Estimating Land Values using Residential Sales Data

Stanley D. Longhofer and Christian L. Redfearn

Lincoln Institute of Land Policy
Working Paper

Lincoln Institute Product Code: WP09SL1

Abstract

In this research, we attempt to develop a statistically rigorous, but straightforward and tractable method by which vacant land sales on the periphery of a metropolitan area can be used to estimate underlying implicit land values throughout the city, including areas in which there are few or no vacant lot sales. Our methodology involves a two-step process. In the first step, we use single-family residential sales data to estimate a constant-quality price surface using a locally-weighted regression (LWR) model. This standardized property value surface is obtained by pricing an identical housing bundle at every location, where this bundle includes both the typical dwelling characteristics and the typical lot characteristics.

The second step of our analysis involves using vacant land sales in neighborhoods on the periphery to "pin down" the absolute level of the standardized value surface. We do this by imputing the value of the standard dwelling at the locations of the vacant lot sales, and subtract the observed selling prices of the lots to come up with the "value" of the standard dwelling. Subtracting this imputed dwelling value from our standardized price surface provides us with estimates of absolute land values at each location throughout the city. The land surface we estimate is consistent with our understanding of the local market from which are data are obtained (Wichita, Kansas) and is decidedly non-monotonic, with neighborhoods of significant premium in land prices over others. We also find a number of regularities at odds with the basic urban model in a setting we had anticipated to behave much like the proto-typical monocentric city: polycentric land pricing and land at the periphery varying widely rather than being price according to some nominal "agricultural" use. Though our discussion focuses primarily on the land price recovery exercise, these are potentially interesting findings that speak directly to validity of the basic urban model in its standard form.

About the Authors

Stanley D. Longhofer holds the Stephen L. Clark Chair of Real Estate and Finance in the Barton School of Business at Wichita State University, where he founded the Center for Real Estate in 2000. As Director of the Center for Real Estate, Dr. Longhofer provides research services and educational programs to real estate professionals throughout the region, and is the author of the Center's annual *Wichita Housing Market Forecast*. His research on real estate, mortgage discrimination, financial contracting, and bankruptcy has been published in a number of leading academic journals. Currently, Dr. Longhofer's research focuses on housing market dynamics. Prior to coming to Wichita State, Dr. Longhofer spent five years as a financial economist at the Federal Reserve Bank of Cleveland. He also served as a founding member of the Federal Reserve System's Fair Lending Advisory Group, consulting with bank examiners on fair-lending issues.

Barton School of Business
Wichita State University
1845 Fairmount
Wichita, KS 67260-0077
316-978-7120
316-978-3263 (fax)
stan.longhofer@wichita.edu

Christian L. Redfearn is an associate professor in the School of Policy, Planning, and Development at the University of Southern California. He joined the faculty at USC after completing his Ph.D. in economics at the University of California, Berkeley. An urban economist, Prof. Redfearn is engaged in research projects that focus on the evolution of metropolitan land and real estate markets. Specific examples of his current research involve neighborhood evolution, assessment of spatial amenity capitalization, housing price measurement issues in a complex urban setting, trading property rights and externalities in historic preservation districts, the hierarchy of urban land markets, as well as the spatial organization of metropolitan employment and its persistence over long periods of time. He has published recently in the *Regional Science and Urban Economics*, the *Journal of Urban Economics*, the *Journal of Regional Science*, *Real Estate Economics*, and *Environment & Planning A*.

School of Policy, Planning, and Development
University of Southern California
Los Angeles, CA 90089
213-821-1364
213-740-6170 (fax)
redfearn@usc.edu

We gratefully acknowledge financial support from the Lincoln Institute for Land Policy through its David C. Lincoln Fellowships in Land Value Taxation and the USC Lusk Center for Real Estate. We also thank Chris Morlan with the Sedgwick County Appraiser's Office for assistance with our data.

Table of Contents

Estimating Land Values using Residential Sales Data

1. Introduction

The efficiency gains that would accrue from the use of a land tax–as opposed to a tax on a property's entire value, including both land and improvements–are widely recognized. Because land is immobile and its supply is fixed, agents will not modify their investment/consumption decisions in response to a land tax, thereby avoiding the deadweight losses associated with most taxes (Vickrey, 1999).[1] As a consequence, many scholars have suggested that state and local jurisdictions consider incorporating land tax regimes into their overall menu of taxing instruments (Tideman, 1999; Bowman and Bell, 2004)

A practical barrier to the adoption of a land tax, however, is the fact that real estate is a bundled good, containing both land and structures.[2] Most methods for estimating the value of real estate ultimately rely on observing market transactions. Unfortunately, these transactions almost invariably involve the bundled good, not the land and structures separately. As a consequence, it can be difficult to obtain reliable, market-based estimates of underlying land values in developed areas. Without good land value estimates, it will be hard for taxing jurisdictions to implement a land tax. Furthermore, it is worth questioning whether the benefits of a land tax would still exist if land value estimates were not based on true market values.

In this research, our goal is to develop a statistically rigorous, but straightforward and tractable method by which vacant land sales on the periphery of a metropolitan area can be used to estimate underlying implicit land values throughout the city, including areas in which there are few or no vacant lot sales.

Our analysis involves a two-step process. In the first step, we use single-family residential sales data from Sedgwick County (Wichita), Kansas to estimate a constant-quality price surface using a locally-weighted regression model. The key advantage of using locally-weighted regression is its inherent flexibility to recover local pricing of dwelling attributes. By estimating the price function independently at each sale location the Law of One Price is not imposed; rather the data are allowed to reveal pricing. As a result, this methodology avoids the significant omitted variable problems that would otherwise plague the location dummy variable coefficients estimated in a traditional hedonic model. Using the coefficient vectors from the set of local regressions, we estimate a property value surface for a "standardized housing unit" at each and every location in the city.

[1] See Cohen & Coghlin (2005) for a simple explanation of the deadweight costs associated with most taxes and the way that a land tax would avoid these distortions.
[2] There are other reasons for questioning whether a land tax could effectively replace the more general real property tax. See Mills (1998) for a good discussion of these issues.

This standardized property value surface implicitly measures the relative premium for different locations in the city, but does not "pin down" precise land values for each property. That is, the standardized surface is obtained by pricing an identical housing bundle at every location, where this bundle includes both the typical dwelling characteristics and the typical lot characteristics. We recover land prices by fitting this standardized surface to the few land sales we do observe. Although there are few sales of vacant lots in the core, there are many sales at the periphery. Moreover, in many of the neighborhoods on the edge of the city, both vacant lot and completed dwelling sales occur at roughly the same time.

Using this fact, the second step of our analysis involves using vacant land sales in neighborhoods on the periphery to pin down the absolute level of the standardized value surface estimated in the first step. With this calibration, we can theoretically use the standardized value surface to estimate land values throughout the city, including in those neighborhoods without vacant lot sales.

The key assumption in taking these steps is that local amenities and drawbacks are capitalized into land and not the structure. The locally-weighted regressions used in estimating the standardized property price surface allows a standard structure to be priced everywhere, which explicitly holds constant the physical characteristics the improvements to the property at each location. Spatial variation in the price surface of this standard dwelling, therefore, then must come from capitalization of local amenities and drawbacks into the underlying land prices.

In order to "pin down" the relative value surface, we use our locally-weighted regression results to estimate the value of the standard dwelling at the location of our vacant lot sales. We then adjust these standardized values to reflect the actual characteristics of the lots, and finally subtract the observed sale prices of these lots to extract the implied value of the standard improvements on each lot. We then estimate the value of the standardized dwelling (exclusive of location) as the average of their imputed values at each vacant lot sale location.

Using this standardized dwelling value we are able to make use of amenity capitalization in dwelling sales to recover land prices throughout the metropolitan area. The land surface we estimate is consistent with our understanding of the Wichita market and is decidedly non-monotonic, with neighborhoods of significant premium in land prices over others. We also find a number of regularities at odds with the basic urban model in a setting we had anticipated to behave much like the prototypical city: polycentric land pricing and land at the periphery varying widely rather than being price according to some nominal "agricultural" use. Though our discussion focuses primarily on the land price recovery exercise, these are potentially interesting findings that speak directly to validity of the basic urban model in its standard form.

In the next section we discuss the challenges associated with estimating lot values using single-family residential sales data, as well as recent related research. We contrast this research with our own empirical strategy. Section 3 contains a discussion of our empirical results, while Section 4 concludes by discussing the broader implications and caveats of our analysis.

