

FHFA STAFF WORKING PAPER SERIES

The Price of Residential Land for Counties, ZIP Codes, and Census Tracts in the United States

Morris A. Davis
William D. Larson
Stephen D. Oliner
Jessica Shui

February, 2020 (revised)
January, 2019 (original)

Working Paper 19-01

FEDERAL HOUSING FINANCE AGENCY
Division of Housing Mission & Goals
Office of Policy Analysis & Research
400 7th Street SW
Washington, DC 20219, USA



Working Papers prepared by staff of the Federal Housing Finance Agency (FHFA) are preliminary products circulated to stimulate discussion and critical comment. The analysis and conclusions are those of the authors alone, and should not be represented or interpreted as conveying an official FHFA position, policy, analysis, opinion, or endorsement. Any errors or omissions are the sole responsibility of the authors. References to FHFA Working Papers (other than acknowledgment) should be cleared with the authors to protect the tentative character of these papers.

Please address correspondence to William Larson (william.larson@fhfa.gov).

The Price of Residential Land for Counties, ZIP Codes, and Census Tracts in the United States

Morris A. Davis, William D. Larson, Stephen D. Oliner, Jessica Shui
FHFA Staff Working Paper 19-01
February, 2020 (revised)
January, 2019 (original)

Abstract

We use data on appraised land value from a data set of more than 14 million appraisals to produce annual estimates of the average price of land used in single-family housing and the average share of house value attributable to land. We generate results for a balanced panel of more than 1,000 counties, 9,000 ZIP codes, and 19,000 census tracts in the United States between 2012 and 2018, covering 87% of the population and 85% of all single-family homes; we also produce cross-section results using pooled data for 2012-2018 that have even broader geographic coverage. The cross-section results confirm predictions about land prices from canonical models of urban economics and provide a number of other stylized facts. The panel results show that land prices rose faster than house prices over 2012-2018 in large metro areas, boosting the land share of house value, while the land share fell in smaller metros. For the U.S. as a whole, the land share increased notably over 2012-2018, reflecting the outsized influence of the largest metro areas. The land prices and land shares we generate are intended to aid researchers in urban and regional economics and policy-makers monitoring the health of the nation's single-family housing market and are available for free download at <https://www.fhfa.gov/papers/wp1901.aspx>.¹

Keywords: land prices · land leverage · price gradient · standard urban model · price dynamics

JEL Classification: R14, R21, R32

¹The indices are works in progress and all data, tables, figures, and other results in this working paper are subject to change.

Morris A. Davis
Department of Finance and Economics
Rutgers Business School
Rutgers University
1 Washington Park Room 627
Newark, NJ 07102
mdavis@business.rutgers.edu

William D. Larson
Federal Housing Finance Agency
Office of Policy Analysis & Research
400 7th Street SW
Washington, DC 20219, USA
william.larson@fhfa.gov

Stephen D. Oliner
American Enterprise Institute for Public
Policy Research
1789 Massachusetts Avenue, NW
Washington, DC 20036
stephen.oliner@aei.org

Jessica Shui
Federal Housing Finance Agency
Office of Policy Analysis & Research
400 7th Street SW
Washington, DC 20219, USA
jessica.shui@fhfa.gov

1 Introduction

Researchers have taken to describing a single-family house as a physical structure occupying some land: See Bostic, Longhofer, and Redfearn (2007), Davis and Heathcote (2007) and Davis and Palumbo (2008), for example. Because housing structures are infrequently renovated and construction costs change relatively slowly from year to year, rapid change in the value of housing typically occurs when the underlying land is appreciating or depreciating. For this reason, the housing boom and bust of 1998-2012 has been described as a land boom and bust (Davis, Oliner, Pinto, and Bokka, 2017).

Although the importance of studying and monitoring the price of land in residential use is now well understood, few studies have produced data on land prices at a relatively fine level of geography. Broadly speaking, researchers have used one of two methods to estimate the price of residential land. Both of these methods require data that have, until recently, been hard to acquire. The first method uses data from sales of vacant or near-vacant land. Three examples of the first method are Haughwout, Orr, and Bedoll (2008), Nichols, Oliner, and Mulhall (2013) and Albouy, Ehrlich, and Shin (2018). These authors all use data from the Costar Group, Inc. Haughwout, Orr, and Bedoll (2008) estimate the price of land inside the New York metro area; Nichols, Oliner, and Mulhall (2013) produce price indexes for land for 23 metro areas; and Albouy, Ehrlich, and Shin (2018) estimate the average value of urban land in nearly all metropolitan areas in the United States.

The second method measures the price of land as the difference between house value and the replacement cost of the structure on the land. Davis and Palumbo (2008) apply this method to data from the American Housing Survey to generate the average price of land for 46 metro areas. Davis, Oliner, Pinto, and Bokka (2017) use proprietary data on house prices and construction costs from a number of sources to generate the level of land prices and changes in land prices at the ZIP code level for the Washington, DC metropolitan area.²

In this paper, we use a huge database of home appraisals to produce annual panel data for the price of land in single-family residential use for 1,132 counties, 9,194 ZIP codes, and 19,252 census tracts from 2012 through 2018.³ These estimates of land prices cover 87% of

²See Nichols, Oliner, and Mulhall (2013) and Davis, Oliner, Pinto, and Bokka (2017) for additional references in this literature.

³The January 2019 version of this working paper reported land prices for 964 counties, 8,344 ZIP codes, and 11,494 census tracts.

the U.S. population and 85% of all single-family homes.⁴ To our knowledge, ours is the first study to produce these estimates at a fine geography for nearly the entirety of the United States.⁵ Our source data are from the Uniform Residential Appraisal Report submissions to the Government Sponsored Enterprises (GSEs), Fannie Mae and Freddie Mac. The reports are required by the GSEs before they purchase or securitize a mortgage. These data contain more than 14.7 million unique appraisals submitted between 2012 and 2018.

Our raw estimates of land values in this data set are based on “cost-approach” appraisals; we set land value equal to the appraised value of the house less an estimate of depreciated replacement cost of the housing structure.⁶ A common concern about this residual method of estimating land values is that it assumes that the sum of the replacement cost of the housing structure and the value of the land (if it were vacant) is equal to the market value of housing. We show this assumption does not hold when a housing structure has become functionally obsolete and is due to be torn down or extensively remodeled. To address this issue, we calibrate a simple option model for tearing down and rebuilding a house. Simulations of the calibrated model suggest that the value of housing is well approximated as the sum of the replacement cost of the structure and the market value of the land if vacant for about the first 20 years of the life of the structure. As a conservative application of this result, we only use appraisals for relatively new homes with an effective age of no more than 15 years. This filter eliminates about 34% of the appraisal observations, preserving only the appraisals where we believe the implied land value is an unbiased estimate of the market value of the land if the land were vacant. We eliminate an additional 19% of our observations when the appraisals are nearly equal to a public tax-assessor value, due to concerns about the accuracy of assessed values as a measure of market value for reasons we discuss later. After applying these filters, our working sample contains 7.9 million appraisals.

We generate two sets of estimates of land value, “as-is” and “standardized.” Our as-is estimates report the value of land per-acre, without any adjustments or corrections. Our

⁴Additionally, by appropriately aggregating county-level values we create panel estimates for 597 Core-Based Statistical Areas (CBSAs), all 50 states and the District of Columbia, and the United States as a whole. Note that we report pooled cross-section estimates of land prices for many more localities, including 909 CBSAs, 2,292 counties, 19,012 ZIP codes, and 58,326 census tracts.

⁵Albouy, Ehrlich, and Shin (2018) produce estimates for nearly all Primary Metropolitan Statistical Areas (PMSAs) in the United States, but they do not report on or make available data for any finer level of geography. On average, they observe 212 direct land sales per PMSA.

⁶We exclude data on vacant land sales due to the difficulty of controlling properly for differences in the characteristics of the vacant land, such as whether water and sewer lines are in place.

standardized estimates report the price of land per quarter-acre, roughly the median sized lot in our data, after adjusting for the fact that the price of land per acre tends to fall as acreage increases, the so-called “plattage effect.” We generate the standardized estimates using a two-step procedure. First, we use the procedure of Davis, Oliner, Pinto, and Bokka (2017) to adjust for the effect of lot size on land prices and compute the price of land per quarter acre for each assessed property in our working sample. Then, we use a procedure called Kriging, as described by Basu and Thibodeau (1998), to interpolate this standardized price-per-quarter-acre of land to lots under all remaining single-family housing units in a given geography (county, ZIP code, or census tract).⁷ Using a 20% hold-out sample, we show in our data that Kriging offers a lower root mean square error in interpolating land prices than some other commonly used methods of spatial interpolation. In an Appendix, we analytically derive a land-price gradient from a simple, calibrated urban model and show that when we simulate data from that model, Kriging delivers the correct land-price gradient from model-simulated data. The procedure we use to compute the as-is estimates is exactly the same as with the standardized estimates, except we add one step at the end to undo the correction for plattage effects.

The primary goal of the paper is to generate land values and indices covering the United States at a fine geography for use by researchers and policy-makers. To that end, the aggregated land-price data we generate in this paper are available for download at the web site of the Federal Housing Finance Agency, at <https://www.fhfa.gov/papers/wp1901.aspx>. For each county, ZIP code, and census tract where we compute land prices, and for broader geographies built up from the county data, we post a variety of useful statistics, including land shares of total property value and both as-is and standardized land prices.

In addition, we use the data to develop and confirm important stylized facts about land prices in the United States. First, as shown by Albouy, Ehrlich, and Shin (2018) and others, the level of land prices at the center of a metro area varies greatly. For example, the price of land at the center of metro areas with more than 2 million single-family housing units is more than 25 times greater than the price of land at the center of metro areas with less than 500 thousand housing units. Second, the rate at which land prices decline from the city center also varies across metro areas. Finally, the price of land covaries with certain

⁷We obtain our universe of single-family housing units in a given geography from the assessor data licensed from Corelogic. These data contain nearly all parcels for all land use types in the counties for which CoreLogic has acquired data.

variables in accordance with predictions of classic models of urban economics. In particular, measured at the average price per acre in a ZIP code, residential land prices are negatively correlated with lot size and are positively correlated with the size of the housing structure on a lot. ZIP codes with higher residential land prices also tend to have a smaller share of land area with vacant structures or land, a smaller share of land area devoted to agriculture, and a greater share devoted to commercial and industrial structures.

2 Option Model of Housing Teardowns

In this section, we build a simple model for when a land owner should optimally tear down his or her house and rebuild. We use the model to develop a rule-of-thumb to determine the oldest existing homes for which we can derive unbiased estimates of land value from a cost-approach value decomposition.⁸

In the model, the land owner owns property with a building of size S on a lot of size L . The lot size is fixed in perpetuity, but the building size can be changed. This property earns rents of

$$q^H S^{1-\phi} L^\phi \tag{1}$$

where q^H is the rental price per unit of housing service provided and the Cobb-Douglas aggregate of structures and land, $S^{1-\phi} L^\phi$, is the number of units of housing services.

Each period, the land owner must decide whether or not to demolish the building and rebuild on vacant land, or to let the building depreciate some and revisit the choice next period. Since L is fixed in perpetuity, we can summarize the decision problem of the land owner as one over a choice of S only. Denote $V(S)$ as the value of owning a property with a building of size S and similarly let $V(0)$ denote the value of the property as vacant land.

When the land is not vacant and a structure of size S sits on the property, the land owner chooses either to let the property sit as is and collect rents, or to knock the property down

⁸Although we are the first to develop such a rule-of-thumb, many other papers have studied the option value of development, for example Titman (1985), Clapp, Eichholtz, and Lindenthal (2013) and McMillen and O’Sullivan (2013).

and make the land vacant. This problem has the expression

$$V(S) = \max \{ q^H S^{1-\phi} L^\phi + \beta V(S(1-\delta)), V(0) \} \quad (2)$$

β is the factor by which the land owner discounts the future and δ is the rate at which the housing structure depreciates. The first term in the max operator is the value of the property with the structure left intact. This term includes the discounted value of owning a property with a structure of size $S(1-\delta)$ next period. The second term is the value of the property when it is made vacant, assuming there are no demolition costs that must be paid to clear the land of the structure.

Denote p^S as the price per unit of newly-built structure. When the land is vacant, the land owner chooses to build the optimally-sized structure to maximize the value of the land. The choice determines the rents earned this period, plus the discounted value of the property in the future after accounting for depreciation of the structure, less the cost of building the structure. This choice satisfies:

$$V(0) = \max_{\mathcal{S}} \{ q^H \mathcal{S}^{1-\phi} L^\phi - p^S \mathcal{S} + \beta V(\mathcal{S}(1-\delta)) \} \quad (3)$$

The solution to this model can be characterized by two variables: \bar{S} , the size of the structure that is built when the land is vacant,⁹ and \underline{S} , the smallest structure that exists (i.e. a structure of any smaller size is demolished). Using the relationship $\underline{S} = \bar{S}(1-\delta)^T$, where T represents the maximum age of any housing structure, $T = (\log \underline{S} - \log \bar{S}) [\log(1-\delta)]^{-1}$.

We solve this model and calibrate it as follows. We set the discount rate to $\beta = 0.90$; the annual depreciation rate to $\delta = 0.023$ (Harding, Rosenthal, and Sirmans, 2007); land's share of value to $\phi = 0.30$; the price per unit of structure, a normalization, to $p^S = 1$; and then we find the level of rent q such that the value of housing for a newly built optimally-sized structure is $V(\bar{S}) = 100$. This calibration delivers simulation results that the oldest house is 80 years old; the smallest (most depreciated) house size at the time of demolition $\underline{S} = 10.8$ with the value at that house size equal to the value of the underlying land $V(0) = V(\underline{S}) = 30$; and the optimal house built on vacant land is $\bar{S} = 70$.

⁹This is the argmax of equation (3).

Figure 1 shows how housing value (blue line) and the replacement cost of structures (red line) change with the building age in this model. The replacement cost of structures declines at a constant rate from 70 when newly built to 10.8 after 80 years, at which point the structure is torn down. The value of the vacant land is always 30 (green line). The value of housing declines gradually over time from 100 when newly built to 30 after 80 years. When the structure is torn down, the value of housing (30) is less than the sum of the value of land and the replacement cost of the structure ($40.8 = 30 + 10.8$).

So, according to this calibrated model, at what point is the house value no longer well-approximated as the sum of the value of the vacant land and the cost of the depreciated housing structure? To answer this question, in Figure 2 we compute a “land share” of house value two different ways, and determine the age of the housing structure at which these two methods stop producing similar results. The first (correct) method, the blue line, computes land’s share of value as the ratio of the value of vacant land to the value of housing, $V(0) / V(S)$. This method shows that land’s share of value increases monotonically from 30% for newly built homes to 100% for homes about to be torn down. The second method, the red line, computes land value residually as house value less the replacement cost of structures, $V(S) - S$. This is meant to approximate how land is measured residually when given an appraised value of housing, $V(S)$, and an estimate of the depreciated reconstruction cost of the housing structure, S . Land’s share of housing is this residually-measured land value divided by house value. Figure 2 shows that this measure of land’s share of housing ranges from 30% for newly constructed homes to only about 60% for homes about to be torn down. The value of housing at the point of teardown is 30, entirely equal to land value. A residually-measured estimate of land value at the point of teardown would be biased down and only equal to $30 - 10.8 = 19.2$. Figure 2 shows that the two methods produce nearly identical estimates of land share of value for structures that are younger than 20 years old. This result guides restrictions to our sample of data that we describe later.

Of course, different models will produce different results. The point of this section is to not write down the most realistic model of land ownership and teardowns. Rather, it is to gain intuition about the value of the option to tear down a house and how that option affects residually measured estimates of land’s share of house value. Our general results should be robust to any model where an optimal teardown occurs decades after a house is built. The reason is as follows: When a house is relatively newly built, the expected date of a teardown

is so far away in the future that the option of tearing down the house has little value. Since this option has little value and depreciation is low, the value of housing is well approximated as the sum of the replacement cost of the structure and the value of the vacant land.

