. 2005). The noise level at which such effects are observed does not have to be high. It has been shown that people exposed to traffic noise with a 24-hour average of 55 dBA are found to be at a higher risk for hypertension (Barregard, Bonde et al. 2009; Bodin, Albin et al. 2009), and those exposed to 60 dBA or greater are found to be at a higher risk for stroke (Sorensen, Hvidberg et al. 2011). This impact is expected to grow worse with the increasing number of vehicles in the urban road network and the diminishing number of night-time quiet hours unless mitigating measures are put in place (XXX). Given that much of this increased exposure occurs in areas where people live and work any mitigating measures must be based on a cost-benefit analysis.

An important first step towards such an analysis is to examine real estate price fluctuations as a function of noise levels using hedonic regression. Such work has provided important insight on the nature of this relationship. Thus, Bateman et al. (2001) found that the negative eect of trac noise ranged from 0.08% to 2.22% per dB with a mean around 0.55% per dB. In a review of 65 noise evaluation studies Navrud (2002) found an average decrease of 0.64% per dB. **We need to briefly review other relevant papers here.** However, with the exception of a few studies that have focused on air traffic noise exposure **(we should check on this)**, the relationship between road traffic noise levels and real estate prices has rarely been explored in the United States, even though the former is perhaps the greatest source of noise in residential neighborhoods (XXX). One major reason for the thinness of the scholarly literature on this topic in the US lies in the difficulty of modeling noise propagation over the landscape.

To properly analyze the spatial association between real estate prices and exposure to traffic noise at a landscape level, it is necessary to create a noise surface map at sufficiently detailed spatial resolution to account for the complex and heterogeneous interaction between the noise source and the resistance of the landscape to noise propagation. Implementing such a model can be very difficult. For one thing, the data needed for the model is very extensive and may not even be readily available (e.g., building footprint and height data). Furthermore, it is computationally very intensive. Fortunately, recent developments in Geographic Information Systems and distributive computing have reduced these difficulties, making it much easier to create a noise surface map at landscape level.

In this paper, we seek to contribute to the hedonic literature by examining the relationship between real estate prices and variation in road traffic noise exposure. We approach this issue by means of an empirical case study of the St. Paul, Minnesota urban area. In particular, we construct a locally flexible model that allows for a non-stationary relationship between house prices and our explanatory variables over geography and time. Such Locally Weighted Regression (LWR) models have become increasingly common in published work and have been shown to be more accurate than spatial autoregressive models with misspecified weights matrices.

Our work makes three contributions to the field. First, we utilize new estimates of traffic noise that has never before been used in the hedonic literature. The data explicitly model noise propagation as a function of nearby buildings and landcover, as well as present the results at a fine spatial resolution (10m x 10m). Such estimates, combined with spatially explicit modeling techniques allow us to make our second contribution to the literature. We present some of the first non-parametric estimates of how the impact of traffic noise spatially varies within an urban area in the United States. We find that while the impact of other important control variables in the hedonic function vary over space, the impact of traffic noise does not. Lastly, we present the first estimates of how the Great Recession has affected the implicit price of traffic noise. Contrary to other published work, our results suggest that consumers' marginal willingness to pay to avoid traffic noise actually increased after the recession.

2 Data

The 2010 US Census lists the population of the Twin Cities Metropolitan Region (Minneapolis and St. Paul and their surrounding areas) as almost 3 million residents spread over seven counties. This study examines single family residential home transactions in the Census-defined urban areas of three of the seven counties Dakota, Ramsey and Washington County (see Figure 1). We obtained data from approximately forty thousand sales transactions between 2005 and 2010 (n=42,095) from the 2010 MetroGIS Regional Parcel Dataset published by the Metropolitan Council. In addition to the geographic location and date of the house sale, we collected and/or constructed structural and locational variables commonly used in the hedonic literature. Table 1 provides a brief description of the variables and some basic summary statistics. Figure 1 shows an overview of the study area as well as the spatial distribution of the house sales prices. Table 2 shows a simple correlation matrix of the quantitative variables.

Table 1: Variable Description and Summary Statistics

	min	25%	50%	mean	75%	max	σ
Sale Price (thousands $)	98.8	195.0	241.0	265.5	314.9	675.0	102.8
House Size (feet^2)	390	1158	1628	1750	2188	4000	704
Lot Size (acres)	0.02	0.17	0.25	0.25	0.31	0.60	0.11
Owner Occupancy (0 = no, 1 = yes)	0.0	1.0	1.0	0.8	1.0	1.0	0.4
Year House was Built	1850	1950	1973	1967	1993	2010	32
Traffic Noise (dB)	25.1	50.5	55.3	56.4	62.2	91.5	8.8
Median Income in Census Tract (thousands $)	13.9	54.6	69.1	72.6	89.7	143.3	23.6
Elementary School Standardized Test Scores	336	360	365	365	370	552	11
Distance to Central Business District (km)	1.1	6.8	13.2	14.7	21.8	37.1	8.9
Distance to nearest Park (km)	0.0	1.1	2.2	2.6	3.8	9.7	1.9
Distance to nearest Lake (km)	0.0	0.4	0.8	0.9	1.3	4.4	0.7
Distance to nearest Shopping Center (km)	0.0	0.9	1.5	1.9	2.3	10.8	1.6

2.1 Structural Attributes

According to Wilhelmsson (2000), the most common structural attributes included in real-estate hedonic pricing studies are living area, number of bathrooms, age, garage and lot size. Although the 2010 MetroGIS Regional Parcel Dataset includes structural data on living area, age, garage presence, lot size and owner occupancy for every transaction, we have variables for the number of bedrooms, bathrooms, and size of the garage only for those house sales in one county. We feel confident in our results even without these independent