2. Estimating Land Values using Residential Sales Data

Given the long academic interest in land value taxation, it is somewhat surprising that relatively little empirical work has been done to address the practical challenges in accurately decomposing land and building values from residential sales data. Brunori and Carr (2002) review state and local practices regarding the separate valuation of land and improvements for property tax purposes, and find that nearly all local property tax assessor offices value land and improvements separately, regardless of whether they are required by state law to do so. Although these authors report that the majority of these offices indicate a high degree of confidence in their land value estimates, they do not explore how the assessors actually derive these values.

Bell and Bowman (2006) address this deficiency, using nine local assessor offices as case studies to examine how land values are estimated in practice. They conclude that assessor offices generally use some combination of (1) an "abstraction" method, in which land values are calculated as the residual of total property value less the depreciated value of the improvements (an application of the cost approach of appraisal); (2) an "allocation" method, in which land values are simply estimated as a fixed percentage of overall property value; and (3) a "contribution" approach, which essentially involves some type of multiple regression analysis of vacant lot and/or improved property sales. They conclude that most assessors apply a combination of all three methods, along with a great deal of subjective judgment.

Gloudemans (2000, 2002) and Gloudemans, Handel and Warwa (2002) all attempt to use non-linear regression (hedonic) techniques to estimate land values from improved parcel sales data. Specifically, these papers model total property value as additive in its land and building components but multiplicative within the characteristics of each of these components. Because land and building values are separable in this model, it is possible to use the regression coefficients to separately estimate land and building values. They conclude that their land value estimates perform quite well relative to standard computer assisted mass appraisal benchmarks (the average assessed value to sale price ratio and the coefficient of dispersion of this ratio). In the end, these articles conclude it is feasible to use multiple regression techniques to estimate land values even when there are few or no vacant sales. As we discuss further below, however, these articles do not address the extent to which their land value estimates are biased by omitted physical structure characteristics that are correlated with geography.

Ashley, Plassmann and Tideman (1999) is in many respect most closely associated with our analysis. They estimate total property values of commercial properties in downtown Portland using a simple hedonic specification. They then use a quadratic spatial smoothing technique to estimate land value. In contrast, our work uses locally-weighted regressions to estimate a total relative property value surface over a city, which we then "pin down" to actual values based on actual land values at the periphery of the city.

The Challenges with Hedonic Land Value Estimates

In order to understand the motivation behind our analysis and how it compares with prior research, it is worth reviewing the basics of hedonic property value regression and the challenges these assumptions imply for estimating land values from these regressions. In its most basic form, the hedonic pricing equation takes a form such as

$$V = \gamma_0 + \sum_l \gamma_l X_l + \sum_n \delta_n I_n + \sum_s \gamma_s Z_s \quad , \qquad (1)$$

where V is the value (sale price) of the property, X is an l-dimensional vector of land characteristics (including lot size, street type, lot amenities, etc.), I is an n-dimensional vector of neighborhood dummy variables, and Z is an s-dimensional vector of structure characteristics (building size, number of bedrooms, construction quality variables, etc.). It is important to note that Z may include a number of interaction effects between structure and lot characteristics. For example, a very small lot size might affect the value of additional square feet of the structure. Similarly, the value of an added bathroom or more square feet may differ based on the neighborhood in which the property is located.[3] Conceptually, we can decompose the value of a parcel into its lot value, L, and building value, B, with $V = L + B$. Using the physical lot and building characteristics from (1) above, we can write

$$L = \alpha_0 + \sum_l \gamma_l X_l + \sum_n \delta_n I_n \qquad (2)$$

and

$$B = \beta_0 + \sum_s \gamma_s Z_s \quad , \qquad (3)$$

with $\gamma_0 = \alpha_0 + \beta_0$. In principle, one would like to apply the hedonic estimates from (1) to (2) and (3) to decompose the total property value into land and building values. Upon inspection, a number of problems with this approach become evident. Most obviously is the challenge inherent in allocating the intercept term γ_0 into its land and building components, α_0 and β_0. This is not just a theoretical problem; the constant term in direct regressions using sale price as the dependent variable can often exceed any reasonable estimate of total land value, indicating that it captures both land and building value components.

A second problem is apparent from the use of value (sale price) as the dependent variable in (1). It turns out that the natural log of price is used as the dependent variable in most hedonic regressions. Not only does this transformation address the heteroskedasticity concerns so prevalent with housing data, it also results in regression coefficients that have a more natural and intuitive interpretation. The valuation model that underlies the log-linear specification, however, is multiplicative, not additive. As a result, land and

[3] We have assumed here that structure characteristics do not affect underlying land values. The intuition behind this is that land values reflect the highest-and-best use of the site. To the extent that the physical characteristics of the structure are not optimal for the site, this affects the value of the building, not the land itself. In an extreme case, this could imply that the building value is negative (up to the value of demolition and site preparation costs) but the site value is still its value as vacant land in its highest-and-best use.

building values are not separable in a traditional log-linear regression model, even if the "constant term" problem could be addressed.[4]

The most serious concern with using traditional hedonics to infer land values, however, is the likelihood that the estimated neighborhood coefficients, δ_n, will be biased because of a failure to accurately model the building value characteristics, Z_s, in (1) above. The physical housing structures in most residential neighborhoods are generally quite homogeneous. For example, homes within a given neighborhood are likely to have similar sizes, floor plans, and construction materials, reflecting the vintage of when they were built. Unless an extensive number of neighborhood/building characteristic interaction terms are incorporated into Z_s, these variables will inevitably be highly correlated the neighborhood indicator variables, I_n, suggested the potential for substantial bias in the estimated δ_n.[5]

In many hedonic applications, this may not be an inordinate problem, as the neighborhood dummy variables are implicitly included to control for just such vintage effects. If the focus of one's analysis, however, is to use these estimates to infer land values, potential omitted variable bias becomes a major concern. It is worth noting that this omitted variable bias is just as likely to exist in the non-linear regression models employed by Gloudemans (2000, 2002) and Gloudemans, Handel and Warwa (2002) as it is in the simple formulation given in (1) above.[6]

This need to include a large number of interaction terms in hedonic regressions is really a reflection of the fact that global pricing of attributes does not hold in urban areas with housing submarkets. Goodman and Thibodeau (1998, 2003, 2007) have demonstrated the existence of submarkets and the significant differences in the pricing of housing characteristics with urban areas. Redfearn (2009) showed that the Law of One Price could be easily rejected using data from Los Angeles. McMillen and Redfearn (2008) demonstrate that not only do prices of attributes vary spatially, but also they vary in a way that is consistent with rational microeconomic behavior. In light of these findings, we abandon standard hedonic analysis in favor of a more flexible approach to controlling for quality differences across dwellings.

Our approach is based on work by Cleveland and Devlin (1979) and is called locally-weighted regression (LWR). LWR was first used in a real estate context by Stock (1981)

[4] As discussed above, Gloudemans (2000, 2002) and Gloudemans, Handel and Warwa (2002) attempt to address this problem using non-linear regression techniques, explicitly modeling property values as additive in land and building values but multiplicative within the components of these values.
[5] It should also be noted that data limitations may make it virtually impossible to include the requisite interaction terms, as there may be many neighborhoods with insufficient sales to permit estimation of both location fixed effects and interaction effects.
[6] Indeed, the fact that Gloudemans (2002) includes a "vacant land" factor and find that vacant parcels are worth 30 percent less than otherwise identical improved parcels suggests that an omitted variable bias is a significant problem for their land value estimates.

and Meese and Wallace (1991). The basic notion is that the implicit pricing of housing characteristics occurs locally, within submarkets. That is, the Law of One Price holds where buyers pursue similar dwellings, forcing sellers to adjust pricing accordingly. Because housing is a bundled good, there is little in the way of market forces to impel prices to be constant across all dwellings. Indeed, if prices for pools were "too high" in one neighborhood, there is no practical way in which owners of pools where prices for them are low to trade them in other markets. Hence, local completion exists and local pricing should be consistent.