3 Data

In each mortgage appraisal, there are typically three separate approaches to estimating the value of the underlying property. The first is the sales comparison (or “comps”) approach, by which an appraised value is generated based on recent comparable transaction prices. A second “income” approach estimates the value of the property as the discounted flow of imputed rental income. Finally, the “cost” approach attempts to separately estimate the cost of the components of the property, the land and the structure, and assumes the estimated value of the property is the sum. We use cost-approach appraisals in our analysis.

The data on cost-approach appraisals are from Uniform Residential Appraisal Report submissions collected by the GSEs. After data cleaning, including the removal of duplicate and/or resubmitted appraisals, we have approximately 14.7 million unique cost-approach appraisal records submitted between 2012 and 2018.¹⁰ Among the appraisal records, 52% specify a single source for the estimates of replacement costs, while 15% do not specify any source and the remaining 32% are associated with at least two sources. Marshall & Swift (57%) and “local information” (31%) are the two most commonly appearing sources, followed by R.S. Means (13%) and “internet” (10%), with others (e.g. tax records and new construction information) making up about 1%.

We wish to include in our sample only those cost-approach appraisals where the land component is an unbiased estimate of the market value of land. This likely occurs when two conditions hold: First, the option value of redevelopment is low and second, appraisals are not anchored to biased tax assessments (for reasons we discuss later).

To address any biases arising from the option value of redevelopment, we use the *effective*

¹⁰In our pre-cleaning of the data to arrive at this count, we exclude appraisal records with 1) lots smaller than 500 square feet or larger than 2 acres; 2) missing property value or property value less than \$10,000; 3) cost-approach-estimated site value missing or less than \$200; 4) land-price-per-acre smaller than \$200; 5) site value greater than cost-approach-estimated property value; 6) missing depreciation information or depreciation at least three times greater than the contract price or the appraised value; 7) land share of property value less than 1% or greater than 99%; 8) structure-land area ratio equal less than .01 or greater than 4; 9) construction date before 1850 or after 2018.

age variable in the appraisal dataset to eliminate observations from our working sample. Effective age is computed by appraisers as follows:

$$\text{Effective Age} = \text{Economic Life} \times \left(\frac{\text{New Replacement Cost} - \text{Depreciated Structure Value}}{\text{New Replacement Cost}} \right) \quad (4)$$

For example, for a structure with an assumed economic life of 80 years, a depreciated structure value of \$100,000, and a replacement cost of the structure as new of \$150,000, the effective age would be $80 \times (50/150) = 26.7$ years.¹¹

The results from our calibrated model of section 2 suggest that the residual method produces accurate estimates of the value of land for structures less than 20 years old. Given an assumed constant depreciation rate of 2.3% per year and a maximum economic life of a given house of 80 years, the calibrated model implies the maximum effective age from equation (4) for reliable cost-based appraisals of land should be approximately

$$80 \times [1 - (1 - 0.023)^{20}] = 29.8 \text{ years} \quad (5)$$

In our analysis, we conservatively restrict our data to appraisals reporting an effective age to half of this value, 15 years or less, corresponding to an age of 8.9 years in the calibrated model. This filter eliminates 34% of the working sample.

With respect to a different potential bias in the appraisal data set, for a variety of reasons appraisers sometimes anchor their reported appraisal to existing estimates of value such as the contract price or tax assessments. The anchoring of appraisals to contract prices is not problematic for our purposes as contract prices should reflect market value in most if not all cases. However, if an appraiser anchors to a tax assessment this may impart a significant downward bias to estimates of the value of land. Lutz, Molloy, and Shan (2011) and others show that tax assessments are often biased relative to transaction values for two reasons. First, the rate of change of tax assessments is sometimes capped to prevent these assessments from rising too quickly in rapidly-appreciating housing markets. Second, assessments can be right-censored near the market value: assessments less than or equal to the market value are unchallenged by the property owner, but assessments significantly greater than the market value are likely can be challenged by the property owner and adjusted down to the market

¹¹The U.S. Bureau of Economic Analysis uses a service life of 80 years for new 1-4 unit residential structures; see Katz and Herman (1997).

value.

To illustrate the size and nature of potential biases, the top panel of Figure 3 shows a histogram of the ratio of appraised value to tax-assessed value for all housing units for which we have an appraisal and a tax assessment.¹² The top panel shows the ratio for total property value, inclusive of land and structures, while the bottom panel shows the ratio for the land component alone. Both panels clearly show that appraised values tend to be higher than tax assessments. In addition, the spike at the value of 1.0 in the bottom panel shows that a sizeable mass of appraisers anchor their appraisals of land value to the tax-assessed value of land. To remove this potential source of bias from our results, we exclude from our working sample any property within 2% of the assessed value for either the land component or the total value.¹³ This filter eliminates an additional 19% percent of the working sample, yielding a final sample of 7.9 million valid cost-approach appraisals from which we estimate land values.

4 Computing Land Values and Land Shares

4.1. Overview

Our working sample, while quite large, covers a minority of all single-family properties. In this section, we describe how we estimate land values for *all* single-family properties.¹⁴ We produce two estimates of land value for every parcel. In the first set of estimates, which we call our standardized estimates, we compute the value of land as if every parcel was exactly 0.25 acres. Standardizing lot size corrects for the fact that the price-per-acre of land tends to decrease with acreage, all else equal, a relationship known as the plattage effect. In the second set of estimates, we estimate the value of land for the parcel without correcting for plattage effects. These are our as-is estimates, which we report on a per-acre basis. The public-use data files posted on the FHFA website provide both of these measures of land

¹²Not all properties in the appraisal database have a tax assessment.

¹³For most counties, we have tax assessment data for 2012, 2013, 2017 and 2018; for some counties we have fewer years. In all cases, we assign values for missing years using a straight-line interpolation for both house value and land value. If either imputed value lies within 5% of the value recorded in the appraisal data, the appraisal is dropped from our working sample.

¹⁴We have been asked why we interpolate the data in our working sample to produce estimates of land value for all parcels when we only report averages. The intuition behind our procedure is that the relatively young homes in our working sample can be geographically clustered, and by interpolating land values to all parcels (and then computing averages) we undo any effects of this clustering. Later on, we show using a hold-out-sample analysis that our estimates of land value are more accurate than simple averaging of the data in the working sample, providing some validation of the intuition.

values. For each measure, we report the average value across all single-family parcels by county, ZIP code, and census tract, along with aggregations of the county-level data to CBSAs, states, and the U.S. as a whole.¹⁵

4.2. Standardized Estimates of Land Value

In this section, we estimate the value of land for every parcel in a given geography as if the parcel was exactly one-quarter acre. We start by converting the value of land for parcels in our working sample to values as if all parcels were one-quarter acre. This is more complicated than (say) taking the value of land on a one acre parcel and dividing by four. Instead, we use a regression procedure similar to that described by Davis, Oliner, Pinto, and Bokka (2017). For each county, we pool the data in all years. Then, county by county, we regress the log of the value of land on the log of lot size, including ZIP-code fixed effects and year dummies as regression controls. Denote the county-specific coefficient on the log of lot size arising from this regression as β . For each parcel in our working sample, we compute the predicted log price per quarter acre as the observed price plus our (county-specific) estimate of β times the difference of the log of a quarter-acre and the log of the actual size of the lot.¹⁶

Next, we merge our working sample of data with the 2017 vintage of single-family parcels from assessor data licensed from Corelogic. These data contain the near-census of parcels in the counties for which it has acquired rights to the data.¹⁷ After this merge, we assign a standardized, quarter-acre land value to each single-family property in the assessor data that is not included in our working sample using an interpolation procedure called Kriging.

The Kriging procedure is commonly used in the hard sciences. For our purposes, we need a method that can be used in urban areas with steep and varying gradients over short distances, and in rural areas with relatively flat gradients and geographically sparse transactions. We also need a method that is computationally manageable to loop over thousands of areas that

¹⁵When ZIP codes span multiple counties, the reported value is the average of values in each represented county, weighted by the share of single-family housing stock.

¹⁶The county-level estimates of β are reported in the dataset for the paper posted on the FHFA website. The dataset also includes the estimated parameters from similar regressions used in the estimation of land shares, as discussed in section 4.5.

¹⁷We maintain this fixed sample of properties in order to avoid composition bias, effectively turning the sequence of average land values we report in our panel data into a Laspeyres measure.

include millions of parcels. Kriging satisfies these requirements.¹⁸ To our knowledge, there have been no studies in the land-price literature that have evaluated the relative accuracy of different spatial imputation methods. Later on, we compare the accuracy of Kriging to a number of alternative interpolation procedures in a 20% holdout sample. In this sample, Kriging produces more accurate estimates.

In Appendix A, we discuss the Kriging procedure in detail. Here we provide a brief summary. Before we do so, the key takeaway is that Kriging, like other estimators, uses a weighted-average of n nearest neighbors to generate predicted land prices. What makes Kriging different is its algorithm to generate the weights.

Derivation of those weights proceeds in five steps. The first step involves calculating pairwise differences in values between each pair in the sample within a certain distance range, approximately 6.9 miles (0.1 degrees in coordinate distance) in our case.¹⁹ The next step establishes 15 bins of distances and computes the average “semivariance,” defined as half of the squared difference in land values, of all the points in each distance bin. The third step fits a 3-parameter curve that preserves monotonicity to this set of 15 binned averages. This is referred to as a “variogram;” we estimate one variogram per county per year. The fourth step applies this estimated curve to estimate covariances between values in an unsampled location and a number of nearby sampled locations – we choose 20 nearby neighbors. The fifth step uses these fitted covariances to construct the weights on the nearby sampled locations.

Figure 4 shows a heat map of land value for the city of Washington, DC and provides a look at how Kriging interpolates land value. Panel (a) shows the standardized land values in our working sample and panel (b) shows how the Kriging interpolation algorithm interpolates these values to the entire geography of Washington, DC, including areas with few observations. As can be seen from the legend of panel (b), interpolated standardized land values vary widely, from less than \$150,000 per quarter-acre to more than \$750 thousand per

¹⁸Basu and Thibodeau (1998) conduct an analysis of spatial autocorrelation in housing prices by comparing predictions from hedonic models to models with spatially autocorrelated errors. They find that traditional hedonic models are more accurate when unexplained price variation is spatially uncorrelated; otherwise, Kriging is more accurate. Our method of standardizing-then-Kriging land prices is analogous to their approach. This technique is also referred to as “Regression Kriging.”

¹⁹Our data are partitioned by county (and by year in the panel). This implies the Kriging procedure only considers pairwise points where both points are in the same county. The 6.9 mile cutoff will not bind in any county where the maximum distance between two locations in that county is less than 6.9 miles.

quarter-acre. Panel (c) shows the average value of the standardized land prices by ZIP code and panel (d) shows the average value for as-is land prices (we describe how we compute as-is land prices later).²⁰ In both cases, the mean is computed using the estimated land prices for properties in the working sample and the interpolated land prices for all other properties. Interestingly, the as-is land prices in panel (d) are uniformly higher than the standardized prices in panel (c). This occurs because lot sizes in Washington, DC are generally much smaller than a quarter-acre, so standardizing to the much larger quarter-acre lot size reduces value.²¹ Maps like this help to illustrate the value that households place on location-specific attributes such as school quality, access to transportation and other natural and man-made amenities.

4.3. Accuracy of the Kriging Procedure

In this section of the paper, we evaluate the accuracy of the Kriging procedure to interpolate land values to parcels that are not in the working sample. For this analysis, we omit a randomly selected 20% of our working sample and determine the root mean-square error (RMSE) of the Kriging interpolation procedure for that omitted 20%. In every year, we compute the RMSE for each county in our sample. Across counties, the median RMSE is about 40 percent. This might seem high, but it is in fact well within our priors. To understand why, denote true land, structure and house value for a given parcel as p_L^*L , p_S^*S and p_H^*H such that

$$p_L^*L = p_H^*H - p_S^*S \quad (6)$$

where p_L^* , p_S^* and p_H^* are the price per unit of land, structures and housing, respectively, and L , S and H are the quantities. Suppose that the value of structures is not measured with error but housing is measured with multiplicative error e such that observed house value p_H^oH is equal to $p_H^*H(1 + e)$. We can then write an expression for the percentage deviation of observed land value $p_L^oL = p_H^oH - p_S^*S$ from the truth as

$$\frac{p_L^oL - p_L^*L}{p_L^*L} = \left(\frac{p_H^*H}{p_L^*L} \right) e \quad (7)$$

²⁰Panel (b) shows that it is possible to estimate a land price for every location in the city but panels (c) and (d) illustrate that we do not report land values for some ZIP codes in Washington, DC. We do not report land prices for a ZIP code or tract if the number of direct land-price observations in our working sample for that geography is < 10 observations.

²¹Restated, due to plattage effects the value of a single quarter-acre lot is much less than the value of, say, four lots of 1/16 acre each.

Equation (7) says that the percentage measurement error in land values is equal to the inverse of land's share of house value times the percentage measurement error in house values. Using round numbers, if the standard deviation of measurement error in house prices is about 10% and if land's share of housing is about 25%, then the standard deviation of measurement error to land prices should be roughly 40%.²²

We compare the median RMSE from Kriging to three other commonly-used spatial interpolation procedures: Null, Nearest Neighbor, and Inverse-Distance Weights. Null sets the interpolated value of the target parcel equal to the unconditional average value in the relevant area (county, ZIP, or tract, respectively); Nearest Neighbor sets the interpolated value equal to the average value of the 20 most proximate observations; and Inverse-Distance Weights computes the weighted-average value of the 20 most proximate observations with the weights inversely proportional to the squared distance between the neighbor and the target parcel. Table 1 reports median RMSEs across counties for Kriging and the other spatial interpolation methods we consider for our 20% hold-out sample. As the table shows, measured this way, Kriging provides the greatest interpolation accuracy in every sample year. A comparison of the RMSE of Kriging and Null illustrates why we interpolate land values to all parcels and then compute the average, as compared to simply computing the average of the land values we directly observe. The simple average of the land values in the working sample – the Null estimate – is not as accurate at predicting the value of land of the held-out parcels as Kriging. Therefore, our Kriging procedure should also produce more accurate estimates of the average value of land than the Null estimator.

We also evaluate the accuracy of the Kriging procedure for our application by seeing if it can recover land values that endogenously arise from a simple rendition of the standard monocentric city model that we can compute analytically. We simulate two data sets from this model, one in which land values are measured perfectly and one in which land values are measured with error and then check the extent to which Kriging can replicate the analytic gradient of land prices. In the data set in which land values are perfectly measured, Kriging nearly exactly replicates the analytic gradient. In the data set in which land values are measured with error, Kriging produces relatively small average errors. For more details, see appendix B.

²²Note that Case and Shiller (1989) estimate the residual standard deviation to be about 15% for housing in repeat-sales of individual properties.

4.4. As-Is Estimates of Land Value

In some applications it is inappropriate to control for plattage effects. For example, standard urban models with optimizing households, i.e. Alonso (1964), Mills (1967), Muth (1969), and Brueckner (1987), predict that lot sizes will be larger in areas where land is plentiful and smaller where land is scarce. Such differences in optimal lot sizes contribute to observed differences in actual land prices per-acre across a metro area. Testing the predictions of these models using land values standardized to a quarter-acre lot may be inappropriate. Additionally, many accounting exercises require actual, as opposed to standardized, land values including tabulations used for national accounting or local tax assessments. For these reasons, we compute land prices per acre without adjusting for plattage effects, i.e. our as-is estimates of land value. To do this, we simply undo the correction for plattage effects for all parcels.²³ In our data files, we report as-is estimates of land value on a per-acre basis alongside the average lot size of single-family homes in the assessor dataset. These can be multiplied to yield estimates of the average value of land for single-family residential lots in each area.

