Highly Disaggregated Topological Land Unavailability*

Chandler Lutz
Copenhagen Business School

Ben Sand
York University

First Available Draft: August 10, 2017

November 9, 2017

Abstract

Using high resolution satellite imagery data and GIS software, we compute the percentage of undevelopable land – Land Unavailability – at levels high levels of geographic disaggregation down to the zip code level. Our Land Unavailability measure expands on the popular proxy from Saiz (2010) by (1) using higher resolution satellite imagery from the USGS; (2) more accurate geographic boundaries; and (3) multiple levels of disaggregation. First, we document the importance of using precise boundary files and disaggregated data in the construction of land unavailability. Less precise boundary files lead to measurement error in land unavailability that can violate standard instrumental variable assumptions, while larger aggregated areas (e.g. MSAs in California and the Southwest) have larger variance in Land Unavailability and thus yield less precise two-stage least squares estimates. Next using data at the zip code level with nearly complete coverage of the contiguous US we show that Land Unavailability is uncorrelated with housing demand proxies, validating Land Unavailability as an instrument for house prices. Further we find that within local housing markets (e.g. after controlling for aggregated Land Unavailability) that higher Land Unavailability is associated with lower house price growth but higher absolute prices, congruent with inter-market mobility and financially constrained households substituting expensive for cheap housing during a boom.

JEL Classification: R30, R31, R20;

Keywords: Topological Land Unavailability, Real Estate, Housing Market

*Lutz: cl.eco@cbs.dk. Sand: bmsand@yorku.ca

1 Introduction

Housing is the largest financial asset for the typical US household (Tracy and Schneider, 2001). Economists have thus naturally connected fluctuations in housing markets with the causes of the Great Recession (Mian and Sufi, 2009, 2011, 2014), business cycle dynamics (Leamer, 2007), household consumption decisions (Bostic et al., 2009; Mian et al., 2013; Mian and Sufi, 2015; Aladangady, 2017), the efficacy of fiscal and monetary policy (Agarwal et al., 2017; Gabriel and Lutz, 2014; Gabriel et al., 2017), education and life cycle choices (Charles et al., 2015), industry composition and wages (Beaudry et al., 2012, 2014), firm formation (Adelino et al., 2015), corporate investment (Chaney et al., 2012), and financial market behavior (Lutz et al., 2016). As several potential sources of endogeneity obfuscate these economic relationships, researchers have sought exogenous variation and subsequently an instrument in their search for causal inference. A popular instrument in the housing literature is the topological Land Unavailability proxy of Saiz (2010).[1] The Saiz Land Unavailability measure is constructed by computing the percentage of land that is not developable due to either (1) a steep slope (e.g. mountainous land) or (2) water or wetlands (e.g oceans, lakes, etc.). Using this instrument economists typically pursue a two-stage least squares approach where they regress house price growth on land unavailability in the first stage and then the outcome of interest on predicted house price growth in the second stage.[2] This process hence yields the causal relationship between house prices and the outcome of interest.

In this paper, we build on Saiz (2010) and extend his work in several directions to build a new and updated Land Unavailability proxy. But first we review the Saiz methodology. Saiz constructs his measure of Land Unavailability by computing the percentage of undevelopable land within a 50 km radius circle each MSA's first central city an MSA centroid. Yet MSAs are small in the Northeast, where house price growth was relatively

[1]Saiz's paper more broadly studies cross-geography housing market elasticities using both land unavailability and a proxy for housing market regulation. Yet the proxy for housing market regulation is likely endogenous leaving topographical land unavailability as the candidate instrument.

[2]In a related approach, researchers also interact Saiz Unavailability with national factors correlated with demand growth, such as national interest rates or proxies for labor demand shocks. See for example Chaney et al. (2012) and Aladangady (2017).

muted during the 2000s, but large in California and the Southwest, areas that experienced high house price volatility before and during the crisis. Hence, the difference between MSA polygon size and the 50 km circle used in the construction of Saiz Land Unavailability is correlated with house price growth, potentially violating the exclusion required for two-stage least squares (2SLS) estimates that use Saiz Unavailability as an instrument. Indeed, the exclusion restriction requires that the measurement error from the instrument is uncorrelated with the second stage errors (Fuller, 2009). Using canonical examples that measure the impact of house price growth on consumption or employment, results suggest that the Saiz unavailability measurement error (due to variation in polygon size) is correlated with second stage residuals. In contrast, precise boundary files, like the ones used in our construction of Land Unavailability, mitigate measurement error related to polygon size and thus eliminate this source of bias in 2SLS estimates.

In the construction of our Land Unavailability measure, we use more accurate satellite data that is now available from the United States Geographical Survey (USGS). Our Land Unavailability measure also exploits more precise and geo-spatially consistent polygon areas in our calculation of undevelopable land. This easily allows us to extend our computation method to various units of economic geographic aggregation including MSAs, counties, and zip codes.

Using our zip code GIS data, we show that Land Unavailability is more variable in large MSAs, where MSAs are the typical the geographic aggregate for Land Unavailability in the housing and urban literature. Higher variance for larger MSAs means that there is more uncertainty in models that predict house price growth using Land Unavailability. Assuming that estimated first stage regression estimates are not equal to the true population estimates (estimated with error), this implies that 2SLS estimates that use Land Unavailability will be less precise for larger MSAs, the same MSAs that experienced large house price volatility during the 2000s. Thus more disaggregated proxies for Land Unavailability may be more relevant for researchers.

We also examine the correlation between Land Unavailability and proxies for housing demand as the use of Land Unavailability has depends on its exogeneity relative to demand factors. Indeed, there has been debate in the literature on validity of Land Unavailability as an instrument and its exogeneity.[3] Using zip code level data, we find that Land Unavailability is not positively correlated with housing demand factors and the inclusion of housing demand factors does not mitigate the predictive power of Land Unavailability.

Finally, we further examine the predictive power of Land Unavailability with regard to house prices at the zip code level. To our knowledge, we're the first to undertake such and evaluation. We find zip code Land Unavailability is a strong predictor of house price prices. Yet our results also show that Land Unavailability in the areas surrounding the a given zip code is a key predictor of house prices, highlighting how households move within cities in response to a neighborhood price shock. After controlling for more aggregated Land Unavailability or region fixed effects, higher zip code Land Unavailability is associated with lower house price growth but higher absolute prices. This result is in line with mobile financial constrained home buyers within local housing markets.

The rest of this paper is organized as follows: Section 2 describes the data; in section 3 we provide an overview of the Saiz methodology and consider the consequences of the measurement error related to the Saiz proxy in 2SLS applications; the Land Unavailability measure developed in this paper is covered in section 4; we discuss aggregated versus disaggregated Land Unavailability in section 5; the validity of Land Unavailability is assessed in section 6; section 7 documents the predictive power of Land Unavailability; and section 8 concludes.

2 Data Sources

The United States Geographical Survey (USGS) provides the two main datasets that we use to measure slope, water, and wetlands land unavailability.[4] The first is the USGS National Elevation Dataset (NED) 3DEP 1 arc-second Digital Elevation Model (DEM). The 1 arc-second DEM data provide continuous coverage of the United states at approx-

[3]See, for example, Mian and Sufi (2011, 2014) and Davidoff et al. (2016).
[4]https://viewer.nationalmap.gov/launch/

imately a resolution of 30 meters.[5] The original Saiz uses 3 arc-second DEM data with a resolution of approximately 90 meters. The DEM data allow us to calculate slope files and hence the percentage of land unavailable due to a steep slope. Our second main dataset is the USGS 2011 Land Cover Dataset.[6] These data use Landsat imagery to classify land use in the US. The relevant categories for this paper are water (oceans, lakes, rivers, etc.) and wetlands. From the Land Cover data, we measure the portion of undevelopable due to wetlands and water.

2.1 Other Data

In addition the above data, our study also includes Shapefiles for various geographies from the US Census Bureau and satellite imagery from Google Maps.

Our data also include a number of key housing and control variables: House prices are from the FHFA (repeat-sales house prices at the MSA level) and Zillow (hedonic house prices available down to the zip code level); from the 2000 US Census at the zip code level we retain the percentage of people with a college education, percentage of foreign born, housing density, a zip code level amenities index from large internet company that aggregates information on access to restaurants and bars, retail shopping, public transit and other amenities. From the County Business Patterns data we compute the (Bartik, 1991) shock of labor demand. We also map the county Bartik Shock to the zip code level using the Missouri Data Bridge.

3 A Review of the Saiz 2010 Methodology

The groundbreaking work of Saiz (2010) provides the foundation for this paper as it was the first to use detailed satellite imagery and GIS methods to compute proxies of land unavailability. Saiz (2010) uses the USGS 90 meter DEM to compute the percentage of land unavailable due to a steep slope. Specifically, he notes that land with a slope above 15 percent faces architectural impediments to construction. The second dataset

[5] For a sample file, see https://www.sciencebase.gov/catalog/item/5903e5b0e4b022cee40c773d. The Coordinate Reference System (CRS) used for these data is GRS80.

[6] For a sample, see https://www.sciencebase.gov/catalog/item/581d5a13e4b0dee4cc8e5120. The CRS used for these data are NAD83.

that Saiz uses is the 1992 Land Cover dataset. Using this dataset, combined with digital contour maps, Saiz measures the percentage of land that is unavailable due to oceans, lakes, rivers, etc. Saiz computes the percentage of unavailable land from a 50 kilometer radius around the centroid of each MSA's first central city.

