A Statistical Learning Approach to Land Valuation: Optimizing the Use of External Information

David Albouy* Minchul Shin†

September 19, 2022

Abstract

We develop a statistical learning model to estimate the value of vacant land for any parcel, regardless of improvements. Rooted in economic theory, the model optimizes how to combine common improved property sales with rare, but more informative, vacant land sales. It estimates how land values change with geography and other features, and determines how much information either vacant or improved sales provide to nearby areas through two levels of spatial correlation. For most neighborhoods, incorporating improved sales often doubles the certainty of land value estimates. Relative to conventional estimators, our method reduces problems from excess variance and sample selection.

JEL codes: C11, C43, R1, R3

Keywords: land values, hierarchical modeling, spatial data, Bayesian estimation

Some funding provided by the Lincoln Institute for Land Policy, Reference #: T24R001-TDA080819. We thank Alok Ranjan for valuable research assistance. Seminar participants at the Lincoln Institute for Land Policy, Jönköping University, and the University of Copenhagen provided valuable comments. The views expressed in this paper are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

*University of Illinois. Email: albouy@illinois.edu

†Federal Reserve Bank of Philadelphia. Email: minchul.shin@phil.frb.org

1 Introduction

Valuing land accurately and objectively is possibly the most formidable technical barrier to taxing land value (Mills, 1998). Sales of vacant land are rare, especially in built-up areas, where values are the highest. Furthermore, vacant properties may be difficult to compare, since land values can change considerably over short distances or periods of time (Ashley et al., 1999; Gloudemans et al., 2002; Bell and Bowman, 2006). On the other hand, improved, i.e., non-vacant, properties sell much more frequently, but are even more difficult to compare. Such properties pair land with improvements that range from a bungalow to a super-tall skyscraper.

Recent advances in data science and statistical methods may mitigate many challenges to valuing land when properly applied. Our goal here is to optimally combine small numbers of vacant property sales with large numbers of improved property sales to provide a credible prediction of what price any lot would be worth if it were vacant. In the process, we learn more about the economics of how much an improved property's value is due solely to its land.

The statistical learning approach we develop here is unique in how it values vacant and improved lots jointly, making use of their neighboring locations. This method can be applied in many settings as the data required are available to most assessors. Assessors commonly make subjective judgments as to what properties are comparable. In contrast, the statistical learning approach determines what lots are comparable more objectively, based on how well the value of vacant and improved properties predict each other's values throughout the data. The approach even accounts for idiosyncrasies that may make any one observed transaction price for a property less than fully representative of the typical price that a property would sell for, if it were sold repeatedly. These "shrinkage" methods are used to reduce the influence of outliers on estimates of ind[1]

Mathematically, our learning approach uses a multilevel Bayesian framework that constructs a posterior distribution of unknown parameters, assumed to have a diffuse and uninformative prior distribution. These parameters have a hierarchical structure. A large number of lower-order parameters describe a flexible land-value surface. These are generated from a conditional distribution, which depends on a smaller number of higher-order hyper-parameters. These hyper-parameters teach us more about the abstract qualities of land values, and how they interact with

[1] Angrist et al. (2017) use a Bayesian technique to jointly model experimental and non-experimental estimates of teacher-value added. Differences between the two are modeled using a model of student selection into schools.

improved property values. The parameters are estimated sequentially in a loop through an iterative standard Markov Chain Monte Carlo (MCMC) method.

This framework is practically useful for several reasons. First, the MCMC posterior simulator we develop — a variant of a Gibbs samplers — works quite well for a large model with over 900 parameters. Second, it offers a convenient yet coherent way to construct a full predictive distribution that accounts for parameter uncertainty. This allows one to characterize uncertainty around the land values predicted by the model. It also provides safeguards around issues of "overfitting" from using too little data to identify too many parameters. In a sense, land values are "shrunk" not only using vacant land observed further away, but also using land values one would predict from improved properties observed nearby.[2]

This flexible, non-parametric technique resembles others based on moving averages, kernel density, or Kriging.[3] What makes our model different from others is that it simultaneously models a vacant land-value function with an improved property-value function, estimating the two correlated non-parametric functions jointly. This joint estimation reduces the uncertainty of land-value estimates drawn from the data. In our example below, the extra information contained in the improved sales data reduces this uncertainty by over 50 percent.

2 An Econometric Model of Land Values

2.1 The underlying value of land

The econometric model centers around a vacant land value index r, which may be used to value any plot i. Fundamentally, this value should depends on location, lot characteristics, and legal and other institutional factors, which can affect development opportunities. The value of the location includes its proximity to places of work as well as the neighborhood amenities it provides access to. Lot characteristics may include its size, as well as dimensions. Legal and institutional factors include regulations such as zoning, approval processes, and the ability of third parties to

[2] The only other application of Bayesian methods to land values is that of Ecker and Isakson (2005). This uses Bayesian methods to estimate at what lot size the price-size function changes from convex to concave. Big data techniques are innovated by Davis et al. (2017)

[3] For an early example of semi-parametric techniques used to value land, see Thorsnes and McMillen (1998)

block development.[4] It should also depend on time, although we save this for a later extension in subsection 6.4.

To ease exposition, consider the following linear model, similar to Epple et al. (2010), where r is the logarithm of the price of land per square foot:

$$r_i = \delta_0^r + \delta_1^r d_i + \delta_2^r A_i + \eta_j^r \equiv (Z_i)'\delta^r + \eta_j^r \tag{1}$$

The term η_j represent "area effects," indicating a discrete area where the lot is located. In Epple et al. (2010), such areas refer to different municipalities. They may also represent geographically finer neighborhoods, such as census tracts, which we also consider.These effects may capture both institutional features, as well as other location characteristics described above. In the multilevel Bayesian framework, these area effects are not fixed effects, but are lower-order parameters drawn from an underlying stochastic process determined by the hyper-parameters.

In addition, d_i is a metric for distance to a central location, which is meant to capture continuous location effects within the discrete area. The term A_i contains the log acreage of the lot, along with other lot characteristics. Together, all of these fundamental determinants of land values other than the area effects are compiled into the vector Z_i.[5]

2.2 Information from transactions data

A key to understanding our statistical learning approach is that it assumes that every transaction measures the true underlying value imperfectly. Each sale contains a signal of the true value, obscured by noise. Let y_i^v be the observed transaction price of vacant lot i in logarithmic form. We assume this price is determined by

$$y_i^v = r_i + (X_i^v)'\beta^v + e_i^v \tag{2a}$$

The added covariates in (2a) are

[4] The relevance of such factors is considered in work by Kok et al. (2014) and Gyourko and Krimmel (2021). These characteristics may interact: when land is costlessly sub-dividable, the particular shape or size of a lot may matter less than when there are barriers to subdividing them.

[5] These observable variables may not map cleanly to the fundamental determinants of land values. Lot shapes could be determined by differing legal restrictions within a municipality; lots further from the city center may have services that make them more or less desirable. Note that while roughly two-thirds of census tracts in our sample lie completely within municipal boundaries, in general, they are not coterminous. Some tracts span multiple municipalities, although only 10 percent of tracts were less than 85 percent in a single municipality.

- X_i^v, potentially observable features of a vacant sale transaction that might affect the price, such as seasonality or whether it was brokered or auctioned;
- e_i^v, an error term which refers to all unobserved features of the transaction, including bargaining abilities of the buyer or seller, or measurement error in the record.

In the end, what matters is that the index should be better at predicting an out-of-sample transaction — e.g. an adjacent location or even a later one in the same location — than the actual transaction price we do observe. The transaction price of any comparable property is subject to various idiosyncrasies which may obscure the correct land value.

While r_i captures underlying land values, it is affected by how the transaction characteristics in X_i^v are recorded. Thus, it should be normalized to be zero to reflect the land values one wants to measure. For instance, a tax assessor may desire an arms-length, non-auction sale made in April. Adjustments should also be made so that e_i^v may be assumed to have a mean of zero, e.g. assuming the bargaining abilities are on average zero, as are recording errors. In principle, some variables in X_i^v and in Z_i might be exchanged, depending on what variables one wants to include in the valuation of land. One may want to include or exclude the presence of infrastructure or the (potentially improvable) quality of the terrain.[6]

While sales prices of vacant land likely inform us the most about true land values, they are indeed much rarer than sales of improved properties. In the data, areas as large as a census tract typically have no vacant transactions in a given year. Thus, the major innovation we provide is to pair equation (2a) with a second equation for the transacted price of an improved lot, also in logarithmic form, y_i^m:

$$y_i^m = \phi_i r_i + (X_i^m)' \beta^m + u_i + e_i^m \tag{2b}$$

The determinants of improved land sales prices in equation (2b) are more varied and complex than in (2a), but share many parallels:

- r_i, the value of the vacant land itself, but in proportion to some rate ϕ_i. In theory, the linear in logarithms formulation implies a Cobb-Douglas production function, where the ϕ_i parameter represents the cost share of land in production, assuming non-land inputs do not

[6] In practice, vacant land often has some minor private improvements such as grading or landscaping, although these may not be particularly valuable to a new owner. Access to public improvements – water, sewage, roads, electricity, etc. – does not pose any particular problems in valuing vacant land as such, particularly with regards to land taxation.

vary in price spatially. The subscript i indicates how this share could depend on features of the property. It should be less than one, for all but truly vacant land.[7]

- X_i^m includes observable improvements on the land, such as the type of structure or built square feet. Like X_i^v, it should also include features of the transaction. The point of this term is merely to control for these observable features.[8]

- u_i captures the determinants of improved property values outside of vacant land costs and improvements. This is modeled using the same variables as in (1).

$$u_i = (Z_i)'\delta^u + \eta_j^u \tag{3}$$

  For instance, we would expect tight land-use regulations to raise the price of improved sales relative to vacant ones (Albouy and Ehrlich, 2018). For the purpose of valuing land, this is largely a nuisance term, as it captures confounders from trying to estimate the land value parameters δ^r and η_j^u from improved property data.

- e_i^m is the error term. It accounts for measurement error, as well as transaction characteristics. In addition, it may reflect unobserved characteristics of an improved property, such as the color of the exterior walls.

Because the price index in equation (2a) fixes the loading of r_i to 1, it is meant to reflect the intrinsic value of vacant land. After accounting for controls, observed vacant land sales vary proportionally with the land value index r_i. Improved sales should vary less, in proportion to ϕ_i, insofar as land values are orthogonal with u_i.

2.3 Identification

On their own, the identification requirements in estimating (2a) are fairly standard. These involve properly specifying location and lot characteristics, while dealing with omitted variables in e_i^v. A compounding challenge (and opportunity) lies in identifying the parameters in equations (2b) and (3), especially the factor loading parameter ϕ_i. The estimable model acknowledges that in the

[7]Evidence for a Cobb-Douglas relationship is seen in Thorsnes (1997), Epple et al. (2010), and Combes et al. (2021). The relationship may be generalized to depend non-linearly on r_i, viz., according to a function $\Phi_i(r_i)$, such as a a polynomial. For instance, a quadratic function $\phi_{i1}r_i + \phi_{i2}(r_i)^2$, would represent a Constant Elasticity of Substitution (CES) form.

[8]In principle, the variables in X_i^v related to transactions, should be a subset of X_i^m

improved sales u_i is estimated jointly with $\phi_i r_i$. In other words, the improved sales data estimate the term $m_i = \phi_i r_i + u_i$.

The parameters are identified by the fact that u_i is not included in the equation for vacant land sales. Identifying r_i without vacant land data requires imposing some restrictions on the structure of ϕ_i and on u_i. For instance, a common rule-of-thumb method used to value vacant land imposes a particular value for ϕ_i, such as 0.25, and assumes $u_i = 0$. The key here is that we may use the relationship between m_i and r_i to identify both ϕ_i and parameters in u_i. This means that in lieu of estimating (2a) and (2b), the method effectively estimates the reduced-form equations:

$$y_i^v = (Z_i)'\delta^r + \eta_j^r + (X_i^v)'\beta^v + e_i^v \tag{4a}$$

$$y_i^m = (Z_i)'\delta^m + \eta_j^m + (X_i^m)'\beta^m + e_i^m \tag{4b}$$

Focusing on a single type of improved property, $\phi_i = \phi$, we have that the reduced-form δ parameters obey

$$\delta^m = \phi\delta^r + \delta^u \tag{5a}$$

This equation reminds us that a challenge to estimating the effects of observable variables on vacant land values is complicated by both the scaling factor, ϕ, and an additional component, δ^u. Similarly, the area effects for improved properties are given by

$$\eta_j^m = \phi\eta_j^r + \eta_j^u \tag{5b}$$

To identify ϕ, we assume that the additional term for the area effects includes an idiosyncratic component orthogonal to the land component

$$\eta_j^r \perp \eta_j^u, \tag{5c}$$

much like a random effect. In practice, we make no such assumption for δ^u and δ^r, although it is possible in principle.

One way to understand how η_j^r and η_j^u work in (5b) is to decompose the area effects that determine y_i^v and y_i^m into two orthogonal factors. The first orthogonal factor, η_j^r, affects both vacant and improved prices, and the second orthogonal factor η_j^u is unique to improved lots. This exclusion restriction is plausible as long as observed simultaneous increases in both y_i^v and y_i^m are entirely attributable to an increase in the vacant land value, η_j^r. This identification assumption excludes any factor that affects the value of vacant land without affecting the value of improved properties. Variation in the value of vacant land has to pass through to the value of improved properties, as governed by the parameter ϕ. Thus, we assume there can be no systematic relationship between how municipalities affect vacant values with how they affect improved values' deviation from vacant values.[9]

A violation of this restriction might occur if a factor that affects the productivity of construction is correlated with the value of land. For instance, if areas with higher land values have restrictive zoning that lower productivity, then the estimate of ϕ will be biased up. Higher land prices will appear to push up housing prices more than they actually do if we fail to account for the higher regulatory costs created by restrictive zoning.

