How “Fixed” are Fixed Effects?
A Novel Approach to Measuring the Value of Location

Stanley D. Longhofer
W. Frank Barton School of Business
Wichita State University
stan.longhofer@wichita.edu

Christian L. Redfearn
Price School of Public Policy
University of Southern California
redfearn@usc.edu

In this paper, we develop a novel method for measuring the value of location within a metropolitan area. By estimating overlapping “semi-local” regressions and exploiting the symmetry among them, we are able to construct a surface of the relative value of location within a metropolitan area. This methodology maintains the simplicity of using location fixed effects while resolving the omitted variable bias that arises from the traditional application in hedonic regressions. Using data from Maricopa County (Phoenix), we find that the premium for location is decidedly non-monocentric, asymmetric, and highly dynamic. This dynamism conflicts with the typical assumptions associated with the use of fixed effects in the construction of aggregate price indexes, and potentially complicates the interpretation of difference-in-difference methods.

Current Draft: December 2022

1. Introduction

The Alonso-Muth-Mills model of cities establishes the theoretical foundations for household trade-offs between location and other goods (Henderson, 2014). In the standard urban model, the location premium is a function of the access to the single spatial amenity in the model: the central business district (CBD), in which all employment is found. In reality, most employment is scattered throughout a metropolitan area, with less than 10 percent of all employment in the central business district (CBD) in large MSAs (Demographia, 2020). Moreover, seven out of eight trips from home are not commutes to work (Gordon and Lee, 2015). As such, the value of access embedded into a parcel's location is more than simply the opportunity cost of the drive to the CBD. Rather, a large collection of amenities and disamenities beyond a property's lines acts to influence the premium paid for its location. Because of this, the standard urban model on its own can fail to accurately capture the nuances of location pricing within cities.

More importantly, the various amenities and disamenities that are capitalized into location premia can change markedly over time. Unfortunately, few hedonic analyses of house prices control for these temporal changes explicitly and comprehensively. Instead, it is common for the collection of spatial amenities to be controlled for using static locational fixed effects. In a cross-section, these controls net the spatial amenities that are capitalized into a location. Over time, however, cities change, as do the neighborhoods within them. As such, location fixed effects actually represent the average contribution of changing local amenities and disamenities to the value of house's location over time. If we are interested in identifying the temporal impacts of changing fundamentals or policies, the use of static location fixed effects will permit omitted local trends to be commingled with other variables that are correlated with time.

Urbanization, suburbanization, and gentrification all represent broad trends that essentially reorganize spatial hierarchies – and all represent misspecifications that would invalidate the use of location fixed effects. Beyond these large secular trends, cities and neighborhoods evolve as local fundamentals evolve within them: better schools, reductions in crime, and shifting retail opportunities should all be capitalized into location premia. And where these changes in location premia are omitted, the fact that the distribution of physical housing characteristics varies systematically across a city suggests that the impacts that *should* be reflected in location premia may instead end up in the estimated shadow prices of the physical housing characteristics. Worse yet, these temporal changes in the value of location may complicate the interpretation of difference-in-difference approaches or repeat-sales indexes, which typically require that local fixed effects are actually fixed over time. Where they are not fixed, the identification of the impact of, say, a spatial policy change, may in fact result from spurious changes in the value of location. While is it common to impose time-invariant locational fixed effects, this assumption is seldom tested. Instead, the substantial dynamism that occurs within metropolitan areas is implicitly ignored in using these location fixed effects. The question we ask in this paper is how "fixed" are these fixed effects?

In this paper, we develop a novel methodology for disentangling the location premia from the many other sources of heterogeneity embedded in a house price: hex-based "semi-local" regressions (SLR). This methodology provides the geographic flexibility offered by more complex tools such as locally weighted regressions, while maintaining the simplicity and computational speed afforded by traditional cross-sectional hedonic techniques.

Using this SLR methodology, we undertake a formal analysis of how the distribution of location premia changes over time. We find that the spatial and temporal variation of location premia is neither monocentric nor stable over time. Because these findings pose challenges for many of the empirical approaches that use house price data as part of urban research, we introduce a possible way forward that embraces local dynamism rather than imposes its absence.

To implement our SLR methodology, we first partition the single-family residential Maricopa County, Arizona into a grid of hexes. We then run standard log-linear hedonic regressions on overlapping subsamples of the data. The choice of hexes to select subsamples is driven by their symmetry. The hex structure allows us to construct subsamples of sales which are drawn from a center hex and up to six adjacent hexes. Using data from each seven-hex subsample, we run a standard hedonic regression, controlling for property characteristics, time, and the location fixed effect associated with the relative premia of adjacent hexes. We construct a similar subsample in and around each hex, obtaining a set of local implicit prices and location premia. Finally, we exploit the symmetry among hexes to estimate a relative location premia surface for the entire metropolitan area.

Our SLR technique provides a nice balance between the flexibility of locally weighted regression (LWR) models with the simplicity of a traditional hedonic specification. Recent research has demonstrated how LWR models can be used to more-accurately estimate the shadow prices of property characteristics over space, and thereby better estimate the location values of different areas in a city.[1] Such models are complex and computationally intensive, however, and this complexity can obscure the central urban economic questions they are intended to illuminate.

Like the LWR methodology, our SLR technique allows the shadow prices of a property's physical characteristics to vary over space. Because of this, the resulting locational value surface is free from the bias that can arise from the spatial variation in property physical characteristics over the city. In contrast to LWRs, however, SLRs are simple and quick to implement, making them a viable tool for addressing a variety of policy questions for which accurate location premia estimates may be required. The outcome of this framework and semi-local regressions is the goal of this research: a robust estimate of the spatial distribution of the premium paid for location. We contrast the location premia derived through this SLR approach and show how it differs in meaningful ways from a value surface estimated using standard hedonic techniques. Moreover, we are able to apply our technique to different time periods, allowing us to document how relative values of different locations withing a metropolitan area change over time.

The primary question we posed in the title of the paper: "How 'Fixed' are Fixed Effects?" has a simple answer: Our results show that static location fixed effects mask significant variation that occurs over time within metropolitan areas. These results are problematic for empiricists looking to use house prices as signals. Recovering the independent impact of a local policy change, for example, will require a more sophisticated assessment of what constitutes good "location" controls.

Our paper proceeds in several steps. In the next section, we discuss how space and location premia are estimated using standard hedonic techniques and why this traditional approach is subject to bias. In Section 3, we develop our methodology for using hex-based semi-local regressions to estimate the relative location premia of neighborhood across a metropolitan area. We implement this SLR technique in Section 4, using it to estimate the relative location premia of neighborhoods in Maricopa County, Arizona, and show how these locational value estimates differ from those derived using traditional hedonic fixed effects. In Section 5 we segment our data into different epochs to demonstrate how these location premia change

[1] See Agarwal, et al. (2021), Bindanset & Lombard (2014), Borst & McCluskey (2008), Cohen, Coughlin and Zabel (2020), Longhofer and Redfearn (2022), and Malone and Redfearn (2022), among others.

over time within Maricopa County. Section 6 concludes with a discussion of the implications for empirical work when location premia change over time.

2. The Value of Location

While von Thunen's model well approximated the location of heavy manufacturing and industry in early industrialization, land pricing in more dispersed and polycentric cities today reflects the varied bundle of amenities that are located throughout the region. In 1850, job density in cities was located around ports, rail lines, and water. Wealthier households and different firm technologies today demand a much larger set of amenities and transportation costs have enabled both households and firms to locate well outside the traditional CBD. It is the implied access from a given residential location to a large and diverse set of spatial amenities that net to the premium that is paid for a given location – that which is valued above the value of the service flow from the property characteristics within the property lines.