As an example of the geographies that Saiz uses within each MSA, we plot Google satellite imagery for the Los Angeles-Long Beach MSA in figure 1. Here, the blue outlined area represents the polygon boundary for the Los Angeles-Long Beach MSA. The orange polygons are the central cities within the Los Angeles-Long Beach (Los Angeles, Long Beach, Pasadena, and Lancaster). The red dots are the centroids of each central city polygon, and the red circle represents a 50 km radius around the first central city centroid for the MSA (in this case, the Los Angeles central city). The 50 km circle around the first central city centroid is the area used by Saiz to assess Land Unavailability. Clearly, the location of the first central city centroid determines the land used in the calculation of Saiz unavailability: The Saiz circle with a 50 km radius captures most of the Los Angeles area, but does not cover the central city around the Lancaster and Palmdale areas, two cities with a combined 2000 population of over 230,000, or eastern Los Angeles around Pomona. The Saiz circle also does not cover the disjointed polygons representing the Catalina islands. More generally, larger polygons are less likely to be covered by the Saiz circle. We can see this in figure 2 which plots Google satellite imagery for the Riverside-San Bernardino MSA. The layout of figure 2 mirrors figure 1. The Riverside-San Bernardino is extremely large and the Saiz circle only covers a small fraction the total land area. While much of the uncovered land area is sparsely populated, the Saiz circle does miss the central cities in Palm Springs, Palm Desert, and Temecula. Interestingly, as the first central city for Riverside-San Bernardino used to compute the Saiz circle (the Riverside central city) sits near the southwestern border of the polygon, the Saiz circle spills into eastern Los Angeles and all the way to Anaheim. In fact for the Riverside-San Bernardino MSA, nearly 25 percent of the Saiz circle extends outside of the MSA polygon boundary and reaches areas, like Anaheim, with notably different economics and geographies relative to Riverside-San Bernardino. More broadly, many central cities are located near MSA polygon boundaries and this is especially for coastal MSAs. Figures 3 and 4 plot the geographic polygons and Saiz circles for MSAs along the southern Florida coast and Lake Michigan respectively. Notice that all of the first central cities are adjacent to the coast around both the Florida coast and Lake Michigan. The coastal nature of the first central cities in these areas matches our expectations given the historical importance of waterways for economic development. Yet in the computation of land unavailability using the Saiz circles, this implies that nearly half of land is going to be unavailable due to water (assuming a straight coastline that splits the Saiz circle), irrespective size or shape of the polygon or how it extends inland. Take for example Benton Harbor, MI (St. Joseph on the Google Map; the east side of Lake Michigan). The Saiz circle completely covers this rather small MSA polygon but also extends markedly into Lake Michigan. Using the Saiz circle to measure Land Unavailability suggests that 47 percent of land in Benton Harbor is undevelopable due to water (Saiz calculates that total land unavailability in Benton Harbor is 50 percent). Hence, the geographic proximity of the Benton Harbor's first central city relative to the coast, plays a crucial role in determining its overall land unavailability. The geographic importance of an MSA's first central city relative to the coast plays an equally important role for larger polygons like Chicago (southwest of Lake Michigan in figure 4) or rectangular or oddly shaped polygons like those along the Florida coast (figure 3). Overall, figures 3 and 4 show that the computation of land unavailability using the Saiz circles for coastal MSAs will deem a substantial portion land undevelopable due to water.

Together, the above figures highlight instances where a circle with a 50 km radius around the city centroid may produce land unavailability calculation peculiarities. Generally, the Saiz circles are going to under cover MSAs that span large geographic areas, but cover more land area than comparatively smaller polygons. Obviously if differences between MSA polygons and the Saiz circles are random, it will not bias regression estimates that examine the relationship between the house price growth and land unavailability computed using the foregoing technique. Unfortunately however, MSAs in California and the Southwest generally are larger in geography and these areas also experienced large house price growth in the 2000s. In contrast, in the Northeast for example, MSAs are generally smaller and experienced lower housing volatility during the 2000s. Figure 5 extends the above figures and plots all MSA polygons for the Saiz dataset in blue and the corresponding circles with a 50 km radius centered around the first central city centroids in red. Clearly, MSAs in the Northeast are smaller and well covered by the 50 km radius circles, while those in Southwest are much larger compared to the circles. Figure 6 further highlights the geographic coverage of the Saiz circles relative to the MSA polygons. Here, we plot the differences between the MSA polygons and the saiz circles: Red areas are differences between MSA polygons and the Saiz circles within the MSA boundaries; yellow these differences outside the polygon boundaries; and blue areas are the differences outside of the polygon in the oceans, the Gulf of Mexico, or the Great Lakes. First, congruent the above figures for the Florida Coast and the Great Lakes, many coastal MSAs have nearly 50 Land Unavailability due to due water as their first central cities lie along the coast. Next, the difference in the inland coverage of the Saiz 50 km radius circles across US geographies is stark: In the Northeast the circles are typically larger than the MSAs and cover more land than the MSA polygons (yellow), while MSAs in the Southwest and California, cities that experienced large housing variance in the 2000s, are notably larger than their corresponding circles.

In housing analyses using Land Unavailability, researchers often employ the following first stage regression:

$$\Delta \ln HousePrice = \beta_0 + \beta_1 Unavailability + \varepsilon \tag{1}$$

We run this regression using 2002 - 2006 FHFA house prices and the Saiz Unavailability proxy. As expected, $\hat{\beta}_1$ is positive and significant (White t-stat = 11.39), indicating that Land Unavailability predicts house price growth. We also retain the residuals from the regression in equation 1. The correlation between these residuals and the difference in the area between the MSA polygons and their 50 km radius circles is 0.33 (White t-stat from a regression = 4.58), meaning that areas where this difference is large larger residuals for the regression in equation 1 and positively predict house price growth. Furthermore, the correlation between the size of the difference in MSA polygons and the Saiz circles and the Saiz Unavailability is 0.16 (White t-stat from a regression = 2.35). These findings, together with those in figures 5 and 6, show that Saiz Unavailability is highly correlated with size of the difference between MSA polygons and their corresponding circles thus that the regression estimator, β_1 in equation 1, is biased. Even though β_1 is biased, two-stage least squares (2SLS) estimates remain consistent as long as the IV exclusion restriction holds (in this case, the correlation between the second stage errors and the measurement error from the instruments is zero).[7] Yet if the exclusion restriction with regard to the measurement error does not hold; that is if the measurement error from the Saiz Unavailability proxy is correlated with the errors from the regression of interest, then the IV estimates will be inconsistent. This violation of the exclusion restriction will obviously depend on the second stage regression of interest and we examine this issue within common applications next section.

3.1 Consumption, House Prices, and Saiz Land Unavailability

The relationship between house price fluctuations and consumption in the lead up to the Great Recession and during its aftermath has been widely studied by economists due to its wide-reaching economic implications (Bostic et al., 2009; Mian et al., 2013; Mian and Sufi, 2015; Aladangady, 2017). The regression of interest is

$$\Delta \ln C_{it} = \beta_{0t} + \beta_1 \Delta \ln HP_{it} + u_{it} \tag{2}$$

If $\Delta \ln HP_{it}$ is used as proxy for housing wealth for household i, β_1 can be interpreted as the elasticity of consumption with respect to housing wealth. Equation 2 is thus of key importance as a standard representative-agent model with consumption risk insurance implies β_1 is zero.

As many unobserved factors or reverse causality may contaminate causal estimates of

[7] We can think of the (Saiz, 2010) proxy as Land Unavailability measured with error. Thus, our aim is to estimate $Y = X\beta + \varepsilon$ using instruments $\tilde{W} = W + \eta$, where all matrices have the typical dimensions. The 2SLS Instrumental variables estimates is $\hat{\beta}_{IV} = (\tilde{W}'X)^{-1}\tilde{W}'Y$. Substituting $Y = X\beta + \varepsilon$ and $\tilde{W} = W + \eta$, we have $\hat{\beta}_{IV} = \beta + (W'X + \eta'X)^{-1}(W' + \eta')\varepsilon$. Then if the exclusion restriction holds with respect to W' and η, $(\frac{\tilde{W}'\varepsilon}{n} \xrightarrow{p} 0)$, $\hat{\beta}_{IV} = \beta$. Indeed, a key assumption for the use of instrumental variables in error in variables problems is that the measurement error of the instrument is uncorrelated with the second stage errors (Fuller, 2009).

β_1 in equation 2, a natural approach is to pursue an instrumental variable research design using Saiz Unavailability as an instrument. The first stage is given above in equation 1. We estimate equation 2 using county level data as in Mian et al. (2013). To do this, we map Saiz MSA Land Unavailability (recall Saiz Land Unavailability is only available at MSA level) to the county level using the corresponding shape files and the number of housing units as weights. Zillow county-level data and county-level auto sales proxy house prices and consumption respectively. The growth in these variables is taken from 2002 to 2006 (boom) and from 2006 to 2009 (bust).

