Research Division
Federal Reserve Bank of St. Louis
Working Paper Series

Semi-Parametric Interpolations of Residential Location Values: Using Housing Price Data to Generate Balanced Panels

Jeffrey P. Cohen
Cletus C. Coughlin
and
John M. Clapp

Working Paper 2014-050A
http://research.stlouisfed.org/wp/2014/2014-050.pdf

December 2014

FEDERAL RESERVE BANK OF ST. LOUIS
Research Division
P.O. Box 442
St. Louis, MO 63166

The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Federal Reserve Bank of St. Louis Working Papers are preliminary materials circulated to stimulate discussion and critical comment. References in publications to Federal Reserve Bank of St. Louis Working Papers (other than an acknowledgment that the writer has had access to unpublished material) should be cleared with the author or authors.

Semi-Parametric Interpolations of Residential Location Values: Using Housing Price Data to Generate Balanced Panels

Jeffrey P. Cohen, Cletus C. Coughlin, and John M. Clapp*

December 2014

Abstract

We estimate location values for single family houses by local polynomial regressions (LPR), a semi-parametric procedure, using a standard housing price and characteristics dataset. As a logical extension of the LPR method, we interpolate land values for every property in every year and validate the accuracy of the interpolated estimates with an out-of-sample forecasting approach using Denver sales during 2003 through 2010. We also compare the LPR and OLS models out-of-sample and determine that the LPR model is more efficient at predicting location values. In a balanced panel application, we use GMM estimation to examine how the location value estimates are affected by airport infrastructure investments.

Keywords: Land Values, Semi-Parametric Estimation, Local Polynomial Regressions, Balanced Panel, Fixed Effects

JEL Classification: C14, R51, R53, H41, H54

*University of Connecticut, Federal Reserve Bank of St. Louis, and University of Connecticut, respectively. The authors appreciate the assistance of Brett Fawley, David Lopez, Diana Cooke, and Lowell Ricketts. Clapp and Cohen acknowledge support from the Center for Real Estate, University of Connecticut. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Introduction

Identically-sized lots and houses in distinct locations in a metropolitan area likely have different market values, a difference largely attributed to the value of location since the structure can be renovated or even rebuilt at a similar cost, regardless of its location. The relatively high variability in land value has been well-known by real estate professional and researchers for many years.[1] However, finding and implementing a theoretically sound and practical method for separating the value of the land (i.e., location) from the value of the housing structure has remained a challenge.[2]

We investigate the separate valuation of residential land and structure using housing price sales data. The methods we develop are important where land values rather than housing prices are required.[3] For example, tax assessors recognize that accurate property valuation must address the very different determinants of location value versus reproducible structural characteristics.[4] Separate estimates of land and building value are used to adjust property tax assessments for structure depreciation and for changes over time in land value. Moreover, the

---

[1] For example, Diamond (1980) stressed that the price of urban residential land depended primarily on location features and amenities.

[2] Throughout the paper, we use the terms land prices and location values interchangeably. Location value highlights that, as suggested by theory, the right to build at a specific location commands a price.

[3] The relative volatility over time of the land value component contributes to macroeconomic risks as suggested by Davis and Palumbo (2008) and by Bourassa et al. (2011).

[4] Longhofer and Redfearn (2009) note that property taxes are typically based on total property value. Nevertheless, tax assessors separately estimate and report the two components (Gloudemans, Handel and Warwa, 2002).

ratio of structure to land is used by investors to choose the time and intensity of redevelopment (Hendriks, 2005; Dye and McMillen, 2007; Clapp and Salavei, 2010; Ozdilek, 2012).

The complex interaction between land values and the values of improvements includes an option premium for the right but not the obligation to renovate a small old house in a neighborhood with McMansions on identically sized lots. Moreover, Longhofer and Redfearn (2009) argue that the implicit prices of structural attributes vary over space because many neighborhoods were developed with relatively homogeneous structural characteristics that adjust slowly as the land values in the neighborhood change. They propose locally weighted regressions with variation in implicit prices modeled by a smoothing function.

We introduce an alternative approach that considers the interaction between structure and land. Our approach combines local polynomial regressions (LPR) with a linear ordinary least squares model. The former provides estimates of the location values of each property over time, while the latter, using the characteristics of the structure, provides estimates of the value of each structure. A backfitting method ensures orthogonality between location and structure.

To generate location values, our empirical analysis requires only a standard hedonic dataset that include sales price, location (latitude and longitude), and housing characteristics. For years in which a property was not sold, we interpolate its value as a logical extension of LPR, where estimation is typically on a grid that spans the data; estimated values are typically interpolated to all sales falling within the grid. The interpolations produce a balanced panel of estimated land values at each location and any point in time where a property has sold. The balanced panel spans the years and locations covered by the hedonic dataset.

Our interpolation to a balanced panel is motivated by the fact that accurate property valuation requires separate valuation of land and structure and by many uses of panel data in

studies of house value. For example, where valuation of school attributes is based on multiyear data, house level fixed effects have required repeat sales, and these sales suffer from sample selectivity.[5] Our method for interpolating values allows use of all sales, including houses selling only once.

Any study where the identification strategy depends on mean differencing of panel data might apply our interpolation methods to construct a balanced panel. Our empirical analysis uses housing sales data for Denver over the period of 2003-2010; data include sales price, location, and various housing characteristics.[6] In our application, we focus on the effect of changes in airport infrastructure at the Denver International Airport on land prices. However, our analysis is easily adapted to any amenity that might affect land values. For example, Yinger (2009) uses structural hedonic models to identify the effect of distance to an environmental hazard. Our method suggests using changes in the hazard over time to test robustness in a reduced form model.

Our spatial smoothing LPR methods are most closely related to Gibbons, Machin and Silva (2013), who improve on boundary fixed effects models by using spatial smoothing for more slowly varying cross-boundary trends related to demographic sorting and other factors. Similarly, Brasington and Haurin (2006) use spatial statistics as part of their identification strategy. They use a nearest neighbor spatial weight matrix that "acts like a highly localized

[5] Nguyen-Hoang and Yinger (2011) provide a detailed review of several multi-year fixed effect models intended to measure school quality capitalization. Section 2.2.2 evaluates house level fixed effects, such as Figlio and Lucas (2004), as compared to attendance zone fixed effects models, such as Black (1999) and Bayer, Ferreira and McMillan (2007).

[6] We have chosen this time frame due to the availability of airport infrastructure investment data.

dummy variable," controlling influences such as a nearby abandoned property (p. 260). Our spatial controls cover much larger distances, but a second stage estimation using our balanced panel might control more localized characteristics.

The major contributions of our research include our application of a semi-parametric estimation technique using local polynomial regressions (LPR) to separate the value of location from the value of structures. Second, we calculate the NRMSE for the LPR and ordinary least squares (OLS) approaches, and perform a statistical test to determine that the LPR approach is more efficient than OLS. Third, we develop an interpolation method to construct balanced panel data: that is, we estimate location value at every location for every year in our sample. Fourth, the panel data allows us to use a GMM estimation approach to determine the spillover effects of changes in airport infrastructure capital on location values, while addressing potential endogeneity of the regressors. More generally, our method for estimating a balanced panel of land values may be applied in different contexts.

Following this introduction, the paper consists of several sections. First is a selective literature review, followed by a brief summary of the data used in our analyses. Next, we summarize a semi-parametric approach developed by Clapp (2004) for separating land prices from improvements. Our extension of Clapp (2004) is the interpolation of location values for additional years in which there are no sales of a particular property. We present the interpolation procedure and the resulting land values. We also compare the predictive accuracy of the LPR approach against OLS. We then examine the extent to which various types of airport infrastructure expenditures spill over into land values. We complete the paper with a summary of our main findings.

Literature Estimating Location Values and Spillover Effects

U.S. housing prices (i.e., the total price that includes land and structures) experienced a dramatic increase in the years leading up to 2006. This boom in housing prices was followed by a major bust that began in 2006. When one takes a closer look at the U.S. boom and bust, one sees much heterogeneity across regions. Such heterogeneity is described in detail in Cohen, Coughlin and Lopez (2012). A related finding during the boom and bust is that land prices have been more volatile than structure prices.[7]

In this paper, our focus is on residential location values in Denver, a metropolitan area that did not experience the boom and bust extremes of many areas.[8] Quarterly housing prices in Denver in the years 2000 through 2012, which includes the boom, bust, and nascent recovery in the U.S. housing market, are shown in Figure 1**.** Additionally, and explained in detail later, Figure 1 depicts the levels of the time dummy variables in an OLS (hedonic) regression of Denver housing prices for the years of our dataset. It is noteworthy that the general trend in housing prices tracks the trend in the time dummy variables.

[Insert Figure 1 here]

In an illustration of one way to use our residential location value estimates, we explore the effect of infrastructure spending at the Denver International Airport on our estimates. A

[7] Recently, Nichols, Oliner and Mulhall (2013) have found such a result, a finding that is consistent with prior research by Davis and Heathcote (2007), Davis and Polumbo (2008), and Sirmans and Slade (2012).

[8] While our approach can be applied generally, our focus on Denver reflected a desire to avoid areas with extreme booms and busts in housing prices for our initial analysis. Plus we felt that an airport removed from the city would make our infrastructure-related analysis cleaner.

large literature examines how and to what extent transportation infrastructure spills over or is capitalized into housing prices, including McMillen and McDonald (2004), Weinberger (2000), and Forest, Glen, and Ward (1996). These papers use a hedonic pricing approach to assess the effects of the transportation infrastructure on housing prices (or commercial rents, in the case of Weinberger). Gatzlaff and Smith (1993) use both a repeat sales approach and a hedonic housing prices approach to examine the effects of a new rail line in Miami, and find that it had only a minimal impact. However, none of these papers directly address the issue of transportation infrastructure spillovers by distinguishing the effects on land values.[9]

The separation of urban land value from structure value is made challenging by the scarcity of vacant land sales in an urban setting. Hendriks (2005) evaluates three methods used by appraisal professionals for this purpose: fractional apportionment (FAT), rent apportionment (RAT) and price apportionment (PAT) theories. He raises substantial questions about each, recommending that appraisers caution their clients about the unreliability of apportionment methods. Our local regression method (LRM) is most closely related to PAT since it uses sales prices together with location and property characteristics to allocate value (i.e., predicted price from a hedonic model) between land and structure.

Longhofer and Redfearn (2009), who examine how in practice one might disentangle the value of land from the value of structures on the land, argue that land and structures are inseparable, as does Hendriks (2005). Both appeal to an argument that houses within a neighborhood are reasonably homogeneous, in terms of the general size of structure relative to lot size. The Longhofer and Redfearn approach requires data on vacant land sales, and they

[9] Additional references related to infrastructure spillovers can be found when we present our illustration.

estimate land values city-wide using locally weighted regressions. In some applications, a lack of vacant land sales data may pose challenges to implementing this approach.[10]

Clapp and Salavei (2010) focus on a different approach than the one in Longhofer and Redfearn (2009). Specifically, they implement an "option value" approach where existing structure relative to optimal structure at any time will influence the value of the land. There are high adjustment costs, including foregone rents from the existing structure and construction costs, so reaching the redevelopment "trigger point" takes time. Therefore, a property with a given set of characteristics will also have covariant location value and implied prices for these characteristics.