The LWR begins with the identification of an observation and selects "neighboring sales" to be included in a regression that will estimate implicit prices for that house. "Neighboring" sales needn't be just those that are closest physically; in this work, "close" is defined not only across latitude and longitude, but also size and date. In this way, the typical buyer search process is captured: buyers look at a particular point in time, within neighborhoods, at dwellings of a particular size. The observations used in the local regressions presented below are those that are the 70 "closest" observations in this combined measure of similarity.[7] The weighting of these observations is by a tri-cubic kernel, so that weights decline in an accelerating manner. The weights are scaled so that the furthest observation within the 70 observations gets zero weight.[8]

The LWR are run at every dwelling sale observation – at each point, a set of implicit prices is estimated. Rather than a single price estimate, a surface of prices estimates is obtained. This seems impossible at first glance – how can N observations be used to estimate N×(k+1) parameters (where k is the number of covariates in the local regressions)? This is possible because of the imposed smoothness of the local regressions. As the number of observations in the local regressions is expanded, the resulting surface gets smoother. Indeed, OLS can be thought of a local regression with a uniform weighting kernel that uses all the observations; at each point, the same parameter estimates would be recovered. As such, the LWR nests the hypothesis that implicit prices are uniform throughout the market – the data can reveal this regularity, it is not imposed.

Given the local estimates, a standard dwelling can be priced at each location. In this way, quality is held constant. The resulting price surface is that of the standard dwelling everywhere within the geographic support of the data. Because the standardized dwelling structure is the same at all locations, the resulting values become a relative land value surface shifted by the value of the standard dwelling improvements. Of course, we never observe this structure value. Rather, we must use the sale prices of unimproved lots to calibrate the height of the value surface, backing into the constant value of the standard dwelling. If our assumption that all amenity capitalization is correct – that capitalization of amenities is into land and not structures – then the pattern of land prices should be echoed in the standardized price surface. Subtracting off this constant value of the

[7] The specific distance measure is the Euclidean distance between dwellings across latitude, longitude, size in square feet, and sale quarters. Each variable is normalized to have mean 0 and standard deviation of 1 to prevent the natural scaling of the variables to cause over-weighting of any one of the variables.
[8] The specific kernel seems to have little impact on any of the results presented in this paper. The same is true of the specific elements of the measure of closeness, so long as latitude and longitude are included.

standard improvements, therefore, provides with an estimate of the land price surface for Wichita.

3. Data and Empirical Results

We conduct our analysis using data on residential sales in between 2003 and 2005 provided by the Sedgwick County (Kansas) Appraiser's Office. Sedgwick County is the home to Wichita, the largest city in Kansas and the largest MSA contained entirely within the state. Although the Wichita MSA contains four counties, the periphery of the city itself lies within Sedgwick County and primarily consists of undeveloped agricultural land. Indeed, in many respects Wichita approximates the prototypical "flat featureless plain" of urban economic theory, with a perfectly elastic supply of land and no natural or legal barriers to new development. At the 2000 census, Wichita's population was 344,284, a 9.75 percent increase since the 1990 census.

Vacant lot sales were identified as parcels without a dwelling or those specifically identified as a "land" sale; only lots with a land use identified as "single-family residential" or "vacant residential land" were included in the analysis. We eliminated parcels larger than one acre in size, parcels that sold for more than $5 per square foot, as well as a handful of parcels that were spatially disjoint from the rest of the data used in the analysis. After imposing these restrictions, 2,389 vacant lot sales remained in our data set. Table 1 shows the summary statistics for these lot sales.

Improved parcel sales were identified in a similar fashion. In addition to the restrictions place on vacant lot sales discussed above, we eliminate properties with less than 400 square feet of finished living area, certain irregular architectural styles (earth contact homes, twin homes, cottages, etc.), and those with a sale price exceeding $30 per square foot of land. Our final improved parcel sample contains 18,176 sales used in the local regressions. Of these, 6,178 sales (those occurring in 2005) are used to estimate our 2005 price surface. Summary statistics for the entire sample of sales are shown in Table 2. An expositional description of all the variables used in our analysis is provided in Table 3. The challenge facing this line of research can be clearly seen in Figure 1, which shows the spatial distribution of the improved property sales (blue dots) and vacant lot sales (orange dots) in our cleaned data. Of particular note is the dearth of vacant lot sales in the entirety of the core of Wichita. Indeed, there are only a handful of parcel sales not at the absolute periphery. Moreover, it is worth questioning whether those lot sales that do occur in the developed areas of the city really provide a good estimate of residential land values; many of these sales are to adjoining property owners or have been sold for purposes other than to build a single-family house.[9]

We begin our empirical analysis by running traditional hedonic regressions on the improved property sales in our sample, running separate regressions for six sectors of the city (defined as the four inner quadrants as well as the far east and west sides of the

[9] It is worth noting that in Wichita it is virtually unheard of to purchase a developed property with the intent to tear down and rebuild a new home on the lot. The periphery of the city is so close to the core that there is little benefit to doing inner-city tear downs.

city).[10] The results, shown in Table 4, demonstrate the challenge associated with trying to estimate land values by decomposing a hedonic equation such as (1) above. Of particular note is the wide variation in the estimated price elasticities of additional above-grade living area (ranging from 0.36 in the Northeast sector to 0.64 in the West sector), lot size (0.01 in the Northeast and Northwest sectors vs. 0.11 in the Southwest sector), age (-0.03 in the East and West sectors vs. -0.14 in the Southwest sector), and the impact of the architectural style of the home (a ranch house commands a 21 percent premium in the West sector, while it has a only an 11 percent premium in the Southeast sector). These regression results show that the shadow prices of the physical characteristics of houses can vary substantially across a metropolitan area, suggesting that an incredibly high number of interaction variables would be required to decompose land and building values hedonic regression coefficients.

Figures 2 and 3 show the spatial variation in the age and above-grade living area, respectively, of the property sales in our sample. As one would expect, the oldest and smallest homes are located in the core area of the city, while the homes are newer and larger on the periphery. Similar spatial variation can be seen across other housing characteristics as well. Given the wide variation in the shadow prices of these characteristics across different parts of the city – along with the paucity of vacant lot sales in the developed parts of the city – it becomes evident that attempts to use traditional hedonic methods to isolate the underlying land values of the improved parcel sales would be nearly impossible.

Indeed, it is easy to reject the assumption that these data should be pooled and that one set of metropolitan shadow prices should be imposed on dwelling characteristics. While the six regional regressions make this clear, even they pool inappropriately; these regressions provide average effects, while uncovering true land values demands local pricing. Our LWRs use the same variables, but on a local sample around each target dwelling. The "window-size" determines how large the local sample will be; while the "kernel" determines how much weight each of the local sample observations receives in the local regressions. In our regressions, we use a fixed window size of 70 observations – the 70 "closest" observations. We use a distance metric to reflect search, in which households looking at similar dwellings drive attribute prices to a local Law of One Price. Distance is given by:

$$d_{ij} = f(\text{x - coordinate, y - coordinate, living area, bedrooms, date of sale})$$

Where f is the Euclidean distance between subject property i and property j across the five dimensions in the distance equation. All variables are standardized so that each has mean 0 and standard variance of 1. Dwellings that are "close" under this approach are physically near one another (x, y), are the same size (above-grade living area and number of bedrooms), and are sold close to each other temporally. For the 70 closest dwellings a weighted least squares regression is estimated using the tri-cubic function as weights:

$$w_{ij} = \left[1 - \left(\mathrm{d_{ij}}/\max(\mathrm{d_{ij}})\right)^3\right]^3$$

[10] Within each sector are a number of smaller MLS zones; MLS zone and quarter dummies are included in these each of the sector regressions.

In this way, the dwellings most like the subject dwelling get the most weight, while the "farthest" observation gets no weight.

Table 5 shows the summary statistics from our 6,178 LWRs.[11] These results show that the mean price elasticity of above-grade living area is 0.39. But there is substantial variation across the city, ranging from 0.01 to 0.85 (10th to 90th percentiles). Similar variation can be found for all of the estimated coefficients. The average coefficients, however, are all quite typical of what one would find in a traditional hedonic regression.
Estimating a Standardized House Value Surface

As discussed above, our LWR results provide location-specific estimates of the value of each of a home's physical characteristics. In this section, we define a hypothetical standardized house and use the LWR coefficients to predict the value of this standardized house across the city. It is important to note that this house may not (likely will not) actually exist in many parts of the city. Nevertheless, it is simple to predict this value at each point by simply applying the LWR coefficients to the standardized house's characteristics.

We define a standard house as having the "typical" characteristics of a home at the periphery of the city. In particular, we used the median characteristics of the improved parcel sales in MLS zones that contain at least 25 vacant lot sales. Our intuition for using this standardized structure is that it will be representative of the homes that will be built on the vacant lots, thereby facilitating the calibration exercise that follows. As we discuss below, however, the specific choice of standardized structure is not completely innocuous.

