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
January 2019

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
January 2019

Abstract
We use data on the appraised land value from a data set of more than 16 million appraisals to produce annual estimates of the average price of land used in single-family housing. We generate a balanced panel of data on land value per acre for more than 900 counties, 8,000 ZIP codes, and 11,000 census tracts in the United States between 2012 and 2017, covering 85% of the population and 83% of all single-family homes. Based on the predictions of a calibrated optimal model of teardowns, we limit our sample of appraised land values for homes with an effective age of no more than 10 years. We standardize these land values to a per-acre basis, and then spatially interpolate to the universe of single-family homes using Kriging. Our analysis produces three main findings for our sample: land prices in most areas increased between 2012 and 2017; land prices tended to rise faster than house prices; and land appreciated most rapidly in areas with relatively high density of structures. The land prices 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.

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

JEL Classification: R14, R21, R32

1These indices are works in progress and all data, tables, figures, and other results in this working paper are subject to change.
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 currently in residential use is now well understood, until recently 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 land in current residential use. Both of these methods require data that have been, until recently, 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 land values for all 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 964 counties, 8,344 ZIP codes, and 11,494 census tracts over the 2012 through 2017 period. Land prices are estimated for

2See Nichols, Oliner, and Mulhall (2013) and Davis, Oliner, Pinto, and Bokka (2017) for additional references in this literature.
areas representing more than 85% of the U.S. population and 83% of all single-family homes homes.\(^3\) To our knowledge, ours is the first study to produce these estimates at a fine geography for nearly the entirety of the United States.\(^4\) Our source data are the Uniform Residential Appraisal Report submissions to the Government Sponsored Enterprises (GSEs), Fannie Mae and Freddie Mac. These reports are required by the GSEs before they guarantee a mortgage against default. These data contain more than 16 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.\(^5\) 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 value of housing. This is not true when the market value of the structure is below its replacement cost, as would be the case when a housing structure has become functionally obsolete and is due to be torn down or extensively remodeled, or when housing demand has fallen dramatically and construction activity in the area has ceased (see Glaeser and Gyourko, 2005). 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 at least the first 10 years of the life of the structure. We therefore limit our data to housing units with an “effective age” of no more than 10 years. This filter eliminates about half 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.

For each estimate of land value in these data, we use the procedure described in Davis, Oliner, Pinto, and Bokka (2017) to adjust for the effect of lot size on land prices (the so-called “plattage effect”) and then compute the land price per acre. Then, we use a procedure called Kriging as described by Basu and Thibodeau (1998) to interpolate land price per acre

---

\(^3\)We also create a pooled cross-section of land prices for many more localities than the panel. This database includes land prices for 1,758 counties, 15,450 ZIP codes, and 38,539 census tracts. The pooled cross section also includes partial-year data for 2018 that are excluded from the annual panel dataset.

\(^4\)Alouy, 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 only 212 direct land sales per PMSA.

\(^5\)We exclude data on vacant land sales.
for every single-family housing unit in a given geography (county, ZIP code, or census tract) that does not have a GSE appraisal report. 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 using data simulated from the model.

The primary goal of the paper is to generate land prices indices covering most of the United States at a fine geography for use by researchers and policy-makers. To that end, the aggregate land price data presented 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. The data include land price per acre by year, county, ZIP code, and census tract.

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 2.5 million single-family housing units is more than 20 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; the rate of decline depends on the size of the metro area, the amount of regulation and the nature of the area’s topography. The price of land covaries with certain variables in accordance with what standard urban models predict. In particular, measured at the average price per acre in a county, residential land prices are negatively correlated with lot size and are positively correlated with floor-area ratios, farm land prices, and other land uses in the area.

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.

---

6We obtain the universe of single-family housing units in a given geography from the “Assessor Data” licensed from CoreLogic. The CoreLogic data contain the near-universe of parcels in the counties for which it has assembled data.
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

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

where $q^H$ is the rental price per unit of housing service provided and $S^{1-\phi} L^\phi$ is the number of units of such housing service produced under a Cobb-Douglas production function.

Each period, the land owner must decide whether or not to demolish the building and rebuild on vacant land, or to let the building deteriorate further 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 and make the land vacant. This problem has the expression

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

$\beta$ is the factor by which the land owner discounts the future and $\delta$ 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:

$$V(0) = \max_S \left\{ q^H S^{1-\phi} L^\phi - p^S S + \beta V(S(1-\delta)) \right\}$$
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,\(^7\) 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, $T = \left( \log \bar{S} - \log \underline{S} \right) [\log (1 - \delta)]^{-1}$.

We solve this model and calibrate it as follows. We set the discount rate to $\beta = 0.90$; the annual depreciation rate to $\delta = 0.023$; 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 that the optimal house built on vacant land is $\bar{S} = 70$.