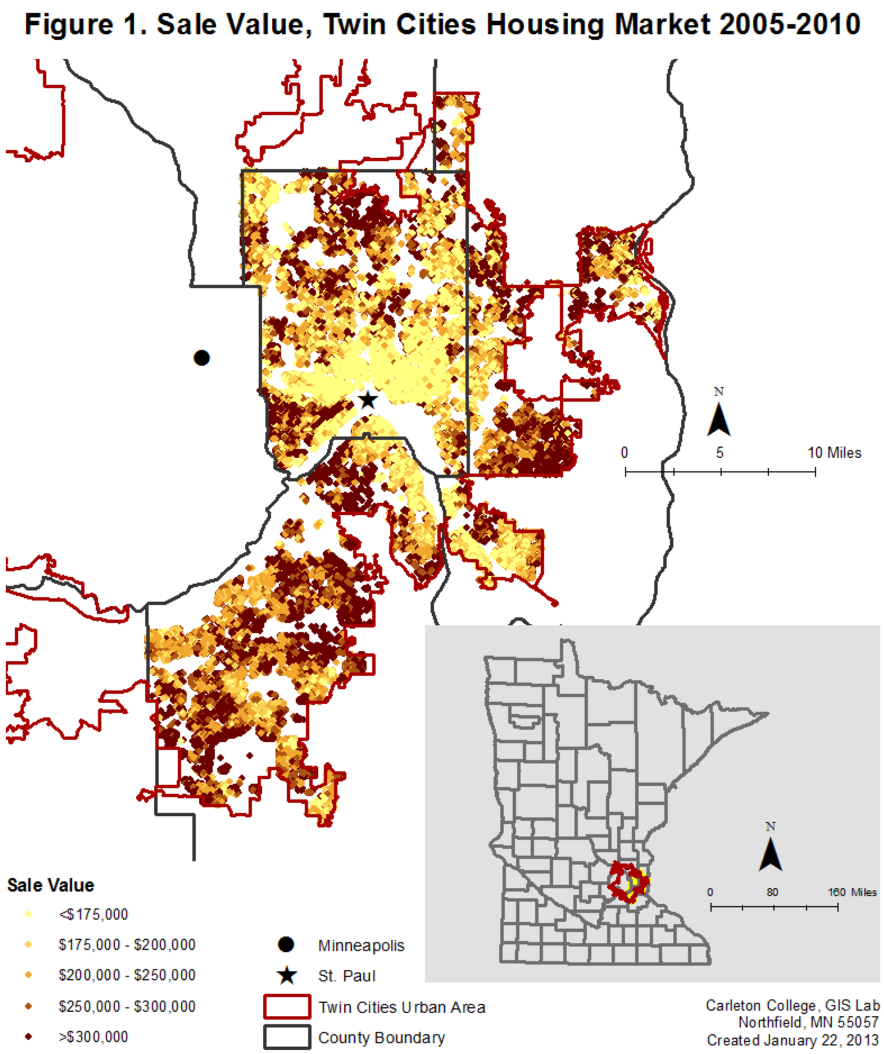

Figure 1: This figure shows the spatial extent of our study area as well as the significant spatial variation in single family house sales prices.

Table 2: Pearson Correlation Coefficient Matrix of Quantitative Variables

	PRICE	*YEAR*	*HOME*	*LOT*	*NOISE*	*INCOME*	*MCA*	*CBD*	*LAKE*	*PARK*	*SHOP*
Sale Price	1										
Year built	0.45	1									
Home size	0.74	0.52	1								
Lot size	0.38	0.47	0.45	1							
Traffic noise	-0.13	-0.19	-0.07	0.00	1						
Median Income	0.52	0.57	0.50	0.41	-0.18	1					
MCA 3rd grade	0.29	0.34	0.28	0.29	-0.04	0.44	1				
Dist. to CBD	0.31	0.58	0.45	0.43	-0.04	0.53	0.35	1			
Dist. to lake	-0.04	-0.17	-0.10	-0.25	-0.03	-0.12	-0.10	-0.09	1		
Dist. to park	0.25	0.42	0.28	0.22	-0.13	0.38	0.22	0.57	0.20	1	
Dist. to shop	0.27	0.37	0.27	0.15	-0.19	0.36	0.18	0.46	0.15	0.38	1

variables for most of our study data because sensitivity analysis conducted in areas with the additional structural variables revealed very similar estimates when the variables were included and excluded. These results are in section 6 of the the paper. Additionally, other hedonic work has been published using a similar set of explanatory variables in this area, see for instance, Sander and Polasky (2009).

2.2 Noise Data

Insert noise data description and methodology here. Nega et al. (2012)

2.3 Other Locational Attributes

A common real estate addage states that the three most important things about real estate are: location, location and location. Knowing where each house is located allows us to also construct a vector of other attributes associated with the sales transaction. For instance, using GIS software we are able to calculate the Euclidean distance to numerous points of interest within the dataset, such as the nearest central business district, shopping centers, parks, and lakes. Additionally, a variable denoting the median household income for the surrounding census tract is created through the use of TIGER shapefiles from the 2010 Census Bureau and data from the 2010 American Community Survey (ACS). Lastly, we associate each transaction with its elementary school and include the average 3rd grade Minnesota Comprehensive Assessment (MCA) score for the local elementary school during the year of purchase. Test scores were obtained from the Minnesota Department of Education website. The school district and elementary school attendance boundary spatial information was taken from the Minnesota Geospatial Information Office Clearinghouse Data Catalog.[1]

[1] `http://www.mngeo.state.mn.us/chouse/data.html`

Table 3: Mean Variable Values Across Time

Variable	2005	2006	2007	2008	2009	2010	All Years
Dependent Variables							
Sales Value ($ thousands)	279	282	275	263	231	233	266
Structural Variables							
Home Size (sq. ft.)	1,737	1,736	1,743	1,813	1,727	1,775	1,750
Lot Size (acres)	0.25	0.25	0.25	0.26	0.25	0.26	0.25
Year Built	1967	1966	1966	1971	1968	1968	1967
Locational Variables							
Traffic noise (dB)	58.5	58.6	58.6	52.5	52.8	52.8	56.4
3rd grade MCA scores	364.8	364.8	365.2	363.5	362.3	366.1	364.5
Median Income ($ thousands)	72.0	72.1	72.3	74.8	72.6	73.2	72.6
Dist. to CBD (km)	14.7	14.6	14.8	15.0	14.3	14.5	14.7
Dist. to lake (km)	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Dist. to shop center (km)	1.9	1.9	1.8	2.0	1.9	1.9	1.9
Dist. to park (km)	2.6	2.6	2.6	2.8	2.6	2.6	2.6
Number of Observations	10,991	8,885	6,549	5,503	5,961	4,206	42,095

2.4 Time

Given that we have data from before and after the recession of 2008-09, we subset the data in Table 3 by year of sale to look for differences. In addition to the noticeable decline in nominal sales values, there is also a drop in the number of sales across years. For example, in 2010 there were only 4,206 property sales within the dataset, a more than 50% reduction from 2005. While the mean sale price declined from roughly $280,000 pre-crash to $230,000 post-crash, most of the structural and neighborhood variables are relatively consistent across years. However, traffic noise is noticeable different in the pre- and post-crash market transactions (the mean drops from 58 dB to 52 dB).

3 Basic Econometric Model

Consistent with past research, this study implements a semi-logarithmic hedonic pricing model. Our aim is to estimate the marginal willingness to pay for different attributes, in particular changes in the traffic noise associated with the house. Equation (1) expresses the basic hedonic model,

$$ln \text{ Sale Price}_i = \beta_0 + \beta_1 \text{Noise}_i + \beta_2 S_i + \beta_3 N_i + \beta_4 T_i + \text{error}_i, \tag{1}$$

where Noise_i is the noise level for house i, S_i is a vector of the house's structural attributes, N_i is a vector of the neighborhood attributes, and T_i is a vector of year fixed effects. This functional form allows us to interpret the regression coefficients as price semi-elasticities for the underlying attributes. For instance, we can interpret the coefficient on noise as the average percentage change in price for a one decibel increase in the traffic noise in our dataset.