In practice, dummy variables for neighborhoods or zip codes in hedonic regressions are traditionally used to account spatial variation in location premia. Nevertheless, researchers that employ locational dummy variables seldom test if these fixed effects are constant over time. Moreover, even if these fixed effects were constant over time, they may not accurately measure actual location premia. This is because the shadow prices of the physical characteristics of the structures vary across the metropolitan area. As a result, the estimated location fixed-effect coefficients from a hedonic regression that covers the entire metropolitan area will likely suffer from omitted variable bias.

To see this, consider the traditional hedonic pricing equation that takes a form such as

$$V = \alpha_0 + \sum_j \beta_j X_j + \sum_n \delta_n I_n + \sum_t \tau_t I_t \tag{1}$$

Here, V is the value (sale price) of the property, X is a j-dimensional vector of property characteristics (building size, number of bedrooms, construction-quality variables, lot size, etc.), I is an n-dimensional vector of neighborhood dummy variables, and I is a t-dimensional vector of time dummies.

One might be tempted to assume that the estimated neighborhood coefficients, δ_n, from such a regression could be used to measure the relative values of different locations across the metropolitan area. After all, the whole purpose of these variables is to isolate neighborhood fixed effects so that the other coefficients of the regression can be more accurately estimated. Unfortunately, this is not correct. The reason is that the physical housing structures within a given residential neighborhood are generally quite homogeneous. For example, homes within a neighborhood are likely to have similar sizes, floor plans and construction materials, reflecting the vintage of when they were built. At the same time, past research has shown clearly that the Law of One Price does not hold with respect to housing attributes across a metropolitan area (McMillen and Redfearn, 2010). That is, the shadow prices of the physical characteristics of homes vary considerably across the metropolitan area. As a result, unless an extensive number of neighborhood/building-characteristic interaction terms are incorporated into X, these variables will inevitably be highly correlated with the neighborhood indicator variables, I_n, leading to the likelihood that the estimated δ_n will be biased.

Thus, it is clear that fixed-effect coefficients from a market-wide hedonic regression should not be used to estimate relative location premia because both the structure of the housing stock (size, age, style, etc.) and the shadow prices of these attributes vary considerably across the metropolitan area. As a result, location dummy variables will tend to incorporate both true location premia as well as market preferences for the type of housing stock found in each neighborhood.

This is not just a theoretical problem. Figure 1 shows how the physical characteristics of housing varies across neighborhoods in Maricopa County.[2] As is typical of many cities, homes on the periphery tend to be larger and newer than those in the center. The highest "quality" homes (a measure developed by the Maricopa County Assessor) tend to be located in the north, southeast, and especially the northeast edges of the city. The same is true for neighborhoods with high percentages of homes in gated communities. Similar clustered spatial variation can be seen in nearly all physical characteristics one might include in an hedonic regression. Indeed, while it is common to use distance from the CBD to control for land prices, the age of a dwelling does a reasonable job replicating the role of this distance in standard models in Maricopa because of the obvious correlation between the two.

[Insert Figure 1 here]

The spatial correlation of housing attributes would not be a problem if the shadow prices of these characteristics were the same across all neighborhoods. Figure 2, however, shows that this is not the case. The panels in this figure show the spatial distribution of the shadow prices of these same four physical

[2] Our hex-based method of defining neighborhoods will be explained below.

characteristics from semi-local hedonic regressions centered around each neighborhood.[3] The first panel in this figure shows that price elasticity of additional living area is highest in the northeast part of the city and lowest just southwest of downtown and the northwest and southwest edges of the city. Similarly, the impact of age is highest in the center of the city and on the northwest and southwest edges, the highest quality homes are typically found on the periphery and in the highly affluent northeast quadrant, and gated communities are located primarily in neighborhoods on the edges of the city.

[Insert Figure 2 here]

Combined with the heterogeneity of property characteristics across the city, these varying shadow prices suggest that in a traditional hedonic regression – which once again implicitly assumes that the hedonic coefficients are spatially invariant – the locational fixed effect coefficients will necessarily be biased as the regression erroneously attributes the impact of these physical characteristics to location.

While Figure 1 makes it clear that the physical characteristics of parcels vary widely across a metropolitan area, it also demonstrates that the housing stock tends to be relatively homogeneous across neighborhoods that are very close to each other. As a result, the neighborhood fixed-effect coefficients should suffer must less from omitted variable bias in regressions that only include "nearby" neighborhoods.

We propose to use this fact develop a relative location value surface for an entire metropolitan area using the location dummy variables from a series of "semi-local" regressions. To do this, we overlay a hex grid on the metropolitan area, assigning each parcel to the hex "neighborhood" in which it is located. By adjusting the distances between the hex centroids, we can arbitrarily alter the size of these neighborhoods, allowing them to be as large or as small as we desire.

For each neighborhood thus defined, we run a semi-local regression that includes all of the sales from neighborhood as well as those that are in immediately adjacent neighborhoods, including fixed effects to account for differences in values across these neighborhoods. Because adjoining neighborhoods are more likely to have homes that are relatively similar to one another, the fixed-effect coefficients from these regressions are more likely to capture true locational premia rather than the impact of housing stock differences across neighborhoods.

[3] The details of this semi-local regression methodology will be described shortly.

Due to the hexagonal structure of each neighborhood and the choice to limit each regression to include sales only in adjacent neighborhoods, each of these semi-local regressions will include (at most) seven neighborhoods with six neighborhood dummy variables. Thus, on their own these regressions will be inadequate to uncover relative location premia for an entire metropolitan area. Because we run separate regressions for each neighborhood in the metropolitan area, however, we can use the resulting matrix of dummy variables to estimate the relative location premia for all neighborhoods across the metropolitan area. We turn to describing the specifics of this methodology in the next section.

3. Semi-Local Regression Methodology

Imagine a hex grid of neighborhoods across a metropolitan area. Columns are indexed by letters and rows are indexed by numbers as shown Figure 3. Each neighborhood is adjacent to (at most) six other neighborhoods. For example, neighborhood C2, highlighted in green, is adjacent to neighborhoods B2, B3, C2, C4, D3, and D4.

[Insert Figure 3 here]

Consider how we might estimate the relative value of neighborhood B2 (highlighted in darker dark grey) compared to C3, denoted as ρ_{B2}^{C3}. First, we could run a regression that includes sales from neighborhood C3 and each of its neighbors, including dummy variables for each of these adjacent neighborhoods. Let δ_{B2}^{C3} be the coefficient on the dummy variable for neighborhood B2 in the regression centered on neighborhood C3; other coefficients will be denoted similarly. Because neighborhoods C3 and B2 are quite near one another, the housing stock in each should be relatively similar, as should the shadow prices of the physical housing characteristics. As a result, δ_{B2}^{C3} provides a direct estimate of the relative value of land in neighborhood B2 compared to C3 with little, if any, omitted variable bias.

This is not the only way we might estimate this relationship, however. Suppose instead that we ran a regression centered around neighborhood B2. In this case the negative of the coefficient δ_{C3}^{B2} would also estimate the relative value of land in B2 compared to C3. If these coefficients were estimated without error, we would have $\delta_{B2}^{C3} = -\delta_{C3}^{B2}$: the two coefficients would be the exact negatives of each other. In practice, however, each regression will include data that the other does not. For example, sales from neighborhoods C4, D3, and D4 will be included in the regression centered around C3 but not in the one centered around B2, while sales from neighborhoods A1, A2, and B1 will be included in the regression centered around neighborhood B2 but not the one for C3. Thus, these two estimates of ρ_{B2}^{C3} will likely differ from each other slightly.

