Journal of Real Estate Research

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rjer20

A Problem with Distance Variables and Alternatives for Their Use

Michael C. Farmer, Syed Fuad, Kusum J. Naithani & Donald J. Lacombe

To cite this article: Michael C. Farmer, Syed Fuad, Kusum J. Naithani & Donald J. Lacombe (29 Mar 2024): A Problem with Distance Variables and Alternatives for Their Use, Journal of Real Estate Research, DOI: 10.1080/08965803.2024.2325244

To link to this article: https://doi.org/10.1080/08965803.2024.2325244

JOURNAL OF REAL ESTATE RESEARCH
https://doi.org/10.1080/08965803.2024.2325244

A Problem with Distance Variables and Alternatives for Their Use

Michael C. Farmer[a], Syed Fuad[a], Kusum J. Naithani[b] and Donald J. Lacombe[c]

[a]Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, TX, USA; [b]Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA; [c]School of Financial Planning, Texas Tech University, Lubbock, TX, USA

ABSTRACT
Researchers commonly use distance variables to: (i) estimate the direct influence of a landmark on an outcome of interest, such as a neighborhood park on home price; or (ii) control for omitted spatial influences that affect predictions of key policy variables. While both uses continue, the use of distance as a control, such as distance to Central Business District (CBD), is now more common. Using distance to a given position such as CBD is added to multivariate analysis as a method to capture all remaining, or omitted spatial effects that influence the dependent variable. We show that there is a latent and inherent identification problem with the distance variable; and we show that this extends to the use of distance as a control. These biases affect more than the distance variable. They generate inconsistent estimates for *all* other spatially distributed variables in a model. We then introduce an alternative control that captures unmodeled influences that vary across space, and we show that this fully stabilizes all model parameter estimates and measures of model efficiency.

KEYWORDS
Hedonic modeling; distance variables; CBD; spatial estimation; multicollinearity; regional policy

JEL CODES
R30; R31; R20; C21; C18; P25

Introduction

The regional sciences often have used distance variables to explain spatially distributed events. Distance variables have analyzed issues as diverse as the nesting choices of ground birds to issues that affect residential home purchases of young families. It is hypothesized, respectively, that the nesting choices of a bird or the home purchases of a young family respond to the distance of the bird nest to a windmill (Grisham et al., 2014; Zuta et al., 2012) or the distance of a family home to a school (Metz, 2015).

Distance variables make an attractive baseline for statistical modeling. In the case of home choice, proximity to various amenities clearly motivates housing choices and price. Over the decades, there is a tradition to estimate home choice response to the distance to a Superfund site (Ihlanfeldt & Taylor, 2004; Taylor et al., 2016), an environmental hazard (Brasington & Hite, 2005), a historic district (Noonan et al., 2007), or an employment center (Harrison & Rubinfeld, 1978). More recently, Letdin and Shim (2019) compare the value of distance to work versus distance to urban amenities to sort home location

CONTACT Syed Fuad syed.m.fuad@ttu.edu

choices among demographically distinct households. Nase et al. (2019) use distance to the nearest train station and highway exit to explain rental values of Amsterdam office buildings; and Tontisirin and Anantsuksomsri (2021) use distance to multiple landmarks to isolate real estate use responses to access to Bangkok transit stations. Other notable and recent papers that use distance variables are Dumm et al. (2016), Robinson et al. (2017), Lawani et al. (2019), Dauth and Haller (2020), Sun et al. (2021), van Vuuren (2023), and Abbiasov and Sedov (2023).

Commonly, the distance to the Central Business District (CBD) is employed to control for the omitted effects of position on parameter estimates of spatially distributed explanatory variables, such as home age or crime rate. Examples include: Schuetz (2009), Meltzer and Schuetz (2010), Osland (2010), and Dermisi and McDonald (2010), Yilmazkuday and Yilmazkuday (2016), Jayet et al. (2016), Salvati et al. (2016), and Nguyen and Diez (2017). More recently, distance to CBD was used as a control in Borck and Schrauth (2021) and Blanco (2023).

The literature does suggest some discomfort with the distance variable. Evidence by Palmquist's (2005) use of a discrete zonal (0/1) variable to assess home price differentials around a Superfund site by proximity within one mile (or not) replaces a continuous distance variable. Subsequently, Cameron (2006) explicitly challenges the use of continuous distance variables as a spatial control. On logical grounds, distance to the CBD for example makes little sense if industrial areas 8 miles southeast of CBD were compared to residential neighborhoods 8 miles to the northeast, even controlling for other explanatory variables. Finally, Ross et al. (2011) used constructed data to test distance variable consistency. Simulations illustrated ubiquitous instability in parameter values of distance variables as more were added to a model. Taken together, Cameron (2006) and Ross et al. (2011) point to a concern for omitted variable bias leading to inconsistent parameter estimates of all spatially distributed variables. It is not an efficiency concern *per se*.

Perhaps due to these concerns, there has been a decline in the use of distance variables as the primary causal or policy variable in a model. Yet the alternative use of a distance variable to control for unmodeled spatial influence continues to enjoy currency (see Bondemark & Merkel, 2023; Borck & Schrauth, 2021; Diao et al., 2023). Also, other works extend Palmquist's dichotomous zone variable by adopting multiple zones: Diao et al. (2023) and Bogin and Doerner (2019) employ, respectively, three dichotomous zones – within 0–1, 1–2, or 2–3 kilometers from a landmark, or two zones – within 0–5, or 5–15 miles from a landmark. We submit these extensions are merely discretized distance variables subject to the same raft of concerns. Though these works reveal discomfort with the use of continuous distance variables, they leave an incomplete understanding of the problems with distance variables. Though use of continuous distance variable has declined, the distance variable remains in common use. Our scan of regional science journals finds one or more articles in nearly every issue since 2020 that use a distance variable as a meaningful estimate of the value of proximity to a given landmark. These include high impact journals; and articles that use a distance variable are present among most cited recent works.

Returning to Cameron's (2006) observation that measured distance is free of direction, it follows that the direction free Euclidean distance makes a poor control for unmodeled spatial information. After all, Euclid deliberately constructed the distance formula to abstract away from position. So, in the search for a valid control to capture omitted spatial influences directly defined by position, the control must capture relative direction between any two positions – not distance.

The control we construct unpacks the Euclidean distance formula into its four nested variables. Consider the measured distance from point A to B *(Distance$_{AB}$)*, in long-lat space as:

$$Distance_{AB} = \left[(Longitude_A - Longitude_B)^2 + (Latitude_A - Latitude_B)^2\right]^{\frac{1}{2}}$$

This formula embeds four nested variables that measure relative position between any two points based on direction: difference in longitude and in latitude, and the square of each:

$$\left\{(long_A - long_B); (long_A - long_B)^2; (lat_A - lat_B); (lat_A - lat_B)^2\right\}$$

A consistent control for omitted, spatially correlated, influences that affect the dependent variable should allow the fixed position to be changed without any change to other model parameters. As the control only needs to account for omitted influences due to relative position, we expect a control to operate effectively at any position. All four variables above can anchor to a different fixed position. Our prediction is that all parameter estimates remain the same except for the control variables themselves. We illustrate this correction through two empirical examples.

The empirical work concludes by integrating the correction with spatial econometrics. The spatial lagged explanatory variable (SLX) model is added. The SLX adds a new set of independent variables to the model; so, the outcome is qualitatively identical. All parameter values remain the same, including lagged variables, regardless of the fixed point chosen. Yet, the fixed point corrections are beyond the spillover influences of adjacency *per se* and focus on the impacts of specific position within the data set.

This control strategy has wide implications for the regional sciences and affiliated disciplines. Clearly, the key policy variable or variable of interest in these disciplines is almost invariably spatially distributed, e.g., crime rates, school quality. The correction herein captures unmodeled (omitted) cross-correlated spatial effects among the independent variables to secure consistent parameter estimates for all spatially distributed variables in a model. This is a powerful procedural tool that permits a clean sorting of the weighted influences of each independent variable on the dependent variable, corrected for omitted spatial effects. This is especially relevant if the spatially distributed independent variable is the key policy variable under examination.

Concerns for Distance Variables

There is seldom a single feature over the landscape of a study area where its proximity influences home price. Rather, it is reasonable to expect that the distance to one type of landmark will co-vary with the distance to many others. We expect, for example, proximity to a recreational park to co-vary with proximity to amenities common to residential neighborhoods, but not to industrial processing centers, and vice versa.