Standard urban economic theory[2] predicts that the price of land will vary over space within our dataset to account for locational amenities. As such, we add interaction terms between lot size and distance to the nearest central business district to account for spatial variation in the price of land. The first column in Table 4 shows the results of this regression. The negative coefficient of -0.0027 on the noise variable (a one dB increase in noise decreases house price by 0.27 percent) is in line with the estimates described earlier in the literature. The significant coefficients on the lot size interaction terms suggest that the value of land varies over space. The significant coefficients on the non-linear interaction terms suggests that the nature of the spatial variation in the value of land may be complex.

Table 4 also attempts to begin to understand how the economic recession may have influenced the hedonic function in our study area. The other columns in the table show the regression output from separating the data into sales before and after September of 2008 (the month that the US Federal government took over Fannie Mae and Freddie Mac, Lehman Brothers filed for bankruptcy, and the American International Group (AIG) narrowly avoided bankruptcy only through a $85 billion loan from the US Federal government). The coefficients for some of the hedonic variables of interest (noise, schools, land, for instance) appear different across the two time periods.

Taken in aggregate, the results in Table 4 suggest that the hedonic price function may vary over space and time. While methods exist for parameterizing this variation (such as spatial expansion as suggested by Casetti (1972)), we have no a priori knowledge of how to parameterize the variation over space and time. As such, we turn to a semi-parametric form of hedonic regression, a flexible modeling approach which lets the data reveal how relationships vary, rather than specifying them beforehand.

4 Locally Weighted Regression Model

Locally Weighted Regression (LWR) techniques (also known as Geographically Weighted Regression) are described in detail by Brunsdon et al. (1998). It is a weighted least squares methodology in which regression coefficients are estimated over space as a function of the local data as described in Equation (2),

$$\hat{\beta}_i = (X'W_iX)^{-1}X'W_iY, \tag{2}$$

where X is a $n \times m$ matrix of independent variables, W_i is the $n \times n$ weights matrix, and Y is the $n \times 1$ vector of dependent variable values. The weights matrix, W_i is a diagonal matrix where element w_{jj} denotes the weight that the j^{th} data point will receive in the regression coefficients estimated at location i in

[2]Cite Muth, Mills, etc. here.

Table 4: Basic Regression Results- Dependent Variable = ln Sale Price

	All Sales			**Pre-September 2008**			**Post-September 2008**		
Variable	Coefficient	t-statistic		Coefficient	t-statistic		Coefficient	t-statistic	
(Intercept)	4.07E+00	42.1	***	3.38E+00	29.2	***	4.67E+00	21.83	***
Structural variables									
Year built	1.81E-03	37.6	***	1.94E-03	37.2	***	1.54E-03	14.15	***
log(Finished square feet)	5.52E-01	174.6	***	5.51E-01	162.5	***	5.57E-01	76.48	***
Lot size	1.59E+00	11.5	***	1.77E+00	12.1	***	1.14E+00	3.41	***
Lot size^2	-1.62E+00	-6.5	***	-1.93E+00	-7.2	***	-8.69E-01	-1.47	
Owner Occupancy	2.60E-02	10.3	***	2.47E-02	8.4	***	1.46E-02	2.53	*
Neighborhood variables									
Distance to CBD	5.44E-05	21.4	***	5.38E-05	20.1	***	5.55E-05	8.75	***
Distance to shop	4.00E-05	18.0	***	3.82E-05	16.0	***	4.57E-05	8.94	***
Distance to college	-3.26E-05	-70.7	***	-2.90E-05	-58.7	***	-4.23E-05	-39.39	***
MCA 3rd grade	6.84E-04	6.7	***	1.77E-03	9.5	***	3.87E-04	2.72	**
Median income	2.02E-06	31.4	***	1.79E-06	25.3	***	2.33E-06	15.78	***
Environmental variables									
Traffic Noise	-2.77E-03	-22.7	***	-2.27E-03	-17.8	***	-4.25E-03	-14.01	***
Distance to lake	-7.89E-05	-18.3	***	-8.01E-05	-17.6	***	-8.46E-05	-8.01	***
Distance to park	-8.80E-06	-4.6	***	-8.31E-06	-4.1	***	-8.93E-06	-1.94	.
County									
Dakota county	Omitted								
Ramsey county	3.50E-01	7.5	***	4.06E-01	7.2	***	2.79E-01	3.29	***
Washington county	4.55E-01	50.9	***	4.52E-01	47.4	***	4.52E-01	21.63	***
Sale Year									
2005	Omitted			Omitted					
2006	6.22E-03	2.4	*	7.07E-03	3.0	**			
2007	-2.83E-02	-9.9	***	-2.65E-02	-10.1	***			
2008	-1.46E-01	-47.2	***	-1.35E-01	-44.5	***	Omitted		
2009	-2.29E-01	-73.6	***				-4.73E-02	-4.30	***
2010	-2.61E-01	-76.3	***				-7.73E-02	-6.55	***
Interaction terms									
Land size * Dist. to CBD	-1.36E-04	-7.2	***	-1.57E-04	-7.9	***	-7.90E-05	-1.72	.
Land size^2 * Dist. CBD^2	-3.61E-09	-4.1	***	-4.34E-09	-4.7	***	-1.73E-09	-0.82	
Land size * Dist. to CBD^2	3.29E-09	6.2	***	3.78E-09	6.8	***	2.06E-09	1.59	
Land size^2 * Dist. to CBD	1.56E-04	4.9	***	1.89E-04	5.6	***	6.56E-05	0.85	
R squared	0.751			0.765			0.691		
Adjusted R squared	0.751			0.765			0.689		
sample size	42095			31,253			10,765		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 the dataset. We employ a bi-square weights function and a k-nearest neighbor bandwidth approach as described in equation (3),

$$w_{jj} = \left[1 - \left(\frac{d_{ij}}{d_k}\right)^2\right]^2 \text{ if } d_{ij} < d_{ik}, \text{ otherwise} = 0, \tag{3}$$

where d_{ij} denotes the distance between observations i and j, and d_{ik} is the distance from observation i to the k^{th} nearest observation. This function assigns weights close to 1 for data points near observation i, weights positive but closer to zero for observations farther away, and zero for all $n-k$ observations farther away than the k^{th} nearest observation. We estimate LWR coefficients using bandwidths ranging from as small as 50 observations and as large as 10,000 observations. We choose the LWR bandwidth my minimizing the Generalized Cross Validation score as detailed in equation (4),

$$n * \sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{(n - v_1)^2}, \tag{4}$$

where y_i is the dependent variable value, $\hat{y}_i$ is the predicted dependent variable value for observation i, and v_1 is the "effective number of model parameters." v_1 =tr($\mathbf{S}$), where the matrix $\mathbf{S}$ is the "hat matrix" which maps y onto $\hat{y}$,

$$\hat{y} = \mathbf{S}y,$$

and each row of $\mathbf{S}$, r_i is given by:

$$r_i = X_i(X'W_iX)^{-1}X'W_i.$$

5 Results

Similar to the basic econometric model described in Section 3, the LWR model estimates the logged sales price of a house as a function of structural, locational and temporal variables. In order to account for changing market conditions over time in our data, when estimating LWR coefficients we use only houses sold within the past 12 months. Thus, the coefficient estimates at a particular house are estimated using data from other sales nearby in both time and geography, similar to the methodology used by Nappi-Choulet and Maury (2011).

