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

June 27, 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

The value of noise pollution is typically estimated through simulated or related markets because no explicit private market for noise abatement exists. Nelson (1982) is perhaps the best known earliest review of 14 housing market studies of the economic impact of traffic noise on property values in Canada and the United States. More recently, a number of reviews have been produced for European government agencies. Bateman et al. (2001) reviewed the literature for the Scottish Government while Navrud (2002) provided a "state of the art" review for the European Commission Environment Directorate General.

The Hedonic Price Method is commonly used to value noise by examining real estate price fluctuations as a function of noise levels. Based on the canonical work by Rosen (1974), researchers typically estimate a hedonic price function specifying the logged sale price of a property as a linear function of the traffic noise exhibited at the house. Such a specification leads to reporting the impact of traffic noise as a percentage decrease in house price per one decibel (dB) increase in traffic noise. According to Bateman et al. (2001), the negative effect of traffic noise has 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.

Nelson (2008) produced an updated review of the traffic noise literature for the edited volume Hedonic Methods in Housing Markets Economics. The chapter describes several areas of continued interest in the empirical literature, most notably: the measure of traffic noise, spatial modeling of the hedonic price function, and housing market dynamics.

1.1 Measure of Traffic Noise

I'm going to need some help writing about noise measures. Perhaps Tsegaye can lend a hand?

Noise data is typically not collected for each every house in a hedonic study. For example, Huang and Palmquist (2001) collect noise data using 128 monitoring stations throughout a suburb of Seattle, WA. Such a strategy forces the researchers to interpolate the data between the monitoring stations in order to associate each house with a noise level. While the researchers do not describe how the interpolation was done, they implicitly admit the potential uncertainty of the method by categorizing traffic noise exposure into 2.5 dB wide categories. Other researchers, such as Theebe (2004) use 5 dB wide bins, and Nelson (2008) even claims that some studies use bins as wide as 10 dB. Such coarse measurements can lead to serious errors in variables and imprecise estimates of the impact of noise.

Other researchers obtain noise levels from external sources without much description of the variable source. For instance, Duarte and Tamez (2009) simply list a "City Council Sonic Map", Blanco and Flindell (2011) cite the Department for Environment Food and Rural Affairs Noise mapping England project. Andersson et al. (2010) state, "Information about noise levels is from a study on the health effects of traffic noise conducted in Lerum [Sweden] in 2004 (Ohrstrom et al. 2005). Separate noise calculations were made for railway and road noise for all the houses in Lerum."

Traffic noise at a location can vary significantly depending on changing traffic conditions. Such variation has led to multiple measures of noise. Some measures eliminate the fluctuations in sound and instead represent the equivalent continuous constant sound pressure at a site. Other indices use noise values that are only exceeded a given percentage of time. For instance, Huang and Palmquist (2001) use dB L_{10} as their noise measure, where L_{10} denotes the value that is exceeded only 10 percent of the time. Still others adjust nighttime noise exposure levels to account for potential increased perceived nuisance. **?** even surveyed inhabitants about their perceptions of the level of noise and their annoyance. The work finds similar results for significant levels of measured noise as compared subjective measures of noise.

What do we have? There is still a lot of work to be done in this area, better understanding the measures others use as well as what we have and how they *compare. I'm surprised at how little information people actually provide about their noise data.*

Spatial resolution of the data. Theebe (2004) had noise levels for square areas of 100 x 100m.

1.2 Spatial Analysis

Advances in Geographic Information System (GIS) technology and digitized datasets have made research easier to conduct over larger geographies, with larger samples, and while explicitly considering the spatial nature of the housing market. Estimation of the hedonic function using Ordinarly Least Squares (OLS) is typically inappropriate because the housing data commonly exhibit spatial autocorrelation. In particular, spatially autocorrelated error terms will lead to inconsistent estimates of OLS regression standard errors and inappropriate measures of statistical inference.

It is now common to use spatial econometric techniques in the hedonic literature, for instance estimating a spatial lag (in which house prices are modeled as a function of neighboring house prices), or a spatial error model (in which the degree of spatial autocorrelation is also estimated). In the noise literature, for example, Theebe (2004) econometrically estimates a spatial error model, while Andersson et al. (2010) uses a spatial lag model.

One common reason for significant spatially autocorrelated error terms is hedonic function misspecification. A portion of the hedonic regression errors may be explained by omitted spatial variables. Thus, nearby houses will tend to have similar spatial variable values and then similar error terms. Second, model misspecification can occur due to spatial non-stationarity of the hedonic function. That is, housing submarkets exist and the impact of changing housing characteristics will vary across submarkets. Failing to properly account for such heterogeneity will lead to spatially autocorrelated error estimates.