As noted above, for the exclusion restriction to hold when the Saiz unavailability is measured with error, the correlation between the Saiz Unavailability measurement error and the regression error from equation 2 needs to be zero. In most applications, it is not possible to check the exclusion restriction with respect to measurement error, but in this paper we compute Land Unavailability using precise geographic boundaries (our process is described in depth below). Thus we are able to calculate the measurement error in Saiz Unavailability emanating from discrepancies between the polygon shapefiles and Saiz circles. Specifically, for each county we assume that the land covered by the Saiz circle is perfectly measured by Saiz unavailability. Then, the rest of the area in the polygon (red areas in figure 6) is proxied by Saiz Unavailability and measured with error. Thus using precise polygon boundary files, we can calculate the measurement error associated with the Saiz circles within the polygon.[8] To get a sense of this measurement error, figure 7 plots the within polygon difference between county polygons and their corresponding Saiz circles. The setup of figure 7 mimics the MSA data in figure 6. Comparing figures 6 and 7, the geographic pattern of the differences between the Saiz 50km radius circles and the polygons for counties is even more pronounced: the Saiz circles with a 50 km radius largely cover polygons in the Northeast, but are much smaller than polygons in the western US. As there are many more counties than MSAs, the Saiz 50km radius circles also cover more area compared to many polygons in the Northeast, but completely

[8]Noting that within the Saiz circle, Land Unavailability is equally weighted by area and assuming that Saiz Unavailability perfectly measures Land Unavailability within the Saiz circle, we have $\text{True Unavailability} = \text{Saiz}\frac{\text{SaizArea}}{\text{TotalArea}} + (\text{Saiz} + \text{Error})\frac{\text{NotSaizArea}}{\text{TotalArea}} = \text{Saiz} + \text{Error}\frac{\text{NotSaizArea}}{\text{TotalArea}}$.

misses some large polygons in the Southwest. Thus, moving from east to west in figure 7 coincides with a noticeable increase of in the within polygon differences between county polygons and the Saiz circles.

Columns (1) and (2) of table 1 show the first and second stage estimates from equations 1 and 2 using Saiz Unavailability, where column (2) presents the IV estimates. The results match the literature: Higher Saiz Unavailability predicts 2002 to 2006 house price growth and higher house price growth yields increased auto sales growth. The regressions are weighted by the number of households in 2000 and heteroskedasticity robust standard errors are clustered at the state level. We also retain the residuals from the IV model in column (2) and measure the Saiz Unavailability error using the polygons in figure 7 using precise boundary files and the procedure documented above. We then regress the residuals from the IV model in column (2) on the Saiz Land Unavailability error. This regression is weighted by the number of households and bootstrapped standard errors are in parentheses. Recall that for the IV exclusion restriction to hold, that the correlation between the errors from the IV regression and the Saiz Land Unavailability must be equal to zero. Instead, column (3) shows that the second stage residuals are correlated with the Saiz Unavailability measurement error, providing evidence that the exclusion restriction is violated. In columns (4) - (6) of table 1, we repeat this analysis using the our Land Unavailability measure (described below) that employs precise polygon boundaries. Column (4) shows that Land Unavailability is strongly predictive of county house price growth. The coefficient on our Land Unavailability measure (column (4)) is larger than that for the Saiz proxy (column (1)) and the R^2 increases 15 percent as we move from column (1) to column (4). In column (5) we find an elasticity between house price growth and consumption growth of 0.32. This coefficient is nearly 40 percent larger (and has a higher t-statistic) than the 2SLS estimate from column (2) that uses Saiz Unavailability as an instrument. Last and most importantly, the coefficient in the regression of the residuals from column (5) on Saiz Unavailability Error is not statistically different from zero, supporting that the exclusion restriction holds and that the Saiz Unavailability measurement error is not correlated with the measurement error when using our Land

Unavailability measure as an instrument. Indeed, while the standard errors are similar in columns (3) and (6), the coefficient estimate in column (6) falls by nearly half and the R^2 in drops by 75 percent.

Table 2 re-estimates the models in table 1 with data from 2006 - 2009, the same period used by Mian et al. (2013). Column (1) documents that Saiz Unavailability predicts house price growth, while the IV model in column (2) show that lower house price growth causes a decrease in consumption, using Saiz Unavailability as an instrument. Column (3) tests the exclusion restriction. Here we find that Saiz Unavailability error is uncorrelated with the second stage error. The results suggest that the exclusion restriction holds. Columns (4) - (6) instead use our Land Unavailability measure. Land Unavailability is a relevant instrument (column (1)), while column (2) shows that the coefficient on 2006 - 2009 house price growth increases 50 percent when using Land Unavailability, rather than Saiz Unavailability, as and instrument. Finally, column (6) regresses the IV residuals from column (5) on the Saiz Unavailability Error and results suggest that the exclusion restriction is not violated when Land Unavailability as an instrument.

3.2 Non-tradable Employment, Housing Net Worth Shocks, and Saiz Land Unavailability

Mian and Sufi (2014) similarly use the Saiz instrument in models that estimate the impact of the housing net worth shock on non-tradable employment in their analysis of the Great Recession.[9] We replicate their analysis using both Saiz Unavailability and Land Unavailability as instruments.[10] The results are in table 3, where the layout of table 3 mirrors tables 1 and 2. Overall, using both Saiz Unavailability and Land Unavailability, the results are congruent with Mian and Sufi (2014): More negative housing net worth shocks cause lower non-tradable employment growth from 2007 - 2009. Yet comparing the tests for the exclusion restriction in columns (3) and (6) suggests that the exclusion restriction is violated when using Saiz Unavailability, but not when using Land Unavailability.

[9]Mian and Sufi use the Saiz elasticity measure that includes both Saiz Unavailability and a proxy for land regulation. We just consider the Saiz Unavailability portion of the instrument here.

[10]For non-tradable employment, we use employment in the restaurant and retail industries from Mian and Sufi (2014).

4 Construction of Land Unavailability

A key aim of this paper is to calculate the percentage of undevelopable land in a geographic area, where the levels geographic aggregation span MSAs, counties, commuting zones, zip codes, etc. We follow Saiz (2010) and use digital elevation model and land cover data to compute land unavailability based on either steepness of slope or presence of water. Yet our approach differs from Saiz as we buffer each geometric polygon by 5 percent of land area, rather than compute a circle around the polygon's centroid. Using a buffer allows the topological area used in the construction of land unavailability to more closely match the area of the polygon and also allows for a consistent approach across different units of geographical aggregation (e.g. MSAs versus zip codes). The 5 percent buffer is calculated as 5 percent of the square root of polygon land area in meters.

For an instructive example, consider the map of the Los Angeles-Long Beach MSA in figure 8. As above, the blue outline corresponds to the polygon boundary for the MSA and the red circle, the area used to calculate land unavailability in Saiz (2010), has a radius of 50 km and is centered at the first central city's centroid. The yellow outline is a 5 percent buffer around the Los Angeles MSA and represents the geographic boundary used to calculate land unavailability in this paper. A number of observations are readily apparent in a comparison of the geographic areas covered by the circle with a 50 km radius centered at the centroid (red) and buffered polygon (yellow): (1) The buffered polygon provides complete coverage even though the polygon is awkwardly shaped; (2) the buffered multi-polygon allows for disjointed multi-polygons and buffers each individual polygon, allowing for islands that the US Census agglomerates in geographic units; and (3) the buffered polygon extends to the ocean and thus accommodates land unavailability when a polygon touches an ocean or other large body of water not covered by the shapefile. This approach also easily extends to various levels of geographic aggregation and hence are able to compute Land Unavailability at levels of aggregation used by economists and researchers.

Despite the differences in computation method, our proxy for Land Unavailability is highly correlated with that from Saiz (2010). Figure 9 shows a scatter plot of our land unavailability measure compared with Saiz. The slope of 0.80 and and R^2 of 0.70 highlights the similar nature of our two measures.

5 Aggregated vs Disaggregated Land Unavailability

The foregoing research that exploits Land Unavailability necessarily uses MSA Land Unavailability as that is the only level of aggregation available from the Saiz dataset. Yet housing markets vary substantially within large geographic areas such as MSAs and, as noted above, MSAs with larger land area also experienced larger house price growth during the 2000s. If Land Unavailability also varies more with larger MSAs and assuming that Land Unavailability is measured with error, then first stage predictions will be more uncertain and second stage estimates will be less precise for the large MSAs that also experienced high price growth during the 2000s.[11] We explore this possibility in table 4 and figure 10. First in table 4, we regress MSA size on within MSA measures of zip code Land Unavailability spread including the (1) variance, (2) range, and (3) interquartile range. In all three cases, the spread in zip code Land Unavailability within the MSA is highly correlated with size, meaning that larger MSAs have more variation in Land Unavailability within their polygon boundaries. Figure 10 further highlights this point as here for each MSA size decile, we plot the smoothed density of the range of zip code Land Unavailability within each MSA. Clearly as MSAs grow in size (as we move from the first decile to the tenth decile), the density of the Land Unavailability range shifts rightward, matching our above regression results and suggesting that the use of highly aggregated Land Unavailability may lead to less precise 2SLS regression estimates.

6 The Validity of Land Unavailability as an Instrument

The use of Land Unavailability as an instrument relies on its exogeneity relative to other proxies for housing demand. Specifically, if higher Land Unavailability is exogenous and predicts higher house price growth, then Land Unavailability should not be positively correlated with factors of housing demand. In the literature, there has been debate on this issue. Mian and Sufi (2011, 2014) claim that Land Unavailability is exogenous while

[11] This uncertainty may be magnified if there is population heterogeneity related to the predictive power Land Unavailability as an instrument.