Note that if the aim is only to generate a good out-of-sample prediction about a transaction price of vacant land (not the underlying value), y_i^v, then this identification assumption (5c) does not play a role (at least under the linear-Gaussian assumption). Independence of η_j^r and η_j^u does not affect the predictive performance, as ϕ will reflect a statistical dependence. The assumption is needed to interpret r_i and u_i distinctly in economic terms.

2.4 Bayesian estimation methods and variance structures

Bayesian methods posit that each parameter in the model is known up to some quantifiable level of uncertainty, modeled by a probability distribution. We begin with an extremely uninformative prior knowledge on these parameters, and update these beliefs using the available data. This creates a posterior distribution of parameters that is much more precise than the prior. As alluded to earlier, these parameters have a hierarchical structure.

To simplify the mathematical exposition, we focus on the lower-order parameters that describe the area effects, η_j. The hierarchical method draws the area effects, η, from a conditional distribution, determined by a set of hyper-parameters, which themselves have their own distribution. As an extreme example, suppose that an entire municipality, say $j = 1$, has no vacant land sales. Bayesian methods let us construct a reasonable distribution of the area effect based on vacant

[9] In principle, we could improve our estimate of ϕ using (5a) by assuming that δ^r operate like random effects, if the orthogonality condition is warranted. It is entirely possible for these effects to be correlated. More desirable municipalities, with higher land values in η_j^r could have stricter building codes or zoning requirements, pushing up η_j^u.

land sales seen in other areas. In addition, local improved sales provide a separate local signal on land values via $\phi_i r_i$ term in equation (3). The parameter ϕ_i has a distribution estimated from observations in municipalities outside of $j = 1$ that have both vacant and improved sales.

Bayesian methods also shrink estimates to mitigate the influence of outliers. Say that for municipality 2, there was only a single land sale. That land sale may not represent an entire discrete area for idiosyncratic reasons. A standard frequentist approach — namely, a fixed-effects model — would be pinned to estimating the land value for that area from that one sale. In the Bayesian method, the prior belief, constructed from other areas, would be updated to more closely reflect that one observation, but not completely. The degree of updating depends on how high the variance of the area effects, η_j^r, is relative to the variance of sampling errors in land transactions, e_i^v. The greater that sampling error – which seems to be large for vacant lots — the more suspicious we are that a small number of vacant sales will be representative of all local lots. This makes local improved sales more useful, although these two are effectively shrunken. Naturally, the more sales of vacant land in an area there are, the less important are sales of improved properties, as well as vacant land in other areas.[10]

Importantly, our method lets draws of η_j be correlated across space. Indeed, high value areas are likely adjacent. There are potentially many parameters that would describe the correlation structure. We assume that the correlation rises or falls with distance according to an exponentially declining distance metric

$$cov(\eta_j^r, \eta_{j\prime}^r) = \sigma_{\eta,r}^2 \exp(-d_{jj\prime}/k_{\eta,r}) \tag{6a}$$

$$cov(\eta_j^u, \eta_{j\prime}^u) = \sigma_{\eta,u}^2 \exp(-d_{jj\prime}/k_{\eta,u}) \tag{6b}$$

where $d_{jj\prime}$ is the Euclidean distance (in miles) between areas j and $j\prime$. Stacking η_j^r for all j's, from 1 to J, produces

$$\eta^v = [\eta_1^r, \eta_2^r, ..., \eta_J^r]' \sim N(0_{J\times 1},\ \Sigma\left(\sigma_{\eta,r}^2, k_{\eta,r}\right)),$$

where $0_{J\times 1}$ is a $J \times 1$ vector of zeros and the $J \times J$ covariance matrix, Σ, is parameterized by two scalars, $\sigma_{\eta,r}^2$ and $k_{\eta,r}$. $\sigma_{\eta,r}^2$ governs the variance of each element in η^r. The parameter $k_{\eta,r}$ governs the correlation between two η's. Holding distance between two areas fixed, a larger value of $k_{\eta,r}$ implies stronger correlations. As $k_{\eta,r} \to \infty$, $corr(\eta_i^r, \eta_j^r) \to 1$. On the other hand, as $k_{\eta,r} \to 0$,

[10] A large number of vacant land sales may be used help to estimate how ϕ_i may vary across different kinds of properties.

$corr(\eta_i^r, \eta_j^r) \to 0$. Similarly, for the idiosyncratic improved effect:

$$\eta^u = [\eta_1^u, \eta_2^u, ..., \eta_J^u]' \sim N(0_{J\times 1},\ \Sigma\left(\sigma_{\eta,u}^2, k_{\eta,u}\right)),$$

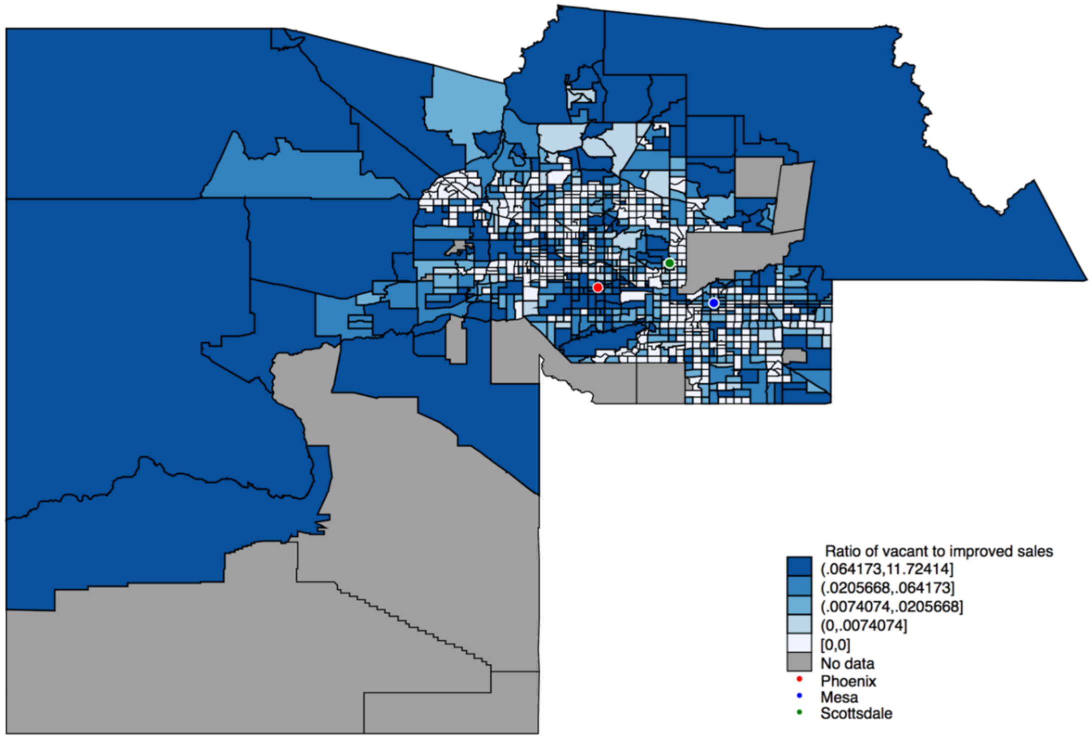
The variance term in $\sigma_{\eta,u}^2$ limits how much improved property sales can get at vacant land values. Even with an infinite number of such sales, and a known ϕ, the η_j^u term hides the value of vacant land.

The model is completed by specifying the prior distribution on other unknown parameters. The appendix provides a far more detailed account of how we solve it. As is standard, the main regression parameters $(\delta, \beta, \eta, \phi)$, have normal conjugate prior distributions and thus normal conditional posterior distributions. The variance parameters σ^2 take on a marginal posterior distribution that are inverse gamma, thereby taking only positive values. The hyper-parameters for these conjugate priors, which reflect our initial uncertainty, are set to minimize any impact on the prior distribution. We use a Metropolis-Hastings-within-Gibbs sampler algorithm, iterating over the blocks of parameters in 11 steps. The sampler begins from a set of estimates based on a more conventional, unshrunken estimate, using ordinary least squares (OLS). The MCMC method allows us to estimate numerically over an unbounded distribution, and is particularly useful for the spatial correlation parameters, k, for which the conditional probabilities are not closed form.

3 Incorporating data into the model

We illustrate how this empirical framework may be used for a large county for a single year, using the parsimonious specification shown above. The estimation sample includes transactions in 2018 from the Maricopa County Assessors' Office General Parcel Data and shapefiles from the US Census Bureau. We use transactions within 35 miles of Phoenix or Mesa, eliminating records with very small (less than 1/120 acre) or large lots (larger than 1 acre) and unreasonable values (e.g., the property type is not "vacant land" on the Affidavit of Sale). After cleaning, we are left with 1,153 vacant land sales. To focus on a single property type, we use improved sales for residential properties only, of which there are 99,174 after cleaning. This limited model may be expanded straightforwardly with sufficient data work and computing resources. We leave the challenges and opportunities of incorporating data over several years and property types for future work.

Figure 1: Ratio of vacant to improved sales transactions by census tract in Maricopa County, 2018



The land value index (1) is modeled as a function of variables that are available in both the vacant and improved sales. As mentioned above, these enter in both reduced-form equations, and model the intrinsic value of a vacant lot.

- d_i is the logarithm of 1 plus the minimum Euclidean distance in miles to the city hall in either Phoenix or Mesa, whichever is closest. We designate these two centers as our central business districts, or "CBDs," although this area is has decentralized commuting patterns.
- A_i is the log recorded lot size in square feet, along with 6 additional indicators for whether a lot is located on a street corner, in a cul-de-sac, in a gated community, on a lake, on a mountain, or on a paved road.
- η_j^r, the area effects, are determined either by
    1. 25 possible municipalities, accounting also for unincorporated areas, missing values, and with the city of Phoenix as the excluded category.[11]

[11] There are missing values in `citycode`. We treat them as a separate city, name it as "ZZ". In sum, we have 26 possible values for city code.

Table 1: Mean Land Characteristics for Vacant and Improved Properties

Type of Property Property	No. of Trans.	Price/ sq ft	Distance to CBD	Lot Size in sq ft	On Corner	In Cul-de-sac	In Gated Comm.	On Lake	On Mtn	Paved Road
Vacant	1,133	10	19	20,668	0.14	0.10	0.15	0.00	0.04	0.92
Improved	92,674	340	17	7,470	0.09	0.04	0.06	0.01	0.01	0.98

2. 887 possible census tracts as defined in the 2010 Census. This is the number of tracts in Maricopa County, excluding tracts without any transactions whatsoever (although these could be included). 60 percent of these tracts have no vacant sales, but do have improved sales.

Figure 1 shows the ratio of vacant transactions we observe relative to the number of improved transactions. In peripheral areas, we occasionally see more vacant transactions than improved ones, although in more central areas we see that a majority of tracts have no vacant transactions whatsoever.

Table 1 describes the observed land characteristics for both vacant and improved properties. Vacant lots in our sample are on average only two miles further than improved ones, but are nearly three times as large. Our sample also contains many vacant lots inside of gated communities. While vacant lots are less likely to be on a paved road than improved properties, still over 90 percent are.[12]

The controls for the residential properties used in the vacant land sales equation, (2a), X_i^v, include 3 variables:

- Three indicator variables for quarters when the transaction was recorded: Q2, Q3. The excluded category is Q1. Our data show no transactions in Q4.
- An indicator takes one if the transaction involves multiple parcels.

The controls for the residential properties used in the improved property sales equation (2b), X_i^m, include 15 variables:

- The log square footage of the built dwelling, i.e. the residential space.

[12]One concern is that the value of vacant land may be different in older areas with few sales, relative to newer, peripheral areas, where sales are more frequent. While the data generally show this latter pattern, we generally found that relative vacant land quality in older areas was not significantly different than in younger areas, although they were more frequently on a paved road or in a gated community.

- The recorded age of the building, and its square.
- Three indicator variables for whether the structure has two stories; three stories; four or more stories. The excluded category is a single story.
- Six indicators for a recorded measure of structure quality, ordered 2 through 7. The lowest category, 1, is excluded.
- The same three controls included in X_i^v.

4 Estimation results

Our model estimates are presented in a series of tables. Table 2 included the key hyper-parameters. To save space, we discuss the values of the reduced-form coefficients in Table 4, and the control variables in Table 5 in the appendix. Our tables contain estimates using either municipalities or census tracts for the discrete area classification. The tables show the posterior mean, posterior standard deviation, and 90% credible interval for the highest posterior density. This is the narrowest interval involving values of highest probability density.