Some traffic noise research simply explicitly assumes that their data comprise a single market. For example, Kim et al. (2007) assumes that by restricting analysis to residential properties surrounding the Inner Circular highway in the central area of Seoul from 2002 to 2004, that the price of traffic noise will be constant. Other research, such as Theebe (2004) used political boundaries to segment the market into five separate subarkets in the western part of the Netherlands.

Another strategy is to let the data reveal different relationships within smaller areas of the study. For instance, Day et al. (2007) and Duarte and Tamez (2009) use semi-parametric techniques to allow the hedonic relationship to smoothly vary over space. We use techniques similar to the Geographically Weighted Regression method described in Duarte and Tamez (2009) to allow the hedonic function to vary over space without having a priori conceptions of where the function will be different.

1.3 Temporal Variation

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.

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 estimate regression coefficients over time using a moving window of 12 months in the hedonic analysis over the years 2005 - 2010. Such an analysis allows us to see if and when the implicit price of traffic noise changes within our study timeframe. Our results suggest that the percentage change in house price from an extra decibel of traffic noise actually became more negative after the Great Recession than before.

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 sptially explicit modeling techniques allow us to make our second contribution to the literature. We present one of the first estimates of how the impact of traffic noiseeee varies over space. 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 hedonic price of traffic noise.

2 Data

The 2010 US Census lists the population of the Minneapolis and St. Paul Metropolitan area in Minnesota 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

Table 1: Simple Correlation 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

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 structural and locational variables commonly used in the hedonic literature.

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 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. 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

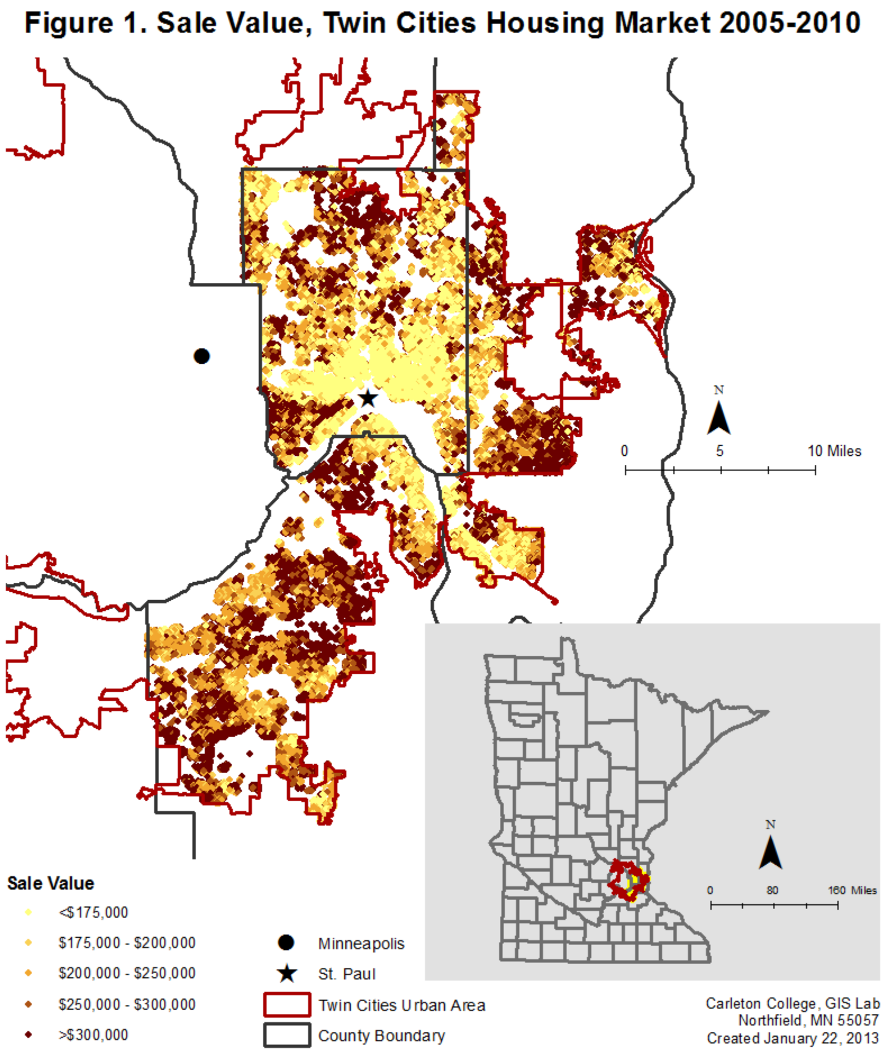

Figure 1: This figure shows the spatial extent of our study area as well as the significant spatial variation in single family house sales prices.