Figure 2 shows the relationship between the number of nearest houses receiving positive weights in the Locally Weighted Regression and the Generalized Cross Validation score across three different models. The first model simply uses the basic structural characteristics as the explanatory variables (size of the house in square feet, size of the lot in acres, a categorical variable denoting the architectural style of the house, and whether or not the house is owner-occupied). The second model adds locational variables to the previously described model (test scores at the local elementary school, census tract median income, and distances to the central business district, nearest park, nearest shopping center, and nearest lake). The third model adds city fixed effects. Generally speaking, the third model performs slightly better than the second model, which performs better than the first model. The minimum GCV score is obtained using Model (3) and a bandwidth of 200 nearest houses. These results suggest that we are better able to predict house prices using the information contained in nearby (in both time and space) house transactions. They are also roughly similar to the results of Nappi-Choulet and Maury (2011), who reported results using the geographically nearest 300 to 500 observations rather than all across Paris.

Figure 2 also displays the GCV scores for the three models estimated at the global scale. The GCV scores for the global models follow a similar patter as the LWR models, with Model (3) performing better than Model (2) and Model (1) having the highest GCV score. It is also worth noting that the LWR models have significantly lower GCV scores for a range of bandwidths when compared to the global models. These results suggest that it is important to model our data in a spatially explicit manner.

Regression output from a LWR model is voluminous in comparison to a standard regression model, as coefficients are estimated at each location within the dataset. We report our results in a manner similar to Duarte and Tamez (2009). Table 5 displays the results of LWR Model 3 and the similar global regression model. Again, note that the LWR model yields a substantially smaller GCV score (0.027 vs. 0.041).

The first column in Table 5 displays the estimates from a standard semi-log OLS regression with traffice noise, structural characteristics, location-based amenities, and both city and year of sale dummy variables. The coefficient in the second row of the table shows the estimated impact of an additional dB of traffic noise on the natural log of the house sale price to be -.23 percent. This value is within the range of other estimates as described earlier in section 1. Coefficients for the dummy variables (on house style, city, and year of sale) are not reported for the sake of brevity, but are available upon request. The second column in the table reports the estimated coefficient standard errors. With the exception of the coefficient on the distance to the nearest park, all reported coefficients are statistically different from zero at conventional levels.

The third column displays the mean estimated coefficients from the LWR model. In this model coefficients are estimated at each observation within the dataset using the nearest 200 house sales within the past 12 months. The fourth column displays the standard deviation of the estimated coefficients to give a sense of how the coefficients vary when allowed to do so through the use of locally weighted regression techniques.

Column (5) in Table 5 shows the percentage of the local regressions for which the coefficient's p-value is less than 0.1. All regressions contained statistically significant estimates of the impact of increased living space, and a majority contained significant estimates for lot size and year of construction. Roughly half of the regressions estimated an impact of traffic noise statistically different from zero, while approximately one-third of the local regressions estimated statistically significant estimates of the locational variables (school test scores, census tract median income, distance to local amenities).

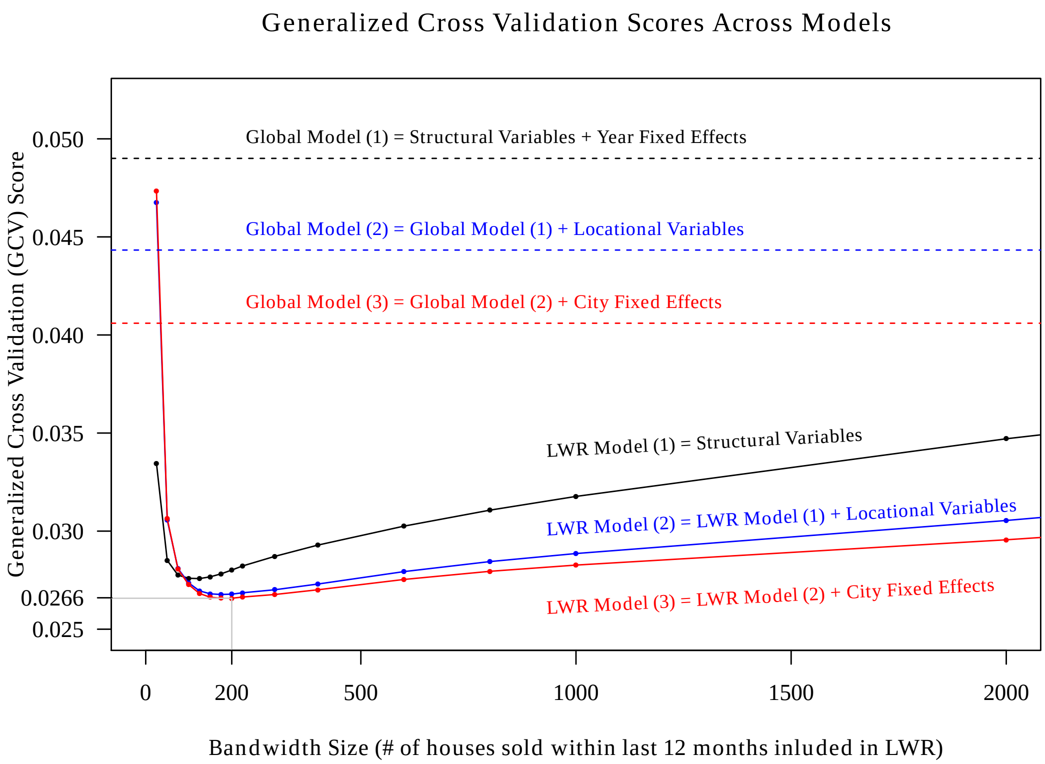

Figure 2: This figure shows the relationship between bandwidth size and the GCV score for three different Locally Weighted Regression (LWR) models. For comparison, the GCV scores for each model when estimated at a global scale are also shown. Note that the LWR models all have significantly smaller GCV scores than the global models. The minimum GCV score is obtained by LWR Model (3) at a bandwidth of 200 nearest houses.