Longhofer and Redfearn (2009) use a nonparametric approach, locally weighted regressions. They allow the valuation of location and structural characteristics to vary smoothly over space. The spatial smoothing method is similar to Clapp (2004) except that he holds the implicit prices of structural characteristics constant and requires orthogonality between structure prices and location values. However, Longhofer and Redfearn (2009) attribute all spatial variation in implicit structure prices to a second stage land valuation equation, so the difference between the two valuation methods may not be great.

The Clapp (2004) LPR approach separates the value of land and improvements with a semi-parametric method. We use LPR in the present paper, and we incorporate several extensions. We implement a procedure to interpolate location values for each property in all years in which the property was not sold. This is a logical extension of the LPR method, where

[10] In the context of commercial real estate, Haughwout, Orr, and Bedoll (2008) estimate land prices using a dataset that includes purchases of vacant land as well as plots with unoccupied structures slated for demolition and subsequent replacement by new constructions.

estimation on a grid and interpolation to all datapoints is standard. We also conduct statistical tests for assessing the accuracy of the interpolation procedure and for comparing the predictive accuracy of the LPR versus the OLS approaches.

Data Summary

Descriptive statistics for the housing data are presented in Table 1 for Denver. There were over 326,000 observations for single family residential homes that sold between 2003 and 2010 in Denver. The distribution of sales across years was fairly uniform through 2006, then transactions declined by as much as 50 percent. Still, due to the large sample size, the smallest number of yearly sales, in 2010, was over 20,000. The distribution of sales across counties was reasonably uniform. There were approximately 3.1 bedrooms, with approximately 2.3 full baths and 0.33 half-baths in the typical house sold in Denver over this period. Well over half the houses sold had a garage, a basement and a fireplace. The average sale price was approximately $250,000. Using the latitudes and longitudes of the houses and the airport, the average house in our dataset was located about 20.5 miles from the center of the airport. The closest house was 5 miles from the airport while the furthest house was 56 miles away.

[Insert Table 1 here]

Method for Separating Land and Structure Values[11]

The preceding housing data are used in our method to disentangle location values from structure prices. We follow the LRM and "option value" approach of Clapp (2004) and Clapp and Salavei (2010), respectively. Location value (i.e., the value of the right to build a single family residence at a given location) exhibits more variation across both time and space than structural values,

[11] The LRM description and Technical Appendix below parallel the discussion in Cohen *et al*. (2013).

which can be reproduced at the current cost of construction once the redevelopment trigger point has been reached.[12]

First, a parametric method – the standard hedonic model – is used for generating implicit prices for all housing characteristics (structure and location), and a price index independent of these characteristics. We regress the log of sales price (*lnSP*) on a vector of house structure characteristics (Z), locational characteristics (S), and time (*t= 1…T)*) which is represented here in the form of annual time dummies, Q_t:

$$\ln \mathrm{SP}_{\mathrm{it}} = \gamma_0 + Z_{it}\alpha + S_i\beta + \gamma_1 Q_1 \cdots + \gamma_T Q_T + \varepsilon_{it} \qquad (1)$$

where ε is assumed to be an iid, normally distributed (for the purposes of hypothesis testing) noise term.[13] For the estimation of equation (1), we drop one of the time dummies, which becomes the base year.

The cumulative log price index for a standard house in the area is measured by the parameters on the annual time dummies, γ. Using our analysis, we plot the price index in Figure 1 as the exponential of each of the time dummies, with 2010 as the base year (which has a value of 100 in Figure 1). In constructing the price index, we assume the structure and location parameters do not vary over time. But since they are not constant over time, over any time interval *T* we are measuring the average implicit prices, α and β. This forces any changes over

[12] Davis and Palumbo (2008) decompose property value into structure and land components, and find significant changes in land value over time and across metropolitan areas. They subtract the cost of construction from sales prices, while we use the implicit value of the structure.

[13] The natural log of sales price is the dependent variable because logarithms control for heteroscedasticity and some nonlinearity, and enhance degrees of freedom. Hastie and Tibshirani (1990), pp. 52-55, discuss degrees of freedom for smoothing models.

time into the estimates of the γ parameters; they can be considered an approximation to a pure time component that shifts the constant of the regression, γ_0.

The LPR model differs from equation (1) primarily by estimating the equation at each point on a grid composed of equally-spaced time, latitude, and longitude points that span the data. In our model, there are 20 time, 15 latitude, and 15 longitude points, for a total of 4500 "knots" (or target points) on the grid. The size of the bandwidth determines whether or not an observation will be used to estimate the function value at the knot. For this paper, the bandwidth is chosen to be {.3σ(time), .3σ(latitude), .3 σ(longitude)} and the bandwidth is adjusted upward at any target point where there are fewer than 20 observations within one bandwidth. The technical appendix contains a discussion of cross-validation bandwidth selection, methods for dealing with insufficient density of transactions at any target point, estimation of standard errors, and other details of the LPR model.

We focus on the nonlinear space-time relationships. The semi-parametric LRM model enters because of the "curse of dimensionality." As a practical matter, there would typically be five or six variables for structural characteristics (e.g., interior area, bathrooms) on the left hand side of equation (1). If all were represented by even a coarse grid, the data would be sparse near any point. The semi-parametric solution assumes linearity for the equation (1) parameters, α, on all the housing characteristics.[14] An LPR model is used in the LRM method to estimate these coefficients conditional on the location of the house. This approach addresses the concern of Longhofer and Redfearn (2009) by requiring statistical independence between the estimated coefficients on Z and the nonlinear part of the model.

[14] Of course, a nonlinear relationship (e.g., with building age) is typically modeled with a quadratic term.

To implement this logic, the LRM method begins by using ordinary least squares to estimate equation (1) followed by LPR estimation to revise the $\hat{\alpha}$'s to assure independence from the location value estimates: the coefficients are the "Robinson" coefficients, $\hat{\alpha}_R$. The Robinson coefficients are estimated after conditioning the SP and Z variables on latitude, longitude and time. Then, we subtract the estimated value of structural characteristics to obtain the partial residuals:

$$partres_{it} = lnSP_{it} - Z_i \hat{\alpha}_R \qquad (2)$$

where *partres* is the partial residual after subtracting structure value estimated with LPR.

A nonparametric part of the LRM model is:

$$partres_{it} = q\left(S_i, t_i\right) + \varepsilon_{it} \qquad (3)$$

where S_i is a vector consisting of the latitude and longitude and t_i is the date of sale for house *i*. The "backfitting" method iterates between equations (2) and (3) until there is negligible change in $\hat{\alpha}_R$.[15]

To summarize, the LRM estimate may be taken as a reasonable approximation to location value over time, $q\left(S_i, t_i\right)$, because we subtract an average value of structural characteristics, $Z_i \hat{\alpha}_R$ where the estimation method requires statistical independence between location value and

[15] When we apply the kernel weighting scheme to equation (3), this solves a problem in the standard hedonic model of the bunching of transactions within the quarters. The t_i variable is based on day, month, and year of the transactions and the t_0 target is the middle of the year.

improvement value.[16] We use LPR to estimate $q(\mathrm{S}_i, t_i)$ at each of 4500 target points (or "knots") on a grid that spans the data. We interpolate from the target points to our observations, a standard approach in the literature on nonparametric estimation. We innovate by using tri-linear interpolation, which we discuss below.

Results and Performance of the LPR Approach

Table 3 presents coefficients from a standard hedonic regression, equation (1), and compares them to the Robinson coefficients estimated by taking partial residuals from equation (1) and backfitting with equations (2) and (3). All the OLS coefficients have plausible signs and

[16] Some, such as Davis and Palumbo (2008), have suggested that location value should be estimated as property value less construction costs. To get to this quantity, one would add back $Z_i \hat{\alpha}_R$ and then subtract construction costs. An approximation to construction costs can be obtained by assuming that they are invariant within the metropolitan area and that they change slowly over time as the costs of material and labor change, and therefore the level of construction costs at time zero is the same for all properties in the city. The Marshall Valuation Service (MVS) is one approach to approximation of this level. Then percentage changes over time can be approximated by using a construction cost indexes such as those published by Engineering News-Record (ENR, http://enr.construction.com/economics/). With these adjustments, location value is estimated by:

$$\hat{q}(\mathrm{S}_i, t_i) + Z_i \hat{\alpha}_R - C_{it}$$

where C_{it} is an estimate of construction costs for house *i* at time *t*. This procedure may be considered as a robustness check.

magnitudes. The time dummy coefficients display a pattern consistent with the Denver house price index (Figure 1). Structural characteristics have magnitudes consistent with the literature. In particular, value decreases with structure age at a decreasing rate, a typical result for the housing market. Conversion of the age coefficients to an index equal to 100 for a new house show depreciation of about 1.5% per year declining to near zero at age 30, when the house is worth 80% of its initial value. After that values rise back to 100% at about age 60; this is likely due to renovations of older houses and to restrictions imposed by the quadratic functional form.

[Insert Table 2 here]

The Robinson coefficients handle location value (a function of latitude and longitude) in the nonparametric part of the model and they require orthogonality between the two parts of the model. The structural coefficients typically change by between 10 and 20 percent, except for the relatively unimportant coefficient on number of stories. But the backfitting method dramatically changes the way location is modeled. The county dummy coefficients change by amounts ranging from -46% to 165% as measured by the OLS coefficient divided by the Robinson coefficient. The highly constrained hedonic specification for location – the quadratic in latitude and longitude – is replaced by the nonparametric part of the LRM model, equation (3).

We conduct an exercise to compare the accuracy of the two models - OLS and LPR - in estimating location values in Denver between 2003 and 2010. For both models, we run multiple simulations of an out-of-sample forecast to produce estimates of location values, which are then added to the respective structural values to produce an estimated sales price. We compare the estimated sales price to the actual sales price and use normalized root mean squared error to determine which model, OLS or LPR, most accurately estimates location value. The exact steps taken are outlined below.

OLS

To forecast the location values using ordinary least squares, we omit a random 20% of the full set of observations (326,744). The remaining 80% is used to run the hedonic regression. We forecast the log of sales price of the omitted 20% using the coefficients of the 80% hedonic regression. Finally we compare the estimated log of sales price against the actual log of sales price of the 20% using normalized root mean squared error. We repeat this procedure 30 times to account for sample bias.

LPR

To forecast the location values using the LPR technique, we omit a random 20% of the full sample of observations (326,744). We then obtain the Robinson coefficients with a regression using the remaining 80% of the sample. We forecast the structural values of the remaining 20% using the coefficients from the 80% Robinson coefficients regression. To obtain the partial residuals of the 80% (used later in the interpolation procedure) we subtract the fitted structural values of the 80% from the actual log of the sales price of the 80%.

For the interpolation procedure to work, the 20% subset must be completely contained within the 80% subset: the maximum time, longitude, and latitude of the 20% must be less than the maximum time, longitude, and latitude of the 80%. Similarly, the minimum time, longitude, and latitude of the 20% must be greater than the minimum time, longitude, and latitude of the 80%. We remove the observations that fail to meet this requirement from the 20% subset.