Table 2: 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

the Euclidean distance to numerous points of interest within the dataset, such as the nearest central business district, shopping centers, parks, and lakes.

We also 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. Lastly, 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).

2.4 Time

Given that we have data from before and after the recession of 2008-09, we subset the data in Table 2 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 displays a noticeable difference 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.

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

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. We can interpret the regression model coefficients as the price semi-elasticities of the underlying attributes. For instance, we can interpret the coefficient on noise as the percentage increase in price for a one decibel increase in the traffic noise associated with the transaction in our dataset.

Standard urban economic theory[1] 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 3 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 3 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 3 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 (1974)), 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.

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

Table 3: 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

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 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.

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 a bandwidth of 200 nearest houses and Model (3).

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 4 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 4 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.

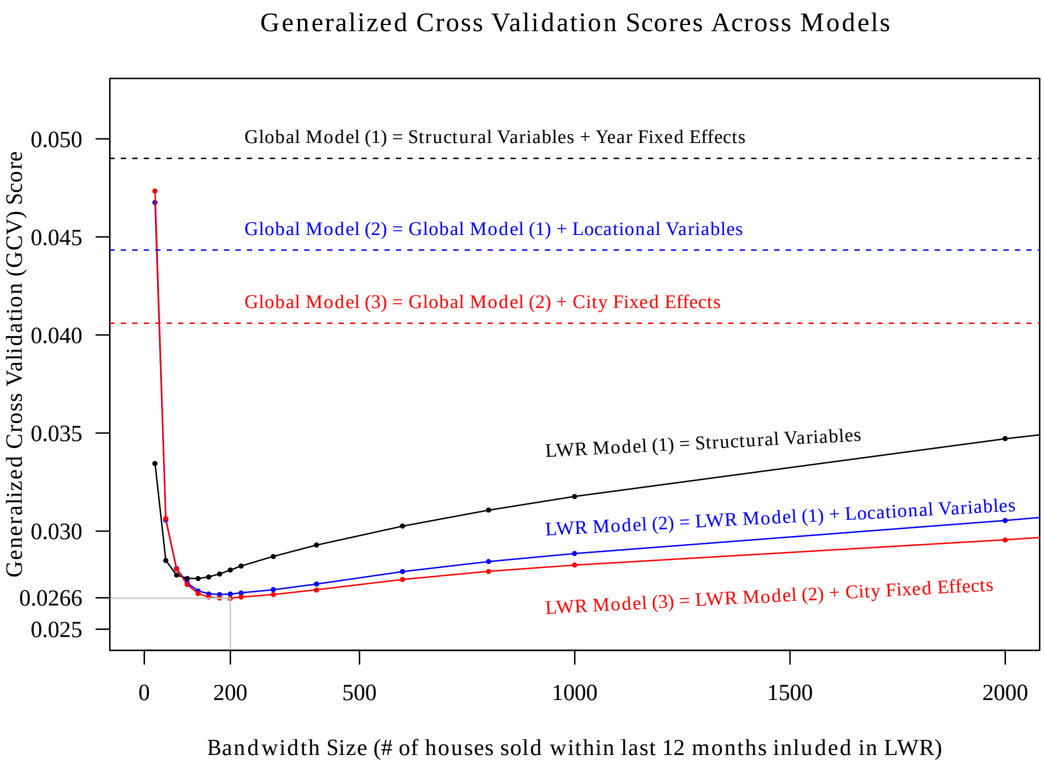

Figure 2: This figure shows the relationship between bandwidth size and the GCV score for three different Locally Weighted Regression (LWR) models. For comparison, the GCV scores for each model when estimated at a global scale are also shown. Note that the LWR models all have significantly smaller GCV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

Table 4: 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 $\sigma_{\hat{\beta}}$

Column (5) in Table 4 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).

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 over 30,000 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 Space Matter?