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 at 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 at 80 years. When the point at which 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, $V(0) / V(\bar{S})$. 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, $V(\bar{S}) - S$. This is meant to approximate how land is measured residually when given an appraised value of housing, $V(\bar{S})$, and an estimate of the depreciated reconstruction

---

\(^7\)This is the argmax of equation (3).
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 sample restrictions on the set of homes we use to estimate land values.

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. 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, 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 uses the discounted flow of imputed rental income to arrive at an estimated value of the property. Finally, the “cost” approach attempts to estimate the cost of the components of the property, the land and the structure, giving the estimated value of the property as the sum. We use cost-approach appraisals in our analysis. In the cost approach, our understanding based on conversations with industry experts is that appraisers typically use the residual method to estimate the land component: land value is set equal to an $a\ priori$ estimate of the value of the property less the depreciated replacement cost of the structure component.

Our data on cost-approach appraisals are from Uniform Residential Appraisal Report submissions as collected by the GSEs. After data cleaning, including the removal of duplicate and/or resubmitted appraisals, we have approximately 16.4 million unique cost-approach
appraisal records submitted between 2012 and 2018. 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 restrict the sample of homes in our study based on reported effective age. A standard definition for effective age is

\[
\text{Economic Life} \times \left( \frac{\text{New Replacement Cost} - \text{Depreciated Structure Value}}{\text{New Replacement Cost}} \right)
\]

(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

\[
80 \times \left( \frac{50}{150} \right) = 26.7 \text{ years}.
\]

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

\[
80 \times \left[ 1 - (1 - 0.023)^{20} \right] = 29.8 \text{ years}
\]

(5)

In our analysis, we conservatively restrict our data to appraisals reporting an effective life of 10 years or less, corresponding to an age of 6 years in the calibrated model. Our data contains 7.8 million cost-approach appraisals with an appraised land value, lot size, and a structure with an effective age of no more than 10 years.

---

8We exclude appraisal records with 1) lots smaller than 500 square feet or larger than 2 acres; 2) property value missing or 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 10; 9) construction date before 1850 or after 2018.

9The 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).
4 Standardization and Interpolation of Land Values

To create a value for land within a county or subaggregate, we use a two-step procedure. First, we standardize all land value estimates in the data to a per-acre basis. Then, we assign to all single-family parcels in the county a value of land using an interpolation algorithm. We report the average value of land per acre, averaged across all single-family parcels, for each desired geography within the county, including the county itself, ZIP codes, and census tracts.\(^{10}\)

To standardize all estimates of land value to a per-acre basis, we first correct for plattage effects using a procedure similar to Davis, Oliner, Pinto, and Bokka (2017). For each county, we pool all data in all years and regress the log of the value of land on the log of lot size including ZIP code fixed effects and year dummies. Denote the coefficient on the log of lot size in this pooled regression as $\beta$. For each parcel in our sample, we then compute the predicted log price per acre as the observed price plus our estimate of $\beta$ times the difference of the log of one acre and the log of the size of the lot.

We then assign a value of land to each single-family property in the county that lacks a GSE appraisal by using an interpolation procedure called Kriging. Before describing Kriging, note that we merge our data set of appraised properties to the universe of single-family parcels from the “Assessor Data” licensed from CoreLogic. The CoreLogic data contains the near-census of parcels in the counties for which it has acquired rights to the data.

Kriging 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 include millions of parcels. Kriging satisfies these requirements.\(^{11}\) 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 in this section, we compare Kriging to a number

\(^{10}\)When ZIP codes span multiple counties, the reported value is the average of values in each represented county, weighted by the single-family housing stock share.

\(^{11}\)Basu 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.
of alternatives and show that Kriging produces more accurate estimates in a 20% holdout sample.

In Appendix A, we discuss the Kriging procedure in detail. Here we provide a brief summary. Before we do so, note that the bottom line 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, 5 miles in our case. 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 3 provides a look at the output of the Kriging procedure for the District of Columbia. The red dots show the location of observations in our sample located inside the city boundaries. The figure itself shows a heat map of spatially interpolated (log) land prices per acre. Figure 4 shows the mean of (log) land prices by ZIP code. The mean of each ZIP code is computed using the estimated land prices for properties with a GSE appraisal and the interpolated land prices for properties without an appraisal.