Using the LPR technique, we estimate the location values of the 80% from the partial residuals calculated earlier. Then, using tri-linear interpolation, we forecast the location values of

the 20% (an appendix for the tri-linear interpolation procedure is available upon request from the authors). We add the forecasted location values to the previously forecasted structural values to get an estimate of the log of sales price of the 20%. Finally, we compare the estimated log of sales price against the actual log of sales price of the 20% using normalized root mean squared error. We repeat this procedure 30 times to account for sample bias.

t-Test

The average NRMSE over all of the 30 trails for the OLS and LPR are 0.0413 and 0.0411, respectively. The next question we address is whether or not this difference is statistically significant. We use Welch's t-test for the difference in the means of the NRMSE rather than the ordinary Student's t-test because we cannot assume that the variances of the NRMSE of the two populations (methods) are equal – the denominator of the t-statistic is not based on a pooled variance estimate. We calculate the t-statistic to be 1.989. With 54 degrees of freedom at the 5% significance level (one-tail test critical value = 1.674), we reject the null hypothesis and conclude that on average, the NRMSE of OLS is significantly greater than that of the LPR.

Location Value Interpolation and Airport Infrastructure Spillovers

We perform two extensions using the location value data. First, we implement an interpolation procedure to obtain estimates of land values for each house in our sample in every year over the period 2003-2010, and examine the accuracy of these land value estimates. In other words, the interpolation procedure estimates the value of land at each location, at every point of time. There are 2,613,744 (equal to 326,744 x 8) points in the grid. The 326,744 sales observations are initially used to generate 326,744 location values, as is standard in the literature. Then we must interpolate to fill in the remaining values for years in which a specific location did not have a

sale. Thus, the number of interpolated points is 2,613,744 – 326,744 = 2,287,208. Second, once we have interpolated these location values, we utilize them to assess the spillover effects of changes in various categories of airport infrastructure stock values on changes in location values.

Land Value Interpolation

The interpolation procedure uses the grid of equally spaced knots from the local polynomial regression. Because this grid spans the data, each point to be interpolated is surrounded by 8 knots. To understand this grid, one might imagine a cube (or more generally, a prism with 6 sides and 8 vertices), with the point to be interpolated within the cube. The method of tri-linear interpolation (Bourke, 1999) approximates the land value at any point inside the cube (or prism) using the values on the lattice points.

A pictorial representation of our interpolation results is provided in Figures 2-4. These figures show the location values in quintiles and facilitate comparisons between 2003 and 2006, 2006 and 2010, and 2003 and 2010, respectively. The quintiles are not calculated by separate years, but rather for the entire period. For a given figure, the more the figure is shaded by blue and green, the higher the estimated location values. Thus, Figure 2 suggests that location prices tended to rise somewhat between 2003 and 2006.

[Insert Figure 2 here]

Subsequently, land prices tended to decline for a few years and then flattened out. Note that in Figure 3, the extent of blue and green shading declines between 2006 and 2010.

[Insert Figure 3 here]

Finally, note that in Figure 4 the direction of a general change in location values is difficult to discern. While the blue area is larger in 2010 than in 2003, the red area is also larger in 2010 than in 2003.

[Insert Figure 4 here]

To confirm these results and indicate the general movement of land prices, here are the medians in natural logs by year: 2003 – 9.90, 2004 – 9.92, 2005 – 9.95, 2006 – 9.94, 2007 – 9.90, 2008 – 9.84, 2009 – 9.83, and 2010 – 9.84. These estimates indicate the following year-over-year changes in location values: 2003/2004: 2 percent, 2004/2005: 3 percent, 2005/2006: -1 percent, 2006/2007: -4 percent, 2007/2008: -6 percent, 2008/2009: -1 percent, and 2009/2010: 1 percent. We conclude that there was little change overall between 2003 and 2010, so the changes evident in the Figures relative to the airport can be interpreted accordingly.

Balanced Panel Illustration: Estimating the Spillover Effects of Airport Capital Stocks on Land Values

Possible applications of a balanced panel of location values data abound. One way for us to illustrate the usefulness of generating such a balanced panel data set is to demonstrate how location values are impacted by airport infrastructure capital stocks over time. We estimate the following model after obtaining interpolated land values for each house in each year:

$$L_{i,t} = c_0 + c_1 * A_{1,\,i,t} + c_2 * A_{2,\,i,t} + c_3 * A_{3,\,i,t} + c_4 * A_{4,\,i,t} + c_5 * A_{5,\,i,t} + \alpha_i + \tau_t + \varepsilon_{i,t} \qquad (4)$$

In this model, $L_{i,t}$ is level of the real interpolated land value for property i in year t; $A_{1,\,i,t}$ through $A_{5,\,i,t}$ represent airport infrastructure stocks for property i in year t for airfields,

terminals, parking, roads/rails/transit, and "other", respectively; $A_{1,\,i,t}$ through $A_{5,\,i,t}$ are weighted by the distance from house i to the airport in year t; α and τ are individual and time fixed effects, respectively; and $\varepsilon_{i,t}$ is an iid error term with mean zero, constant variance and zero covariance across observations; and for Denver, i=1, 2,…, 178,731; t=2003, 2004,…, 2010. [17] All time invariant unobservables are controlled for by α_i, and general time effects by τ_t. Infrastructure effects are identified by changes over time at the individual location level.

We employ a year-over-year change approach, which leads to the following model:

$$\Delta L_{i,t} = c_1 * \Delta A_{1,\,i,t} + c_2 * \Delta A_{2,\,i,t} + c_3 * \Delta A_{3,\,i,t} + c_4 * \Delta A_{4,\,i,t} + c_5 * \Delta A_{5,\,i,t} + \theta + \Delta\varepsilon_{i,t} \quad (5)$$

where Δ is the one-year-over-year change for property i. This year-over-year change approach enables us to identify the impacts of airport infrastructure on location values, as described in greater detail below. Note that examining year-over-year changes causes the cross-sectional fixed effects to drop out. The year-over-year analysis also leads to a new set of time-specific fixed effects, θ, which includes an intercept term.

The units of observation are the interpolated land values for individual houses with a transaction during the years 2003-2010. We identify the effect of improvements in year t off of change from before to after the "event," FAA's five major categories of investments.

[17] The level of the real interpolated land value is calculated in the following order: 1) add the value for the coefficient on the time dummy to the interpolated value from the LPR approach; 2) divide this land value, which is in logs, by a deflator: and 3) take the exponential of the real value. There is evidence that land values increase with the square root of lot size, so the fact that we are using logs is important since it prevents excess acreage from having the same effect on value as the building pad.

Specifically, we expect a dampened effect due to distance from the airport. While we initially control for unobservables with cross-sectional fixed effects, these effects drop out due to the transformation of the variables to year-over-year changes as in equation (5). The time fixed effects variables remain (i.e., the θ in equation (5)). Also, we don't have the identifying demographic groups that Clapp and Ross (2004) had to address the possibility that individuals with jobs involving travel may prefer to reside near the airport. A strategy here for allowing sorting is to allow the individuals who value a bigger and better airport to bid up the price of housing (which is reflected in land values). We accomplish this by allowing a lag after airport investments.

In applications such as ours, there may be other variables that spill over into land values, such as airport noise (a disamenity). Any heterogeneity due to airport noise can be captured through our individual-level fixed effects (FE). Since our model in equation (5) is based on the FE, we do not need to collect demographics at the CBG level. We focus on individual transactions because houses within any particular CBG will differ in their access to the expanded airport. By lining all the transactions up around each given year of airport expansion/depreciation, and including calendar year FE along with the individual level FE, we control for omitted variables other than the expansions and depreciation.[18]

[18] However, we do not have enough "events" in enough MSAs to do the statistical tests used in event studies in the finance literature. The most important explanatory variables are distance from the airport interacted with the amount and type of expansion. It may also be the case that some expansions don't increase congestion but only make the terminal facilities more attractive.

Finally, we estimate equation (5) for the one-period change model, using the Arellano-Bond estimator. This approach uses Generalized Method of Moments (GMM), to instrument each one-period change infrastructure variable with the corresponding lagged value of the one-period change infrastructure variable. This approach follows Arellano and Bond (1991).

We use both airport infrastructure and housing sales data from Denver for this application of our balanced panel estimates. The Denver International Airport opened in 1995. During our sample period, substantial additional airport infrastructure investment as well as depreciation occurred. We use the perpetual inventory method, together with real investment and depreciation assumptions (see the data appendix), to calculate the stocks of infrastructure capital for 2003-2010. These capital stock estimates for various categories of airport infrastructure are shown in Table 3.

[Insert Table 3 here]

Table 4 presents the GMM estimates of equation (5). Given the complex nature of the urban area southwest of the airport, we choose to focus this part of our application on the properties in the northwest quadrant of the airport. Figures 5-7 show the locations of the properties examined and their associated land values for 2003 and 2006, 2006 and 2010, and 2003 and 2010, respectively. Relative to the entire sample, land values used for this part of our analysis tend to be less than in the other part of the Denver area. The southwest region is close to downtown Denver, and there are likely a broad variety of economic factors that can be expected to influence land values. Although there were 54,439 home sales between the years 2003-2010 in the northwest region, it is less developed than the southwest region and there are fewer other factors (such as other types of infrastructure, business activity, etc.) that might be expected to influence land values.

[Insert Figures 5, 6, and 7 here]

After interpolating the land values in all years for the houses sold in the northwest region, we have over 381,000 land value observations. However, we base this estimation approach on a set of properties that exclude a "buffer" of approximately 20%, which leaves approximately 322,000 properties in our analysis for the northwest quadrant. This buffer approach mitigates any potential "border" issues that may bias the coefficient estimates by using the full sample.

To address potential concerns about lack of orthogonality between the airport infrastructure growth rates and the error terms, we perform a GMM procedure for the one-year change model. We use lagged one-year changes (for each infrastructure type) as instruments in the one-year change version of equation (5). There is an issue worth noting. Generally, if the number of instruments exceeds the number of parameters to be estimated, the J statistic is a test that can be used for over-identification of the model. The J statistic in our GMM estimation is 1.74*10^(-18) (essentially 0), implying exact identification (Hansen, 1982). But in fact, in our estimation the number of orthogonality conditions is the same as the number of parameters, so the test for over-identification is not necessary.

Table 4 presents the GMM results with the one-year change in land values as the dependent variable. The one-year changes in each infrastructure category, inversely weighted by each property's (i.e., observation's) distance to the airport, are the independent variables (as shown in equation (5)). In the very short run, the coefficients for airfields, parking, intermodal transportation, and roads, rail, and transit are positive, while the coefficient for other infrastructure is negative. All variables reveal a highly statistically significant (P-value = 0.0000) relationship with land prices, which is not surprising given the large number of observations.

[Insert Table 4 here]

Precisely how the airport infrastructure expenditures in each of the five categories spill over to location values is unclear. The lumpiness of expenditures both within and across categories precludes straightforward interpretations. Moreover, both of these parking- related improvements are much different in nature and transparency for travelers and businesses than some of the "other" activities, such as de-icing equipment improvements, can lead to other externalities such as pollution runoff.