There are yet other ways of estimating the relationship ρ_{B2}^{C3}. For example, the coefficient δ_{B3}^{C3} estimates the relative value of B3 compared to C3, while δ_{B3}^{B2} estimates the relative value of B3 compared to B2. The difference between these two coefficients, $\delta_{B3}^{C3} - \delta_{B3}^{B2}$, is therefore another way of estimating the value of land in B2 compared to C3, as is the difference between the coefficients δ_{C2}^{C3} and δ_{C2}^{B2}. Still more estimates come from the coefficients of the two regressions centered around neighborhoods B3 and C2 (see Figure 4).

[Insert Figure 4 here]

In all, we have six different ways of estimating ρ_{B2}^{C3}, the true relationship between neighborhoods B2 and C3. As noted above, if all of the coefficients from these regressions were estimated without error, each method would give us an identical value for ρ_{B2}^{C3}. In practice, however, each is estimated imprecisely because each uses data from a different subset of neighborhoods. Equally weighting the results of these six methods gives us the following estimate of the relative value of land in neighborhood B2 compared to C3 as follows:

$$\rho_{B2}^{C3} = \frac{[\delta_{B2}^{C3} - \delta_{C3}^{B2} + (\delta_{B3}^{C3} - \delta_{B3}^{B2}) + (\delta_{C2}^{C3} - \delta_{C2}^{B2}) + (\delta_{B2}^{B3} - \delta_{C3}^{B3}) + (\delta_{B2}^{C2} - \delta_{C3}^{C2})]}{6}$$

Based on this technique, we propose to estimate a relative land value surface across an entire metropolitan area as follows:

1. Create a hexagonal grid to define “neighborhoods” across the metropolitan area.
2. Estimate semi-local regressions for each neighborhood, with each regression including data from the neighborhood in question and each of the abutting neighborhoods.
3. Use the location fixed-effects coefficients from all of these semi-local regressions to estimate the relative land value premium of each neighborhood compared to its adjacent neighborhoods (the ρ_j^i’s).
4. Select a “base neighborhood” as the starting point against which the entire relative land value surface will be anchored.

5. Moving outward from the base neighborhood, use the estimated ρ_j^i's to calculate the relative value of each neighborhood compared to the base neighborhood.

To further elaborate on step 5 above, suppose that we had selected C3 as our base neighborhood. Each hex in the city can be assigned to a ring indicating its relative distance to the base neighborhood. In Figure 5 below, hexes B2, B3, C2, C4, D3, and D4 are all in Ring 1, while hexes A1, A2, A3, B1, B4, C1, D2, E3, and E4 are in Ring 2, and hexes D1 and E2 are in Ring 3. Other hexes are assigned to rings in a similar fashion.

[Insert Figure 5 here]

The value of hexes in Ring 1 relative to the center hex are simply their estimated ρ's: ρ_{C2}^{C3} is the value of neighborhood C2 compared to the center, ρ_{D3}^{C3} is the value of neighborhood D3 compared to the center, and so forth.

Moving on to Ring 2, where possible we calculate each hex's value using the most direct path to the center available. Thus, the relative value of neighborhood C1 can be estimated directly through its relationship with C2: $\rho_{C1}^{C3} = \rho_{C2}^{C3} + \rho_{C1}^{C2}$. In contrast, the relative value of neighborhood D2 can be most-directly calculated in two ways. Going through hex C2, we could estimate it as $\rho_{C2}^{C3} + \rho_{D2}^{C2}$. Alternatively, we could go through neighborhood D3 and estimate the relationship as $\rho_{D3}^{C3} + \rho_{D2}^{D3}$. Averaging these estimates, we have the relative land value of neighborhood D2 compared to C3 as

$$\rho_{D2}^{C3} = \frac{\rho_{C2}^{C3} + \rho_{D2}^{C2} + \rho_{D3}^{C3} + \rho_{D2}^{D3}}{2}$$

Continuing in a similar fashion across every neighborhood allows us to estimate a land value surface for the entire metropolitan area, all relative to a base location.[4]

4. Empirical Analysis

We demonstrate our new methodology using data on residential sales from Maricopa County, Arizona. Maricopa is home to Phoenix, a large and rapidly growing metropolitan area. It attracted out-sized capital flows during the housing bubble and its house prices reflected that on both the boom side and the bust.

[4] In our algorithm, we first try to value a hex using any neighbors in the next inner ring. Because metropolitan areas develop in unusual patterns, however, there are some contiguous hexes that only have a neighbor in the same ring. In these cases, trace the value path through this same-ring neighbor.

Since hitting bottom in 2009, the housing market there has tripled in price, but has only recently returned to pre-bust price levels. This is a housing market for which we anticipate a great deal of local variation. The aggregate churn in prices suggests a potential for greater local variation in price appreciation.

Our home sales data were provided through the Lincoln Institute for Land Policy and collected by the Maricopa County Assessor's Office, which collects and maintains property characteristic and sales data for use in their computer assisted mass appraisal (CAMA) efforts following guidelines established by the International Association of Assessing Officers (IAAO). Improved parcel sales were identified as transactions classified with a "Single-Family Residential" property type containing a single living unit with at least 400 square feet of finished living area. We eliminated parcels with lots smaller than 6,000 square feet (the minimum developable lot size under the current zoning code) and those larger than 100,000 square feet.

We next created a hex grid that spans all of the sales in our data. Our baseline grid has 50 hexes per row and 37 total rows of hexes. The resulting east-west distance between hex centers is just under 2 miles, and the land area of each hex is just over 4 square miles. Each sale is assigned a neighborhood based on the hex in which the property is located. In order to ensure sufficient sales for every semi-local regression, we restricted our analysis to include only hexes with at least 50 sales over the 12-year period.

Our final Maricopa County sample includes 630,531 single-family home sales from 2007 through 2018, spread across 441 different hex neighborhoods. Table 1 provides an expositional description of the variables used in our analysis, while Table 2 shows the summary statistics for each of these samples. Over this time frame, the typical home sale in Maricopa County involved a 21-year old house with 1,919 square feet of living area sitting on an 8,141 square foot lot. The median sale price was $225,000, or $116.02 per square foot.[5]

[Insert Tables 1 and 2 here]

Using these data, we run a tradition hedonic price regression using a standard set of variables including total living area, lot size, proxies for building quality (number of bathroom fixtures and quality rating), and indicator variables for special lot characteristics (gated community, golf course lot, arterial fronting, and mountain lot). The specification also includes time (year of sale) and location (hex id) fixed effects.

[5] Note that the median-priced sale is not the same as the median-sized house.

The results of this regression are shown in the first column of Table 3.[6] Overall this would generally be considered a strong model. The adjusted R-square is very high, and each of the estimated coefficients has the right sign and is a reasonable magnitude.[7]

[Insert Table 3 here]

The remaining columns of Table 3 show the summary statistics of our 441 semi-local regressions (SLRs).[8] Recall that each of these regressions is centered around a specific hex and includes data from that hex and each of its neighboring hexes. Each includes the same explanatory variables as the traditional hedonic, but the location fixed effect variables are limited to only those for the six neighboring hexes, because each SLR includes only data from these abutting hexes. Before we turn to the estimated coefficients, it is worth noting that Table 3 suggests that the individual SLRs appear to be very good hedonic regressions in their own rights. The 10th percentile of the adjusted R-squares is still 0.737. Moreover, the 90th percentile of the p-values for most of the regressors is 0.000, meaning that the estimated coefficients are significant at a level greater than 0.1 percent in more than 90 percent of all of the SLRs. From this we feel confident about the power of each individual SLR to estimate the local shadow prices of physical characteristics and location premia.

Comparing the SLRs results with those from the traditional hedonic, we see that the median estimated SLR coefficients are generally quite similar to those from the traditional hedonic model. These typical values mask significant variation over the city, however. As shown graphically earlier in Figure 2, the price elasticity of living area ranges from 0.440 to 0.677 (10th to 90th percentile values), while the price elasticity of age ranges from -0.199 to -0.038. Similar variation can be seen across other regressors as well.