4.5. Land Shares

While the price of land is useful in itself, the share of house value attributable to land is also of interest to researchers and policy-makers. For example, as has been remarked by Davis and Heathcote (2007), Davis, Oliner, Pinto, and Bokka (2017) and others, rapid changes in the share of housing attributable to the value of land are indicative of positive shocks to the demand for housing. The reason is that land is inelastically supplied, and in many locations it takes considerable time for builders to increase the number of housing units.

For each housing unit in an area, we estimate the value of land and the value of housing. The land share we report is the average value of land across housing units divided by the average value of housing across those same housing units. For each single-family housing unit in the assessor data, we use the as-is estimates of the actual value of land for each housing unit. Our procedure to compute the value of housing for all the single-family units in the assessor data is more complicated and the remainder of this section provides details. But the procedure resembles the one used to estimate land values: estimate the standardized home value for homes with admissible appraisals, use Kriging to interpolate values to all homes in

²³Referring to the text in the first paragraph of section 4.2., for each parcel we set the as-is estimate equal to the standardized estimate plus β times the difference of the log of the size of the lot (in acres) and the log of a quarter-acre.

the area, then unstandardize to create as-is values for all homes.

To start, for *all* of the housing units in the appraisal data and not just the relatively new units in our working sample, indexed by i , we regress the log of appraised house value (Y_i) on the log of effective age (X_i^A), the log of lot size (X_i^L) and the log of structure size (X_i^S) with intercept β_0 and coefficients β_A , β_L and β_S and error term v_i .²⁴

$$Y_i = \beta_0 + X_i^A \beta_A + X_i^L \beta_L + X_i^S \beta_S + v_i \quad (8)$$

We estimate this model for each county. Given the coefficient estimates in Equation 8, we compute the expected value for each house in the appraisal data if the effective age was (counterfactually) 15 years, the lot size was 0.25 acres and the structure size was 2,000 square feet. Call this standardized house value \tilde{Y}_i , computed as

$$\tilde{Y}_i = Y_i + (15 \text{ years} - X_i^A) \beta_A + (0.25 \text{ acres} - X_i^L) \beta_L + (2,000 \text{ sq ft} - X_i^S) \beta_S$$

We use Kriging to interpolate the standardized house value \tilde{Y}_i to all parcels in the assessor data.

In the last step, we use information about the lot size and structure size in the assessor data to convert the estimate of standardized house value to an un-standardized estimate of house value, one that correctly takes on-board the age and size of the structure. This step is complicated for two reasons: (1) Effective age is not reported in the assessor data and (2) the lot size and structure size reported in the assessor data does not necessarily equal the lot and structure sizes reported in the appraisal data.

Starting with the second issue, we run regressions of X_i^L and X_i^S in the appraisal data set on the same variables in the assessor data, call them χ_i^L and χ_i^S , at the county level

$$\begin{aligned} X_i^L &= a^L + \chi_i^L \gamma^L + u_i^L \\ X_i^S &= a^S + \chi_i^S \gamma^S + u_i^S \end{aligned}$$

where the u terms are the errors in the regression. We then compute for each value in the

²⁴We also include ZIP-code and year fixed effects.

assessor data the predicted values of lot size and square footage

$$\begin{aligned}\tilde{\chi}_i^L &= a^L + \chi_i^L \gamma^L \\ \tilde{\chi}_i^S &= a^S + \chi_i^S \gamma^S\end{aligned}$$

Next, using the appraisal data we run county-specific regressions of effective age on square footage, lot size, and ZIP-code fixed effects of the form

$$X_i^A = \delta_0 + X_i^L \delta_L + X_i^S \delta_S + \nu_i \quad (9)$$

We predict effective age in the assessor data, $\tilde{\chi}_i^A$, using the coefficients from Equation 9 regression and the corrected values of lot size and square footage

$$\tilde{\chi}_i^A = \delta_0 + \tilde{\chi}_i^L \delta_L + \tilde{\chi}_i^S \delta_S$$

Finally, using parameters estimated in Equation 8, we compute house value for each parcel in the assessor data that is not in the appraisal data as

$$\tilde{Y}_i + (\tilde{\chi}_i^A - 15 \text{ years}) \beta_A + (\tilde{\chi}_i^L - 0.25 \text{ acres}) \beta_L + (\tilde{\chi}_i^S - 2,000 \text{ sq ft}) \beta_S$$

5 Results

We only report results if our working sample includes at least 50 observations within a county in the relevant time period. When we report results for ZIP codes and census tracts, we require at least 50 observations within the county and at least 10 observations within the appropriate geography, also in the relevant time period. For this reason, we generate two data sets. In the first, our “pooled cross-section,” we create our county-level working sample by combining all data from 2012 through 2018 in the appraisal database. This provides the maximum number of admissible appraisals. In the regression to produce standardized land prices, we include a year fixed effect alongside the plattage adjustment and center the estimates using 2013 prices. As reported in Table 2, our pooled cross-section includes data for 2,292 counties, 19,012 ZIP codes, and 58,327 census tracts. In the second dataset, we do not pool data by year and simply report annual estimates for a balanced panel. Given our minimum data requirements, we report land prices for 1,132 counties, 9,194 ZIP codes, and 19,252 census tracts each year from 2012 through 2018. The annual and pooled sam-

ples cover 87% and 98% of the U.S. population residing in the 50 states plus the District of Columbia, respectively, and 85% and 97% of the single-family housing units. After constructing county-level statistics, CBSA, state, and national statistics are calculated by aggregating over counties for which land values are available, weighted by the single-family housing stock. We report state and national data for all 50 states plus the District of Columbia in both the pooled and panel datasets. We are also able to calculate values for 909 CBSAs for the pooled sample and 597 in the balanced panel.

Table 3 shows some basic statistics from our as-is estimates of land value and land shares. The top panel reports estimates from the pooled cross-section of counties and the bottom panel reports estimates from the annual panel data of counties (but pooled). The data show an enormous range in the average price per acre of land in single-family, residential use from less than \$10 thousand per county at the 1st percentile of the pooled cross-section data to nearly \$1.2 million at the 99th percentile. The land shares from this data set also show huge variation, from 7% at the 1st percentile to more than 50% at the 99th percentile. The panel data display similar variation, but the price of land and land shares are uniformly higher, as this data set drops many rural counties with cheap land and low land shares. Returning to the pooled cross-sectional data, the average price of land per acre by county is \$135 thousand and the median price per acre is \$50 thousand, indicating a significant right-skewness. The standard deviation of the average price per-land by county is \$1.156 million, 8.5 times the average value.

In the remainder of this section, we present other stylized facts related to the land-price and land-share data. For expositional purposes, we present pooled estimates first in order to validate our data in terms of known cross-sectional relationships between land prices and other variables. We then proceed to the annual panel, where we present several new findings. Overall, there are five main categories of stylized facts that we present: 1) land-price gradients and levels; 2) parallel information for land shares; 3) spatial variation in housing-structure density; 4) barriers to building housing and urban decline; and 5) changes over time to land shares. For most of the stylized facts we discuss we use as-is land prices, as these reflect the variation in lot sizes due, at least in part, to optimizing decisions by builders and households.

5.1. Land Prices: Gradients and Levels

The traditional monocentric city model predicts land prices fall with distance to the Central Business District (CBD) because households are willing to pay less per unit of housing as commuting costs rise. Since the marginal cost of an additional unit of structures is roughly constant within the city, the solution to the zero-profit condition for housing producers requires variation in the price of land. Therefore, the negative house price gradient translates to a negative gradient for land prices.

To illustrate the relation between land prices and proximity to the CBD, Figure 5 shows the land price per acre (pooled sample) for ZIP codes within 25 miles of the CBDs of two Core Based Statistical Areas (CBSAs), Washington-Arlington-Alexandria, DC-VA-MD-WV and San Francisco-Oakland-Hayward, CA. The same land price data are presented as maps (top panel) and as plots of prices as a function of the radial distance to the center of the CBD (bottom panel). These figures show a clear, downward sloping land-price gradient. In the very central ZIP codes there are not enough single-family housing transactions in our working sample to reliably construct land prices. Apart from these excluded areas, land prices average about \$15 million per acre in the most expensive ZIP codes in the Washington, DC metro area and about \$20 million per acre in the San Francisco metro area. By about 10 miles from the CBD, the gradients are flat on average but with significant variation depending on the direction away from the CBD.

The relationship between land prices and proximity to the CBD illustrated in Figure 5 is not unique to these two cities. Panel (a) of Figure 6 shows land prices as a function of distance to the CBD of a metro area for all ZIP codes with centroids within 25 miles of a central ZIP-code centroid. The panel shows a downward sloping relationship, but with a large amount of variation. Panel (b) of Figure 6 shows the same data as in panel (a), but separates the cities into three size groups. This picture more clearly makes that point that the level and rate of change of land prices with respect to distance from the CBD depends on city population. Indeed, in the largest cities, land prices are about 3 times higher than in the middle-size group and about 25 times higher than for the smallest group of cities.

The monocentric city model also predicts that land-supply restrictions have a negative effect on land prices holding population and amenities constant. This occurs because land supply restrictions have a negative effect on the spatial efficiency of the city which becomes capi-

talized into land prices. But, as discussed by Saks (2008), Saiz (2010) and Davidoff (2016), such restrictions also create amenities that may exceed the efficiency loss, with the net effect potentially being a reduced-form positive correlation between land restrictions on the one hand, and house and land prices on the other. Panel (c) of Figure 6 shows land-price gradients by population and the top and bottom halves of regulatory burden as measured by the Wharton Residential Land Use Regulation Index (Gyourko, Saiz, and Summers, 2008); and panel (d) shows the land price gradients by topographic interruptions (Saiz, 2010). These panels show supply restrictions are positively correlated with the levels of land prices but have little correlation with the slopes of the gradients.

5.2. The Land Share Gradient

Figures 7, 8 and 9 highlight some of our results using data on land shares across ZIP codes. Similar to Figure 5, Figure 7 shows results for the Washington, DC and San Francisco CBSAs; and similar to Figure 6, Figure 8 shows how land shares vary with distance to the CBD, and how that relationship changes based on population, regulation and topography. Comparing the results of Figures 7 and 8 to those of Figures 5 and 6, we draw three conclusions. First, similar to land values, land shares vary widely across geography. Second, land shares also fall with distance to the CBD, though the rate of decline is generally modest. Some decline is to be expected as our model predicts that land's share of home value rises with age until the option to rebuild is exercised, and in most cities older homes are located closer to the CBD than newer homes.²⁵ Finally, greater regulation and topographical interruption are associated with higher land shares.

Figure 9 shows the relationship of the land share of value for all homes in each ZIP code (x-axis) against the land share for just the newer homes in our working sample in that ZIP code (y-axis). The figure shows that, on average, newer homes have lower land shares, as the intercept of the regression line is just about 0 and the slope is 0.86. This figure also corroborates the predictions of the optimal teardown model: land share of value increases with the age of the structure, all else equal.

²⁵The mix of older and newer homes helps explain the steep decline in the land share for very large cities with low regulation, shown in panel (c) of Figure 8. The only cities in this group are Dallas and Houston. In both cities, the number of newer homes rises sharply beyond 10 miles from the CBD, which boosts the structure share and reduces the land share. The other city groups in panel (c) have much more limited increases in newer homes with distance from the CBD.

5.3. Land Prices, Structure Density, and the Reservation Use

Economists typically model the production of housing as a function of structures and land inputs. Earlier, we showed that the price of land tends to fall with distance to the CBD. As the price of land rises, with no change in structure cost, profit maximizing builders should use less land in housing developments. We can test this prediction directly. The assessor data contain information on both interior square feet and the lot size. We use these two variables to construct the ratio of the interior square footage to the lot size, known as the “floor-area ratio” (FAR), for single-family homes in all areas where we estimate land value.

Panel (a) of Figure 10 graphs FARs for single-family housing against land prices for the counties in our data. At a low land price per acre, structure density is low, with the cheapest land containing FARs at about 0.05 (i.e. a house with about 2,200 interior square feet on a one-acre lot) while the most expensive land contains FARs near 1 (i.e. a 3 story, 2,400-square-foot row house on a 2,400-square-foot lot). Panel (b) demonstrates that on average, lot sizes decrease with land prices. Both panels show that builders economize on land when its price is high.

Panel (c) shows the relation of the average price of agricultural land by county, as measured by the U.S. Department of Agriculture, for counties with some land in agricultural use, and the average value of land in residential use in those counties. Urban economics models predict that these values should be linked, as land at the edge of urban areas can be used for either agricultural or residential use. Not surprisingly, panel (c) shows a strong positive relationship.

5.4. Barriers to Building Housing and Urban Decline

Figure 11 shows the correlation across CBSAs of as-is land prices per acre in millions of 2013 dollars, left column, and land shares of house value, right column, against covariates of housing-supply elasticities. Each dot in every graph represents a CBSA from the pooled cross-sectional data. The top and middle panels display, respectively, the correlation with the Wharton regulatory index of Gyourko, Saiz, and Summers (2008) and with topographic interruptions as measured by Saiz (2010). These panels show that, on average, the price of land and the share of housing attributable to land value increase across CBSAs with both regulatory burden and topographic difficulty in building housing. The bottom panel shows the correlation of land prices and land shares with the fraction of the housing stock with a

value less than replacement cost in 1990 (“Urban Decline”), as first documented by Glaeser and Gyourko (2005). The bottom panel shows that metropolitan areas that are in relative decline, as evidenced by a larger fraction of the housing stock that is below replacement cost, also have relatively low land prices and low land shares.

5.5. Changes in Land Shares

Figure 12 shows how land shares changed in the data between 2012 and 2018. Panel (a) plots a histogram of all the ZIP-level changes and panel (b) plots changes in the land share as a function of the land share in 2012. Both panels suggest that land shares increased slightly, on average, during the housing recovery but that there was wide variation in the change across ZIP codes. Panel (c) shows changes as a function of the initial land share in 2012, for cities with more than 2 million housing units (blue line), cities with 500 thousand to 2 million housing units (red line), and cities with less than 500 thousand housing units (green line). This panel shows the share of house value attributable to land tended to increase in cities with at least 500 thousand units and decline in smaller cities, as the blue and red lines are everywhere above 0 and the green line is everywhere below it.

Although not apparent from the ZIP-level results in Figure 12, the land share for the United States in the aggregate increased from 35.7% in 2012 to 39.2% in 2018. This 3.5 percentage point increase was driven by a rapid increase in land prices at the national level of 8.6 percent per year that outstripped the national rise in house prices. Figure 13 reconciles the relatively balanced changes in land shares shown in Figure 12 with the substantial increase at the national level.

The red bars in Figure 13 show the change in the land share over 2012-2018 for counties sorted into 25 bins based on the number of single-family housing units from 2013-2017. We use counties rather than ZIPs because counties are the most granular units that we aggregate to higher-level geographies. Each bin represents the experience of about 90 counties. The blue dots show the aggregate value of land in single-family, residential use in 2012 for all the counties included in each bin. These dots represent the weight for each bin in the national aggregate. The figure tells an interesting story: Almost all of the bottom 18 bins ranked by county-level population, 1,650 counties in all, had declining land shares. In contrast, the top seven bins all experienced an increase in land shares, including very large increases in the highest and third-highest bins. The highest bin alone, which is dominated by counties in California, has much more weight in the national aggregate than do the lowest 18 bins

together. Thus, even though land's share of home value was flat or falling over 2012-2018 in most counties in the United States, the large rise in the land share in the most populous counties drove up the national share.