We evaluate the accuracy of the Kriging procedure by omitting a randomly selected 20% of our sample and determining the root mean square error (RMSE) of the Kriging interpolation procedure for that 20%. In every year, we compute the RMSE for each county in our sample. Across counties, the median RMSE is between 51 and 54 percent. This might seem high, but it is in fact well within our priors. Denote true land, structure and house value for a

---

12 Our data are partitioned by county (by year). This implies that the Kriging procedure never considers any pairwise points where one of the points is not in the county. Restated, the 5-mile cutoff will not bind in any county where the maximum distance between two locations in that county is less than 5 miles.
given parcel as \( L^* \), \( S^* \) and \( H^* \) such that

\[
L^* = H^* - S^*
\]

(6)

Suppose that structures are not measured with error but housing is measured with multiplicative error \( e \) such that observed house value \( H^o \) is equal to \( H^* (1 + e) \). We can then write an expression for the percentage deviation of observed land value \( L^o = H^o - S^* \) from the truth as

\[
\frac{L^o - L^*}{L^*} = \left( \frac{H^*}{L^*} \right) e
\]

(7)

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. Case and Shiller (1989) estimate the residual standard deviation to be about 15% for housing in repeat-sales of individual properties. Given land’s share of house value is about 30%, this suggests the standard deviation of measurement error to land prices should be about 50%.

We compare the median RMSEs 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 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.

We also evaluate the accuracy of the Kriging procedure for our application by seeing if it can recover land values that are the outcome of a simple rendition of the standard monocentric city model that we can compute analytically.\(^{13}\) 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. We ask if Kriging can exactly replicate the analytic gradient of land prices in the model. In the data set in which land values are perfectly measured, Kriging nearly exactly

\(^{13}\)See Alonso (1964), Mills (1967), Muth (1969), and Brueckner (1987).
replicates the analytic gradient. In the data set in which land values are measured with error, Kriging produces relatively low average errors. For more details, see appendix B.

5 Results

Table 2 presents observation counts for the calculated land price series. We are able to calculate land prices for 964 counties, 8,344 ZIP codes, and 11,494 census tracts in our balanced panel between 2012 and 2017. Our pooled cross-section, calculated using data pooled between 2012 and 2018 and re-based to 2013 using the year dummy coefficients from the county-level regressions described above, includes data for 1,758 counties, 15,450 ZIP codes, and 38,539 census tracts.\(^\text{14}\) The annual and pooled samples cover 85\% and 94\% of the U.S. population residing in the 50 states plus the District of Columbia, respectively, and 83\% and 93\% of the single-family housing units.

In this section, we present stylized facts related to the land-price data. For expositional purposes, we present pooled estimates first in order to validate our data in terms of known relations 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 in turn: 1) the land price gradient; 2) spatial variation in housing-structure density; 3) spatial variation in land use; 4) dynamics of land prices; and 5) dynamics relating housing and land prices.

5.1. The Land Price Gradient

The traditional monocentric city model predicts land prices fall with distance to the 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 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

\(^{14}\)As noted in the introduction, we had only partial data for 2018 when the empirical work for the paper was completed. 2018 data are therefore usable in a pooled context but not in a panel.

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 center ZIP codes, shown in grey on the maps, there are not enough single-family housing unit transactions to construct an index. Land prices where single-family units are plentiful enough to construct an index start at about $10 to $30 million per acre in the most expensive ZIP codes of both cities. By about 10 miles from the CBD, the gradients are mostly flat, though there is still some slope in the San Francisco CBSA.

These relations between land prices and proximity to the CBD are not unique to these two cities. Figure 6, panel (a) pools land prices and proximities for all ZIP codes within city population ranges and shows that average land prices fall with distance to the CBD. The radius at which the gradient flattens may be usable as a metric of what defines the effective boundaries of the city: When the gradient flattens, the value of the locations in the flat sections are no longer governed primarily by the forces described by the traditional monocentric city model and perhaps another model is necessary.

The monocentric city model also predicts that land-supply restrictions have a negative effect on land prices holding population and amenity constant. This occurs because land supply restrictions have a negative effect on the spatial efficiency of the city which becomes capitalized into land prices. But such restrictions also create amenities that may exceed the efficiency loss, with the net effect being a reduced-form positive correlation between land restrictions on the one hand, and house and land prices on the other (see Saks, 2008, Saiz, 2010, and Davidoff, 2016). Figure 6, panels (b) and (c) show land-price gradients by population and top and bottom halves of regulatory burden (Gyourko, Saiz, and Summers, 2008) and topographic interruptions (Saiz, 2010), respectively. These panels show supply restrictions to be positively correlated with the levels of the land price but not the slopes of the gradients.

5.2. Land Prices, Housing Structure Density, and the Reservation Use

The traditional production function for housing includes structure and land inputs. Because of the downward-sloping land price gradient, combined with constant marginal structure production costs, housing producers shift towards land-intensive housing production far from the CBD where the price of land is low. Our master assessor dataset contains 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 each area covered by our land data.