This is where the problem arises. For a distance variable to be identified in a multivariate model, the measured distance to that landmark from a subject observation must exhibit truly independent variation from every other measured distance to every other landmark. The problem with distance variables is that there is no *independent* variation among them. The shift in measured distance to one landmark *precisely* predetermines the change in measured distance to every other. This is easily apprehended through two examples.

Consider a home positioned equidistant between a well-appointed park and a toxic dump site. A home price change between any two points along this line might be measured with perfect accuracy in a multivariate model. For example, a one-unit move in either direction may alter home price by exactly \$700. Though the total effect of this move may be estimated with perfect accuracy, it is impossible to partition the total effect between the two influences. A move closer to the park may be the result of a \$100 benefit of being closer to the park and a \$600 benefit of being further from the dump; or a \$600 benefit of being closer to the park and a \$100 benefit of being further from the dump. The two cannot be distinguished with the information available because change in the distance to the park and distance to the dump co-vary by an exact process.

A visual extension of this example is shown in Figure 1. Consider a home sale at position O and another at position P. Positions A, B, and C are landmarks, and each positively influences home value. As home position varies, we cannot partition the marginal value change from position O to position P for each distance to A, B, or C. It may be useful to consider O as the sample mean longitude and latitude position of a home valued at the average home price in the sample. In that way, a move from O to P is the expected change in overall value in a move from position O toward C. Yet like the park-and-toxic dump problem, this estimated value change in moving closer to C is also due to changing distances away from B and A. The independent effects are not identified.

As Figure 1 illustrates, the observed change in home sale price at P, which is closer to C, precisely determines the change in measured distance to A and B. There is simply no random or independent functional relationship among proximity among these three points; or no independent variation among distance variables exists – the very definition of non-identification. Any estimation of distances to points A and B are fully determined by the change in distance to C.

An important observation is that the overall prediction accuracy of different models would be unchanged from model to model. That is, the \$700 gain of one unit closer to the park or the price change akin to value change from points O to P above are estimated with the same accuracy, regardless of the array of distanced variables used. This

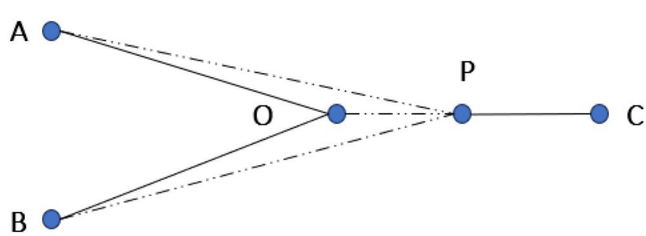

Figure 1. Graphical representation of the interdependence among distance variables.

is an intuitive explanation that the problem is not one of model efficiency but more a problem of omitted variable inconsistency.

Finally, this lack of independence between distance variables extends to high covariance among all spatially distributed independent variables. To illustrate the problem, consider homes of a different age. Older homes are positioned closer to downtown; and they are also more valuable. A failure to craft some control for location of this spatially distributed variable may be due only to downtown proximity or designation as an historical district (Noonan et al., 2007). Only a control for location will avert a spurious conclusion that aging houses are intrinsically more valuable.

Non-Identification of Distance Variables

Formally, the above examples point to the non-identification of distance variables. We formally define this only to establish testable hypotheses for the empirical examples. Recall, a model variable is unidentified if it provides no new information about variation in the dependent variable. Once one distance variable enters a model, no new information about variation in the dependent variable is available by adding others, that is, $(\Pr Y|X_1) = (\Pr Y|X_1, X_2)$. Non-identification arises when the probability of Y given X_1 is unchanged by the addition of X_2, such as different mixes of distance to A, B, and C above. Technically this condition indicates a 'collapse of the design matrix', which means the distance variable is unidentified.

We underscore that the lack of unique information when adding a distance variable to a model is not to be confused with multicollinearity, which occurs when there is a scalar relationship between one or more columns of the explanatory variable matrix. In perfect multicollinearity, the OLS estimator cannot be used because the $(X'X)^{-1}$ matrix is not invertible. However, a functional relationship can exist between two columns of the design matrix, such as age and age squared of a house, that typically does allow $(X'X)^{-1}$ to be inverted. So, the core definition of nonidentification is satisfied if $(\Pr Y|X_1) = (\Pr Y|X_1, X_2)$, even if $(X'X)^{-1}$ inverts. Once the position of activity, Y, and distance to landmark, X_1, is known, no new information about the activity at position Y is available from the addition of other distance variables. Simply, that is because there is no independent variation between distances to landmarks X_1 and X_2. This is borne out by our empirical results. Modeled parameters change as new distance variables are added to a model even though explanatory power does not meaningfully change, if at all.

The Appendix A reviews the trigonometric identity whereby the longitude and latitude position of one point can be defined directly by the longitude and latitude positions of any other two points. That is, information about the longitude and latitude of the dependent variable (i.e., its position) and longitude and latitude of one independent (distance) variable precisely defines the longitude and latitude of any other landmark and no independent information exists.

Control for Position: A Spatial Fixed Point

Deaton and Hoehn (2004) presented a well cited example of the use of distance to the central business district (CBD) as a control. The variable was used as a control for broader effects that impact the value of a park to home values in multiple neighborhoods across a large metropolitan area. The authors intended to isolate the effects of this neighborhood amenity in otherwise similar neighborhoods, except for the neighborhood location. The logic is sound, but the variable is not identified; and, therefore, this strategy continues to lead to inconsistent estimates, the sole purpose for its application. We do retain the logic of this practice and develop a consistent control below.

Figure 2 offers an intuitive explanation why a control for position requires the four variables used. It also helps to explain why a large change in explanatory power does not result – or why the fixed point corrects for position as a strategy to capture omitted spatial influences on parameters than on overall explanatory power. The base records long-lat position, and the height records expected value from an estimated mode, in this case sales price. So, each height-position point represents the expected value attributable to position from an estimated model. If that model includes spatially distributed variables such as crime rate or tree cover, parameter estimates of these variables will be fitted to capture positional information as well. The goal, especially for the policy variable of interest, is to disentangle this effect from parameter estimates for every model variable. It is this omitted information that forms the chief rationale for the use of distance as an anchor, such as CBD. Yet a distance variable can make a partial correction for overall position because, as we have seen, distance is not oriented by direction.

The contours of this value map are not surprising. For convenience, *Mean* is the expected value at the mean position for every independent variable from which any regression estimate is calibrated. *Optimal* is merely a position of convenience to show the highest expected value in the data set. Clearly, value change is not linear across the distance δ from Mean to Optimal. Indeed, virtually all positions of distance δ realize a different marginal value change, even though the model parameters exert some control for differences in crime rate, tree cover and home age. Again, injecting distance from Mean as a model variable to any other position does not control this. In both linear regression or maximum likelihood, a control that passes through *Mean* must account for overall long-lat position and capture the curvature of expected value generated by the model.

On Figure 2, consider the value change from the small total downward latitude change to Landmark from Mean, and consider the longer longitudinal movement to the right to Landmark. Moving downward to the lower latitude of Landmark from Mean, value decreases at an increasing rate of decent. Yet moving to the right, longitude value

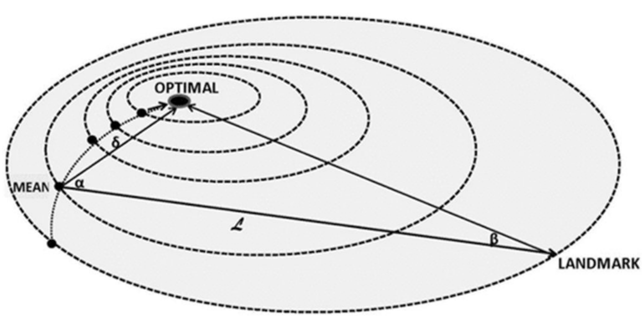

Figure 2. Graphical representation of the value changes from the mean position to any landmark decomposed into second order changes in longitude and latitude.

change is at first slightly increasing, but then slowly decreases; or value change by longitude would be positive and then would be negative. Clearly, the four position variables will capture any value change attributable to position.

Every point will generate an expected value of latitude and longitude difference based on its position. The average change for each of the four directional variables will be fitted to the model to estimate all parameters. Of note, as the reference position changes, say from Mean to Optimal these four positional parameters change, but every position will control for the same average change in value between positions. As a result, only one reference position, that is, only one set of the four directional control variables is sufficient; addition of more than one reference positions *simultaneously* in a single model will result in perfect linear dependency among the two sets of control variables. Moreover, all other parameters values and efficiency measures would be unchanged. Adding these four control variables captures available home price variation attributed to omitted spatial position.