The differences in the coefficient estimates in Table 4 suggest, given the respective stocks of infrastructure capital, that the impact of additional spending varies substantially across the categories. The coefficient estimates in Table 4 imply a somewhat small (but statistically significant) effect of infrastructure improvements on land values at a given location. For example, with a $1,000 increase in the airfields capital stock, the average property's land value rises by approximately $0.29.[19] A similar $1,000 increase in the terminals capital stock leads to an average land value that is $0.34 higher. A $1,000 increase in parking infrastructure leads to a $0.24 increase in average land values. With a $1,000 increase in intermodal roads, rail, and transit capital stocks, land values rise by $2.58. Perhaps the magnitude for intermodal infrastructure is dramatically higher than that for the airport-specific investments because roads, rail, and transit may be utilized by individuals who are not necessarily travelling by air, as well as by air travelers. Finally, the coefficient on "other" infrastructure implies a $1,000 increase in this type of capital leads to a $0.30 decrease in land values. Since this category includes the categories of airport infrastructure not captured by the four other categories, it is somewhat of a

[19] Of course, numerous properties are affected by the expenditure, so the aggregate impact in this illustrative case substantially exceeds $1000.

"black box". One might conjecture that deicing equipment, for instance, has detrimental effects on the environment through runoff, which could adversely affect land values.

One important caveat concerning the interpretation of the results in Table 4 is that the parameter estimates from regressions in general reflect the effects of "small" improvements in infrastructure, however, the lumpy nature of many infrastructure projects implies there may not be much effect of a small change in the capital stock. For instance, a new or renovated terminal does not provide any benefits until it is completed, while there may be investments on it over large periods of time that are counted in the capital stocks at the time the investments are made. For this reason, it is important to consider the changes in infrastructure over time, which is the approach we follow for the results in Table 4.

Another important issue is the timing of the spillover effect. There may be some initial price adjustments at the time of the expansion announcement. However, the time path of any adjustment process is not clear. There may not be a full price adjustment until several years later once the investments are in place and functioning.[20] Obviously, our illustration does not fully address this issue.

Conclusion

We present a theoretically sound, semi-parametric estimation procedure - local polynomial regressions - to estimate location values. In addition to being grounded in statistical theory, the estimation procedure can be implemented in a straightforward manner using datasets that are commonly used in studies of housing markets. All that is required is data on sale prices, sales

[20] Jud and Winkler (2006), Agostini and Palmucci (2008), and McMillen and McDonald (2004) found evidence of adjustment effects of various types of transportation infrastructure.

dates, and on the associated structural and location characteristics of the properties. We compare the LPR and OLS models using an out-of-sample forecasting procedure, and determine through a difference in means test on the respective NRMSE that the LPR model is more efficient at predicting location values.

A major contribution of our analysis is our development of an interpolation procedure that is useful in estimating a balanced panel of location value observations in each year for all properties that sold during the sample period. We accomplish this with trilinear interpolation with the same grid as for the LPR estimates so that values for year in which a property did not sell are interpolated in the same way as those in which it did. We validate the accuracy of the interpolated estimates with an out-of-sample forecasting approach.

Finally, we illustrate a potential application of our interpolation procedure, together with the LPR estimates, for properties near Denver's airport. Our balanced panel GMM regressions on yearly changes in location values suggest that different airport infrastructure investments have positive, albeit different in magnitude, impacts on location values. These impacts attenuate with distance from the airport.

Our interpolation procedure to generate balanced panels of location values from LPR estimation has the potential for many other applications. These include the effects of school spending on location values and assessment of the impacts other types of public goods, such as parks, on location values. In addition, it is easy to envision the usefulness for other applications, such as house price dynamics driven mostly by changes in land value or taxation of land separately from structures. Finally, the density of location values for a given area allows for the generation of numerous land price gradients.

References

Agostini, C.A., Palmucci, G.A. (2008) The Anticipated Capitalisation Effect of a New Metro Line on Housing Prices. *Fiscal Studies*, 29: 233-256.

Arellano, M. and Bond, S. (1991). Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, 58: 277-297.

Bayer, P., Ferreira, F., & McMillan, R. (2007). A Unified Framework for Measuring Preferences for Schools and Neighborhoods. *Journal of Political Economy*, 115(4), 588-638.

Black, S. (1999). Do Better Schools Matter? Parental Valuation of Elementary Education. *Quarterly Journal of Economics*, 114(2) 577-599.

Bourassa, S.C., Hoesli, M., Scognamiglio, D., Zhang, S. (2011) Land Leverage and House Prices. *Regional Science and Urban Economics*, 41: 134-144.

Bourke, P. (1999) Interpolation Methods. Available online at: http://paulbourke.net/miscellaneous/interpolation/ [Accessed 09 May 2014].

Brasington, D., & Haurin, D. R. (2006). Educational Outcomes and House Values: A Test of the value added Approach. *Journal of Regional Science*,*46*(2), 245-268.

Clapp, J. M. (2004) A Semiparametric Method for Estimating Local House Price Indices. *Real Estate Economics,* 32: 127-160.

Clapp, J.M., Ross, S.L. (2004) Schools and Housing Markets: An Examination of School Segregation and Performance in Connecticut. *Economic Journal*, 114: F425-F440.

Clapp, J.M., Salavei, K. (2010) Hedonic Pricing with Redevelopment Options: A New Approach to Estimating Depreciation Effects. *Journal of Urban Economics,* 67: 362-377.

Cohen, J.P., Coughlin, C.C., Lopez, D.A., Clapp, J.M. (2013) Estimation of Airport Infrastructure Capitalization for Land Value Capture Purposes. *Working Paper WP13JC2,* Lincoln Institute of Land Policy.

Cohen, J.P., Coughlin, C.C., Lopez, D.A. (2012) The Boom and Bust of U.S. Housing Prices from Various Geographic Perspectives. *Federal Reserve Bank of St. Louis Review,* 94: 341-368.

Cohen, J.P., Coughlin, C.C. (2008) Spatial Hedonic Models of Airport Noise, Proximity, and Housing Prices. *Journal of Regional Science,* 48: 859-878.

Cohen, J.P., Morrison Paul, C.J. (2003) Airport Infrastructure Spillovers in a Network System. *Journal of Urban Economics,* 54: 459-473**.**

Davis, M.A., Heathcote, J. (2007) The Price and Quantity of Residential Land in the United States. *Journal of Monetary Economics*, 54: 2595-2620.

Davis, M.A., Palumbo, M.G. (2008) The Price of Residential Land in Large US Cities. *Journal of Urban Economics*, 63: 352-384. Data located at Land and Property Values in the U.S., Lincoln Institute of Land Policy. Available at: http://www.lincolninst.edu/resources/.

Diamond, D.B. (1980) The Relationship between Amenities and Urban Land Prices. *Land Economics*, 56: 21-32.

Dye, R.F., McMillen, D.P. (2007) Teardowns and Land Values in the Chicago Metropolitan Area. *Journal of Urban Economics*, 61: 45-64.

Figlio, D. N., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *American Economic Review*, 94(3), 591-604.

Forrest, D., Glen, J., Ward, R. (1996) The Impact of a Light Rail System on the Structure of House Prices: A Hedonic Longitudinal Study. *Journal of Transport Economics and Policy*, 30: 15-29.

Gatzlaff, D.H., Smith, M.T. (1993) The Impact of the Miami Metrorail on the Value of Residences near Station Locations. *Land Economics*, 69: 54-66.

Gibbons, Stephen, Stephen Machin, and Olmo Silva. Valuing School Quality Using Boundary Discontinuities. *Journal of Urban Economics,* 75 (2013): 15-28.

Gloudemans, R. J., Handel, S., & Warwa, M. (2002). An Empirical Evaluation of Alternative Land Valuation Models. Lincoln Institute of Land Policy.

Hansen, L.P. (1982) Large Sample Properties of Generalized Method of Moments Estimators. E*conometrica,* 50: 1029-1054.f

Hastie, T. J., Tibshirani, R.J. (1990) Generalized Additive Models. Chapman & Hall/CRC Monographs on Statistics & Applied Probability.

Hendriks, D. (2005) Apportionment in Property Valuation: Should We Separate the Inseparable? *Journal of Property Investment & Finance,* 23: 455-470.

Haughwout, A., Orr, J., Bedoll, D. (2008) The Price of Land in the New York Metropolitan Area. *Current Issues in Economics and Finance,* 14: 1-7. Federal Reserve Bank of New York.

Ingram, G.K., Hong, Y.H., eds. (2012) *Value Capture and Land Policies*. Cambridge: Lincoln Institute of Land Policy.

Jud, G., Winkler, D. (2006) The Announcement Effect of an Airport Expansion on Housing Prices. *Journal of Real Estate Finance and Economics,* 33: 91-103.

Longhofer, S.D, Redfearn, C.L., (2009). Estimating Land Values Using Residential Sales Data. *Working Paper WP09SL1*, Lincoln Institute of Land Policy.

McMillen, D.P., McDonald, J. (2004) Reaction of House Prices to a New Rapid Transit Line: Chicago's Midway Line, 1983-1999. *Real Estate Economics*, 32: 463-486.

Nguyen-Hoang, P., & Yinger, J. (2011). The Capitalization of School Quality into House Values: A Review. *Journal of Housing Economics*, 20(1), 30-48.

Nichols, J.B., Oliner, S.D., Mulhall, M.R. (2013) Swings in Commercial and Residential Land Prices in the United States. *Journal of Urban Economics*, 73: 57-76.

Ozdilek, U. (2012) An Overview of the Enquiries on the Issue of Apportionment of Value between Land and Improvements. *Journal of Property Research*, 29: 69-84.

Pavlov, A.D. (2000) Space-Varying Regression Coefficients: A Semi-parametric Approach Applied to Real Estate Markets. *Real Estate Economics*, 28: 249-283.

Ruppert, D.,Wand, M.P., Holst, U., Hossjer, O. (1997) Local Polynomial Variance Function-Estimation. *Technometrics* 39(3): 262-273.

Sirmans, C.F., Slade B.A. (2012) National Transaction-based Land Price Indices. *Journal of Real Estate Finance and Economics*, 45:829-845.

Wand, M. P., Jones, M.C. (1995) Kernel Smoothing. Chapman and Hall/CRC Monographs on Statistics & Applied Probability.

Weinberger, R.R. (2000) Commercial Rents and Transportation Improvements: The Case of Santa Clara County's Light Rail. *Working Paper WP00RW2*, Lincoln Institute of Land Policy.

Yinger, J. (2009). Hedonic Markets and Explicit Demands: Bid-Function Envelopes for Public Services, Neighborhood Amenities, and Commuting Costs. *Center for Policy Research Working Paper*, (114).

Data Appendix

-Capital stocks:

We use the perpetual inventory method with annual data on new airport investments in several different categories (airfields; terminals; parking; rail, road and transit; and "other") to obtain separate estimates of capital stocks for each of these categories. The source for the disaggregated airport investment data for Denver is the FAA CATS database, Form 127, which is located at http://cats.airports.faa.gov/Reports/reports.cfm (most recently accessed 9/28/14).

Specifically, we deflated the investment series using a national deflator for government investment obtained from the 2013 Economic Report of the President, and the initial (or seed) value for the capital stock for each category. This seed value is obtained as the average of the investment data for the years 2001 through 2004, multiplied by the estimated service life for each category of investment. The depreciation rate was assumed to be the inverse of the service life, and the capital stocks followed a straight line depreciation path. Consequently, the depreciation rate = 1/service life.