Perhaps most relevant to our present application is the variation in coefficients of the lot-based indicator variables: Gated community, Golf course lot, and Mountain lot. Notice from Table 1 that these characteristics were present in only 4.2, 3.3 and 0.8 percent of the sales, respectively. Even more importantly, these lot characteristics tend to be clustered spatially, a fact which can be clearly seen for

---

[6] The year of sale and neighborhood fixed effect coefficients are not shown.

[7] The strong significance of the variables is not surprising given the large number of observations, although one would expect these estimates to still be significantly different than zero even with a much smaller sample size.

[8] Note that the p-values shown in these columns are the summary statistics of the p-values from the 441 SLRs and thus show the range of p-values obtained from these regressions. Importantly, they are *not* the p-values of the coefficients reported immediately above them.

gated communities in the last panel of Figure 1. Indeed, these variables are omitted in many of the SLRs because there are no lots with these characteristics in any of the neighborhoods included in that SLR. The spatial correlation of these desirable lot characteristics further indicates that the location values estimated through a traditional hedonic regression will be unreliable.

[Insert Figure 6 here]

Figure 6 shows the estimated location value surfaces derived from our SLR technique (left panel) and traditional hedonic fixed-effect coefficients (right panel), both of which have been normalized to a common central location (north of Midtown and west of the Phoenix Mountain Preserve). While the overall pattern of estimated location values across the metropolitan area are similar across these two methods, there are marked differences. These differences are easier to see in Figure 7, which maps the difference between the SLR and hedonic value surfaces. This figure makes it clear that the traditional hedonic undervalues locations south and west of downtown and overvalues those to the northeast, including the affluent communities of Scottsdale, Paradise Valley, and North Scottsdale.

[Insert Figure 7 here]

Upon reflection, this makes sense. While these areas are the most desirable in the area and command the highest location premia (as seen in both panels of Figure 6), they also have the highest quality *structures*. That is, part of the higher value of homes in these neighborhoods is due to the fact that they have the most expensive houses, independent of the land on which they sit. The traditional hedonic intermingles these effects, essentially attributing part of the structure value to the locations themselves, thereby overvaluing these locations. The opposite is true in the less affluent areas southwest of downtown Phoenix.

Before we use this methodology to address our central question, we note that we conducted a number of robustness checks on our hex-based semi-local regression methodology.[9] First, we tested whether the size of each hex affected our ability to estimate relative location premia. While using more hexes makes it possible to identify more nuanced variation in neighborhood premia, the general patterns observed in Figures 6 and 7 remained unchanged.

We also tested whether our minimum number of sales per hex restriction meaningfully impacted our results. As one would expect, the power of each SLR does increase with a larger sample size. The

[9] The results of all of our robustness checks are available from the authors upon request.

estimated location premia, however, did not appear to change based on limiting the analysis to hexes with more or fewer sales.

Finally, our results were also robust to changes in the choice of staring hex for our calculations. Essentially, choosing a different center hex to do the relative value surface calculations is analogous to choosing a different omitted neighborhood when using fixed effects in a traditional hedonic regression, resulting in a level shift across the entire metropolitan area without affecting relative values across neighborhoods.

5. Are Fixed Effects "Fixed"?

We are now in a position to turn to the central question of our analysis: How "fixed" are fixed effects? In particular, we want to investigate how our estimated location premia change over time. To do this, we segment the data into three epochs. Epoch 1 spans 2007 to 2010 (the "bust"), Epoch 2 spans 2011 to 2014 (the "recovery"), while Epoch 3 spans 2015 to 2018 (the "acceleration").

Figures 8 and 9 illustrate how location premia evolved over these epochs. In Figure 8, the left panel maps the SLR value surface during Epoch 1 while, while the right panel maps the *changes* in these location premia between Epochs 1 and 2. Figure 9 does the same for Epochs 2 and 3, with the left panel showing the Epoch 2 value surface and the right panel showing the changes between the epochs. In interpreting these figures, remember that all values are relative a specific location near the center of town.[10] As a result, any negative premia differences shown in the right panels of these two figures do not necessarily indicate that property values fell in these neighborhoods. Rather, they mean that the premia of such locations declined *relative to the central baseline neighborhood* between the two epochs.

With this interpretative note in mind, the changes depicted Figures 8 and 9 are striking, indicating that the changes in location premia over time are quite large and spatially clustered. Moreover, these changes seem to be explainable based on what we know about the evolution of the housing market in Phoenix during the housing crash, recovery, and acceleration. While the Phoenix metropolitan area was heavily affected by the subprime crisis, the impacts were not uniform across the city. Rather, the arrival of subprime lending brought more capital and higher prices in areas that had previously had less access to mortgage debt. The crash thereby resulted in particularly sharp declines in prices in lower priced

[10] For those who are interested, the center of this benchmark hex is latitude 33.53389, longitude -112.1123.

neighborhoods, and many of these neighborhoods struggled to recover for years following the end of the Financial Crisis.

[Insert Figure 8 here]

This is consistent with the spatial patterns of the location premia shown in Figures 8 and 9. During the bust (Epoch 1) the values of neighborhoods in southwestern Phoenix fell sharply relative to the higher priced, wealthier neighborhoods to the northeast, a pattern that can be seen in the relative location premia depicted in the left panel of Figure 8. During the recovery (Epoch 2), the impact of lower cost mortgages and lower cap rates was experienced first by luxury markets, and the right panel of Figure 8 shows that these neighborhoods did indeed see the greatest increase in relative values, even as location premia in much of the rest of the city declined.

[Insert Figure 9 here]

As the recovery accelerated during Epoch 3, however, its benefits began to reach those neighborhoods in southwest Phoenix that had struggled to rebound from the subprime crisis. Consistent with this, the right panel of Figure 9 shows that these neighborhoods saw their relative location premia rise dramatically, while the affluent neighborhoods in the northeast and saw their relative premia fall. Once again, this does not mean that home values fell in these affluent neighborhoods during Epoch 3. Rather, it means that the values of other neighborhoods increased at a relatively faster pace.

[Insert Figure 10 here]

Figure 10 provides another way of looking at how the location premia changed across these eras. The left panel plots the relative premia changes between Epochs 1 and 2, while the right panel plots the changes between Epochs 2 and 3. Each plot includes a red 45 degree line and a blue line trend line. On average, location premia fell across the Phoenix area compared to our baseline hex between Epochs 1 and 2. While many neighborhoods did see relative gains, there is no discernable pattern based on a neighborhood's relative premium during Epoch 1. In contrast, the right panel of Figure 10 shows that neighborhoods with the lowest relative values during Epoch 2 saw the greatest relative gains between Epochs 2 and 3, while the opposite occurred for neighborhoods with the highest relative values in Epoch 2. Thus, it appears that relative neighborhood location premia became more equal over this third, "acceleration" epoch.

One might wonder whether the temporal changes in locational premia we have uncovered are an anomaly, arising only because of some factors unique to Phoenix during this era. To test this, we conducted the same analysis using data from Wichita, Kansas as well. Whereas Phoenix is a fast-growing area with significant geographic barriers to growth in some directions, Wichita approximates the quintessential flat, featureless plain of urban economic theory, and grows at a slow, steady pace. As a result, Wichita provides a nice contrast against which to test the conclusions drawn from a dynamic market like Phoenix. As in Phoenix, we found that relative location premia in Wichita vary over time in systematic ways.[11] Given that Wichita is a very stable market that typically experiences very slow, steady aggregate price appreciation, the fact that relative location premia change over time here is compelling evidence that our results reflect a pattern that can be expected to arise in many if not all local markets.