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 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). 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 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 4 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 4. Column (6) displays the probability of obtaining a larger standard deviation of LWR coefficient estimates based on the 100 simulations performed. For most variables, we never obtained a larger coefficient standard deviation 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 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 4 another way. 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 substantially larger than those 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 Do the Regression Coefficients Vary over Time

We know that house prices fell dramatically during and after the Great Recession of 2008-09. We know less about the patterns of change in the hedonic price functions during and after the fall in prices. For instance, did all house prices simply drop (mostly affecting the intercept term of the hedonic function), or did the implicit price of structural characteristics like the size of the house or its lot decrease, or perhaps the value of environmental amenities disproportionately dropped? This section presents some of the first evidence of how the hedonic price function for housing changed over time surrounding the Great Recession of 2008-09. We ask whether the LWR coefficients exhibit a temporal pattern.

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).

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

The results of these regressions will help us understand the patterns of change in the hedonic price of different characteristics. In particular, significant values

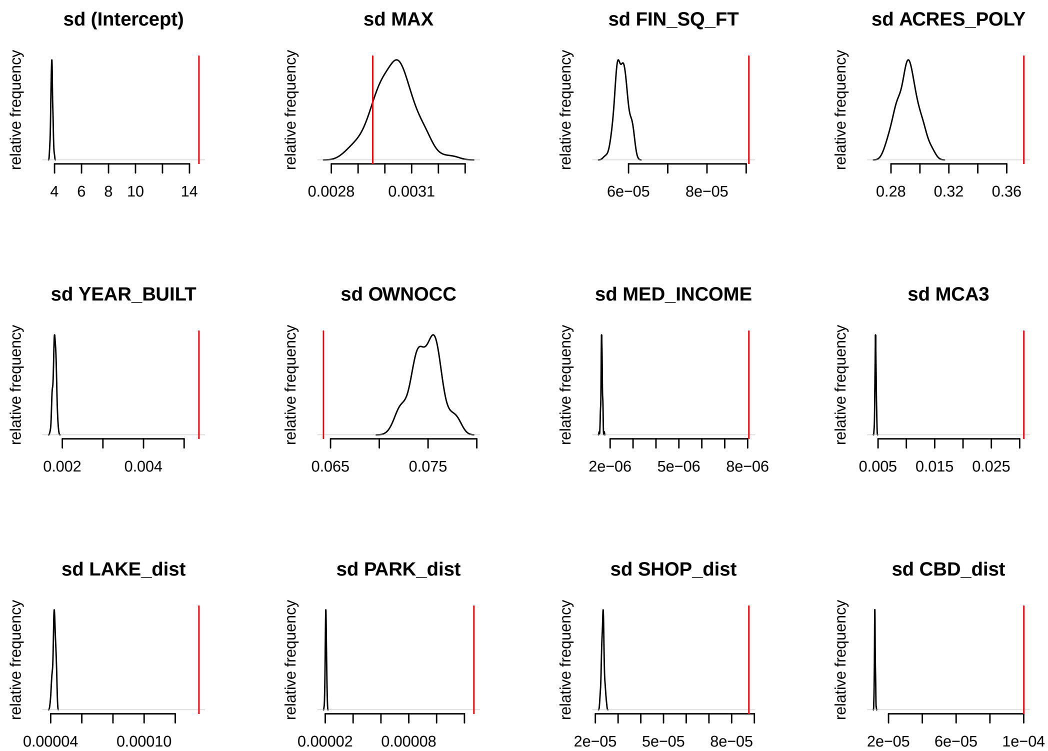

Figure 3: The Monte Carlo Simulation Distribution of LWR Coefficient Standard Deviations and True Coefficient Standard Deviations.

for $\hat{\alpha}_2$ suggest that there was a structural break in the LWR coefficients post September 2008, and significant values for $\hat{\alpha}_3$ will imply a difference in the temporal trend of the LWR coefficients.

Note that the tables have not yet been formatted.

5.2.1 Traffic Noise

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 show that accounting for the month of the house sale significantly predicts to the estimated LWR coefficients for traffic noise. The linear regression results presented in Table 5 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 a significant negative trend.