This correction allows two very strong predictions. Given the spatially distributed model attribute variables, adding the four directional variables will control for overall position by generating the same estimated coefficient values for every model variable regardless of the fixed reference point used. Secondly, the control for this omitted spatial variation in a model will leave unchanged values of model efficiency, such as R^2, Log Likelihood, or AIC as fixed position changes.

Empirical Examples

We selected data sets with the intention to cover extreme ranges: high resolution, low numbers; and low resolution, large numbers. One data set has only 365 observed sales in Lubbock, Texas and covers an area only 4.8 miles wide and 8 miles long. This data set has direct observation of demographics via survey. The other has 13,327 observed sales in Columbus, Ohio; yet local demographics rely on the 2000 census with data aggregated at the census tract level, when the average census tract size was 4,000 households; so, this data has very low demographic resolution. Concern from several readers and conferences suggested that large, high resolution data sets may realize a spurious response to this control. In other words, what makes the correction work is the high quality of data as spatial variation is already well fitted to the existing model. Our goal is to demonstrate the range of the fixed point correction as a general arithmetic principle, not to draw causal assertions from these results.

Lubbock, Texas

Data in Lubbock are 365 residential sales from June 2006 to December 2008 reported in Farmer et al. (2012). MLS data includes sales price, square footage, lot size, house age, presence of a garage. We also collect household level demographic data including income, family size, owner occupancy of residence, employment status, etc.

Panel A in Figure 3 maps the study area of seventeen neighborhoods, named by the Lubbock Realtor Association and scattered across a five-by-eight-mile area. The population is *prima facie* homogeneous: 83% white; mean family size is 3.85; and incomes fall

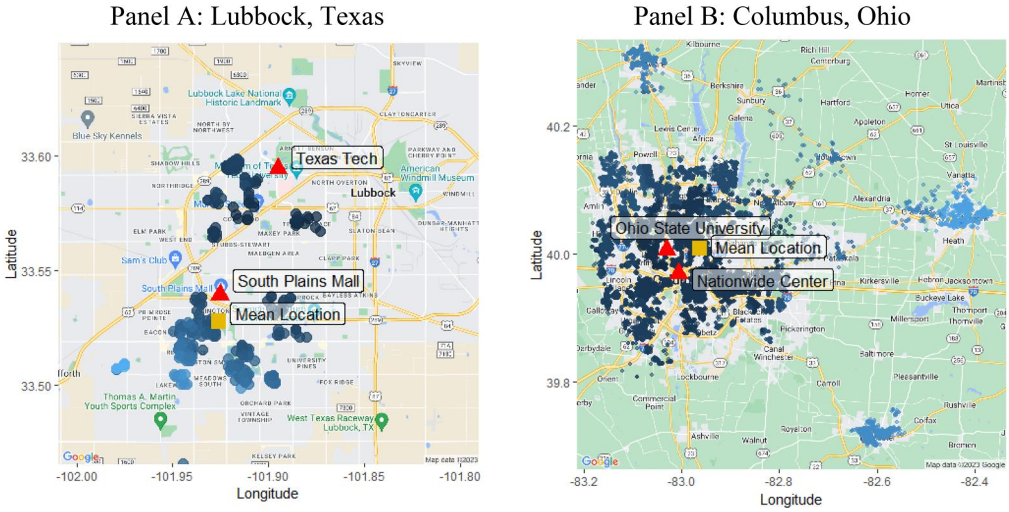

Figure 3. Map of study areas showing the locations of residential sales, the mean position of all residential sales (yellow box), and two landmarks (red triangles).
Note: Shades of blue of residential locations indicate average distance from the two landmarks (darker shades correspond to closer 'average' proximity to landmarks).

within the 40 to 95 of national percentiles. Homes are detached single family units. More than 80% are owner-occupied. Unemployment rates are 2%; and 58% of households had two working adults. Panel A in Table 1 presents the summary statistics of the variables from Lubbock, TX.

Panel A in Figure 3 shows two prominent locations: Texas Tech University (Tech), which is a strong employment draw and close to the city center; and South Plains Mall (Mall), which is an employment center and regional retail shopping area. Distance from each property to each landmark was defined by longitude and latitude distances, or simple Euclidean distance.

Columbus, Ohio

Data for Columbus is a sample of 13,327 single-family detached house sales in central Ohio in 2000. The data have been used in multiple studies prior to the release of the 2010 census. These include Brasington (2007), Brasington and Sarama (2008), and FARES (2002). The original data set records sales prices from five metropolitan areas in the greater Columbus area as well as home age, square feet, presence of a second story and demographic census tract averages, racial heterogeneity, income, and crime levels. Panel B in Table 1 presents the summary statistics of the variables from Columbus, OH.

Panel B in Figure 3 shows two prominent locations in Columbus: Ohio State University (OSU), which is a strong employment draw; and the Nationwide Center (NWD), which is an employment and business center with some retail shopping and recreation. The observations in the central ring are in Franklin County, and far more diverse than the neighborhoods of Lubbock, Texas. So, we test datasets of different sizes, different resolutions, and different levels of diversity.

Table 1. Summary statistics of two data sets.

	Panel A: Lubbock, TX				Panel B: Columbus, OH			
	Mean (S.D.)	Median	Max	Min	Mean (S.D.)	Median	Max	Min
Sales price ($)	228,598 (120,081)	197,000	1,050,000	61,000	136,988 (96907)	119,000	2,225,001	30,000
Square foot	2,711.06 (933.10)	2,550	6908	974	1,601.53 (648.82)	1,452	9,933	29
Lot size (Sq. yard)	10,747.53 (9,826.97)	8,964	150,000	887	11,886.65 (26,511.79)	7,940	940,025	194
House age (Years)	22.04 (16.72)	19.00	69.00	1.00	38.71 (30.94)	35.00	200.00	0.00
Income ($)	104,115 (39,548)	100,000	150,000	40,000	53,360 (23,651)	49,583	200,001	6,136
Garage (0/1)	0.964 (0.186)	1.00	1.00	0.00	–	–	–	–
Env. Proxy	34.22 (29.74)	25.97	126.32	0.00	–	–	–	–
Second story (0/1)	–	–	–	–	0.587 (0.492)	1.00	1.00	0.00
Pct white (%)	–	–	–	–	80.42 (22.63)	88.92	100.00	0.00
Offenses per district	–	–	–	–	97.51 (63.35)	127.05	735.34	2.45
Distance to Tech (decimal degree)	0.082 (0.091)	0.085	1.737	0.011	–	–	–	–
Distance to Mall (decimal degree)	0.040 (0.087)	0.033	1.682	0.009	–	–	–	–
Distance to OSU (decimal degree)	–	–	–	–	0.161 (0.138)	0.121	0.668	0.010
Distance to NWD (decimal degree)	–	–	–	–	0.163 (0.131)	0.131	0.672	0.008
Longitude (degree)	−101.925 (0.029)	−101.927	−101.873	−102.296	−82.96 (0.17)	−82.99	−82.37	−83.65
Latitude (degree)	33.528 (0.091)	33.522	33.599	31.898	40.01 (0.10)	40.02	40.32	39.70
N	*365*				*13,327*			

Econometric Results and Discussion

Lubbock, Texas

Shiroya (2010) examined if the Lubbock data could embed discernable submarkets despite its very small numbers and remarkable, prima facie demographic homogeneity. Shiroya hypothesized that an environmental variable, *progressed landscape* – a diverse vegetation structure from short bushes, short trees and a higher tree canopy (MacArthur & MacArthur, 1961) might sort otherwise homogeneous data into submarkets based on this taste parameter, following Sieg et al. (2002) who sorted housing choice over a very large, multiple city area in southern California. Field methods used for the progressed landscape are described in Farmer et al. (2012) and Leuenberger (2015). We use this environmental variable, clearly spatially distributed, to assess if the fixed point correction could parse remaining spatial information over such a small space and small sample size with seemingly homogenous residents, regardless of position of the fixed point correction.

Columbus, Ohio

The challenge for the correction in Columbus is much the opposite. The dual policy variables in Columbus track the separate influences of racial composition and crime rates on home price. Demographic data are at the resolution of the 2000 census tract, far lower than the house-by-house demographics in Lubbock, TX. More importantly, wealth and home age reach some of their highest values in the older cities that ring Columbus, so there is no obvious spatial path through the study area of these variables. In that sense, the concern is that the correction might not harmonize parameters and fail as a control to correct to consistent parameters.