6 Conclusion

Although it is widely recognized that booms and busts in house prices reflect, in part, booms and busts in underlying land prices, until recently little data was available to study land prices. We help fill this gap by using a very large data set of appraisals to generate annual panel data from 2012 through 2018 of the average price of land used in single-family homes for 1,132 counties, 9,194 ZIP codes, and 19,252 census tracts. We also calculate pooled cross-sectional estimates of land prices for about twice as many counties and ZIP codes and about three times as many census tracts. Overall, we document a number of properties of the level and growth rate of land prices that are generally consistent with predictions of traditional models of urban economics. We expect that future researchers will use the data we generate to build on our results, and current and future policy-makers will monitor these data to better understand emerging risks in housing markets.

References

- ALBOUY, D., G. EHRLICH, AND M. SHIN (2018): “Metropolitan Land Values,” *Review of Economics and Statistics*, 100(3), 101–120.
- ALONSO, W. (1964): *Location and land use: Toward a general theory of land rent*. Harvard University Press.
- BASU, S., AND T. G. THIBODEAU (1998): “Analysis of spatial autocorrelation in house prices,” *Journal of Real Estate Finance and Economics*, 17(1), 61–85.
- BOSTIC, R. W., S. D. LONGHOFFER, AND C. L. REDFEARN (2007): “Land leverage: Decomposing home price dynamics,” *Real Estate Economics*, 35(2), 183–208.
- BRUECKNER, J. K. (1987): “The structure of urban equilibria: A unified treatment of the Muth - Mills model,” *Handbook of Regional and Urban Economics*, 2, 821–845.
- CASE, K. E., AND R. J. SHILLER (1989): “The Efficiency of the Market for Single-Family Homes,” *American Economic Review*, 79(1), 125–137.
- CLAPP, J. M., P. EICHHOLTZ, AND T. LINDENTHAL (2013): “Real option value over a housing market cycle,” *Regional Science and Urban Economics*, 43(6), 862–874.
- DAVIDOFF, T., ET AL. (2016): “Supply constraints are not valid instrumental variables for home prices because they are correlated with many demand factors,” *Critical Finance Review*, 5(2), 177–206.
- DAVIS, M. A., AND J. HEATHCOTE (2007): “The price and quantity of residential land in the United States,” *Journal of Monetary Economics*, 54(8), 2595–2620.
- DAVIS, M. A., S. D. OLINER, E. J. PINTO, AND S. BOKKA (2017): “Residential land values in the Washington, DC metro area: New insights from big data,” *Regional Science and Urban Economics*, 66, 224–246.
- DAVIS, M. A., AND F. ORTALO-MAGNÉ (2011): “Household Expenditures, Wages, Rents,” *Review of Economic Dynamics*, 14(2), 248–261.
- DAVIS, M. A., AND M. G. PALUMBO (2008): “The price of residential land in large U.S. cities,” *Journal of Urban Economics*, 63(1), 352–384.
- GLAESER, E. L., AND J. GYOURKO (2005): “Urban decline and durable housing,” *Journal of political economy*, 113(2), 345–375.
- GYOURKO, J., A. SAIZ, AND A. SUMMERS (2008): “A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index,” *Urban Studies*, 45(3), 693–729.

- HARDING, J. P., S. S. ROSENTHAL, AND C. SIRMANS (2007): “Depreciation of housing capital, maintenance, and house price inflation: Estimates from a repeat sales model,” *Journal of Urban Economics*, 61(2), 193–217.
- HAUGHWOUT, A., J. ORR, AND D. 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).
- HENGL, T. (2007): *A Practical Guide to Geostatistical Mapping of Environmental Variables*. Luxembourg: Office for Official Publications of the European Communities.
- KATZ, A. J., AND S. W. HERMAN (1997): “Improved estimates of fixed reproducible tangible wealth,” *Survey of Current Business*, 77(5), 69–92.
- LUTZ, B., R. MOLLOY, AND H. SHAN (2011): “The housing crisis and state and local government tax revenue: Five channels,” *Regional Science and Urban Economics*, 41(4), 306–319.
- MCMILLEN, D., AND A. O’SULLIVAN (2013): “Option value and the price of teardown properties,” *Journal of Urban Economics*, 74, 71–82.
- MILLS, E. S. (1967): “An Aggregative Model of Resource Allocation in a Metropolitan Area,” *American Economic Review*, 57(2), 197–210.
- MUTH, R. F. (1969): *Cities and housing; the spatial pattern of urban residential land use*. University of Chicago Press.
- NICHOLS, J. B., S. D. OLINER, AND M. R. MULHALL (2013): “Swings in commercial and residential land prices in the United States,” *Journal of Urban Economics*, 73(1), 57–76.
- SAIZ, A. (2010): “The Geographic Determinants of Housing Supply,” *The Quarterly Journal of Economics*, 125(3), 1253–1296.
- SAKS, R. E. (2008): “Job Creation and Housing Construction: Constraints on Metropolitan Area Employment Growth,” *Journal of Urban Economics*, 64(1), 178–195.
- TITMAN, S. (1985): “Urban land prices under uncertainty,” *American Economic Review*, 75(3), 505–514.

A Appendix: Spatial Interpolation Methods

This section describes various interpolation methods discussed in the paper. Our land price index approach begins by dividing the universe of parcels N within a geography into those with values that are observed (i.e. included in our working data) and those with values that are not observed. For convenience, below we denote the number of observed values in a geography as N^s . All land prices we report are calculated as the simple average of the price estimate (or actual value, when available) of each individual parcel within the county or county subaggregate (i.e. census tract). Accordingly, differences between land values within a given geography arise due the method used to estimate prices for parcels with unobserved values.

At the end of the day, we estimate the value of land when land value is not directly observed as a weighted average of land values of a subset of parcels within the county where land prices are directly observed. The estimated price for parcel i is calculated as the average of the n nearest (by proximity) neighbor parcels, indexed by j , with weights $\lambda_{i,j}$.²⁶ For a particular method, the estimated price for parcel i is:

$$\hat{p}_i = \sum_{j=1}^n \lambda_{i,j} p_{i,j} \quad (10)$$

The weights are assumed to sum to unity, or $\sum \lambda_{i,j} = 1$.

1.1. Spatial Statistics

Spatial statistical methods do not consider spatial relations in outcomes, only proximity. Here, we discuss three spatial statistics that are commonly used to interpolate prices spatially. The general form of each of these statistics is below, where h is the distance between the location to be imputed, i , and another nearby sampled location, j , that is one of the n nearest locations. The exponent c gives the degree of decay in the weight that is due to distance between the parcels.

²⁶So $p_{i,1}$ is the closest observed price to location i ; $p_{i,2}$ is the second closest observed price to location i ; and so on.

$$\lambda_{i,j} = \frac{1}{h_{i,j}^c} \left(\sum_{j=1}^n \frac{1}{h_{i,j}^c} \right)^{-1} \quad (11)$$

Null Estimator

The null estimator (“Null”) sets $n=N^s$ and $c = 0$, giving $\lambda_{i,j} = 1/n$. This gives the estimate of an individual parcel as the sample average.

Nearest Neighbor

The nearest neighbor estimator (“NN”) also sets $c = 0$, giving $\lambda_{i,j} = 1/n$. But n is typically set to the 5 to 25 nearest observed prices. This gives the estimate of a parcel as the sample average of nearby parcels.

Inverse-distance weights

The inverse-distance weight estimator (“IDW”) sets n generally within the range of 5 to 25 nearest observed prices as with the NN estimator. The calculation of λ then involves assuming an exponent c . This exponent is commonly set equal to $c = 2$, giving a relation between points that declines with the square of the distance. In this case, $\lambda_{i,j} = \frac{1}{h_{i,j}^2} \left(\sum_{j=1}^n \frac{1}{h_{i,j}^2} \right)^{-1}$

1.2. Geostatistics

In addition to spatial statistics, we include a single geostatistical estimator: ordinary kriging. As with the nearest neighbor estimator, n nearest neighbors are weighted and summed to generate predicted prices. λ is calculated based on the strength of the observed relationship between observations of different proximities within the sample.

Derivation of the weights proceeds in five steps. The first step involves calculating pairwise differences in values between each pair in the sample within a certain distance range. The next step collapses and bins the semivariances (half of the squared differences) into averages by the distance between the points. The third step fits a curve, often referred to as a “variogram,” to this set of binned averages. The fourth step applies this estimated curve to estimate covariances between values in an unsampled location and nearby sampled locations. The fifth steps uses these fitted covariances to construct the weights.²⁷

²⁷For a more in-depth overview of kriging, see Hengl (2007).

As an illustration of this procedure, we present kriging steps for land prices in Washington, DC, pooled between 2012 and 2018. There are about 110,000 parcels and 16,000 sampled standardized land prices. To start, differences and semivariances γ are calculated for each pair of points in N .²⁸

The results from the first two kriging steps are shown in Figure A.1. The hollow circles represent semivariance averages within each of the 15 distance bins. Distances are reported as miles but are in terms of latitude/longitude degrees in the programs.²⁹

In step three, a functional form for the relationship between the semivariances and the distances (the hollow circles) must then be assumed and fit to the data. The fitted curve, as shown by the blue line, is typically upward sloping, indicating the greater the distance the higher the variance. The spherical functional form has three parameters, a_0 , a_1 , and r that we estimate

$$\gamma(h; a_0, a_1, r) = \begin{cases} a_0 + a_1\left(\frac{3}{2}\left(\frac{h}{r}\right) - \frac{1}{2}\left(\frac{h}{r}\right)^3\right), & 0 < h < r \\ a_0 + a_1, & h \geq r \end{cases} \quad (12)$$

The three parameters combine to give the “sill” which is the value to which the variogram asymptotically approaches as the distance between points approaches infinity, or $a_0 + a_1$; the “nugget” which is the value of the variogram when distance approaches zero, or a_0 , and the “range,” r , which is the value of h when the variogram reaches the sill.³⁰

In addition to the binned semivariances, Figure A.1 also shows a fitted spherical functional form for Washington, DC, with $\hat{a}_0 = 0.06$, $\hat{a}_1 = 0.54$ and $\hat{r} = 8.97$ miles. This function is used

²⁸The semivariance for prices at two points i and j is half of the squared difference, $\frac{1}{2}(p_i - p_j)^2$. Semivariances are used instead of variances to facilitate construction of weights using covariances in later steps, and this is the standard in the geospatial literature. Isotropy (i.e. the direction between the points does not affect the strength of the relationship) is a standard assumption, which we make here in order to express proximity using a single variable, h .

²⁹In other words, we take the square root of the sum of the squared differences between two sets of coordinates. Since our distances are generally small, we use this simplified distance measure as a proxy for the actual distance, which varies due to changes in latitude.

³⁰The spatial statistics previously discussed are special cases of models that can be fit to a variogram. For instance, when $a_1 = 0$, the kriging estimator reduces to the nearest neighbor weights. When the function is an exponential with $a_0 = 0$ and an exponent of 2, then the kriging estimator reduces to inverse-distance weights. The advantage of the kriging estimator is that it does not place these parameter restrictions on the spatial relation, *a priori*.

to estimate the semivariance between any two hypothetical points, facilitating interpolation of prices in unsampled locations.³¹

The fourth step specifies the function transforming the fitted semivariances to covariances. For any two points j and k , a distance $h_{j,k}$ is calculated. Then, when combined with the fitted variogram parameters, the covariance, C , between the points is estimated as:

$$C(h_{j,k}; \hat{a}_0, \hat{a}_1, \hat{r}) = \hat{a}_0 + \hat{a}_1 - \gamma(h_{j,k}; \hat{a}_0, \hat{a}_1, \hat{r}) \quad (13)$$

The final step involves constructing the matrices and performing the operations necessary to arrive at the weights. First, construct c_i , an $n \times 1$ vector of estimated covariances between an unsampled location i and its n nearest points. Then, we construct C_i , an $n \times n$ covariance matrix of the n nearest points. These matrices are augmented in the standard fashion with a Lagrange multiplier and column/row vectors of ones and a zero to normalize the weights to sum to one. These give the weights in λ_i , an $n \times 1$ vector.

$$\begin{bmatrix} \lambda_i \\ \mathcal{L} \end{bmatrix} = \begin{bmatrix} \mathbf{C}_i & \mathbf{1} \\ \mathbf{1}' & 0 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{c}_i \\ 1 \end{bmatrix} \quad (14)$$

1.3. Comparison

We compare the fit of each spatial interpolation method by comparing actual standardized land prices for a 20% hold-out sample to predicted land prices estimated using an 80% training sample. We consider a number of values for the number of nearest neighbors and the overall distance boundary considered.

These results are shown in Table A.1 in terms of the mean, median, and standard deviation of RMSEs across the 2292 counties in the pooled sample. The mean RMSE in the 20% hold-out sample is 0.517. Nearest neighbor and inverse-distance weights give increasing accuracy, with mean RMSEs falling to 0.418 and 0.413, respectively.

Each of the kriging estimates gives similar average fit, with means RMSEs ranging between 0.389 or 0.394. We interpret this finding as indicating that neither the boundary nor the

³¹Other functional forms are common, especially the exponential function, $\gamma(h; a_0, a_1, r) = a_0 + a_1(1 - \exp(-\frac{h}{r}))$. We consider the exponential function as well in some exercises.

number of nearest neighbors considered tends to affect the estimates at the ranges considered. Overall, we interpret these findings as lending support to our decision to use the 20 nearest neighbors and a 6.9 mile boundary in our county-specific kriging procedure.

B Appendix: Monte Carlo Simulation of Standard Urban Model

Assume that a city lies on a featureless plane with a region called the central business district (CBD) at its center. This district provides all employment and because commuting is costly in a way we specify precisely, households wish to live near the CBD. Spatial equilibrium requires identical households to have identical utility in all locations in the city. As we show, this implies that households consume less housing at a higher price near the CBD. Additionally, the housing production function implies housing is produced with high density and structure intensity near the CBD where land prices are high, and with low density near the edge of the city.

To be precise, assume a person consuming c units of consumption and h units of housing receives utility of

$$(1 - \alpha) \ln c + \alpha \ln h \tag{15}$$

If a person lives distance d from the city center, their wage after commuting is $w(1 - td)$ where t is the percentage of income that must be paid to commute for each unit of distance d . Denote the rental cost per unit of housing d units from the city center as q_d^h . A person living d units from the city center faces the budget constraint of:

$$w(1 - td) = c + q_d^h h \tag{16}$$

A person choosing to live in location d units away from the CBD maximizes utility (15) subject to the budget constraint (16) by choosing optimal consumption c_d and housing h_d of

$$c_d = (1 - \alpha) w (1 - td) \tag{17}$$

$$q_d^h h_d = \alpha w (1 - td) \tag{18}$$

This means maximized utility at distance d from the center can be written as \mathcal{U}_d

$$\begin{aligned}\mathcal{U}_d &= (1 - \alpha) \ln [(1 - \alpha) w (1 - td)] + \alpha \ln [\alpha w (1 - td) / q_d^h] \\ &= \kappa_u + \ln w + \ln (1 - td) - \alpha \ln q_d^h\end{aligned}\tag{19}$$

where κ_u is a constant equal to $\alpha \ln \alpha + (1 - \alpha) \ln (1 - \alpha)$. In equilibrium, we assume all locations have to provide the same utility, for example location d and d' must satisfy

$$\mathcal{U}_d = \mathcal{U}_{d'}$$

Then from equation (19) this implies

$$\begin{aligned}\ln (1 - td) - \alpha \ln q_d &= \ln (1 - td') - \alpha \ln q_{d'}^h \\ \frac{q_{d'}^h}{q_d^h} &= \left(\frac{1 - td'}{1 - td} \right)^{\frac{1}{\alpha}}\end{aligned}\tag{20}$$

Equation (20) governs the rate at which housing rental prices per unit change with distance from the CBD, roughly t/α percent per unit of d .