Panel (a) of Figure 7 graphs FARs for single-family housing against land prices. 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) and the most expensive land containing FARs near 1 (i.e. a 3 story, 2,400 square foot row-house on a 2,400 square foot lot). Holding structure constant, lower structure density is associated with larger lots. Accordingly, we see in panel (b) that lot sizes decrease monotonically with land prices.

As land prices fall and lot sizes increase, the land use approaches its reservation use. Agricultural production occurs at the edge of most cities. Because land use is governed by the bid-rent curves for various sectors, and low-density residential land use is typically the dominant developed use at the edge of the city, the single-family land price for ZIP codes at the edge of the city should be closely linked to the agricultural land price. Panel (c) shows the agricultural land price per acre, as measured by the U.S. Department of Agriculture, is highly correlated with single-family residential land prices.

5.3. Variation in Land Use

Land prices and land use are also correlated in the neoclassical model. Each economic sector has its own production function, and these production functions include land as an input directly (e.g. agriculture) and/or indirectly (interior space, e.g. offices, retail, manufacturing, and housing). Profit maximization results in a marginal willingness to pay for land based on its proximity to the CBD, referred commonly in urban economics as a “bid-rent” curve. In part because different sectors have different bid-rent curves, sectors sort into regions of the city. Traditionally, agriculture is the most land intensive and thus occupies the cheapest land far from the CBD. Conversely, office space is the least land intensive and occupies the densest structures in the CBD to take advantage of agglomeration externalities. In practice, at each distance from the CBD, there is a mixture of land uses within a particular area, though land use tends to follow the theoretical predictions.

Our land prices are for single-family housing units, yet due to the spatial equilibrium condition, these should be highly correlated with land use types for other sectors. The panels in Figure 8 show the fraction of the land area in respective sectors as a function of the single-family land prices we have estimated. Land use classifications are generated by mapping the
CoreLogic land use codes to create the aggregate categories presented in the figures.

Panel (a) shows the three residential land use types as a function of the single-family land price per acre. All three residential land use types increase in share with the land price up to about $5 million per acre, at which point, single-family detached land use declines while the other two continue increasing. Panel (b) shows industrial and agricultural land use as a function of the single-family land price per acre. Agricultural land use monotonically declines with the single-family land price, highlighting the reservation use of agricultural land in urban areas. Conversely, industrial land use almost monotonically increases with the land price. Panel (c) presents three different commercial land use types, retail, office, and other. Retail and office land use rise with the single-family land price until about $15 million per acre, at which point, office increases dramatically, mainly at the expense of retail. Panel (d) shows vacant structure and vacant land use. Both are declining as a function of the single-family land price, reflecting the increasing opportunity cost of leaving a potentially productive asset in an unproductive state.

5.4. Land Price Dynamics

Bogin, Doerner, and Larson (2018) document patterns of house price appreciation in different parts of the city and across cities of various sizes and types. Based on the insights of Davis and Palumbo (2008) and other papers mentioned earlier, this should reflect different rates of appreciation of land prices after controlling for the value of housing that is attributable to the replacement cost of structures.

Why does the price of land increase at different rates in different places? Table 3 shows the results of four models for land-price changes over 2012-2017 estimated using ordinary least squares. Model 1 contains CBSA-specific fixed effects and serves the purpose of focusing on within-city land price dynamics. Models 2 through 4 do not contain CBSA fixed effects and therefore are able to capture partial correlations across cities based on city-level attributes.

Model 1 shows the FAR and initial average house value to be positively associated with land price appreciation, whereas the distance to the CBD is negatively associated with land appreciation. Combined, these results indicate that areas near the CBD have land prices that have been appreciating faster than areas far from the CBD, holding city price levels constant.
Model 2 evaluates the effect of city size on land price appreciation, with appreciation positively related. Model 3 estimates the appreciation gradient as it relates to city size. The coefficient on the distance variable is 0.0036 and the coefficient on the distance × housing units (log) variable is -0.00028, giving a flat gradient in a city with 340,000 housing units, and negatively (positively) sloped appreciation gradients in larger (smaller) cities. Model 4 shows each of the previous two models’ results to be robust, while also demonstrating housing market regulation to be positively correlated with land price appreciation. In sum, these results echo the findings of Bogin, Doerner, and Larson (2018). The recent data suggest land prices are appreciating in much the same manner as house prices—fastest near the centers of large cities.

5.5. House and Land Price Dynamics

The stylized fact that land and house prices have been rising together should not come as a surprise, as land prices and house prices are linked through the housing production function. Previous studies have shown land prices to be more volatile than the price of housing, most notably in the land-leverage literature (e.g. Bostic, Longhofer, and Redfearn, 2007, and Davis and Heathcote, 2007).