Key Results

Tables 2 and 3 compare six hedonic home price models in Lubbock, TX and Columbus, OH, respectively. The first model uses only home and neighborhood attributes with no distance variables. The second and third models use a different distance variable to different positions in the data set; the fourth model uses both distance variables; and the fifth and sixth models use the fixed point correction, alternatively anchored to each of the two positions as a control for omitted spatial effects. Five critical tests are evaluated:

- We expect parameter estimates of spatially distributed variables to stabilize only under models with fixed position controls (models 5 and 6).
- We expect parameter estimates of non-spatially distributed variables such as square footage or presence of a second story or garage to be relatively stable across models for each city.
- We expect measures of asymptotic efficiency $(Adjusted\ R^2, AIC, \mathrm{Log}Likelihood)$ to converge across all models whether they include one distance variable, no distance variable, or the fixed point correction; and convergence will strengthen asymptotically with sample size.

Table 2. Model comparison for Lubbock, Texas.

	Model 1: Baseline	Model 2: Distance (Tech)	Model 3: Distance (Mall)	Model 4: Distance: Mall&Tech	Model 5: Fixed Point 1 Tech	Model 6: Fixed Point 2 Mall
Square Foot	1.048*** (0.041)	1.047*** (0.041)	1.051*** (0.041)	1.041*** (0.040)	1.027*** (0.039)	1.027*** (0.039)
Lot size (Sq. yard)	0.119*** (0.027)	0.118*** (0.028)	0.122*** (0.028)	0.109*** (0.027)	0.115*** (0.026)	0.115*** (0.026)
House Age (Years)	−0.120*** (0.011)	−0.120*** (0.012)	−0.120*** (0.011)	−0.137*** (0.011)	−0.143*** (0.012)	−0.143*** (0.012)
Garage (0/1)	0.021 (0.020)	0.022 (0.020)	0.020 (0.020)	0.039** (0.019)	0.040** (0.019)	0.040** (0.019)
Env. Proxy birdsXspecies	0.013 (0.008)	0.013 (0.008)	0.012 (0.008)	−0.003 (0.008)	−0.0002 (0.008)	−0.0002 (0.008)
Income ($)	0.085*** (0.026)	0.085*** (0.026)	0.086*** (0.026)	0.087*** (0.025)	0.078*** (0.024)	0.078*** (0.024)
Dist. (Tech) (decimal degree)	–	−56.236 (110.836)	–	−1.852*** (0.350)	–	–
Dist. (Mall) (decimal degree)	–	–	0.125 (0.113)	1.933*** (0.359)	–	–
Long (Tech) (decimal degree)	–	–	–	–	7.814*** (1.058)	–
Lat (Tech) (decimal degree)	–	–	–	–	0.838** (0.392)	–
$Long^2$ (Tech) (decimal degree)	–	–	–	–	60.975*** (8.207)	–
Lat^2 (Tech) (decimal degree)	–	–	–	–	−2.245*** (0.460)	–
Long (Mall) (decimal degree)	–	–	–	–	–	1.609*** (0.467)
Lat (Mall) (degree)	–	–	–	–	–	1.041*** (0.372)
$Long^2$ (Mall) (decimal degree)	–	–	–	–	–	60.975*** (8.207)
Lat^2 (Mall) (degree)	–	–	–	–	–	−2.245*** (0.460)
AIC_c	214.8401	217.10318	218.0762	247.7204	284.5346	284.5346
LogLik	−99.4200	−99.55159	−100.0381	−113.8602	−130.2673	−130.2673
Adj. R^2	0.842	0.841	0.842	0.853	0.865	0.865

Note: 365 observations. Robust standard errors are in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

Table 3. Model comparisons for Columbus, Ohio.

	Model 1: Baseline	Model 2: Distance (OSU)	Model 3: Distance (NWD)	Model 4: Distance: OSU&NWD	Model 5: Fixed Point 1 OSU	Model 6: Fixed Point 2 NWD
Square Foot	0.774*** (0.011)	0.753*** (0.010)	0.754*** (0.010)	0.755*** (0.010)	0.756*** (0.010)	0.756*** (0.010)
Lot size (Sq. yard)	0.042*** (0.005)	0.061*** (0.005)	0.060*** (0.005)	0.058*** (0.005)	0.048*** (0.005)	0.048*** (0.005)
House Age (Years)	−0.032*** (0.002)	−0.041*** (0.002)	−0.040*** (0.002)	−0.039*** (0.002)	−0.037*** (0.002)	−0.037*** (0.002)
Second Story (0/1)	−0.053*** (0.006)	−0.043*** (0.006)	−0.043*** (0.006)	−0.046*** (0.006)	−0.044*** (0.006)	−0.044*** (0.006)
Income ($)	0.434*** (0.009)	0.367*** (0.009)	0.383*** (0.009)	0.351*** (0.009)	0.369*** (0.009)	0.369*** (0.009)
Offenses per District	−0.046*** (0.005)	−0.090*** (0.005)	−0.086*** (0.005)	−0.086*** (0.005)	−0.084*** (0.005)	−0.084*** (0.005)
Pct White (%)	0.057*** (0.005)	0.079*** (0.005)	0.080*** (0.005)	0.071*** (0.005)	0.081*** (0.005)	0.081*** (0.005)
Dist. (OSU) (decimal degree)	–	−0.493*** (0.020)	–	−1.245*** (0.094)	–	–
Dist. (NWD) (decimal degree)	–	–	−0.476*** (0.022)	0.822*** (0.100)	–	–
Long (OSU) (decimal degree)	–	–	–	–	0.137*** (0.027)	–
Lat (OSU) (decimal degree)	–	–	–	–	0.145*** (0.026)	–
$Long^2$ (OSU) (decimal degree)	–	–	–	–	−0.778*** (0.059)	–
Lat^2 (OSU) (decimal degree)	–	–	–	–	−1.229*** (0.131)	–
Long (NWD) (decimal degree)	–	–	–	–	–	0.099*** (0.025)
Lat (NWD) (decimal degree)	–	–	–	–	–	0.237*** (0.027)
$Long^2$ (NWD) (degree)	–	–	–	–	–	−0.778*** (0.059)
Lat^2 (NWD) (degree)	–	–	–	–	–	−1.229*** (0.131)
AIC_c	5024.078	4448.907	4557.208	4383.439	4605.793	4605.793
LogLik	−2503.039	−2214.454	−2268.604	−2180.72	−2289.896	−2289.896
Adj. R^2	0.714	0.726	0.724	0.727	0.723	0.723

Note: 13,327 observations. Robust standard errors are in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.001$.

- We expect two of the parameters of the fixed point correction variables $\{(long_A - long_B), (lat_A - lat_B)\}$ to change as fixed positions for the correction changes.
- Fixing the influence of location may pick up some new information, resulting in marginally higher efficiency measures, especially with high resolution data.

Lubbock, Texas

The Lubbock, Texas demographics across the study space are more uniform than those in Columbus. The study space is also much smaller with higher resolution data. Table 2 compares parameter estimates for Lubbock model with the natural log of home prices as the dependent variable and natural log of continuous independent variables; model 1 is the baseline model without any distance or fixed point variables, models 2–4 include distance variables and models 5 and 6 include fixed point correction variables (we also report the results from the linear models in Appendix Tables A1–A3.). Value changes in distance to Texas Tech and the Mall show substantial change between models 2–4 and 3–4; for Texas Tech the parameter increases from −56.236 to −1.852 and for the Mall, it increases from 0.125 to 1.933 with a change in the statistical significance (t-value changes from 1.106 to 5.384). Across models 2 and 4, the estimates for home construction variables which are not spatially distributed (square footage, lot size, house age) are quite stable.

The effect of different distance variables on spatially distributed variables (the environmental proxy) in the model is especially acute. The most notable parameter instability arises in the policy variable itself. The parameter estimates for the environmental proxy decreases from 0.012 and 0.013 in models 2 and 3 respectively to -0.003 in model 4; most noteworthy, the estimated effect of the environmental variable in the models with fixed point correction (models 5 and 6) is -0.0002, which is substantially different from the distance corrected and baseline models. Moreover, the environmental policy variable has a t-value of over 1.50 across models 1–3 yet *highly insignificant* in the fixed position correction models (5 and 6) with t-values of 0.025.