The service lives for the different categories were as follows: service lives of airport terminals and airfields = 25 years; service life of parking = 40 years; service life of roads/rail/transit = 44 years; service life of "other" = 25 years. The justification for these service lives can be found in the following sources. The 25- year number for airfields and terminals came from Airports Council International and was used in Cohen and Morrison Paul (2003). The service life for parking can be found at the following site: http://www.chamberlinltd.com/extending-the-service-life-of-parking-structures-a-systematic-repair-approach/. For the roads, rail, and transit variable, we take the average of these two service lives and use 44 years for the service life. See

http://www.bea.gov/scb/account_articles/national/0797fr/table3.htm: The highways and streets service life is 60 years (0.0152) and the state and local railroad equipment service life is 28 years (0.0590).

-Land price indexes

We interpolated land values for all years for each house, using a method devised by Clapp (2004) and subsequently modified by Brett Fawley, Diana Cooke, and us. Details are available from the authors upon request.

Subsequently, we add back the values of the time dummy variables from the hedonic regressions, then deflated the land values by the CPI for Denver.

Technical Appendix

The purpose of this appendix is to explain the details of the local polynomial regression (LPR) and local regression model (LRM) methods, including calculation of bandwidths and standard errors. These methods are summarized by an algorithm. Parts of this discussion parallel Cohen et al. (2013).

A standard hedonic model provides a point of departure. Regress the log of sales price (*lnSP*) on a vector of house structure characteristics (*Z*), locational characteristics (*S*), and time (*t*), which is represented here in the form of annual time dummies, Q_t:

$$\mathit{ln}\mathrm{SP}_{\mathrm{it}} = \gamma_0 + Z_i\alpha + S_i\beta + \gamma_1 Q_1 \cdots + \gamma_T Q_T + \varepsilon_{it} \qquad \text{(A1)}$$

where ε is an iid noise term that is assumed to be normally distributed for the purposes of hypothesis testing.

By way of contrast, a local polynomial regression fits a surface to the observations conditional on the function values estimated at each knot on a grid. The LRM is a partial linear model designed to allow substantial nonlinearity in the spatial and time dimensions: it fits a value surface at each point in time as an alternative to estimating the set of parameters for *S* and *t* in equation (A1). It retains the linear portion of equation (A1) for structural characteristics. The LRM views price index and value surface estimates as descriptive exercises that are not designed to test hypothesis about parameters. Writing the model as follows emphasizes the nonlinear and nonparametric aspect of the LRM:

$$\mathit{ln}\mathrm{SP}_{it} = f\left(Z_i, S_i, t_i\right) + \varepsilon_{it} \qquad \text{(A2)}$$

We allow the function *f( )* to be nonlinear because local house prices rarely move in a straight line over time and a nonlinear spatial pattern is well known.

LRM estimation methods can be introduced by imagining that a number, q, of identical houses transact at a given point in space and time, denoted by the fixed vector ($\mathbf{z}_0$, $\mathbf{s}_0$, t_0). Then, an obvious way of estimating equation (A2) at the fixed point would be to average those prices:

$$\hat{f}(\mathbf{z}_0, \mathbf{s}_0, t_0) = \frac{\sum_{i=1}^{q} ln\mathrm{SP}_{it_0}}{q} - \frac{\sum_{i=1}^{q} \varepsilon_{it_0}}{q} \tag{A3}$$

The error term results from negotiation between heterogeneous buyers and sellers. Since the average error term will tend to zero as the sample size gets large, we will have a consistent estimator of a point on the value surface at the given point in time.

Actual sales prices are spread out in space and time as well as over the range of housing characteristics, *Z*. If the data were densely distributed over these characteristics, then we could average prices that are "close to" any particular point in characteristic space ($\mathbf{z}_0$), physical space ($\mathbf{s}_0$) and time (t_0). This averaging process is very much in the spirit of nonparametric smoothing. Nonparametric smoothing implements this local averaging idea by down-weighting observations that are more distant from the fixed point:

$$\hat{f}(\mathbf{z}_0, \mathbf{s}_0, t_0) = \sum_{i=1}^{q} \frac{K_h(\cdot) ln\mathrm{SP}_{it_0}}{\sum_{i=1}^{q} K_h(\cdot)} - \frac{\sum_{i=1}^{q} \varepsilon_{it_0} K_h(\cdot)}{\sum_{i=1}^{q} K_h(\cdot)} \tag{A4}$$

where the weighting function, $K_h(\cdot)$, is defined such that greater distances (e.g., larger values for $S_i - \mathbf{s}_0$) imply lower values for *K*; *h* is bandwidth, a set of parameters that govern the selection of points "close to" the target vector.[1]

[1] Equation (A4) is the well-known Nadaraya-Watson (NW) smoother. See Clapp (2004) for details on the choice of the kernel weighting (i.e., density) function. Experts in this field have found that the choice of bandwidth is much more important than the choice of a kernel density function.

Bandwidth selection is a trade-off between high variance (bandwidth is too small) and high bias (bandwidth is too large). This paper uses a cross validation method for bandwidth selection: See Wand and Jones (1995, Chapter 4). Locally adaptive bandwidths are allowed by increasing bandwidth until 20 observations are within one bandwidth of the fixed point.

The number of variables in the local polynomial regression (LPR) should be small because of the curse of dimensionality. Therefore, we estimate equation (A1) and obtain partial residuals, *partres*$_{it}$, by subtracting $Z_i \hat{\alpha}$ from $lnSP_{it}$. Then the LPR part of the local regression model (LRM) is applied to *partres*$_{it}$.

Equation (A4) is a special case of local polynomial regression (LPR), given a specific point in space and time, $\mathbf{x}_0 = (\mathbf{s}_0, t_0)$, the data, $\mathbf{X}_{it} = (S_i, t_i)$ and $Y_{it} = partres_{it}$. Local polynomial regression now takes the form of equation (A5):[2]

$$Y_{it}(\mathbf{x}_0) = \beta_o + (\mathbf{X}_{it} - \mathbf{x}_0)\beta_1 + (\mathbf{X}_{it} - \mathbf{x}_0)^2 \beta_2 + \ldots + (\mathbf{X}_{it} - \mathbf{x}_0)^p \beta_p + \varepsilon_{it} \qquad \text{(A5)}$$

Here, the β_j (j=1,…,p) are column vectors with number of elements equal to the columns of $\mathbf{X}_{it}$: β_0 is a scalar.[3] Note that, when $\mathbf{X}_{it}$ equal $\mathbf{x}_0$ then equation (A5) reduces to β_0, the parameter of interest. Thus, LPR fits a surface to the Y-values conditional on the values of $\mathbf{x}$ given by $\mathbf{x}_0$:

[2] The exponents in equations (A5), (A6) and (A8) are taken element-by-element.

[3] The parameters other than β_0 allow for curvature around $\mathbf{x}_0$; a weighted average of neighboring points, equation (A4), would ignore curvature. Also, comparing equations (A6) and (A3) show how LPR takes local averages.

E.g., $\mathbf{x}$ is a grid of equally spaced points that span the data; the level of Y is estimated conditional on each knot of the grid.

The treatment of time is much more flexible in equation (A5) than it would be in the OLS model, equation (A1). LPR treats time as an addition to the spatial dimension: that is, we grid time as finely as the data permit at each point in space. For example, to estimate the value function at 10 points in time, and at each point of a 30x30 spatial grid, we need 9,000 regressions. Each estimator gives high weight to observations that are nearby in space and time and lower weight to those that are farther away.

Kernel weights are applied when estimating equation (A5):

$$\text{Min}(\hat{\beta})\sum_{i=1}^{n}\{Y_{it}-\beta_0-\ldots-(\mathbf{X}_{it}-\mathbf{x}_0)^p\beta_p\}^2 K_h(\mathbf{X}_{it}-\mathbf{x}_0) \tag{A6}$$

where the weights are applied to each of the variables including the constant term (the vector of ones).[4] Applying OLS, the parameters estimated using equation (A6) can be estimated as follows:

$$\hat{\beta}(\mathbf{x}_0)=(\mathbf{X}_{\mathbf{x}}^{\mathbf{T}'}W_x\mathbf{X}_x)^{-1}\mathbf{X}_x^{T'}W_xY \tag{A7}$$

$$\mathbf{X}_x=\begin{bmatrix}1 & \mathbf{X}_{1t}-\mathbf{x}_0 & \cdots & (\mathbf{X}_{1t}-\mathbf{x}_0)^p\\ \vdots & \vdots & \ddots & \vdots\\ 1 & \mathbf{X}_{nt}-\mathbf{x}_0 & \cdots & (\mathbf{X}_{nt}-\mathbf{x}_0)^p\end{bmatrix} \tag{A8}$$

$$W_x=diag\{K_h(\mathbf{X}_{1t}-\mathbf{x}_0),\ldots,K_h(\mathbf{X}_{nt}-\mathbf{x}_0)\}\ . \tag{A9}$$

[4] The metric for time is different from that for space (and also different for structural characteristics). Cross-validation (CV) is used to select optimal bandwidths: If CV indicates that more distance is needed for estimation at any knot, than a larger bandwidth will be chosen in the spatial dimension. This addresses a concern raised by Pavlov (2000).

This regression is repeated for each point on the $\mathbf{x}_0$ grid. It is important that $\hat{\beta}_0$ is the main parameter of interest at each knot because the terms in the polynomial collapse to zero at the knot. I.e., $\hat{\beta}_0$ is the smoothed value for the dependent variable.

LPR is a weighted OLS regression at the point $\mathbf{x}_0$, so we can test hypotheses on the $\hat{\beta}_0$'s by assuming that they are multivariate normal with the following covariances:

$\left(Var - Cov(\hat{\beta})\right) = \left(\mathbf{X}_x^T W_x \mathbf{X}_x\right)^{-1} \mathbf{X}_x^T W_x \mathrm{V} \mathrm{W}_x \mathbf{X}_x \left(\mathbf{X}_x^T W_x \mathbf{X}_x\right)^{-1}$,where V is a diagonal matrix of variances for ε_i.

The algorithm for estimating Robinson coefficients with backfitting

A backfitting method is used to estimate the coefficients in the linear part of the model. This is done to obtain orthogonality between the coefficients on structural characteristics and the locational characteristics. The backfitting method can be summarized as follows:[5]

- Use LPR to calculate *ln*SP given time and space (*S*,*t*). The structural characteristics (*Z*) are not used in this regression. The predicted values are E(lnSP| *S*,*t*).
- For each element in the vector of structural characteristics (*Z*) use LPR to calculate predicted values: E(*Z* | *S*,*t*).
- Subtract these estimates from the original values to determine sales price (lnSP*) and value of structure (*Z* *) independent of time and space: lnSP*=lnSP-E(lnSP| *S*,*t*) and *Z**= *Z* - E(*Z* | *S*,*t*).
- Use OLS to estimate the parameters of the structural characteristics (the "Robinson coefficients") by regression lnSP* on *Z**: lnSP*= *Z* * α_R +e.

[5] The *i*,*t* subscripts are suppressed for ease of notation.

- Iterate the previous steps until there is no economically significant change in the estimated Robinson coefficients, $\hat{\alpha}_R$.

