Commercial and Residential Land Prices Across the United States

Joseph B. Nichols*
Federal Reserve Board

Stephen D. Oliner
Federal Reserve Board

Michael R. Mulhall
J.P. Morgan Chase

Revised: July 2010
Previous drafts: May 2009, February 2010

Abstract

We use a national dataset of land sales to construct land price indexes for 23 MSAs in the United States and for the aggregate of those MSAs. We construct the price indexes by estimating hedonic regressions with a large sample of land transactions dating back to the mid-1990s. The regressions feature a flexible method of controlling for spatial price patterns within an MSA. The resulting price indexes show a dramatic increase in both commercial and residential land prices over several years prior to their peak in 2006-07 and a steep descent since then. These fluctuations in land prices have tended to be considerably larger than those in well-known indexes of commercial real estate and house prices. Because those existing indexes price a bundle of land and structures, this comparison implies that land prices generally have been more volatile than structures prices over this period.

*Contact author. Email and phone: joseph.b.nichols@frb.gov, (202) 452-2983. We thank Christopher Reynolds for producing the location maps that we use in the paper. We received valuable comments from numerous Federal Reserve colleagues at an early stage of this study and from Edward Millner and other participants at the AREUEA 2009 mid-year meeting. We also thank Andreas Lehnert, Hui Shan, and Shane Sherlund for helpful conversations on estimation issues and Dan Sichel for useful comments on a recent draft of the paper. The data for this study were provided by the CoStar Group, Inc. The views expressed herein are those of the authors alone and should not be attributed to the Board of Governors of the Federal Reserve System or other members of its staff.

1. Introduction

All types of economic activity require land, either directly or indirectly. The direct use of land is obvious in industries such as farming and construction. But all other forms of commerce ultimately require land as well because workers, equipment, and buildings need to be located somewhere. Even a cutting-edge high-tech company like Google has a corporate campus in Silicon Valley and more than 65 other offices, research facilities, and data centers around the world.[1]

The importance of land is reflected in available estimates of its aggregate value. A broad measure of the value of residential and commercial land in the United States can be derived from the Flow of Funds (FOF) accounts published by the Federal Reserve Board. The implied FOF estimate – which covers land held by households, nonprofit organizations, and businesses other than farms and financial corporations – equals the market value of real estate minus the replacement cost of structures. At the end of 2010:Q1, this estimate of land value in the United States was nearly $4 trillion.[2]

With such a large aggregate value, changes in land prices can have a substantial effect on the net worth of businesses and households. In this regard, Davis and Heathcote (2007) estimate that swings in residential land prices accounted for most of the variation in house prices over 1975-2006 for the United States as a whole. Davis and Palumbo (2008) reach the same conclusion for a large set of metropolitan areas over a somewhat shorter sample period, as do

---

[1] See www.google.com/corporate/address.html for a listing of Google's locations.

[2] The data for this estimate are in tables B.100, B.102, and B.103 of the *Flow of Funds Accounts of the United States*, available at www.federalreserve.gov/release/z1. Barker (2007), Case (2007), and Davis (2009) present analogous estimates from earlier vintages of the Flow of Funds data. Note that all such estimates should be regarded as subject to considerable measurement error. The Flow of Funds accounts stopped publishing an explicit measure of land value in 1995, at least in part because of concerns about the accuracy of the estimates. Nonetheless, the discontinued series on land value can be reconstructed from other series published in the accounts. For other estimates of national land values that do not rely on Flow of Funds data, see Davis and Heathcote (2007), Barker (2007), and the Bureau of Labor Statistics (2007), along with the earlier work by Goldsmith (1951) and Manvel (1968).

Bostic, Longhofer, and Redfearn (2007) in their detailed analysis of home price changes within a single metropolitan area (Wichita, Kansas).

Land also serves as a form of collateral for loans, especially for construction loans. If the borrower defaults before a construction project is finished, the lender's recovery will depend in large part on the value of the land pledged as collateral. Commercial banks in the United States have substantial exposure to land prices through their lending activities. At the end of 2010:Q1, U.S. commercial banks held about $380 billion in construction and land development loans, and more than 18 percent of these loans were delinquent – by far the highest delinquency rate among the major types of bank loans.[3]

Despite the importance of land as a component of wealth, as a source of variation in real estate prices, and as collateral for loans, only a handful of studies have calculated land price indexes for the nation as a whole or for a broad set of cities.[4] Davis and Heathcote (2007) and Davis and Palumbo (2008) estimate price indexes for residential land, while Davis (2009) estimates indexes for both residential and commercial land. These indexes, however, are not based on transaction prices. Instead, Davis and his coauthors infer land prices as a residual from data on real estate prices and construction costs, combined with the assumption that the prices of existing structures always equal their replacement cost. This identifying assumption likely is reasonable over long spans of time, but it may not be valid over shorter periods.[5]

[3] These figures represent aggregates from the Consolidated Reports of Condition and Income (Call Reports) submitted by domestic commercial banks to the Federal Financial Institutions Examination Council. See Lee and Rose (2010) for an in-depth analysis of the Call Report data.
[4] That said, there are numerous studies of land prices for single cities or other narrow geographic areas. See, for example, Brownstone and DeVany (1991), Colwell and Munneke (1997, 2003), Guntermann and Thomas (2005), Haughwout, Orr, and Bedoll (2008), Isakson (1997), Kowalski and Paraskevopoulos (1990), Peiser (1987), and Wieand and Muth (1972). These studies cover commercial land or a combination of commercial and residential land. For studies that focus exclusively on residential land, see Bryan and Sarte (2009), Downing (1970), Greenlees (1980), Ihlanfeldt (2007), Rosenthal and Helsley (1994), and Voith (2001).
[5] Indeed, the standard neoclassical theory of investment with adjustment costs, as embodied in Tobin's q, links the volume of investment spending to the size of the gap between the market value of an asset and its replacement cost.

In contrast, Sirmans and Slade (2009) use transaction prices to calculate national land price indexes. However, they do not estimate price indexes for individual metropolitan statistical areas (MSAs), an important limitation given the substantial local variation in real estate markets. In addition, Sirmans and Slade use a very small set of variables to control for the dispersion in land prices in their regressions, do not allow the effects of these variables on land prices to differ across MSAs or property types, and do not weight the data to ensure that their price indexes are nationally representative. Our study addresses all of these methodological issues and calculates land price indexes not only at the national level but also for individual MSAs. We provide the first transaction-based indexes of land prices for a broad swath of MSAs across the United States.

Using source data obtained from the CoStar Group, Inc., we construct a dataset that includes roughly 160,000 land transactions in 23 MSAs.[6] These MSAs include the major population centers in the United States and some smaller cities. For most MSAs, the data span the period from the mid- or late 1990s to the end of 2009. We use these data to estimate a set of hedonic equations for residential and commercial land prices.[7] The explanatory variables in the equations include a number of characteristics of the property and the sale transaction, a flexible specification of the effects of location on price within an MSA, and half-yearly dummy variables to capture the changes in land prices over time after controlling for the other factors. The specification of locational effects combines the property's distance from the central business

---

Davis and Heathcote (2007) and Davis and Palumbo (2008) assume, in effect, that this gap is always zero for residential structures, and Davis (2009) makes the same assumption for both residential and commercial structures. This assumption is valid only if the stock of these structures can be adjusted very rapidly in response to changes in economic conditions, which seems unlikely given the planning and construction lags inherent in real estate projects.

[6] In addition to our paper and Sirmans and Slade (2009), other studies that have used the CoStar land-price data include Haughwout, Orr, and Bedoll (2008) for New York City and Colwell and Munneke (2003) for Chicago.

[7] The commercial land sales in our dataset cover not only properties slated for office and retail development but also industrial sites. To avoid confusion about the types of land in our dataset, we will use the label "commercial/industrial" rather than "commercial" for the rest of the paper.

district, which has been used in many studies, and the semiparametric specification introduced in Colwell (1998) that can accommodate a much wider range of spatial patterns.

Our primary focus is on the land price indexes implied by the coefficients on the half-yearly dummy variables. For the 23 MSAs as an aggregate, we present price indexes for a composite of residential and commercial/industrial land along with separate indexes for these two broad types of land. We also report the analogous price indexes for each MSA in all periods for which sufficient data are available.

The results show that land prices trended up at a moderate pace from 1995 until about 2002, and then accelerated sharply. From the second half of 2002 to the second half of 2007, our composite index of residential and commercial/industrial land prices for the 23 MSAs jumped nearly 130 percent, with even larger increases in the MSAs along the East Coast and in the Far West. However, prices have tumbled over the past few years, leaving the composite index for the 23 MSAs in the first half of 2009 more than 30 percent below its peak. Separate indexes for commercial/industrial and residential land prices display the same broad pattern, with especially large swings for the residential index. These moves, for the most part, outstrip the variation since 2002 in well-known national indexes of house prices and commercial real estate prices. Because those indexes price a bundle of land and structures, this comparison implies that land prices generally have been more volatile than prices of structures over this period. This greater volatility confirms the findings in Davis and Heathcote (2007) and Davis and Palumbo (2008) for residential property for an earlier period. Intuitively, land prices should be more volatile than the prices of structures because the latter are tied, at least loosely, to construction costs, while land prices have no such anchor.

The composite land price index in the second half of 2009 has begun to stabilize, and the movements of the composite commercial/industrial and the composite residential index have begun show interesting deviations from the movements of existing property price indexes. Residential home prices rallied briefly in the second-half of 2009, in part in response to a new-home buyer tax credit. Prices of land for residential construction, which had no such tax benefit, continued to decline. The commercial land price index actually started to move upward at the end of 2009, while commercial property prices provided conflicting signals, but overall continued their decline.

The remainder of the paper is organized as follows. The next section describes our dataset on land transactions. Section 3 presents our approach to modeling the effects of location on land prices, and section 4 describes all other aspects of the empirical methodology. Section 5 presents the estimation results. The final section summarizes our conclusions and lays out the next steps in this line of research.

2. Data

The data for our analysis were obtained from the CoStar Group, a major provider of information on commercial real estate in the United States (www.costar.com). Among its various data products, CoStar maintains a database on sales of commercial property and land in the United States (the "COMPS" database). CoStar obtains the transaction data from public records, interviews with parties to the transactions, and field inspection of the properties. Currently, the COMPS database includes more than one million transactions.

We analyze the transactions in COMPS explicitly identified as land sales. CoStar defines land sales as transactions that involve vacant property or property with unoccupied structures that are slated for demolition. These criteria ensure that the value of any existing structures

should be incidental to the total value of the property. A separate field in COMPS indicates whether the intended use of the land is residential, industrial, or commercial. To be included in COMPS, a residential land parcel must consist of at least five single-family lots or be large enough to support multifamily buildings with at least five units. There is no lower size limit in COMPS for industrial or commercial land parcels.

Each transaction record contains the sales price, address, and the longitude and latitude of the land parcel, along with a series of text fields describing the characteristics of the parcel. We use the information in these text fields to create a number of indicator variables for our hedonic price regression. The indicators include various improvements to the raw land (whether it has been paved, graded, finished, fully improved, previously developed, or platted); other characteristics of the property (the presence of existing structures and the presence of soil or building contamination); the proposed final use of the land (private development, public use, open space, or land to be held for investment); and a few characteristics of the transaction itself (whether it is a foreclosure sale or an expansion of a neighboring property).

For the analysis in this paper, we constructed a dataset with nearly 160,000 land transactions in 23 MSAs.[8] These MSAs include five cities in the Northeast corridor (Boston, New York, Philadelphia, Baltimore, and Washington DC); three areas in Florida (Orlando, Tampa/St. Petersburg, and South Florida); six cities on the West Coast (Los Angeles, San Diego, San Francisco, Sacramento, Seattle, and Portland); and nine cities in the interior of the country (Atlanta, Chicago, Dallas, Denver, Detroit, Houston, Las Vegas, Phoenix, and Tucson). Prior to 1995, the transaction data for most of these MSAs is either sparse or nonexistent. Accordingly, we standardized the starting point of the analysis in 1995, except for six MSAs for which the

[8] Although our complete database includes transactions in 138 MSAs, many of these MSAs have relatively few observations, and we restrict our attention to the MSAs with the richest data.

data begin between 1997 and 1999. The transaction data are available through the end of 2009. As described in Appendix A, we eliminated transactions with missing data and applied a variety of screens intended to improve the quality of the dataset. For example, we removed about 10,000 observations that appeared not be at arms length or that otherwise did not represent true market prices. We also removed all observations for an MSA in a given half-year when we judged the sample size to be insufficient to generate a reliable price index for that period.[9]

Table 1 provides information on the sample used for our analysis. As shown, for the 23 MSAs taken together, we have roughly 159,000 sales transactions in total, slightly more than half of which are commercial/industrial transactions. About 2,500 sales are in MSAs that have more than 40 transactions in a given half-year when the commercial/industrial and residential land sales are pooled, but do not have more than 30 transactions for each type of land. Accordingly, we exclude these observations when we use the hedonic regression to estimate separate MSA-level price indexes for commercial/industrial land and residential land. The "Commercial/Industrial" and "Residential" columns in the table show the number of observations used to estimate the MSA-specific price indexes for each type of land.