We thus conclude that relative location premia do change over time and do so in way that are not simply random. As a consequence, locational fixed effects are not, in fact, “fixed.” The patterns above make it clear that the relative position of location premia migrate over time in ways that invalidate the use of traditional location fixed effects. By using location fixed effects, the dynamics made evident in the figures above are left to find other covariates that are correlated. This will likely complicate the results of empirical approaches that rely on their stability to recover their parameters of interest.

6. Implications and Directions for Future Research

At some basic level, we find studying cities interesting because they change. This is abundantly clear to anyone following fights about gentrification in recent years, sprawl and suburbanization in the 1990s, the end of American heavy manufacturing in the 1980s, and white flight in the decades before that. Most recently, Covid caused a large movement in housing demand from dense, core locations, to larger houses in periphery locations. All of these broad dynamics provide larger contexts in which other more local changes occur. Change over time in access or exposure to levels of school quality, crime, retail, and employment is typical within metropolitan areas.

In order to focus on a particular mechanism or dynamic, it has become common practice to impose location fixed effects, which implicitly assume that neighborhoods are unaffected by the changes around them. One might argue that location fixed effects are an effective tool because, despite all the changes that occur within a city, there is much that remains similar over time. The question in this paper is whether it is reasonable to use fixed effects in the presence of local change. We find that local neighborhood premia

[11] The results for Wichita are omitted for brevity but are available from the authors upon request.

change considerably over time within a major metropolitan area, suggesting that so imposing location fixed effects is inappropriate and potentially problematic.

In this paper, we develop a novel semi-local regression structure that allows us to extract a location premia surface. In an earlier paper, we made use of locally weighted regressions (LWR) to explicitly estimate the value of the land for each parcel in our data (Longhofer and Redfearn, 2022). While theoretically sound and computationally tractable, the LWR approach was viewed by practitioners as complex, and it requires some judgement calls about window size, weighting, and kernel choice. In this effort, we have developed a hex-based system that retains the essential flexibility of the LWR, while using a widely accepted hedonic technology. The hex-based sampling provided us a more flexible model, and its symmetry afforded us a chance to estimate a robust location premium surface. While we found marked spatial variation in the implicit prices on all of the housing characteristics, our larger surprise was the extensive and large movements in the implicit prices paid for location. It is clear among the results that there is no basis for assuming that spatial fixed effects are in fact fixed.

As we explored different hex sizes, we found that the smaller the hex size, the greater the variation in the location premium. In exploring the results, we found patterns fully consistent with patterns of luxury submarkets in Paradise Valley and Scottsdale, with exposure to subprime in lower price neighborhoods, and with access to more amenities in the burgeoning areas around ASU in Tempe. Clearly, it seems plausible that there are good controls to be found among the hexes – many places seem to retain similar ranks over time. But while perhaps possible to find these "good" controls, assuming that one can choose any or all of the neighborhoods in Maricopa and assume that their relative relationship regarding location premia seems decidedly premature. It should be noted that if it was the case that traditional local fixed effects really were fixed, the semi-local regression and hex approach could find them. Traditional pooled hedonic regressions are a special case of the SLR approach. As such, employing the SLR approach in the case when fixed effects are appropriate should yield estimates that, while perhaps noisy, would be quite similar to those found through traditional techniques – nothing like the clustered and evolving value surfaces we see in our data.

These results quickly raise other issues. First, does it matter? Does the imposition of fixed effect when location premia change matter for the implicit prices, for the implied price index, or for other empirical approaches like difference-in-differences (DiD). We showed how much variation there is implicit prices, but others have shown this before. We did not explore differences in aggregate house price indexes largely because it was not clear how best to do this. The most common quality-controlled price index – the repeat sales approach – assumed that all differences outside the price for the housing are constant. It then becomes a more nuanced conversation about what is being priced in a price index. But clear when the bundle of local amenities changes and/or their implicit prices change, the interpretation of the repeat sale index changes.

As for the implications of DiD in the presence of changing location premia, we have begun work on developing simple placebo tests. A deeper look is beyond the scope of this paper, but our initial forays into this suggest that imposing location fixed effects can produce a large number of false negatives and false positives. While preliminary, these results are consistent with a growing literature on the fragility of DiD approaches. It may be that this fragility is born in part by the inappropriate imposition of statis in a world in which metropolitan change is the norm.

We do not have an easy solution to these challenges. DiD and other quasi-experimental approaches have been quite popular in many urban and housing literatures. But in the end, the likely way forward is a return of choosing controls using stronger theoretical and empirical basis for inclusion. In this paper, we have provided compelling evidence that membership in the same metropolitan area alone is not a solid foundation for a good set of controls for the many local amenities and disamenities that we know are capitalized in houses.

7. References

Agarwal, Sumit, Ying Fan, Daniel P. McMillen and Tien Foo Sing. 2021. “Tracking the pulse of a city—3D real estate price heat maps.” *Journal of Regional Science*, 61(3), 543-569.

Bidanset, P. E. and J.R. Lombard, 2014. Evaluating spatial model accuracy in mass real estate appraisal: A comparison of geographically weighted regression and the spatial lag model. *Cityscape*, 16(3), 169–182

Borst, R. A. and W. J. McCluskey, 2008. Using geographically weighted regression to detect housing submarkets: Modeling large-scale spatial variations in value. *Journal of Property Tax Assessment & Administration*, 5(1), 21–21

Cohen, J. P., C. C. Coughlin and J. Zabel, 2020. Time-geographically weighted regressions and residential property value assessment. *The Journal of Real Estate Finance and Economics*, 60(1), 134– 154.

Demographia United States Central Business Districts 2020, 2020. http://www.demographia.com/db-cbd2000.pdf

Gordon, P. and Lee, B., 2015. “Spatial Structure and Travel: Trends in Commuting and Non-commuting Travels in US Metropolitan Areas.” *In Handbook on Transport and Development*, Edward Elgar Publishing.

Henderson, J.V., 2014. *Economic Theory and the Cities*. Academic Press.

Longhofer, Stanley D. and Christian L. Redfearn, 2022. “Estimating Land Values using Residential Sales Data.” *Journal of Housing Economics*, 58(A), December 2022.

Malone, T. and C. L. Redfearn, 2022. To measure globally, aggregate locally: Urban land, submarkets, and biased estimates of the housing stock. *Real Estate Economics*, 50(3), Autumn (Fall) 2022.

McMillen, Daniel P. and Christian L. Redfearn, 2010. “Estimation and Hypothesis Testing for Nonparametric Hedonic House Price Functions.” *Journal of Regional Science*, 50(3), pp. 712-733.

8. Tables

Table 1 – Description of Variables

Variable	Description
Sale price	Sale price of the vacant lot or improved property sale
Age of improvements	Year of sale minus the year the improvements were constructed; may be negative if the home was sold while the improvements were under construction
Arterial fronting	A lot located on an arterial road
Bathroom fixtures	Number of bathroom fixtures (bathrooms sinks, toilets, showers, tubs, etc.)
Living area in SF	Total square feet of finished living area
Lot size in SF	Total square feet of land area in the parcel
Golf course lot	A lot adjacent to the fairway or green of a golf course
Mountain lot	A lot located on a mountain
Quality	Residential quality class; values range from 0 to 7 with 3 being average and 7 the highest

Table 2 – Summary Statistics

Variable	Mean	Std.Dev.	1st Percentile	25th Percentile	Median	75th Percentile	Max.
Sale price	$280,132	$269,124	$30,000	$145,000	$225,000	$330,850	$1,333,000
Living area square feet	2,136	917	864	1,511	1,919	2,545	5,137
Price per square foot	$123.64	$65.86	$23.85	$82.48	$116.02	$152.22	$337.02
Land square feet	11,306	10,309	6,015	7,005	8,141	10,265	55,494
Age of improvements	25.9	18.7	1	10	21	39	72
Bathroom fixtures	8.2	3.1	3	6	8	10	17
Quality	3.451	0.681	2	3	3	4	6
Gated community	0.042	0.201	0	0	0	0	1
Golf course lot	0.033	0.179	0	0	0	0	1
Arterial fronting	0.059	0.236	0	0	0	0	1
Mountain lot	0.008	0.088	0	0	0	0	0