Our final standardized dwelling is a 1,528 square foot ranch-style home on an 11,049 square foot lot. It is one year old and has 3 bedrooms, 3 baths, a fireplace, central air, four additional plumbing fixtures, and an attached garage. It is not on a "premium" lot nor does it front a major street. Notice that these characteristics make this standardized dwelling comparable with the 75th percentile of all improved property sales in our sample (Table 2).

Using these characteristics, we price this dwelling at the location of each of the local regressions, creating a spatial distribution of pricing for an identical dwelling by construction. Once again, our premise is that spatial variation in the price of the standard dwelling derives from local amenities and drawbacks.

Figure 4 shows the spatial distribution of observed sales prices across the city, while Figure 5 shows the spatial distribution of our standard house. In order to understand these figures, note that Figure 4 can be interpreted as the actual market value of properties across the city, because it derived from actual sale prices. In contrast, Figure 5 shows the relative land value surface. Because the standard structure is the same at each point in space (and hence has the same value across the city) differences in the height of this surface reflect relative land values across the city.

[11] Complete results from each of these regressions are available upon request.

Several points are worth noting in comparing these surfaces. First, as might be expected given the age and size of the homes found in these locations, the highest observed sale prices occur on the far east and northwest sides of the city (Figure 4). In contrast, land values (Figure 5) are highest on the near east side and lower on the periphery of the city. Upon reflection, this is not surprising. At the periphery very large homes are built on relatively cheap land. Closer in, smaller homes are found on more valuable land. Indeed, the "peaks" in this figure correspond with the location of the more affluent older neighborhoods in the city.

Calibrating the Relative Value Surface

Though we are interpreting the spatial variation in this standardized value surface as the manifestation of variation in local land premium, we can not immediately recover the underlying land values. This is because the standardized dwelling remains a bundle of both land and improvements. To recover land values, we need to use the information in the vacant land sales to "pin down" this surface. We do this by calculating the value of the standard dwelling on each of the parcels in our vacant lot sample by interpolating from the values of standard dwellings on nearby improved parcel sales.

Next we adjust the standard dwelling value to reflect the actual characteristics of the vacant lots. As with improved properties, the price of vacant lots varies both with location and lot characteristics. For example, an otherwise homogeneous neighborhood may have some "premium" lots that are adjacent to a golf course or water feature. Similarly, we need to adjust size differences between our standardized dwelling and the actual vacant lot. Once again, we make these adjustments using the LWR coefficients from nearby improved parcels. Doing this gives us an estimated value of the standard dwelling on the vacant lot with its actual characteristics. Note that this differs from our earlier exercise, in that our relative value surface assumed the same lot characteristics across the entire city.

We now have (1) the actual selling price for each vacant lot (L) and (2) the estimated market value of each of these lots if it had the standard set of improvements on it (V). By subtracting the actual lot price from this estimated market value, we obtain an estimated value of the improvements: $B = V - L$. In theory, this difference should be the same at each location, because the set of improvements on the lots has been constrained to be identical. In practice, unobserved heterogeneity in the lot prices and noise in estimating the market values leads to variation across lots. We therefore estimate the value of the standard improvements as $\overline{B}$, or the average of the estimated B's for the vacant lot parcels.

To repeat, we tie down the relative value surface estimated above by extending it to our vacant lot parcels. We then adjust these values to reflect the actual characteristics of these parcels, and then subtract the actual sale prices of these lots to obtain an estimated value of the standard improvements on each lot. Averaging over all vacant lots in our sample gives us an estimated value of the standard improvement across the city.

The result of this process yielded an implied structure value of $157,000. This value should be spatially invariant, with local amenities capitalized into land and not structure prices. The standardized dwelling value surface from Figure 5 less the cost of the standardized dwelling, $\overline{B}$, yields our estimated value of land for each 2005 sale in our sample.

Figure 6 shows the spatial distribution of our estimated land values in dollars. The blue dots represent lots with estimated values in excess of $10,000, and so forth. While the majority of the estimated lot values are significantly positive, there are a number of parcels for which we estimate negative land values. We discuss this apparent anomaly next.

Can Land Values Be Negative?

The first blush answer to this question is "no."[12] Rather, we believe this result reflects the subtle challenges with estimating and calibrating a land value surface. One of the challenges with our methodology is determining the proper standardized dwelling to impose across the metropolitan area. Land values are determined by the highest-and-best use of the land, not their current use or even some arbitrarily imposed use (such as our standard dwelling). Two immediate questions come to mind regarding whether the resulting land values from this exercise will be reasonable.

The first is the possibility that the highest-and-best use of a parcel may not be single-family residential. If developers anticipate a parcel may be redeveloped for some other, higher-valued, use, the true underlying land value will reflect this fact, and our estimated land values will be too low for such parcels.[13] More broadly, this problem simply reflects the fact that urban land values are actually the upper envelope of the values under different uses. The implication is that a comprehensive measure of land values requires consideration of all possible land uses, something that is beyond the scope of this project. The second question involves a related but slightly subtler issue: What is the optimal dwelling for a particular lot? Throughout this analysis we have argued that any variation in our relative value surface reflects differences in underlying land values at different locations, because the physical dwellings on each parcel have been constrained to be the same. While this notion has great intuitive appeal, carrying out our thought exercise to its logical extreme poses some knotty questions.

Suppose that a fire burned down a home in a developed neighborhood and the lot were cleared of the debris. The relevant question for determining the value of the land is what type of home would be rebuilt on this land. In some parts of the inner city, they typical home is a 60-year old, 2 bedroom, 1 bath home with 900 square feet on a 5,000 square

[12] The obvious exception is when there is environmental contamination or some other liability associated with owning the land that exceeds the land's value. We ignore that possibility in the present discussion.
[13] Note that this problem only suggests a downward bias in our land value estimates; if the land's value is lower under alternative uses, then our estimates should be reasonably correct, subject to the caveats discussed below.

foot lot. One the periphery are neighborhoods with relatively new homes with 4 or more bedrooms, an equal number of baths and well over 3,000 square feet of living area. Would our "typical" 1,528 square foot home be built on these lots should they suddenly become vacant? Likely not. More importantly, does our estimated standardized value surface really reflect the price a buyer would pay for such a house? To the extent it does not, the value loss should be attributable to the structure, not the land.[14] Once again, the implication is that our estimated value surface is likely underestimated.

Another way of thinking of this problem is that by estimating our value surface with a standard set of improvements we may be making out-of-sample predictions in many parts of the city. As a result, even if our theoretical exercise is correct, we may not be able to use our LWR coefficients to accurately estimate this surface at each location.

This discussion suggests that a "better" value surface would allow the imposed structure to vary across the city to reflect the "ideal" housing in any given area. We have begun to work on this problem, but the challenges associated with pinning down this moving value surface (because the structures are no longer constant) are considerably more complex.

4. Conclusions and Directions for Future Research

The original goal of this project was to develop a rigorous yet tractable method for extracting urban land values in developed areas with few vacant land sales. Our intuition was straightforward: while dwelling sales could not be directly used to estimate land values, they could be used to estimate location premium. These premiums could then be used in conjunction with land sales to recover underlying land prices even in neighborhoods with no vacant land sales. The statistical tool we use to estimate the location premium was locally-weighted regression (LWR).

This is the first contribution of our project. Although the use of more flexible modeling statistical methods is becoming more common, "standard" hedonic regression remains the dominant tool of empiricists using housing data. In some applications, this is entirely appropriate, but the imposition of fixed implicit prices on housing characteristics is a strong assumption that is generally untested. Using LWR techniques, we are able to estimate unique shadow prices for various housing components at each location throughout the city. These LWRs confirm that, indeed, the implicit prices of a home's physical characteristics vary considerably across the metropolitan area. In light of this result, it is clear that the use of fixed coefficients is a misspecification that can bias parameters used to inform policy.

Using the estimated LWR coefficients, we then proceed to predict the value of a standard dwelling unit across the city. Because the physical characteristics of this structure are the same at each and every location (and hence should have the same value), the resulting value surface is, in fact, a measure of the relative value of land at each location. That is, we assume that local amenities are capitalized into land and not structures, meaning that

[14] In the appraisal literature, this type of value loss to the structure is known as external or economic obsolescence.

spatial variation in the price surface derived from a standard dwelling must be due to variation in land values.

Our final step is to "pin down" this relative value surface using vacant land sales (which primarily occur on the periphery of the city). We do this by imputing the value of the standard dwelling at the locations of the vacant lot sales, and subtract the observed selling prices of the lots to come up with the "value" of the standard dwelling. Subtracting this imputed dwelling value from our standardized price surface provides us with estimates of absolute land values at each location throughout the city.