4.1 Core estimates

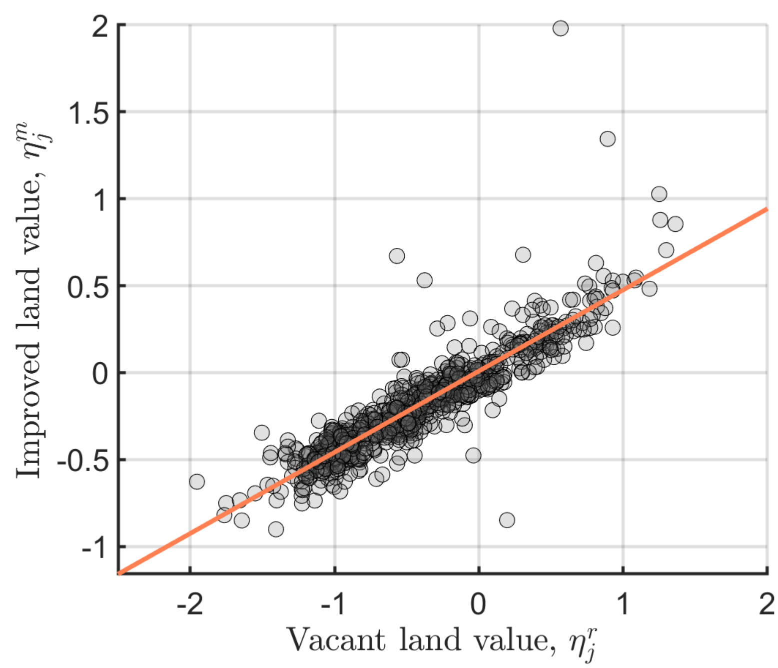
The first row of Table 2 shows the key loading parameter, ϕ, has an estimated distribution centered at 0.38 in the municipal model, and 0.43 in the census-tract model. The higher value of ϕ in the census-tract model seems to be due to more precise geography weakening attenuation effects. These numbers are somewhat larger than estimates in the literature for the cost share of land in housing production, e.g. Epple et al. (2010), Combes et al. (2021). Figure 2 helps illustrate how ϕ is estimated in the census tract model. It plots the improved-property effects against the vacant-land effects. The slope of the fitted line, which reflects equation (5b), gives a value of ϕ close to that reported in the table.

The estimated standard deviation of the local effects across municipalities, $\sigma_{\eta,r}$, is 0.71. Across census tracts it is substantially larger, at 1.31, due to there being over 30 times more tracts than municipalities. The estimates also indicate that there is considerable spatial correlation in the area land effects, η_j^r. For example, the correlation between η_j^r and $\eta_{j'}^r$ is 0.75 ($= \exp(-5/17.67)$) when the distance between municipality centroids is 5 miles. This spatial correlation is stronger across

Table 2: Parameter Estimates from Full Bayesian Estimation, Part 1: Core estimates

Dependent variable: log(price) per sqft	Municipal model				Census tract model			
			Lower	Upper			Lower	Upper
Parameter names	Mean	Std. Dev	(10th p)	(90th p)	Mean	Std. Dev	(10th p)	(90th p)
Cost share of land, ϕ	0.38	0.09	0.27	0.48	0.43	0.02	0.41	0.45
Land area variance, $\sigma_{\eta,r}$	0.71	0.31	0.44	1.13	1.31	0.24	0.97	1.62
Land area spatial correlation, $k_{\eta,r}$	17.67	20.94	3.43	44.97	58.57	18.81	32.82	85.71
Improved area extra variance, $\sigma_{\eta,u}$	0.23	0.10	0.13	0.37	0.14	0.005	0.14	0.15
Improved area extra correlation, $k_{\eta,u}$	30.39	26.94	4.49	67.79	0.07	0.05	0.01	0.15
Vacant transaction variance, $\sigma_{e,v}$	0.72	0.02	0.70	0.74	0.60	0.01	0.59	0.62
Improved transaction variance, $\sigma_{e,m}$	0.34	0.001	0.34	0.34	0.26	0.001	0.26	0.27

Figure 2: Posterior mean of improved η_j^m versus vacant η_j^r value effects for census tracts



census tracts, at 0.92. The statistically significant spatial correlation among η_j^r's implies that much is to be learned about land values from adjacent locations, after accounting for noise.

Not surprisingly, the finer tract-level model fits the data considerably better, as measured by the sampling error. The mean estimate of the standard deviation of the sampling error for vacant transactions, $\sigma_{e,v}$, is 0.72 for the municipal model and lower, at 0.60, for the tract model. These measures are similar to or smaller than the standard deviations corresponding to the areas, $\sigma_{\eta,r}$. This implies that a single vacant transaction in a census tract is still more informative than the estimate the model would provide using only surrounding data. Each vacant observation should cause a substantial update of η_j^r relative to the prior values. Indeed, the influence on the mean estimate should be proportional to the inverse of the variance, a.k.a. the precision $\tau \equiv 1/\sigma^2$.

The sampling error for improved property sales is less variable than for vacant sales: $\sigma_{e,m} < \sigma_{e,v}$. However, in determining how much this informs land values, the model in (2b) implies this number should be scaled up by $1/\phi$, which provides values of 0.89 and 0.60 in the two models.[13] These numbers already imply that a vacant sale is more informative about land values than an improved sale, even before considering additional un-observables in η_j^u. This is counter-balanced by the fact that improved sales are about 80 times more common.

The next step is then to consider the additional un-observables in the improved properties, η_j^u. These limit how much improved sales can inform the land value index. While the estimated values for $\sigma_{\eta,u}$ are relatively small, they, too, need to be scaled up by $1/\phi$, producing larger numbers of 0.61 and 0.32. Thus, even one local vacant sale may be about as informative as a very large number of improved sales. This is worth bearing in mind as 60 percent of census tracts had no clean vacant sales in 2018 — this was true for only 4 percent of tracts going back to 2007. Meanwhile, only about 2 percent of tracts have no clean improved sales in 2018.[14] In the end, how much improved sales can refine land value estimates depends on many features of the data, including the number of observations in each area j, and the strength of spatial correlations. Thus, we numerically quantify the gains from the joint estimation in the next subsection.

4.2 Illustrating the benefits of joint estimation

So far we have described only abstractly the benefits of simultaneously estimating the values of vacant land and improved properties in a Bayesian framework. Here we quantify the benefits precisely, taking advantage of the fact that the model produces a complete probability distribution for every parameter, including every area effect. We simply compare the posterior standard deviation of each η_j^r with both the improved and vacant data to the larger standard deviation with only the

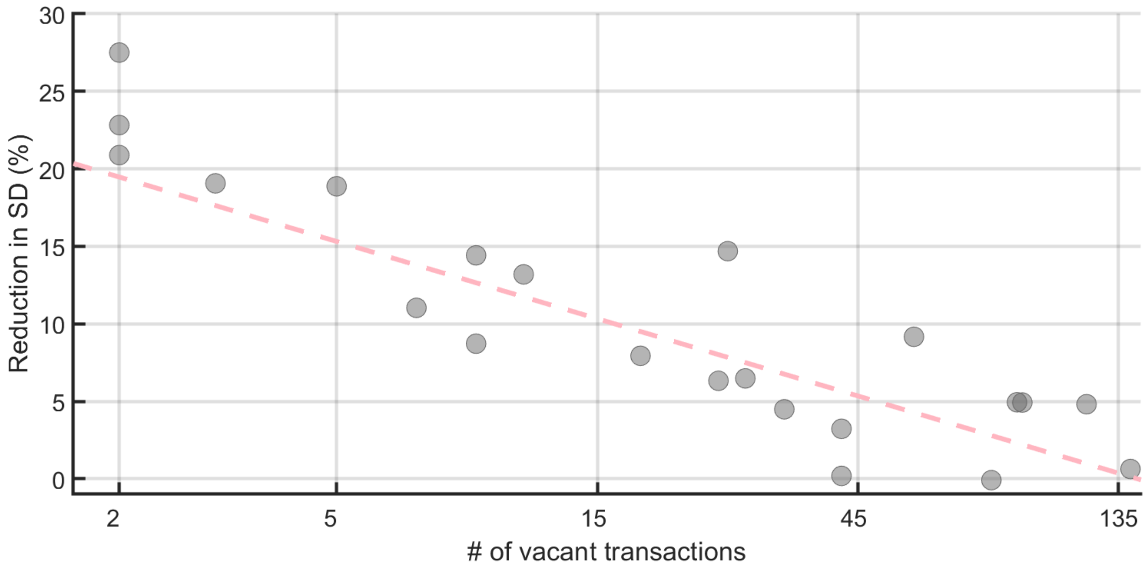
[13]To see this point, ignore the variables X and Z in the reduced-form equation in (4a) (4b) and substitute in (5b). Rearranging we have

$$\eta_i^r = y_i^v - e_i^v = \frac{1}{\phi} y_i^m - \frac{1}{\phi}\eta_i^u - \frac{1}{\phi} e_i^m$$

Thus, in estimating η_i^r the error term $(1/\phi)e_i^m$ is comparable to e_i^v. Furthermore, as the sample size grows large, the unobserved area effect $(1/\phi)\eta_i^u$ persists, even as the influence of the error terms, e_i^v and e_i^m vanish.

[14]Note that the spatial covariance in the land index implied by $k_{\eta,r}$ gets weaker moving from the municipal to the tract-level model, while the opposite happens for the idiosyncratic component for improved properties, as implied by $k_{\eta,u}$. This may be due to the smaller number of vacant land transactions, and might be improved by using periods over time.

Figure 3: Efficiency gain in using improved sales to estimate municipal area effects of land values, η_j^r



vacant data.[15] The percent reduction quantifies how much the joint estimation technique increases our certainty of the local land-value index.

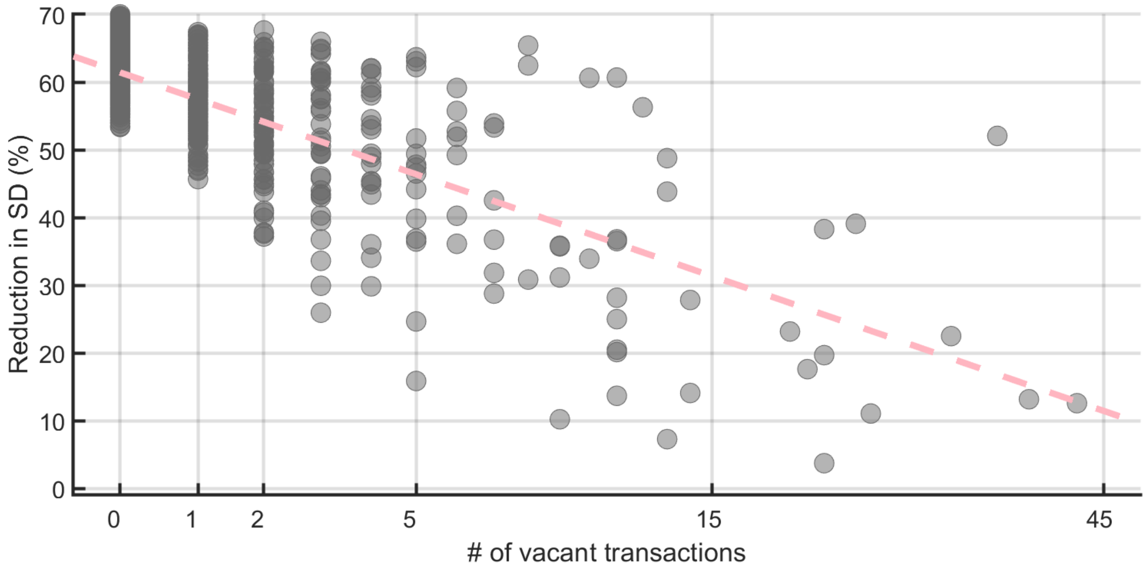
Figure 3 plots these percentage reductions for the municipal model, where the geography is coarse, but vacant transactions are relatively abundant. The x-axis arranges the municipalities according to the number of vacant transactions available in area j.[16] For municipalities with the fewest number of transactions, the standard deviation falls about 20 percent. Naturally, the gain is smaller for municipalities with more vacant transactions, falling to two percent in those with over 100 vacant transactions. Nevertheless, there is almost always a gain.

Figure 4 plots the percentage reductions for the census-tract model. With finer geography, and few or no vacant transactions in each area. This means that information from improved property sales is potentially much more valuable. The figure confirms this expectation: the posterior standard deviation for most of the tracts falls by over 30 percent. For tracts with no transactions, the gain is always over 50 percent.

[15] Note that this differs from σ_η^r, which captures the data generating process, conditional on knowing the parameters with certainty.

[16] Table 6 in Appendix contains related information about this figure.

Figure 4: Efficiency gain in using improved sales to estimate census-tract area effects on land values, η_j^r



5 Valuing vacant land

There are several ways to use this model to predict the value of properties. The first way is to predict the value of a certain transaction. This way may be used to cross-validate the model using observations from outside of our estimating sample. In principle, it could be used by a developer or other investor to determine whether or not a certain offer price for a property is likely over- or under-valued. Our focus here is on vacant lots, although the methodology is easily applied to improved properties. The second way is to predict the value of the underlying land itself, i.e. the index, which would likely be more useful for levying a land tax.. As mentioned previously this requires normalizing X_*^v, and perhaps defining Z, to reflect the type of value one wants. As Bayesian methods avoid over-estimating between-group dispersion, and try to zero out noise in particular transactions, this latter measure may have considerable public appeal.