This relationship holds in our data as well. Figure 9 plots the average annual growth rate of land prices (horizontal axis) against the average annual growth rate of house prices from Bogin, Doerner, and Larson (2018) (vertical axis) between 2012 and 2017, a period of recovery from the housing bust. The 45 degree line shows cases where growth rates are identical. The figure shows the range of changes to land prices is larger than that of house prices. A simple OLS regression relating house and land price appreciation gives an elasticity of 0.44 over this sample, indicating a 10% change in land prices is associated with a 4.4% change in house prices.\footnote{The equation $\Delta \text{House Prices}_i = \alpha + \beta \Delta \text{Land Prices}_i + \epsilon_i$ is estimated over $N = 8,154$ counties indexed by $i$, with changes denoting the log-difference between 2017 and 2012 values. Estimates give $\hat{\alpha} = 0.026(0.001)$, $\hat{\beta} = 0.440(0.014)$, and $R^2 = 0.56$, with heteroskedasticity-robust standard errors in parentheses.}

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 2017 of the average price of land per acre used in single-
family homes for 964 counties, 8,344 ZIP codes, and 11,494 census tracts. We also calculate pooled cross-sectional estimates of land price per acre for 1,758 counties, 15,450 ZIP codes, and 38,539 census tracts to facilitate analysis of gradients and long-run, location-specific spatial variation. In our work, we use a new model of the option to tear-down and rebuild a house to guide the maximum effective age of the housing stock with which to calculate unbiased estimates of land prices. For properties without appraisal data, we interpolate land prices using a Kriging procedure, provides more accurate estimates than other interpolation methods.

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


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 county into those that are sampled $N^s$ and unsampled $N^u$. All indices 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 indices within a geography arise due the method used to estimate prices for unsampled parcels.

The price of an unsampled parcel is estimated as a weighted average of prices of a subset of observed parcels within the county. The estimated price for parcel $i$ is calculated as the average of the $n$ nearest (by proximity) neighbor parcels, indexed by $j$, with weights $\lambda_{i,j}$.

For a particular method, the estimated price for parcel $i$ is:

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

(8)

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.

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

(9)

---

16So $p_{i,1}$ is the closest observed price to location $i$; $p_{i,2}$ is the second closest observed price to location $i$; and so on.
Null Estimator
The null estimator ("Null") sets $n=N^*$ 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 $\lambda$ 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. $\lambda$ 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.\textsuperscript{17}

As an illustration of this procedure, we present kriging steps for land prices in Washington, DC, pooled between 2012 and 2017. There are about 122,000 parcels and 4,000 sampled standardized land prices. To start, differences and semivariances $\gamma$ are calculated for each pair of points in $N$, resulting in over 9 million pairwise combinations that are within 10 miles

\textsuperscript{17}For a more in-depth overview of kriging, see Hengl (2007).
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. Exact values are shown in the table, including the distance, the number of pairs, and the average semivariance. Distances are reported in terms of distance in latitude/longitude degrees. The bin at a distance of 0.05 therefore corresponds to a distance of 3.11 miles, with an average semivariance of 1.33 over 992,298 pairwise semivariances. The average semivariance of 1.33 corresponds to an average variance of 2.66, or an average standard deviation of 1.63, in terms of the log land price per acre, for any hypothetical pair of points with the distance between them referenced by the bin.

In step three, a functional form for the relationship between the seminvariances 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

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

(10)

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.

---

18 The 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, $h$.

19 In 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 Euclidean distance.

20 The 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.
In addition to the binned semivariances, figure A.1 also shows a fitted spherical functional form for Washington, DC, with \( \hat{a}_0 = 0.53, \hat{a}_1 = 0.89 \) and \( \hat{r} = 4.35 \) miles. This function is used to estimate the semivariance between any two hypothetical points, facilitating interpolation of prices in unsampled locations.\(^{21}\)

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:

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

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 \( \lambda_i \), an \( n \times 1 \) vector.

\[
\begin{bmatrix}
\lambda_i \\
\mathcal{L}
\end{bmatrix} = 
\begin{bmatrix}
C_i & 1 \\
1' & 0
\end{bmatrix}^{-1}
\begin{bmatrix}
c_i \\
1
\end{bmatrix}
\]

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 1758 counties in the pooled sample. The mean RMSE for the entire sample of observations is 0.771, representing the average variation around the county mean. This is similar to the mean of the training sample, which gives a mean RMSE of 0.767 when compared to the 20% of the observations in the hold-out sample. Nearest neighbor and

\(^{21}\) Other 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.

inverse-distance weights give increasing accuracy, with mean RMSEs falling to 0.569 and 0.522, respectively.