The algorithm for estimating location values

- Calculate partial residuals by subtracting the estimated values of the structural characteristics from lnSP: *partres*=lnSP- $Z\hat{\alpha}_R$.
- Use LPR to estimate location values by regressing partial residuals on time and space values: $\hat{q}(\mathrm{S},t)$ =E(*partres*| *S*,*t*)

Table 1: Descriptive Statistics, Denver Single Family Home Sales, 2003-2010

Variable	Mean	Std. Dev.	Variance	Minimum	Maximum
Sale Price (Log)	12.4315	0.5486	0.3010	6.9078	15.4249
Yr2003	0.1393	0.3463	0.1199	0	1
Yr2004	0.1530	0.3600	0.1296	0	1
Yr2005	0.1528	0.3598	0.1295	0	1
Yr2006	0.1384	0.3453	0.1192	0	1
Yr2007	0.1232	0.3286	0.1080	0	1
Yr2008	0.1112	0.3144	0.0989	0	1
Yr2009	0.0955	0.2940	0.0864	0	1
Yr2010	0.0865	0.2811	0.0790	0	1
No of Bedrooms	3.1587	0.8403	0.7061	1	13
No of Full Baths	2.2924	0.8853	0.7838	1	12
No of Half Baths	0.3259	0.4938	0.2439	0	5
No of Fireplaces	0.7669	0.7313	0.5348	0	10
Garage Dummy	0.9103	0.2858	0.0817	0	1
Basement Dummy	0.8031	0.3977	0.1581	0	1
Stories Dummy	0.4839	0.4997	0.2497	0	1
Adams County Dummy	0.1900	0.3923	0.1539	0	1
Denver County Dummy	0.2268	0.4188	0.1754	0	1
Douglas County Dummy	0.1782	0.3827	0.1465	0	1
Arapahoe County Dummy	0.2114	0.4083	0.1667	0	1
Jefferson County Dummy	0.1935	0.3951	0.1561	0	1
Longitude	-104.9384	0.1441	0.0208	-105.4648	-103.765
Latitude	39.6936	0.1440	0.0207	39.1305	40.242
Longitude Squared	11012.0871	30.2435	914.6665	10767.1109	11122.82
Latitude Squared	1575.6048	11.4253	130.5373	1531.1990	1619.42
Lat*Lon	-4165.3879	16.7768	281.4625	-4206.4213	-4098.49
Age	34.0694	27.3488	747.9552	0	145
Age Squared	1908.6075	2870.4277	8238049	0	21025
Land Square Feet (Log)	9.0213	0.6916	0.4765	6.2146	18.1084

Observations = 326,744

Table 2: Hedonic Regressions, Denver SFR Home Sales, 2003-2010

Valid cases:	326744	Dependent variable:	Sales Price (Log)
Missing cases:	0	Deletion method:	None
Total SS:	98336.134	Degrees of freedom:	326717
R-squared:	0.591	Rbar-squared:	0.591
Residual SS:	40228.422	Std error of est:	0.351
F(24,178706):	18150.942	Probability of F:	0

				Robinson's Method		
Variable	**Coeff.**	**T-Value**	**P-Value**	**Coeff.**	**T-Value**	**P-Value**
Constant	-4233.666469	-13.527658	0.00			
Yr2003	0.056438	21.222047	0.00			
Yr2004	0.085489	32.724294	0.00			
Yr2005	0.121738	46.576639	0.00			
Yr2006	0.111406	41.845352	0.00			
Yr2007	0.065053	23.879809	0.00			
Yr2008	-0.051856	-18.623067	0.00			
Yr2009	-0.041102	-14.261565	0.00			
No of Bedrooms	0.014671	15.866329	0.00	0.017519	19.144798	0.00
No of Full Baths	0.164688	152.781209	0.00	0.146963	136.717485	0.00
No of Half Baths	0.151836	91.099069	0.00	0.132745	79.822351	0.00
No of Fireplaces	0.170455	161.099640	0.00	0.145072	137.001850	0.00
Garage Dummy	0.170616	71.760975	0.00	0.170824	72.495021	0.00
Basement Dummy	0.138199	80.941398	0.00	0.128803	75.681740	0.00
Stories Dummy	0.014131	8.167266	0.00	0.026279	15.247442	0.00
Adams County Dummy	0.003398	1.021937	.307	-0.04711	-10.518532	0.00
Denver County Dummy	0.183828	61.887987	0.00	0.069461	18.394318	0.00
Douglas County Dummy	0.016086	3.821778	0.00	-0.27454	-38.868442	0.00
Arapahoe County Dummy	-0.021698	-6.430422	0.00	-0.20686	-43.457920	0.00
Longitude	-107.719767	-21.667117	0.00			
Latitude	-72.155186	-15.752523	0.00			
Longitude Squared	-0.799777	-37.149268	0.00			
Latitude Squared	-1.115896	-32.965914	0.00			
Lat*Lon	-1.529108	-41.419968	0.00			
Age	-0.014724	-160.94147	0.00	-0.01629	-164.96856	0.00
Age Squared	0.000122	156.981420	0.00	0.000126	151.223767	0.00
Land Sq Feet (Log)	0.183304	159.507617	0.00	.235341	196.404030	0.00

Table 3: Airport Infrastructure Capital Stocks, Denver International Airport, 2003-2010

(constant (2003) millions of dollars, net of depreciation)

	2003	2004	2005	2006	2007	2008	2009	2010
Airfield	759.3	735.2	710.6	695.0	678.8	664.0	675.6	670.2
Terminal	592.1	581.4	570.7	591.8	596.6	601.3	590.9	589.1
Parking	88.9	88.4	87.3	96.5	131.2	136.0	139.6	138.9
Road, Rail & Transit	104.4	107.1	111.9	112.3	110.4	113.0	111.7	114.3
Other	385.2	379.1	367.8	358.9	348.3	350.6	345.1	349.6
Total	1929.8	1891.3	1848.4	1854.4	1865.3	1865.0	1863.0	1862.1

Table 4: GMM Estimation of One-year Change of Land Value on Airport Infrastructure Capital Stocks (normalized by distance to the airport), Denver International Airport

Note: land values are in real terms

Dependent Variable: One-year change Land_Level
Method: Panel Generalized Method of Moments

Sample: 2005 2010 IF LAT>39.8439 AND LONG<-104.6733
Periods included: 6
Cross-sections included: 53742
Total panel (balanced) observations: 322452
2SLS instrument weighting matrix
Instrument specification: Constant, One-year change_AIRFIELD(-1) One-year change_TERMINAL(-1) One-year change PARKING(-1) One-year change RD_RL_TRN(-1) One-year change OTHER(-1)
Constant added to instrument list

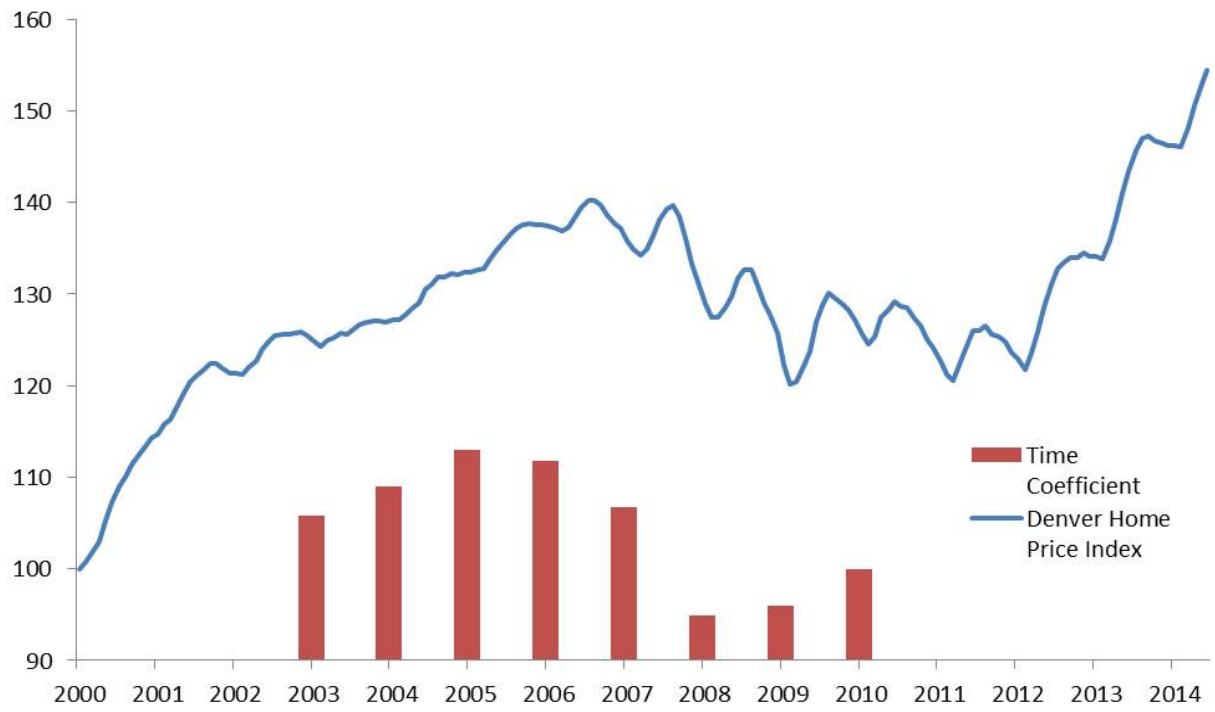
Variable	Coefficient	Std. Error	t-Statistic	Prob.
Constant	-2603.124	1.906908	-1365.102	0.0000
One-year change_AIRFIELD	0.000285	7.02E-06	40.59259	0.0000
One-year change_TERMINAL	0.000339	2.94E-06	115.1157	0.0000
One-year change_PARKING	0.000235	5.08E-06	46.18320	0.0000
One-year change_RD_RL_TRN	0.002581	3.71E-05	69.61141	0.0000
One-year change_OTHER	-0.000302	3.86E-06	-78.24314	0.0000

Effects Specification

Period fixed (dummy variables)

