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

June 24, 2013

1 Introduction

- Introduction to environmental amenities of a property
- Outline how the impact of environmental quality on the value of a property is specific depending on locational factors.
- Intro to traffic noise as an environmental good.
- Brief description regarding the negative impacts of traffic noise (brief summaries of past studies on the topic)
- Outline how traffic noise is measured decibel levels on a logarithmic scale
- Description of the housing market crash in the fall of 2008
- How the crash effected Twin Cities metro real estate market
- Research question: Does traffic noise have a significant impact on properties within the Twin Cities metro real-estate market and if so, does this effect vary across space while accounting for time?

2 Past Research

Given the public policy ramifications of the issue, it is not surprising that multiple reviews of the economic impacts of traffic noise exist. Nelson (1982) is perhaps the best known earliest review of 14 studies 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.

Since no explicit private market for noise abatement exists, the value of noise pollution is typically estimated through simulated or related markets. The Hedonic Price Method is commonly used to value noise by examining real

estate price fluctuations as a function of noise levels. The estimation of the noise impact in the literature is typically reported as a percentage decrease in house price per one decibal (dB) increase in local traffic noise. According to **?**, the negative effect of traffic noise has ranged from 0.08% to 2.22% per dB with a mean around 0.55. Navrud (2002) found an average decrease of 0.64% per dB in a review of 65 noise evaluation studies. Nelson (2008) produced an updated review of the traffic noise literature for the edited volume Hedonic Methods in Housing Markets Economics. Nelson describes several areas of interest in the empirical literature, most notably: spatial modeling of the hedonic price function, long-run vs. short-run housing dynamics, and the appropriate measure of traffic noise. We now explain each of these areas in turn and our work's contribution to the field.

2.1 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. Ordinarly Least Squares estimation of hedonic functions commonly violate standard assumptions of uncorrelated error terms.

? econometrically estimates a spatial error model

Researchers such as , and Andersson et al. (2010) attempt to correct for spatial autocorrelation through the use of spatial

One challenge and opportunity for researchers is to properly define the spatial extent at which to conduct their analysis.

Spatial heterogeneity

Spatial autocorrelation

data collection and spatial modeling easier than in the past. However, the larger sample sizes and spatial extents create issues of heterogeneity as the relationship between characteristics and housing prices may differ across housing submarkets. Day (2003) investigated the impact of traffic noise in Glasgow, Scotland and found four separate submarkets with varying relationships between noise and prices. Interestingly, traffic noise coefficients were estimated to be negative in three out of four submarkets, while the fourth revealed a significantly positive relationship. The positive relationship is hypothesized to be related to increased transportation access and mobility in an outer suburb of the city. In a similar study, Day et. al (2007) found traffic noise estimates to be significantly negative in five of eight submarkets using appriximately 10 thousand residential sales in 1997 in Birmingham, England.

Most researchers who allow the hedonic function to vary over geography do so through explicit parameterization done by the researchers (usually using political boundaries).

Theebe (2004) also used market segmentation to estimate more than 160,000 property sale values in the western part of the Netherlands sold between 1997 and 1999. First, Theebe split the full sample into five parts according to political jurisdictions. Through spatial autoregression techniques, Theebes results

indicated that sales values are unaffected by noise levels below 65 dB. Above 65 dB, NDSI is relatively small, ranging from 0.3 to 0.5% per dB. An analysis of submarket data yielded larger results although these findings are less precise. While Theebe covered a large study area, the noise levels are obtained for square areas of 100 x 100m using the cumulative 24 hours energy level index Leq. This study uses a Leq index at a square area of 10 x 10m, therefore each property has a unique and continuous noise decibel measurement. Consequentially, the analysis on smaller submarket samples will be relatively more precise in terms of traffic noise estimates than Theebe (2004).

While Day (2003) and Theebe (2004) considered the existence of submarkets through market segmentation and used estimators that accounted for spatial autocorrelation, these segmentations are limited by space. By defining clear boundaries of submarkets, Duarte et. al (2009) noted that this clear delimitation does not allow interdependence between submarkets to be considered. Through locally weighted regression (LWR), the authors resolved the issue of spatial autocorrelation and considered the existence of submarkets while providing a soft border between localities. By examining 3,196 appraisals of multifamily dwellings assessed in 2005 in Barcelona, Spain, Duarte et. al investigated a range of neighborhood and environmental attributes. A Monte Carlo test statistically validated the spatial variation of local factors. The results from the LWR indicate that the average NDSI for observations located in the 65-70 dB is negative while average NDSI for observations exposed to 70-75 dB is 0.26%. Finally, dwellings exposed to higher levels of noise (75-80 dB) have an average -0.40% NDSI. The authors explained this paradox by considering areas with intermediate levels of noise have closer access to services and transport within the city without experiencing an extreme noise externality. In this study, a similar implementation of locally weighted regression is utilized in order to resolve for spatial autocorrelation while allowing for interdependency across submarkets. Furthermore, two Monte Carlo tests are utilized to statistically support (1) the use of a non-global model and (2) the spatial variation of estimated parameters.