5.1 Estimating the value of individual lots

The Bayesian model produces a probability distribution for any vacant-land transaction given its observable characteristics. In turn, this distribution may be used to compute any number of statistics about probable land values. For a vacant lot with characteristics $[d_*, a_*, X_*^v]'$ in city j, its value can be represented as

$$y_*^v = (Z_*^v)\delta^v + \eta_j^r + (X_*^v)'\beta^v + e_*^v, \quad e_*^v \sim N(0, \sigma_{e,v}^2) \tag{7}$$

There are two sources of uncertainty. The first is from e_*^v, which captures uncertainty about transactions unexplained by the estimated parameters model. The second is from uncertainty in the estimates of the parameters themselves. Thus, the predictive distribution of the vacant land price y_*^v can be expressed formally as a product of these two corresponding conditional probabilities:

$$p\left(y_*^v \middle| [Z_*, X_*^v]', \mathcal{D}\right) = \int p\left(y_*^v \middle| [Z_*, X_*^v]', \delta^r, \eta_j^r, \beta^v, \sigma_{e,v}^2\right) \times p\left(\delta^r, \eta_j^r, \beta^v, \sigma_{e,v}^2 | \mathcal{D}\right) d\delta^r d\eta_j^r d\beta^v d\sigma_{e,v}^2$$

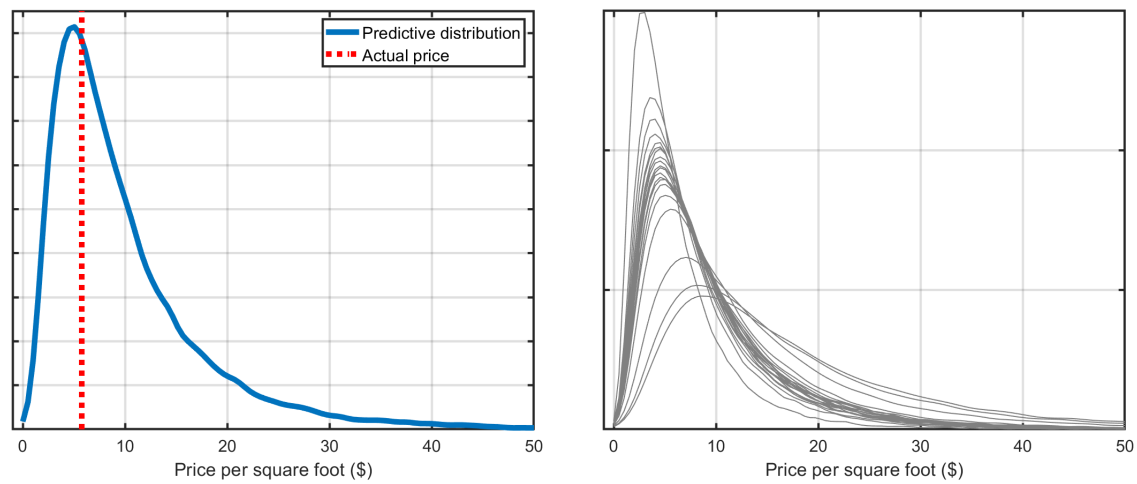
$\mathcal{D}$ is the data used to estimate the posterior distribution of unknown parameters. The first multiplicand is based on equation 7; the second is based on the posterior distribution of unknown parameters. Although there is no closed-form for this predictive distribution, it is possible to simulate land values from the distribution. These simulated draws can be used to approximate values of interest, such as point and interval predictions.

As an example, suppose we want to value a vacant lot in Phoenix that is three miles from the center and has a lot size of 7,910 square feet. The actual transaction price of this lot was \$4.04 per square foot. Figure 5 presents the estimated predictive distribution of the value of this lot based on the municipal model. The actual value is near the mode of this distribution. This predictive distribution is skewed to the right, with a mean and median of \$10.70 and \$8.02, respectively. This predictive distribution characterizes the uncertainty around these point estimates. For instance, one can construct an α%-credible interval, which contains the true value with α percent posterior probability. The 80% credible interval based on this model is [\$1.30, \$15.50], which is rather wide, implying that there the range of possible values we could observe in single transaction could differ by an order of magnitude. The right panel in Figure 5 expands on the left panel, presenting distributions for 20 distinct parcels at different locations with different characteristics. While each predictive distributions have a distinct location and shape, they all show considerable ranges.

These simulations highlight how a single transaction or comp may be far from representative. Researchers and assessors get a much more reliable sense of land values by taking into account both nearby and improved transactions. As we will discuss later, some portion of uncertainty can be attributed to the simplistic nature of our model, and may be reduced by including more co-variates

Figure 5: Predictive distribution for vacant land transaction prices (municipal model)

(a) Predictive distribution and actual outcome (b) Price distributions at 20 different locations



or a more detailed spatial structure.

5.2 Land value index

We now turn to using the model to provide index values for individual plots of land in (1). One may also aggregate these plots by municipality or other groupings. Recall that the land index may depend on variables other than location, seen in A_i. For our purposes, we set Z_i and X_i to area j's average value for observed transactions, denoted Z_j and X_j.

The posterior distribution of the vacant land-value index r in municipality j with a particular characteristics Z_j and X_j is given by integrating over the posterior distribution of parameters. We denote this distribution as

$$p(r_j|Z_j, X_j, \mathcal{D}) = \int p(r_j|\delta^r, \eta_j^r, Z_j, X_j, \mathcal{D})p(\delta^r, \eta_j^r|Z_j, X_j, \mathcal{D})d\delta^r d\eta_j^r \tag{8}$$

To procure dollar values, we set the index to be the posterior mean of the exponent of r_j,

$$\text{Vacant land value for } j = E[\exp(r_j)|Z_j, X_j, \mathcal{D}] = \int \exp(r_j)p(r_j|Z_j, X_j, \mathcal{D})dy_{ij}. \tag{9}$$

Table 3 presents mean land-value measures by municipality. The first column uses the means of the vacant transaction. The second and third columns present the mean land-value indices using two different sets of land characteristics, Z. Column 3 uses the lot characteristics of the vacant

Table 3: Land Values per Square Foot in Maricopa County by Municipality: Sample and Index Values for Vacant and Improved Properties

City Code	Municipality Name	Transaction Sample Mean (1)	Estimated Index Values Vacant (2)	Estimated Index Values Improved (3)	No. of transactions Vacant (4)	No. of transactions Improved (5)
AV	Avondale	10.3	8.1	8.0	33	1778
BU	Buckeye	4.7	5.0	6.7	28	3160
CC	Cave Creek	7.0	8.1	9.8	9	64
CF	Carefree	5.2	7.4	9.6	7	104
CH	Chandler	14.4	16.4	20.8	18	6266
EL	El Mirage	2.4	5.3	6.4	5	710
FH	Fountain Hills	7.7	9.7	11.8	42	785
GI	Gilbert	9.0	10.6	14.2	26	6745
GL	Glendale	9.7	10.3	8.5	57	4221
GO	Goodyear	5.5	6.1	6.4	42	3008
GU	Guadalupe	4.2	8.7	9.8	3	18
LP	Litchfield Park	12.3	10.7	11.1	2	275
MC	unincorporated	3.9	3.7	5.3	142	7257
ME	Mesa	8.8	9.7	11.3	79	11151
PE	Peoria	10.1	11.3	10.5	90	4876
PV	Paradise Valley	34.4	31.5	29.0	6	142
QC	Queen Creek	6.6	8.0	10.2	25	1445
SC	Scottsdale	19.0	15.4	21.8	118	7193
SU	Sun City	9.1	10.6	9.8	88	4764
TE	Tempe	23.4	18.9	20.2	11	2563
TO	Tolleson	6.7	6.8	7.1	2	60
YO	Youngtown	4.7	6.7	6.7	2	171
ZZ	Unknown	17.4	9.2	11.7	9	256
PH	Phoenix	11.1	9.9	11.1	309	32162
	Maricopa County (arithmatic)	10.1	9.7	11.1	1153	99174
	Maricopa County (geometric)	6.9	6.9	6.9	1153	99174

transactions; column 4 instead uses those for improved ones. Recall, that while we list only the mean values, the empirical model produces an entire distribution.

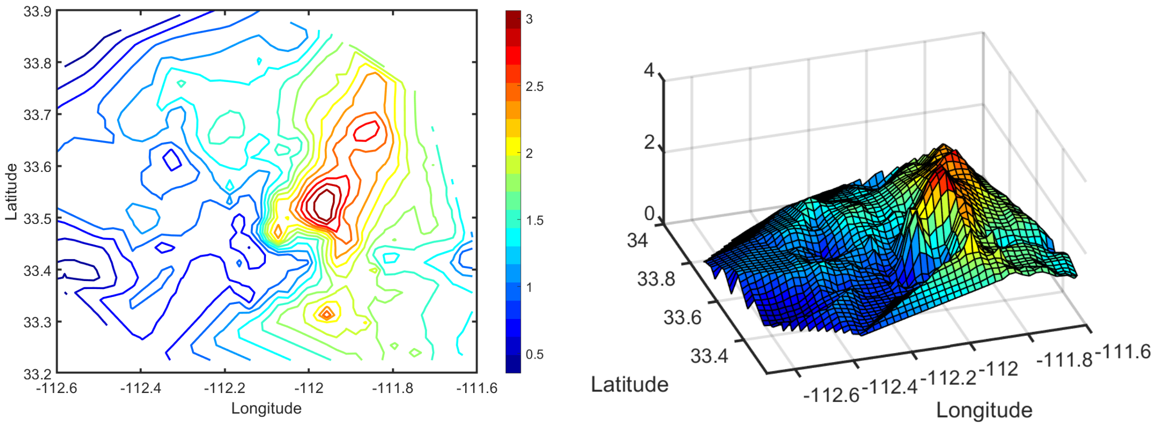
The land-value index for vacant transactions in (2) is typically closer to the county mean than the values taken directly from the sample in (1), a result of shrinkage.[17]. In column 3, we see that land values for improved properties are generally higher than vacant lots, because they have better lot characteristics, Z, mainly smaller plot sizes, as described in Table 1.

In some municipalities, like Chandler or Gilbert, the vacant land-value index is higher than the sample average when improved properties have high estimated values. The opposite is true in low-value unincorporated areas. These refinements result from our joint estimation procedure. They imply that vacant lots in high-value areas are often more negatively selected than vacant lots

[17]The index mean is slightly lower than the sample mean for the arithmetic mean, but is no different for the geometric mean. This appears to be the result of modeling the index in logarithms.

Figure 6: Posterior mean of vacant land value over space

(a) Contour plot (b) 3D plot



in low-value areas.

The values we report are based only on transacted properties, although in principle one could estimate land values for all lots, including those for which there is no transaction record. At the same time, one must recognize the shortcomings with transactions data: one can only imperfectly control for characteristics that occur more often in transacted properties than in non-transacted ones.

Figure 6 presents a 3-dimensional contour map of land values using the geographically finer census-tract model. The benefit of having random location effects on a finer grid is clear: the estimated surface has much more realistic spatial variation than the one implied by the coarser municipality model. In addition, we can see that the estimated location effects exhibit a high degree of non-linearity, although without any obviously erratic patterns. While several areas show local maxima for land values, eventually land values do eventually decrease as one moves towards the edge of the county.

6 Extensions

The model in the previous section is kept simple to convey the gains from combining information from improved with vacant land sales. In this section, we propose a few possible extensions.

6.1 Heterogeneous land-value gradients

The basic model does little to model continuous changes in land values, such as within a municipality or a tract. One way to enrich the spatial structure within our model is to allow for heterogeneity in slope parameters in (1) using the following equation

$$r_i = (\delta_0^r + \eta_{1,j}^r) + (\delta_1^r + \eta_{2,j}^r)d_i + \delta_2^r A_i \tag{10}$$

Following a similar logic from before, this then leads to an expanded set of reduced-form equations describing the transaction prices:

$$y_i^v = \delta_0^r + \delta_1^r d_i + \delta_2^r A_i + \eta_{0,j}^r + \eta_{1,j}^r d_i + (X_i^v)'\beta^v + e_i^v \tag{11a}$$

$$y_i^m = \delta_0^m + \delta_1^m d_i + \delta_2^m A_i + \eta_{0,j}^m + \eta_{1,j}^m d_i + (X_i^m)'\beta^m + e_i^m \tag{11b}$$

The additional parameters, $\eta_{1,j}^r$ and $\eta_{1,j}^m$, capture possibly different slopes using distance from the CBD within each municipality. This specification allows that gradients vary across municipalities. For example, one could expect that the gradient is steeper in a more central municipality, while the gradient gets flattened as we move further away from the center.

6.2 Heterogeneous land shares

Another way to enrich the spatial structure within our model is to allow for heterogeneity in the land-share parameter ϕ_i. This may be done by letting ϕ_i vary by observable characteristics. Because it is modeled stochastically, one can apply a hierarchical structure that would be difficult to model in frequentist settings. For example, one may allow the cost-share parameter to vary by municipality j so that

$$\eta_j^m = \phi_j \eta_j^r + \eta_j^u, \quad \eta_j^r \perp \eta_j^u. \tag{12}$$

Recall η_j^r and η_j^m are multivariate normal random vectors. Whether ϕ_j is space-varying or not is an empirical matter. One may validate this possibility with the data. For example, we can estimate the ϕ_j with the similar prior distribution for η_j,

$$\boldsymbol{\phi} = [\phi_1, \phi_2, ..., \phi_K]' \sim N(\mu_\phi,\ \Sigma\left(\sigma_\phi^2, k_\phi\right)), \tag{13}$$

where μ_ϕ is a $K \times 1$ vector and the $K \times K$ covariance matrix is parameterized by two scalars: $cov(\phi_i, \phi_j) = \sigma_\phi^2 \exp(-d_{ij}/\phi_v)$. The parameter σ_ϕ^2 governs the variance of each element in $\boldsymbol{\phi}$, and k_ϕ governs the correlation between two ϕ_j's.