Note that we can also work out how the quantity of housing changes as a function of distance to the CBD. We start by using the definition of utility and substituting in optimal consumption but keeping housing

$$\mathcal{U}_d = (1 - \alpha) \ln [(1 - \alpha) w (1 - td)] + \alpha \ln h_d\tag{21}$$

Once we impose $\mathcal{U}_d = \mathcal{U}_{d'}$, this gives us

$$\begin{aligned}(1 - \alpha) \ln (1 - td) + \alpha \ln h_d &= (1 - \alpha) \ln (1 - td') + \alpha \ln h_{d'} \\ \rightarrow \frac{h_{d'}}{h_d} &= \left(\frac{1 - td'}{1 - td} \right)^{-\frac{1-\alpha}{\alpha}}\end{aligned}\tag{22}$$

Now that we have worked out how housing quantities h and prices per unit q_d^h vary from the city center, we can also work out how the quantities and prices of land and structures change with distance. Temporarily suppressing the distance subscripts, assume competitive

builders build housing using land l and structures s according to a CES production function

$$h = [(1 - \theta) s^\rho + \theta l^\rho]^{\frac{1}{\rho}} \quad (23)$$

with $\rho \in (-\infty, 1]$. Assume each unit of housing generates revenue of q^h ; further, assume each unit of land costs q^l and each unit of structure costs 1. Builders maximize

$$q^h [(1 - \theta) s^\rho + \theta l^\rho]^{\frac{1}{\rho}} - s - q^l l \quad (24)$$

The first-order conditions for optimal structures are

$$1 = q^h h^{1-p} (1 - \theta) s^{p-1} \quad (25)$$

$$\rightarrow s = [q^h (1 - \theta)]^{\frac{1}{1-p}} h \quad (26)$$

This implies that once we know q^h and h , we also know s . Note that because we know s , we also know $q^l l = q^h h - s$. Now consider the first-order condition for optimal land:

$$q^l l = q^h h^{1-p} \theta l^p \quad (27)$$

and thus

$$l = \left[\frac{q^l l}{q^h h^{1-p} \theta} \right]^{\frac{1}{p}} \quad (28)$$

Given a set of parameters, we can compute how quantities and prices and expenditures on housing, structures and land change with distance from a CBD. For a rough calibration, we set $\alpha = 0.25$ based on the median housing budget shares of renters as documented by Davis and Ortalo-Magné (2011). The other parameters we set to match some approximate features of a city. We set $t = 0.02$ such that people 10 miles from the CBD consume about double the housing than people at the CBD but spend 20% less.³² We jointly set $\theta = 0.90$ and $\rho = -2.0$ such that land's share of value rises from about 15% 10 miles from the CBD to about 55% at the CBD. We normalize the price per unit of housing to 1 at the CBD and normalize the quantity of housing consumed at the CBD such that the total expenditure at the CBD is for a \$1 million house. As noted earlier, we assume the price per unit of housing structure

³²This is referring to single-family homes.

is 1.0 everywhere in the metro area. Table A.2 shows prices, quantities and expenditures on housing, structures and land as well as land's share of house value, the quantity of land once we normalize the size of a single-family plot at the CBD to 0.25 acres, and land price per acre.

We simulate two data sets based on the calibration of this model. In both data sets, we draw 100 observations for houses in neighborhoods uniformly between 0 and 3.5 miles from the CBD; 200 observations for houses in neighborhoods uniformly between 3.5 and 7.5 miles from the CBD; and 300 observations for houses in neighborhoods uniformly between 7.5 miles and 10 miles from the CBD. In the first data set we assume no quantities or prices are measured with error. This enables us to see the accuracy of the Kriging procedure with regards to this application in an ideal environment.

In the second data set, we allow for i.i.d. measurement error in both the value of housing and the value of structures.³³ This simulation gives us some intuition for how the Kriging procedure performs under conditions where land is imperfectly measured. Denote $\widetilde{q_d^h h_d}$ as observed housing value and \widetilde{s} as observed structures costs. $\widetilde{q_d^h h_d}$ and \widetilde{s} are determined as

$$\begin{aligned}\widetilde{q_d^h h_d} &= q_d^h h_d (1 + e_d^h) & e_d^h &\sim U[-0.10, 0.10] \\ \widetilde{s} &= s (1 + e_d^s) & e_d^s &\sim U[-0.10, 0.10]\end{aligned}$$

with e_d^h and e_d^s drawn independently. We then compute observed land value residually,

$$\begin{aligned}\widetilde{q_d^l l_d} &= \widetilde{q_d^h h_d} - \widetilde{s} \\ &= q_d^h h_d - s + (q_d^h h_d e_d^h - s e_d^s)\end{aligned}$$

Denote the term in parentheses as e_d^l , measurement error in land value. Even though the standard deviation of e_d^h and e_d^s are relatively small (5.7 percent each), the standard deviation of measurement error as a percent of true land value, measured as $e_d^l / (q_d^l l_d)$, in this second data set is much larger, 27 percent. The measurement error for land value is magnified because land value is residually measured and accounts for a relatively small fraction of home value (as we discuss earlier in the paper).

Table A.3 compares the estimates from Kriging for land price per acre by distance to CBD

³³This measurement error can be thought of as a deviation from model-determined prices.

to the numbers we compute analytically in Table A.2 for the simulated data set measured without error (data set 1) and the simulated data set with measurement error (data set 2) for 0 to 9 miles to the CBD. When the data are measured with error, the Kriging results are less accurate, though the average error across the distance bins is relatively small at 4.2%.

Figure 1: House Value, Structure Cost, Vacant Land by Age as Predicted by Simple Tear-down Model

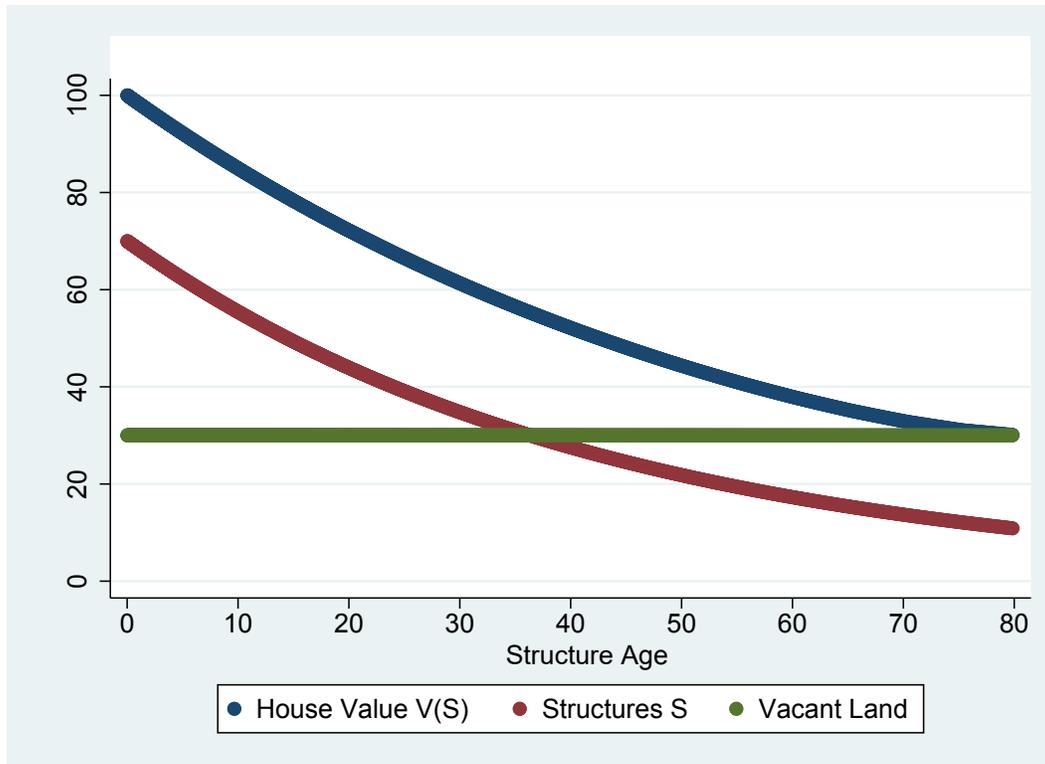


Figure 2: Land Share of House Value Computed Correctly and Incorrectly as Predicted by Simple Teardown Model

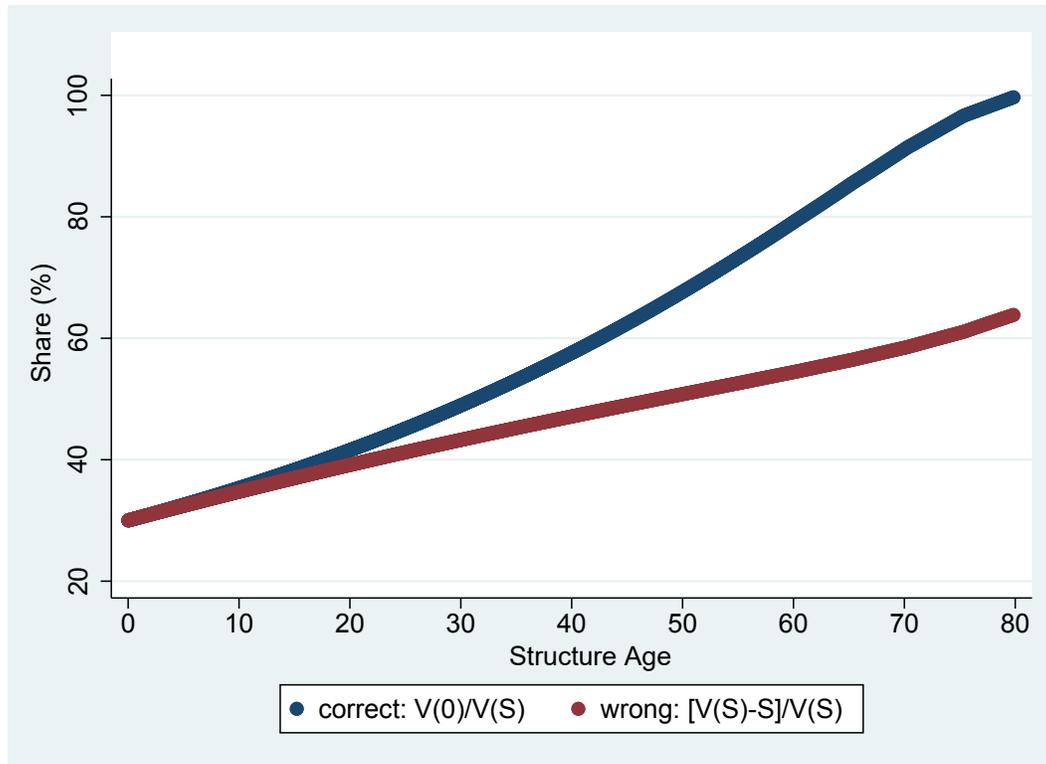
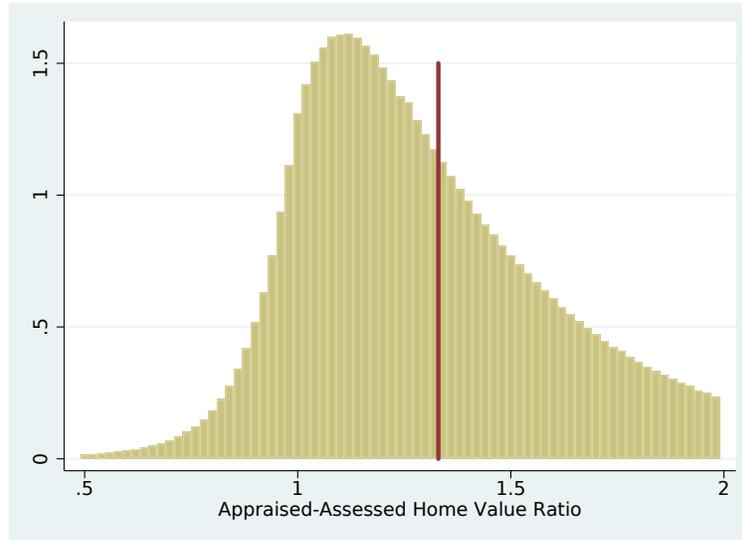
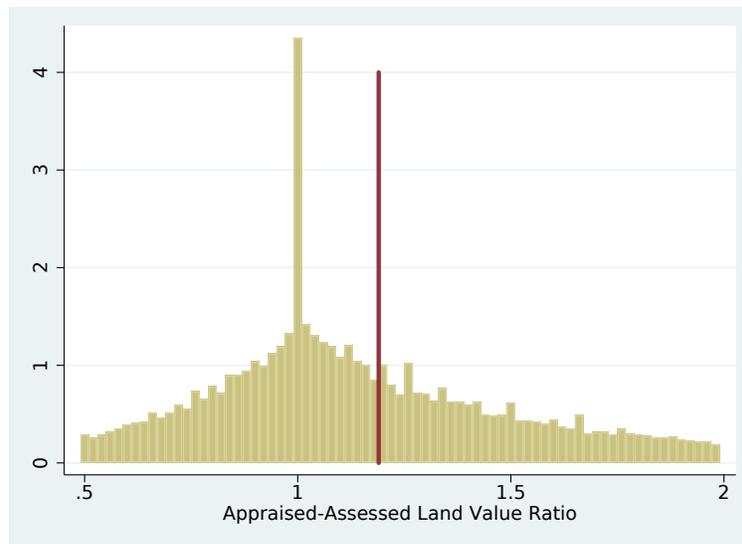


Figure 3: Appraisal Anchoring to Assessed Land and Property Values, Washington, DC

(a) Property Value Ratios, All Appraisals
(Median Value of 1.33)



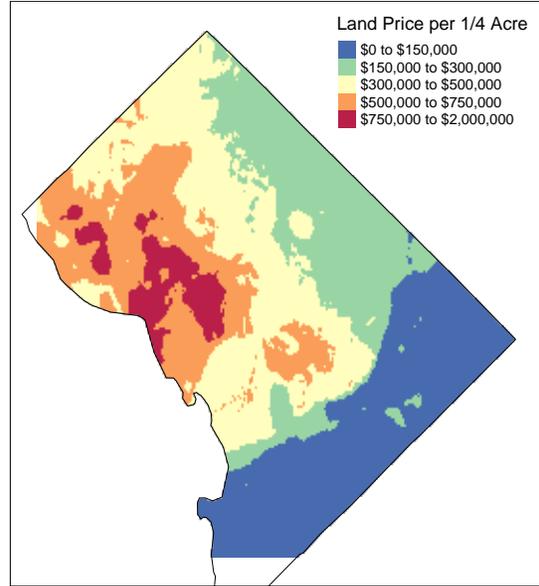
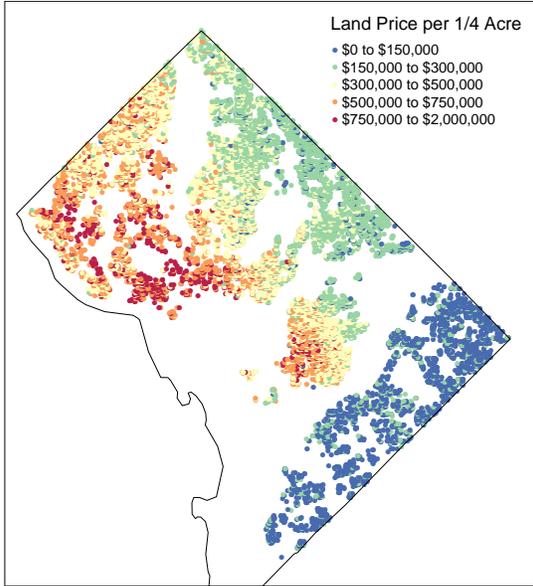
(b) Land Value Ratios, All Appraisals
(Median Value of 1.19)



Notes: This figure presents the ratio of appraised values to assessed values for the entire property (top panel) and for the land component (bottom). Appraisals are defined as anchored to assessments using methods described in the text. Values less than 0.5 and more than 2 are omitted from the display of the histogram but are included in the calculation of median values. Each red bar shows the median.