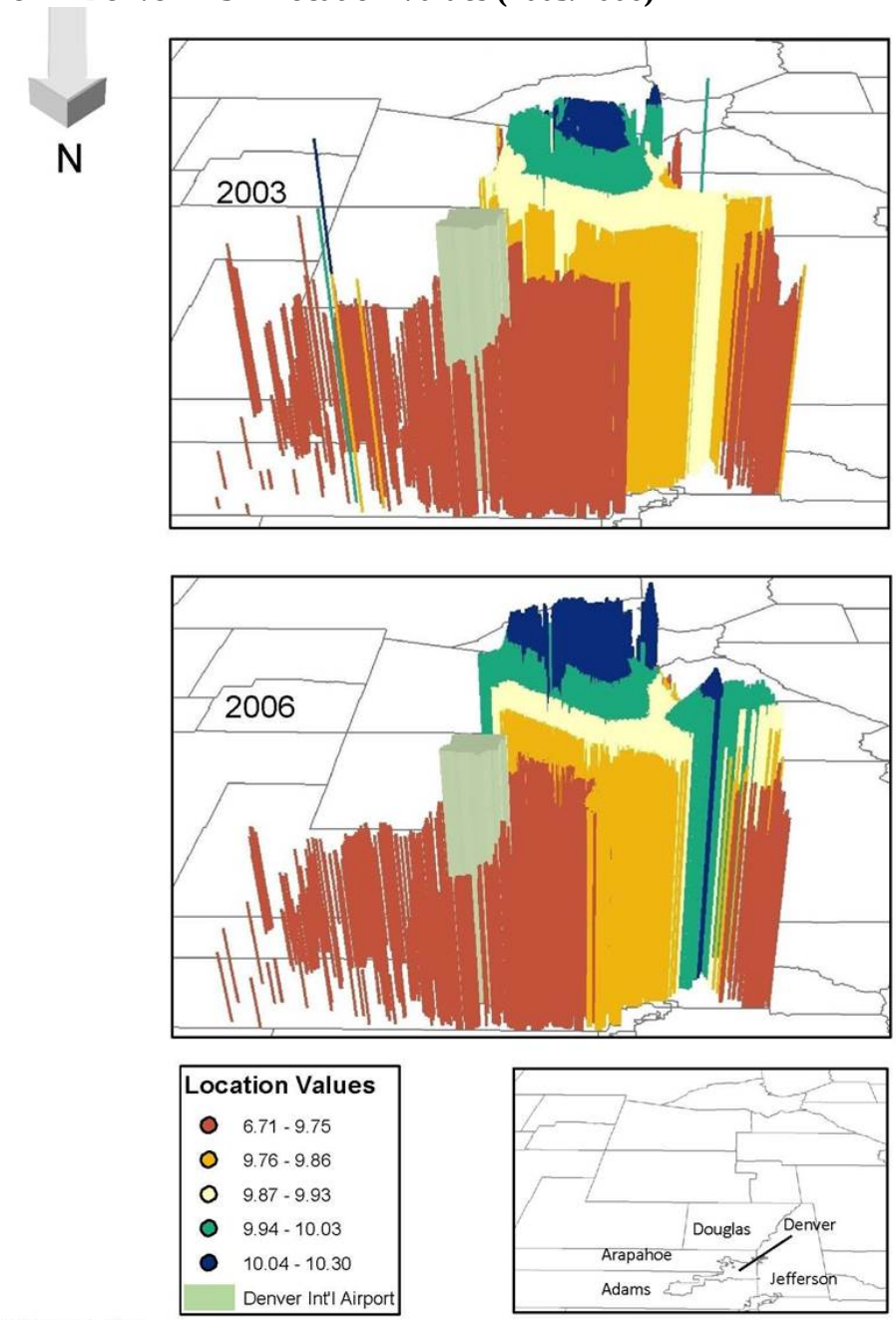
R-squared	0.979117	Mean dependent var	-2321.416
Adjusted R-squared	0.979116	S.D. dependent var	1722.607
S.E. of regression	248.9371	Sum squared resid	2.00E+10
Durbin-Watson stat	0.716351	J-statistic	1.74E-18
Instrument rank	11		

Figure 1 – Single Family Home Sale Prices, Denver



Source: Denver Home Price Index is from Federal Reserve Economic Data (FRED); Time dummy coefficient estimates are obtained from regressions in Table 3, normalizing 2010 (the omitted year) to equal 100.

Figure 2 – Denver MSA Location Values (2003/2006)



* Perspective from north looking south

Figure 3 – Denver MSA Location Values (2006/2010)

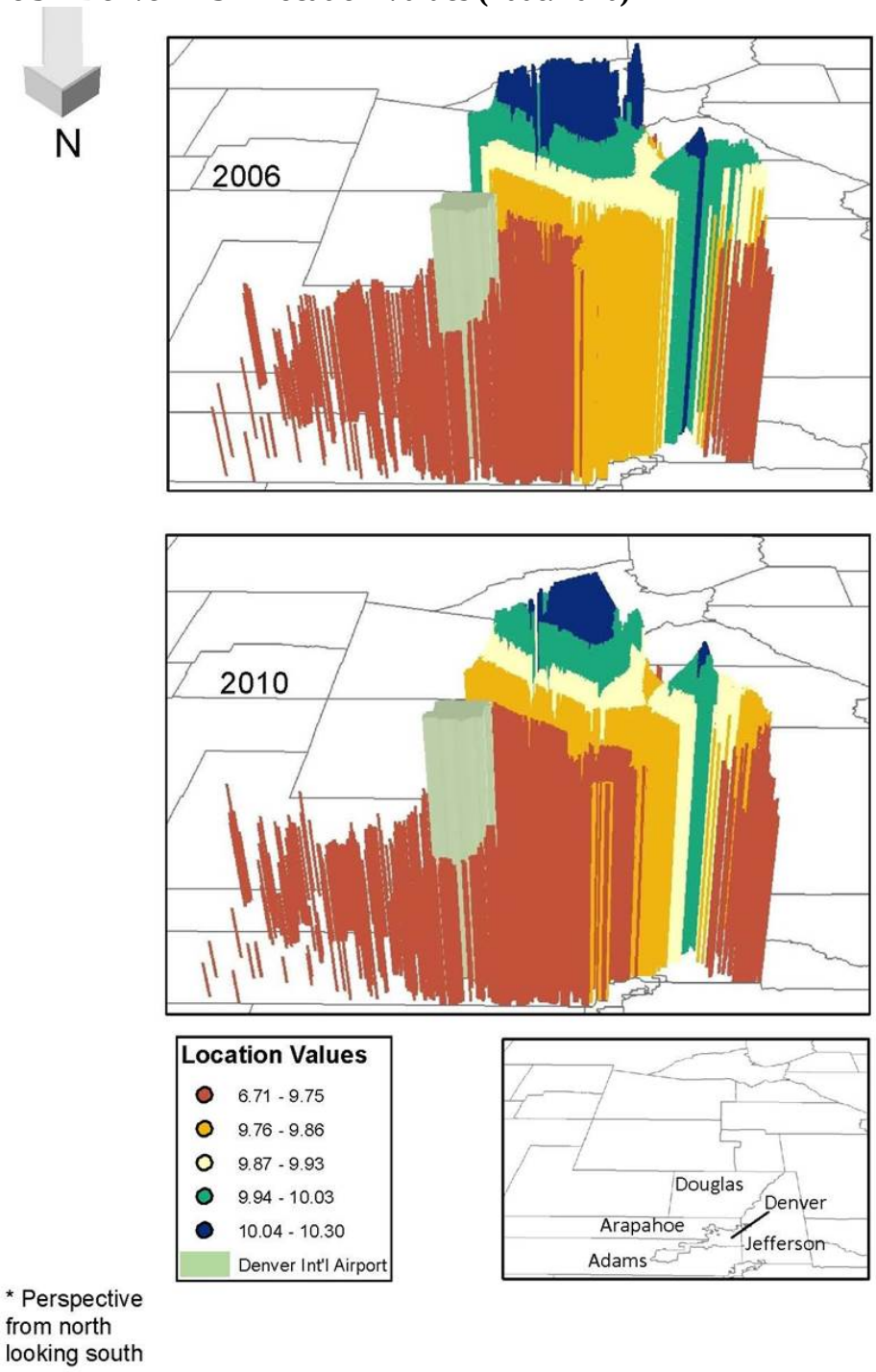
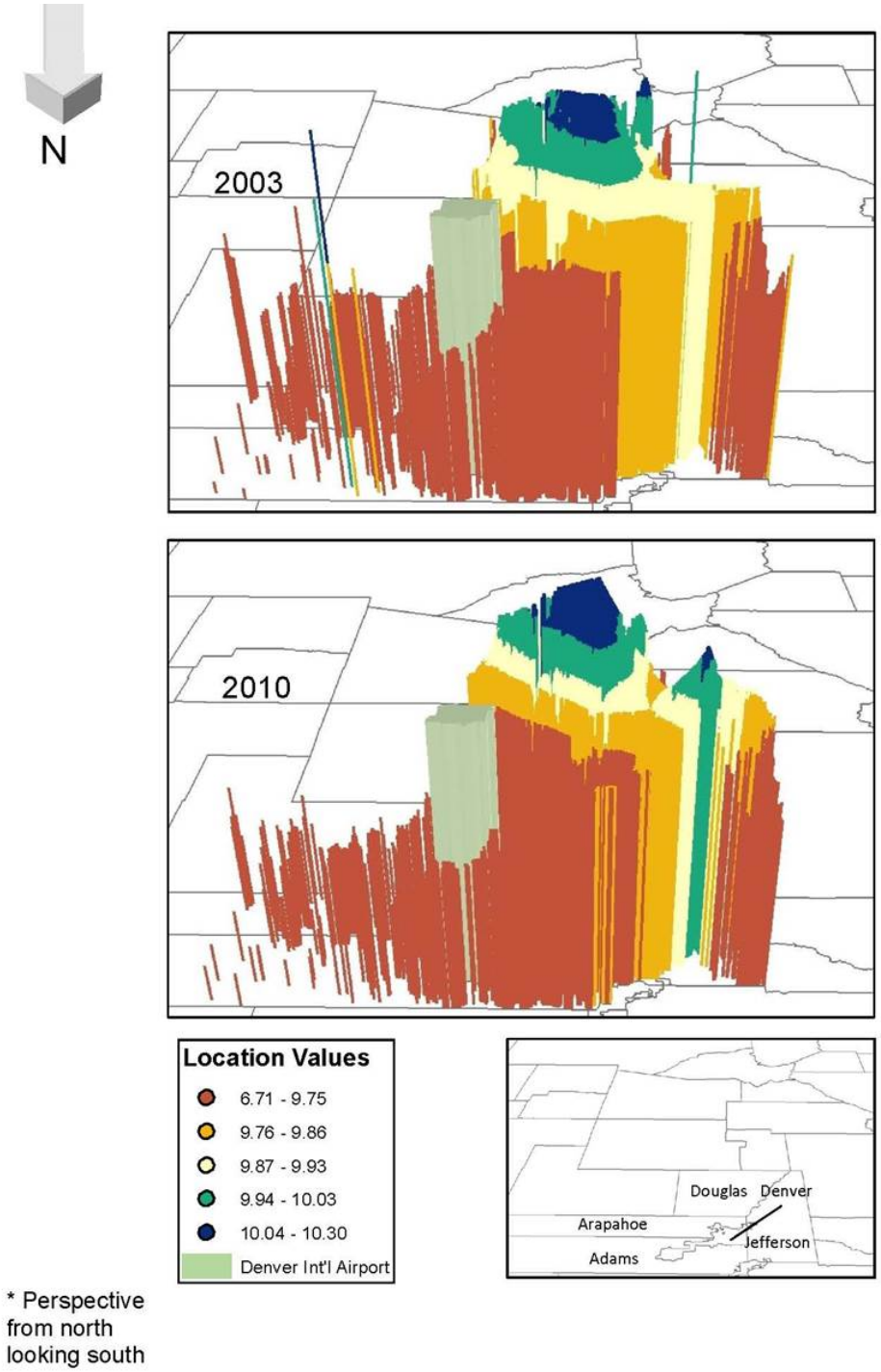
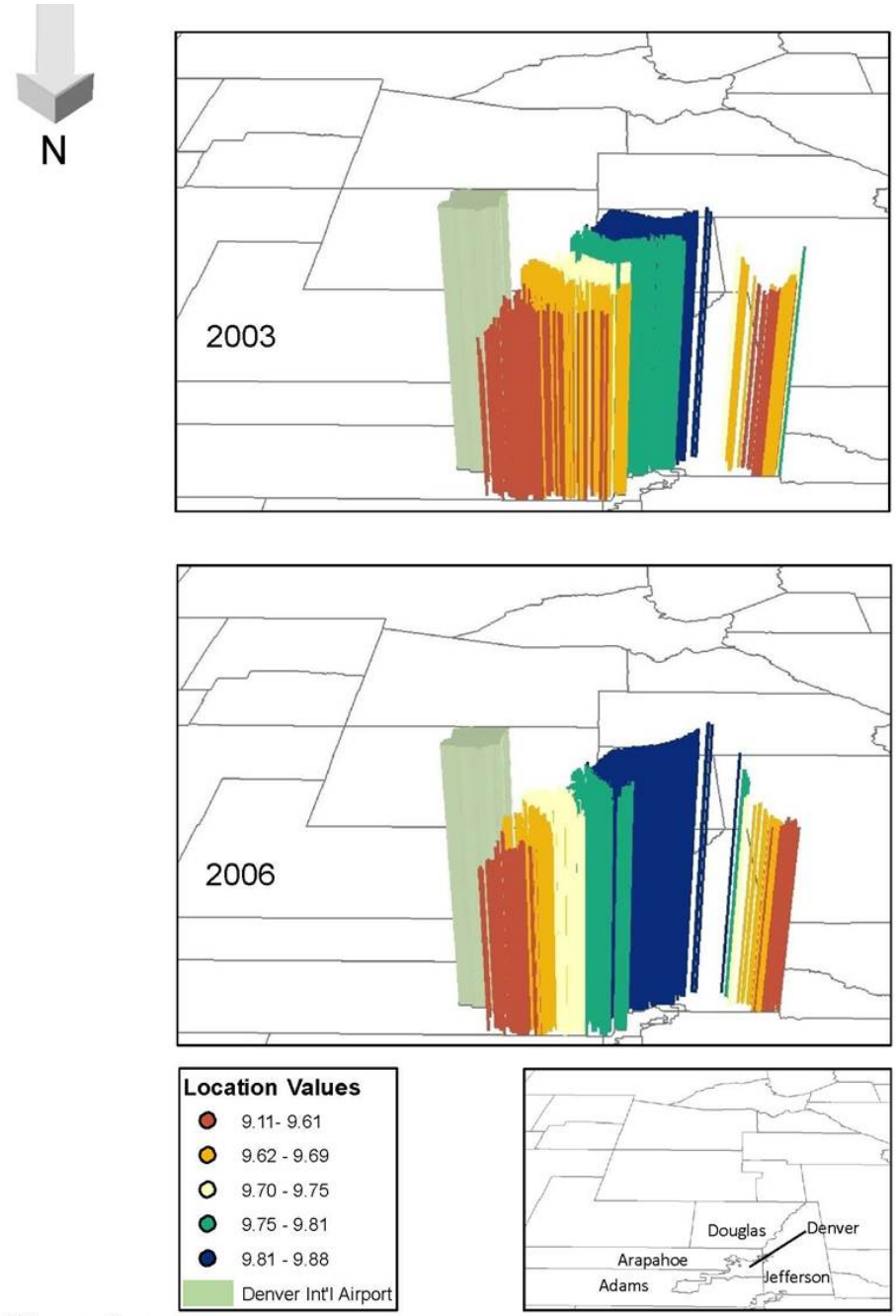