Importantly, *all* identical parameter estimates are realized for every model variable in the two fixed position control models, confirming the key hypothesis. Aligned with predictions, there is a noticeable improvement in the efficiency measures with the inclusion of the fixed point correction variables in models 5 and 6. The stability of efficiency across models with unstable parameter estimates (models 2–4) for spatially distributed variables occurs as predicted. Finally, the parameters of the correction variables vary as position of the correction for position changes with changing the fixed point.

Columbus, Ohio

The Columbus, Ohio study area reflects a more common scale of regional science examination: it is a much larger space, with far more observations and a more diverse population. Table 3 demonstrates the potential magnitude of problems with nonidentification of distance variables. Like the last model, both dependent and continuous independent variables are logged. Between models 2 and 4, the estimated effect on home price for the distance to the University (OSU) decreases by 152.53%, from −0.493 to −1.245; from model 3 to 4, the parameter for the distance to NWD changes from −0.476 to 0.822.

As expected, the fixed point correction models fully stabilize all parameter estimates. In uncorrected models, the key spatially distributed variables of rate of criminal offences, and racial composition (percentage white) show some volatility. Home age and square footage show the least volatility as the spatial distribution of the oldest homes tends to ring the city. (Columbus is ringed by once independent cities. As the state capital, these municipalities were annexed. Newer homes were built within Columbus city limits.) Estimates for criminal offenses per district vary the most from the baseline model (without distance or fixed point correction) but vary by about 5% among models with some distance variable, although these stabilize with the addition of the fixed position corrections (models 5 and 6).

A key policy variable is racial composition (percentage of white residents) which also demonstrates parameter instability across models without fixed point correction. The estimate for the value of a one percent (percentage point) increase in the percentage of white households reaches both its highest and most significant value in the corrected models. Corrected models tend to generate the highest or lowest estimated values for spatially distributed models than models with distance variables. Comparing value changes in house price between neighborhoods that have a majority non-white population and all others, this corrected model parameter predicts an increase in home value close to 0.081% for every percentage point increase in the number of white households.

Once again, the overall information in the models appears unaltered and there is remarkable stability in efficiency measures across models with unstable parameter estimates (models 2–4), although expectedly we observe a relatively poor fit because of the low resolution data. Models generate virtually identical R^2 statistic, even with different parameter estimates for spatially distributed variables. The Log Likelihood and AIC statistics for models 2–6 (with distance variables and fixed point corrections) are also similar. Expectedly, there is a small improvement in efficiency measures in these models with low resolution data, with little additional information being picked up with fixing the influence of location (models 5 and 6). Finally, parameters of *all* variables only stabilize with the inclusion of fixed point correction variables.

Alternative Modelling Choices

Some alternatives are worth noting. Driving time and a single zonal distance variable might have a different outcome than a distance variable. Driving time has a weakness in common with Euclidean distance in that both form a lattice over the data that tracks omitted influences across space, much like illustrated on Figure 2. Our experiments with single zonal variables (sports center; hazardous waste site) show none of the inconsistencies of distance variable while multi-zonal variables perform as continuous distance variables, simply discretized.

An area of immediate concern is the extent to which this spatial economic tool compares, duplicates, replaces, or resolves a different problem altogether than other spatial econometric tools. We test the spatial lagged-X variable model (SLX) below and present results with and without the fixed point control in parallel to the test above.

Table 4 compares results for Lubbock and Columbus of a new model with lagged independent variables, with and without the fixed point correction. The overall

Table 4. Spatially lagged-X (SLX) model estimates.

	Lubbock, TX		Columbus, OH	
	W/o fixed point correction	Fixed point correction (Tech)	W/o fixed point correction	Fixed point correction (OSU)
Square footage	0.932*** (0.042)	0.915*** (0.040)	0.558*** (0.012)	0.552*** (0.012)
Lot size (sq. yard)	0.108*** (0.029)	0.125*** (0.028)	0.077*** (0.007)	0.079*** (0.007)
House age (years)	−0.112*** (0.022)	−0.090*** (0.021)	−0.072*** (0.005)	−0.072*** (0.005)
Garage (0/1)	0.027 (0.018)	0.036** (0.017)	–	–
Env. Proxy birdsXspecies	−0.007 (0.009)	−0.004 (0.008)	–	–
Income ($)	0.050** (0.024)	0.050** (0.022)	0.053** (0.021)	0.060*** (0.021)
Second story (0/1)	–	–	−0.032*** (0.007)	−0.029*** (0.007)
Offenses per district	–	–	−0.086*** (0.019)	−0.080*** (0.019)
Pct white (%)	–	–	−0.032** (0.016)	−0.033** (0.016)
Long (decimal degree)	–	8.588 (16.160)	–	−4.811* (2.559)
Lat (decimal degree)	–	50.269*** (15.346)	–	−1.153 (3.167)
$Long^2$ (decimal degree)	–	186.790 (165.059)	–	−2.363 (6.283)
Lat^2 (decimal degree)	–	231.736* (127.959)	–	10.715 (14.562)
Lag(Square Footage)	0.052 (0.088)	−0.119 (0.093)	0.577*** (0.019)	0.551*** (0.019)
Lag(Lot size (Sq. yard))	0.023 (0.043)	0.056 (0.045)	−0.127*** (0.009)	−0.112*** (0.009)
Lag(House age (Years))	−0.013 (0.026)	0.014 (0.030)	0.060*** (0.005)	0.055*** (0.005)
Lag(Garage (0/1))	−0.060 (0.046)	0.073 (0.049)	–	–
Lag(Env. Proxy birdsXspecies)	0.018 (0.014)	−0.027* (0.015)	–	–
Lag(Income ($1000))	0.331*** (0.068)	0.305*** (0.066)	0.372*** (0.024)	0.302*** (0.024)
Lag(Second story (0/1))	–	–	−0.162*** (0.012)	−0.139*** (0.012)
Lag(Offenses per district)	–	–	0.061*** (0.019)	0.025 (0.019)
Lag(Pct white (%))	–	–	0.102*** (0.016)	0.124*** (0.016)
Lag(Long)	–	−2.772 (16.486)	–	4.923* (2.560)
Lag(Lat)	–	−44.157*** (15.575)	–	1.270 (3.167)
Lag($Long^2$)	–	−150.713 (168.660)	–	1.759 (6.284)
Lag(Lat^2)	–	−153.972 (132.019)	–	−11.617 (14.564)
AIC	305.2332	381.7352	3437.649	3199.875
LogLik	−138.6166	−168.8676	−1702.824	−1575.937
Adj. R^2	0.866	0.884	0.744	0.749

Note: Robust standard errors are in parentheses. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$, $^{****}p < 0.001$.

performance of the correction is unchanged. The effect of the SLX model on the parameters is to partition the parameters into the direct effects and the indirect influences of their nearest neighbors – the sums of which are largely equivalent to the singular effects in the original models. The SLX model appears to be managing neighborhood effects of individual house characteristics such as size and age, and immediate inter-neighborhood effects of features such as crime rate, race or school quality. The correction for overall position is an independent effect and does not appear to inject any noticeable, unique or duplicatory pattern in the spatial or non-spatial parameters between the fixed point corrections and the SLX regression estimates. So, the SLX is a complementary model component that exerts a wholly independent effect from position per se. The two practices are not substitutes.

Implications of Findings

The examples above provide strong empirical tests that demonstrate the problem with distance variables and the utility of a robust location correction to repair that problem when used as a control for overall position in the study area. In the case of a reasonably good model, the fixed point correction in a model adds new information that explains more variation overall and provides consistency that distance variables cannot provide to account for the spatial position of observations. If the model fails to account for large systematic spatially distributed variation uncorrelated with model variables, the correction may add some explanatory power, but that should signal model revision. An outcome of this study is that low resolution data, rather than a small number of observations, is more likely to exhibit this property and itself acts as an additional diagnostic for the researcher.

The correction itself is a simple mathematical consequence of position in standard linear regression or two variable distributions used in maximum likelihood estimation. Distance variables, however, clutter the issue as they co-vary with other spatially distributed variables. That is, distance variables occlude the capacity to examine location-specific effects in a study area as the distance variable is itself free of positional information. Yet the problem caused by the position of an observation, which distance variables try to solve when used as an anchor position, is ubiquitous.

We view the fixed point correction as a routine contribution which is easy to implement and should be accompanied by a test using another potion to replicate results. Future research to compare, contrast and integrate the fixed point correction alongside spatial Durbin models or spatial error models is needed to enhance our understanding the effects and nature of spatial variation in the regional sciences broadly, including (perhaps especially) models of causal inference.