6.3 A nearly non-parametric spatial land value function

A serious limitation in determining land values of particular lots is that they can vary over rather small geographies. Even adjacent lots may differ in values because of their views or their neighbors, characteristics which may be difficult to observe or quantify. As detailed above, using a finer geography increases the value of using improved sales to estimate vacant values. As one procures additional data, one can make the geography increasingly fine, in the limit letting areas j be the same as lots i. In this case, the model becomes:

$$
\begin{aligned}
y_i^v &= (X_i^v)'\beta^v + Z_i'\delta^r + \eta_i^r + e_i^v \\
y_i^m &= (X_i^m)'\beta^m + Z_i'\delta^m + \eta_i^m + e_i^m,
\end{aligned}
$$

where η_i^r and η_i^m have an i-subscript. If we adopt the same class of covariance function for η_i^r and η_i^m, then this model becomes a variant of the Gaussian process prior model, also known as "Kriging" in geo-statistics. This model aims to estimate for some arbitrary location l_i

$$
\begin{aligned}
y_i^v &= (X_i^v)'\beta^v + Z_i'\delta^r + f(l_i) + e_i^v \\
y_i^m &= (X_i^m)'\beta^m + Z_i'\delta^m + g(l_i) + e_i^m,
\end{aligned}
$$

where $f(l_i)$ and $g(l_i)$ are spatially varying non-parametric functions for the intrinsic values of land and improvements, respectively. One important distinction from the conventional Gaussian process prior model is that our would allow $f(l_i)$ and $g(l_i)$ to be codependent, so that improved sales are informative about $f(l_i)$.

6.4 Modeling evolution over time

We can also extend our model to allow for time evolution of the land value when we have transactions collected over time.[18] There are a number of ways of attempting this, such as by letting the area

[18] For an overview of challenges in estimating spatial hedonic models over time, see Thanos et al. (2016).

effects vary with time:

$$y_{i,t}^{v} = (X_{i,t}^{v})'\beta^{v} + Z_{i,t}'\delta^{r} + \eta_{j,t}^{r} + e_{i,t}^{v}$$
$$y_{i,t}^{m} = (X_{i,t}^{m})'\beta^{m} + Z_{i,t}'\delta^{m} + \phi_t\eta_{j,t}^{r} + \eta_{j,t}^{u} + e_{i,t}^{m},$$

where we denote t as a time index. The key is how to model $\eta_{j,t}^{r}$ and $\eta_{j,t}^{u}$. Ideally, one would like to allow correlation across both space and time so that we can borrow information from nearby transactions both in terms of calendar time and physical distance. A simple modeling strategy would be to decompose $\eta_{j,t}^{r}$ into two additive components,

$$\eta_{j,t}^{r} = \mu_j^{r} + \psi_t^{r}$$

where μ_j^{r} is a temporally-fixed spatial component and ψ_t^{r} is a spatially-fixed time component. The spatial component can be modeled in a similar fashion as before:

$$\mu^{r} \sim [\mu_1^{r}, \mu_2^{r}, ..., \mu_J^{r}]' \sim N\left(0_{J\times 1}, \Sigma(\sigma_{\mu,r}^{2}, k_{\mu,r})\right)$$

where $0_{J\times 1}$ is a $J \times 1$ vector of zeros and the $J \times J$ covariance matrix, Σ, is parameterized to allow for spatial correlation. We can model the time-varying component using a standard time-series model. A minimalist way is to model a random-walk process,

$$\psi_t^{r} = \psi_{t-1}^{r} + v_t, \quad v_t \sim N(0, \sigma_{\psi,r}^{2}), \quad \psi_0^{r} \sim N(0, \sigma_{0,\psi,r}^{2}).$$

The time-varying component resembles the space-varying component except that the correlation structure depends on a measure of temporal distance, rather than spatial distance. This distance dependent correlation leads to an automatic shrinkage/smoothing by assuming that nearby vacant land values are similar to each other. As we illustrated earlier, this type of shrinkage is helpful when there are not many transactions available for a certain area or time.[19]

There is also the possibility that the cost share of land, ϕ_t, will vary over time. For example, land and may appreciate at a faster rate than stuctures — a phenomenon related to the "land-leverage"

[19]When $\mu_j^{r} = 0$, the above model reduces to a class of models developed and studied by Schwann (1998), Francke and De Vos (2000), Francke (2010) for constructing real estate price indices. The common idea in this type of models is that there is a serially correlated latent variable that smooths its estimate over time. One can also enrich the model by including location/cluster-specific trends, $\psi_{t,k}^{r}$ where k is a group indicator. In this way, time trends are clustered within a subset of locations as in Ren et al. (2017) and Francke and van de Minne (2017).

hypothesis (Bostic et al., 2007) — causing the share ϕ_t to rise during market upswings. We leave modeling this phenomenon to future work.[20]

7 Conclusion

While we believe the model we present is novel and helpful, it leaves room for more elaborate specifications. Given this room, and the limits in our data, we consider the above results to be still preliminary and suggestive. Nevertheless, they mark a progression towards attaining more precise and accurate estimates of land values, albeit at the cost of increased computation.

While the iterative estimation procedure is not the most transparent among available procedures to value land, the output may have considerable appeal. It estimates an intuitive parameter describing how much a property's cost is due to its land. Furthermore, it optimally combines data on improved values, which are abundant and local, and vacant land, which are accurate but sparse. Furthermore, the methodology has safeguards against over-inferring typical land values from any one idiosyncratic observation, something which less technical methods would have a hard time judging. At the same time, the method accounts for how vacant lots typically differ from the typical properties one may need to assess.

At the practical level, these methods provide particular promise in providing land value estimates that assessors and citizens will find acceptable due to their overall accuracy. By accounting for a greater range of uncertainty than conventional models, citizens might also find them less imposing if assessors can find a way to communicate such uncertainty properly.

A number of goals lie ahead. First, it would be an interesting exercise to estimate and compare models proposed in Section 6. In addition, we can evaluate the model by testing the out-of-sample predictions, holding some observations out of the estimation sample. This may be used to evaluate point, interval, and even density prediction of transaction prices for both vacant and improved properties.

[20]One distinction we examined is if ϕ differed in older, more established neighborhoods, relative to newer, more peripheral neighborhoods. We classified a tract as young if the average transacted property was less than 25 years old. To improve accuracy, we used four years of data. The mean value of the distribution of ϕ for young areas is 0.37 (std dev 0.02), whereas for older areas, the mean value is 0.42 (std dev of 0.02). This appears to be inconsistent with the view that land quality of vacant lots sold in older areas is of lower quality than in younger areas. Rather, it appears consistent with the land leverage hypothesis, if cost shares were similar at the time of construction.

Interestingly, the land-value index could be conditioned to depend on variables deemed worthy of land-value taxation. This includes access to public services: location in certain school districts, possibly organized by average test scores; proximity to major highways, hospitals, or other public services. These may be the most politically acceptable for land value taxation as they reflect benefits provided by local governments. At the same time, we could include crime rates or air quality, to provide discounts for residents living in less favorable conditions. One might also purposefully exclude the estimated effects of zoning or land-use regulations, which may artificially lower land values, to encourage local communities or developers to use land in more profitable ways.

References

Albouy, D. and G. Ehrlich (2018): "Housing Productivity and the Social Cost of Land-use Restrictions," *Journal of Urban Economics*, 107, 101–120.

Angrist, J. D., P. D. Hull, P. A. Pathak, and C. R. Walkers (2017): "Leveraging Lotteries for School Value-Added: Testing and Estimation," *Quarterly Journal of Economics*, 132, 871–919.

Ashley, R., F. Plassmann, and N. Tideman (1999): *Improving the Accuracy of Downtown Land Assessment*, Lincoln Institute of Land Policy.

Bell, M. E. and J. H. Bowman (2006): *Methods of Valuing Land for Real Property Taxation: An Examination of Practices in States that Require Separate Valuation of Land and Improvements*, Lincoln Institute of Land Policy.

Bostic, R. W., S. D. Longhofer, and C. L. Redfearn (2007): "Land Leverage: Decomposing Home Price Dynamics," *Real Estate Economics*, 35, 183–208.

Combes, P., G. Duranton, and M. Gobillon (2021): "The Production Function for Housing: Evidence from France," *Journal of Political Economy*, 129, 2766–2816.

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.

Ecker, M. D. and H. R. Isakson (2005): "A unified convex–concave model of urban land values," *Regional Science and Urban Economics*, 35, 265–277.

Epple, D., B. Gordon, and H. Sieg (2010): "A New Approach to Estimating the Production Function for Housing," *American Economic Review*, 100, 905–24.

Francke, M. and A. van de Minne (2017): "The Hierarchical Repeat Sales Model for Granular Commercial Real Estate and Residential Price Indices," *Journal of Real Estate Finance and Economics*, 55, 511–532.

Francke, M. K. (2010): "Repeat Sales Index for Thin Markets: A Structural Time Series Approach," *Journal of Real Estate Finance and Economics*, 41, 24–52.

Francke, M. K. and A. F. De Vos (2000): "Efficient Computation of Hierarchical Trends," *Journal of Business and Economic Statistics*, 18, 51–57.

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

Gyourko, J. and J. Krimmel (2021): "The Impact of Local Residential Land Use Restrictions on Land Values across and within Single Family Housing Markets," *Journal of Urban Economics*, 126, 103374.

Huang, A., M. P. Wand, et al. (2013): "Simple Marginally Noninformative Prior Distributions for Covariance Matrices," *Bayesian Analysis*, 8, 439–452.

Kok, N., P. Monkkonen, and J. M. Quigley (2014): "Land Use Regulations and the Value of Land and Housing: An Intra-Metropolitan Analysis," *Journal of Urban Economics*, 81, 136–148.

Mills, E. S. (1998): "The Economic Consequences of a Land Tax," *Land Value Taxation: Can It and Will It Work Today*, 31–48.

Ren, Y., E. Fox, and A. Bruce (2017): "Clustering Correlated, Sparse Data Streams to Estimate a Localized Housing Price Index," *Annals of Applied Statistics*, 808–839.

Schwann, G. M. (1998): "A Real Estate Price Index for Thin Markets," *Journal of Real Estate Finance and Economics*, 16, 269–287.

Thanos, S., J. Dube, and D. Legros (2016): "Putting Time into Space: the Temporal Coherence of Spatial Applications in the Housing Market," *Regional Science and Urban Economics*, 58, 78–88.

Thorsnes, P. (1997): "Consistent Estimates of the Elasticity of Substitution between Land and Non-Land Inputs in the Production of Housing," *Journal of Urban Economics*, 42, 98–108.

Thorsnes, P. and D. P. McMillen (1998): "Land Value and Parcel Size: A Semiparametric Analysis," *Journal of Real Estate Finance and Economics*, 17, 233–244.

Appendix

A Additional tables

In this section, we report and discuss additional tables regarding the estimation of municipal and tract-level models presented in the main text.

A.1 Parameter estimates: Other lot characteristics

Table 4: Land determinants. Although the Phoenix-Mesa metropolitan area that includes Maricopa County is not very centralized, it is interesting to note the price gradient away from the CBD which we place in Phoenix and Mesa. The coefficient on continuous distance for both vacant and improved land is essentially zero in the municipal model. While the value of the municipal effects falls with distance from the center, the same is not true for distances within the municipalities, on average.

The same logic applies to the estimates from the census tract model. Although the distance coefficients have a positive sign, they should not be interpreted as the standard land price gradient as η_j^r and η_j^u non-parametrically model the relationship between the value and the distance to CBD. Instead, they capture the linear location effects not accounted by the discrete area effects, η_j. Overall, we find that both the vacant and improved land values have a non-linear but decreasing relationship with the distance to CBD. To confirm this, we present the scatter plot showing the relationship between total spatial component $(\delta_1^v d_j + \eta_j^v)$ and distance to CBD in Figure 7. The overall land value gradient is decreasing in distance to CBD as the discrete area effects account for a much larger share of spatial variation.

The other lot characteristic is its area. Here we see the usual "plattage" pattern, which shows value per square foot falling with lot size. The value falls much faster for improved lots than for vacant lots. This is the opposite of what we would expect from a cost-driven story: the value of improved lots should drop at a slower rate with size. The negative coefficient is close to -1, which would imply that the value of a parcel does not depend on its size. Our estimation results suggest that the estimate for improved transactions likely suffers from severe omitted variable issues – larger lots may have much lower quality improvements – or mis-specification issues, possibly from the log-log form. These issues deserve further consideration.

Other land price determinants have reasonable estimates indicating that there is a premium for the vacant land located on a street corner, in a cul-de-sac, in a gated community, and on a paved road.[21] For improved properties, there is premium for being located on street corner, in a gated community, on a lake, on a mountain, and on a paved road.

Table 5: Land controls. There is small, but significant seasonality in the transaction data. For example, the price of the land sold in Q2 is slightly larger than that sold in other times. There is a premium for a transaction that is associated with multiple parcels. An improved land is cheaper if it is older or it is low quality.

[21] The posterior distribution for the parameter associated with "located on a lake" in the vacant land equation turns out to be essentially the same as its prior distribution. This is because there is no vacant land transaction that is located on a lake, and therefore the prior distribution for the associated parameter did not get updated.