Davidoff et al. (2016) contends that Land Unavailability is positively correlated with housing demand. Our study differs from previous attempts to assess the exogeneity of Land Unavailability as instrument as we use a more highly disaggregated dataset with nearly complete coverage of the contiguous United States. Previous studies that aim to assess the exogeneity of Land Unavailability employ data at the MSA level. Yet housing markets, demand factors, and Land Unavailability can vary tremendously within MSAs, making MSAs an inappropriate level of aggregation with which to judge Land Unavailability exogeneity. MSAs also only cover a fraction of US land area and thus bias any correlations between Land Unavailability and housing demand factors towards areas with higher levels of historical development (e.g. the Northeastern US). Indeed, for a city to be classified as an MSA, it must have at least 50,000 people. As Land Unavailability increases the cost of home and building construction, then MSAs are unlikely to be located in areas with high Land Unavailability. To see this, consider figure 11 that plots Google satellite imagery for the US and coverage for Saiz Unavailability. Again, red circles represent a 50 km radius around each MSA first central city centroid in the Saiz dataset. The figure clearly shows a strong negative correlation between the instances of MSAs and Land Unavailability due to rugged terrain. This pattern is clearly visible in the Rocky Mountain region, where for example in Colorado, five MSAs sit at the base of the Rockies. Yet even in the populated pacific states, California, Washington, and Oregon, there is a negative relationship MSA instantiation and terrain slope. Indeed, the northern ascent of California MSAs is limited by the Mendicino and Shasta National Forests, while Seattle lies between Olympic National Park and Wenatchee National Forest. Thus, judging the exogeneity of Land Unavailability using only MSAs will lead to biased results.

We hence examine the correlations of between Land Unavailability and proxies of demand at the zip code level with near complete national coverage. The proxies of demand that we consider at the zip code level include the county Bartik shock mapped to the zip code level from 2000 - 2006, covering the same time period as the large increase in US housing demand during the 2000s, the zip code amenities index, and the 2000 Census zip code share of foreign born, share of college educated, and housing density. We examine the correlation between these variables and Land Unavailability in table 5. Column (1) is the correlation coefficient between Land Unavailability and the variable in each row. Columns (2) and (3) are the output of separate regressions of Land Unavailability on the variable in each row and show the heteroskedasticity robust p-value (using standard errors clustered at the three-digit zip code level) and the R^2 statistic. The regressions are weighted by the number of households in 2000. The results show that Land Unavailability is not positively correlated with proxies of demand and the R^2 statistics are all small in magnitude. Indeed, there is nearly no correlation between Land Unavailability and the Bartik Shock, the share of college educated, and the housing density. The correlations between the share of foreign born or amenities and Land Unavailability are negative and significant, opposite of the sign needed for Land Unavailability to be positively correlated with demand. Indeed, many immigrant enclaves for example are near the southern US boarder, areas with low Land Unavailability, or in traditional port cities. This negative correlation is therefore not a concern for the use of Land Unavailability as an instrument in a typical panel framework as region fixed effects get differenced away. We also examine the change of the foreign born share from 2000 to 2011 in the last row of table 5. The correlation coefficient is just -0.037 and the regression in columns (2) and (3) yields a p-value on Land Unavailability of 0.43 and an R^2 of approximately zero. Altogether, these results suggest that Land Unavailability is not positively correlated with proxies of demand.

Table 6 shows zip code level regressions of 2002M01 - 2006M12 Zillow house price growth on Land Unavailability and proxies of demand. We use 2002 - 2006 as this time period is associated with a large increase in US prices and housing demand. The regressions in columns (1) - (4) of table 6 document how the relationship between house price growth and Land Unavailability changes after the inclusion of housing demand proxies. The regressions are weighted by the number of households in 2000 and heteroskedasticity robust standard errors are clustered at the three-digit zip code level. Column (1) regresses Zillow house price growth on Land Unavailability without any controls and finds an elasticity between Land Unavailability and log house prices of 0.224. Column (2) also includes the Bartik labor demand shock, the college share, and the foreign born share. The coefficient on Land Unavailability jumps to 0.307 as Land Unavailability is negatively correlated with the foreign born share. The standard error also falls to 0.032 as all of the additional variables in column (2) have predictive power with respect house prices. As expected, the Bartik shock and foreign share positively predict house prices. In column (3) we add the zip code amenities index, where the coefficient on Land Unavailability increases to 0.32 due to the negative correlation between Land Unavailability and amenities. Last, column (3) adds the housing density. The coefficient on Land Unavailability is unchanged. Altogether, these regressions show that the inclusion of housing demand factors does not erode the predictive power of Land Unavailability with regard to house prices.

One potential concern with the Land Unavailability proxy is that the one component of the total undevelopable land, Land Unavailability due to water, may correlated with proxies of demand. For example, construction of housing in coastline California zip codes is obviously restricted by the ocean, but properties near the ocean are the most desirable. As long as the relative demand between ocean and non-ocean properties does not change over time, in a differenced regression property distance to the ocean will not violate the IV exclusion restriction. However, this is difficult to check empirically. Alternatively, we can ensure that the predictive power of Land Unavailability with regard to house prices is not solely due to proximity to water. We do this in columns (5) - (8) of table 6 and regress house price growth on the components of Land Unavailability, unavailability due to a steep slope, water, and wetlands, using the full specification from column (4). First, note that the correlation between Water Unavailability and Slope Unavailability is -0.13, meaning that the root cause of Land Unavailability varies substantially across zip codes and that areas with building impediments due to rugged terrain are typically not near water. Columns (5) - (7) individually estimate the relationship between house price growth and the components of Land Unavailability, while column (8) shows a full specification that includes all components that comprise Land Unavailability. The results show that Slope, Water, and Wetlands Land Unavailability all positively predict 2002 - 2006 house price growth individually and in the full specification. Thus, the predictive power of Land Unavailability is driven by all of its components and does not hinge on a single factor. Last, note that the predictive power of Slope and Water Unavailability is similar in column (8). The t-statistics on the coefficients for Slope and Water Unavailability are 7.39 and 7.36 respectively.

7 The Predictive Power of Land Unavailability

We assess the predictive power Land Unavailability at the zip code level, the lowest that house price data is available in the US. Undoubtedly as zip codes are often quite small, the supply response of housing to changes in demand not only depends on the Land Unavailability of the zip code itself, but also on the Land Unavailability of areas surrounding the zip codes. Thus, we also consider Land Unavailability at the 4 and 3 digit levels (e.g. the first four and three digits of the zip code), as well as for counties and commuting zones. The four digit and three digit zip codes offer drastically different levels of aggregation. For example, in our dataset there are nearly 6000 4 digit subgroups, but only 877 three digit zip code areas. In comparison, there are 1473 counties and 445 commuting zones with Zillow zip code house price data. We report the predictive effects of Land Unavailability for 2002 to 2006 house price growth at different levels of aggregation in table 7. Column (1) is identical to column (1) of table 6 and shows that a one percentage point increase in Land Unavailability is associated with a 0.224 percent increase in zip code level Zillow house prices. Column (2) adds the 4 digit zip code Land Unavailability. Here, the coefficient on zip code Land Unavailability becomes insignificant while the coefficient on the 4 digit zip code Land Unavailability is 0.383. Notice also that the R^2 in column (2) doubles to 0.065. The coefficient on the 3 digit zip code Land Unavailability in column (3) is even larger at 0.570 and the R^2 doubles again to 0.136. Columns (4) and (5) show that county and commuting zone Land Unavailability are even stronger predictors of zip code house price growth. The coefficients on county and commuting zone Land Unavailability are 0.754 and 0.877 respectively, while the R^2 statistics in columns (4) and (5) are 0.288 and 0.351. Clearly, the areas surrounding a zip code are important for Land Unavailability. Yet after controlling for a Land Unavailability in wider geographic area (e.g. three digit zip codes, counties, or commuting zones in columns (3) - (5)), the coefficient on zip code Land Unavailability becomes negative and significant. This highlights the dynamics of Land Unavailability within local housing markets: After controlling for wider geographic Land Unavailability higher Land Unavailability predicts *lower* house price growth. While this result may seem antithetical compared to extant research that finds a positive predictive relationship between Land Unavailability and house prices, it is instead consistent with substitutability within local housing markets.

For example, consider the case of California. The local housing markets in California are among the most diverse in the US: Eight of the top ten US regions in terms of 2002M01 within region house price variance (in levels), using either three or four digit zip codes, are in California. California commuting zones are also the largest in the nation by area. Thus there is out-sized potential for housing substitutability in California, relative to the rest of the country.