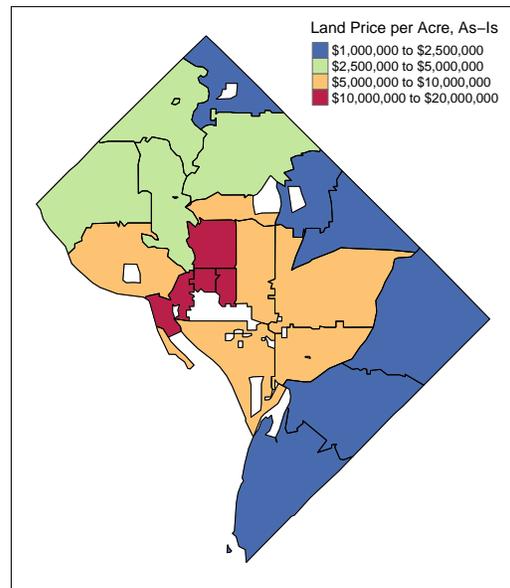
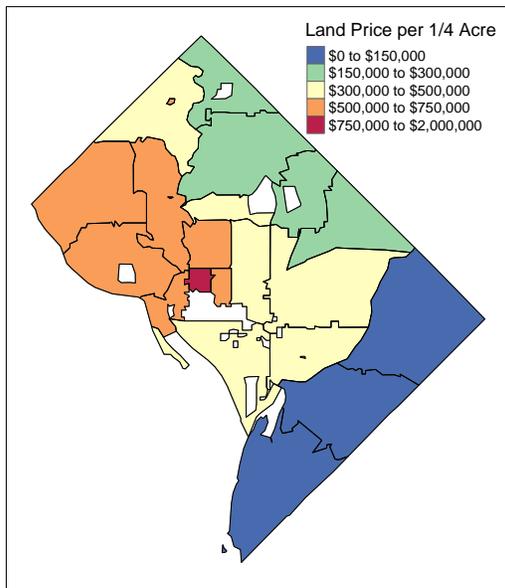
Figure 4: Washington, DC Kriging Example

(a) Standardized Land Prices - Working Sample (b) Standardized Land Prices - Kriged Surface



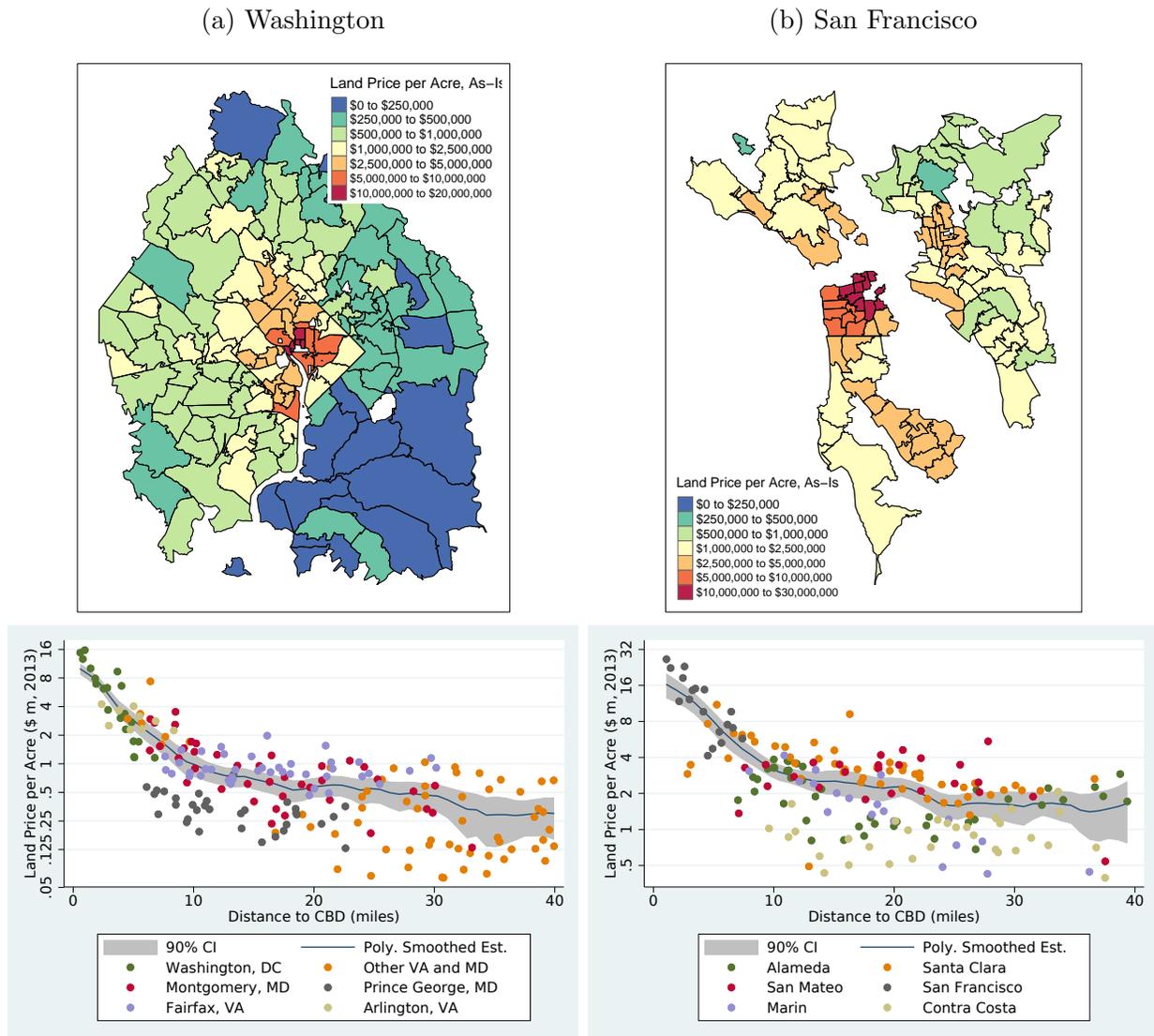
(c) Standardized Land Prices - ZIP code Avg

(d) As-Is Land Prices - ZIP code Avg



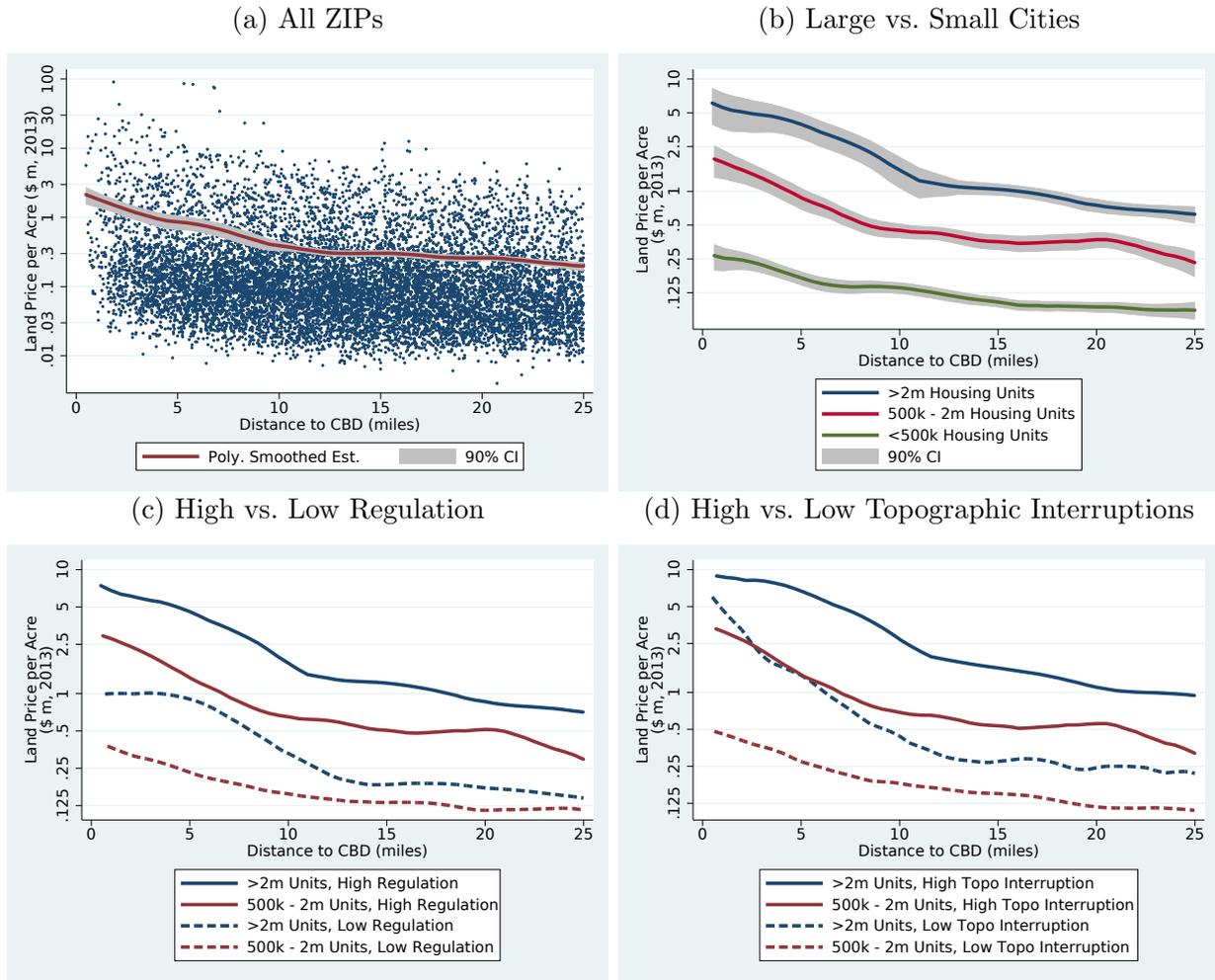
Notes: This figure shows the average level of land prices by ZIP code for the District of Columbia in our pooled sample over the period 2012-2018.

Figure 5: Land Prices, pooled cross-section (2013 prices)



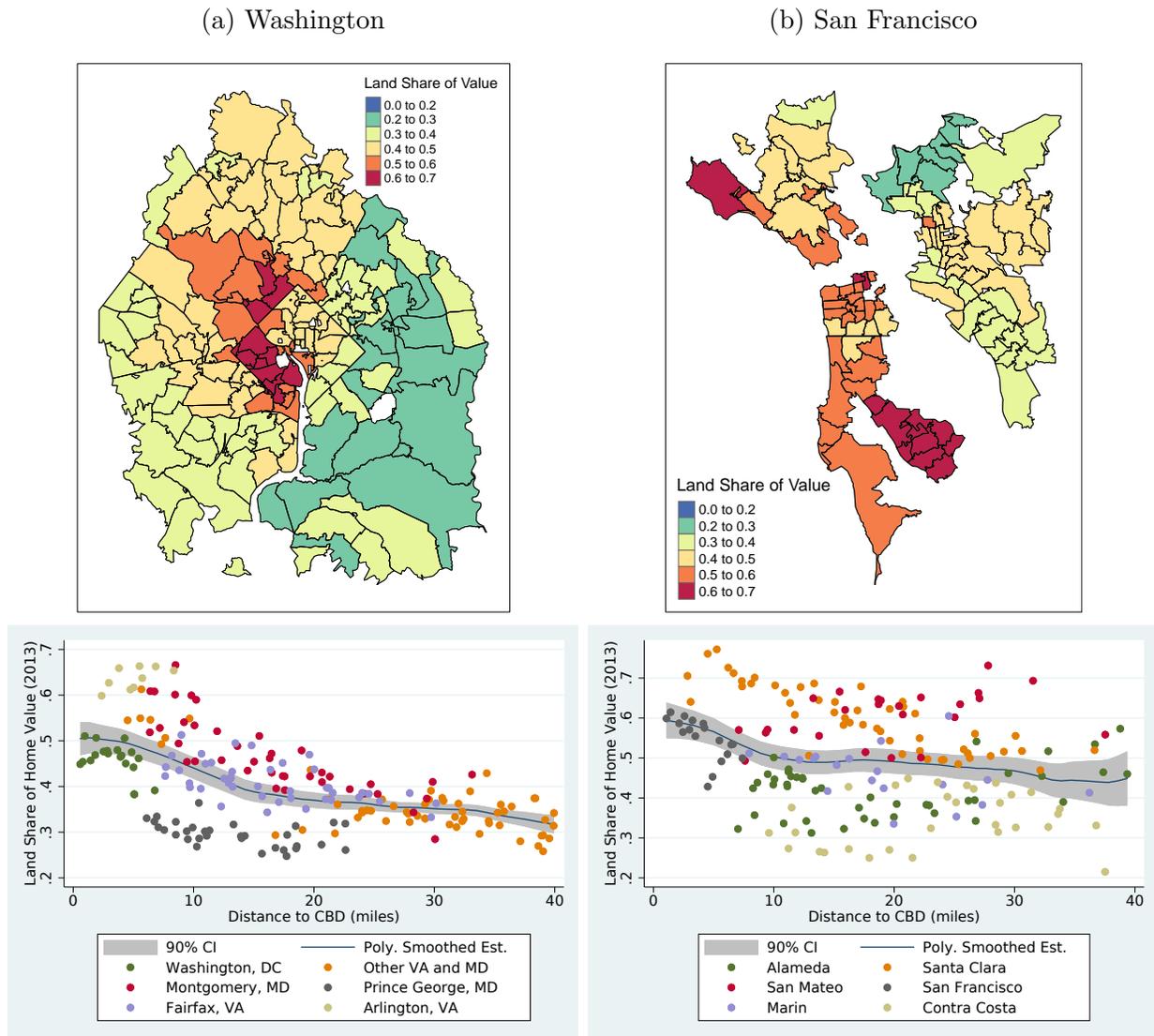
Notes: The sample in the maps includes all ZIP codes in the relevant counties. Graphs include a subsample of observations from the maps, including all ZIP codes with centroids within 40 miles of an identified central ZIP code centroid. The line is fit using a local polynomial smoother and the confidence intervals represent the 5th/95th percentile estimates.