At the MSA level, the sample sizes vary widely, ranging from about 2,200 for Houston to more than 16,500 for Phoenix. This variation reflects, at least in part, differences in the amount of development activity across the MSAs and does not correlate closely with the relative shares of the MSAs in national stocks of residential or commercial/industrial land. Because of this disconnect, we employ MSA-level weights, described in section 4, to obtain nationally representative results.

---

[9] In particular, we included the residential or commercial/industrial transactions for an MSA in a given half-year only when we had more than 30 observations for that property type. For regressions in which we estimated a single MSA-level price index that covered both property types, we included a given half-year only when we had more than 40 observations in total. We settled on these minimum half-yearly sample sizes after observing the volatility in the price indexes generated without regard to sample size.

The other key dimension of the sample is the number of observations over time. Figure 1 shows one measure of this time pattern – the median sample size across the 23 MSAs for each half-year since 1995; the figure presents separate series for commercial/industrial and residential transactions.[10] As shown by the dashed line, the median MSA has at least 50 residential land transactions in every half-year through the first half of 2006, and more than 80 transactions in most of these periods. However, transaction volume plunged after mid-2006 with the collapse in housing activity. For commercial/industrial land, transaction volume was well maintained through 2007 but dropped sharply in 2008 and the first half of 2009. One factor contributing to this sharp decline in sample size was the shift toward non-arms-length transactions, which include the seizure of land by banks in foreclosures. As noted above, we exclude these transactions from our analysis. Although transaction volume for both commercial/industrial and residential land recovered somewhat in the second half of 2009, the level remained relatively low. Given the sample-size cutoffs we impose at the MSA level, the reduced volume of sales implies that we cannot calculate land price indexes for some MSAs in recent years, though we are able to calculate indexes through 2009:H2 for the 23 MSAs taken together.

3. Locational Effects on Land Prices

The location of a property is a prime determinant of its value. Home values reflect the proximity to good schools, employment opportunities, recreation, and transportation, along with the amount of crime in local neighborhoods and the quality of services provided by local government. Similarly, the value of commercial property varies with its proximity to transportation hubs and arteries, the presence of complementary business activities (such as retail

[10] We show the median sample size across MSAs for each period, rather than the average sample size or the total number of observations, to reduce the distorting effect from the increase in the number of MSAs in the sample from 1995 to 1999. Note also that the figure presents the median sample size across MSAs before applying the cutoff rules described in the previous footnote. We do this to display the underlying time patterns in transaction volume.

stores interspersed with restaurants), and the income level and other demographics of the local population. These locational effects would be expected to be even stronger for undeveloped land than for existing residential or commercial real estate, as the market value of the buildings at those sites would be tied – at least loosely – to the construction costs for the metropolitan area as a whole. Land values have no such anchor and are driven by the anticipated profit of developing a specific site.

Given the likelihood of important locational effects, empirical studies of land prices have included one or more variables to capture these effects. A simple and very common specification uses the property's distance from the central business district (CBD) of its MSA as a primary measure of its location.[11] This specification, however, imposes strong restrictions on locational effects. Consider two properties located at different points on a circle centered at the CBD. If the two properties are identical apart from their location on the circle, this specification implies that they will have the same market value, regardless of differences in their proximity to various amenities. Recognizing that distance from the CBD is unlikely to fully capture locational effects, most studies have included other variables as well. The added variables have included the distance from major roads, rail lines, and airports; distance from suburban business nodes; distance from the coastline; dummy variables for local topography; dummy variables for location within the city limits and within specific counties; measures of the amount of street frontage; and demographic information for the surrounding area.

For our study – which covers 23 separate MSAs – it is not practical to specify a vector of locational variables for each land transaction. Instead, we use the semi-parametric approach in Colwell (1998) to capture locational effects on real estate prices over and above those

[11] See, for example, Brownstone and DeVany (1991), Colwell and Munneke (1997), and Peiser (1987). In a similar specification for the New York City area, Haughwout, Orr, and Bedoll (2008) use the property's distance from the Empire State Building.

determined strictly by distance from the CBD.[12] In brief, Colwell's method superimposes a grid on a map showing the locations of the transactions in the dataset. The grid can be sized to fit the boundaries of this area, and it can be divided into as many component squares or rectangles as the researcher wishes to use. In Colwell's application to downtown Chicago, the grid was a rectangle that contained 36 square pieces (four squares in one direction, nine in the other). His method estimates the price level associated with each of the 50 (5x10) vertices of this grid.

To carry out the estimation, each vertex is treated as a separate variable in the hedonic regression. Let $V_1,\ldots,V_n$ denote the set of n vertices. Prior to estimation, values must be assigned to $V_1,\ldots,V_n$ for each transaction in the data set. This is done by measuring the nearness of the transacted property to the four vertices of the block in which it is located; all other vertices are assigned a value of zero for this observation. For a transaction located exactly at the center of a given block, each of the four surrounding vertices receives a value of 0.25. For transactions located elsewhere in the block, the values assigned to the four vertices are weights that sum to one and that reflect an area-based measure of closeness.

Figure 2 illustrates this weighting scheme. The figure shows a single block in a larger grid, with a transaction at point A. The value assigned to vertex V_1 for this transaction equals the area of the rectangle formed by A and the opposite vertex (the shaded area), divided by the total area of the block. This ratio of areas will converge to one as point A approaches V_1. The values for V_2, V_3, and V_4 associated with point A are calculated in the same manner as for V_1: Form the analogous opposite rectangles and calculate the ratio of the area of each rectangle to the area of the block. All other vertices in the grid surrounding this block have a value of zero for the transaction represented by point A.

[12] Colwell's method is a special case of the bilinear spline function developed by Poirier (1976, chapter 2).

After the values for $V_1,\ldots,V_n$ have been assigned in this manner for each transaction in the dataset, a hedonic price regression can be estimated with the vertices included as explanatory variables.[13] The estimated coefficient for a given vertex represents the height of the price surface at that point on the grid. Colwell shows that the surface defined by these estimated grid points is continuous, piecewise linear along the edges of the individual blocks, and parabolic along slices within each block.

In implementing this method, we laid out the grids for the individual MSAs to conform to the spatial pattern of the land transactions. For five of the MSAs (Atlanta, Dallas, Denver, Las Vegas, and Tucson), we were able to encompass the vast majority of transactions with a rectangular 5x5 grid composed of individual rectangular blocks. With a 5x5 grid, we estimate coefficients for 36 vertices (6x6) on the spatial price surface. As an example, figure 3 shows the 5x5 grid for Dallas, superimposed on a scatter plot of the transactions in our dataset. For the other MSAs, the spatial distribution of transactions did not fit well within a rectangular outline – generally because of the presence of a body of water or mountains – and we drew the outline of the grid to fit these patterns. In each case, the grid consists of between 22 and 28 individual blocks (with between 36 and 42 vertices), arranged in a non-rectangular shape. As an illustration, figure 4 presents the grid used for South Florida, where the land transactions occupy a band that parallels the coastline.

In summary, we take a hybrid approach to modeling locational effects on land prices. We include distance from the CBD as an explanatory variable in the regressions, which we augment with the semi-parametric grid method described above to capture the features of the spatial price surface that do not lie on a constant gradient away from the CBD.

---

[13] If the regression includes a constant term, one vertex must be omitted to avoid perfect colinearity among the explanatory variables. The colinearity arises because the vertex weights for any observation sum to one.

4. Specification and Estimation of the Hedonic Price Equation

We use the data described in section 2 to estimate a flexible hedonic regression for land prices. The dependent variable in the regression is the natural log of the price of land per square foot. The explanatory variables can be broken into three broad categories: property and transaction characteristics other than location ($X_1 ,\ldots, X_J$), measures of location ($Z_1 ,\ldots, Z_K$), and a set of half-yearly time dummies ($D_1 ,\ldots, D_T$). The regression pools the data for all 23 MSAs, but it allows the coefficient on each explanatory variable to differ both across MSAs and across the residential and commercial/industrial land transactions within an MSA. In effect, the pooled regression stacks 46 separate regressions (23 MSAs with two broad types of land in each MSA) and provides a convenient framework for testing a wide range of coefficient restrictions across the MSAs and property types. Each regression in the stack can be written as

$$(1) \qquad \ln P_{m,l} = \alpha + \sum_j \beta_{m,l,j} X_j + \sum_k \theta_{m,l,k} Z_k + \sum_t \gamma_{m,l,t} D_t + \varepsilon_{m,l}$$

where $P_{m,l}$ is the vector of observations of price per square foot in the m^{th} MSA for the l^{th} broad land type (l equals either residential or commercial/industrial), and $\varepsilon_{m,l}$ is the corresponding vector of errors. The subscripts on the β, θ, and γ coefficients show that these coefficients are allowed to vary across MSAs and property types.

The vector of property and transaction characteristics ($X_1 ,\ldots, X_J$) includes the natural log of the size of the parcel in order to test the so-called “plattage effect”. Plattage refers to the common empirical finding that the price of a land parcel rises less than proportionally with its size. This relationship likely arises because there is an optimal scale for buildings of a given type, which implies that parcels larger than the size needed to support the optimal building scale earn a lower return.

The other variables in $X_1, \ldots, X_J$ are the indicator variables mentioned in section 2. These variables, which are largely the same as those used by Haughwout, Orr, and Bedoll (2008), include:

- Type of property: separate dummy variables for land used for single-family housing, multifamily rental housing, other residential uses (principally condominiums), and industrial sites, with commercial land as the omitted category. The commercial category includes office, retail, and other commercial land uses.
- Condition of the property: separate dummy variables for whether the land has been graded, paved, finished, fully improved, platted and engineered, or previously developed; whether it has an existing structure; or whether the improvements are reported as unknown. The omitted condition is unimproved land. This group of variables also includes a dummy for whether environment problems exist as defined by reported soil or building contamination.
- Intended use of the property: separate dummy variables for property intended for public use, to be kept as open space, to be held for investment, or with an unknown use. The omitted category is private development.
- Characteristics of the transaction: separate dummy variables for property sold in foreclosure or purchased as part of the buyer's plan to expand a neighboring property.

The vector of location characteristics ($Z_1, \ldots, Z_K$) includes the MSA-specific sets of grid vertices described in the previous section, which provide a flexible way to control for the spatial price patterns within an MSA. We specify a separate grid for residential land and commercial/industrial land in each MSA. In addition to these grid vertices, we include the natural log of distance of each property from the population-weighted center of its MSA. The coefficient on this distance measure is allowed to vary by MSA and by broad property type,

consistent with the treatment of other variables in the regression. Finally, we include a set of MSA fixed effects, with New York City as the omitted MSA.

We weight the observations in the regression prior to estimation. Weighting is required because, as discussed in section 2, the number of land sales by MSA in our dataset does not reflect the relative shares of the MSAs in national stocks of commercial/industrial and residential land. To illustrate how we correct for this divergence, assume that a particular MSA accounts for 20 percent of the residential land sales in the dataset but for only five percent of total residential land in the 23 included MSAs. The dataset, therefore, overweights this MSA's residential land sales by a factor of four. As a correction, we would apply a weight of 0.25 to each residential land sale from this MSA. A weight for commercial/industrial land can be defined in the same manner; generally, this weight will differ from the weight for the MSA's residential land. Ideally, we would construct weights based on information on the available land area devoted to commercial/industrial and residential real estate by MSA. In the absence of such land data, we weighted the commercial/industrial land observations using estimates from Torto Wheaton Research of the total space in commercial and industrial buildings by MSA and the residential land observations using the number of occupied single-family and multifamily housing units from the 2000 Census. We re-adjust these weights on a period-by-period basis to account for the exclusion of MSAs with insufficient observations during particular time periods.

The pooled regression is estimated by maximum likelihood with a variance-covariance matrix ($\boldsymbol{\Sigma}$) that allows the variance of the error term to differ by MSA. Specifically, we assume that $\boldsymbol{\Sigma}$ is diagonal, with σ_m^2 (m = 1,…,23) as the error variance for every observation in the m^{th} MSA. This structure accommodates potential differences across MSAs in the amount of unobserved heterogeneity in land parcels and in the size of the shocks hitting the MSAs.

Our dataset consists of land sold over a given period and is not a random sample of land parcels, which raises the possibility of sample selection bias. However, the traditional Heckman (1979) procedure to correct for selection bias is not feasible in our case. To implement this procedure, we would need data on vacant land parcels that were not sold during the sample period. Such data would be very difficult to assemble for a large number of MSAs.[14] Moreover, it is unclear whether any such effort would be worthwhile. The results of studies that have applied the Heckman procedure to real estate prices have been mixed to date. Although there is some evidence of selection bias in house prices (see, for example, Gatzlaff and Haurin (1997,1998) and Jud and Seaks (1994)), the few available studies for commercial real estate and land prices have found that selection effects were generally small and insignificant.[15]

5. Results

This section presents the results from the estimation of the hedonic price function in equation 1. We discuss the results in three steps, focusing first on the estimated coefficients for the characteristics of the land parcels and the transactions, then on the spatial price effects, and finally on the estimated variation in land prices over time.