Figure 4 – Denver MSA Location Values (2003/2010)



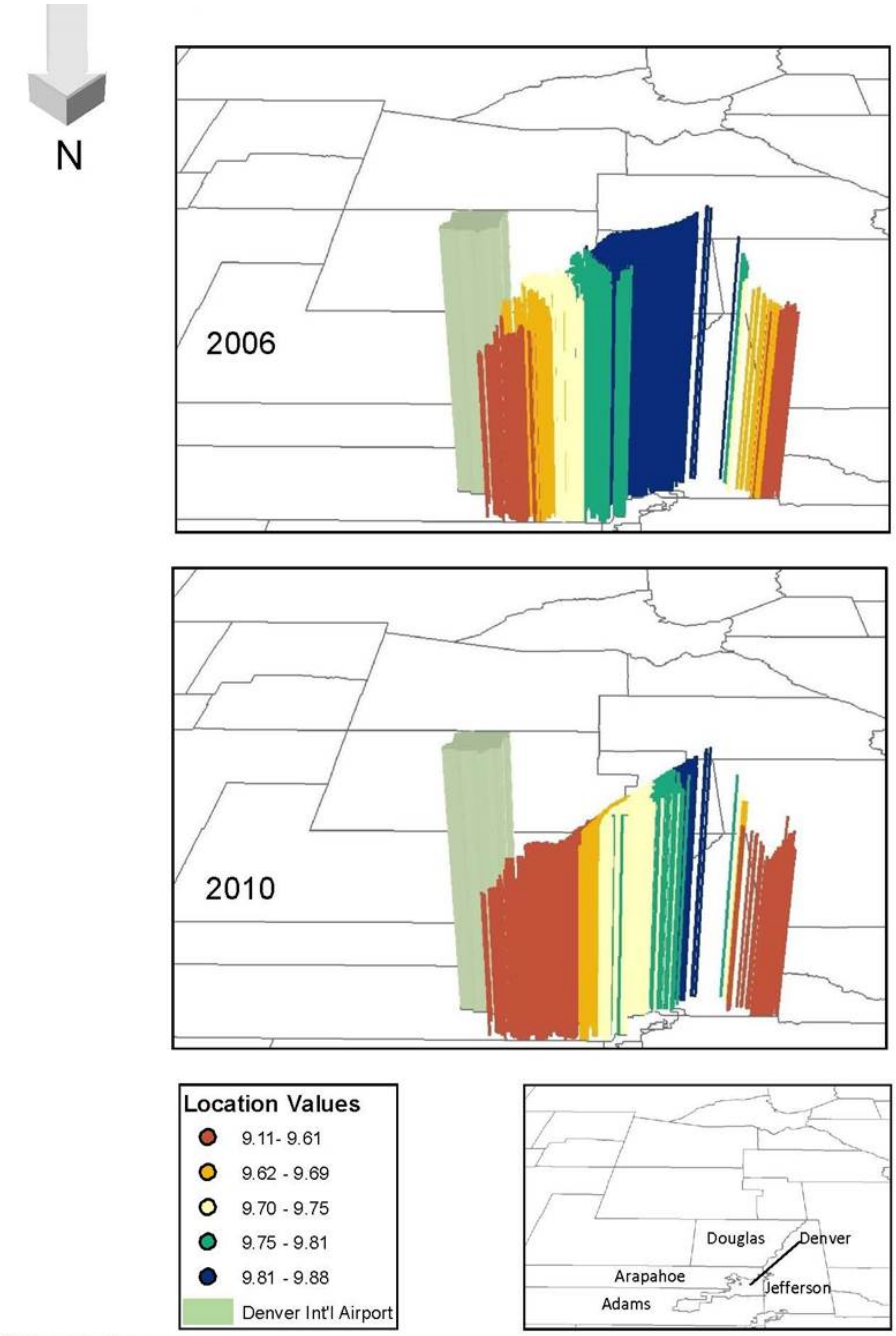
* Perspective from north looking south

Figure 5 – Northwest Portion of Denver MSA Location Values (2003/2006)



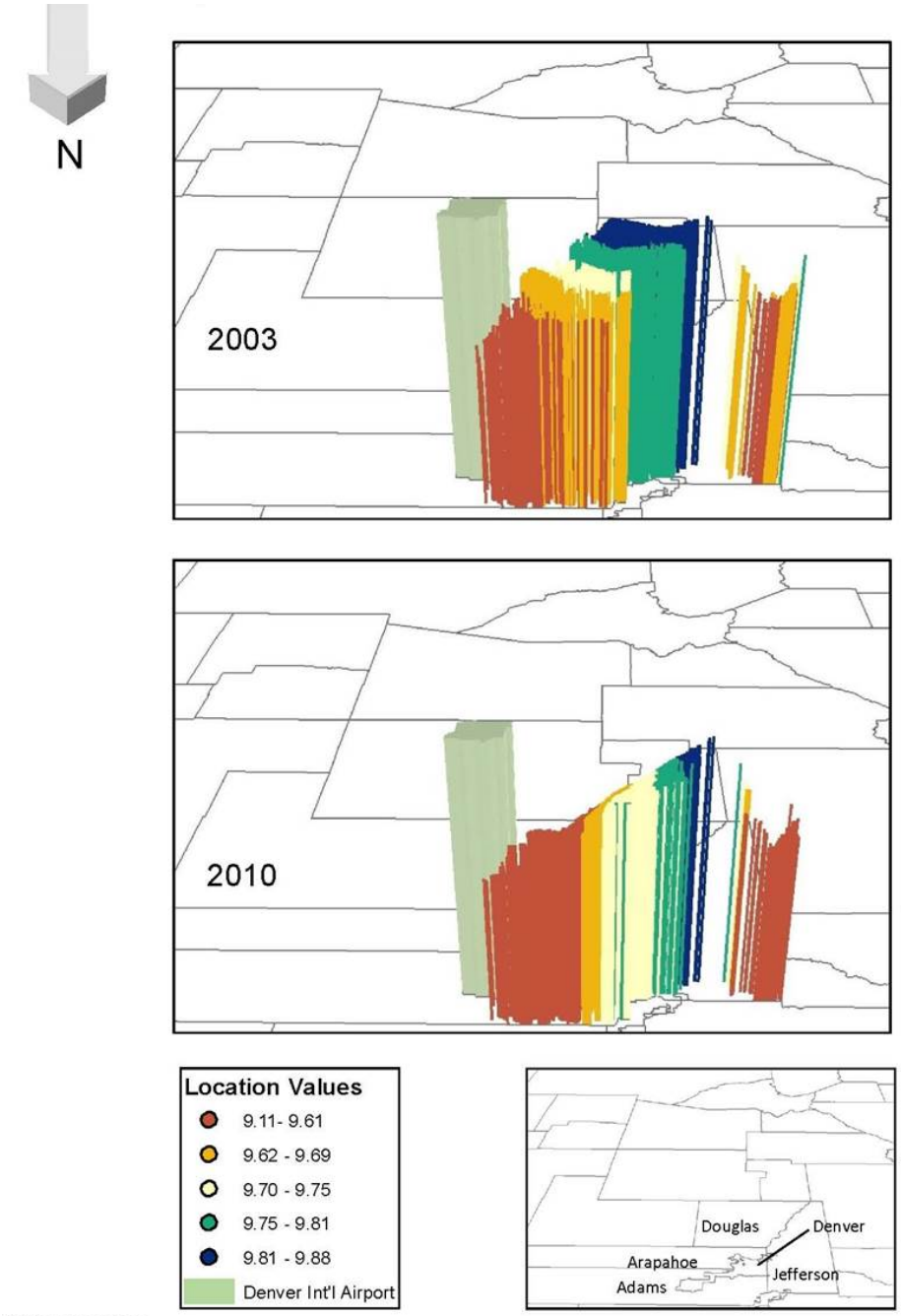
* Perspective from north looking south

Figure 6 – Northwest Portion of Denver MSA Location Values (2006/2010)



* Perspective from north looking south

Figure 7 – Northwest Portion of Denver MSA Location Values (2003/2010)



* Perspective from north looking south