Figure 6: Land Price Gradients



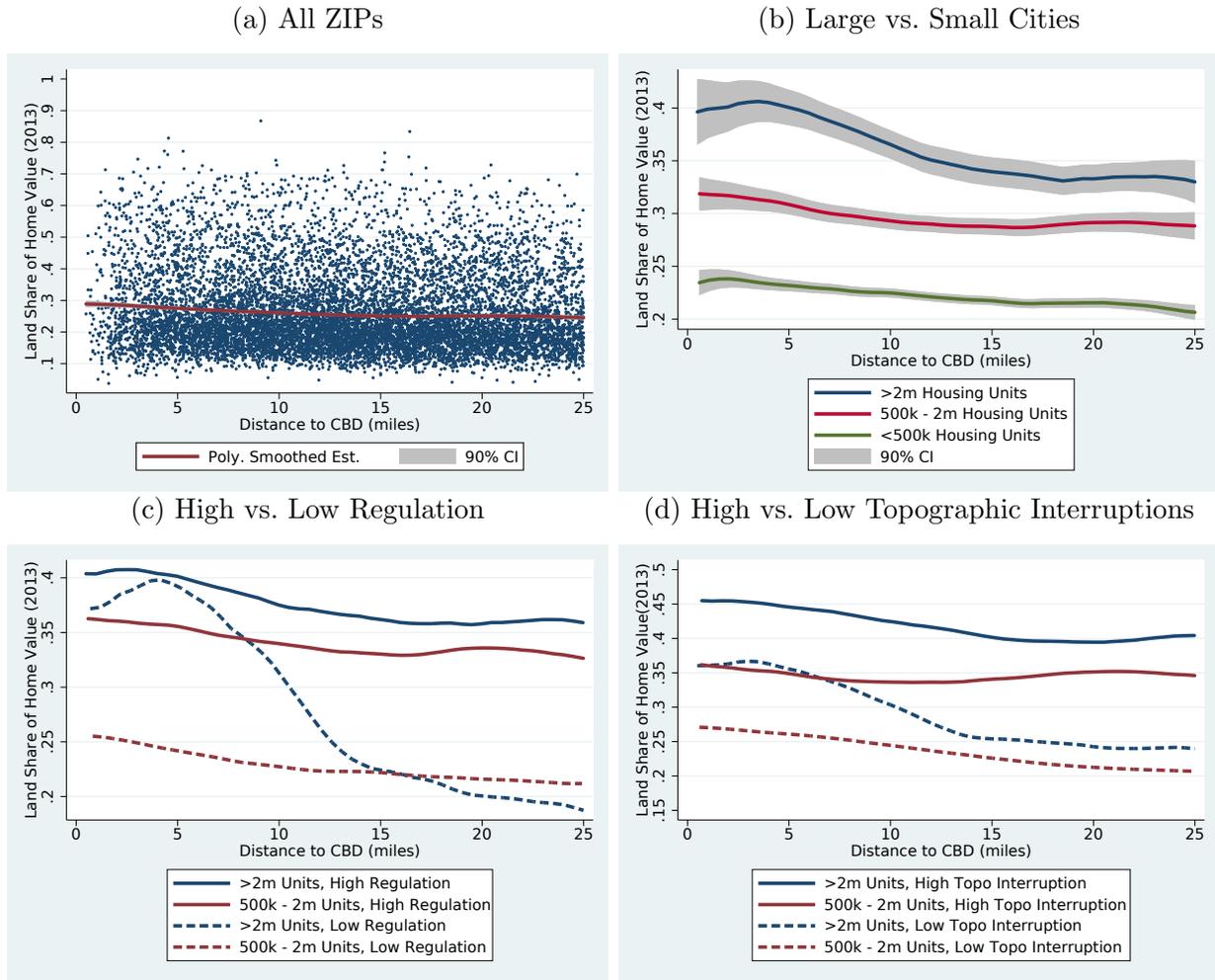
Notes: Samples include all ZIP codes with centroids within 25 miles of an identified central ZIP code centroid. The line is fit using a local polynomial smoother and the confidence intervals in the first panel represent the 5th/95th percentile estimates. Confidence intervals for remaining figures are calculated but omitted from the figures for clarity.

Figure 7: Land Share of Home Value, pooled cross-section (2013 prices)



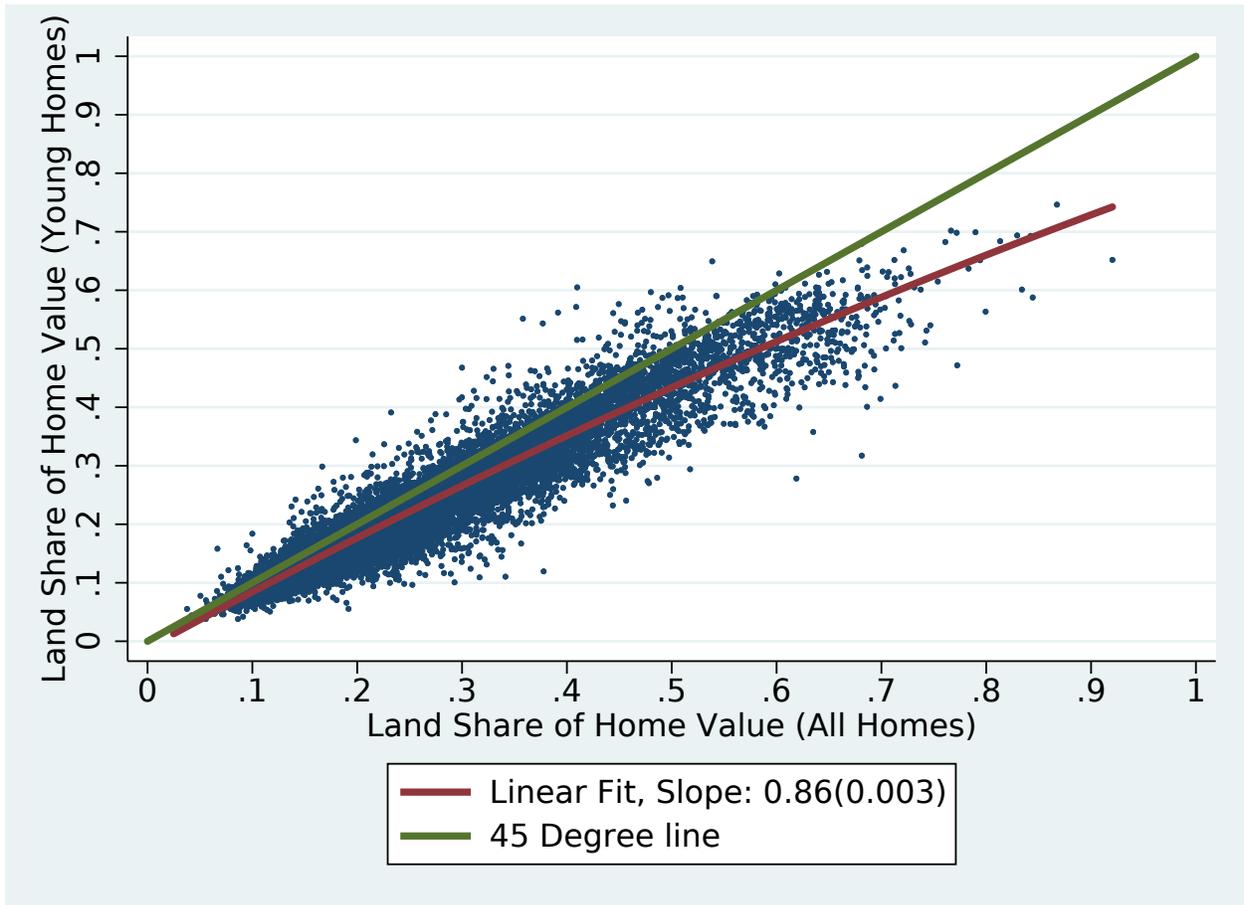
Notes: The sample in the maps includes all ZIP codes in the relevant counties. Graphs include a subsample of observations from the maps, including all ZIP codes with centroids within 40 miles of an identified central ZIP code centroid. The line is fit using a local polynomial smoother and the confidence intervals represent the 5th/95th percentile estimates.

Figure 8: Land Share of Home Value Gradients



Notes: Samples include all ZIP codes with centroids within 25 miles of an identified central ZIP code centroid. The line is fit using a local polynomial smoother and the confidence intervals in the first panel represent the 5th/95th percentile estimates. Confidence intervals for remaining figures are calculated but omitted from the figures for clarity.

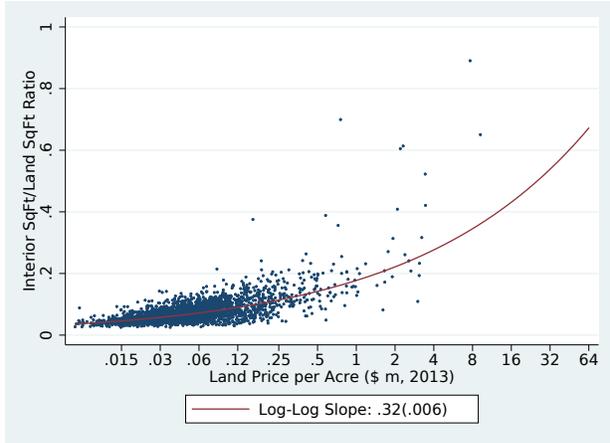
Figure 9: Land Share of Home Value, All Homes vs Young Homes



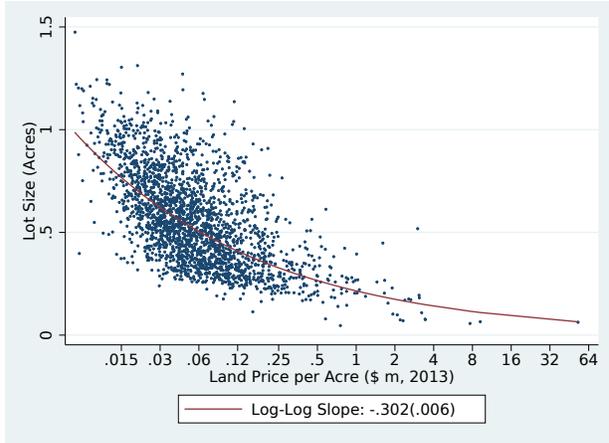
Notes: “All homes” land share is calculated as the average land share for all homes within a ZIP code. The “young homes” land share is the average land share for all homes within a ZIP code with an admissible appraisal (effective age ≤ 15 years and unanchored to tax assessment value).

Figure 10: County Land Prices and Covariates

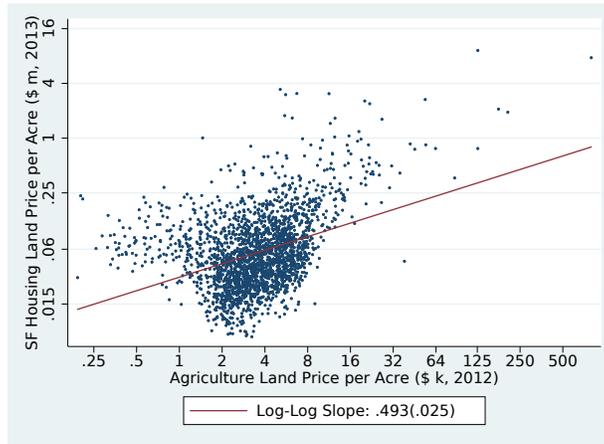
(a) Floor-Area Ratios and Land Prices



(b) Lot Size and Land Prices



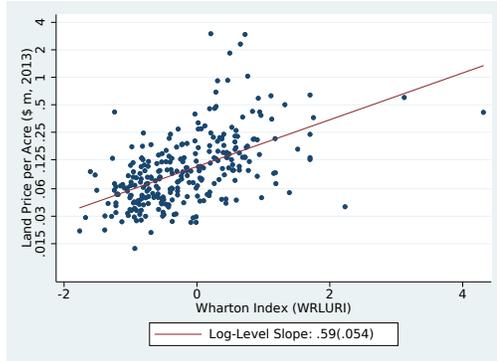
(c) Agricultural and Single-Family Land Prices



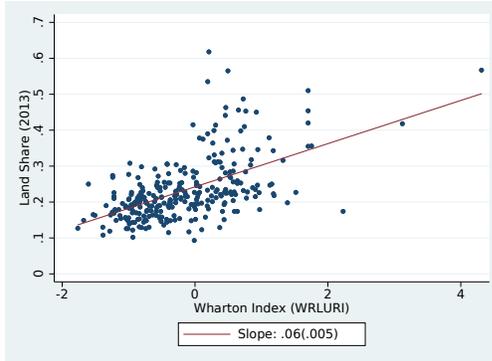
Notes: Sample includes all counties, except the agricultural land sample, which includes only counties with agricultural land. Slope estimates are based on the equation $\ln Y = a + b \ln X + e$, with the fit line based on $\hat{Y} = \exp(\hat{a} + \hat{b} \ln X)$. The value for \hat{b} is presented in the legend, with the standard error in parenthesis. Subfigure (a) omits a single point from the scatterplot: New York County, NY (FIPS 36061), which has a FAR of 1.40. The regression line includes this observation.

Figure 11: CBSA Land Prices, Shares, and Covariates

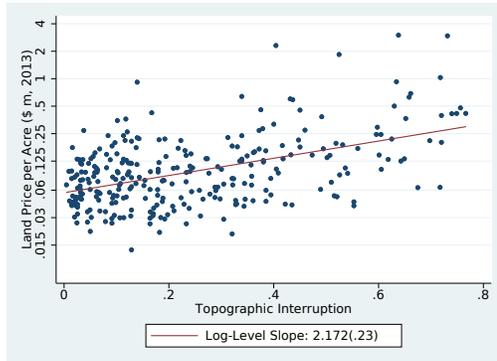
(a) Regulation - Land Prices



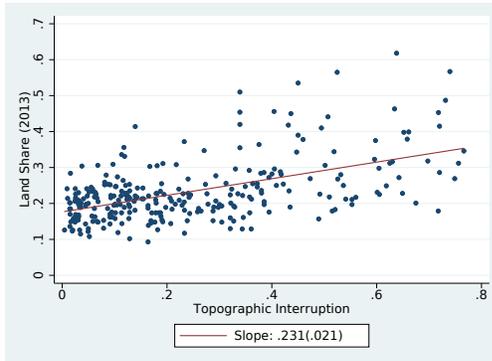
(b) Regulation - Land Shares



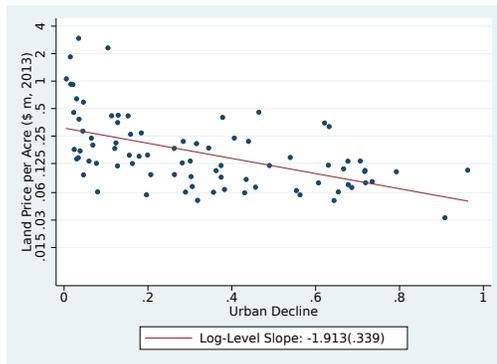
(c) Topo. Interruptions - Land Prices



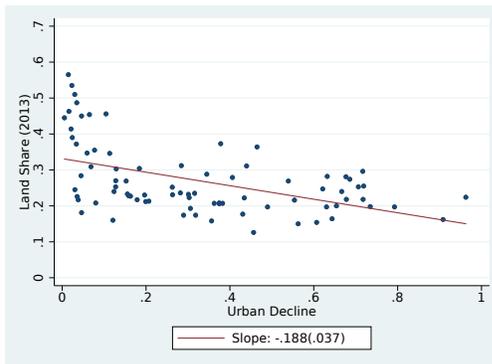
(d) Topo. Interruptions - Land Shares



(e) Urban Decline - Land Prices



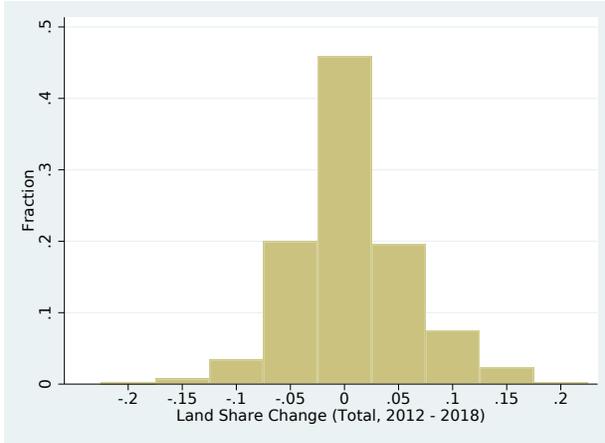
(f) Urban Decline - Land Shares



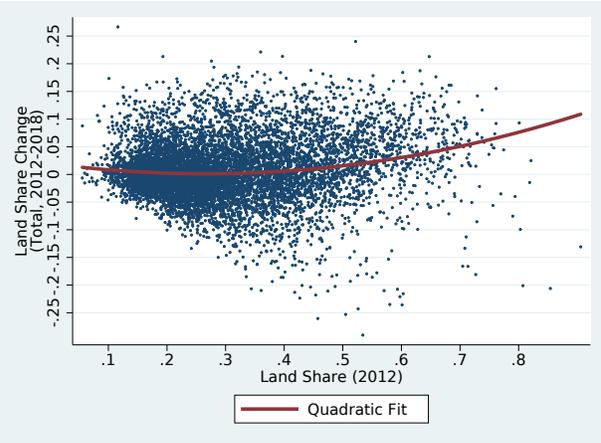
Notes: Sample includes all CBSAs with a measured covariate. Regulation is measured by the Wharton Land Use Regulatory Index (Gyourko, Saiz, and Summers, 2008), where positive values indicate higher regulation. Topographic Interruptions is the fraction of the land area within the city that is unavailable to development due to topographical interruptions (Saiz, 2010). Urban Decline is the fraction of the housing stock with a value less than replacement costs in 1990 (Glaeser and Gyourko, 2005).