Using only California zip codes, columns (1) - (3) of table 8 present regressions of the log first difference in Zillow house prices (HP growth) on Land Unavailability at the zip code level. Controls include the zip code amenities index and share of foreign born in 2000. Fixed effects in columns (1) - (3) account for unobserved differences in house price growth at the county, three digit zip code, and four digit zip code levels, respectively. The results document that house price *growth* was largest in areas with lower Land Unavailability after accounting for region fixed effects (the regression estimates in table 8 are similar even without fixed effects). As noted above, this is result is inconsistent with previous findings and even holds across adjacent California MSAs (Davidoff et al., 2016). Yet agents facing financial constraints (e.g. down payment constraints) make housing consumption decisions based on prices in levels and examining the growth in prices masks the comparability of their choice set. For example zip codes 92507 and 90403 are in Santa Monica and Riverside, within the same commuting zone. The 2002M01 median house price in these areas was $152,500 (Riverside) and $480,800 (Santa Monica). A five percent increase in Santa Monica house prices is $24,040. Comparatively, Riverside house prices would have to increase 15.8 percent to yield the same dollar increase. While this example is extreme it highlights a key point: Dollar increases are more important for households facing financing constraints, and in housing markets characterized by high within region variance substitutability is likely to be magnified. Thus, a more appropriate dependent variable to assess the impact of Land Unavailability in California is a the difference in house price levels. We undertake this analysis in columns (4) - (6) in table 8 where the difference in levels house prices is the dependent variable. The results show that more Land Unavailability is associated with higher house prices, in line with conventional findings. In particular, column (6) controls for four digit zip code fixed effects. As there are 315 four digit zip code regions in California, this regression shows that even after controlling for unobserved differences within precise regions in California that higher Land Unavailability predicts higher prices. Finally, note that many desirable California zip codes are near the Pacific Ocean, perhaps correlating Land Unavailability with demand. So, in table 9 we re-estimate these regressions using only Slope Unavailability. The correlation between Slope Unavailability and Water Unavailability in California is -0.09, implying that California zip codes more constrained by water are relatively flat. The results in table 9 and are consistent with the foregoing findings.

Tables B1 and B2 in appendix B re-estimate the models in tables 8 and 9 using data from the contiguous United States. The results are similar to those for California.

Finally, Tables B3 and B4 in appendix B repeat the predictive house price growth and Land Unavailability regressions from table 7, but instead use house price growth from 2007M01 - 2010M12, covering the downturn in US housing and from 2012M01 to 2016M12, a recent period where US house prices were increasing but the rate of ascent was much lower than during the 2000s boom. We confirm that Land Unavailability predicts cross-section of house prices during the 2000s US housing bust period. Over the recent period from 2012 - 2016, table B4 shows that three digit zip code Land Unavailability retains its predictive power in recent samples.

8 Conclusion

In this paper, we construct a new proxy for Land Unavailability that builds on the work of Saiz (2010). Specifically, our measure uses updated satellite imagery now available from the USGS, more accurate geographic polygons, and is constructed for multiple levels of aggregation. Thus, we construct accurate proxies of Land Unavailability at several common levels of geographic aggregation down to US zip codes.

We show that precise boundary files and more disaggregated Land Unavailability data is preferred. Less precise boundary files yield geographically correlated measurement error that can bias 2SLS estimates. Similarly, there is more variation in Land Unavailability, for example, in larger MSAs that experienced high house price volatility during the 2000s. Thus 2SLS estimates using MSA aggregated data would be less precise for the very MSAs that that experienced high house price growth during the 2000s.

Next, using our zip code land unavailability proxy, we find that Land Unavailability is not correlated with housing demand factors, meaning that Land Unavailability is exogenous and hence validating Land Unavailability as an instrument.

Finally, we examine predictive power Land Unavailability at the zip code level. We that Land Unavailability, especially in surrounding ares, is a strong predictor of local house prices.

References

M. Adelino, A. Schoar, and F. Severino. House prices, collateral, and self-employment. *Journal of Financial Economics*, 117(2):288–306, 2015.

S. Agarwal, G. Amromin, I. Ben-David, S. Chomsisengphet, T. Piskorski, and A. Seru. Policy intervention in debt renegotiation: Evidence from the home affordable modification program. *Journal of Political Economy, Forthcoming*, 2017.

A. Aladangady. Housing wealth and consumption: Evidence from geographically-linked microdata. *American Economic Review*, 107(11):3415–46, November 2017.

T. J. Bartik. *Who Benefits from State and Local Economic Development Policies?* Books from Upjohn Press. W.E. Upjohn Institute for Employment Research, November 1991.

P. Beaudry, D. A. Green, and B. Sand. Does industrial composition matter for wages? a test of search and bargaining theory. *Econometrica*, 80(3):1063–1104, 2012.

P. Beaudry, D. A. Green, and B. M. Sand. Spatial equilibrium with unemployment and wage bargaining: Theory and estimation. *Journal of Urban Economics*, 79:2–19, 2014.

R. Bostic, S. Gabriel, and G. Painter. Housing wealth, financial wealth, and consumption: New evidence from micro data. *Regional Science and Urban Economics*, 39(1):79–89, 2009.

T. Chaney, D. Sraer, and D. Thesmar. The collateral channel: How real estate shocks affect corporate investment. *The American Economic Review*, 102(6):2381–2409, 2012.

K. K. Charles, E. Hurst, and M. J. Notowidigdo. Housing booms and busts, labor market opportunities, and college attendance. 2015.

T. Davidoff et al. Supply constraints are not valid instrumental variables for home prices because they are correlated with many demand factors. *Critical Finance Review*, 5(2): 177–206, 2016.

W. A. Fuller. *Measurement error models*, volume 305. John Wiley & Sons, 2009.

S. Gabriel and C. Lutz. The impact of unconventional monetary policy on real estate markets. *Working Paper*, 2014.

S. A. Gabriel, M. M. Iacoviello, and C. Lutz. A crisis of missed opportunities? foreclosure costs and mortgage modification during the great recession. 2017.

E. E. Leamer. Housing is the business cycle. 2007.

C. Lutz, A. A. Rzeznik, and B. Sand. Local economic conditions and local equity preferences: Evidence from mutual funds during the us housing boom and bust. 2016.

A. Mian and A. Sufi. The consequences of mortgage credit expansion: Evidence from the us mortgage default crisis. *The Quarterly Journal of Economics*, 124(4):1449–1496, 2009.

A. Mian and A. Sufi. House prices, home equity–based borrowing, and the US household leverage crisis. *The American Economic Review*, 101(5):2132–2156, 2011.

A. Mian and A. Sufi. What explains the 2007–2009 drop in employment? *Econometrica*, 82(6):2197–2223, 2014.

A. Mian and A. Sufi. House price gains and us household spending from 2002 to 2006. *Working Paper*, 2015.

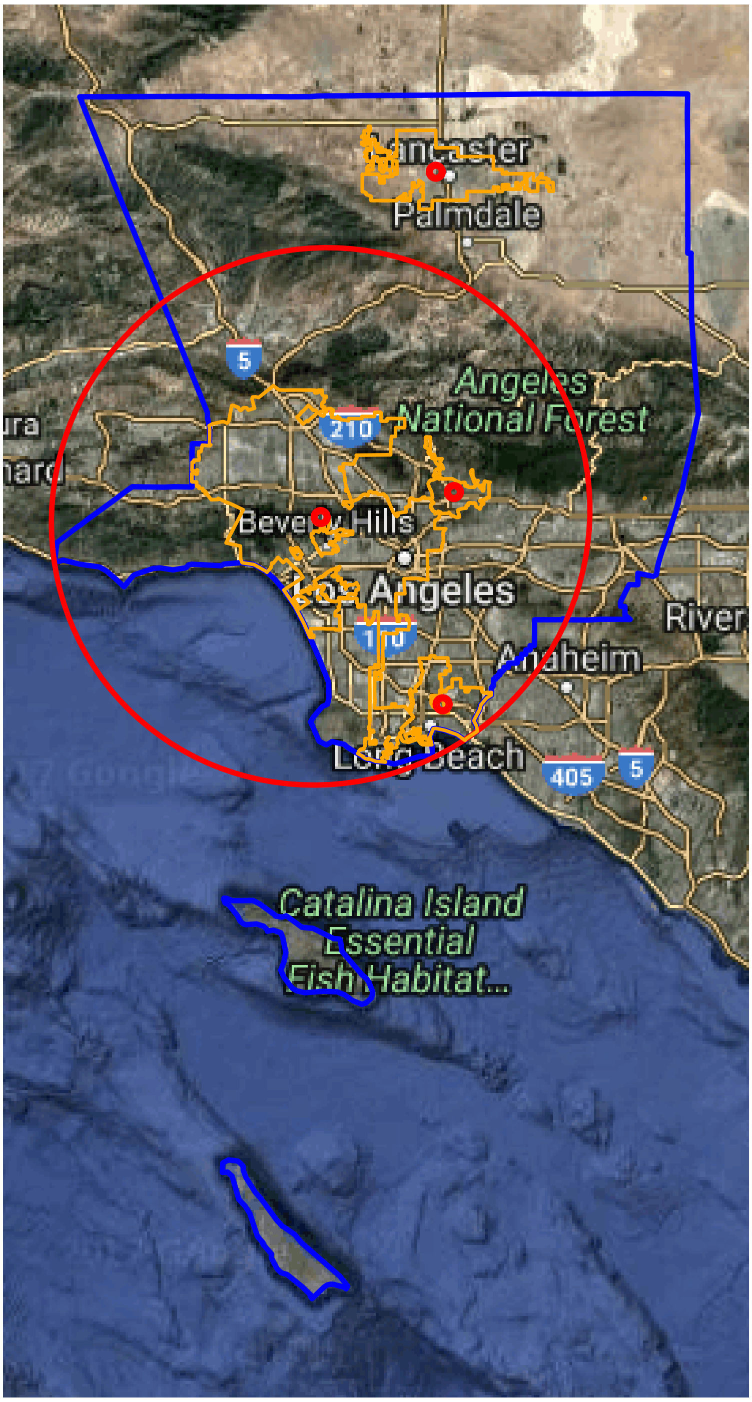
A. Mian, K. Rao, and A. Sufi. Household balance sheets, consumption, and the economic slump. *The Quarterly Journal of Economics*, 128(4):1687–1726, 2013.