Property and transaction characteristics

Table 2 summarizes the estimated coefficients for property and transaction characteristics across the 23 MSAs for residential and commercial/industrial land. Each row reports the median and range of coefficient values across the MSAs, along with the number of MSAs for which the

[14] Strictly speaking, data also would be required for previously-developed properties that are candidates for redevelopment because our dataset include properties with structures that are slated for demolition. Obtaining data for unsold properties that meet this criterion would be even more difficult than collecting data on unsold vacant land.

[15] See Colwell and Munneke (1997) for an analysis of selection effects for commercial and residential land prices and Fisher, Geltner, and Pollakowski (2007) and Munneke and Slade (2000, 2001) for analyses of selection effects for commercial real estate prices.

coefficients were either negative and significant or positive and significant at the five-percent level.

Starting in the first row, the coefficient on the log of parcel size is strongly significant in all 23 MSAs for commercial/industrial land and in all but two MSAs for residential land. The median value of -0.57 for commercial/industrial land indicates that doubling the size of a parcel reduces price per square foot by 57 percent. An equivalent statement is that doubling parcel size boosts the total price of the parcel by only 43 percent. The median result for residential land is essentially the same. These results confirm the plattage effect – that the price of a land parcel rises less than one-for-one with its size.

The next block of the table shows the differences in price per square foot across the different types of residential land and commercial/industrial land. All of these results represent the price per square foot relative to commercial land, the omitted category. As shown, the three types of residential land sell at prices that are not consistently above or below commercial land prices across the 23 MSAs.[16] For example, land for single-family housing sells at a statistically significant discount to commercial land in six MSAs, at a premium in three others, and with no significant difference in the remaining 14 MSAs. In contrast, industrial land sells at a statistically significant discount to land for commercial development in all 23 MSAs. This discount likely arises because the market-driven distribution of land use in a metropolitan area tends to push warehousing and manufacturing activities to areas that are removed from amenities that enhance property values.

Most of the indicators of the condition of a property have the expected effects on price. Considering both residential and commercial/industrial land, property that has been graded,

[16] The range of coefficients for all three types of residential land has an implausibly high maximum value. The high values come from a single MSA, San Diego, and are offset by an equally large negative value for the San Diego MSA dummy variable. We are investigating the source of this odd result.

finished, fully improved, or previously developed sells at an average price that is 7 to 34 percent above that for unimproved land in the median MSA; these price differentials are each statistically significant for half or more of the 23 MSAs. The other property characteristics have less consistently significant effects across the MSAs, though land that has been paved or platted and engineered tends to sell, as expected, for more than unimproved land. One unanticipated result for commercial/industrial land is the tendency for properties with missing information about land improvements to sell at a discount to unimproved land. This result probably reflects correlations with omitted variables that affect prices more for commercial/industrial land than for residential land.

Among the other variables shown in table 2, land that is intended to be held for investment, kept as open space, or has an unknown use tends to sells for less than land purchased for private development, as expected. In contrast, with only a few exceptions, the price of land purchased for public use is not significantly different than that for land purchased for private development. In the bottom block of the table, the most notable result is that foreclosure sales have little systematic effect on land prices. This result may seem surprising in light of widespread evidence that houses sold at foreclosure carry substantial discounts. However, those discounts may reflect, at least in part, physical deterioration as the homes sit vacant or damage inflicted by the prior owner, factors that are not applicable to land. Finally, in the last row of the table, commercial/industrial land purchased to facilitate an expansion plan for the buyer has a lower price in roughly half of the MSAs than otherwise identical property for reasons that may reflect unobserved features of such transactions.

The results in table 2 indicate that the coefficient estimates for a given variable can span a wide range across MSAs and property types. We conducted an extensive set of hypothesis tests

to determine if these observed differences are statistically significant. The details of the tests can be found in Appendix B. However, the results can be summarized quite briefly. For the variables shown in table 2, the tests overwhelmingly reject the null hypothesis that the coefficients for residential land equal those for commercial/industrial land within the MSAs. They also strongly reject the null hypothesis that the coefficients for residential land are the same across all MSAs; the same result holds for commercial/industrial land. These results imply that the property types and MSAs should not be aggregated when estimating the price effects of the variables shown in table 2.

Locational effects

As discussed above, our regression equation features a hybrid specification of the locational effects on land prices. For each MSA in the regression, we estimate separate locational grids for residential land and commercial/industrial land, along with a distance gradient from the CBD for each type of property. Taking all 23 MSAs together, the regression includes more than 1,800 locational variables (23 measures of distance from the CBD and roughly 900 grid vertices for each of the two property types).

The median value of the distance gradient across MSAs is -0.09 for residential land and -0.25 for commercial/industrial land, so that doubling the distance from the CBD reduces land value by as much as one-quarter, all else equal, in the median MSA. However, as with the results for the hedonic characteristics in table 2, the distance effects vary across MSAs. The distance gradient is negative and significant (at the five-percent level) for residential land in 11 MSAs and for commercial/industrial land in 8 MSAs. At the same time, three MSAs (Los Angeles, San Diego, and South Florida) had positive and significant distance gradients for both commercial/industrial and residential land, while two additional MSAs (Tucson and Phoenix),

had a positive and significant gradient for commercial/industrial land.[17] Overall, these results show that a linear price gradient from the CBD is an important feature of land prices in U.S. metropolitan areas.

However, distance from the CBD does not fully characterize locational effects, as indicated by the significant coefficients on many grid vertices. In about three-quarters of the MSAs, at least ten of the grid vertices for residential land are statistically significant at the five-percent level; the same result holds for commercial/industrial land in just under half of the MSAs. Evidently, the classic model of a monocentric city with a linearly declining price gradient generally does not fit the MSAs in our sample, in particular for residential land. This model may break down for any number of reasons, including the existence of transportation corridors and satellite commercial hubs, differences in the quality of services provided by different localities in the MSA, and the influence of topography and coast lines. Figure 5 illustrates the latter effect with the estimated grids for residential land and commercial/industrial land in South Florida. The back edge of the figure lies along the Atlantic coast, with the highest part of the contour representing Miami Beach. The figure clearly shows a price premium for land near the coast after controlling for distance from the CBD.

In sum, both elements of our hybrid approach are needed to model the spatial patterns of land prices. Distance from the CBD is an essential determinant of prices, but there are other features of the price surface that can only be captured with a more flexible specification.

Price indexes

Given our controls for spatial price effects and for the heterogeneous features of the land parcels and the sales transactions, the coefficients on the half-yearly dummy variables trace out

[17] The finding of positive distance gradients is surprising and stands in contrast to the results in previous drafts of this paper. We are currently examining the source of these results.

time-series indexes of land prices. These price indexes are the most important results in the paper.

Before presenting a variety of these indexes, we show explicitly how we calculate them from the estimated hedonic equation. Note that equation 1 can be written as

(2) $$\ln P = A + \sum_t \gamma_t D_t + \varepsilon$$

where A represents all the terms in the equation that are not functions of time and we have suppressed the subscripts for MSAs and property types. Let t_0 denote a chosen base period and let t_1 denote any other period. Then, the difference in the fitted values between the two periods can be written as

(3) $$\ln \hat{P}_{t_1} - \ln \hat{P}_{t_0} = \left(\hat{A} + \hat{\gamma}_{t_1}\right) - \left(\hat{A} + \hat{\gamma}_{t_0}\right) = \hat{\gamma}_{t_1} - \hat{\gamma}_{t_0}$$

so that

(4) $$\hat{P}_{t_1} \big/ \hat{P}_{t_0} \equiv \exp\left(\ln \hat{P}_{t_1} - \ln \hat{P}_{t_0}\right) = \exp\left(\hat{\gamma}_{t_1}\right) \big/ \exp\left(\hat{\gamma}_{t_0}\right).$$

Equation 4 defines the price indexes that we present below. They are functions solely of the estimated coefficients on the half-yearly dummies and are indexed to equal one in a selected period.

The top panel in figure 6 presents the land price index from a restricted version of the hedonic regression that includes a single set of half-yearly dummies. The resulting price index represents a composite index for residential and commercial/industrial land in all 23 MSAs. As shown, this price index trended up from 1995 to 2002, rising at an average annual rate of about 6 percent. Prices then surged, increasing nearly 130 percent from the second half of 2002 to the series peak in the second half of 2007. Since then, the price index has declined more than 35

percent. As of the first half of 2009, the price index had retraced roughly two-thirds of its post-2002 run-up, before leveling off in the second half of the year.

The lower panel shows separate price indexes for commercial/industrial and residential land. To obtain these indexes, we re-estimated the regression using two sets of time dummies, one for commercial/industrial land in all MSAs and another for residential land in all MSAs. Both indexes display the same broad patterns as the composite index in the upper panel – a moderate uptrend through 2002, followed by a sharp acceleration that lasted through 2007 and then a steep reversal. However, the index for residential land displays greater amplitude than the index for commercial/industrial land, and it continued to decline in the second half of 2009, while the commercial/industrial price index turned up. A likelihood ratio test decisively rejects the hypothesis that the residential and the commercial/industrial price indexes are equal.

The basic features of the land price indexes estimated by Sirmans and Slade (2009) are similar to those in figure 6. In particular, both sets of indexes indicate that land prices surged at the national level after 2002 and have fallen sharply in recent years. There are, however, some notable differences in the indexes. The total increase in the Sirmans-Slade indexes from 2002 to the peak, while substantial, is not as large as in our indexes, and their indexes peak earlier than ours. In particular, their composite index of commercial, industrial, and residential land prices peaked in late 2005, when real estate markets were still booming, which seems less plausible than the late 2007 peak in our composite index.

Figure 7 presents our land price indexes for the individual MSAs. These price indexes cover an aggregate of commercial/industrial and residential land in the MSA and are calculated using MSA-specific time dummies in the regression. The upper row of the figure shows the indexes for MSAs on the East Coast while the middle row presents the indexes for West Coast

MSAs and the bottom row shows the indexes for MSAs located in the interior of the country. All the series are indexed to equal 100 in 2002:H2 and every panel has the same scale, so the magnitude of the post-2002 run-up in land prices and the subsequent decline can be compared across MSAs.

The most striking feature of figure 7 is that the swing in land prices generally has been much larger on the coasts and in Nevada and Arizona than elsewhere in the country, a pattern that mirrors the boom-bust cycle in the housing market in recent years. Another notable result is that, in all but a handful of the MSAs, land prices have unwound most or all of the earlier price jump. Although prices continued to drop through the second half of 2009 in roughly half of the MSAs, they flattened out in Washington D.C. and Los Angeles and turned up in several other MSAs.

Table 3 provides additional information on the MSA-level land price indexes. As shown, the price indexes for all but three MSAs reached a peak in 2006 or 2007, with the median peak date across the MSAs occurring in the first half of 2007. The total price increase from the second half of 2002 to the peak ranged from a low of 52 percent in Denver, 76 percent in Dallas, and 78 percent in Detroit to more than 200 percent in Las Vegas, New York, and Washington DC; the median rise across MSAs was 150 percent. Although the range is extremely wide, even the increases at the low end of the range are not small in any absolute sense. The table also shows the extent of the price decline from the MSA-specific peaks to the series low. Eight of the 23 MSAs hit their post-peak low in the first half of 2009, while prices in the remainder continued to decline through the second half of the year. The cumulative peak-to-low decline ranged from 11 percent in Denver to 59 percent in Las Vegas and New York, with a median price drop of 41 percent across the MSAs. We should note, however, that the standard errors for the 2009:H1 and

2009:H2 dummy variables that help determine these price declines are larger than for earlier periods because of the limited volume of land sales in recent years.

Figure 8 and table 4 provide analogous results for the price indexes for commercial/industrial land by MSA. These series, and the residential price indexes discussed below, are calculated from the unrestricted version of the regression equation that allows the coefficients of the half-yearly dummies to vary across both MSAs and property type. The MSA-level movements in commercial/industrial land prices shown in figure 8 are qualitatively similar to those displayed in figure 7 for the aggregate indexes. Both figures show that the swings in land prices generally have been wider in the MSAs on the East Coast and in the Far West than elsewhere in the country.

Comparing tables 3 and 4 reveals further information about the two sets of land price indexes. With regard to the timing of turning points, the peak in commercial/industrial land prices occurred in the first half of 2007 for the median MSA, the same as for aggregate land prices. However, the commercial/industrial index reached its post-peak low for the median MSA in the first half of 2009, while the aggregate price index was still falling in most MSAs through the second half of the year. The price swings also have been a bit less extreme for the commercial/industrial indexes than for the aggregate indexes. For the median MSA, commercial/industrial land prices rose 128 percent from 2002:H2 to the peak, about 20 percentage points less than the median rise for aggregate land prices. Since the peak, the median price decline to the subsequent low was 38 percent for the commercial/industrial index, a shade less than the median drop of 41 percent for the aggregate series; this comparison, though, likely understates the difference that ultimately will emerge because in most MSAs the aggregate price index was still falling at the end of our sample period.