As discussed above, the technical challenges with pinning down the value surface were greater than we anticipated, and there remain a number of conceptual questions that need to be considered more deeply before this technique can be applied in practice to estimate urban land values; it is to these questions that we will turn in ongoing research. In part, the challenges we found were quite revealing with regard to some of the elemental features of the basic urban model.

First, our prior – consistent with the standard urban model – was that the price of land at the periphery would be something close to the price of agricultural land and that variation would be small due to productivity differences. This is far from what we found in the periphery of Wichita. Arm's-length transaction at urban-rural fringe traded for widely different prices due to many factors not in our data sets. For example, one-lot sales looked different from lot sales within master-planned development. Master-planned developments varied as well, apparently as a function of neighbors, with lot sales near other high-end housing trading at higher prices. Furthermore, it appears that peripheral land markets have incorporated anticipated growth between other towns beyond the edge of our data. Finally, land sale prices appear to be more volatile, making pooling data across time questionable.

These regularities are in conflict with the basic urban model: the per-unit price of land appears to be a function of many other factors beyond distance to the CBD and the price of agricultural land. Even in a place that appears to be likely to be a perfect match, a small, city in a featureless plain, land prices are distinctly not monocentric.
Our research began with the assumptions used in the basic urban model and quickly ran into these regularities at odds with them. Nevertheless, our research has clearly demonstrated two key points.

First, we have shown that simple hedonic techniques are unlikely to be able to produce reliable land value estimates in areas where vacant lot sales are lacking. Because the physical characteristics of the housing stock and the shadow prices of these characteristics both vary considerably over a geographic area (reflecting vintage effects, among other things), an extremely large number of interaction terms would need to be included in an hedonic regression model in order to accurately be able to use the resulting regression coefficients to decompose land and building values.

More importantly, however, we have shown that LWR does provide the necessary flexibility to estimate the localized prices required to derive a relative land value surface. When combined with vacant parcel sales, this relative value surface can be calibrated to provide tangible land-value estimates. We find a wide range of land values in Wichita, from negative in the core and older areas, to highly positive in the many desirable neighborhoods of Wichita. We view these land values as preliminary, however.
We require a better understanding of several relationships before submitting our approach to practitioners. First, the irregularities in the land prices are disconcerting from an empirical perspective. It is not clear what constitutes a fair comparison for land under houses in the core. That is, vacant land in the core is odd because it is undeveloped. What led to a favorably located site being left undeveloped? Should this observation be included? Land sales in the periphery represent expectations and risk not present in developed neighborhoods. How should these be controlled for? Finally, the basic premise of constructing a "standardized" price surface using a fixed set of characteristics implies out of sample forecasting in some neighborhoods – the median home in Wichita represents a significant improvement in dwelling characteristics in many locales. How does this influence our results? The next steps are clear and we plan to continue to refine our proposed method for recovering land values in urban areas. The results presented in this report are preliminary in the point estimates, but the method appears to hold the potential we had hoped for when we began.

References

Cohen, Jeffrey P. & Cletus C. Coughlin, 2005. "An Introduction to Two-Rate Taxation of Land and Buildings." *Federal Reserve Bank of St. Louis Economic Review*, May/June, vol. 87, no. 3, pp. 359-74.

Bell, Micheal E., and John H. Bowman, 2006. "Methods of Valuing Land for Real Property Taxation: An Examination of Practices in States that Require Separate Valuation of Land and Improvements." Lincoln Institute of Land Policy Working Paper WP06MB1.

Brunori, David and Jennifer Carr, 2002. "Valuing Land and Improvements: State Laws and Local Government Practices." Lincoln Institute of Land Policy Working Paper WP02DB1.

Goodman, Alan C., Thibodeau, Thomas G., 1998. "Housing Market Segmentation." Journal of Housing Economics vol. 7, pp. 121–143.

Goodman, Alan C., Thibodeau, Thomas G., 2003. "Housing Market Segmentation and Hedonic Prediction Accuracy." Journal of Housing Economics. Vol. 12, pp. 181–201.

Goodman, Alan C., Thibodeau, Thomas G., 2007. "The Spatial Proximity of Metropolitan Area Housing Submarkets." Real Estate Economics. Vol. 35, pp. 209–232.

Gloudemans, Robert J., 2000. "Implementing a Land Value Tax in Urban Residential Communities." Lincoln Institute of Land Policy Working Paper WP00RG1.

Gloudemans, Robert J., 2002. "An Empirical Analysis of the Incidence of Location on Land and Building Values." Lincoln Institute of Land Policy Working Paper WP02RG1.

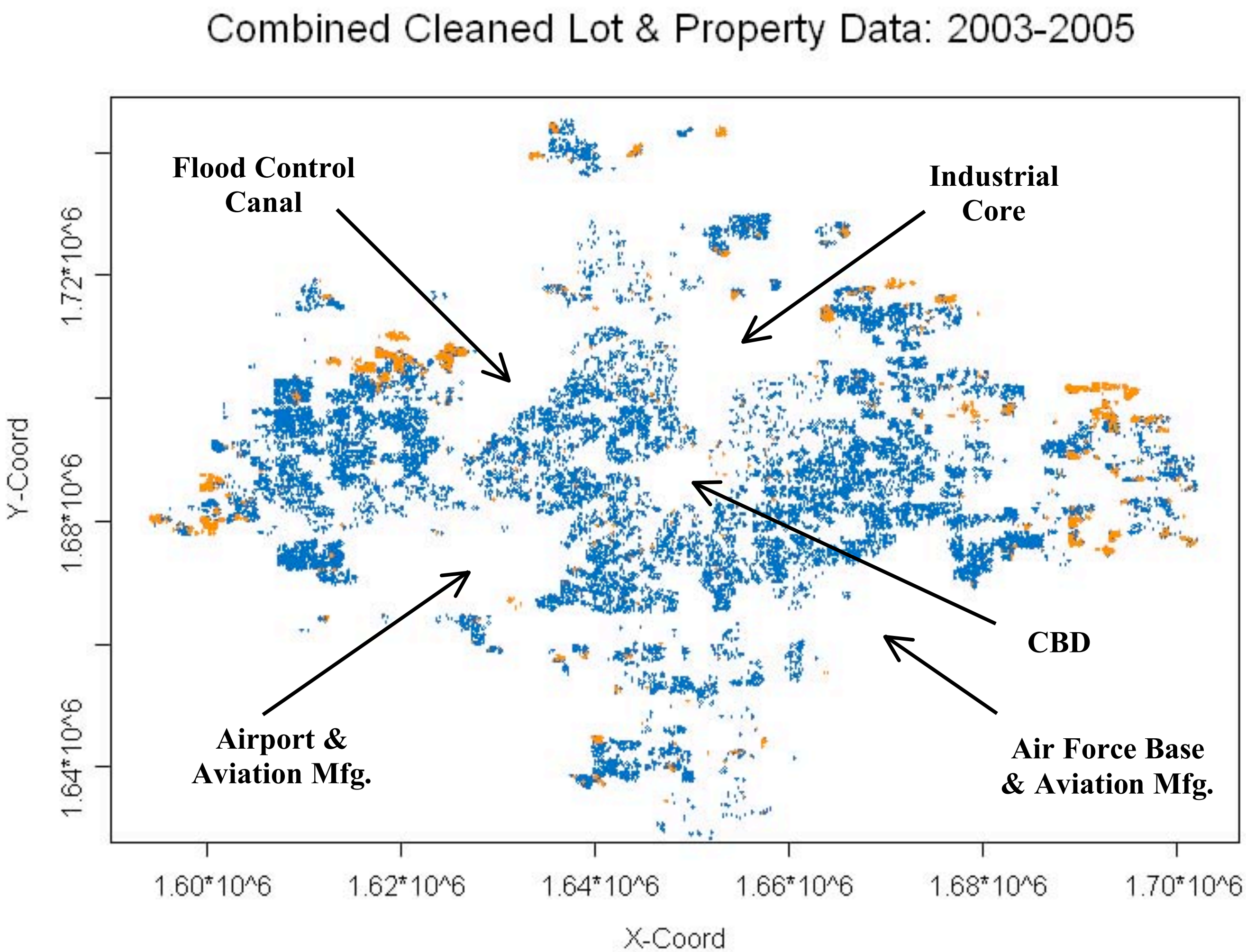
Gloudemans, Robert J., Sheldon Handel and Michael Warwa, 2002. "An Empirical Evaluation of Alternative Land Valuation Models," Lincoln Institute of Land Policy Working Paper WP02RG2.