Table 5: Hedonic Regression Results (dependent variable = ln(Sale Price))

	Global Model		Locally Weighted Regression Model			
	$\hat{\beta}$	$\hat{\sigma}_{\hat{\beta}}$	mean $\hat{\beta}$	$\sigma_{\hat{\beta}}$	(5)	(6)
(Intercept)	8.4E+00	1.3E-01	2.9E+00	1.5E+01	0.64	0.00
Traffic Noise	-2.6E-03	1.6E-04	-2.5E-03	3.0E-03	0.47	0.67
House Size	3.0E-04	2.7E-06	2.6E-04	9.1E-05	1.00	0.00
Lot Size	2.6E-01	1.5E-02	4.0E-01	3.7E-01	0.67	0.00
Year of Construction	1.2E-03	6.5E-05	4.5E-03	5.4E-03	0.79	0.00
Owner Occupancy	2.7E-02	3.2E-03	2.3E-02	6.4E-02	0.37	1.00
Elementary Test Scores	1.5E-03	1.2E-04	2.4E-04	3.1E-02	0.30	0.00
Median Income	3.3E-06	7.7E-08	9.6E-07	8.1E-06	0.35	0.00
distance to CBD	1.3E-05	6.3E-07	1.1E-05	1.0E-04	0.30	0.00
" nearest Park	-3.7E-07	1.1E-06	-1.6E-05	1.3E-04	0.32	0.00
" nearest Lake	2.5E-05	2.1E-06	-2.8E-05	1.4E-04	0.35	0.00
" nearest Shopping Center	-5.3E-06	1.3E-06	1.5E-05	8.8E-05	0.29	0.00
location of sale	city fixed effects		city fixed effects and nearest 200 sales			
timing of sale	year fixed effects		within last 12 months			
GCV score	0.041		0.027			
Moran's I statistic	0.199		0.012			

(5) displays the proportion of $\hat{\beta}_{LWR}$ that can reject the null hypthesis of $\beta = 0$.
(6) displays the proportion of Monte Carlo simulations with larger standard deviations of the LWR coefficients than $\sigma_{\hat{\beta}}$

The lack of statistical significance for many of the local regression coefficients could be due to multiple factors. First, because we are estimating local regressions with only the nearest 200 data points, rather than tens of thousands sales as in the global model, we might be estimating the coefficients with less precision and are unable to differentiate the coefficients from zero. Alternatively, the coefficients themselves might vary, sometimes being close to zero and other times not. This variation might be due to random chance (after all, we have estimated tens of thousands of regressions), or it could be the result of true spatial non-stationarity on the part of the coefficients.

5.1 Does the Hedonic Function Vary over Space?

We believe the use of a LWR estimation procedure is appropriate after conducting two different Monte Carlo simulations with the data. In each simulation we resampled (without replacement) the location of each house in the study and then re-estimated the LWR model with the resampled data. In the first simulation we estimated the LWR model using bandwidths ranging from the nearest 50 to 10,000 sales as before. In 100 consecutive simulations, the smallest Generalized Cross-Validation score was obtained at the largest bandwidth. The smallest of these GCV scores was 0.041, just as was the case with the global (non-LWR) model. In other words, the LWR model never performed better than the global model when the locations of the data were randomized. This seems to be strong evidence that the increased ability to predict our house prices with a local model is not due to chance.

Second, we ran a Monte Carlo simulation as described in Fotheringham et al. (2002) and performed in a similar context by Duarte and Tamez (2009) and Cho et al. (2009). The simulation is as follows:

1. Sample with replacement the location of each data point within the study. (just as in the previous simulation)
2. Estimate a Locally Weighted Regression model on the sampled data *using the optimal bandwidth calculated in original LWR model.*
3. Calculate the mean coefficient estimates as well as the standard deviation of the estimates from the sampled data LWR estimates.
4. Repeat Steps 1.–3. M times.

After $M = 100$ iterations of this simulation, the minimum GCV score for these resampled LWR models is 0.060, substantially larger than all models (global and local) with the actual data. Again, this seems to suggest that our data exhibit spatial non-stationarity in the regression coefficients.

We are also interested in the standard deviation of our estimated LWR coefficients (the fourth column in Table 5 as compared to the distribution of simulated standard deviations. If the standard deviation of the LWR estimates with the true data is larger than the standard deviations from the simulations, this is interpretted as evidence of spatial non-stationarity in the coefficients.

The intuition is as follows. Imagine that the coefficients are constant across space. Then, running local regressions will yield different coefficient estimates across space, not because the coefficients are truly different over space, but purely due to chance and noise in the data. Resampling the location of the data and estimating LWR models repeatedly will generate a distribution of how much the coefficients will vary across space due to chance. However, if the standard deviation of the LWR coefficients is larger than the standard deviations seen from the resampled data, this is consistent with the hypothesis that the coefficients vary over space more than they would due to random chance and noise within the data.

The results of the second Monte Carlo simulation are presented in column (6) of Table 5 and in Figure 3. Column (6) of Table 5 displays the proportion of Monte Carlo replications that resulted in larger standard deviations of the LWR coefficient estimates based. For most variables, we never obtained a larger standard deviation of LWR coefficients in the simulation than we obtained from the LWR model with the actual data. However, the values associated with the traffic noise and owner occupancy variables suggest that the variation in LWR coefficients for these variables in our data may have been due to random chance. These coefficients may in fact be stationary across the study area, while the other coefficients appear to be non-stationary. Future work may want to estimate a mixed-LWR model in which the coefficients on traffic noise and owner occupancy remain stationary while the other coefficients are allowed to vary over space. See Fotheringham et al. (2002) for a description of a mixed-LWR estimation algorithm.

Figure 3 visually shows the results presented in column (6) of Table 5. Each subfigure displays the distribution of LWR coefficient standard deviations obtained from the 100 iterations of the Monte Carlo simulation and the standard deviation obtained with the actual data (in red). Note that it is the case that the standard deviations obtained with the actual data are well to the right of the distribution obtained in the simulation for most of the variables.

In addition to the LWR model with a bandwidth of the nearest 200 houses and a temporal lag of 12 months performing better than other bandwidths and model specifications in terms of the smallest GCV score, it also exhibits far less spatial autocorrelation within the model residuals. We calculate the Morans I statistic to be 0.199 for the global model while our prefered LWR specification reduces the Moran's I statistic to 0.012, a nearly twenty-fold reduction.

5.2 Does the Impact of Traffic Noise Vary over Time?

Little work has been done estimating the changes in hedonic environmental (dis)amenities like noise over time. An event-study literature exists using temporal variation in noise, especially noise from airports (see, for example, Cohen and Coughlin (2009)). However, these articles usually use the change in noise over time to identify the hedonic price of noise, rather than ask whether the hedonic price has changed.