Finally, figure 9 and table 5 present the MSA-level results for the residential price indexes. For most MSAs, prices jumped even more sharply for residential land than for commercial/industrial land. Taking the median across MSAs, residential land prices increased 187 percent from the second half of 2002 to the peak, compared with the 128 percent rise for commercial/industrial land prices. In addition, prices generally peaked earlier for residential land than for commercial/industrial land; for the median MSA, residential land prices peaked in the second half of 2006, a half-year before the peak in the commercial/industrial price series. The steep decline in transaction volume since 2006 limits the number of MSAs for which we can estimate the post-peak behavior of residential land prices. Nonetheless, the MSAs with sufficient data have experienced a sharp drop in land prices, with a median decline from the peak of 57 percent. Prices in most MSAs were continuing to decline in the second half of 2009.

An important issue is how these movements in land prices compare with the price changes for housing and commercial/industrial real estate over the same period. In their assessments of this issue, Davis and Heathcote (2007), Davis and Palumbo (2008), and Sirmans and Slade (2009) found that the price swings for residential land have been wider than those for home prices over various periods. Sirmans and Slade (2009) obtained the same result when comparing their price index for industrial land to standard price indexes for industrial real estate.[18]

Table 6 compares the land price indexes estimated in this study to well-known indexes of commercial real estate and house prices. The land price indexes are the aggregate indexes for commercial/industrial and residential land that were plotted in the lower panel of figure 6. We compare the index of residential land prices to the S&P/Case-Shiller 20-city index of home

[18] However, their comparison of relative price movements for the non-industrial part of the commercial sector was inconclusive because of differences in the sectoral coverage of the indexes.

prices. In addition, we compare the index of commercial/industrial land prices to two indexes of commercial real estate prices: the National Council of Real Estate Investment Fiduciaries (NCREIF) transaction-based index and the Moody's/REAL commercial property price index.[19]

As shown in the table, the index of residential land prices is estimated to have risen roughly three times as much as house prices from 2002:H2 to the respective peaks in the series, and then to have fallen about one and a half times as much from the peak.[20] The greater amplitude of residential land prices relative to home prices is consistent with the results from earlier studies. Because the Case-Shiller index and other house price indexes cover a bundle of land and structures, the results in this paper and elsewhere imply that residential land prices have been more variable than the prices of housing structures. With regard to the price indexes for the commercial/industrial sector, land prices rose substantially more from the second half of 2002 to the peak than did either of the measures of property prices, echoing the results for residential sector, though the differences are less extreme. In contrast, the price decline since the peak has been roughly the same for all three commercial/industrial indexes. This pattern may reflect differences in the access to credit for buyers of land, who borrow almost exclusively from banks, and buyers of existing properties, who had come to depend heavily on the market for commercial mortgage-backed securities, which shut down in 2008 and was still essentially closed at the end of 2009. Further research on the recent experience using MSA-level data would be valuable.

6. Conclusions and Directions for Future Work

[19] Information about both indexes can be found at web.mit.edu/cre/research/credl. For both indexes, the table shows an aggregation of the sub-indexes for office, retail, and industrial properties. We exclude apartment properties to match the coverage of the commercial/industrial land price index.

[20] The S&P/Case-Shiller index excludes multifamily properties, while the residential land index includes multifamily land parcels. This coverage difference, however, does not appear to distort the comparison between the residential land and house price indexes. NCREIF and Moody's/REAL both construct indexes for apartment prices. Averaging the two indexes, the price increase for apartments from the second half of 2002 to the peak was 63 percent and the decline from the peak to the first half of 2009 was 34 percent; both figures closely track the changes in the S&P/Case-Shiller index. Thus, a broader measure of home prices that included apartments would still be considerably less volatile than the index for residential land prices.

This paper constructs land price indexes for a broad set of metropolitan areas in the United States. To calculate the indexes, we estimate a hedonic regression for land prices in 23 large MSAs with a sample of roughly 160,000 land transactions from the mid-1990s to the end of 2009. The regressions control for a variety of characteristics of the land parcels and the sales transactions; they also feature a flexible method of controlling for spatial price patterns within an MSA. Given these controls, the half-yearly dummy variables in the regressions trace out the implied land price indexes.

The resulting indexes show a dramatic increase in both residential and commercial/industrial land prices over several years prior to their peaks in 2006-07 and a steep descent since then. The magnitude of the run-up and the subsequent decline differs across the MSAs, with the largest movements in MSAs on the East Coast and in the Far West. Another key result is that, for the most part, the swings in land prices have been considerably larger than in well-known indexes of commercial real estate and house prices. Because those indexes price a bundle of land and structures, this comparison implies that land prices generally have been more volatile than structures prices over this period.

The results reported in this paper represent the initial findings from a larger research agenda. At this stage, we have documented the substantial swings in land prices but have not analyzed the sources of these movements. The next step is to examine the degree to which land prices have been driven by the availability and cost of financing, the use of leverage in property transactions, supply and demand fundamentals in real estate markets, and broader economic conditions. In addition, we intend to conduct a more rigorous comparison of the movements in the prices of land and structures. This can be done by combining the CoStar data for commercial

real estate and land transactions into single dataset, which would allow us to construct price indexes for commercial/industrial land and structures in an integrated framework.

References

Barker, David. 2007. "Urban Land Rents in the United States". In *Land Policies and Their Outcomes,* edited by Gregory K. Ingram and Yu-Hung Hong. Cambridge, MA: Lincoln Institute of Land Policy.

Bostic, Raphael W., Stanley D. Longhofer, and Christian L. Redfearn. 2007. "Land Leverage: Decomposing Home Price Dynamics." *Real Estate Economics* 35(2): 183-208.

Brownstone, David, and Arthur De Vany. 1991. "Zoning, Returns to Scale, and the Value of Undeveloped Land." *Review of Economics and Statistics* 73(4): 699-704.

Bryan, Kevin A., and Pierre-Daniel G. Sarte. 2009. "Semiparametric Estimation of Land Price Gradients Using Large Data Sets." *Federal Reserve Bank of Richmond Economic Quarterly* 95(1): 53-74.

Bureau of Labor Statistics. 2007. "Technical Information About the BLS Multifactor Productivity Measures." (September 26). Posted at www.bls.gov/mfp/mprtech.pdf

Case, Karl E. 2007. "The Value of Land in the United States: 1975-2005". In *Land Policies and Their Outcomes*, edited by Gregory K. Ingram and Yu-Hung Hong. Cambridge, MA: Lincoln Institute of Land Policy.

Colwell, Peter F. 1998. "A Primer on Piecewise Parabolic Multiple Regression Analysis via Estimations of Chicago CBD Land Prices." *Journal of Real Estate Finance and Economics* 17(1): 87-97.

Colwell, Peter F., and Henry J. Munneke. 2003. "Estimating a Price Surface for Vacant Land in an Urban Area." *Land Economics* 79(1): 15-28.

Colwell, Peter F., and Henry J. Munneke. 1997. "The Structure of Urban Land Prices." *Journal of Urban Economics* 41(3): 321-36.

Davis, Morris A. 2009. "The Price and Quantity of Land by Legal Form of Organization in the United States." *Regional Science and Urban Economics* 39(3): 350-9.

Davis, Morris A., and Jonathan Heathcote. 2007. "The Price and Quantity of Residential Land in the United States." *Journal of Monetary Economics* 54(8): 2595-2620.

Davis, Morris A., and Michael G. Palumbo. 2008. "The Price of Residential Land in Large US Cities." *Journal of Urban Economics* 63(1): 352-84.

Downing, Paul B. 1970. "Estimating Residential Land Value by Multivariate Analysis." In *The Assessment of Land Value*, edited by Daniel M. Holland. Madison, WI: University of Wisconsin Press.

Fisher, Jeff, David Geltner, and Henry Pollakowski. 2007. "A Quarterly Transactions-based Index of Institutional Real Estate Investment Performance and Movements in Supply and Demand." *Journal of Real Estate Finance and Economics* 34: 5-33.

Gatzlaf, Dean H., and Donald R. Haurin. 1998. "Sample Selection and Biases in Local House Value Indices." *Journal of Urban Economics* 43: 199-222.

Gatzlaf, Dean H., and Donald R. Haurin. 1997. "Sample Selection Bias and Repeat-Sales Index Estimates." *Journal of Real Estate Finance and Economics* 14: 33-50.

Goldsmith, Raymond W. 1951. "A Perpetual Inventory of National Wealth." In *Studies in Income and Wealth, vol. 14*, edited by Martin R. Gainsbrugh and Elwyn T. Bonnell: 5-61. New York: National Bureau of Economic Research.

Greenlees, John S. 1980. "Residential Land Price Indices Using Multiple Regression." In *Urban Land Markets: Price Indices, Supply Measures, and Public Policy Effects*, edited by J. Thomas Black and James E. Hoben. Washington DC: The Urban Land Institute.

Guntermann, Karl L., and Gareth Thomas. 2005. "Parcel Size, Location, and Commercial Land Values." *Journal of Real Estate Research* 27(3): 343-54.

Haughwout, Andrew, James Orr, and David Bedoll. 2008. "The Price of Land in the New York Metropolitan Area." Federal Reserve Bank of New York, *Current Issues in Economics and Finance* 14(3): 1-7

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47: 153–61

Ihlanfeldt, Keith R. 2007. "The Effect of Land Use Regulation on Housing and Land Prices." *Journal of Urban Economics* 61(3): 420-35.

Isakson, Hans R. 1997. "An Empirical Analysis of the Determinants of the Value of Vacant Land." *Journal of Real Estate Research* 13(2): 103-14.

Jud, G. Donald, and Terry G. Seaks. 1994. "Sample Selection Bias in Estimating Housing Sales Prices." *The Journal of Real Estate Research* 9(3): 289-98.

Kowalski, Joseph G., and Christos C. Paraskevopoulos. 1990. "The Impact of Location on Urban Industrial Land Prices." *Journal of Urban Economics* 27(1): 16-24

Lee, Seung Jung, and Jonathan D. Rose. 2010. "Profits and Balance Sheet Developments at U.S. Commercial Banks in 2009." *Federal Reserve Bulletin* 96 (May): A1-37. Posted at www.federalreserve.gov/pubs/bulletin/2010/default.htm.

Manvel, Allan D. 1968. "Trends in the Value of Real Estate and Land, 1956 to 1966." In *Three Land Research Studies*, prepared for the consideration of the National Commission on Urban Problems. Washington DC: Government Printing Office.

Munneke, Henry J., and Barrett A. Slade. 2001. "A Metropolitan Transaction-Based Commercial Price Index: A Time-Varying Parameter Approach." *Real Estate Economics* 29(1): 55-84.

Munneke, Henry J., and Barrett A. Slade. 2000. "An Empirical Study of Sample-Selection Bias in Indices of Commercial Real Estate." *Journal of Real Estate Finance and Economics* 21(1): 45-64.

Peiser, Richard B. 1987. "The Determinants of Nonresidential Land Values." *Journal of Urban Economics* 22(3): 340-60.

Poirier, Dale J. 1976. *The Econometrics of Structural Change*. New York: North Holland.

Rosenthal, Stuart S., and Robert W. Helsley. 1994. "Redevelopment and the Urban Land Price Gradient." *Journal of Urban Economics* 35(2): 182-200.

Sirmans, C.F., and Barrett A. Slade. 2009. "National Transaction-based Land Price Indices." Working paper (December 14).

Voith, Richard. 2001. "How Responsive Is the Demand for Residential Land to Changes in Its Price?" *Federal Reserve Bank of Philadelphia Business Review*, Q3: 33-9.

Wieand, Kenneth, Jr., and Richard F. Muth. 1972. "A Note on the Variation of Land Values with Distance from the CBD in St. Louis." *Journal of Regional Science* 12: 469-73.