Mills, Edwin S., 1998. "Is Land Taxation Practical," *Illinois Real Estate Letter*, Fall, vol. 12, no. 4, pp. 1-5.

Tideman, T. Nicolaus, 1999. "Taxing land is better than neutral: Land taxes, speculation, and the timing of development," in *Land/Value Taxation*, Kenneth C. Wenzer, editor, M. E. Sharpe: Armonk, NY, pp. 109-133.

Vickrey, William, 1999. "Simplification, progression, and a level playing field," in *Land/Value Taxation*, Kenneth C. Wenzer, editor. M. E. Sharpe: Armonk, NY, pp. 17-23.

Figure 1



Combined Cleaned Lot & Property Data: 2003-2005
Flood Control Canal
Industrial Core
CBD
Airport & Aviation Mfg.
Air Force Base & Aviation Mfg.
1.64*10^6
1.68*10^6
1.72*10^6
Y-Coord
1.60*10^6
1.62*10^6
1.64*10^6
1.66*10^6
1.68*10^6
1.70*10^6
X-Coord

Figure 2

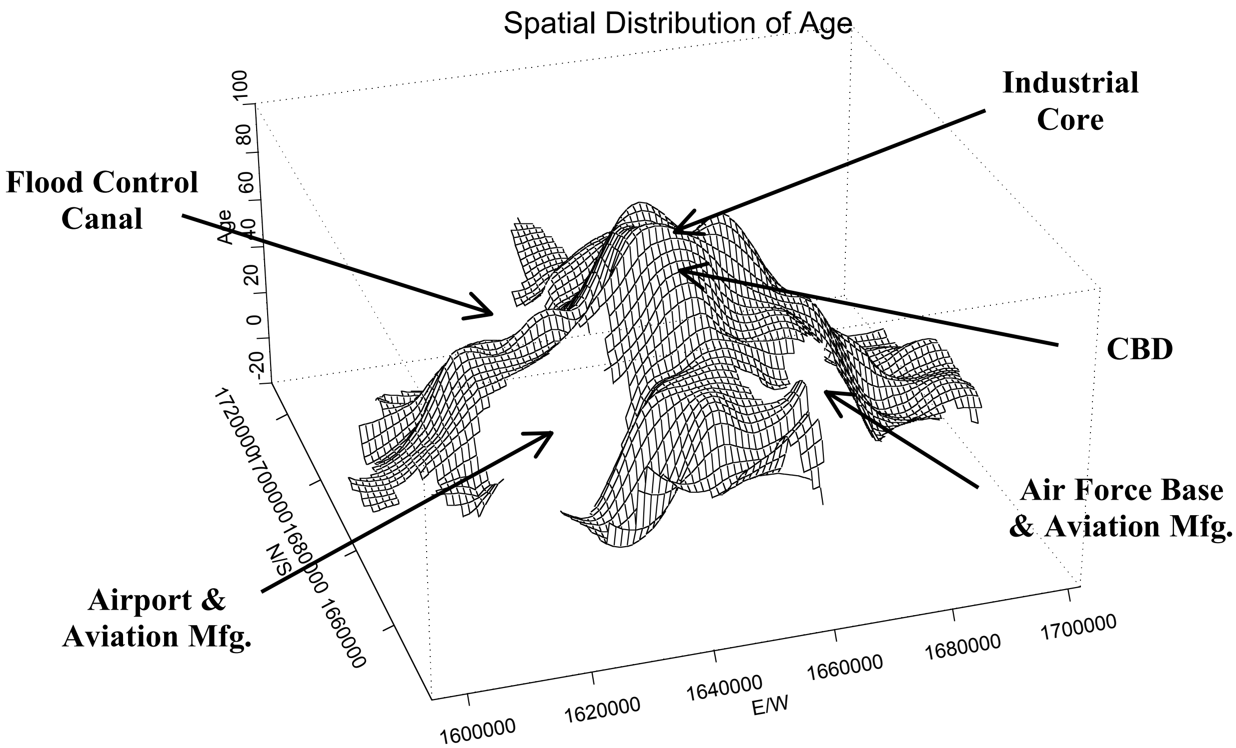


Figure 3

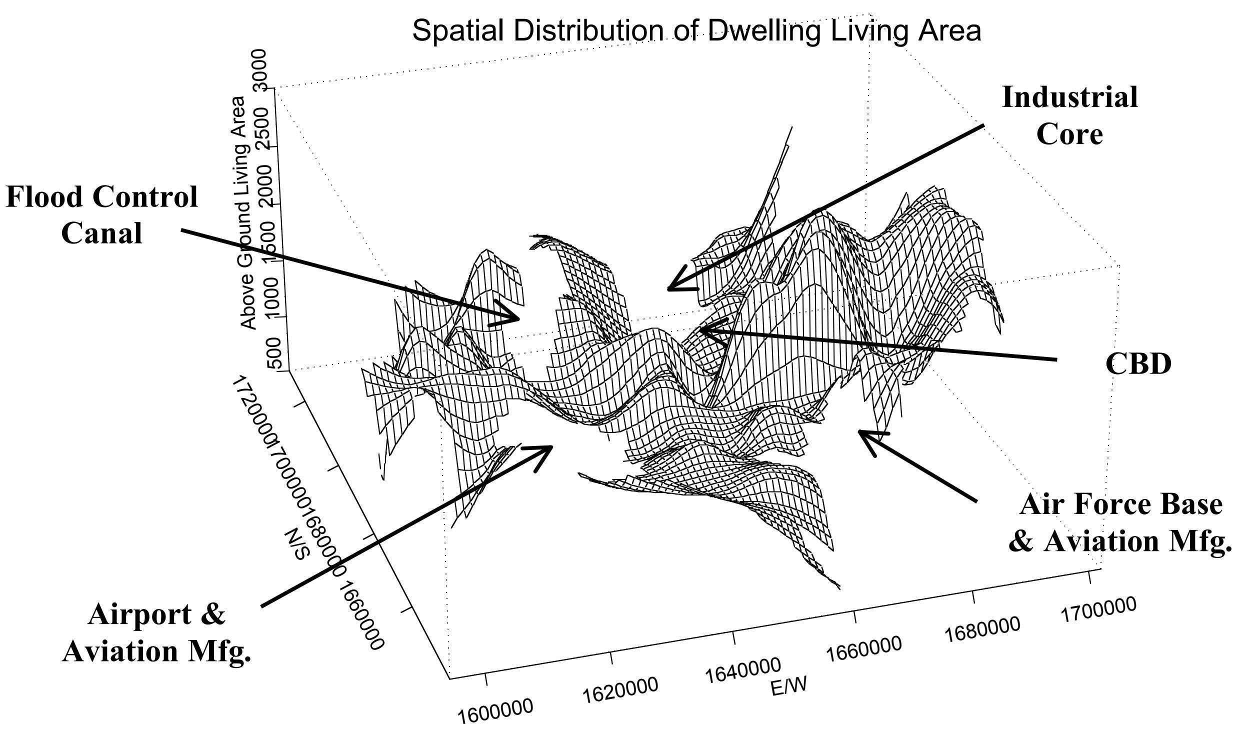


Figure 4

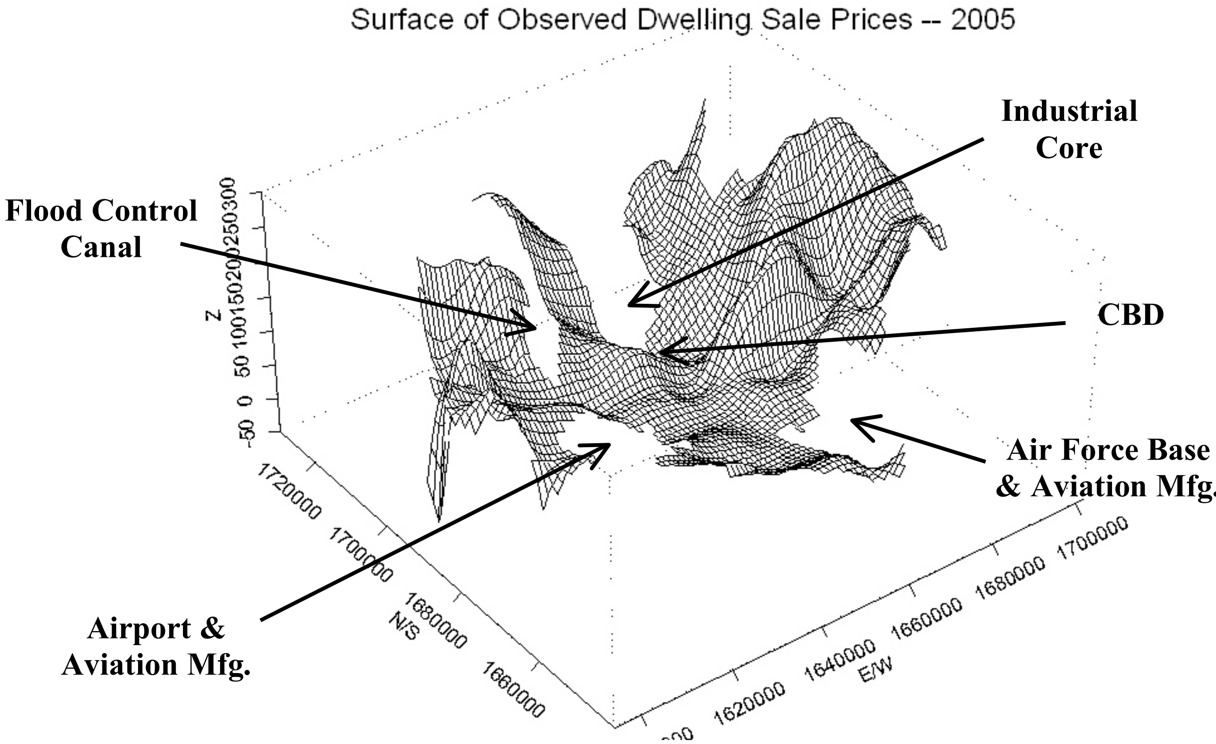


Figure 5

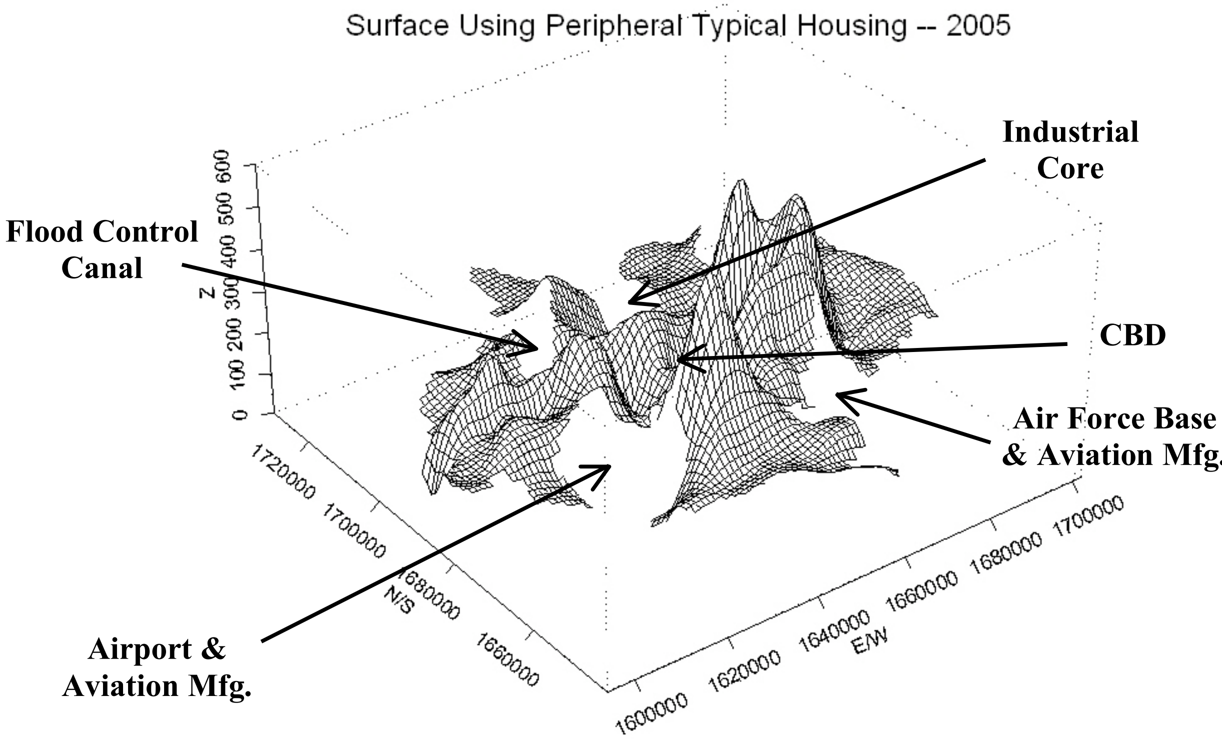


Figure 6 – Estimated Land Values in $1,000s

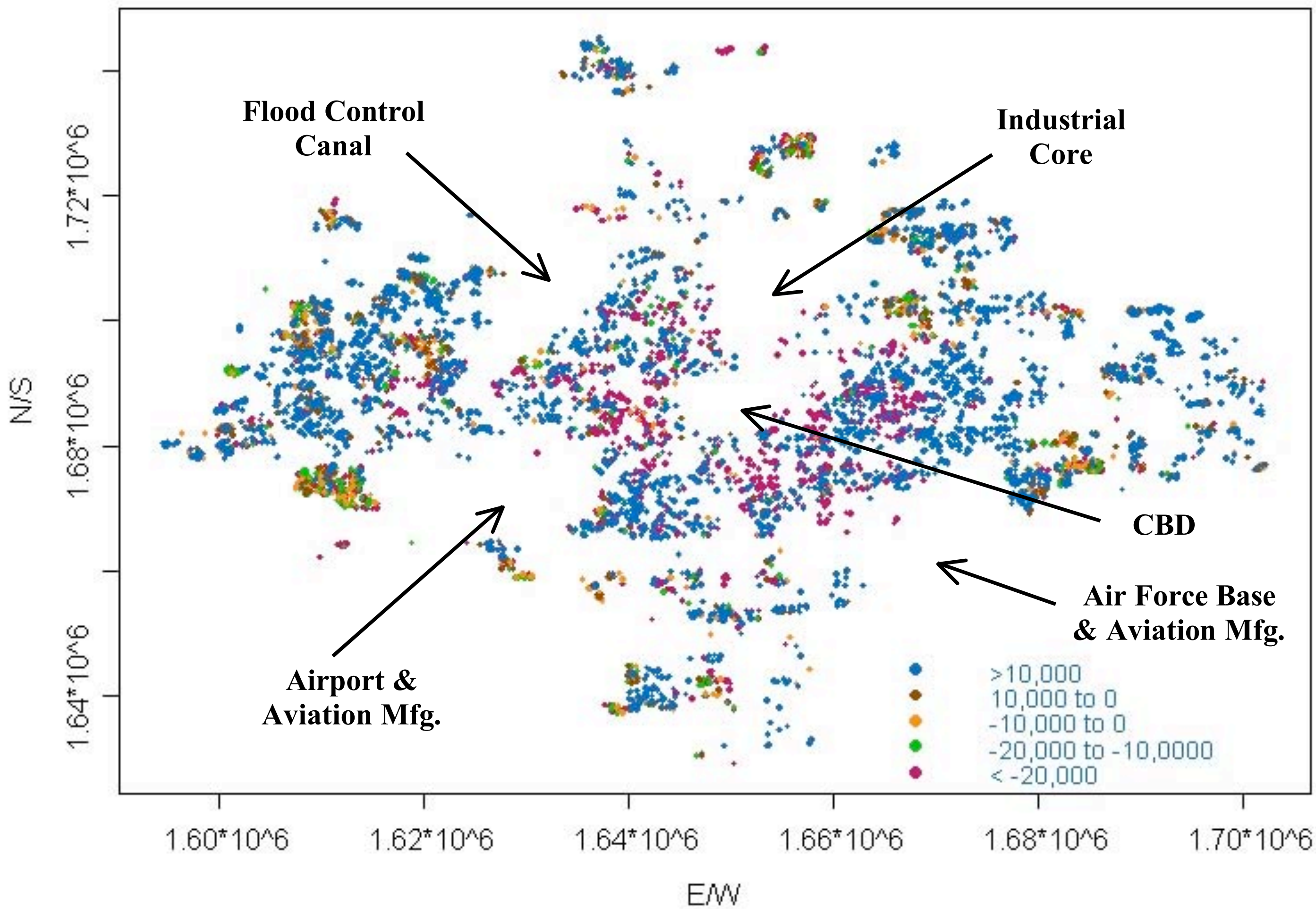


Table 1 – Vacant Lot Sales Summary Statistics

Variable	Mean	Std. Dev.	Min	25th Percentile	Median	75th Percentile	Max
Price	$22,197	$13,904	$350	$14,000	$18,900	$27,000	$170,000
Lot SF	14,277	6,305	3,500	10,278	12,419	16,126	43,125
Price PSF	$1.60	$0.73	$0.08	$1.11	$1.46	$1.95	$4.98
Premium lot	0.081	0.273	0	0	0	0	1
Premium view lot	0.034	0.181	0	0	0	0	1
Cul-de-sac lot	0.009	0.096	0	0	0	0	1
Major street frontage	0.424	0.494	0	0	0	1	1