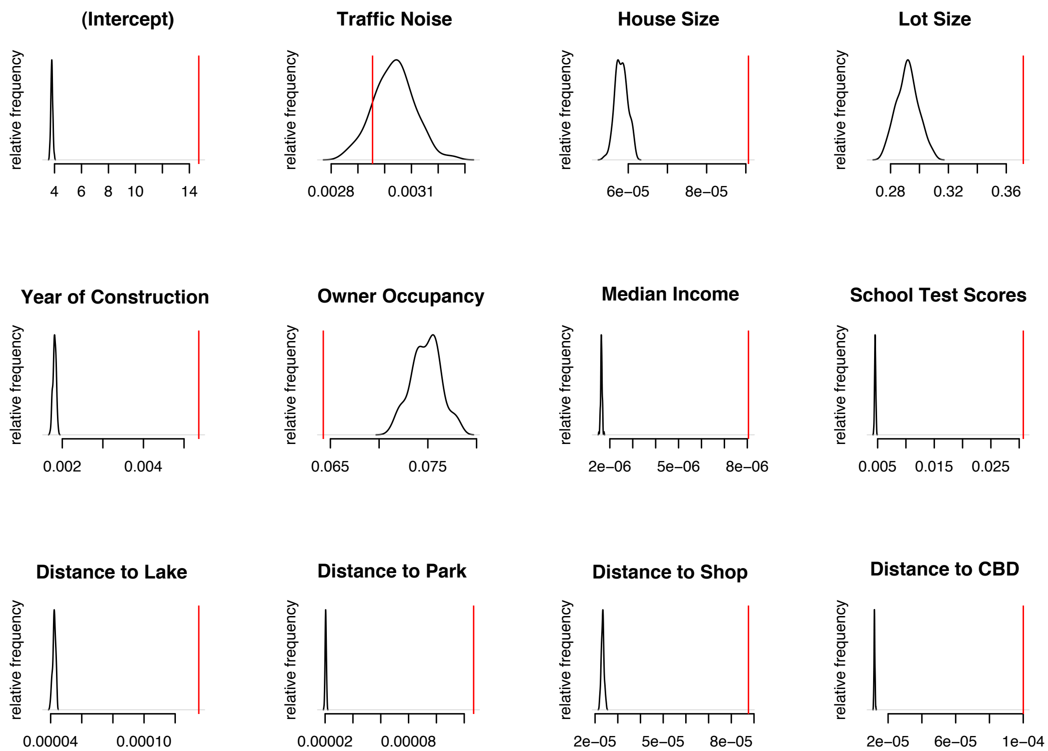

Figure 3: The Monte Carlo Simulation Distribution of LWR Coefficient Standard Deviations and True Coefficient Standard Deviations.

Wilhelmsson (2000) is one example of research asking whether the implicit price of noise has changed over time. The study splits the 290 Swedish house sales into two separate time periods: 1986-89 and 1990-95. A Chow test revealed that the impact of traffic noise was different in the earlier vs. later time period, with noise being considered more of a nuisance in the suburb of Stockholm during the 1990s.

Cho et al. (2011) is a recent example of work asking whether the hedonic regression coefficients change over time in the Nashville, TN area. In particular, the study uses a repeat sales methodology to estimate changes in the implicit prices of hedonic characteristics as a result of the Great Recession of 2008-09 in the United States. The work finds that environmental amenities matter less in sales during the year 2008 when compared to pre-recession sales.

We know of no published research that shows how the hedonic price estimates of traffic noise vary over time surrounding the Great Recession. In this section we seek to compare consumers' willingness to pay for reduced traffic noise exposure before and after the recession. Specifically, we create two new variables, "`months since September 2008`" and a dummy variable, "`post`," for all sales after September 2008. We then regress our LWR coefficients on these two variables and their interaction term as shown in equation (5).

$$\hat{\beta}_{LWR} = \alpha_0 + \alpha_1\texttt{*month} + \alpha_2\texttt{*post} + \alpha_3\texttt{*post*month} + \epsilon \tag{5}$$

The results of this regression will help us understand the patterns of change (if any) in the marginal willingness to pay to avoid traffic noise. In particular, $\hat{\alpha_1} \neq 0$ suggests a linear temporal trend in the implicit price of traffic noise, $\hat{\alpha}_2 \neq 0$ suggests that there was a structural break in the LWR coefficients post September 2008, and $\hat{\alpha}_3 \neq 0$ implies a difference in the temporal trend of the LWR coefficients before and after September 2008. The linear regression results presented in Table 6 show that there is a significant negative shock to the estimated LWR coefficients for traffic noise after September 2008. Additionally, the coefficients also begin to show a significant negative monthly trend after September 2008.

The results in Table 6 contradict the a priori expectation that environmental amenities "matter" less during and after the Great Recession. In fact, these results suggest that the penalty for homes exposed to higher levels of traffic noise increased. That is, the negative hedonic coefficients on traffic noise got more negative after September 2008. It should be noted that this simple model can explain almost 10 percent of the variation in traffic noise LWR coefficient estimates. Future work should investigate how other hedonic coefficients changed during and after the bursting of the housing bubble and resulting economic recession.

5.3 Does the Impact of Traffic Noise Vary Non-linearly?

Some researchers, such as Theebe (2004) conclude that the impact of traffic noise is non-linear in the level of the noise, while others, such as Huang and

Table 6: Regression Results: Dependent Variable = Traffic Noise LWR Coefficients

```
(Intercept) -1.781e-03  4.100e-05 -43.436  < 2e-16 ***
Month        6.982e-06  2.005e-06   3.482 0.000498 ***
Post        -7.998e-04  7.209e-05 -11.094  < 2e-16 ***
Month:Post  -8.868e-05  4.484e-06 -19.776  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.002814 on 31744 degrees of freedom
Multiple R-squared: 0.09289,  Adjusted R-squared: 0.09281
F-statistic:  1084 on 3 and 31744 DF,  p-value: < 2.2e-16
```