Table 1
Characteristics of the Sample

		Number of sales		
MSA	**Starting period**	**Total**	**Commercial/ Industrial**	**Residential**
Total	**---**	**159,216**	**86,165**	**70,667**
Atlanta	1995:H1	15,881	7,879	8,002
Baltimore	1995:H2	3,365	1,740	1,446
Boston	1995:H1	3,342	1,783	1,665
Chicago	1995:H1	11,987	7,072	4,915
Dallas	1995:H2	4,032	3,249	553
Denver	1995:H2	7,663	4,182	3,441
Detroit	1999:H2	2,791	1,748	961
Houston	1999:H2	2,185	1,805	178
Las Vegas	1995:H1	9,959	4,645	5,297
Los Angeles	1995:H1	12,224	8,353	3,871
New York	1998:H2	6,309	3,612	2,667
Orlando	1995:H1	7,071	4,068	2,785
Philadelphia	1998:H1	4,875	2,376	2,344
Phoenix	1995:H1	16,530	7,733	8,797
Portland	1995:H1	4,667	1,739	2,838
Sacramento	1995:H1	2,405	1,606	416
San Diego	1995:H1	3,041	1,957	882
San Francisco	1995:H1	4,134	2,846	1,153
Seattle	1995:H1	8,247	3,315	4,907
South Florida	1997:H2	8,961	5,890	2,984
Tampa/St. Petersburg	1997:H2	6,396	4,023	2,314
Tucson	1995:H1	3,523	1,230	2,199
Washington DC	1995:H1	9,428	3,314	6,052

Note. Los Angeles is defined to include Orange County and the Inland Empire; New York is defined to include northern New Jersey, Westchester County, and Long Island; and San Francisco is defined to include Marin/Sonoma, East Bay/Oakland, and South Bay/San Jose. The number of observations in the "Total" column does not equal the sum of the observations in the "Commercial/Industrial" and "Residential" columns because of the sample construction rules we applied. See the text for details.
Source. Authors' analysis of data from the CoStar Group, Inc. (www.costar.com).

Table 2
Coefficient Estimates Across MSAs

Variable	Residential					Commercial/Industrial				
	Median	Range		# Significant[1]		Median	Range		# Significant[1]	
		Min.	Max.	Neg.	Pos.		Min.	Max.	Neg.	Pos.
Log of parcel size	-.56	-.69	-.23	21	0	-.57	-.69	-.38	23	0
Type of property										
Single-Family	-.19	-.93	25.0	6	3					
Multifamily Rental	.05	-.78	25.2	2	5					
Other Residential	-.32	-1.05	24.8	8	2					
Industrial						-.34	-.54	-.13	23	0
Condition of property										
Graded	.21	-.10	.63	0	15	.07	-.04	.25	0	9
Paved	.15	-.61	.95	1	6	.12	-.44	.35	0	9
Finished	.34	.02	.77	0	19	.08	-.24	.30	0	12
Fully improved	.25	-.27	1.05	0	10	.22	.05	.66	0	12
Platted and engineered	.21	-.99	.87	0	8	.14	-.76	.60	0	4
Previously developed	.25	-.11	.64	0	16	.19	-.05	.41	0	17
Improvements unknown	.01	-.22	.24	4	4	-.07	-.21	.05	12	0
Structure present	.01	-.14	.02	2	1	.07	-.11	.17	0	9
Environmental problems	.00	-1.39	.79	1	0	-.13	-1.39	.32	4	0
Intended use										
Hold for investment	-.22	-.34	-.00	12	0	-.21	-.38	.08	20	0
Public use	-.09	-.49	.82	3	1	.01	-.32	.32	3	5
Open space	-.34	-1.24	.21	13	0	-.39	-1.00	.94	12	0
Unknown	-.10	-.31	-.11	8	0	-.14	-.36	-.08	18	0
Characteristics of transaction										
Foreclosure transaction	.08	-1.06	1.60	0	3	-.08	-1.11	.53	2	2
Sold as part of expansion plan	-.06	-.68	.41	1	2	-.14	-.50	.12	10	0

1. At the five-percent level.
Note. The omitted property type for both the residential and commercial/industrial observations is "office, retail, and other commercial". The omitted condition of property is "unimproved", and the omitted intended use is "private development". Some MSAs lacked observations to estimate every coefficient. For example, six MSAs had no residential land sales reported with environmental problems; the results in that line of the table are based on estimates for the other 17 MSAs.
Source. Authors' analysis of data from the CoStar Group, Inc. (www.costar.com).

Table 3
Price Indexes for Composite of Commercial/Industrial and Residential Land by MSA

	Date of		Percent change	
MSA	Peak	Post-peak low	2002:H2 to peak	Peak to low
Atlanta	2007:H1	2009:H2	84	-38
Baltimore	2007:H2	2009:H1	155	-47
Boston	2007:H2	2009:H2	164	-46
Chicago	2006:H2	2009:H2	85	-41
Dallas	2006:H2	2009:H1	76	-38
Denver	2007:H1	2009:H2	52	-11
Detroit	2005:H2	2009:H1	78	-57
Houston	2005:H2	2009:H1	111	-32
Las Vegas	2006:H1	2009:H1	241	-59
Los Angeles	2006:H2	2009:H1	151	-46
New York	2007:H2	2009:H2	228	-59
Orlando	2006:H2	2009:H2	155	-40
Philadelphia	2005:H2	2009:H1	144	-29
Phoenix	2006:H1	2009:H1	155	-44
Portland	2007:H1	2009:H2	131	-32
Sacramento	2006:H2	2009:H2	132	-44
San Diego	2007:H1	2009:H2	80	-15
San Francisco	2006:H2	2009:H2	194	-52
Seattle	2007:H2	2009:H2	155	-40
South Florida	2007:H1	2009:H2	168	-48
Tampa	2007:H1	2009:H2	150	-36
Tucson	2007:H1	2009:H2	86	-17
Washington DC	2007:H1	2009:H2	332	-58
Median across MSAs	**2007:H1**	**2009:H2**	**150**	**-41**

Note. See table 1 for definitions of selected MSAs.
Source. Authors' analysis of data from the CoStar Group, Inc. (www.costar.com).

Table 4
Price Indexes for Commercial/Industrial Land by MSA

	Date of		Percent change	
MSA	**Peak**	**Post-peak low**	**2002:H2 to peak**	**Peak to low**
Atlanta	2008:H1	2009:H1	82	-38
Baltimore	2006:H1	2009:H2	101	-19
Boston	2005:H2	2009:H1	188	-69
Chicago	2005:H2	2009:H1	89	-58
Dallas	2006:H2	2009:H1	81	-33
Denver	2008:H1	2009:H1	62	-11
Detroit	2005:H1	2009:H2	53	-43
Houston	2006:H1	2009:H1	90	-18
Las Vegas	2007:H1	2009:H1	228	-64
Los Angeles	2006:H2	2009:H1	152	-42
New York	2007:H2	2009:H1	203	-38
Orlando	2006:H2	2009:H2	138	-34
Philadelphia	2007:H2	2009:H1	142	-39
Phoenix	2007:H2	2009:H1	122	-42
Portland	2007:H1	2009:H2	78	-27
Sacramento	2008:H1	2009:H2	142	-38
San Diego	2007:H1	2009:H2	128	-17
San Francisco	2008:H1	2009:H1	135	-38
Seattle	2006:H2	2009:H1	159	-22
South Florida	2007:H2	2009:H1	209	-51
Tampa	2007:H1	2009:H1	102	-16
Tucson	2008:H1	2009:H2	73	-13
Washington DC	2007:H1	2009:H2	251	-58
Median across MSAs	**2007:H1**	**2009:H1**	**128**	**-38**

Note. See table 1 for definitions of selected MSAs.
Source. Authors' analysis of data from the CoStar Group, Inc. (www.costar.com).

Table 5
Price Indexes for Residential Land by MSA

	Date of		Percent change	
MSA	Peak[1]	Post-peak low[2]	2002:H2 to peak[1]	Peak to low
Atlanta	2008:H1	2009:H2	91	-41
Baltimore	NA	NA	NA	NA
Boston	2005:H2	NA	105	NA
Chicago	2005:H2	2009:H1	103	-58
Dallas	NA	NA	NA	NA
Denver	2008:H1	2009:H2	63	-35
Detroit	NA	NA	NA	NA
Houston	NA	NA	NA	NA
Las Vegas	2007:H1	2009:H2	335	-68
Los Angeles	2006:H2	2009:H2	153	-60
New York	2007:H2	2009:H1	330	-76
Orlando	2005:H2	2009:H2	215	-58
Philadelphia	2005:H2	2009:H2	194	-43
Phoenix	2006:H2	2009:H1	247	-57
Portland	2006:H2	2009:H2	249	-51
Sacramento	NA	NA	NA	NA
San Diego	2005:H2	NA	271	NA
San Francisco	2006:H2	NA	236	NA
Seattle	2006:H2	2009:H2	198	-29
South Florida	2006:H2	2009:H2	206	-61
Tampa	2005:H2	2009:H2	187	-62
Tucson	2008:H1	2009:H2	90	-18
Washington DC	2006:H2	2009:H2	192	-24
Median across MSAs	**2006:H2**	**2009:H2**	**187**	**-57**

1. Calculated only for MSAs for which the price index is available through at least 2006:H2 or for which the available data prior to 2006:H2 indicate an earlier peak.
2. Calculated only for MSAs for which the price index is available in 2009:H1, 2009:H2, or both periods.
Note. See table 1 for definitions of selected MSAs. NA indicates that the price index is not available.
Source. Authors' analysis of data from the CoStar Group, Inc. (www.costar.com).

Table 6
Price Indexes for Land, Commercial Real Estate, and Housing

Index	Date of: Peak	Date of: Post-peak low	Percent change: 2002:H2 to peak	Percent change: Peak to low
Commercial/industrial				
Land price index	2007:H2	2009:H1	112	-33
NCREIF transaction-based index	2007:H1	2009:H2	82	-35
Moody's/REAL CPPI	2007:H2	2009:H2	69	-31
Residential				
Land price index	2007:H2	2009:H2	158	-49
S&P/Case-Shiller 20-city home price index	2006:H2	2009:H1	50	-30

Note. The land price indexes are those calculated for the aggregate of all 23 MSAs, which were shown in the lower panel of figure 6. CPPI stands for commercial property price index. The figures for the NCREIF and Moody's/REAL indexes cover office, retail, and industrial properties. The S&P/Case-Shiller index covers single-family homes. All figures in the table are calculated from data that are not seasonally adjusted.
Sources. Land price indexes: Authors' analysis of data from the Costar Group, Inc. (www.costar.com). NCREIF index: MIT Center for Real Estate. Moody's/REAL index: Moody's Investors Service. S&P/Case-Shiller index: Standard and Poor's.