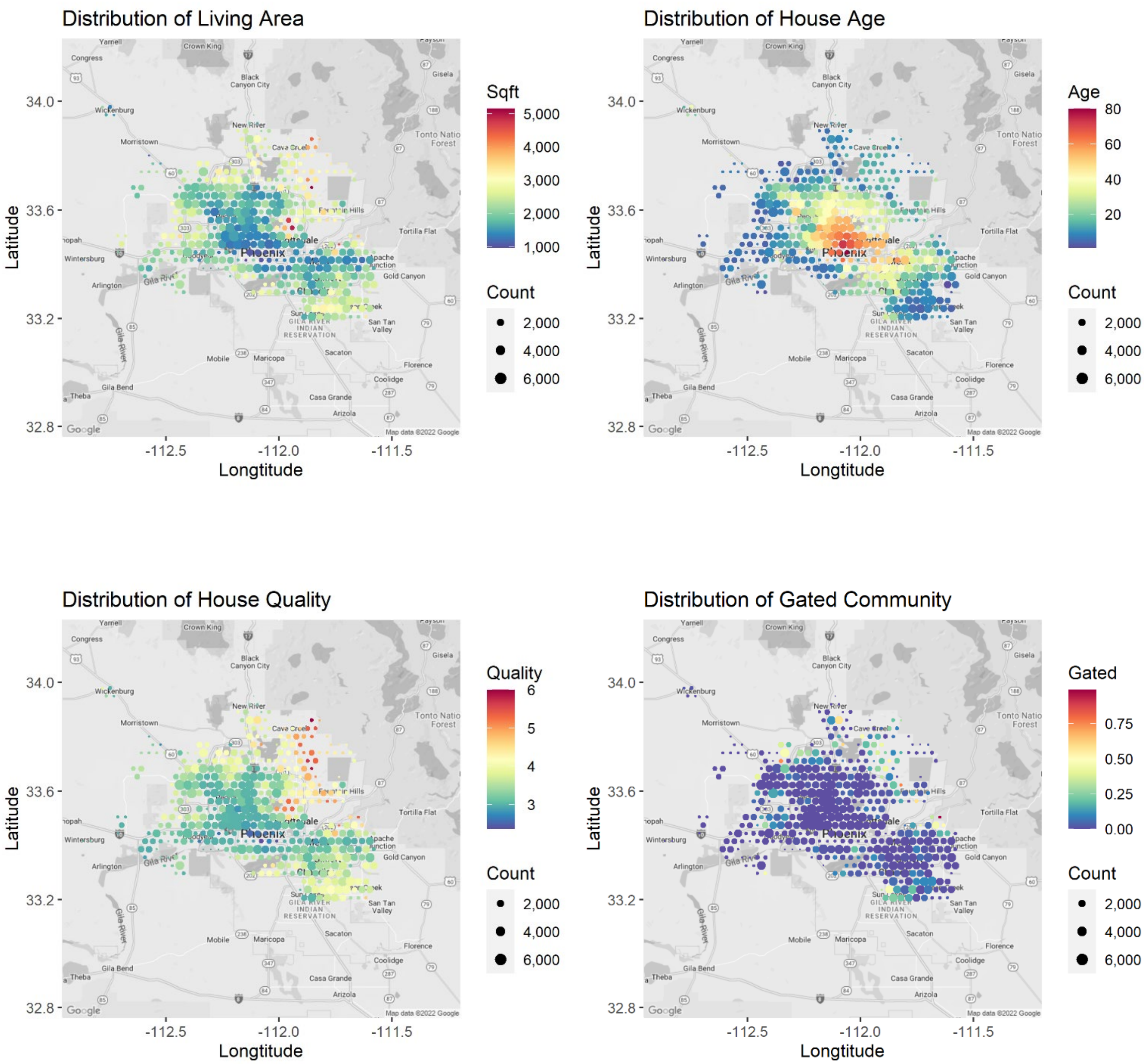
Table 3 – Hedonic and Semi-Local Regression Results – All Years

		Semi-Local Regression Summary Statistics				
Variable	Hedonic	Std. Dev.	Std. Dev.	10th Percentile	Median	90th Percentile
Ln(Living area)	0.583	0.558	0.100	0.440	0.556	0.677
	(0.000)	(0.000)	(0.002)	(0.000)	(0.000)	(0.000)
Ln(Lot size)	0.204	0.203	0.089	0.103	0.217	0.281
	(0.000)	(0.006)	(0.039)	(0.000)	(0.000)	(0.000)
Ln(Age)	-0.112	-0.112	0.065	-0.199	-0.105	-0.038
	(0.000)	(0.012)	(0.070)	(0.000)	(0.000)	(0.000)
Bathroom fixtures	0.013	0.014	0.011	0.002	0.013	0.027
	(0.000)	(0.044)	(0.156)	(0.000)	(0.000)	(0.079)
Quality	0.112	0.113	0.110	0.044	0.105	0.172
	(0.000)	(0.019)	(0.093)	(0.000)	(0.000)	(0.000)
Gated community	0.085	0.056	0.132	-0.087	0.042	0.193
	(0.000)	(0.114)	(0.238)	(0.000)	(0.000)	(0.488)
Golf course lot	0.204	0.148	0.138	0.000	0.132	0.308
	(0.000)	(0.042)	(0.153)	(0.000)	(0.000)	(0.051)
Arterial fronting	-0.065	-0.062	0.050	-0.131	-0.057	-0.006
	(0.000)	(0.092)	(0.218)	(0.000)	(0.000)	(0.426)
Mountain lot	0.031	0.007	0.095	-0.074	0.000	0.097
	(0.000)	(0.226)	(0.285)	(0.000)	(0.085)	(0.694)
Constant	5.843	6.334	1.168	5.130	6.369	7.549
	(0.000)	(0.003)	(0.046)	(0.000)	(0.000)	(0.000)
Observations	630,531	9,366	6,192	1,602	8,192	17,590
Adjusted R-square	0.868	0.806	0.056	0.737	0.817	0.866

Notes: The first column shows the results of a traditional hedonic regression including all observations across the entire county; p-values are shown in parentheses below the coefficients. The remaining five columns show the summary statistics for the 441 semi-local regressions (SLRs) centered on each hex; summary statistics for the p-values are shown in parentheses below the coefficient statistics. These show the range of the p-values obtained in the SLRs, not the p-values of the coefficients reported above them. All regressions included year and neighborhood (hex) fixed effects (not shown). The dependent variable in all regressions is the natural log of sale price.