Table 7: LWR Coefficients vs. Noise Levels

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.748e-03  1.314e-04 -13.301   <2e-16 ***
Noise       -2.760e-06  2.267e-06  -1.218    0.223
Post        -3.553e-03  2.198e-04 -16.165   <2e-16 ***
Noise:Post   3.386e-05  4.006e-06   8.453   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.002829 on 31744 degrees of freedom
Multiple R-squared: 0.08329,  Adjusted R-squared: 0.08321
F-statistic: 961.4 on 3 and 31744 DF,  p-value: < 2.2e-16
```

Palmquist (2001) report a constant impact per decibel increase. In this section we compare the mean LWR noise coefficient estimates across noise levels to look for non-linearities in the impact of traffic noise.

Table 7 estimates the predicted LWR noise coefficient as a linear function of the noise level in decibels and whether or not the sale took place before or after September 2008. The table reveals no statistically measurable linear relationship between the noise coefficient for sales before September 2008 and the level of the noise (that is, the marginal effect is no different at low levels of noise than at high levels of noise). The positive and significant interaction term suggests that the impact may be different for high vs. low levels of noise post-September 2008, but the difference in the predicted marginal effects is so small, that it works about to be a difference of roughly $25 for a $300,000 house at 50 vs. 75 dB.

Figure 4 visually displays the estimated relationship between the level of

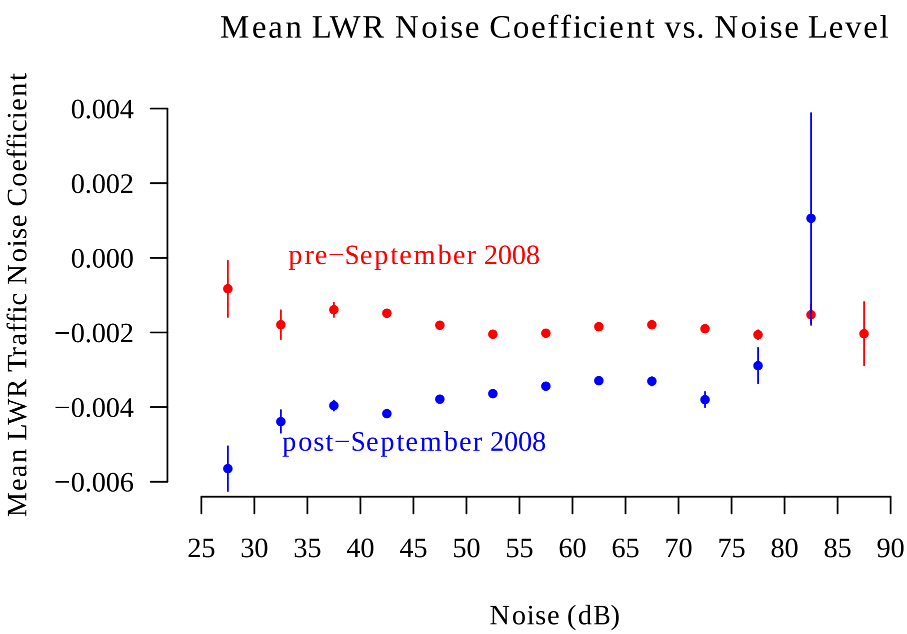

Figure 4: The mean LWR traffic noise coefficient by 5 decibel category and pre- vs. post-September 2008. The vertical lines denote the standard error of the estimated mean by time and level category. Note that over 99 percent of our data have traffic noise levels between 35 and 80 dB and no houses were sold post-September 2008 with noise levels above 85 dB.

traffic noise, time, and the estimated LWR traffic noise coefficients. We categorize the traffic noise data in 5 dB wide bins like Theebe (2004) and then estimate the mean LWR traffic noise coefficients in each category while also controlling for whether the sales took place pre- or post-September 2008. We see little to suggest any meaningful non-linearities in the effect of traffic noise on house prices once we have controlled for the timing of the sales.

6 Discussion

Our work suffers from at least two noticeable weaknesses. First, we are missing some important structural variables that are commonly used in the real estate hedonic literature. To the extent that structural variables like the number of bedrooms, bathrooms, garage size, or construction quality covary with other variables in our dataset, our regression coefficients will suffer from omitted variable bias.

We are somewhat comforted, however, because we do have some additional variables for a subset of our data and their inclusion does not change our results. We have variables for the number of bedrooms, number of bathrooms, and garage size for houses located within Dakota County. Our analysis (Models 1, 2, and 3 described earlier) was repeated for sales located within Dakota County and the results were compared to the estimates obtained from the same analysis when these additional structural variables were omitted. We found strikingly similar estimates of the impact of traffic noise for these data, regardless of whether or not the additional structural variables were included.

Table 8: Mean (Standard Deviation) LWR Traffic Noise Coefficient Estimates in Dakota County with and without Additional Structural Control Variables in Regression.

	Additional structural variables included in regression?	
Locally Weighted Regression Model	no	yes
Model (1) = Structural Variables	-0.00149	-0.00147
	(0.00226)	(0.00309)
Model (2) = Model (1) + Locational Variables	-0.00136	-0.00140
	(0.00184)	(0.00218)
Model (3) = Model (2) + City Fixed Effects	-0.00140	-0.00137
	(0.00219)	(0.00176)

Table 8 shows that the traffic noise coefficients estimated by our LWR models are similar regardless of whether or not the additional structural variables were included as control variables in the regressions. Welch two-sample t-tests fail to reject the null hypothesis of zero mean differences across our three models (comparing LWR estimates with vs. without the additional structural variables included in the model). Paired t-tests find differences for Model 2 and 3, but the estimated differences are zero to four decimal places and in one case the noise impact estimates with the additional variables are slightly larger and in the other case they are slightly smaller. We also conducted simple linear regressions of the traffic noise coefficients without the additional structural variables on the noise coefficients obtained with the additional structural variables. In all three cases the intercept estimates were close zero and the slope terms almost exactly equal to one. Each regression also had R^2 values over 0.7, suggesting that the coefficient estimates obtained with the additional structural explanatory variables are similar to the results obtained without the explanatory variables for Dakota County. We have no reason to believe that our results would be different in the other two counties (Ramsey and Washington) in our study.

The timing of our independent variable collection is also a potential problem. While our house sales data are collected over the course of six years, some of our other variables were collected at specific points in time and assumed to be constant over the study period. In particular, the traffic noise estimates taken from Nega et al. (2012) are for the year 2007. To the extent that traffic flows and composition signicantly changed over time, our traffic noise variables may be inaccurate for those time periods. For instance, the US Department of Transportation reports that total vehicle miles travelled decreased by almost 4 percent year-on-year during the Great Recession.[3] Thus, our noise estimates may be inaccurate for later time periods and this may bias our estimates of the impact of noise on house values. Future work may seek to better estimate the impact of traffic noise over time by generating noise data to reflect changing traffic patterns.

7 Conclusion

We estimated the impact of traffic noise on housing prices using Locally Weighted Regression techniques in the St. Paul, MN metropolitan area. Specifically, we report estimates obtained from regressing the logged sales price on traffic noise and other independent variables at each house within our dataset using only information contained in neighboring houses sold within the last 12 months. We find strong evidence that the hedonic function in our study area differs over space and time. Local regressions using only information contained in the nearest 200 houses generate a substantially smaller Generalized Cross Validation score (0.266) compared to the same model assuming geographical stationarity in the regression coefficients (0.401).