Table 5: 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
```

The results in Table 5 contradict the a priori assumption that environmental amenities will “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 our model can explain almost 10 percent of the variation in traffic noise LWR coefficient estimates.

5.2.2 House Size

In this section we show how the hedonic coefficients on house size (as measured in square feet) change over time. Table 6 shows that the LWR coefficients on house size were positive and trending up before September 2008, then the coefficients drop significantly in both absolute value and their time trend. These results are consistent with the belief that the Great Recession would reduce consumer’s marginal willingness to pay for increases in the size of a house. It should be noted that the model presented in Table 6 explains less than one percent of the variation in our house size LWR coefficients.

Table 6: Regression Results: Dependent Variable = House Size LWR Coefficients

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.741e-04  1.317e-06 208.097  < 2e-16 ***
month        8.742e-07  6.441e-08  13.573  < 2e-16 ***
Post        -1.678e-05  2.316e-06  -7.245 4.43e-13 ***
month:Post  -5.989e-07  1.440e-07  -4.157 3.23e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 9.04e-05 on 31744 degrees of freedom
Multiple R-squared: 0.006092,  Adjusted R-squared: 0.005998
F-statistic: 64.85 on 3 and 31744 DF,  p-value: < 2.2e-16
```

5.2.3 Lot Size

Table 7 shows interesting results that are qualitatively different from our previous results. The post-September 2008 dummy variable coefficient is only marginally significant and small in absolute value (when compared to the regression Intercept term). There is no discernable change in the temporal trend of the lot size LWR coefficients. The model presented in Table 7 explains less than a half of one percent of the variation in lot size LWR coefficients.

Table 7: Regression Results: Dependent Variable = Lot Size LWR Coefficients

```
(Intercept)  0.3828690  0.0054089  70.785  < 2e-16 ***
month       -0.0015113  0.0002645  -5.713 1.12e-08 ***
Post         0.0190349  0.0095117   2.001   0.0454 *
month:Post  -0.0001358  0.0005916  -0.229   0.8185
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.3713 on 31744 degrees of freedom
Multiple R-squared: 0.002806,  Adjusted R-squared: 0.002711
F-statistic: 29.77 on 3 and 31744 DF,  p-value: < 2.2e-16
```

5.2.4 Intercept

We were also interested in how the LWR regression intercept changed over time. For instance, perhaps the house price drops of the Great Recession were manifested primarily through reduced intercept terms. Table 8 is not consistent with that hypothesis. In fact, our results show that the estimated LWR intercept terms are on average significantly higher post-September 2008.

Table 8: Regression Results: Dependent Variable = LWR Intercept Estimates

```
(Intercept)  1.575488   0.213779   7.370 1.75e-13 ***
month       -0.043836   0.010455  -4.193 2.76e-05 ***
Post         2.845774   0.375939   7.570 3.84e-14 ***
month:Post   0.007397   0.023383   0.316    0.752
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 14.67 on 31744 degrees of freedom
Multiple R-squared: 0.003282,  Adjusted R-squared: 0.003188
F-statistic: 34.84 on 3 and 31744 DF,  p-value: < 2.2e-16
```

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 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 9 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 roughly \$25 difference for a \$300,000 house at 50 vs. 75 dB.

Figure 4 visually displays the estimated relationship between the level of 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