Conclusion

Our work illustrates the instability of distance variable parameters. These inconsistencies meet the conditions of unidentified variables; or, $(\text{Pr}Y|X_1) = (\text{Pr}Y|X_1, X_2)$. That is, X_1 and X_2 contain the same information (total explanatory power) as either one independently; or there is no independent variation. In addition, all other parameter estimates of spatially distributed variables are not consistent among the choice of a distance variable.

The unreliable indication of hedonic value of an amenity inferred by a distance variable also explains the ineffectiveness of the distance to the CBD as a means to control for the spurious effects of location. This problem of spurious effects of location on the estimated effects of many key policy variables, raised by Deaton and Hoehn (2004), is very real; but the use of a distance variable to control for unmodeled influences that arise from the specific location of an observation fails because measured distance is not location specific. Yet some correction is still needed.

We suggest the use of a new control, formed by unpacking the Euclidean distance formula into its four nested variables which are specific to location: $\left\{(long_A - long_B); (long_A - long_B)^2; (lat_A - lat_B); (lat_A - lat_B)^2\right\}$, where A is the subject position and B is any random fixed position. Replacing distance to the CBD with these four variables to *any* and *only one B position* fully stabilizes parameter estimates for spatially distributed variables, a result that holds regardless of the fixed position used to construct the correction.

Empirical results corroborate four key hypotheses of this study. The four correction variables constructed from longitude and latitude distances between any observation and a fixed position fully stabilizes parameter estimates for spatially and non-spatially distributed variables in the models, a result that holds regardless of the fixed position chosen to construct the correction.

This work illustrates a severe identification problem with distance variables. Though distance variables do not consistently estimate the effects of a landmark on home price, there remains a need to control for the effects of position on other spatially distributed variables (e.g., home age, crimes rates, local environmental outcomes, education outcomes). By unpacking the Euclidean measure for distance, a control is introduced and tested which fully stabilizes all model parameter estimates and measures of model efficiency.

Notes

1. Triangulation of position C is exactly the condition where AB, BC and AC are all known. Given those three distances, the angular position of C, vis a vis A and B, can be found from $\tan(\angle BAC)$ and $\tan(\angle CBA)$ created by adding C at (x_3,y_3) Introducing position C only defines triangle ABC, thus recovering its own position but adding nothing that tells us more about position A. In fact, there is an infinite number of identical ABC triangles possible around position A, rotating in an 360° circle around A, and forming a sphere around A around the x, y plane.
2. The familiar example is a vehicle that is either red or blue. To examine the effect of color on twilight accidents, knowledge that the range of colors for a type of vehicle is red or blue means that knowledge of the model of the vehicle observed in an accident and the observation that the vehicle is not red or is red perfectly predicts that the vehicle is blue or not blue, respectively; thus making the addition of another color variable redundant and the second color variable an unidentified explanatory variable.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

ORCID

Syed Fuad http://orcid.org/0000-0001-9982-0233

Data Availability Statement

The data that support the findings of this paper are available from Farmer et al. (2012) and FARES (2002).

References

Abbiasov, T., & Sedov, D. (2023). Do local businesses benefit from sports facilities? The case of major league sports stadiums and arenas. *Regional Science and Urban Economics*, *98*, 103853. https://doi.org/10.1016/j.regsciurbeco.2022.103853

Blanco, H. (2023). Pecuniary effects of public housing demolitions: Evidence from Chicago. *Regional Science and Urban Economics*, *98*, Article 103847. https://doi.org/10.1016/j.regsciurbeco.2022.103847

Bogin, A. N., & Doerner, W. M. (2019). Property renovations and their impact on house price index construction. *Journal of Real Estate Research*, *41*(2), 249–284. https://doi.org/10.1080/10835547.2019.12091526

Bondemark, A., & Merkel, A. (2023). Parking not included: The effect of paid residential parking on housing prices and its relationship with public transport proximity. *Regional Science and Urban Economics*, *99*, 103877. https://doi.org/10.1016/j.regsciurbeco.2023.103877

Borck, R., & Schrauth, P. (2021). Population density and urban air quality. *Regional Science and Urban Economics*, *86*, 103596. https://doi.org/10.1016/j.regsciurbeco.2020.103596

Brasington, D. M. (2007). Private schools and the willingness to pay for public schooling. *Education Finance and Policy*, *2*(2), 152–174. https://doi.org/10.1162/edfp.2007.2.2.152

Brasington, D. M., & Hite, D. (2005). Demand for environmental quality: A spatial hedonic analysis. *Regional Science and Urban Economics*, *35*(1), 57–82. https://doi.org/10.1016/j.regsciurbeco.2003.09.001

Brasington, D. M., & Sarama, R. F. Jr, (2008). Deed types, house prices and mortgage interest rates. *Real Estate Economics*, *36*(3), 587–610. https://doi.org/10.1111/j.1540-6229.2008.00223.x

Cameron, T. A. (2006). Directional heterogeneity in distance profiles in hedonic property value models. *Journal of Environmental Economics and Management*, *51*, 26–45. https://doi.org/10.1016/j.jeem.2005.03.003

Dauth, W., & Haller, P. (2020). Is there loss aversion in the trade-off between wages and commuting distances? *Regional Science and Urban Economics*, *83*, 103527. https://doi.org/10.1016/j.regsciurbeco.2020.103527

Deaton, B. J., & Hoehn, J. P. (2004). Hedonic analysis of hazardous waste sites in the presence of other urban disamenities. *Environmental Science & Policy*, *7*(6), 499–508. https://doi.org/10.1016/j.envsci.2004.08.003

Dermisi, S., & McDonald, J. (2010). Selling prices/sq. ft. of office buildings in down town Chicago—How much is it worth to be an old but Class A building? *Journal of Real Estate Research*, *32*(1), 1–22. https://doi.org/10.1080/10835547.2010.12091270

Diao, M., Li, Q., Sing, T. F., & Zhan, C. (2023). Disamenities of living close to transit tracks: Evidence from Singapore's MRT system. *Regional Science and Urban Economics*, *100*, 103894. https://doi.org/10.1016/j.regsciurbeco.2023.103894

Dumm, R. E., Sirmans, G. S., & Smersh, G. T. (2016). Price variation in waterfront properties over the economic cycle. *Journal of Real Estate Research*, *38*(1), 1–26. https://doi.org/10.1080/10835547.2016.12091435

Farmer, M. C., Wallace, M., & Shiroya, M. (2012). Bird diversity indicates value in urban home prices. *Urban Ecosystems*, *16*(3), 131–144. https://doi.org/10.1007/s11252-011-0209-0

First American Real Estate Solutions (FARES). (2002). CRG 03172.

Grisham, B., Borsdorf, P., Boal, C., & Boydston, K. (2014). Nesting ecology of lesser prairie-chickens on the southern high plains of Texas. *Journal of Wildlife Management*, *78*, 857–866. https://doi.org/10.1002/jwmg.716

Harrison, D., & Rubinfeld, D. (1978). Hedonic housing prices and demand for clean air. *Journal of Environmental Economics and Management*, *5*, 81–102. https://doi.org/10.1016/0095-0696(78)90006-2

Ihlanfeldt, K. R., & Taylor, L. O. (2004). Externality effects of small-scale hazardous waste sites: Evidence from urban commercial property markets. *Journal of Environmental Economics and Management*, *47*(1), 117–139. https://doi.org/10.1016/S0095-0696(03)00070-6

Jayet, H., Rayp, G., Ruyssen, I., & Ukrayinchuk, N. (2016). Immigrants' location choice in Belgium. *The Annals of Regional Science*, *57*, 63–89.

Lawani, A., Reed, M. R., Mark, T., & Zheng, Y. (2019). Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Regional Science and Urban Economics*, *75*, 22–34. https://doi.org/10.1016/j.regsciurbeco.2018.11.003

Letdin, M., & Shim, H. S. (2019). Location choice, life cycle and amenities. *Journal of Regional Science*, *59*(3), 567–585. https://doi.org/10.1111/jors.12441

Leuenberger, K. (2015). *The effects of vegetation variability, demographic differences, and spatial factors on urban bird richness* [Master Thesis]. Texas Tech University.

MacArthur, R. H., & MacArthur, J. W. (1961). On bird species diversity. *Ecology*, *42*(3), 594–598. https://doi.org/10.2307/1932254

Meltzer, R., & Schuetz, J. (2010). What drives the diffusion of inclusionary zoning? *Journal of Policy Analysis and Management*, *28*(2), 296–320.