Monte Carlo simulations suggest that the better goodness-of-fit provided by the local models are not likely due to chance and many hedonic impact estimates vary over space within our study area. When the location of our data was randomly assigned and our LWR model was re-estimated, in 100 consecutive simulations the smallest GCV score was obtained when the data were analyzed at a global level rather than local. Additionally, re-estimating the LWR model using a local bandwidth of 200 nearest house sales when the location was randomly assigned yielded substantially smaller standard deviations (i.e. less spatial variation) for the majority of our regression coefficients. This larger variation within our actual location-based data suggests spatial non-stationarity in the regression coefficients. That is, after trying thousands of different combinations of varying levels of local analysis with the spatially redistributed data, we never came close to estimating our observed housing sales prices as well as we can with the local analysis on the actual data.

The impact of traffic noise was one of the few regression coefficients for which we did not find strong evidence of spatial non-stationarity. Such results are contradictory to those presented by Duarte and Tamez (2009) and Hughes Jr. and Sirmans (1992), which found evidence of spatial non-stationarity in the impact of traffic noise in their respective study areas (Barcelona, Spain and Baton Rouge, Louisiana). Additionally, whereas some researchers such as Theebe (2004) suggest that the impact of traffic noise varies with the level of traffic noise, we find little evidence to suggest that the impact of increased traffic

[3]See for instance, `http://www.fhwa.dot.gov/ohim/tvtw/08dectvt`.

noise is different when starting at a low or high noise level. We do, however, find significant temporal variation in the impact of traffic noise in our data. The estimated impact of one additional decibel of traffic noise is a 0.19 percent reduction on average in the sale price of houses before September 2008, vs. 0.37 percent after September 2008.

As mentioned previously, there are multiple areas in which this paper suggests future work on the topic. We found evidence to suggest that, while the locally weighted regression and its associated spatial non-stationarity of the hedonic function are warranted, not all coefficients seem to vary over space. A "mixed" locally weighted regression, in which some variable coefficients vary over space while others are constant, could allow future researchers to more precisely estimate the marginal implicit prices in the hedonic analysis. Second, given our findings of temporal variation in the hedonic function, future researchers may strive to construct datasets reflecting changing economic conditions to estimate the hedonic relationship between housing prices and its constituent characteristics. Lastly, future research should consider applying the findings of Carruthers and Clark (2010), which suggest that locally weighted regressions may allow researchers to more accurately identify the second-stage hedonic functions and thereby better understand the demand for environmental and other non-market goods. Such results will be valuable in future cost-benefit analysis and policy discussions.

References

Ian Bateman, Brett Day, Iain Lake, and Andrew Lovett. The Effect of Road Traffic on Residential Property Values : A Literature Review and Hedonic Pricing Study. Technical Report January, 2001.

Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted regression-modelling spatial non-stationarity. *Journal of the Royal Statistical Society Series D The Statistician*, 47(3):431–443, 1998. ISSN 00390526. doi: 10.1111/1467-9884.00145. URL `http://www.jstor.org/stable/2988625`.

John I. Carruthers and David E. Clark. Valuing Environmental Quality: a Space-Based Strategy. *Journal of Regional Science*, 50(4):801–832, April 2010. ISSN 00224146. doi: 10.1111/j.1467-9787.2010.00662.x. URL `http://doi.wiley.com/10.1111/j.1467-9787.2010.00662.x`.

Emilio Casetti. Generating Models by the Expansion Method: Applications to Geographical Research. *Geographical Analysis*, 4(1):81–91, 1972.

Seong-Hoon Cho, D Christopher, M William, and Seung Gyu Kim. Spatial and Temporal Variation in the Housing Market Values of Lot Size and Open Space. *Land Economics*, 85(1):51–73, 2009.

Seong-Hoon Cho, Seung Gyu Kim, and Roland K. Roberts. Values of environmental landscape amenities during the 20002006 real estate boom and subsequent 2008 recession. *Journal of Environmental Planning and Management*, 54(1):71–91, January 2011. ISSN 0964-0568. doi: 10.1080/09640568.2010.502760. URL `http://www.tandfonline.com/doi/abs/10.1080/09640568.2010.502760`.

Jeffrey P. Cohen and Cletus C. Coughlin. Changing Noise Levels and Housing Prices Near the Atlanta Airport. *Growth and Change*, 40(2):287–313, 2009.

Carlos Marmolejo Duarte and Carlos Gonzalez Tamez. Does Noise Have a Stationary Impact on Residential Values? *Journal of European Real Estate Research*, 2(3):259–279, 2009.

A. Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically Weighted Regression: the analysis of spatially varying relationships*. John Wiley & Sons, West Sussex, England, 2002.

Ju-Chin Huang and Raymond B Palmquist. Environmental Conditions, Reservation Prices, and Time on the Market for Housing. *Journal of Real Estate Finance and Economics*, 22(2-3):203–219, 2001.

William T. Hughes Jr. and C.F. Sirmans. Traffic Externalities and Single-Family Housing Prices. *Journal of Regional Science*, 32(4):487–500, 1992.

Ingrid Nappi-Choulet and Tristan-Pierre Maury. A Spatial and Temporal Autoregressive Local Estimation for the Paris Housing Market. *Journal of Regional Science*, 51(4):732–750, October 2011. ISSN 00224146. doi: 10.1111/j.1467-9787.2011.00713.x. URL `http://doi.wiley.com/10.1111/j.1467-9787.2011.00713.x`.

Stale Navrud. The State-Of-The-Art on Economic Valuation of Noise: Final Report to European Commission DG Environment. Technical report, 2002.

Tsegaye Nega, Carl Smith, James Bethune, and Wei-Hsin Fu. An analysis of landscape penetration by road infrastructure and traffic noise. *Computers, Environment and Urban Systems*, 36(3):245–256, May 2012. ISSN 01989715. doi: 10.1016/j.compenvurbsys.2011.09.001. URL `http://linkinghub.elsevier.com/retrieve/pii/S0198971511000895`.

Heather A. Sander and Stephen Polasky. The value of views and open space: Estimates from a hedonic pricing model for Ramsey County, Minnesota, USA. *Land Use Policy*, 26(3):837–845, July 2009. ISSN 02648377. doi: 10.1016/j.landusepol.2008.10.009. URL `http://linkinghub.elsevier.com/retrieve/pii/S0264837708001324`.

Marcel Theebe. Planes, Tranes, and Automobiles: The Impact of Traffic Noise on House Prices. *Journal of Real Estate Finance and Economics*, 28(2/3):209–234, 2004.

Mats Wilhelmsson. The Impact of Traffic Noise on the Values of Single-Family Houses. *Journal of Environmental Planning and Management*, 43(6):799–815, 2000.