Each of the kriging estimates gives similar average fit, with the mean RMSE either 0.497 or 0.496. We interpret this finding as indicating that the boundary is rarely binding for finding nearest neighbors, and has little effect on the variogram. Overall, we interpret these findings as lending support to our decision to use the 20 nearest neighbors and a 5 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

\[
(1 - \alpha) \ln c + \alpha \ln h
\]

(13)

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 \( q^h_d \). A person living \( d \) units from the city center faces the budget constraint of:

\[
w (1 - td) = c + q^h_d h
\]

(14)

A person choosing to live in location \( d \) units away from the CBD maximizes utility (13)
subject to the budget constraint (14) by choosing optimal consumption \( c_d \) and housing \( h_d \) of

\[
c_d = (1 - \alpha) w (1 - td) \tag{15}
\]

\[
q_d^h h_d = \alpha w (1 - td) \tag{16}
\]

This means maximized utility at distance \( d \) from the center can be written as \( U_d \)

\[
U_d = (1 - \alpha) \ln [(1 - \alpha) w (1 - td)] + \alpha \ln \left[ \alpha w (1 - td) / q_d^h \right] 
\]

\[
= \kappa_u + \ln w + \ln (1 - td) - \alpha \ln q_d^h \tag{17}
\]

where \( \kappa_u \) 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

\[
U_d = U_{d'}
\]

Then from equation (17) this implies

\[
\ln (1 - td) - \alpha \ln q_d = \ln (1 - td') - \alpha \ln q_{d'}^h 
\]

\[
q_{d'}^h / q_d^h = \left( \frac{1 - td'}{1 - td} \right)^{\frac{1}{\alpha}} \tag{18}
\]

Equation (18) governs the rate at which housing rental prices per unit change with distance from the CBD, roughly \( t/\alpha \) 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

\[
U_d = (1 - \alpha) \ln [(1 - \alpha) w (1 - td)] + \alpha \ln h_d \tag{19}
\]

Once we impose \( U_d = U_{d'} \), this gives us

\[
(1 - \alpha) \ln (1 - td) + \alpha \ln h_d = (1 - \alpha) \ln (1 - td') + \alpha \ln h_{d'} 
\]

\[
\rightarrow \frac{h_{d'}}{h_d} = \left( \frac{1 - td'}{1 - td} \right)^{\frac{1}{1-\alpha}} \tag{20}
\]
Now that we have worked out how housing quantities $h$ and prices per unit $q_h$ 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

$$ h = \left[ (1 - \theta) s^\rho + \theta l^\rho \right]^{\frac{1}{\rho}} \quad (21) $$

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

$$ q_h \left[ (1 - \theta) s^\rho + \theta l^\rho \right]^{\frac{1}{\rho}} - s - q_l l \quad (22) $$

The first-order conditions for optimal structures are

$$ 1 = q_h h^{1-p} (1 - \theta) s^{p-1} \quad (23) $$

$$ \rightarrow s = \left[ q_h (1 - \theta) \right]^{\frac{1}{1-p}} h \quad (24) $$

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:

$$ q_l l = q_h h^{1-p} \theta l^p \quad (25) $$

and thus

$$ l = \left[ \frac{q_l l}{q_h h^{1-p} \theta} \right]^{\frac{1}{p}} \quad (26) $$

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.\footnote{This is referring to single-family homes.} 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 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.\textsuperscript{23} This simulation gives us some intuition for how the Kriging procedure performs under conditions where land is imperfectly measured. Denote $q^h_d h_d$ as observed housing value and $s$ as observed structures costs. $q^h_d h_d$ and $s$ are determined as

$$
\begin{align*}
\tilde{q}^h_d h_d &= q^h_d h_d (1 + e^h_d) \\
\tilde{s} &= s (1 + e^s_d)
\end{align*}
$$

with $e^h_d$ and $e^s_d$ drawn independently. We then compute observed land value residually,

$$
\begin{align*}
\tilde{q}^l_d l_d &= \tilde{q}^h_d h_d - \tilde{s} \\
&= q^h_d h_d - s + (q^h_d h_d e^h_d - se^s_d)
\end{align*}
$$

Denote the term in parentheses as $e^l_d$, measurement error in land value. Even though the standard deviation of $e^h_d$ and $e^s_d$ are relatively small (5.7 percent each), the standard deviation of measurement error as a percent of true land value, measured as $e^l_d / (q^l_d l_d)$, 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

\textsuperscript{23}This measurement error can be thought of as a deviation from model-determined prices.