Table 4: Parameter Estimates from Full Bayesian Estimation, Part 2: Land Determinants

Dependent variable: log(price) per sqft	Municipal model				Census tract model			
Variable names	Mean	Std. Dev	Lower (10th p)	Upper (90th p)	Mean	Std. Dev	Lower (10th p)	Upper (90th p)
Vacant land determinants δ^r								
Intercept	3.85	0.34	3.41	4.28	5.48	0.38	4.99	5.97
Log mileage plus one to CBD	0.03	0.05	-0.04	0.09	0.10	0.07	0.01	0.19
Log lot size in square feet	-0.25	0.03	-0.30	-0.21	-0.38	0.03	-0.42	-0.34
Located on street corner	0.08	0.06	0.00	0.16	0.14	0.05	0.07	0.21
Located in a cul-de-sac	0.14	0.07	0.05	0.23	0.11	0.06	0.04	0.19
In a gated community	0.27	0.07	0.19	0.35	0.27	0.06	0.19	0.34
On a lake	-0.02	5.03	-6.46	6.46	-0.05	5.00	-6.42	6.35
Located on mountain	0.01	0.13	-0.15	0.18	-0.15	0.11	-0.29	-0.01
On a paved road	0.38	0.09	0.26	0.50	0.26	0.09	0.14	0.37
Improved property determinants δ^m								
Intercept	6.32	0.04	6.26	6.37	6.70	0.09	6.57	6.82
Log mileage plus one to CBD, $\log(1 + d_i)$	0.01	0.00	0.01	0.02	0.09	0.02	0.06	0.13
Log lot size in square feet	-0.91	0.00	-0.91	-0.91	-0.86	0.00	-0.86	-0.85
Located on street corner	0.06	0.00	0.06	0.07	0.03	0.00	0.03	0.03
Located in a cul-de-sac	0.01	0.01	0.00	0.01	0.00	0.00	-0.01	0.00
In a gated community	0.04	0.00	0.03	0.05	0.01	0.00	0.01	0.02
On a lake	0.17	0.01	0.16	0.19	0.16	0.01	0.14	0.17
Located on mountain	0.04	0.02	0.02	0.07	0.03	0.01	0.01	0.04
On a paved road	0.03	0.01	0.02	0.04	0.06	0.01	0.05	0.07

Table 5: Parameter Estimates from Full Bayesian Estimation, Part 3: Land controls

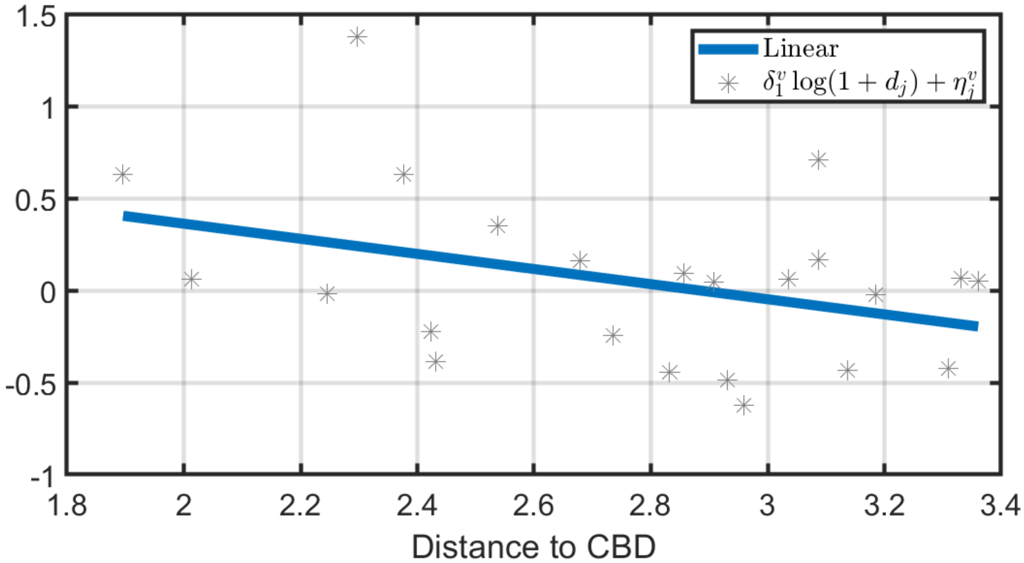
Dependent variable: log(price) per sqft	Municipal model				Census tract model			
Variable names	Mean	Std. Dev	Lower (10th p)	Upper (90th p)	Mean	Std. Dev	Lower (10th p)	Upper (90th p)
Vacant land controls β^v								
Q2	0.08	0.06	0.01	0.16	0.01	0.05	-0.05	0.08
Q3	-0.01	0.05	-0.07	0.06	0.01	0.04	-0.04	0.07
Multiparcel	0.61	0.06	0.52	0.69	0.77	0.06	0.70	0.84
Improved property controls β^m								
Log structure square feet	0.63	0.01	0.62	0.64	0.57	0.00	0.56	0.57
Age of structure/10	-0.01	0.00	-0.01	-0.01	-0.02	0.00	-0.02	-0.01
Age squared/1000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2 story structure	-0.12	0.00	-0.12	-0.11	-0.07	0.00	-0.08	-0.07
3 story structure	0.12	0.01	0.11	0.14	0.00	0.01	-0.02	0.01
4+ story structure	0.22	0.09	0.10	0.34	0.19	0.08	0.09	0.29
Quality grade 2	0.31	0.03	0.27	0.36	0.27	0.03	0.23	0.30
Quality grade 3	0.77	0.03	0.73	0.81	0.52	0.03	0.49	0.56
Quality grade 4	0.91	0.03	0.87	0.95	0.55	0.03	0.51	0.58
Quality grade 5	1.18	0.03	1.13	1.22	0.72	0.03	0.69	0.76
Quality grade 6	1.36	0.04	1.32	1.41	0.92	0.03	0.88	0.96
Quality grade 7	1.17	0.13	1.00	1.34	0.87	0.11	0.74	1.01
Q2	0.06	0.00	0.05	0.06	0.04	0.00	0.04	0.04
Q3	0.04	0.00	0.04	0.04	0.04	0.00	0.04	0.04
Multiparcel	3.22	0.01	3.21	3.23	2.89	0.01	2.88	2.90

A.2 Efficiency gain in using improved sales to estimate municipal effects on vacant land values

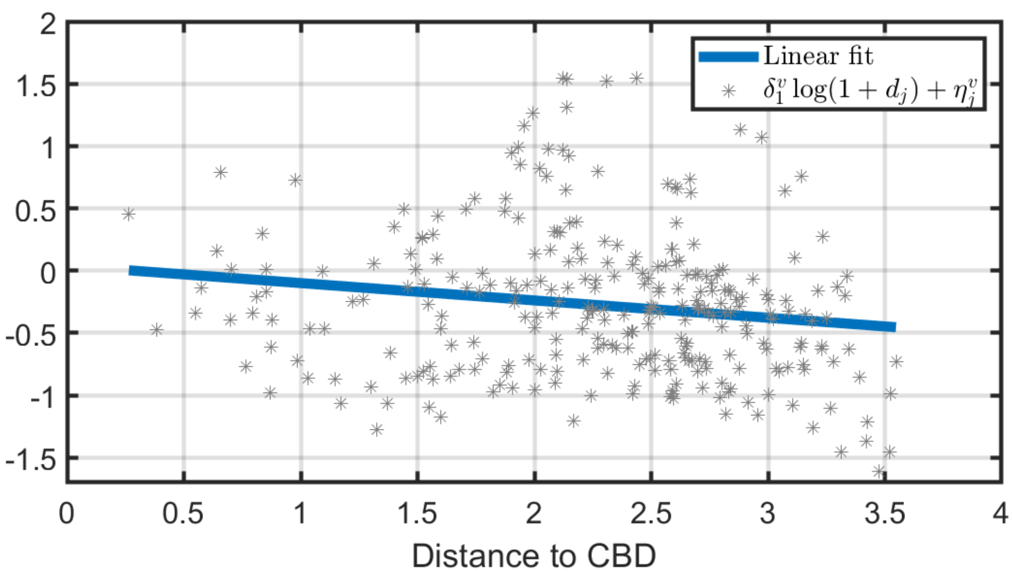
Table 6 reports posterior mean, standard deviation, and statistical efficiency estimating η^r_j. Numbers in this table are used to construct figure 3 in the main text.

Figure 7: Estimated spatial component over distance to CBD

(a) Municipal model



(b) Census tract model



B Un-shrunken estimates

When data availability is not a concern, we can estimate parameters in our model separately. That is, we estimate $(\beta^v, \delta_r, \eta_j^r)$ from the y_i^v equation (4a). Then, we estimate $(\beta^m, \delta^m, \eta_j^m)$ from the y_i^m equation (4b) where $\eta_j^m = \phi\eta_j^r + \eta_j^u$. Having estimated parameters in both equations, we can regress η_j^m on η_j^r to obtain a ϕ estimate. Here we perform a separate estimation of the v and m equations by OLS.

Figure 8 presents scatter plots of $(\eta_j^r$ and $\eta_j^m)$ for residential land values, and commercial land

Table 6: Posterior mean, standard deviation, and statistical efficiency for η_j^r

City code	Municipality name	Individual Estimation (Vacant data only) Mean	SD	Joint Estimation (Vacant and improved data) Mean	SD	SD ratio	No. of observations Vacant	Improved
AV	Avondale	-0.32	0.13	-0.26	0.13	0.96	33	1778
BU	Buckeye	-0.51	0.14	-0.46	0.15	0.94	28	3160
CC	Cave Creek	-0.02	0.20	-0.06	0.22	0.91	9	64
CF	Carefree	-0.03	0.21	-0.08	0.23	0.89	7	104
CH	Chandler	0.57	0.15	0.56	0.17	0.92	18	6266
EL	El Mirage	-0.56	0.22	-0.65	0.27	0.81	5	710
FH	Fountain Hills	0.09	0.14	0.07	0.14	1.00	42	785
GI	Gilbert	0.28	0.12	0.28	0.14	0.85	26	6745
GL	Glendale	-0.28	0.10	-0.25	0.11	0.91	57	4221
GO	Goodyear	-0.51	0.13	-0.55	0.13	0.97	42	3008
GU	Guadalupe	-0.07	0.26	0.07	0.32	0.81	3	18
LP	Litchfield Park	-0.03	0.24	-0.09	0.33	0.72	2	275
MC	unincorporated	-0.70	0.09	-0.71	0.09	0.99	142	7257
ME	Mesa	0.01	0.09	0.01	0.09	1.00	79	11151
PE	Peoria	-0.02	0.10	0.00	0.10	0.95	90	4876
PV	Paradise Valley	1.32	0.26	0.81	0.25	1.04	6	142
QC	Queen Creek	0.02	0.14	0.07	0.15	0.94	25	1445
SC	Scottsdale	0.63	0.09	0.65	0.10	0.95	118	7193
SU	Sun City	-0.10	0.11	-0.07	0.11	0.95	88	4764
TE	Tempe	0.58	0.17	0.62	0.20	0.87	11	2563
TO	Tolleson	-0.45	0.25	-0.17	0.32	0.79	2	60
YO	Youngtown	-0.52	0.23	-0.47	0.30	0.77	2	171
ZZ	Unknown	0.09	0.19	0.21	0.22	0.86	9	256

values, for the sake of comparison. The red lines are fitted by least squares (un-weighted, although weighting by the number of transactions in each j would produce different numbers). For residential land values we get

$$\eta_j^m = \underset{(0.03)}{0.01} + \underset{(0.07)}{0.36} \times \eta_j^r, \quad R^2 = 0.58, \quad n = 23$$

For comparison, if we were to look at commercial properties and land values,

$$\eta_j^m = \underset{(0.10)}{-0.19} + \underset{(0.12)}{0.17} \times \eta_j^r, \quad R^2 = 0.12, \quad n = 18$$

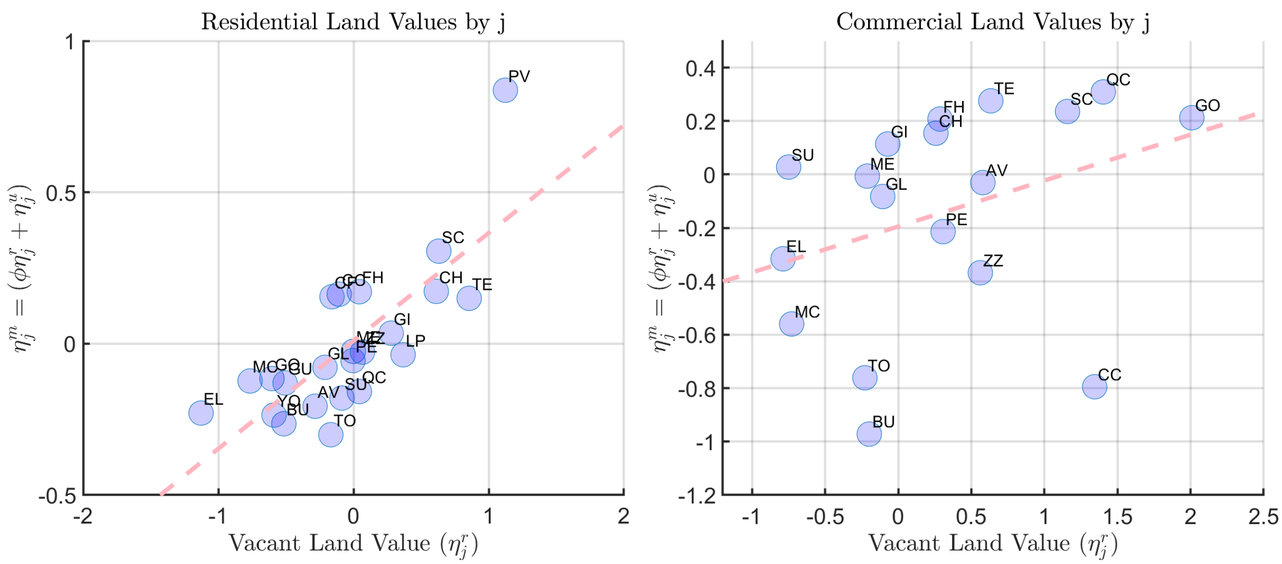
As these are based on a small number of municipalities, the results are imprecise. They do suggest that land makes up a greater share of property costs for residential properties than for commercial. Given the imprecision of the estimates, we focus in the main text on the residential sector, leaving applications incorporating other property types for future work.