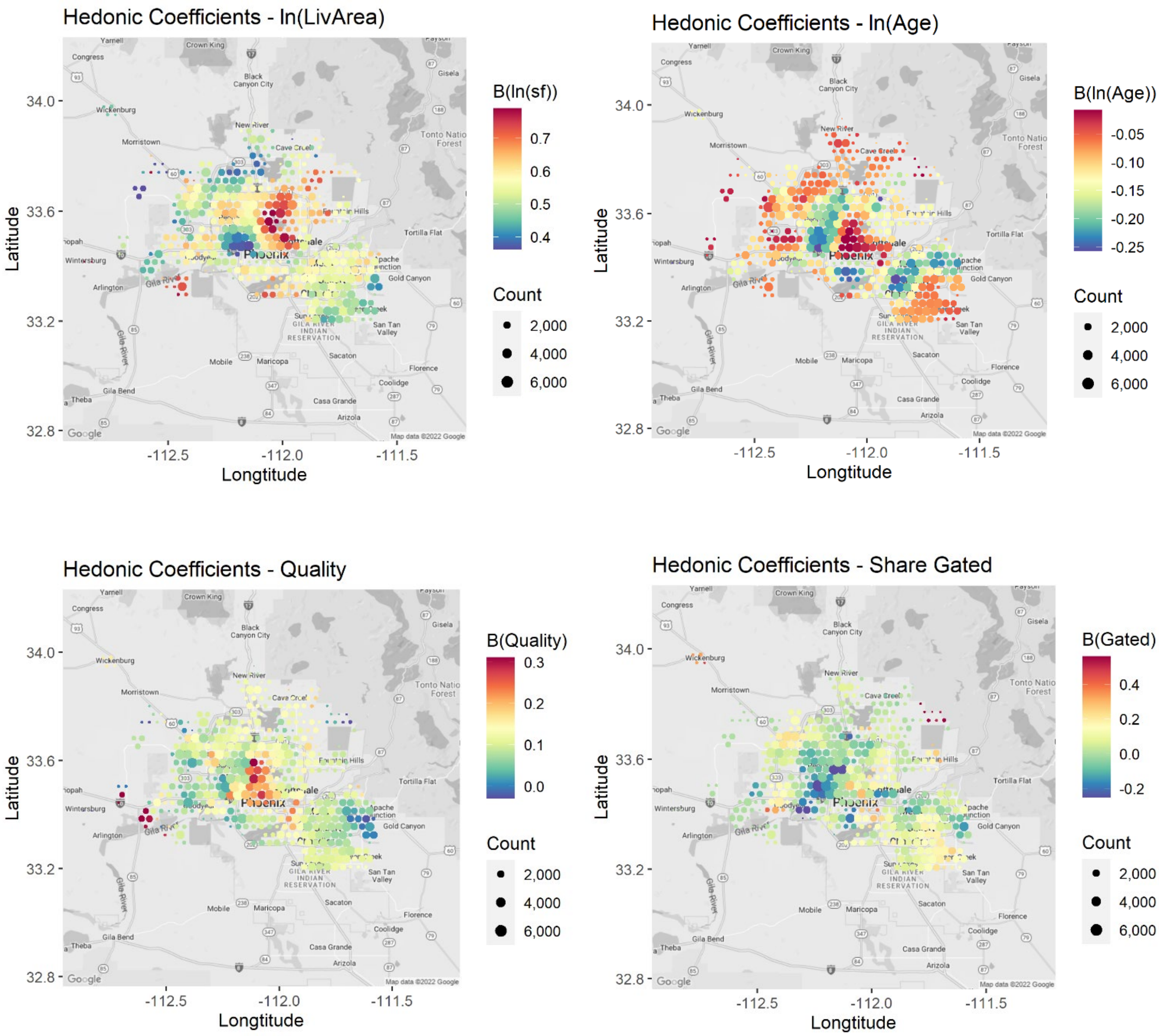
9. Figures

Figure 1 – Spatial Distribution of Physical Characteristics



Notes: Panels show the spatial distribution of the physical property characteristics of sales in each neighborhood (defined by hexes). Living Area and Age are median values within the hex; Quality and Gated Community are mean values within the hex.

Figure 2 – Spatial Distribution of Shadow Prices



Notes: Panels show how the estimated coefficients of various physical housing characteristics differ across neighborhoods (defined by hexes). Coefficients come from semi-local hedonic regressions that includes sales from the neighborhood and any adjacent neighborhoods.