Metz, N. E. (2015). Effect of distance to schooling on home prices. *Review of Regional Studies*, *45*, 151–171. https://doi.org/10.52324/001c.8060

Nase, I., van Assendelft, N., & Remøy, H. (2019). Rent premiums and vertical sorting in Amsterdam's multi-tenant office buildings. *Journal of Real Estate Finance and Economics*, *59*, 419–460. https://doi.org/10.1007/s11146-018-9684-x

Nguyen, T. X. T., & Diez, J. R. (2017). Multinational enterprises and industrial spatial concentration patterns in the Red River Delta and Southeast Vietnam. *Annals of Regional Science*, *59*(1), 101–138. https://doi.org/10.1007/s00168-017-0820-y

Noonan, D. S., Krupka, D. J., & Baden, B. M. (2007). Neighborhood dynamics and price effects of superfund site clean-up. *Journal of Regional Science*, *47*(4), 665–692. https://doi.org/10.1111/j.1467-9787.2007.00525.x

Osland, L. (2010). An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research*, *32*(3), 289–320. https://doi.org/10.1080/10835547.2010.12091282

Palmquist, R. B. (2005). Property value models. *Handbook of Environmental Economics*, *2*, 763–819.

Robinson, S., Simons, R., & Lee, E. (2017). Which green office building features do tenants pay for? A study of observed rental effects. *Journal of Real Estate Research*, *39*(4), 467–492. https://doi.org/10.1080/10835547.2017.12091483

Ross, J. M., Farmer, M. C., & Lipscomb, C. A. (2011). Inconsistency in welfare inferences from distance variables in hedonic regressions. *Journal of Real Estate Finance and Economics*, *43*(3), 385–400. https://doi.org/10.1007/s11146-009-9221-z

Salvati, L., Venanzoni, G., Serra, P., & Carlucci, M. (2016). Scattered or polycentric? Untangling urban growth in three southern European metropolitan regions through exploratory spatial data analysis. *Annals of Regional Science*, *57*, 1–29. https://doi.org/10.1007/s00168-016-0758-5

Schuetz, J. (2009). No renters in my suburban backyard: Land use regulation and rental housing. *Journal of Policy Analysis and Management*, *28*(2), 296–320. https://doi.org/10.1002/pam.20428

Shiroya, M. (2010). A finite mixture approach for household residential choices [Doctoral dissertation]. Texas Tech University.

Sieg, H., Smith, V. K., Banzhaf, H. S., & Walsh, R. (2002). Interjurisdictional housing prices in locational equilibrium. *Journal of Urban Economics*, *52*(1), 131–153.

Sun, W., Guo, D., Li, Q., & Fang, H. (2021). School runs and urban traffic congestion: Evidence from China. *Regional Science and Urban Economics*, *86*, 103606. https://doi.org/10.1016/j.regsciurbeco.2020.103606

Taylor, L. O., Phaneuf, D., & Liu, X. (2016). Disentangling the property value impacts of environmental contamination from locally undesirable land uses: Implications for measuring post-cleanup stigma. *Journal of Urban Economics*, *93*, 85–98. https://doi.org/10.1016/j.jue.2016.03.004

Tontisirin, N., & Anantsuksomsri, S. (2021). Measuring transit accessibility benefits and their implications on land value capture: A case study of the Bangkok Metropolitan Region. *Annals of Regional Science*, *67*(2), 415–449. https://doi.org/10.1007/s00168-021-01053-2 van Vuuren, A. (2023). Is there a diminishing willingness to pay for consumption amenities as a result of the Covid-19 pandemic? *Regional Science and Urban Economics*, *98*, 103858. https://doi.org/10.1016/j.regsciurbeco.2022.103858

Yilmazkuday, D., & Yilmazkuday, H. (2016). Understanding gasoline price dispersion. *Annals of Regional Science*, *57*(1), 223–252. https://doi.org/10.1007/s00168-016-0775-4

Zuta, G., Pe'er, G., Belloq, M. I., & Hansbauer, M. M. (2012). Edge effects and their influence on habitat suitability calculations: A continuous approach applied to birds of the Atlantic Forest. *Journal of Applied Ecology*, *49*(2), 503–512. https://doi.org/10.1111/j.1365-2664.2011.02104.x

Appendix A

A simple proof shows that coordinates, c_1 and c_2, for point C on Figure 1 can be defined directly from the coordinates, $\{a_1, a_2, b_1, b_2\}$. Because triangle ABC is defined, we know the degree angles θ_A at A and θ_B at B. So, given coordinates (a_1, a_2); angle degree θ_A; coordinates (b_1, b_2); angle degree θ_B; and the length AB, then (c_1, c_2) can be recovered.

From the slope, $m = \frac{b_2 - a_2}{b_1 - a_1}$, we find the *angle of inclination* from A to B, θ_{AB} (the angle that forms between line AB and the x-axis), which is:

$$\theta_{AB} = \tan^{-1} m \tag{1}$$

Using θ_{AB} we can define the angle of inclination from A to C, θ_{AC}, from:

$$\theta_A = \theta_{AC} - \theta_{AB} \tag{2}$$

This draws a line from A parallel to the x-axis to divides θ_A into θ_{AC} and θ_{AB}.

To find c_1 and c_2, we also use the distance AC and the law of sines to calculate:

$$\frac{\sin(180 - \theta_A - \theta_B)}{AB} = \frac{\sin(\theta_B)}{AC} \tag{3}$$

Therefore $AC = \frac{AB\sin(\theta_B)}{\sin(180 - \theta_A - \theta_B)}$. With AC and the angle of inclination, θ_{AC}, coordinates (c_1, c_2) can be directly defined by:

$$(c_1, c_2) = (a_1 + AC\cos(\theta_{AC}), a_2 + AC\sin(\theta_{AC})) \tag{4}$$

As an example, let A and B be coordinates, (2, 3) and (2, 5); and both angles A and B be 30°. This makes $AB = 10^{\frac{1}{2}}$ and the slope, $m = -\frac{1}{3}$. So, the angle of inclination from A to B is:

$$\theta_{AB} = \tan^{-1}\left(-\frac{1}{3}\right) \approx 11.565^\circ$$

and the angle of inclination from A to C is:

$$\theta_{AC} \approx (30^\circ - 11.565^\circ) \approx 18.435^\circ$$

and using (3), the distance from A to C becomes:

$$AC \approx 1.826$$

Now, from basepoint $A = (2, 3)$, using (4), coordinates of $C = (c_1, c_2)$ are defined: $(3.789, \ 3.366)$

As a check, points $(2, \ 3)$ and $(3.789, \ 3.366)$ form distance $AC = (3.2 + 0.133)^{\frac{1}{2}} \approx 1.82$, matching AC above. Also, all distances and opposite angles satisfy the law of sines, represented in equation (3).

Critically, point C only conveys information to complete the triangle; or, the information needed to triangulate its own position, given coordinates of A and B.[1] Intuitively, point B can be rotated 360 degrees around point A for the same distance. As C rotates with B, that rotation preserves the same relative distances that define triangle $[ABC]$; or, nothing new is discovered about the position of A by adding point C. No more information about position A beyond $\{(a_1, a_2), (b_1, b_2)\}$ is introduced. The same is true for any position in the study area. Critically, information available from (c_1, c_2) is fully characterized by coordinates (a_1, a_2) and (b_1, b_2) – all that is required for the design matrix to collapse[2] and for any additional distance variable to be unidentified.

Table A1. Linear model comparison for Lubbock, Texas.