Table 2 – Improved Parcel Sales Summary Statistics

Variable	Mean	Std. Dev.	Min	25th Percentile	Median	75th Percentile	Max
Price	$126,067	$70,928	$4,000	$78,585	$114,900	$150,000	$499,000
Above-grade SF	1,379	487	420	1,044	1,276	1,589	6,697
Basement finished SF	170	415	0	0	0	0	2,800
Lot SF	10,730	5,056	2,987	7,653	9,250	11,890	43,556
Age of improvements	29	27	-1	4	22	50	135
Bedrooms	2.86	0.76	0	2	3	3	7
Bathrooms	2.15	0.90	1	1	2	3	8
Add. plumbing fixtures	2.72	1.08	0	2	2	3	9
Ranch-style home	0.7722	0.4194	0	1	1	1	1
Attached garage	0.796	0.403	0	1	1	1	1
Detached garage	0.141	0.355	0	0	0	0	3
Central air	0.960	0.196	0	1	1	1	1
Fireplace	0.606	0.487	0	0	1	1	1
Premium lot	0.060	0.238	0	0	0	0	1
Major street frontage	0.015	0.121	0	0	0	0	1

Notes: Age is calculated as the year of sale less the year the home was built. For new homes sold while under construction, the calculated age may be negative.

Table 3 – Description of Variables

Variable	Description
Price	Sale price of the vacant lot or improved property sale
Above-grade SF	Total square feet of above-grade finished living area
Basement finished SF	Total square feet of below-grade living area finished to the same quality level as the above-grade finished space
Lot SF	Total square feet of land area in the parcel
Age of improvements	Year of sale minus the year the improvements were constructed
Bedrooms	Total number of bedrooms in the home
Bathrooms	Total number of bathrooms, including both full- and half-bathrooms
Additional plumbing fixtures	Number of plumbing fixtures (sinks, bathtubs, etc.) beyond those required based on the number of full- (3 base plumbing fixtures) and half-bathrooms (2 base plumbing fixtures); a proxy for “upgrades” in the home
Ranch-style home	An indicator variable taking the value of 1 if the home has a ranch or walk-out ranch architectural style
Attached garage	An indicator variable taking the value of 1 if the home has an attached garage
Detached garage	An indicator variable taking the value of 1 if the home has a detached garage
Central air	An indicator variable taking the value of 1 if the home has central air conditioning
Fireplace	An indicator variable taking the value of 1 if the home has one or more fireplaces
Premium lot	An indicator variable taking the value of 1 if the home has golf course or water frontage
Major street frontage	An indicator variable taking the value of 1 if the home is on a high-traffic street
Quarter of sale	The quarter in which the sale of the property occurred
MLS zone	A relatively fine geographic variable defined by the local multiple listing service board, corresponding to neighborhoods that are viewed by buyers and sellers as being highly homogeneous

Table 4 – Hedonic regression results by sector

	East	NE	NW	SE	SW	West
Log(Above-grade SF)	0.52***	0.36***	0.51***	0.63***	0.37***	0.64***
	(0.012)	(0.028)	(0.025)	(0.037)	(0.027)	(0.015)
Log(Below-grade SF)	0.02***	0.01***	0.03***	0.02***	0.02**	0.02***
	(0.001)	(0.004)	(0.004)	(0.006)	(0.006)	(0.001)
Log(Lot SF)	0.05***	0.01	0.01	0.09**	0.11***	0.06***
	(0.007)	(0.020)	(0.015)	(0.030)	(0.019)	(0.008)
Age	-0.03***	-0.07***	-0.11***	-0.12***	-0.14***	-0.03***
	(0.002)	(0.009)	(0.008)	(0.017)	(0.009)	(0.003)
Age-squared	-0.01**	0.00	0.00	0.01	0.01	-0.01
	(0.003)	(0.009)	(0.009)	(0.012)	(0.009)	(0.003)
Bedrooms	0.05***	0.10***	0.09***	0.07***	0.11***	0.06***
	(0.003)	(0.011)	(0.010)	(0.011)	(0.010)	(0.004)
Bathrooms	0.03***	0.04**	0.03**	0.04**	0.03	0.05***
	(0.002)	(0.013)	(0.010)	(0.012)	(0.014)	(0.003)
Add. plumbing fixtures	0.07***	0.01	0.05**	-0.06**	0.06***	0.10***
	(0.005)	(0.017)	(0.014)	(0.019)	(0.016)	(0.006)
Ranch	0.17***	0.14***	0.14***	0.11***	0.13***	0.21***
	(0.016)	(0.013)	(0.014)	(0.026)	(0.013)	(0.027)
Attached garage	0.05***	0.15***	0.13***	0.15***	0.13***	-0.01
	(0.014)	(0.013)	(0.012)	(0.023)	(0.014)	(0.025)
Detached garage	0.13**	0.30***	0.38***	0.43***	0.32***	0.31***
	(0.046)	(0.018)	(0.017)	(0.026)	(0.018)	(0.047)
Central air	0.07***	0.18**	0.15***	-0.06	0.00	0.05***
	(0.006)	(0.059)	(0.036)	(0.092)	(0.043)	(0.008)
Fireplace	0.01	-0.25***	-0.04	-0.08	-0.15***	-0.07***
	(0.032)	(0.038)	(0.036)	(0.070)	(0.033)	(0.020)
Premium lot	0.877	0.788	0.811	0.871	0.741	0.886
	4,831	1,684	2,214	1,694	2,083	3,987
Major street	0.05***	0.01	0.01	0.09**	0.11***	0.06***

Table 4 – Hedonic regression results by sector

	East	NE	NW	SE	SW	West
	(0.007)	(0.020)	(0.015)	(0.030)	(0.019)	(0.008)
Constant	-0.03***	-0.07***	-0.11***	-0.12***	-0.14***	-0.03***
	(0.002)	(0.009)	(0.008)	(0.017)	(0.009)	(0.003)
R-square	-0.01**	0.00	0.00	0.01	0.01	-0.01
N	(0.003)	(0.009)	(0.009)	(0.012)	(0.009)	(0.003)

Notes: Standard errors shown in parentheses below the coefficient estimates; output for MLS zone and quarterly indicator variables are omitted; *** significant at the 0.1% level; ** significant at the 1% level; * significant at the 5% level; sectors represent the four quadrants of the inner city and the far east and west sides of the urban area; the dependent variable in these regressions is the natural log of price.

Table 5 – Locally-weighted regression summary statistics

	Mean	Median	St. Dev.	10%ile	90%ile
Log(Above-grade SF)	0.39	0.36	0.34	0.01	0.85
Log(Below-grade SF)	0.03	0.01	0.51	-0.02	0.04
Log(Lot SF)	0.06	0.05	0.18	-0.12	0.25
Log(Age)	-0.09	-0.04	0.28	-0.35	0.05
Bedrooms	0.00	0.00	0.07	-0.07	0.07
Bathrooms	0.06	0.07	0.12	-0.02	0.16
Add. plumbing fixtures	0.02	0.02	0.09	-0.05	0.10
Ranch style house	0.14	0.04	2.92	-0.16	0.22
Attached garage	0.10	0.09	0.18	-0.09	0.30
Detached garage	0.08	0.09	0.16	-0.11	0.25
Central AC	0.21	0.20	0.24	-0.07	0.47
Fireplace	0.05	0.04	0.13	-0.07	0.18
Premium lot	0.05	0.04	0.14	-0.07	0.18
Major street	-0.07	-0.05	0.26	-0.36	0.20
Intercept	8.09	8.43	4.20	4.24	11.58
R-square	0.702	0.713	0.141	0.513	0.877

Notes: Columns show the summary statistics for the estimated regression coefficients from 6,178 locally-weighted regressions; output for MLS zone and quarterly indicator variables are omitted; the dependent variable in these regressions is the natural log of price.

The locally-weighted regression results show the substantial spatial variation of the shadow prices of the physical components of houses in the Wichita area.