---

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 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 without error, Kriging nearly perfectly replicates the analytic results. When the data are measured with error, the Kriging results are less accurate. The average error is 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

- Correct: $\frac{V(0)}{V(S)}$
- Incorrect: $\frac{[V(S)-S]}{V(S)}$

\[ \text{Share (\%)} \]
\[ \text{Structure Age} \]
Figure 3: Washington, DC Kriging Example
Figure 4: Washington, DC Kriging Example
Figure 5: Selected Standardized Land Prices, pooled cross-section (2013 prices)

(a) Washington

(b) San Francisco

Notes: Samples for maps include all ZIP codes with centroids within 10 miles of an identified central ZIP code centroid. Graphs include the same sample extended to 25 miles. The line is fit using a local polynomial smoother and the confidence intervals represent the 5th/95th percentile estimates.
Figure 6: Selected Standardized Land Prices, pooled cross-section (2013 prices)

(a) All Cities

(b) High vs Low Regulation

(c) High vs Low Topographic Interruptions

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 represent the 5th/95th percentile estimates.
Figure 7: Land Prices and Covariates

(a) Floor-Area Ratios and Land Prices

(b) Lot Size and Land Prices

(c) Agricultural and Single-Family Housing 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.
Figure 8: Single-Family Land Prices and Sector Land Use

(a) Residential  
(b) Industrial  
(c) Commercial  
(d) Vacant Land and Structures

Notes: This figure presents ZIP code land use shares as a function of the residential land price from the pooled cross-section. Land use is calculated using land areas in the CoreLogic assessor database. The curve is a local polynomial with the scatterplot omitted for clarity.
Figure 9: House Price and Land Price Changes, 2012-2017

Notes: This figure shows average annual county-level appreciation rates for land and housing prices. The sample is the 964 counties in the panel sample. The curve is a quadratic polynomial.
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 10 miles.
Table 1: Interpolation RMSE (20% hold-out sample)

<table>
<thead>
<tr>
<th></th>
<th>Kriging</th>
<th>IDW</th>
<th>NN</th>
<th>Null-Tract</th>
<th>Null-ZIP Code</th>
<th>Null-County</th>
<th>Hold-Out Obs</th>
</tr>
</thead>
<tbody>
<tr>
<td>2012</td>
<td>0.545</td>
<td>0.558</td>
<td>0.622</td>
<td>0.666</td>
<td>0.659</td>
<td>0.768</td>
<td>148,695</td>
</tr>
<tr>
<td>2013</td>
<td>0.534</td>
<td>0.549</td>
<td>0.609</td>
<td>0.673</td>
<td>0.663</td>
<td>0.764</td>
<td>207,337</td>
</tr>
<tr>
<td>2014</td>
<td>0.538</td>
<td>0.556</td>
<td>0.615</td>
<td>0.666</td>
<td>0.665</td>
<td>0.766</td>
<td>171,719</td>
</tr>
<tr>
<td>2015</td>
<td>0.522</td>
<td>0.540</td>
<td>0.605</td>
<td>0.664</td>
<td>0.661</td>
<td>0.767</td>
<td>224,380</td>
</tr>
<tr>
<td>2016</td>
<td>0.515</td>
<td>0.533</td>
<td>0.592</td>
<td>0.667</td>
<td>0.662</td>
<td>0.763</td>
<td>275,151</td>
</tr>
<tr>
<td>2017</td>
<td>0.530</td>
<td>0.546</td>
<td>0.607</td>
<td>0.682</td>
<td>0.667</td>
<td>0.766</td>
<td>225,927</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th>Counties</th>
<th>ZIP Codes</th>
<th>Census Tracts</th>
<th>Population</th>
<th>S.F. Housing Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balanced Annual Panel</td>
<td>964</td>
<td>8,344</td>
<td>11,494</td>
<td>85.8%</td>
<td>83.2%</td>
</tr>
<tr>
<td>Pooled Cross-Section</td>
<td>1,758</td>
<td>15,450</td>
<td>38,539</td>
<td>94.2%</td>
<td>92.6%</td>
</tr>
</tbody>
</table>

Notes: The county sample include all counties with at least 50 standardized land price observations within the period. The ZIP code and Census Tract samples are the subsets of the relevant geography with at least 10 standardized land price observations that exist within counties with a calculated index. 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 2012-2016 (5-year) ACS estimates.
Table 3: Land Price Change Correlates, 2012-2017

Dependent variable: \( \log(\text{land price 2017}) - \log(\text{land price 2012}) \)