Table 9: 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
```

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. We have the number of bedrooms, bathrooms, and garage area for our houses located within Dakota County. Our analysis (Models 1, 2, and 3 described earlier) was repeated on this geographic subset of the data 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.

Table 10 shows that the traffic noise coefficients estimated by our LWR models are similar regardless of whether the additional structural variables were included or not. Welch two-sample t-tests fail to reject the null hypothesis of zero means 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 egressions of the 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 with slopes almost exactly equal to one and each also had R^2 values

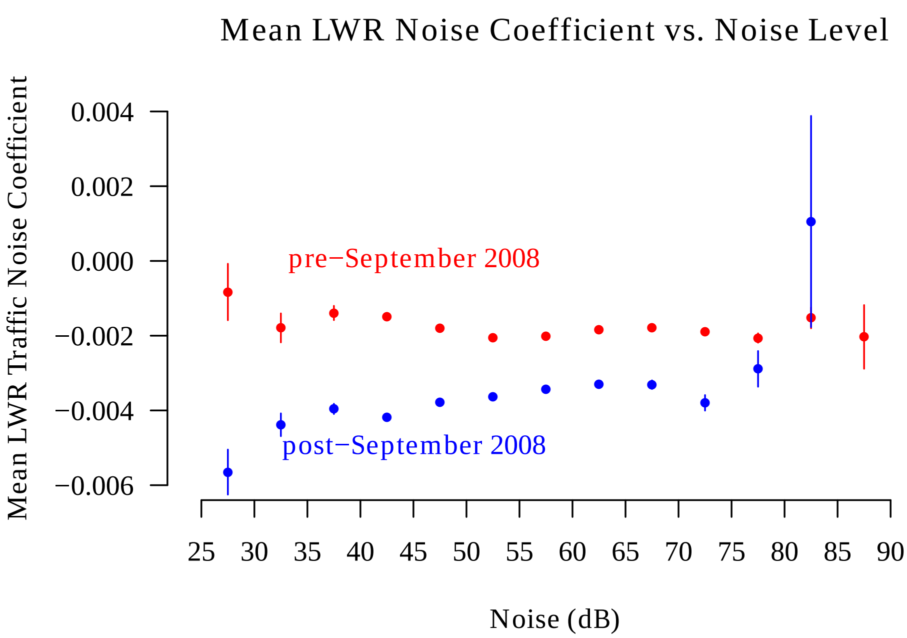

Figure 4: The mean LWR traffic noise coefficient by 5 decibel category and pre- vs. post-September 2008. The vertical lines denote the standard error of the estimated mean by time and level category. Note that over 99 percent of our data have Noise levels between 35 and 80 dB and no houses were sold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

Table 10: Mean and (Standard Deviation) LWR Noise Coefficients in Dakota County both with and without the additional structural variables available only for the county.

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

over 0.7.

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 up to 2 percent year-on-year during the Great Recession. 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 better estimate the impact of traffic noise over time by generating noise data to reflect changing traffic patterns.

What other weaknesses should we discuss?

We didn't explicitly model spatial autocorrelation in the error terms.

We might want to further reduce the number of parameters to estimate through the use of either mixed-LWR or running separate regressions for submarkets determined through hierarchical clustering like Day (2003).

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 estimate linear regressions at each house within our dataset using only information contained in geographically nearby 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 due to chance and that 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 for the majority of our regression coefficients. This larger variation within our true 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.

Contrary to the previous results presented by Duarte and Tamez (2009) and Theebe (2004), we find little evidence to suggest that the impact of traffic noise varies over space or by level of noise within our study area. The traffic noise coefficient was one of only two variables with regression coefficient standard deviations smaller than or similar to the simulated distributions obtained from Monte Carlo experiments. 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 in the sale price of houses before September 2008, vs. 0.37 percent after September 2008. Lastly, after controlling for the Great Recession, we find no significant differences in the impact of traffic noise across the range of commonly seen noise levels (50-70 dB).

References

Henrik Andersson, Lina Jonsson, and Mikael Ögren. Property Prices and Exposure to Multiple Noise Sources: Hedonic Regression with Road and Railway Noise. *Environmental and Resource Economics*, 45(1):73–89, 2010. ISSN 09246460. doi: 10.1007/s10640-009-9306-4. URL http://dx.doi.org/10.1007/s10640-009-9306-4.

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.

J C Blanco and I H Flindell. Property prices in urban areas affected by road traffic noise. *Applied Acoustics*, 72(4):133–141, 2011. URL http://eprints.soton.ac.uk/191013/.

Chris Brunsdon, Stewart Fotheringham, Martin Charlton, and Chris Brunsdont. 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.

Emilio Casetti. Generating Models by the Expansion Method: Applications to Geographical Research. *Geographical Analysis*, 4(1):81–91, 1974.

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.

B Day, I Bateman, and I Lake. Beyond implicit prices: recovering theoretically consistent and transferable values for noise avoidance from a hedonic property price model. *Environmental and Resource Economics*, 37(1):211–232, 2007.

Brett Day. Submarket Identification in Property Markets: A Hedonic Housing Price Model for Glasgow. 2003.

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.

Kwang Sik Kim, Sung Joong Park, and Young-Jun Kweon. Highway traffic noise effects on land price in an urban area. *Transportation Research Part D: Transport and Environment*, 12(4):275–280, June 2007. ISSN 13619209. doi: 10.1016/j.trd.2007.03.002. URL http://www.elsevier.com/wps/find/journaldescription.cws_home/31153/description#description http://linkinghub.elsevier.com/retrieve/pii/S1361920907000260.

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.

Jon P Nelson. Highway Noise and Property Values: A Survey of Recent Evidence. *Journal of Transport Economics and Policy*, 16(2):117–138, 1982.

Jon P Nelson. Hedonic Property Value Studies of Transportation Noise : Aircraft and Road Traffic. In Andrea Baranzini, Jose Ramirez, Caroline Schaerer, and Philippe Thalmann, editors, *Hedonic Methods in Housing Market Economics*, number October 2007, chapter 3, pages 57–82. Springer, New York, NY, 2008.

Sherwin Rosen. Hedonic Prices and Implicit Markets : Product Differentiation in Pure Competition. *Journal of Political Economy*, 82(1):34–55, 1974.

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.