Figure 1
Median Half-Yearly Sample Size Across MSAs

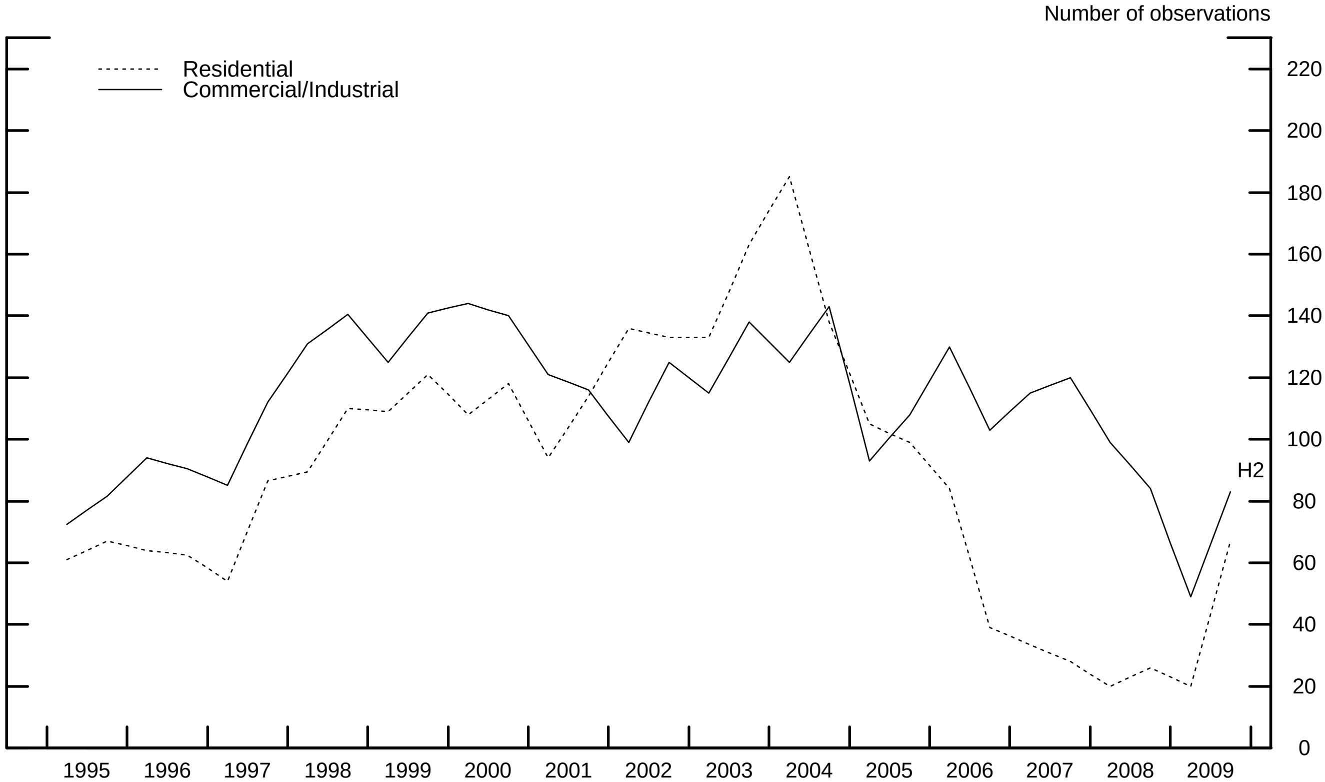


Figure 2
Illustration of Vertex Weights

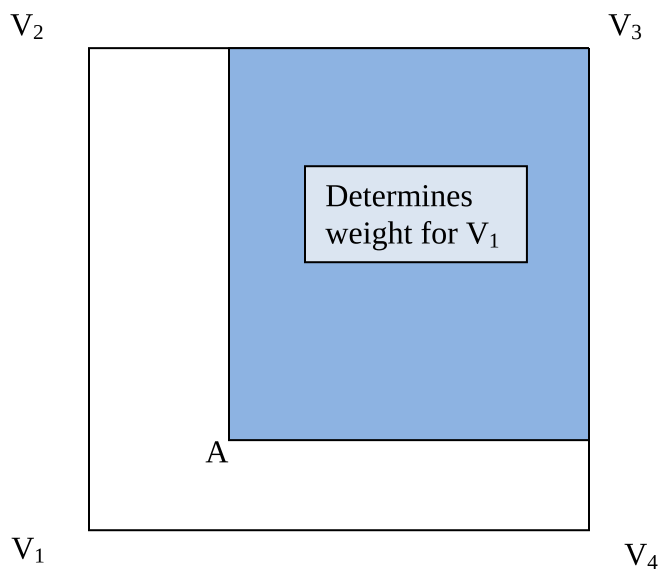


Figure 3
Locational Grid for Dallas

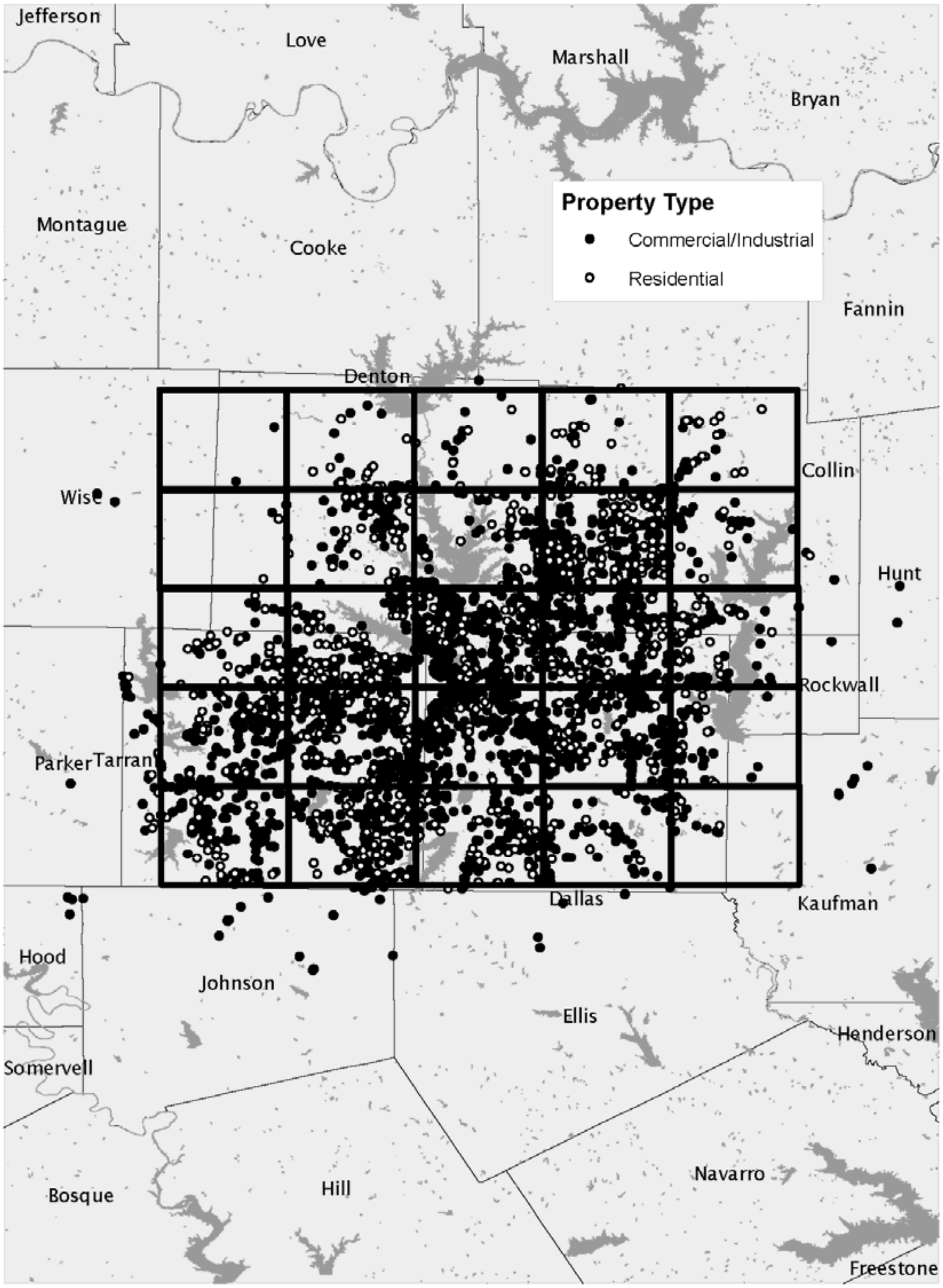


Figure 4
Locational Grid for South Florida

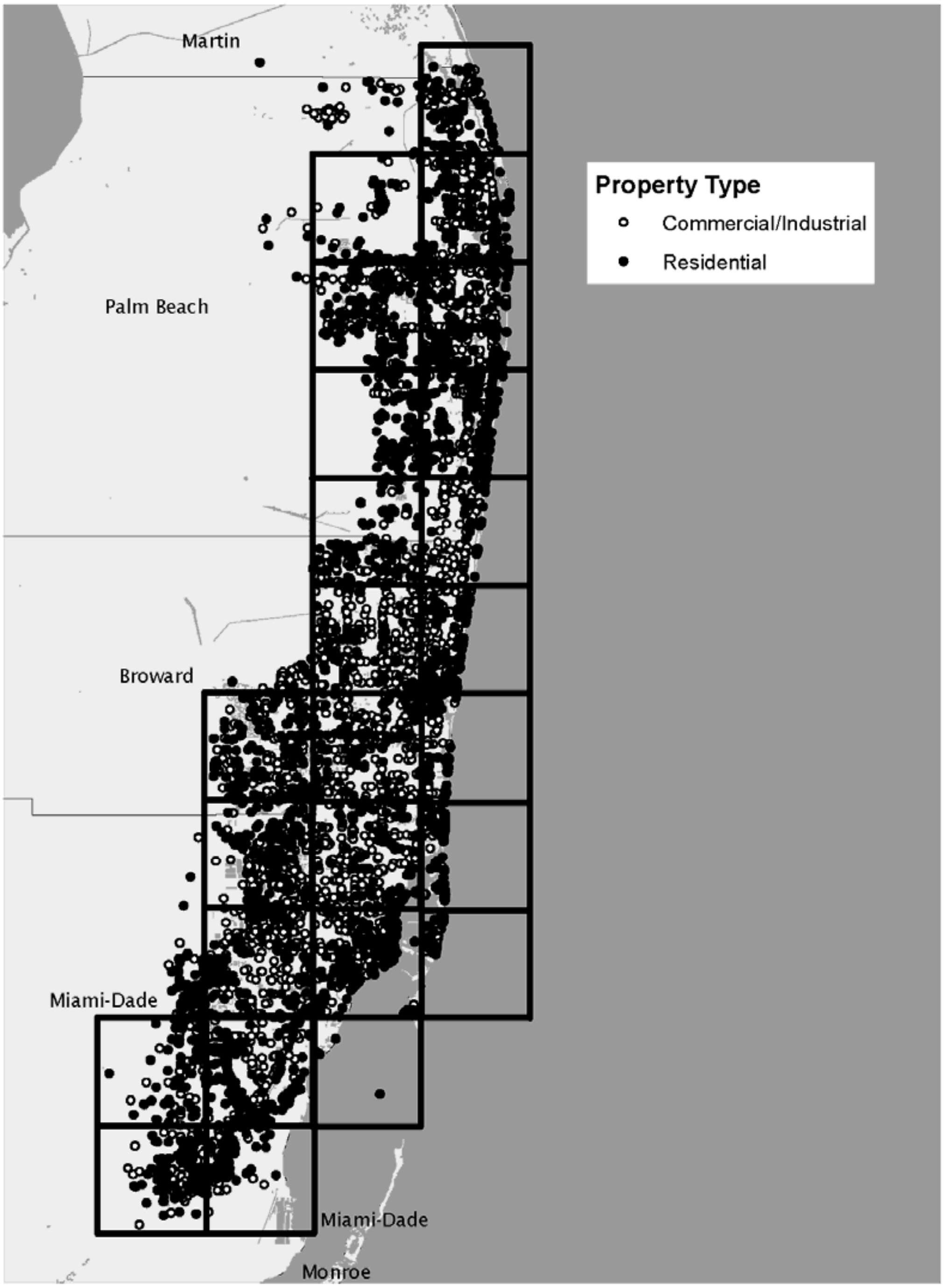
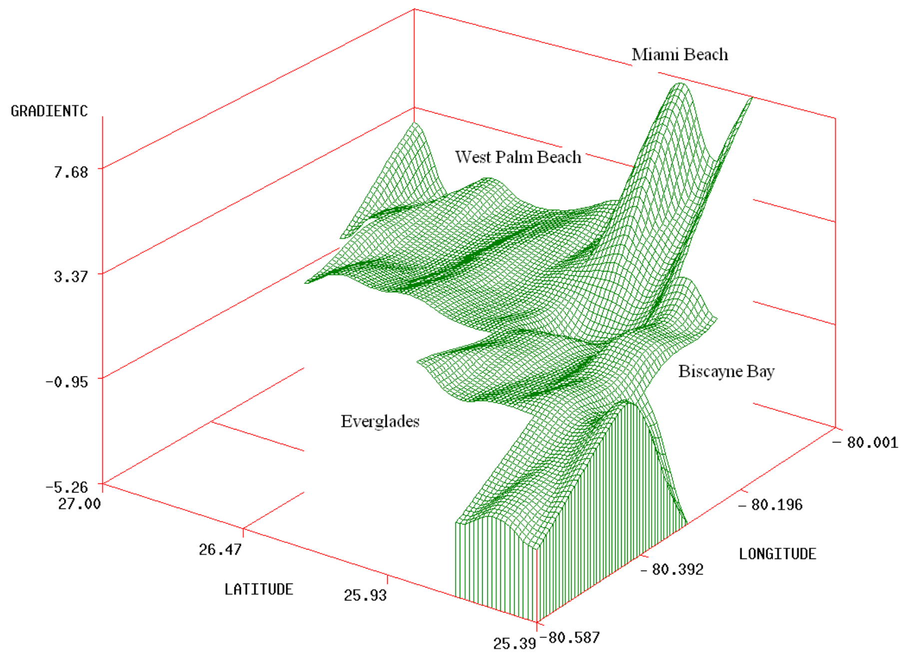