C Posterior sampler

Our empirical model can be written as follows

$$\begin{aligned} y_i^v &= (Z_i)'\delta^r + \eta_j^r + (X_i^v)'\beta^v + e_i^v, \quad e_i^v \sim N(0, \sigma_{e,v}^2), \quad \text{for } i = 1, ..., n_v \\ y_k^m &= (Z_k)'\delta^m + \phi\eta_j^r + \eta_j^u + (X_k^m)'\beta^m + e_k^m, \quad e_k^m \sim N(0, \sigma_{e,m}^2), \quad \text{for } k = 1, ..., n_m \end{aligned} \tag{A.1}$$

Figure 8: Unshrunken estimates of η_j^m versus η_j^r



Note that each equation gets a different index i and k, respectively. This is because it is rare that the same lot is sold as a vacant land and improved land within a year. n_v is the number of vacant land sales and n_m is the number of improved land sales.

Reparametrization of the model. The model can then be written as

$$\begin{aligned} y_i^v &= (W_i^v)'b^v + (C_i)'\eta^r + e_i^v, \quad e_i^v \sim N(0, \sigma_{e,v}^2), \quad \text{for } i = 1, ..., n_v \\ y_k^m &= (W_k^m)'b^m + \phi((C_i)'\eta^r) + (C_i)'\eta^u + e_k^m, \quad e_k^v \sim N(0, \sigma_{e,m}^2), \quad \text{for } k = 1, ..., n_m \end{aligned} \tag{A.2}$$

where $W_i^v = [(X_i^v)', (Z_i)']$, $W_i^m = [(X_i^m)', (Z_i)']$, $b^v = [\beta^{v'}, \delta^{v'}]'$, $b^m = [\beta^{m'}, \delta^{v'}]'$, and C_i is a vector of length J with jth element being indicator variable $C_{i,j} = 1$ if i is in city j otherwise it takes 0. Finally, $\eta^r = [\eta_1^r, \eta_2^r, ..., \eta_J^r]$ and $\eta^u = [\eta_1^u, \eta_2^u, ..., \eta_J^u]$, and

$$\begin{aligned} \eta^v &= [\eta_1^r, \eta_2^r, ..., \eta_J^r]' \sim N(0_{J\times 1}, \ \Sigma\left(\sigma_{\eta,r}^2, k_{\eta,r}\right)) \\ \eta^u &= [\eta_1^u, \eta_2^u, ..., \eta_J^u]' \sim N(0_{J\times 1}, \ \Sigma\left(\sigma_{\eta,u}^2, k_{\eta,u}\right)) \end{aligned} \tag{A.3}$$

where (i,j) element of $\Sigma(\sigma_\eta^2, k_\eta) = cov(\eta_i^r, \eta_j^r) = \sigma_r^2 \exp(-d_{ij}/k_r)$ where d_{ij} is the Euclidean distance (in miles) between ith municipality and jth municipality.

Equation (A.2) and (A.3) represent the empirical model with an unknown parameter vector

$$\theta = [b^{v'}, b^{m'}, \sigma_{e,v}^2, \sigma_{e,m}^2, \phi, \sigma_{\eta,r}^2, k_{\eta,r}, \sigma_{\eta,u}^2, k_{\eta,u}]', \quad \eta = [\eta^{v'}, \eta^{u'}]' \tag{A.4}$$

We construct a posterior distribution of θ and η.

Prior distributions. Our prior distribution on unknown model parameters are set to minimize its impact on the posterior distribution. More specifically, we place a prior distribution for b^v and b^m as

$$b^v \sim N(0_{bv}, 10^5 \times I_{bv}), \quad b^m \sim N(0_{bm}, 10^5 \times I_{bm}), \tag{A.5}$$

where 0_n is a zero vector with length n, I_n is $n \times n$ identity matrix, $N(M, V)$ denotes a multivariate normal distribution with mean M and variance-covariance V. The priors for the standard deviation parameters are set to the Half-t distribution,

$$\sigma_{e,v} \sim \text{Half-t}(2, 25),\ \sigma_{e,m} \sim \text{Half-t}(2, 25),\ \sigma_{\eta,r} \sim \text{Half-t}(2, 25),\ \sigma_{\eta,u} \sim \text{Half-t}(2, 25). \tag{A.6}$$

Note that the Half-t distribution is a scale mixture of simpler Inverse-Gamma distributions, and its density is defined as Huang et al. (2013),

$$\text{If } x \sim \text{Half-t}(\nu, A), \text{ then its density is } p(x) \propto \{1 + (x/A)^2/\nu)\}^{-(\nu+1)/2}, \quad x > 0$$

Prior distribution for ϕ is normal distribution with mean 0 and variance 25. Prior distribution for $k_{\eta,r}$ and $k_{\eta,u}$ are set to normal distribution with mean 10 and variance 25. All parameters in θ are independent a-priori. We also obtain posterior distribution of η where its conditional prior $p(\eta|\theta)$ is defined in Eqn (A.3).

Posterior Inference. Then, our posterior distribution is proportional to the product of the likelihood function and prior distribution function,

$$p(\theta, \eta|\mathcal{D}) \propto p(\mathcal{D}|\theta, \eta)p(\eta|\theta)p(\theta)$$

where $\mathcal{D}$ is the data matrix. As the posterior distribution of θ and η is not in a known parametric family, we construct a posterior simulator that generates random draws from this posterior distribution.

Posterior Simulator. Our posterior simulator is a version of a Metropolis-Hastings-within-Gibbs algorithm. We iteratively generate draws from several conditional posterior distributions. Let g be a g-th iteration. Then, we enter the following g-th iteration with the previous parameter draw, $\psi^{(g-1)} = (\theta^{(g-1)}, \eta^{(g-1)})$:

1. $b^v \sim p(b^v|\psi^{(g-1)}_{-b^v}, \mathcal{D})$
2. $\sigma^2_{e,v} \sim p(\sigma^2_{e,v}|\psi^{(g-1)}_{-(\sigma^2_{e,v}, b^v)}, (b^v)^{(g)}, \mathcal{D})$
3. $b^m \sim p(b^m|\psi^{(g-1)}_{-(b^m, b^v, \sigma^2_{e,v})}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, \mathcal{D})$
4. $\sigma^2_{e,m} \sim p(\sigma^2_{e,m}|\psi^{(g-1)}_{-(b^v, \sigma^2_{e,v}, b^m, \sigma^2_{e,m})}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, \mathcal{D})$
5. $\sigma^2_{\eta,r} \sim p(\sigma^2_{\eta,r}|\psi^{(g-1)}_{-(b^v, \sigma^2_{e,v}, b^m, \sigma^2_{e,m}, \sigma^2_{\eta,r})}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, (\sigma^2_{e,m})^{(g)}, \mathcal{D})$
6. $k_{\eta,r} \sim p(k_{\eta,r}|\psi^{(g-1)}_{(b^v, \sigma^2_{e,v}, b^m, \sigma^2_{e,m}, \sigma^2_{\eta,r}, -k_{\eta,r})}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, (\sigma^2_{e,m})^{(g)}, (\sigma^2_{\eta,r})^{(g)}, \mathcal{D})$
7. $\sigma^2_{\eta,u} \sim p(\sigma^2_{\eta,u}|\psi^{(g-1)}_{-(b^v, \sigma^2_{e,v}, b^m, \sigma^2_{e,m}, \sigma^2_{\eta,r}, k_{\eta,r}, \sigma^2_{\eta,u})}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, (\sigma^2_{e,m})^{(g)}, (\sigma^2_{\eta,r})^{(g)}, (k_{\eta,r})^{(g)}, \mathcal{D})$
8. $k_{\eta,u} \sim p(k_{\eta,u}|\psi^{(g-1)}_{-(b^v, \sigma^2_{e,v}, b^m, \sigma^2_{e,m}, \sigma^2_{\eta,r}, k_{\eta,r}, \sigma^2_{\eta,u}, k_{\eta,u})}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, (\sigma^2_{e,m})^{(g)}, (\sigma^2_{\eta,r})^{(g)}, (k_{\eta,r})^{(g)},$ $(\sigma^2_{\eta,u})^{(g)}, \mathcal{D})$

9. $\phi \sim p(\phi|\psi^{(g-1)}_{-(\phi,b^v,\sigma^2_{e,v},b^m,\sigma^2_{e,m},\sigma^2_{\eta,r},k_{\eta,r},\sigma^2_{\eta,u},k_{\eta,u})}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, (\sigma^2_{e,m})^{(g)}, (\sigma^2_{\eta,r})^{(g)}, (k_{\eta,r})^{(g)}, \sigma^2_{(\eta,u})^{(g)}, (k_{\eta,u})^{(g)}, \mathcal{D})$

10. $\eta^r \sim p(\eta^r|\psi^{(g-1)}_{-(b^v,\sigma^2_{e,v},b^m,\sigma^2_{e,m},\sigma^2_{\eta,r},k_{\eta,r},\sigma^2_{\eta,u},k_{\eta,u},\phi,\eta^r)}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, (\sigma^2_{e,m})^{(g)}, (\sigma^2_{\eta,r})^{(g)}, (k_{\eta,r})^{(g)}, (\sigma^2_{\eta,u})^{(g)}, (k_{\eta,u})^{(g)}, \phi^{(g)}, \mathcal{D})$

11. $\eta^u \sim p(\eta^u|\psi^{(g-1)}_{-(b^v,\sigma^2_{e,v},b^m,\sigma^2_{e,m},\sigma^2_{\eta,r},k_{\eta,r},\sigma^2_{\eta,u},k_{\eta,u},\phi,\eta^r,\eta^u)}, (b^v)^{(g)}, (\sigma^2_{e,v})^{(g)}, (b^m)^{(g)}, (\sigma^2_{e,m})^{(g)}, (\sigma^2_{\eta,r})^{(g)}, (k_{\eta,r})^{(g)}, (\sigma^2_{\eta,u})^{(g)}, (k_{\eta,u})^{(g)}, \phi^{(g)}, (\eta^r)^{(g)}, \mathcal{D})$

where $\psi^{(g-1)}_{-x}$ is a $\psi^{(g-1)}$ vector without elements that correspond to x. We initialize the sampler from the individual estimation without shrinkage described in section B. Then, we iterate above steps G times and obtain G parameters $(\psi^{(g)})$, which can be viewed as draws from the posterior distribution $\psi^{(g)} \sim p(\theta, \eta|\mathcal{D})$. We set $G = 80,000$ for the municipal model and $G = 8,000$ for the census-tract model after discarding first 20,000 and 2,000 MCMC draws, respective. We construct our point estimate for a function of some elements in ψ as its posterior distribution, which can be approximated by our simulated draws,

$$\widehat{f(\psi)} = \frac{1}{G}\sum_{g=1}^{G} f(\psi^{(g)}) \to E[f(\psi)|\mathcal{D}]. \tag{A.7}$$

D Details for the posterior sampler

We describe the posterior sampler in detail.

Step 1 and 2 for b^v and $\sigma^2_{e,v}$ This posterior updating can be done by recognizing that generating b^v and $\sigma^2_{e,v}$ from their conditional distribution is equivalent to generating b^v and $\sigma^2_{e,v}$ from the following model with normal prior for b^v and the Half-t prior for $\sigma_{e,v}$,

$$\widetilde{y}^v_i = (W^v_i)'b^v + e^v_i, \quad e^v_i \sim N(0, \sigma^2_{e,v}) \tag{A.8}$$

where $\widetilde{y}^v_i = y^v_i - (C_i)'(\eta^r)^{(g-1)}$. We write a variable without i index as a stacked version of itself. For example $y^v = [y^v_1, y^v_2, ..., y^v_n]'$ and $W^v = [W^v_1, W^v_2, ..., W^v_n]'$.

We first draw b^v given others,

$$(b^v)^{(g)} \sim N(m_1, V_1) \tag{A.9}$$

where

$$V_1 = \left(W^{v\prime}W^v/(\sigma^2_{e,v})^{(g-1)} + V_0\right)^{-1} \tag{A.10}$$

and

$$m_1 = V_1 \times \left(W^{v\prime}\widetilde{y}^v/(\sigma^2_{e,v})^{(g-1)} + V_0^{-1}m_0\right) \tag{A.11}$$

where we write m_0 and V_0 as a prior mean and variance for b^v and m_1 and V_1 as posterior mean and variance.

Conditional on $(b^v)^{(g)}$ and others, we generate $\sigma^2_{e,v}$ from the inverse gamma distribution

$$(\sigma^2_{e,v})^{(g)} \sim IG((\nu_0 + n_v)/2,\ \widehat{S}_1/ + \nu_0 A_1^{-1},) \tag{A.12}$$

where

$$\widehat{S}_1 = \widetilde{y}^{v'}\widetilde{y}^{v} + (b^{v'})^{(g)}W^{v'}W^{v}(b^{v})^{(g)} - 2W^{v'}\widetilde{y}^{v'} \tag{A.13}$$

and

$$A_1^{-1} \sim G\left((\nu_0+1)/2,\ \left(\frac{\nu_0}{(\sigma_{e,v}^2)^{(g-1)}} + 1/A_0^2\right)^{-1}\right). \tag{A.14}$$

where prior for $\sigma_{e,v}^2$ is

$$\sigma_{e,v} \sim \text{Half-t}(\nu_0, A_0) \tag{A.15}$$

G refers to Gamma distribution and IG refers to inverse gamma distribution.

Step 3 and 4 for b^m and $\sigma_{e,m}^2$. It is very similar to step 1 and step 2 described above.