A. Saiz. The geographic determinants of housing supply. *Quarterly Journal of Economics*, 125(3), 2010.

J. S. Tracy and H. S. Schneider. Stocks in the household portfolio: A look back at the 1990s. *Current Issues in Economics and Finance*, 7(4), 2001.

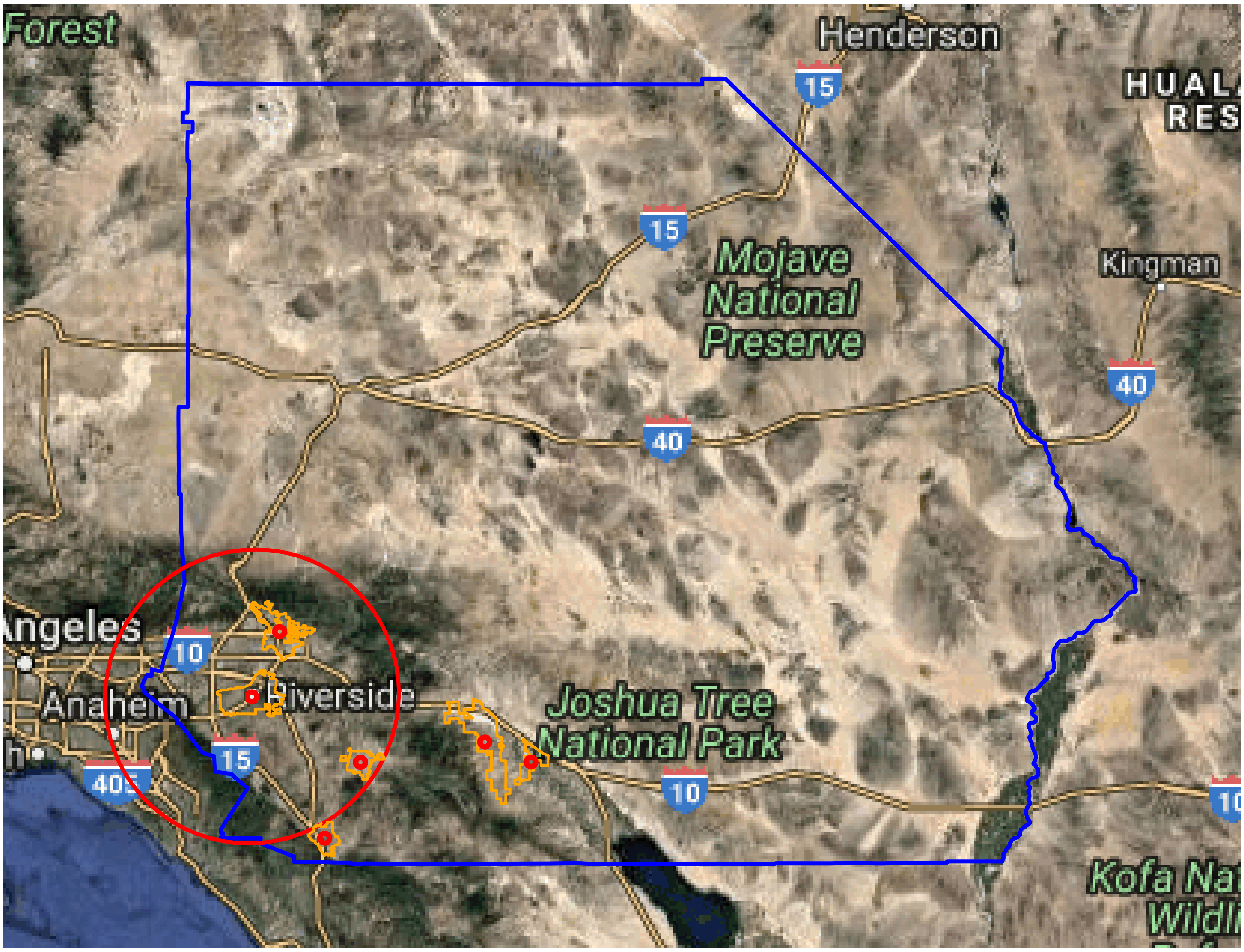
A Figures & Tables

Figure 1: Saiz Land Unavailability Coverage for the Los Angeles MSA



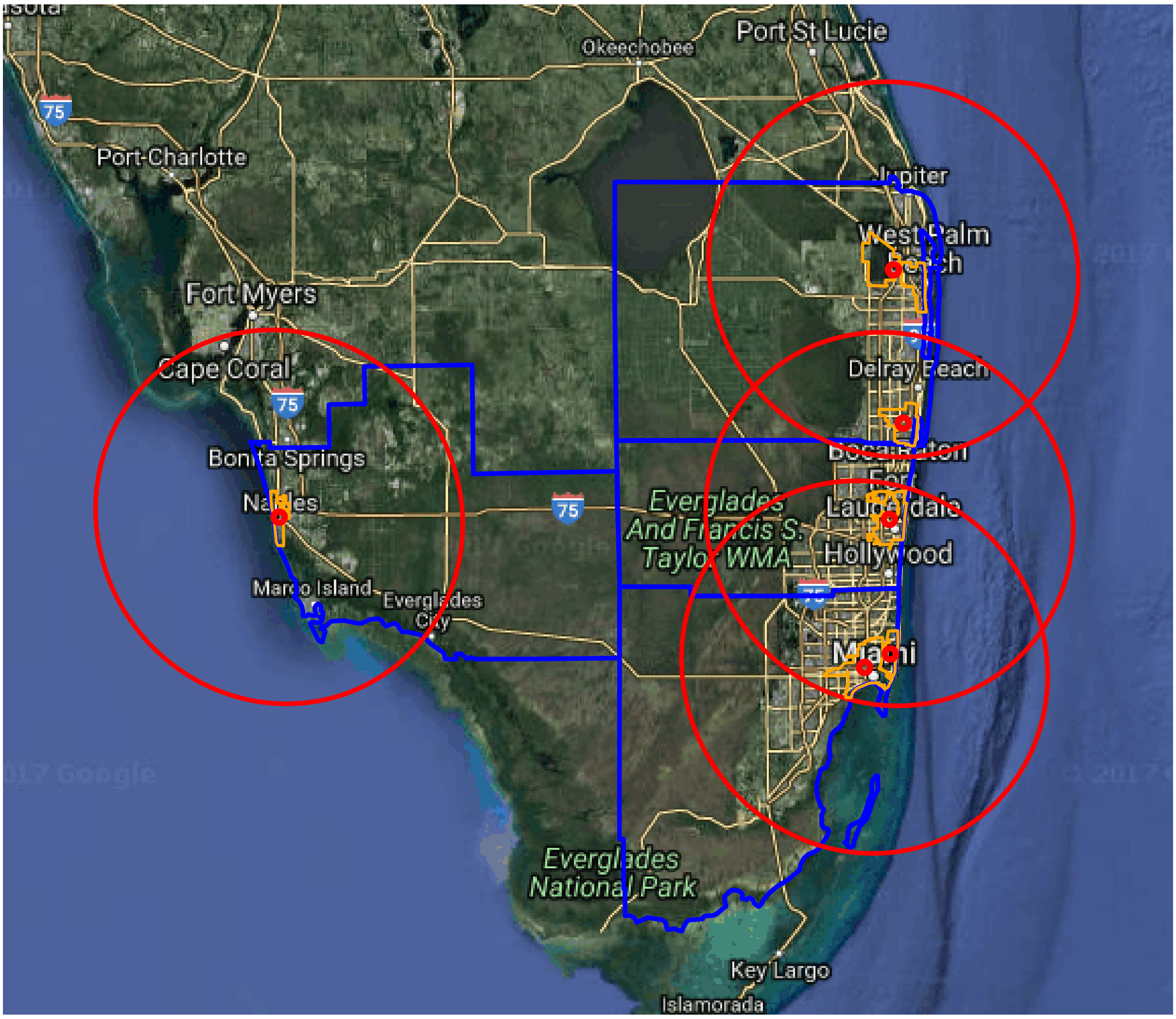
Notes: The blue lines represent the MSA polygons for the Los Angeles-Long Beach MSA. The orange lines signify the central cities within the Los Angeles MSA and the red dots are the centroids for the central cities. The red circle is has a radius of 50 kilometers and is centered around polygon centroid for the first Los Angeles central city (Los Angeles).