3 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 the three of the seven counties Dakota, Ramsey and Washington County (see Figure 1). We obtained data from approximately forty thousand 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. See Table 1 for a complete list of the explanatory variables and their descriptive statistics.

3.1 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).

3.2 Noise Data

Insert noise data description and methodology here.

3.3 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 central business districts, 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).

3.4 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 qualified 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).

4 Basic Econometric Model

Consistent with past research, this study implements a standard 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. 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 decibal 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 ameinities. 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 XXX 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 XXX 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 Freddi 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 XXX 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.

5 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 jth 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 kth 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 kth 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:

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

6 Results

Similar to the basic econometric model described in Section 4, 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. We also estimated the models using sales within the past 6 and 24 months.

Figure XXX shows the relationship between the number of nearest neighbors receiving positive weights in the Locally Weighted Regression and the Generalized Cross Validation score.

Regression output from the 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 Marmolejo Duarte and Gonzalez Tamez (2009). Table XXX displays the results of our prefered LWR model and a similar global regression model.

The first column in Table XXX 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 2. 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 third 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. The fourth column 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.

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 (with 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. In 100 consecutive simulations, the smallest Generalized Cross-Validation score was obtained at the largest bandwidth. 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 **?** and performed in a similar context by Marmolejo Duarte and Gonzalez 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. (we set $M = 100$)

5. Compare the LWR coefficient means and standard deviations from the real data to the simulated values.

Of particular interest is the standard deviation of the LWR coefficients for a variable as compared to the distribution of simulated standard deviations. If the standard deviation of the actual LWR estimates 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. 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 columns 5 and 6 of Table XXX. Column 5 displays the standard deviation of the estimated coefficients from the LWR model estimated and presented in column 3. Column 6 displays the probability of obtaining a larger standard deviation of LWR coefficient estimates based on the 100 simulations performed. For most

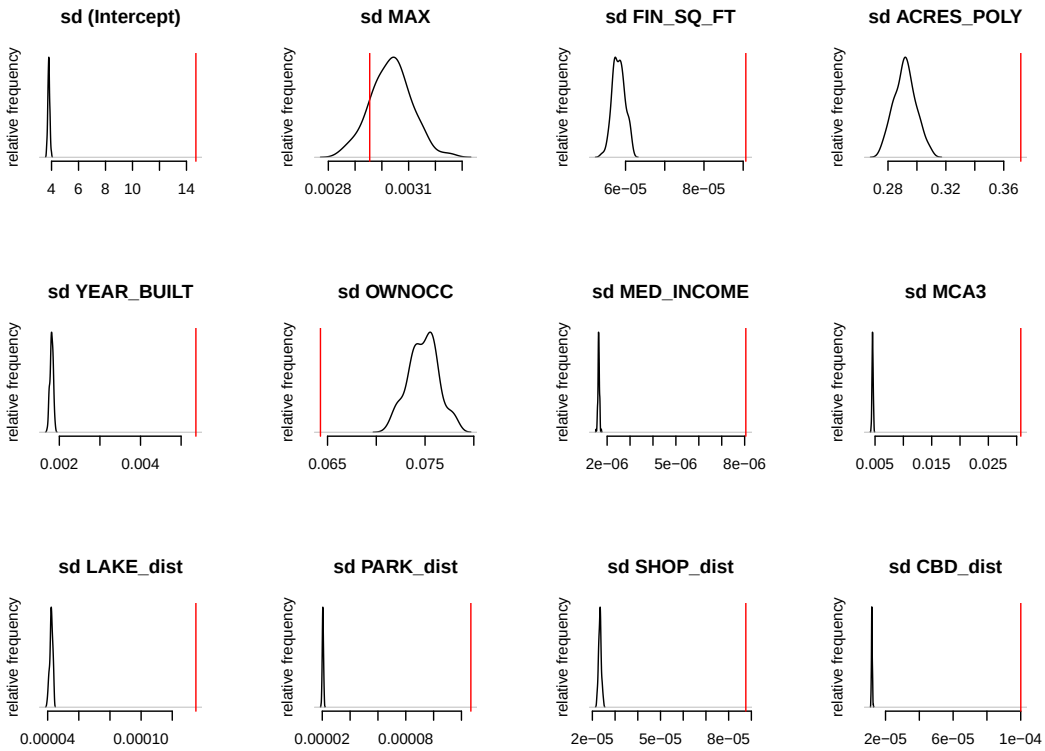

Figure 1: The Monte Carlo Simulation Distribution of LWR Coefficient Standard Deviations and True Coefficient Standard Deviations.

variables, we never obtained a larger coefficient standard deviation in the simulation than we obtained from the LWR model with the true 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 **?** for a description of a mixed-LWR estimation algorithm.

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 on 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.

Coefficients over time.

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, and Iain Lake. 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`.

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

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.

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

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. *Hedonic Methods in Housing Markets Economics*, (October 2007), 2008.

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`.

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.