	Model 1: Baseline	Model 2: Distance (Tech)	Model 3: Distance (Mall)	Model 4: Distance: Mall&Tech	Model 5: Fixed Point 1 Tech	Model 6: Fixed Point 2 Mall
Square Foot	90.069*** (4.193)	90.171*** (4.209)	90.282*** (4.197)	89.101*** (4.178)	88.244*** (3.727)	88.244*** (3.727)
Lot size (Sq. yard)	1,713.553*** (341.849)	1,715.707*** (342.327)	1,720.433*** (341.801)	1,718.132*** (338.544)	1,842.636*** (302.971)	1,842.636*** (302.971)
House Age (Years)	−1,746.757*** (245.015)	−1,737.298*** (246.845)	−1,736.524*** (245.116)	−1,934.757*** (252.810)	−2,176.789*** (237.370)	−2,176.789*** (237.370)
Garage (0/1)	8,593.911 (6,247.244)	8,483.416 (6,263.140)	8,469.489 (6,246.357)	10,739.110* (6,239.257)	10,597.430* (5,547.060)	10,597.430* (5,547.060)
Env. Proxy birdsXspecies	539.479*** (128.475)	541.243*** (128.735)	524.563*** (129.145)	329.158** (145.576)	103.842 (131.697)	103.842 (131.697)
Income ($)	0.195** (0.088)	0.197** (0.088)	0.200** (0.088)	0.199** (0.087)	0.164** (0.078)	0.164** (0.078)
Dist. (Tech) (decimal degree)	–	11,990.920 (34,758.980)	–	−373,448.00*** (132,827.50)	–	–
Dist. (Mall) (decimal degree)	–	–	39,069.460 (35,396.280)	406,974.00*** (135,470.90)	–	–
Long (Tech) (decimal degree)	–	–	–	–	3,242,881.00*** (355,324.60)	–
Lat (Tech) (decimal degree)	–	–	–	–	258,523.40** (123,065.90)	–
$Long^2$ (Tech) (decimal degree)	–	–	–	–	27,193,680.00*** (2,636,368.00)	–
Lat^2 (Tech) (decimal degree)	–	–	–	–	−1,092,402.00*** (135,419.30)	–
Long (Mall) (decimal degree)	–	–	–	–	–	475,652.40*** (152,977.60)
Lat (Mall) (degree)	–	–	–	–	–	357,320.20*** (118,160.40)
$Long^2$ (Mall) (decimal degree)	–	–	–	–	–	27,193,680.00*** (2,636,368.00)
Lat^2 (Mall) (degree)	–	–	–	–	–	−1,092,402.00*** (135,419.30)
AIC_c	9058.171	9060.05	9058.928	9052.912	8968.818	8968.818
LogLik	−4521.086	−4521.025	−4520.464	−4516.456	−4472.409	−4472.409
Adj. R^2	0.762	0.762	0.763	0.767	0.816	0.816

Note: 365 observations. Robust standard errors are in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.001$.

Table A2. Linear model comparisons for Columbus, Ohio.

	Model 1: Baseline	Model 2: Distance (OSU)	Model 3: Distance (NWD)	Model 4: Distance: OSU&NWD	Model 5: Fixed Point 1 OSU	Model 6: Fixed Point 2 NWD
Square Foot	102.358*** (1.012)	102.820*** (1.004)	102.775*** (1.004)	102.807*** (1.004)	102.544*** (1.007)	102.544*** (1.007)
Lot size (Sq. yard)	0.159*** (0.018)	0.183*** (0.018)	0.182*** (0.018)	0.183*** (0.018)	0.168*** (0.018)	0.168*** (0.018)
House Age (Years)	155.757*** (16.787)	138.899*** (16.696)	129.522*** (16.755)	134.977*** (17.012)	150.197*** (16.858)	150.197*** (16.858)
Second Story (0/1)	−19,579.520*** (1,079.657)	−19,461.240*** (1,071.209)	−19,252.940*** (1,071.493)	−19,376.800*** (1,073.495)	−18,981.580*** (1,081.819)	−18,981.580*** (1,081.819)
Income ($)	1.019*** (0.029)	0.911*** (0.030)	0.928*** (0.030)	0.917*** (0.030)	0.927*** (0.030)	0.927*** (0.030)
Offenses per District	0.444 (8.048)	−34.019*** (8.327)	−33.211*** (8.315)	−34.057*** (8.327)	−29.141*** (8.324)	−29.141*** (8.324)
Pct White (%)	201.581*** (23.291)	286.910*** (23.837)	308.005*** (24.245)	296.230*** (25.068)	314.614*** (24.964)	314.614*** (24.964)
Dist. (OSU) (decimal degree)	–	−53,911.880*** (3,696.721)	–	−32,941.670* (17,842.600)	–	–
Dist. (NWD) (decimal degree)	–	–	−57,408.890*** (3,955.116)	−22,931.900 (19,088.360)	–	–
Long (OSU) (decimal degree)	–	–	–	–	19,793.820*** (5,120.252)	–
Lat (OSU) (decimal degree)	–	–	–	–	−3,992.965 (4,826.286)	–
$Long^2$ (OSU) (decimal degree)	–	–	–	–	−87,084.270*** (11,063.380)	–
Lat^2 (OSU) (decimal degree)	–	–	–	–	−215,269.600*** (24,403.490)	–
Long (NWD) (decimal degree)	–	–	–	–	–	15,526.690*** (4,681.487)
Lat (NWD) (decimal degree)	–	–	–	–	–	12,152.250** (5,078.131)
$Long^2$ (NWD) (degree)	–	–	–	–	–	−87,084.270*** (11,063.380)
Lat^2 (NWD) (degree)	–	–	–	–	–	−215,269.600*** (24,403.490)
AIC_c	328244.1	328034.9	328036.9	328035.5	328080.8	328080.8
LogLik	−164113	−164007.5	−164008.5	−164006.8	−164027.4	−164027.4
Adj. R^2	0.690	0.695	0.695	0.695	0.694	0.694

Note: 13,327 observations. Robust standard errors are in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.001$.

Table A3. Spatially lagged-X (SLX) model estimates.

	Lubbock, TX		Columbus, OH	
	W/o fixed point correction	Fixed point correction	W/o fixed point correction	Fixed point correction
Square footage	84.241*** (4.290)	83.131*** (4.086)	82.377*** (1.187)	82.311*** (1.181)
Lot size (sq. yard)	1,670.558*** (317.284)	1,842.429*** (307.741)	0.241*** (0.027)	0.249*** (0.027)
House age (years)	−1,796.722*** (453.380)	−1,471.233*** (444.255)	−200.987*** (32.314)	−203.414*** (32.119)
Garage (0/1)	8,789.248 (5,539.076)	8,891.172 (5,391.989)	–	–
Env. Proxy birdsXspecies	−54.201 (190.064)	91.110 (184.348)	–	–
Income ($)	0.117 (0.079)	0.102 (0.076)	−0.142* (0.073)	−0.134* (0.072)
Second story (0/1)	–	–	−12,167.530*** (1,198.368)	−12,059.500*** (1,191.666)
Offenses per district	–	–	−19.021 (15.628)	−17.235 (15.533)
Pct white (%)	–	–	−114.276 (94.659)	−125.269 (94.112)
Long (decimal degree)	–	−138,144.000 (4,987,900.000)	–	−1,815,156.000*** (479,700.200)
Lat (decimal degree)	–	18,200,290.000*** (4,765,745.000)	–	−1,169,643.000** (593,327.700)
Long^2 (decimal degree)	–	46,897,420.000 (49,812,067.000)	–	829,661.600 (1,177,676.000)
Lat^2 (decimal degree)	–	122,525,154.000*** (39,807,166.000)	–	2,174,723.000 (2,727,757.000)
Lag(Square Footage)	−10.261 (8.331)	−21.216** (8.758)	47.239*** (1.831)	47.887*** (1.826)
Lag(Lot size (Sq. yard))	342.240 (697.510)	1,394.366* (756.879)	−0.216*** (0.034)	−0.205*** (0.034)
Lag(House age (Years))	427.344 (509.465)	607.067 (625.171)	554.823*** (36.381)	544.962*** (36.265)
Lag(Garage (0/1))	9,371.894 (14,294.820)	33,911.940** (16,467.040)	–	–
Lag(Env. Proxy birdsXspecies)	492.728** (224.924)	−87.914 (235.005)	–	–
Lag(Income ($1000))	0.725*** (0.220)	0.553** (0.220)	1.047*** (0.081)	0.921*** (0.082)
Lag(Second story (0/1))	–	–	−27,625.330*** (1,998.959)	−26,239.250*** (2,014.771)
Lag(Offenses per district)	–	–	57.169*** (18.034)	19.226 (18.172)
Lag(Pct white (%))	–	–	371.776*** (98.211)	486.464*** (98.165)
Lag(Long)	–	1,566,617.000 (5,088,079.000)	–	1,827,506.000*** (479,806.000)
Lag(Lat)	–	−17,007,869.000*** (4,822,994.000)	–	1,161,382.000* (593,426.300)
Lag(Long^2)	–	−38,827,983.000 (50,873,830.000)	–	−902,959.800 (1,177,864.000)
Lag(Lat^2)	–	−107,842,482.000*** (41,070,924.000)	–	−2,376,610.000 (2,728,239.000)
AIC	8842.28	8809.958	322061.1	321901.4
LogLik	−4407.14	−4382.979	−161014.6	−160926.7
Adj. R^2	0.803	0.824	0.716	0.720

Note: Robust standard errors are in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$, ****$p < 0.001$.