Figure 2: Saiz Land Unavailability Coverage for Riverside-San Bernardino



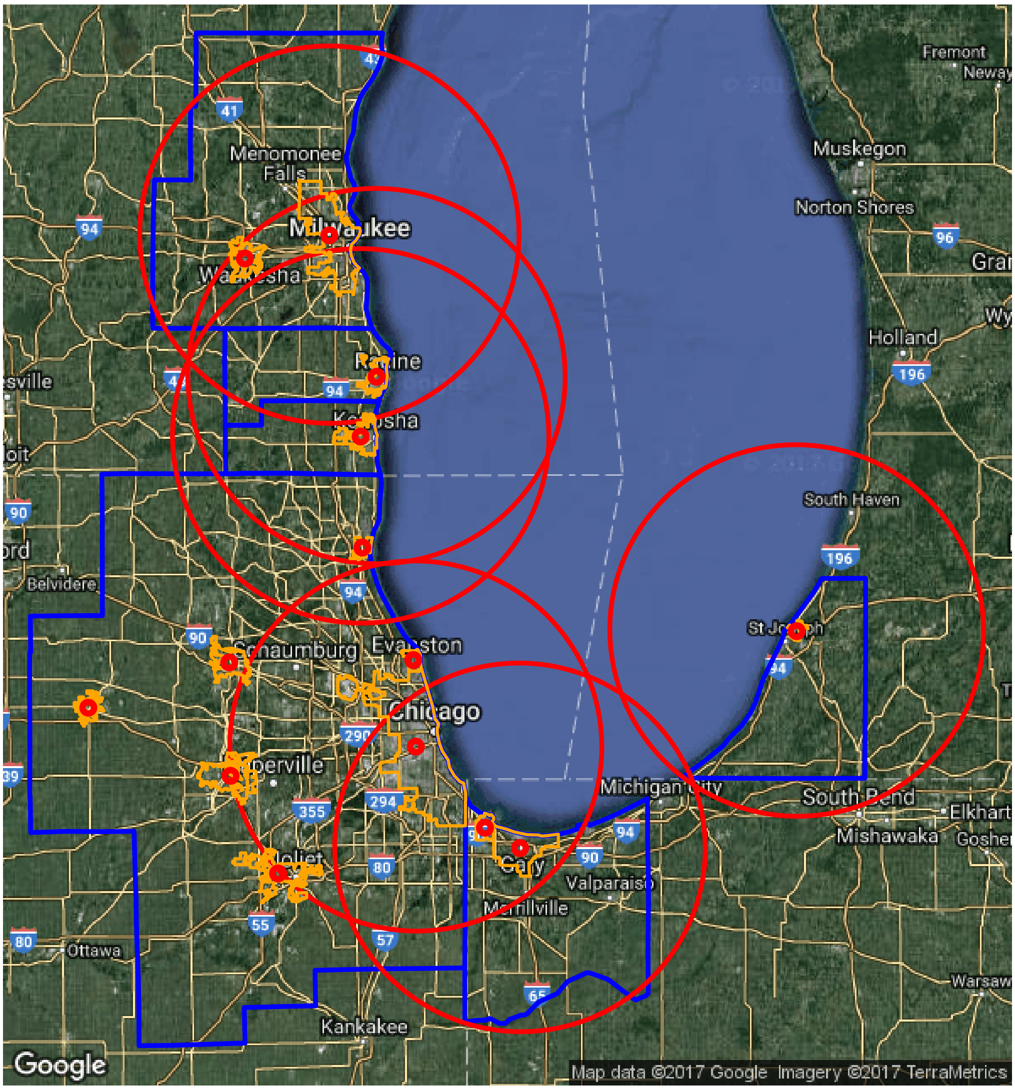
Notes: See the notes for figure 1.

Figure 3: Saiz Land Unavailability Coverage for Coastal Florida MSAs



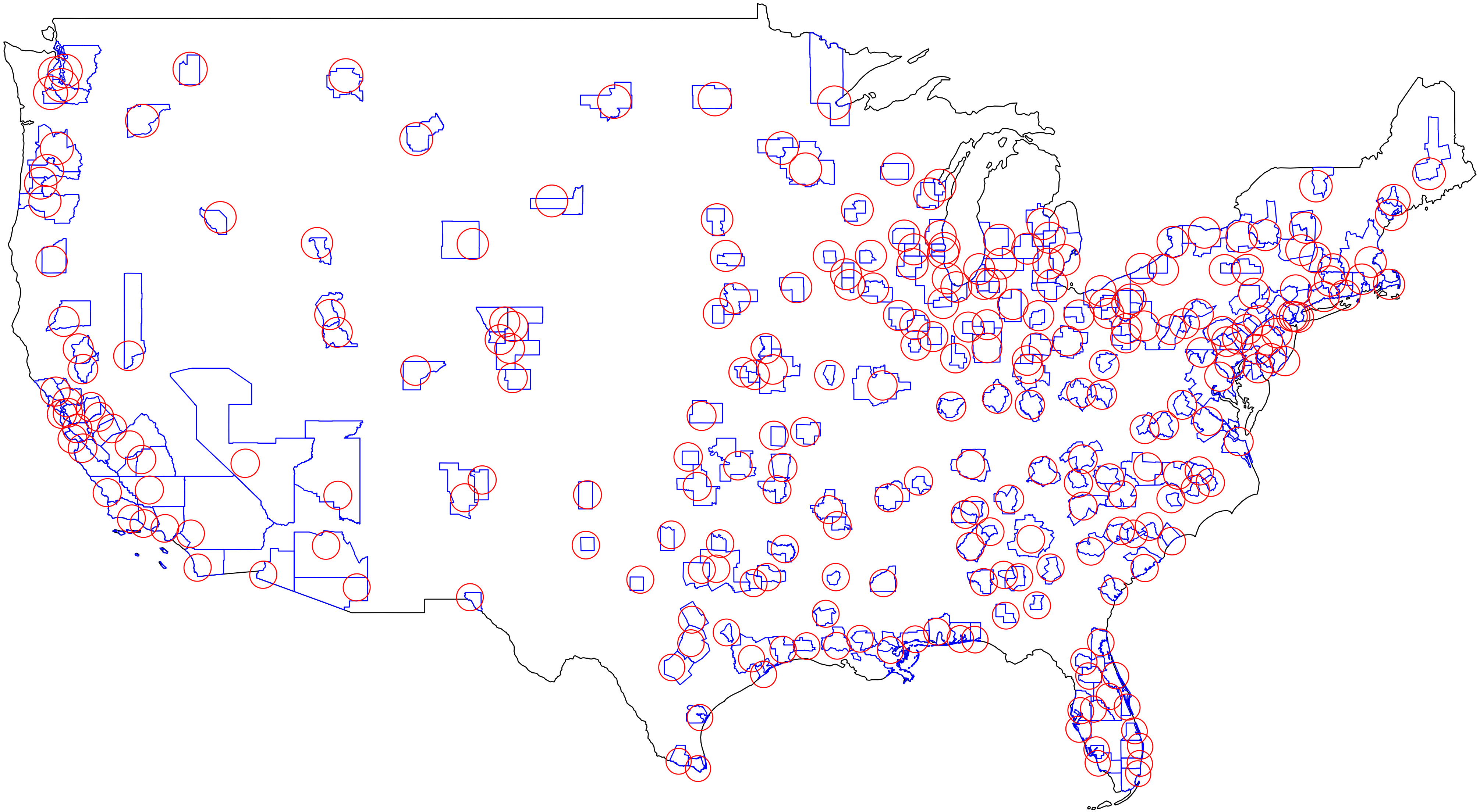
Notes: See the notes for figure 1.

Figure 4: Saiz Land Unavailability Coverage for Lake Michigan MSAs



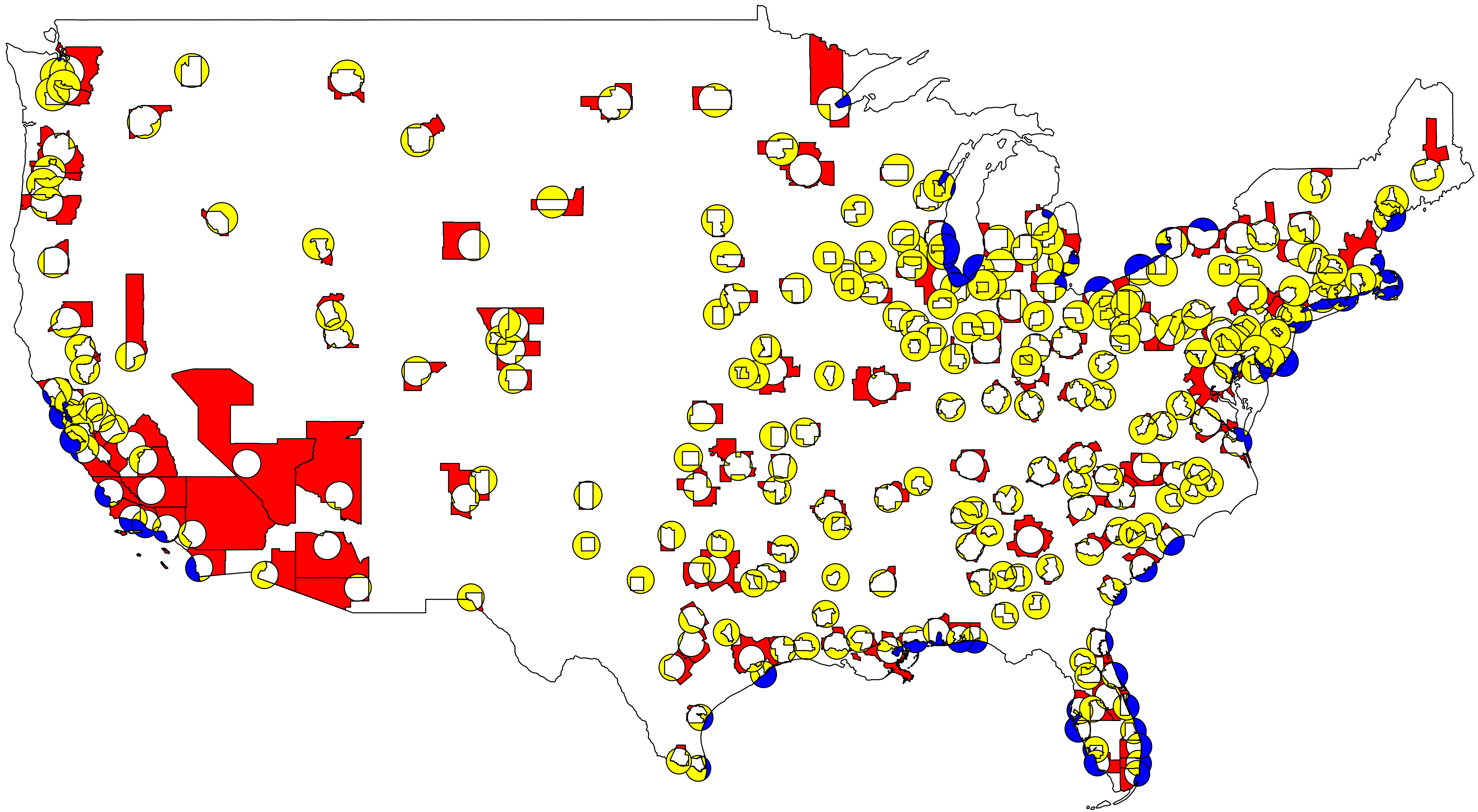
Notes: See the notes for figure 1.