Figure 5
Land Price Surface for South Florida

Commercial/Industrial



Residential

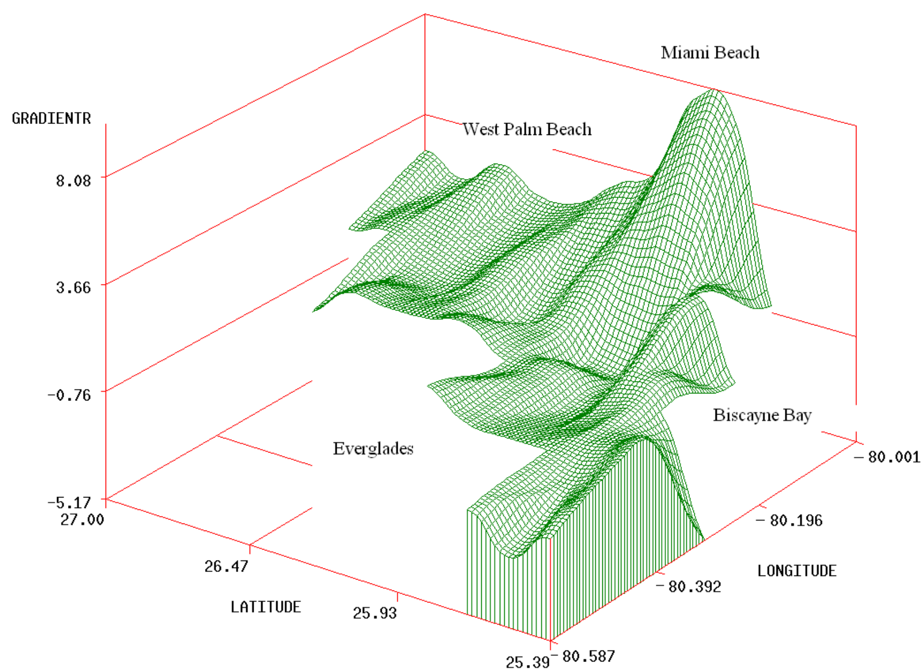
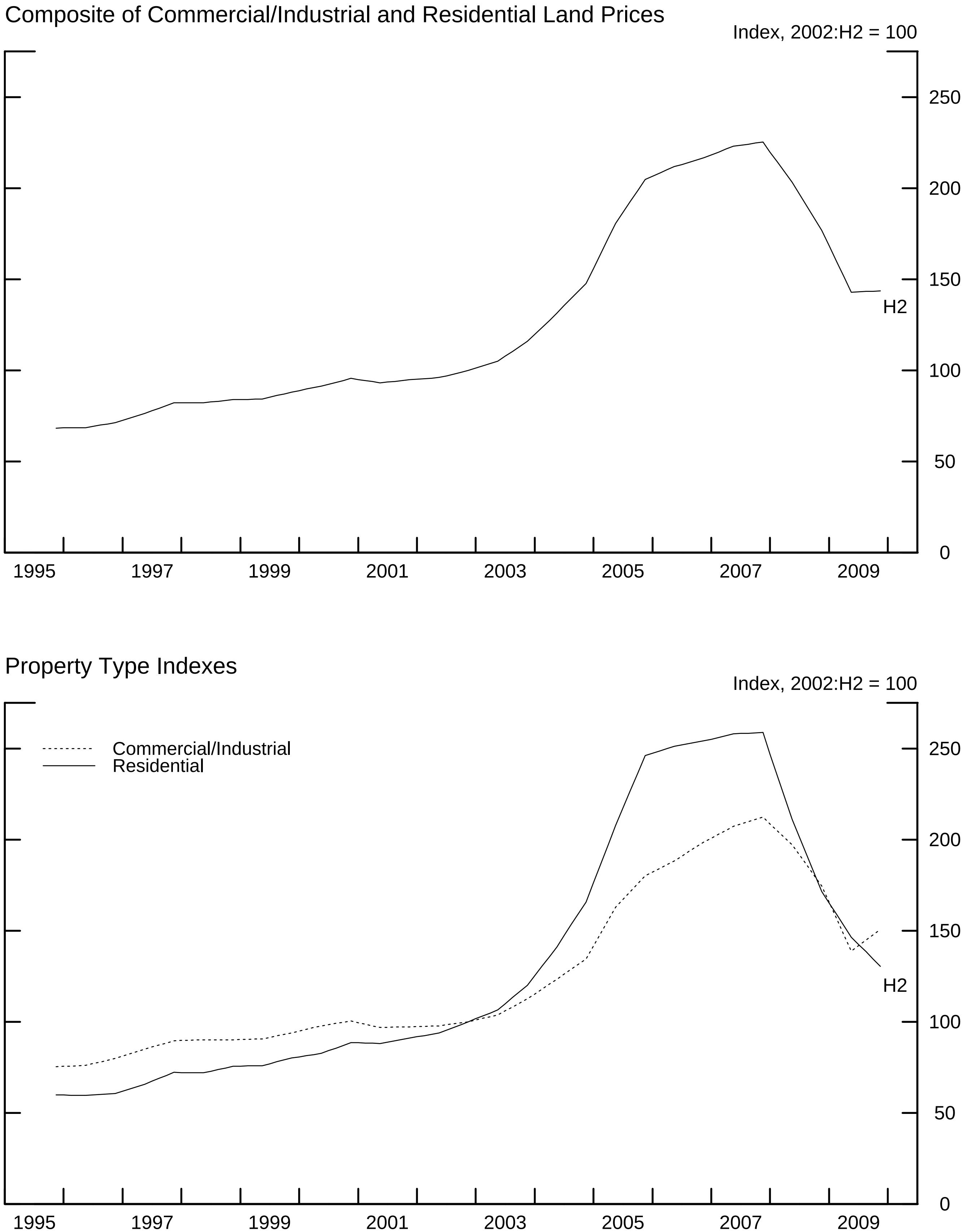
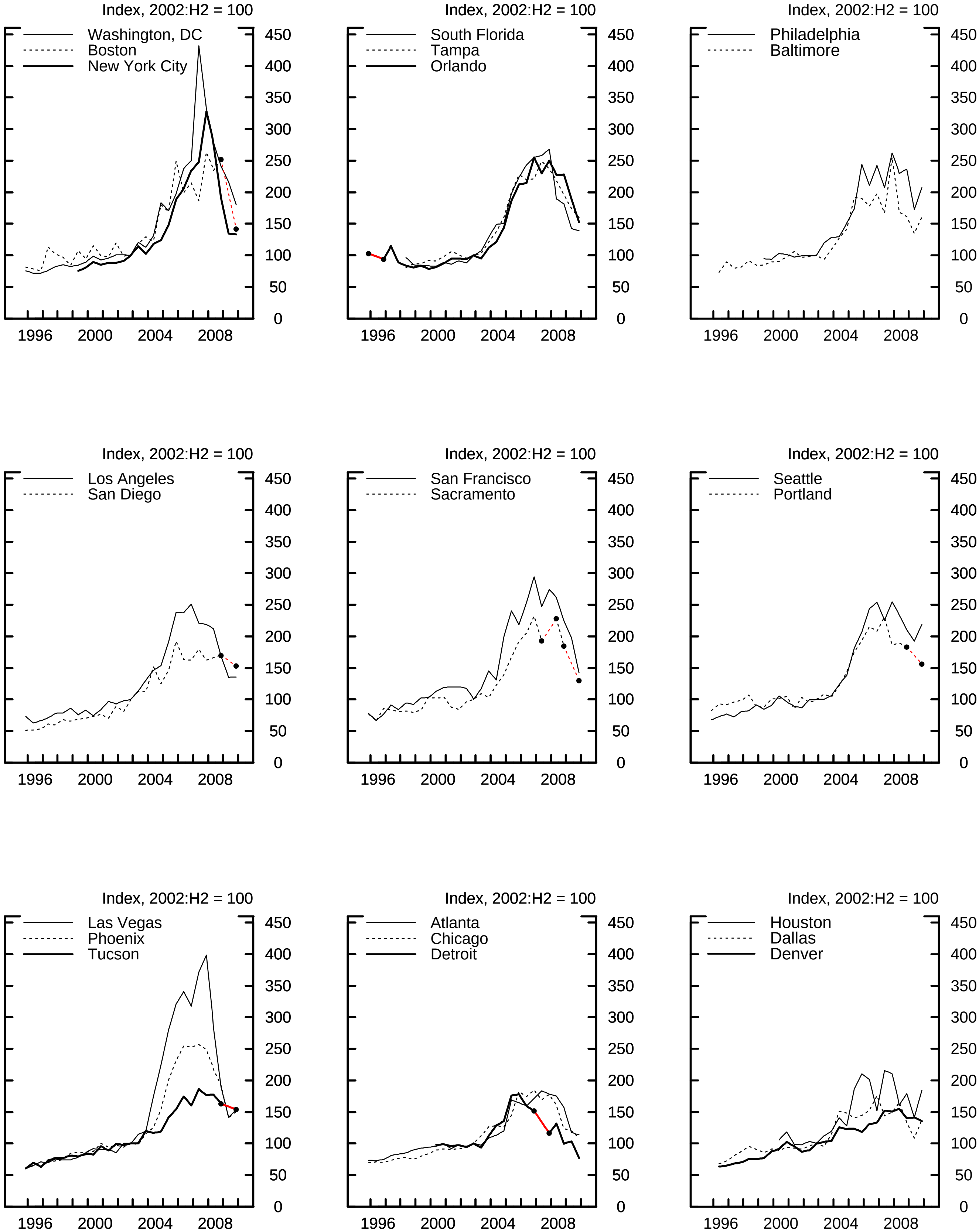


Figure 6
National Land Price Indexes



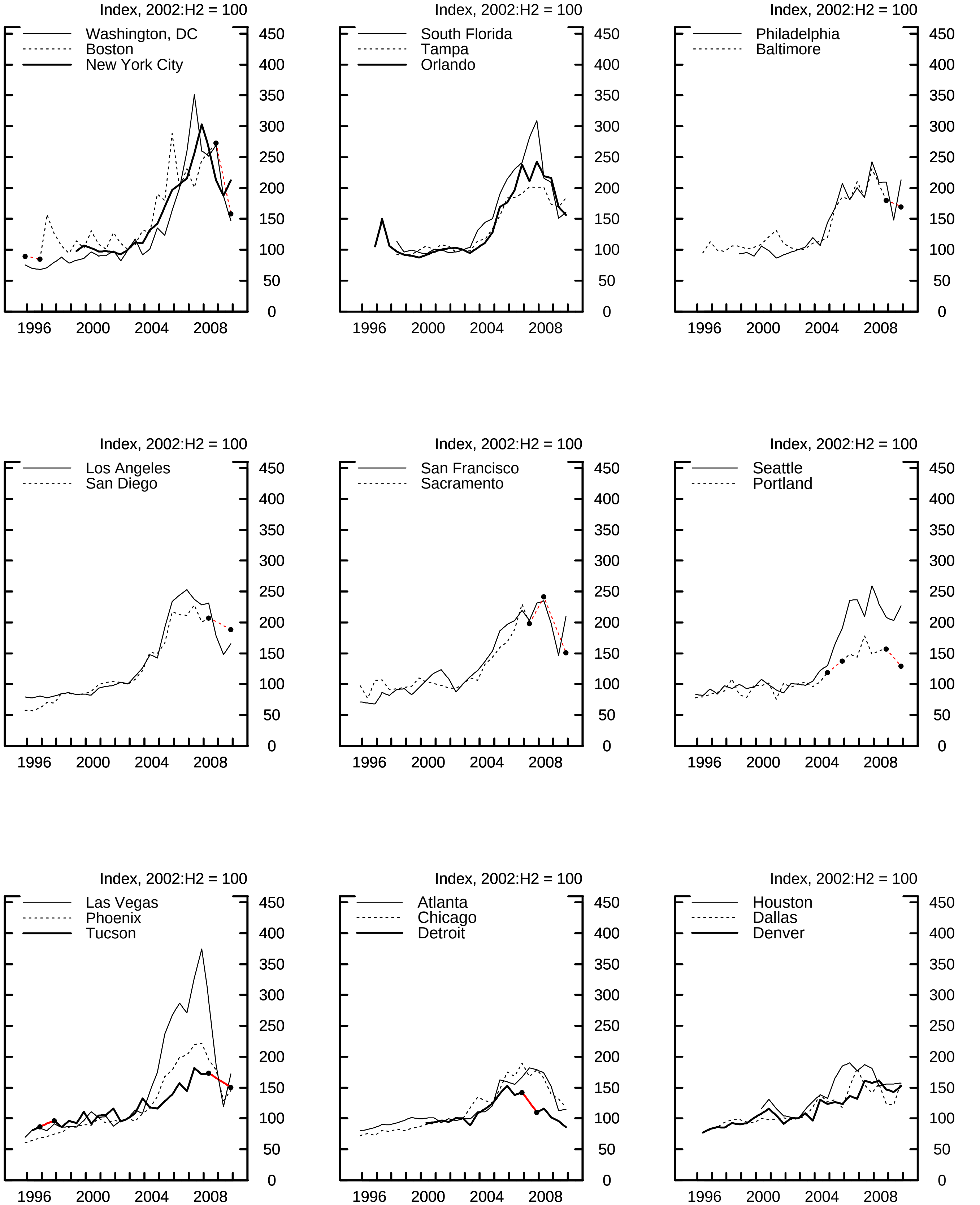
Source. Authors' analysis of data from the CoStar Group, Inc.