Figure 12: Changes to Land Share of Home Value by ZIP Code, 2012-2018

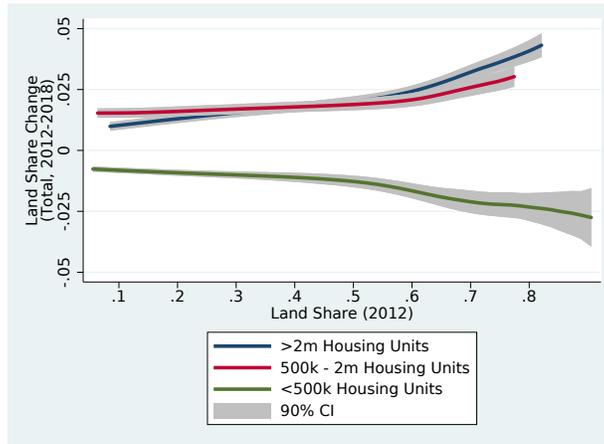
(a) Histogram



(b) By Initial Land Share

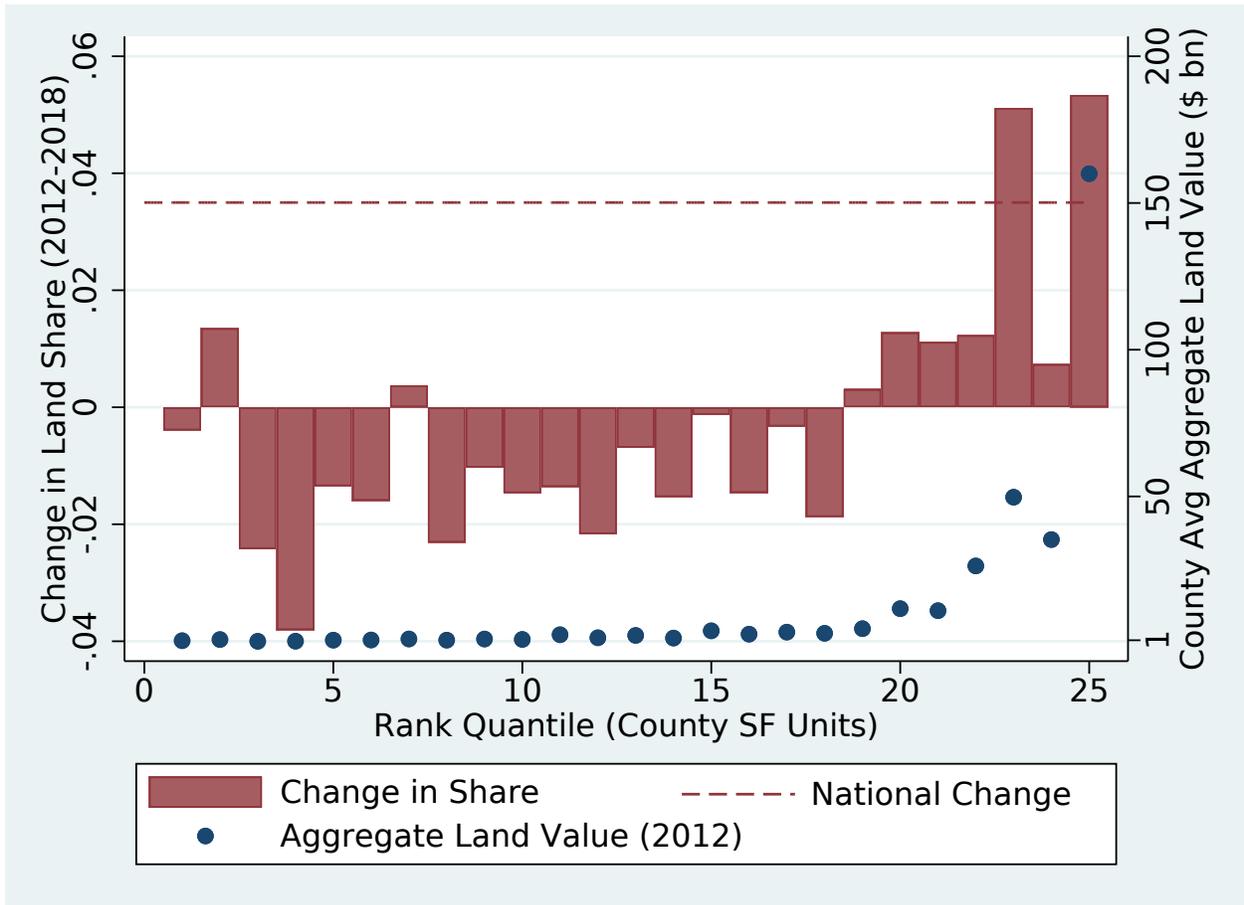


(c) By Population and Initial Land Share



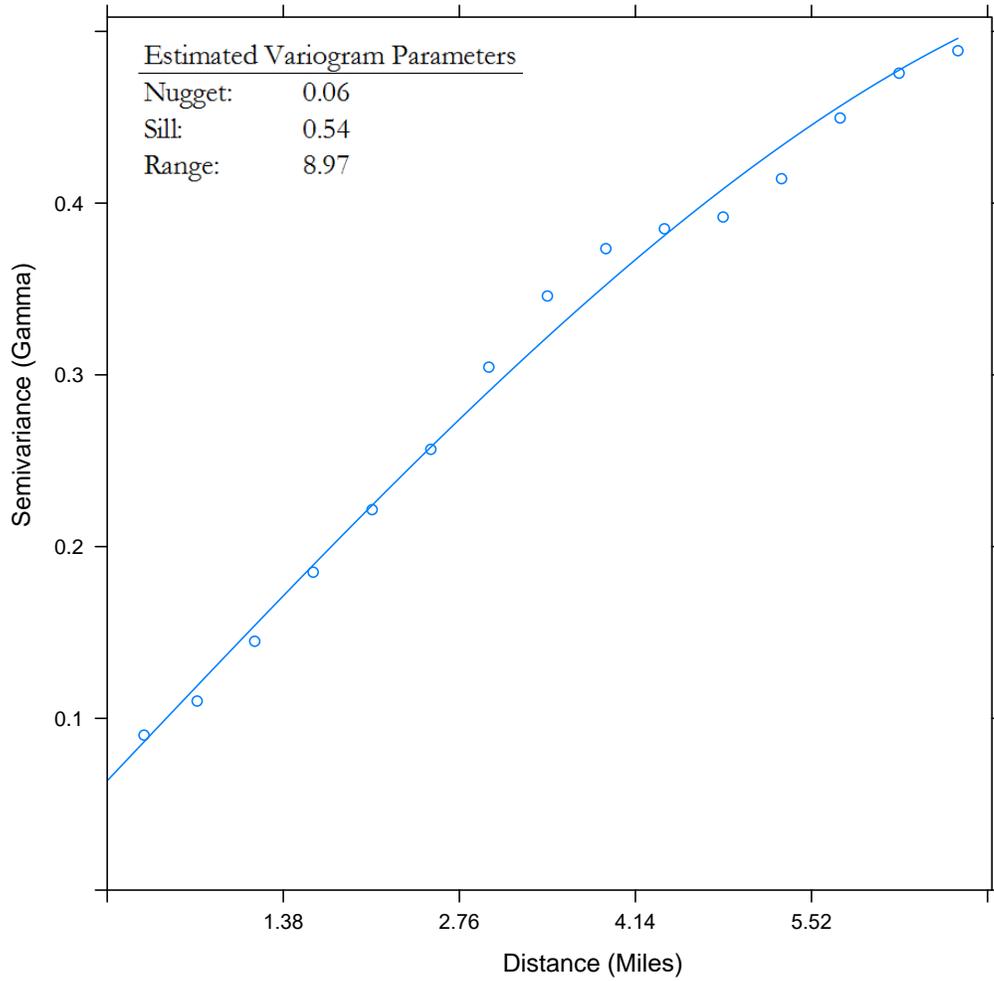
Notes: This figure shows change in average land-share by ZIP code between 2012 and 2018.

Figure 13: Change in Land Share by Quantiles of Housing Units by County, 2012-2018



Notes: The red bars show the change in the land share over 2012-2018 for counties sorted into 25 bins based on the number of single-family housing units from 2013-2017, measured at the county level in the 2013-2017 ACS (5-year sample). The blue dots show the aggregate value of land for all of the counties included in each bin. The change in the land share is the percentage point difference between 2018 and 2012. The red dashed line shows the increase in the land share of the aggregate United States between 2012 and 2018.

Figure A.1: Variogram for Washington, DC



Notes: This figure presents the binned semivariances (hollow circles) and fitted variogram (blue line) for the District of Columbia. The variogram here is estimated for all parcel pairs of extending to 6.9 miles.

Table 1: Interpolation RMSE (20% hold-out sample)

	Kriging	IDW	NN	Null-Tract	Null-ZIP Code	Null-County	Hold-Out Obs
2012	0.411	0.429	0.446	0.453	0.467	0.544	142,953
2013	0.399	0.419	0.431	0.445	0.462	0.543	219,706
2014	0.398	0.416	0.436	0.435	0.455	0.537	200,184
2015	0.390	0.409	0.429	0.431	0.452	0.535	276,177
2016	0.381	0.403	0.405	0.421	0.441	0.526	366,691
2017	0.379	0.401	0.406	0.421	0.439	0.522	329,654
2018	0.390	0.411	0.416	0.428	0.445	0.522	284,085

Notes: Interpolation RMSE calculated as follows. 1) Estimate an interpolated estimate for each hold-out parcel for each year. 2) Calculate an RMSE for each county for each year. 3) Calculate the median RMSE across counties (reported in table). IDW = inverse-distance weights, NN = nearest neighbor.

Table 2: Land Price Index Counts

	Counties	ZIP Codes	Census Tracts	Population	S.F. Housing Units
Balanced Annual Panel	1,132	9,194	19,252	87.3%	85.1%
Pooled Cross-Section	2,292	19,012	58,327	98.0%	97.3%

Notes: The county sample includes all counties with at least 50 admissible appraisals in the relevant temporal aggregation period (annual or pooled). The ZIP code and census tract samples are calculated as subsets of county-level parcel values so long as at least 10 admissible appraisals are available in each temporal aggregation period. When ZIP codes cross county boundaries, the ZIP code index is based on the simple average of all parcels with estimated prices in counties with a calculated index. The population and single-family housing units percentage is the fraction represented by the county index coverage, according to 2013-2017 (5-year) ACS estimates.

Table 3: Land Statistics
As-Is Estimates of Land Value Reported Per Acre

Variable	Pooled Cross Section (2,292 Counties):							Avg.	Std. Dev.
	1st	10th	25th	50th	75th	90th	99th		
Land Value	\$9,574	\$19,173	\$29,561	\$49,777	\$92,560	\$188,454	\$1,180,700	\$135,163	\$1,156,258
Land Share	7.0%	11.3%	14.1%	18.3%	24.1%	32.3%	53.5%	20.3%	9.1%

Variable	Annual Panel, Pooled (1,132 Counties):							Avg.	Std. Dev.
	1st	10th	25th	50th	75th	90th	99th		
Land Value	\$18,425	\$31,782	\$48,214	\$79,773	\$146,560	\$298,008	\$2,199,363	\$177,561	\$477,584
Land Share	10.5%	14.1%	16.9%	21.4%	27.7%	37.3%	56.6%	23.6%	9.6%

Table A.1: Interpolation RMSE (20% hold-out sample), alternative parameterizations

RMSE	Mean	Median	SD
Null - County Average	0.517	0.580	0.225
NN - 20 NN	0.418	0.454	0.161
IDW - 20 NN	0.413	0.456	0.175
Kriging - 10 NN, 6.9 Mile Boundary	0.394	0.429	0.157
Kriging - 20 NN, 6.9 Mile Boundary	0.390	0.426	0.156
Kriging - 30 NN, 6.9 Mile Boundary	0.389	0.426	0.156
Kriging - 20 NN, 3.4 Mile Boundary	0.391	0.426	0.157
Kriging - 20 NN, 10.4 Mile Boundary	0.392	0.427	0.157

Notes: Sample is the pooled cross-section (2292 counties). Interpolation RMSE calculated as follows. 1) Estimate an interpolated estimate for each hold-out parcel. 2) calculate an RMSE for each county for each year. 3) Calculate the median/mean/SD RMSE across counties (reported in table).

Table A.2: Predictions of Calibrated Urban Model

d	q_d^h	h_d	$q_d^h h_d$	s	q_d^l	l_d	$q_d^l l_d$	land share	l_d (acres)	q_d^l per acre
0	1.000	1,000,000	\$1,000,000	\$464,159	0.413	1,295,995	\$535,841	54%	0.25	\$2,143,364
1	0.922	1,062,482	\$980,000	\$480,054	0.354	1,411,219	\$499,946	51%	0.27	\$1,836,505
2	0.849	1,130,281	\$960,000	\$496,838	0.300	1,543,749	\$463,162	48%	0.30	\$1,555,320
3	0.781	1,203,972	\$940,000	\$514,581	0.251	1,697,826	\$425,419	45%	0.33	\$1,298,934
4	0.716	1,284,211	\$920,000	\$533,360	0.206	1,879,309	\$386,640	42%	0.36	\$1,066,527
5	0.656	1,371,742	\$900,000	\$553,260	0.165	2,096,587	\$346,740	39%	0.40	\$857,343
6	0.600	1,467,412	\$880,000	\$574,375	0.129	2,362,219	\$305,625	35%	0.46	\$670,706
7	0.547	1,572,189	\$860,000	\$596,810	0.098	2,696,126	\$263,190	31%	0.52	\$506,049
8	0.498	1,687,183	\$840,000	\$620,680	0.070	3,132,449	\$219,320	26%	0.60	\$362,959
9	0.452	1,813,671	\$820,000	\$646,116	0.047	3,736,426	\$173,884	21%	0.72	\$241,250
10	0.410	1,953,125	\$800,000	\$673,261	0.027	4,655,227	\$126,739	16%	0.90	\$141,134

Table A.3: Land price per acre, as predicted by the model and estimated via Kriging from two data sets

Distance	From Model	From Kriging Procedure			
		Data Set 1		Data Set 2	
		Predicted	% Error	Predicted	% Error
0	\$2,143,364	\$2,137,352	0.28%	\$2,107,724	1.70%
1	\$1,836,505	\$1,836,186	0.02%	\$1,762,402	4.00%
2	\$1,555,320	\$1,555,104	0.01%	\$1,575,406	-1.30%
3	\$1,298,934	\$1,298,750	0.01%	\$1,276,002	1.80%
4	\$1,066,527	\$1,066,499	0.00%	\$998,657	6.40%
5	\$857,343	\$857,259	0.01%	\$878,682	-2.50%
6	\$670,706	\$670,688	0.00%	\$641,687	4.30%
7	\$506,049	\$506,015	0.01%	\$450,198	11.00%
8	\$362,959	\$362,945	0.00%	\$364,878	-0.50%
9	\$241,250	\$241,259	0.00%	\$200,874	16.70%
Mean			0.03%		4.16%