Step 5 $\sigma_{\eta,r}^2$. This is similar to the Half-t updating in step 2. First define

$$\widetilde{\eta}^{v} = chol(R(k_{\eta,r}^{(g-1)}))^{-1}(\eta^{r})^{(g-1)} \tag{A.16}$$

where $R(k_{\eta,r})$ is the correlation matrix implied by $\Sigma(k_{\eta,r})$, and $chol()$ is the Cholesky decomposition that decomposes matrix $X = chol(X)chol(X)'$ where $chol(X)$ is a lower triangular matrix. Then, we have that

$$\widetilde{\eta}_j^{v} \sim i.i.d.\ N(0, \sigma_{\eta,r}^2) \tag{A.17}$$

with $\sigma_{\eta,r} \sim \text{Half-t}(\nu_0, A_0)$. This updating is again given by

$$(\sigma_{\eta,r}^2)^{(g)} \sim IG((\nu_0+J)/2,\ \widetilde{\eta}^{v'}\widetilde{\eta}^{v}/2 + \nu_0 A_1^{-1}) \tag{A.18}$$

where

$$A_1^{-1} \sim G\left((\nu_0+1)/2,\ \left(\frac{\nu_0}{(\sigma_{\eta,r}^2)^{(g-1)}} + 1/A_0^2\right)^{-1}\right) \tag{A.19}$$

Step 6 $k_{\eta,r}$. The conditional posterior distribution of $k_{\eta,r}$ given others is simplified by the following,

$$p(k_{\eta,r}|other, \mathcal{D}) = p(k_{\eta,r}|\eta^{r}, \sigma_{\eta,v}^2), \tag{A.20}$$

and the right hand side term can be written as

$$p(k_{\eta,r}|\eta^{r}, \sigma_{\eta,v}^2) \propto p(\eta^{r}|k_{\eta,r}, \sigma_{\eta,r}^2)p(k_{\eta,r}) \tag{A.21}$$

as long as $p(k_{\eta,r}|\sigma_{\eta,r}^2) = p(k_{\eta,r})$. Note that $p(k_{\eta,r})$ is a prior density function, which is set to be normal density function. The conditional likelihood (or, data-augmented likelihood) function is a multivariate normal density function because

$$\eta^{r} \sim N\left(0,\ \Sigma(k_{\eta,r}, \sigma_{\eta,r}^2)\right). \tag{A.22}$$

We employ Metropolis-Hastings updating with the random-walk proposal,

$$k_{\eta,v}^{new} = k_{\eta,v}^{(g-1)} + c_{k_{\eta,r}}\epsilon, \epsilon \sim N(0,1) \tag{A.23}$$

where $c_{k_{\eta,r}}$ is chosen so that the resulting acceptance probability is approximately between 10% and 40%. This proposal draw is accepted with probability $p_{k_{\eta,r}}$ (i.e., $k_{\eta,r}^{(g)} = k_{\eta,r}^{new}$ with probability $p_{k_{\eta,r}}$ otherwise, $k_{\eta,r}^{(g)} = k_{\eta,r}^{(g-1)}$). The acceptance probability is defined as

$$p_{k_{\eta,r}} = \min\left\{\frac{p((\eta^r)^{(g-1)}|k_{\eta,r}^{new},(\sigma_{\eta,r}^2)^{(g-1)})p(k_{\eta,r}^{new})}{p(\eta^r)^{(g-1)}|(k_{\eta,r})^{(g-1)},(\sigma_{\eta,r}^2)^{(g-1)})p(k_{\eta,r})^{(g-1)})},1\right\}.$$

Step 7 and 8 for $\sigma_{\eta,u}^2, k_{\eta,u}$. These two steps are the same as step 5 and 6. For step 7, we replace $\sigma_{\eta,r}^2$ with $\sigma_{\eta,u}^2$ in step 5. For step 8, we replace $k_{\eta,r}$ with $k_{\eta,u}$ in step 6.

Step 9 ϕ updating is based on the following equation,

$$y_k^m = (W_k^m)'(b^m)^{(g)} + \phi((C_i)'(\eta^r)^{(g-1)}) + (C_i)'(\eta^u)^{(g-1)} + e_k^m, \quad e_k^m \sim N(0,(\sigma_{e,m}^2)^{(g)}), \quad \text{for } k = 1,...,n_m \tag{A.24}$$

Note that b^m and $\sigma_{e,m}^2$ are updated and η^r and η^u are not. We rearrange terms and obtain

$$y_k^m - (W_k^m)'(b^m)^{(g)} - (C_i)'(\eta^u)^{(g-1)} = \phi((C_i)'(\eta^r)^{(g-1)}) + e_k^m, \quad e_k^m \sim N(0,(\sigma_{e,m}^2)^{(g)}), \quad \text{for } k = 1,...,n_m \tag{A.25}$$

And, we take a simple average for each city

$$\underbrace{\frac{1}{n_j}\sum_{k\in\mathcal{I}(j)}\left(y_k^m - (W_k^m)'(b^m)^{(g)} - (C_i)'(\eta^u)^{(g-1)}\right)}_{\widetilde{y}_j^m} = \underbrace{\phi\frac{1}{n_j}\sum_{k\in\mathcal{I}(j)}((C_i)'(\eta^r))^{(g-1)})}_{\phi\widetilde{x}_j^m} + \widetilde{e}_j^m \tag{A.26}$$

where $\mathcal{I}(j) = \{i : y_i \text{ corresponds to land located in area } j\}$ and

$$\widetilde{e}_j^m = \frac{1}{n_j}\sum_{k\in\mathcal{I}(j)} e_k^m \sim N(0,(\sigma_{e,m}^2)^{(g)}/n_j) \tag{A.27}$$

where $n_j = \#\mathcal{I}(j)$. Then,

$$\underbrace{\frac{\widetilde{y}_j}{(\sigma_{e,m})^{(g)}/\sqrt{n_j}}}_{y_j^{m*}} = \underbrace{\phi\left(\frac{\widetilde{x}_j}{(\sigma_{e,m})^{(g)}/\sqrt{n_j}}\right)}_{\phi x_j^{m*}} + e_j^{m*}, \quad e_j^{m*} \sim_{i.i.d} N(0,1) \tag{A.28}$$

with normal prior, $\phi \sim N(m_0, V_0)$. The conditional posterior updating for ϕ is similar to the one in step 1. For those municipalities with no observation, we eliminate corresponding rows from y_j^{m*} and x_j^{m*} before we do the updating.

Step 10, η^r We have

$$y_k^m = (W_k^m)'(b^m)^{(g)} + \phi^{(g)}((C_k)'(\eta^r)^{(g-1)}) + (C_k)'(\eta^u) + e_k^m, \quad e_k^m \sim N(0,(\sigma_{e,m}^2)^{(g)}), \quad \text{for } k = 1,...,n_m \tag{A.29}$$

We move some terms on the right-hand-side to the left and apply the city level average,

$$\frac{1}{n_j}\sum_{k\in\mathcal{I}(j)}\left(y_k^m-(W_k^m)'(b^m)^{(g)}-((C_k)'(\eta^u)^{(g-1)})\right)=\phi^{(g)}\eta_j^v+\frac{1}{n_j}\sum_{k\in\mathcal{I}(j)}e_k^m \tag{A.30}$$

Divide both sides by $\sqrt{(\sigma^2_{e,m})^{(g)}/n_j}$ to obtain

$$\left(\sqrt{n_j}/\sqrt{(\sigma^2_{e,m})^{(g)}}\right)\times\frac{1}{n_j}\sum_{k\in\mathcal{I}(j)}\left(y_k^m-(W_k^m)'(b^m)^{(g)}-((C_k)'(\eta^u)^{(g-1)})\right)=\left(\phi^{(g)}\sqrt{n_j}/\sqrt{(\sigma^2_{e,m})^{(g)}}\right)\times\eta_j^v+e_j^m \tag{A.31}$$

Then the above equation has the following form

$$\widetilde{y}_j^m=z_j^m\eta_j^r+e_j^m,\quad e_j^m\sim_{i.i.d.}N(0,1) \tag{A.32}$$

Stacking this equation over all $j=1,2,...,J$, we get

$$\widetilde{y}^m=\widetilde{Z}^m\eta^r+e^m,\quad e^m\sim_{i.i.d.}N(0,I) \tag{A.33}$$

where $\widetilde{Z}^m=diag([z_1^m,z_2^m,...,z_J^m]')$.

Similarly, we have

$$\left(\sqrt{n_j}/\sqrt{(\sigma^2_{e,v})^{(g)}}\right)\times\frac{1}{n_j}\sum_{i\in\mathcal{I}(j)}\left(y_i^v-(W_i^v)'(b^v)^{(g)}\right)=\left(\sqrt{n_j}/\sqrt{(\sigma^2_{e,v})^{(g)}}\right)\times\eta_j^v+e_j^v \tag{A.34}$$

We write above equation as

$$\widetilde{y}_j^v=z_j^v\eta_j^r+e_j^v,\quad e_j^v\sim_{i.i.d.}N(0,1), \tag{A.35}$$

which is

$$\widetilde{y}^v=\widetilde{Z}^v\eta^r+e^v,\quad e^v\sim_{i.i.d.}N(0,I) \tag{A.36}$$

Starting from the conditional prior

$$\eta_0^r\sim N\left(0,\ \Sigma((\sigma^2_{\eta,r})^{(g)},(k_{\eta,r})^{(g)})\right), \tag{A.37}$$

we compute the posterior distribution of η^r based on the following state space representation

$$\begin{aligned}\widetilde{y}_t&=\widetilde{Z}_t\eta_t^r+e^v,\quad e^v\sim_{i.i.d.}N(0,I)\\ \eta_t^r&=\eta_{t-1}^r\end{aligned} \tag{A.38}$$

where $t=1,2$ and

$$\left(\widetilde{y}_1=\widetilde{y}^v,\ \widetilde{Z}_1=\widetilde{Z}^v\right),\ \left(\widetilde{y}_2=\widetilde{y}^m,\ \widetilde{Z}_2=\widetilde{Z}^m\right). \tag{A.39}$$

We break down the transformed data set into two pieces, $[\widetilde{y},\widetilde{Z}]=\{[\widetilde{y}^v,\widetilde{Z}^v],[\widetilde{y}^m,\widetilde{Z}^m]\}$, and update the posterior distribution of η_t^r sequentially as if data are realized piece by piece. This makes computation of the posterior distribution straightforward as the Kalman filter computes mean and covariance matrix of the following conditional probability densities,

$$\begin{aligned} p(\eta^r \mid \widetilde{y}^v, \widetilde{Z}^v) &= p_N(\eta^r | m_{1|1}, V_{1|1}) \\ p(\eta^r \mid \widetilde{y}^v, \widetilde{Z}^v, \widetilde{y}^m, \widetilde{Z}^m) &= p_N(\eta^r | m_{2|2}, V_{2|2}) \end{aligned} \tag{A.40}$$

where $p_N(x|m, V)$ denotes a density function of the multivariate normal distribution with mean m and covariance matrix V.

Then, we obtain the desired conditional posterior distribution for this step,

$$\eta^r | \mathcal{D}, others \sim N(m_{2|2}, V_{2|2}) \tag{A.41}$$

where $m_{2|2}$ and $V_{2|2}$ are from the Kalman filter based on the state space representation presented in (A.38), which are updated posterior mean and variance-covariance matrix of η_2^r given $[\widetilde{y}, \widetilde{Z}]$.

Step 11, η^u We have

$$y_k^m = (W_k^m)'(b^m)^{(g)} + \phi^{(g)}((C_k)'(\eta^r)^{(g-1)}) + (C_k)'(\eta^u) + e_k^m, \quad e_k^m \sim N(0, (\sigma_{e,m}^2)^{(g)}), \quad \text{for } k = 1, ..., n_m \tag{A.42}$$

We move the first two terms on the right-hand-side to the left and apply the city level average,

$$\frac{1}{n_j} \sum_{k \in \mathcal{I}(j)} \left(y_k^m - (W_k^m)'(b^m)^{(g)} - \phi^{(g)}((C_k)'(\eta^r)^{(g-1)}) \right) = \eta_j^u + \frac{1}{n_j} \sum_{k \in \mathcal{I}(j)} e_k^m \tag{A.43}$$

Divide both sides by $\sqrt{(\sigma_{e,m}^2)^{(g)}/n_j}$ to obtain

$$\left(\sqrt{n_j} / \sqrt{(\sigma_{e,m}^2)^{(g)}} \right) \times \frac{1}{n_j} \sum_{k \in \mathcal{I}(j)} \left(y_k^m - (W_k^m)'(b^m)^{(g)} - \phi^{(g)}((C_k)'(\eta^r)^{(g-1)}) \right) = \left(\sqrt{n_j} / \sqrt{(\sigma_{e,m}^2)^{(g)}} \right) \times \eta_j^u + e_j^m \tag{A.44}$$

Then the above equation has the following form

$$\widetilde{y}_j^m = z_j^m \eta_j^u + e_j^m, \quad e_j^m \sim_{i.i.d.} N(0, 1) \tag{A.45}$$

Stacking this equation over all $j = 1, 2, ..., J$, we get

$$\begin{aligned} &\widetilde{y}^m = \widetilde{Z}^m \eta^u + e^m, \quad e^m \sim_{i.i.d.} N(0, I) \\ &\eta^m \sim N\left(0,\ \Sigma((\sigma_{\eta,u}^2)^{(g)}, (k_{\eta,u})^{(g)})\right) \end{aligned} \tag{A.46}$$

where $\widetilde{Z}^m = diag([z_1^m, z_2^m, ..., z_J^m]')$. Posterior updating for η^m is similar but simpler version of step 10.