Figure 7
Aggregate Land Price Indexes by MSA



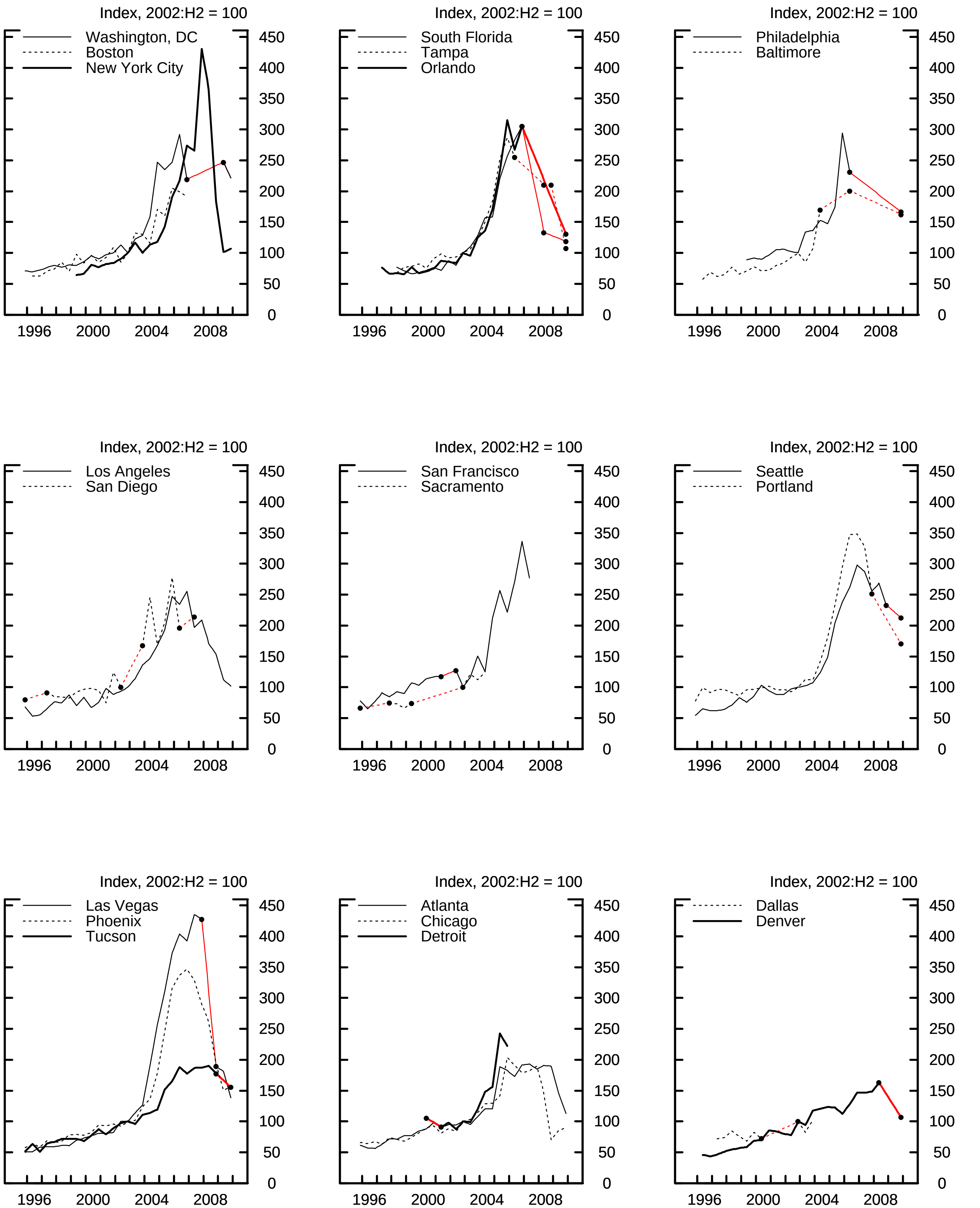
Note. Red segments between dots represent interpolation for missing half-yearly estimates.

Figure 8
Commercial/Industrial Land Price Indexes by MSA



Note. Red segments between dots represent interpolation for missing half-yearly estimates.

Figure 9
Residential Land Price Indexes by MSA



Note. Red sements between dots represent interpolation for missing half-yearly estimates. Houston not shown due to small number of half-yearly estimate

Appendix A

This appendix details the construction of the dataset used in the paper. We apply a sequence of filters to remove transactions that lacked sales prices or supporting documentation, were not at arms length, contained apparent data recording errors, fell too far outside the MSA-specific grid, or were from half-year periods with an insufficient number of observations for a given MSA. The COMPS dataset contains several detailed fields for notes on the property and transaction. We scan these fields for key phrases to identify observations that should be excluded from our analysis. Table A.1 below reports the initial sample size by year and the sequential effects of each filter.

The first screen removed transactions for which the sale price was missing or was less than $250,000. Although there is no minimum property value for inclusion in COMPs, transactions of less than $250,000 were uncommon until recent years, when CoStar significantly increased their coverage of these sales by collecting transaction details from public records. CoStar does not attempt to contact the participants to confirm the details of the transaction for these properties, in contrast to its standard procedure for other transactions. Accordingly, CoStar staff has recommended that we exclude these transactions from our analysis.[1]

Although the CoStar notes state whether a transaction has been classified as non-arms-length, we constructed our own, more comprehensive definition of these and other transactions that do not represent the sale of land at current market prices. We removed transfers of deeds in lieu of foreclosure, foreclosed properties seized by banks, and other

[1] However, we have some concerns abou t the blanket removal of all transactions of less than $250,000 because this procedure filters out low-value transactions for which the data are accurate, potentially inducing sample-selection bias. We are currently developing a less draconian screen for low-value transactions.

properties acquired by banks that lack any notes providing additional clarification. We also exclude any transaction in which the buyer and seller have either the same name or the same address, transfers within a company or family, section 1013 transfers, direct exchanges, transfers of partial interest, gifts, eminent domain acquisitions, the exercise of an existing option, and transactions that include a ground lease. Finally, we excluded any transactions that passed these screens but that CoStar classifies as a non-arms-length or non-market sale or as a sale in which the price does not represent the true value of the land.

We also screened the data for observations where the price per square foot was an outlier that appears to reflect data errors in two separate ways. First, transactions were excluded that had an unusually high or low price per square foot and for which the notes indicate that reported values could not be confirmed, were misleading, or represented considerations other than the land itself. Second, we excluded observations where the reported gross square footage of the site was dramatically different from the reported net square footage. The difference between the two represents the part of the site that is not buildable, but we were concerned that very large differences could indicate a data recording error.

We controlled for spatial variation within each MSA using Colwell's (1998) semi-parametric approach. This technique, described in section 3 of the paper, superimposed a grid over a map of each MSA, where the grid consists of equal-size squares or rectangles. As required by Colwell's method, observations that were more than one square or rectangle outside this grid – those on the periphery of the MSA – were excluded from the analysis.

Our final filter excluded observations from half-yearly periods that lacked sufficient observations to reliably estimate the value of the corresponding time dummy for that particular MSA. As described in section 2 of the paper, we set this threshold at 30 observations for residential land and commercial/industrial land separately when we estimate MSA-level price indexes for both property types and at 40 observations for the two property types taken together when we estimate an aggregate MSA-level price index. The 30-observation threshold applied to each property type is the more restrictive of the two tests, and the table reports the effects of this filter. As shown in column 6, this screen filtered out the most observations in 2007, 2008, and 2009, when transaction volume was well below the levels in earlier years.

A final issue concerns the time lag before a transaction appears in the COMPS database. Our analysis of the historical data indicates that 66 percent of sales are recorded in the database within three months of their sale date, 86 percent after six months, and 93 percent after one year. Based on these results, we judged that sufficient data were available to estimate preliminary price indexes for a given half-year period six months after the end of the period.

Table A.1
Construction of Sample by Year

Year	(1) Initial sample	(2) Price missing or < $250,000	(3) Non-arms length	(4) Outliers	(5) Too far outside grid	(6) Fails MSA obs. test	(7) Final sample
1995	8,216	2,078	189	0	449	141	5,359
1996	9,898	2,153	211	2	351	228	6,953
1997	11,577	1,984	332	0	496	145	8,620
1998	13,956	1,872	574	1	213	108	11,188
1999	16,326	2,286	765	7	155	65	13,048
2000	15,781	1,940	762	5	90	62	12,922
2001	13,767	1,544	735	7	91	160	11,230
2002	15,215	1,391	803	6	108	80	12,827
2003	16,395	1,233	752	8	95	48	14,259
2004	18,177	1,234	790	25	146	114	15,868
2005	16,095	538	734	29	246	130	14,418
2006	15,376	1,280	690	26	885	173	12,322
2007	13,855	3,287	811	18	1,066	333	8,340
2008	11,832	4,045	985	13	384	484	5,921
2009	13,403	7,372	1,477	7	530	460	3,557
Total	**209,869**	**34,237**	**10,610**	**154**	**5,305**	**2,731**	**156,832**

Source. Authors' analysis of data from the CoStar Group, Inc. (www.costar.com).

Appendix B

This appendix presents the results of a large set of likelihood ratio tests for the equality of coefficients in our hedonic regression equation. For each hypothesis, the test statistic is -2 $\ln(\lambda^R/\lambda^U)$, where λ^R is the value of the likelihood function after the equality restrictions have been imposed and λ^U is the value of the unrestricted likelihood function. This statistic is distributed $\chi^2(\rho)$, where ρ equals the number of restrictions imposed by the null hypothesis.

Each column in table B.1 reports the test results for a subset of independent variables. The column labeled "Condition of property," for example, includes the nine dummy variables that fall under this heading, for which the coefficient estimates were summarized in table 2. Similarly, the column labeled "Time effects" includes the full set of half-yearly dummy variables. For any column in the table, the first row shows the results for the most restrictive null hypothesis – that every variable in that column has a single coefficient across all 23 MSAs and both property types. The second row tests a less restrictive null hypothesis – that each variable has a single coefficient across all MSAs for residential land and a separate coefficient across all MSAs for commercial/industrial land. The next row tests the analogous null within MSAs – that each variable has a single coefficient across residential and commercial/industrial land in every MSA but that this coefficient differs across MSAs. The remaining rows test the hypothesis that each variable has the same coefficient across the two property types for MSAs taken one at a time.

Each entry in table B.1 presents the p-value for a particular hypothesis test. p-values smaller than 0.05 indicate that the null hypothesis is rejected at the five-percent

level of significance, while p-values smaller than 0.001 indicate rejection at the one-percent level. The p-values greater than 0.05 have been shaded to highlight the hypotheses that cannot be rejected at the standard five-percent level.

The results in the first three lines of the table provide strong evidence in favor of the highly disaggregated regression model that we estimate instead of a model that aggregates the individual MSAs, the two broad property types, or both. As shown, all of these aggregation hypotheses are rejected at the one-percent level of significance.[2]

The remaining rows of the table show that some of the coefficient restrictions cannot be rejected for individual MSAs. For example, in 13 of the 23 MSAs we cannot reject the hypothesis that distance from the CBD has the same price effect on residential and commercial/industrial land. Similarly, in nine MSAs we cannot reject the hypothesis that parcel size has the same price effect on both types of property. However, in every MSA at least some of the equality restrictions are rejected at the one-percent level. Accordingly, it is not appropriate to estimate an aggregated model for residential and commercial/industrial land prices in any of the MSAs.

[2] Note that we do not test the equality of the locational grids across MSAs. Given the differences in geographic features across MSAs, we would expect the price surfaces to differ as well.

Table B.1
P-values from Hypothesis Tests for Equivalence of Coefficient Estimates

	Parcel size	Condition of property	Intended use	Transaction character-istics	Distance from CBD	Locational grids	Time effects
Across MSAs and property type	<.001	<.001	<.001	<.001	<.001	NA	<.001
Across MSAs, by property type	<.001	<.001	<.001	<.001	<.001	NA	<.001
Across property type, by MSA							
All MSAs	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Atlanta	<.001	.229	<.001	<.001	<.001	<.001	<.001
Baltimore	<.001	.109	<.001	<.001	.595	.431	<.001
Boston	.164	.010	<.001	<.001	.020	.045	<.001
Chicago	.049	<.001	<.001	<.001	.818	<.001	<.001
Dallas	.296	<.001	<.001	<.001	.789	<.001	.557
Denver	.299	<.001	<.001	<.001	<.001	<.001	<.001
Detroit	.137	.417	<.001	<.001	.603	<.001	.087
Houston	.068	<.001	<.001	<.001	.083	<.001	.170
Las Vegas	<.001	<.001	<.001	<.001	.051	<.001	<.001
Los Angeles	<.001	<.001	<.001	<.001	.533	<.001	<.001
New York	<.001	.004	<.001	<.001	.101	.478	<.001
Orlando	.822	.032	<.001	<.001	<.001	<.001	<.001
Philadelphia	.678	<.001	<.001	<.001	.553	.045	<.001
Phoenix	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Portland	<.001	.004	<.001	<.001	.648	<.001	<.001
Sacramento	<.001	.116	<.001	<.001	.276	<.001	<.001
San Diego	.025	.072	<.001	<.001	.384	<.001	.999
San Francisco	.939	.051	<.001	<.001	.951	.059	.255
Seattle	<.001	<.001	<.001	<.001	.007	<.001	<.001
South Florida	<.001	<.001	<.001	<.001	<.001	<.001	<.001
Tampa	.026	.033	<.001	<.001	<.001	<.001	<.001
Tucson	.355	.028	<.001	<.001	<.001	<.001	<.001
Washington DC	<.001	<.001	<.001	<.001	<.001	<.001	<.001

Note. See table 1 for definitions of selected MSAs. The p-values are based on likelihood ratio tests. The likelihood ratio has a chi-squared distribution with degrees of freedom equal to the number of restrictions imposed for that test. NA indicates that we did not perform this hypothesis test. All p-values greater than 0.05 are shaded.
Source. Authors' analysis of data from the CoStar Group, Inc. (www.costar.com)