Figure 3: Defining the Overlapping Regression Samples

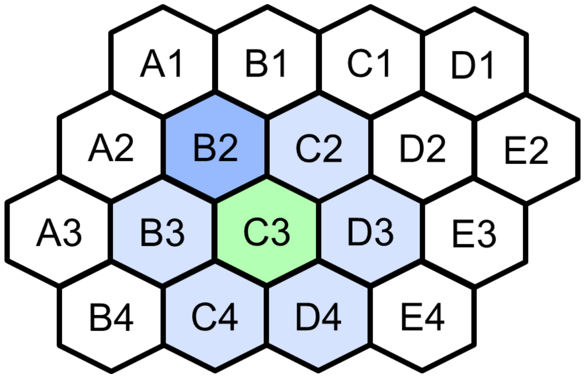


Figure 4: Exploiting the Symmetry between Hexes

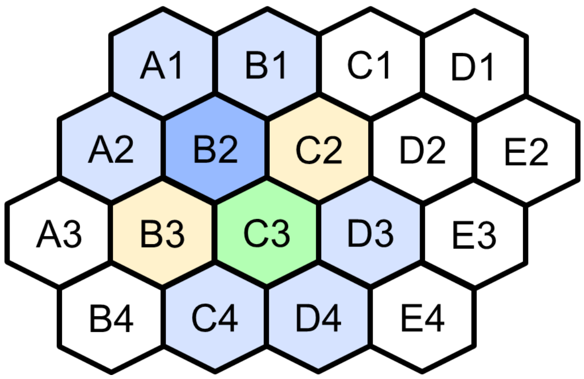


Figure 5: Using Rings to Interpolate a Value Surface

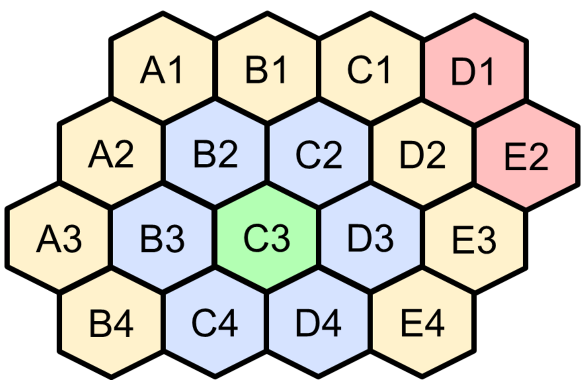


Figure 6: Value Surface Comparisons – All Years

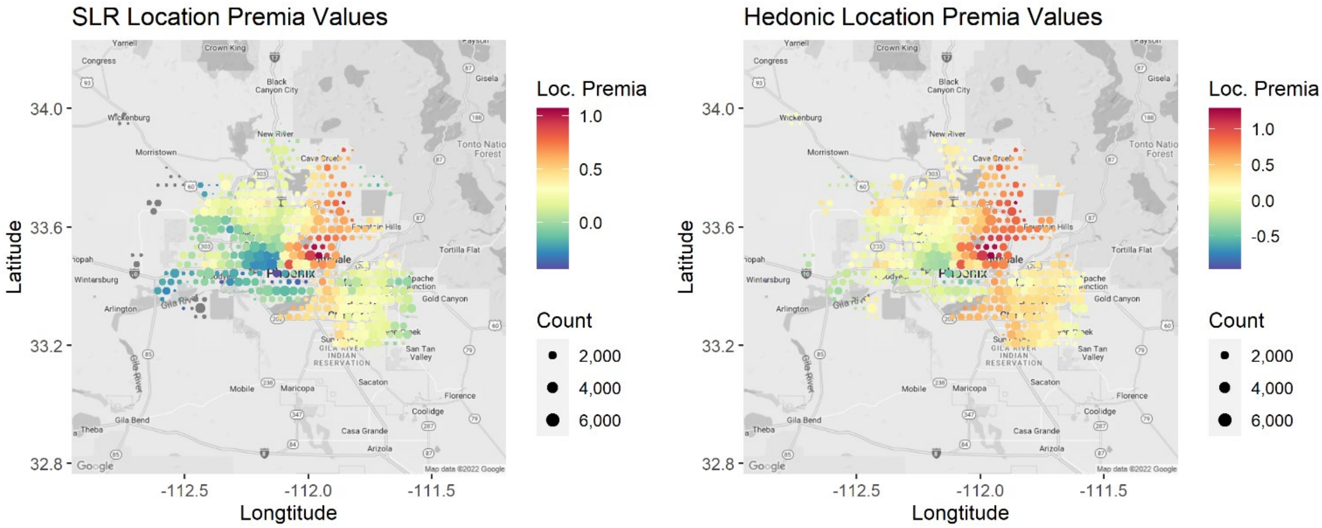


Figure 7: Value Surface Differences (SLR - Hedonic) – All Years

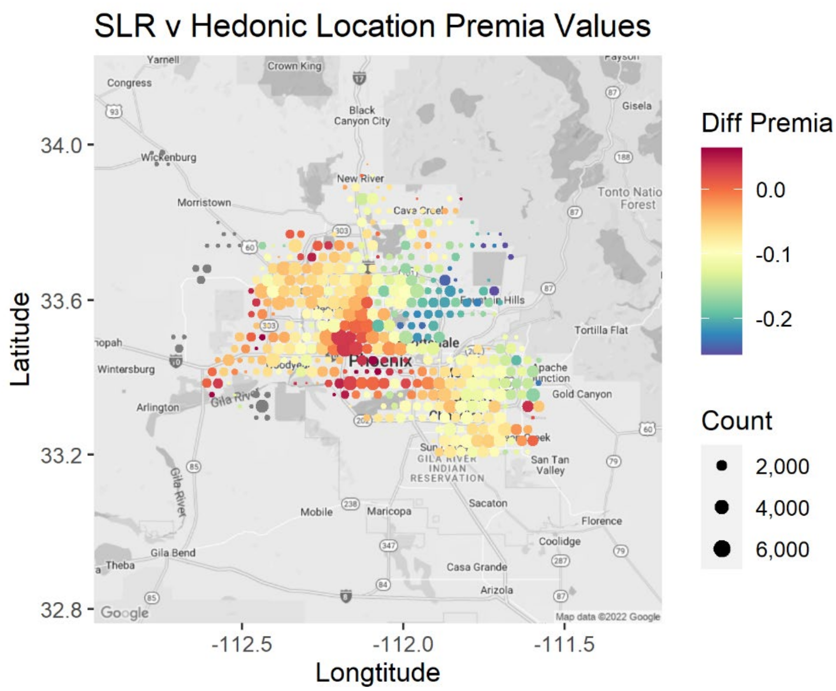
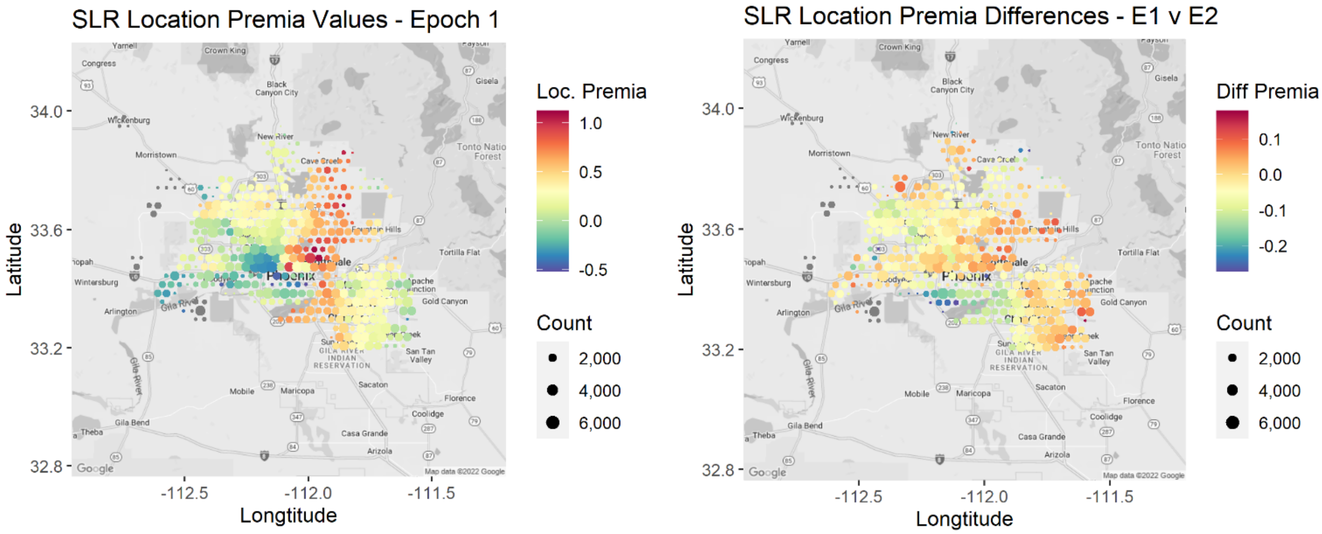
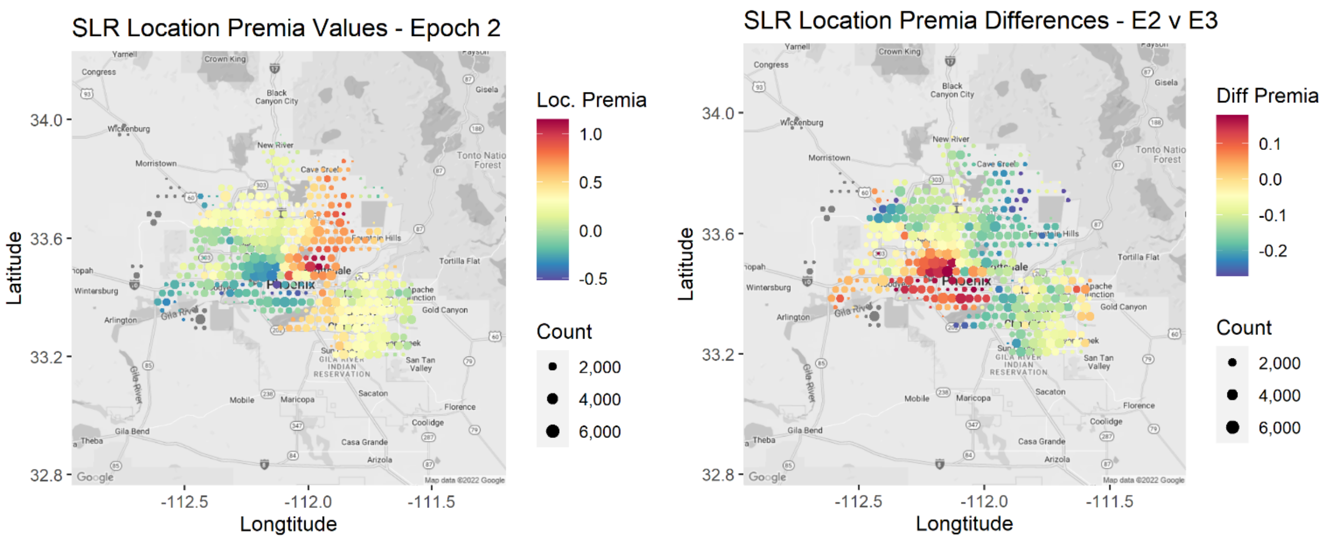


Figure 8: Epoch 1 Location Premia and Change from Epoch 1 to Epoch 2



Note: The left panel of this figure shows the location premia relative to a central location during Epoch 1 (2007-2010). The right panel shows the change in location premia between Epoch 1 and Epoch 2 (2011-2014).

Figure 9: Epoch 2 Location Premia and Change from Epoch 2 to Epoch 3



Note: The left panel of this figure shows the location premia relative to a central location during Epoch 2 (2011-2014). The right panel shows the change in location premia between Epoch 2 and Epoch 3 (2015-2018).

Figure 10: Changes in Location Premia over Time