<table>
<thead>
<tr>
<th></th>
<th>[1]</th>
<th>[2]</th>
<th>[3]</th>
<th>[4]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial land value share</td>
<td>-0.194***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.0361]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floor area - land area ratio</td>
<td>0.0412***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.00927]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Initial average house value</td>
<td>0.0169***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.00283]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance to the CBD (miles)</td>
<td>-0.000496***</td>
<td>0.00363***</td>
<td>0.00434***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[0.000126]</td>
<td>[0.000726]</td>
<td>[0.00105]</td>
<td></td>
</tr>
<tr>
<td>Housing units in CBSA (log)</td>
<td></td>
<td>0.00778**</td>
<td>0.0146***</td>
<td>0.0147***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.00351]</td>
<td>[0.00324]</td>
<td>[0.00447]</td>
</tr>
<tr>
<td>Units (log) × Distance</td>
<td></td>
<td>-0.000285***</td>
<td>-0.000335***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>[5.15e-05]</td>
<td>[7.18e-05]</td>
<td></td>
</tr>
<tr>
<td>WRLURI</td>
<td></td>
<td></td>
<td></td>
<td>0.00961**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>[0.00478]</td>
</tr>
<tr>
<td>Constant</td>
<td>-0.111***</td>
<td>-0.05</td>
<td>-0.136***</td>
<td>-0.139**</td>
</tr>
<tr>
<td></td>
<td>[0.0341]</td>
<td>[0.0438]</td>
<td>[0.0394]</td>
<td>[0.0566]</td>
</tr>
<tr>
<td>CBSA Fixed Effects</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Observations</td>
<td>7,922</td>
<td>8,338</td>
<td>8,272</td>
<td>7,062</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.647</td>
<td>0.038</td>
<td>0.059</td>
<td>0.073</td>
</tr>
</tbody>
</table>

Notes: Robust standard errors in brackets, clustered by CBSA. *** \( p < 0.01 \), ** \( p < 0.05 \), * \( p < 0.1 \). The WRLURI is the Wharton Residential Land Use Regulation Index.
Table A.1: Interpolation RMSE (20% hold-out sample), alternative parametrizations

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full-Sample RMSE</strong></td>
<td>0.771</td>
<td>0.805</td>
<td>0.222</td>
</tr>
<tr>
<td><strong>20% Hold-out Sample RMSE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Null - County Average</td>
<td>0.767</td>
<td>0.799</td>
<td>0.231</td>
</tr>
<tr>
<td>NN - 20 NN, 5 Mile Boundary</td>
<td>0.569</td>
<td>0.613</td>
<td>0.197</td>
</tr>
<tr>
<td>IDW - 20 NN, 5 Mile Boundary</td>
<td>0.522</td>
<td>0.563</td>
<td>0.181</td>
</tr>
<tr>
<td>Kriging - 15 NN, 5 Mile Boundary</td>
<td>0.497</td>
<td>0.536</td>
<td>0.174</td>
</tr>
<tr>
<td>Kriging - 20 NN, 5 Mile Boundary</td>
<td>0.497</td>
<td>0.536</td>
<td>0.175</td>
</tr>
<tr>
<td>Kriging - 25 NN, 5 Mile Boundary</td>
<td>0.496</td>
<td>0.537</td>
<td>0.177</td>
</tr>
<tr>
<td>Kriging - 20 NN, 2.5 Mile Boundary</td>
<td>0.497</td>
<td>0.540</td>
<td>0.180</td>
</tr>
<tr>
<td>Kriging - 20 NN, 10 Mile Boundary</td>
<td>0.497</td>
<td>0.536</td>
<td>0.170</td>
</tr>
</tbody>
</table>

Notes: Sample is the pooled cross-section (1758 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

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

<table>
<thead>
<tr>
<th>Distance</th>
<th>From Model</th>
<th>Data Set 1</th>
<th>Data Set 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Predicted</td>
<td>% Error</td>
</tr>
<tr>
<td>0</td>
<td>$2,143,364</td>
<td>$2,137,352</td>
<td>0.28%</td>
</tr>
<tr>
<td>1</td>
<td>$1,836,505</td>
<td>$1,836,186</td>
<td>0.02%</td>
</tr>
<tr>
<td>2</td>
<td>$1,555,320</td>
<td>$1,555,104</td>
<td>0.01%</td>
</tr>
<tr>
<td>3</td>
<td>$1,298,934</td>
<td>$1,298,750</td>
<td>0.01%</td>
</tr>
<tr>
<td>4</td>
<td>$1,066,527</td>
<td>$1,066,499</td>
<td>0.00%</td>
</tr>
<tr>
<td>5</td>
<td>$857,343</td>
<td>$857,259</td>
<td>0.01%</td>
</tr>
<tr>
<td>6</td>
<td>$670,706</td>
<td>$670,688</td>
<td>0.00%</td>
</tr>
<tr>
<td>7</td>
<td>$506,049</td>
<td>$506,015</td>
<td>0.01%</td>
</tr>
<tr>
<td>8</td>
<td>$362,959</td>
<td>$362,945</td>
<td>0.00%</td>
</tr>
<tr>
<td>9</td>
<td>$241,250</td>
<td>$241,259</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

| Mean     |            | 0.03%      | 4.16%      |