Figure 5: Saiz Land Unavailability Coverage Across US MSAs



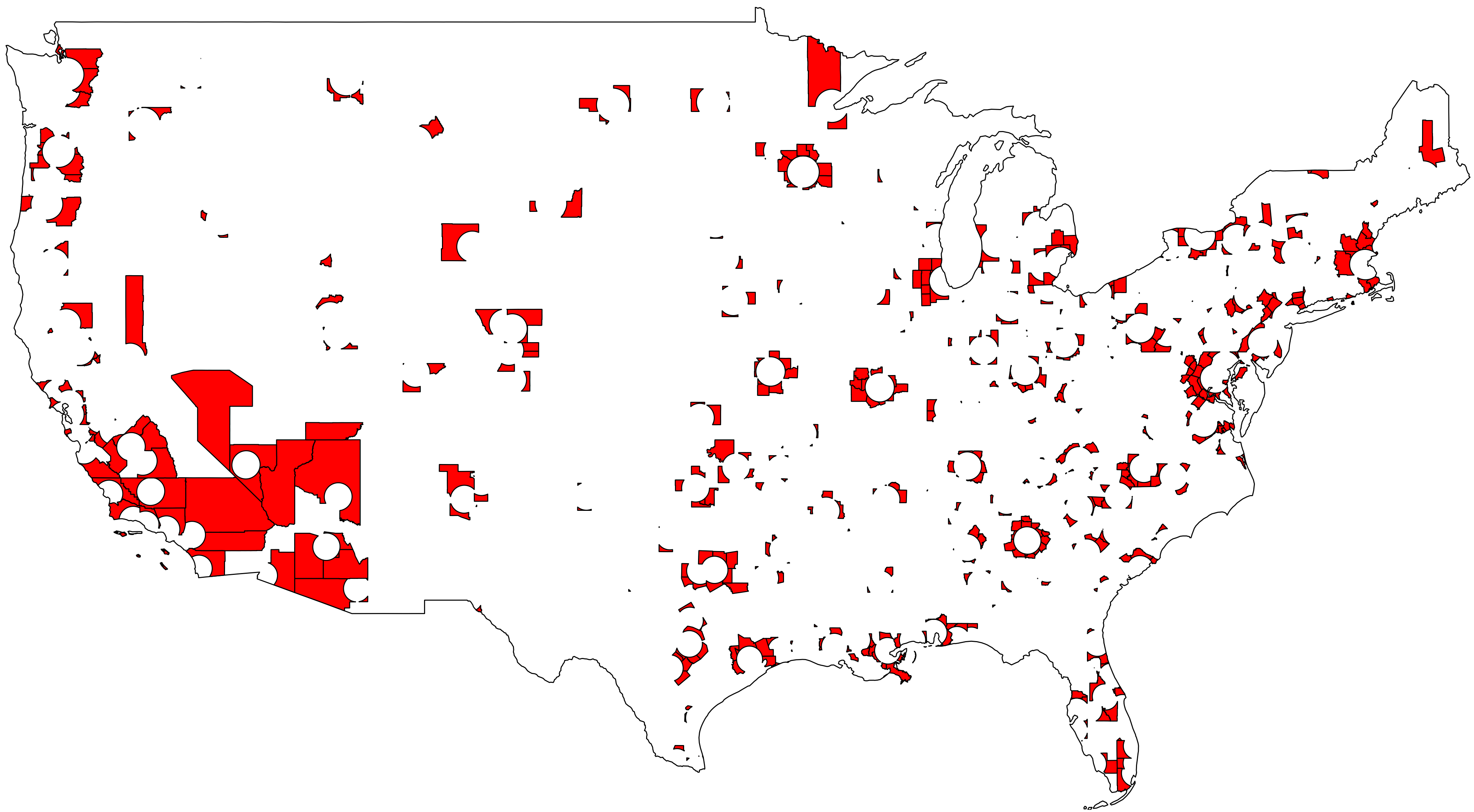
Notes: The blue lines represent MSA polygons. The red circles have a radius of 50 kilometers and are centered around first central city polygon centroid.

Figure 6: Difference in MSA Polygons and Saiz Land Unavailability Coverage



Notes: Red areas are the differences between MSA polygons and the Saiz circles within each polygon. Blue area are the differences between the Saiz circles and the MSA polygons that are outside their respective MSA polygons and lie in oceans, the Great Lakes, or the Gulf of Mexico. Yellow area are the inland differences between the Saiz circles and the MSA polygons that lie outside the polygon boundaries.

Figure 7: Inside Polygon Difference between County Polygons and Saiz Land Unavailability Coverage

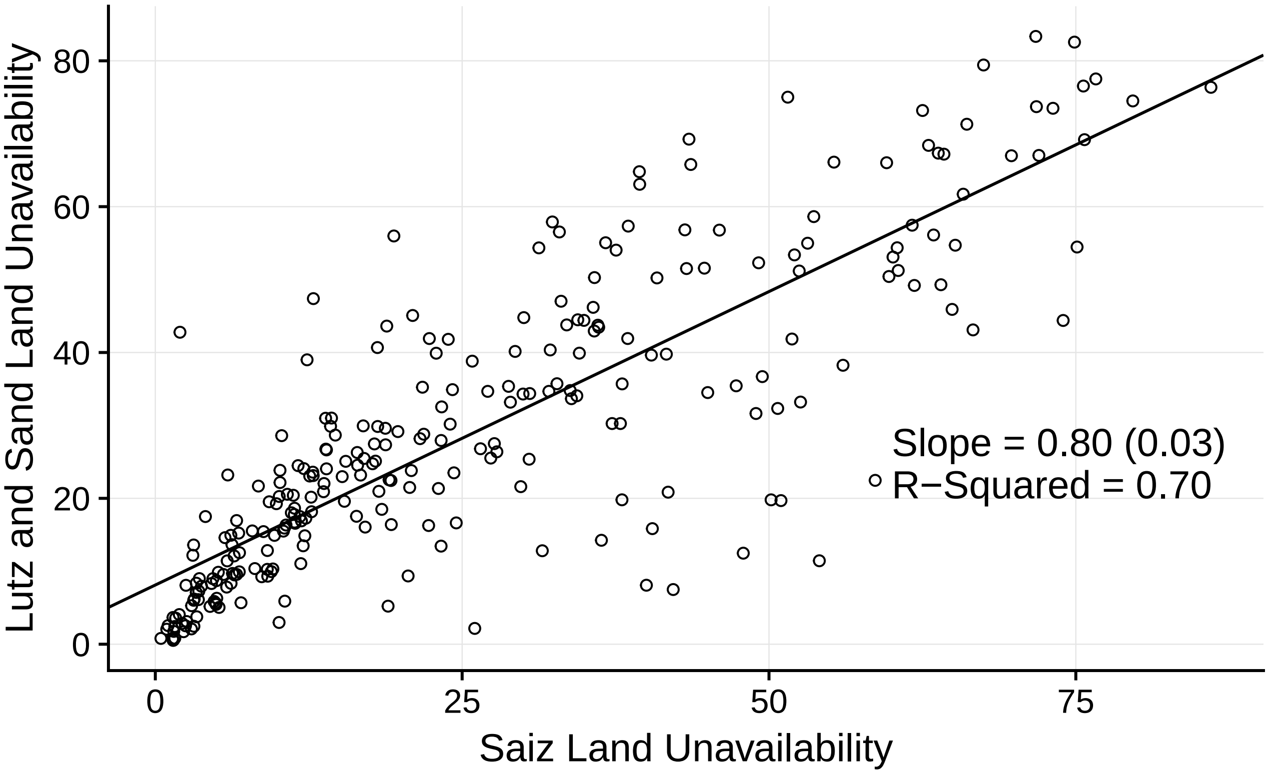


Notes: See the notes for figure 6.

Figure 8: Saiz and Buffered Land Unavailabilty Coverage for the Los Angeles MSA

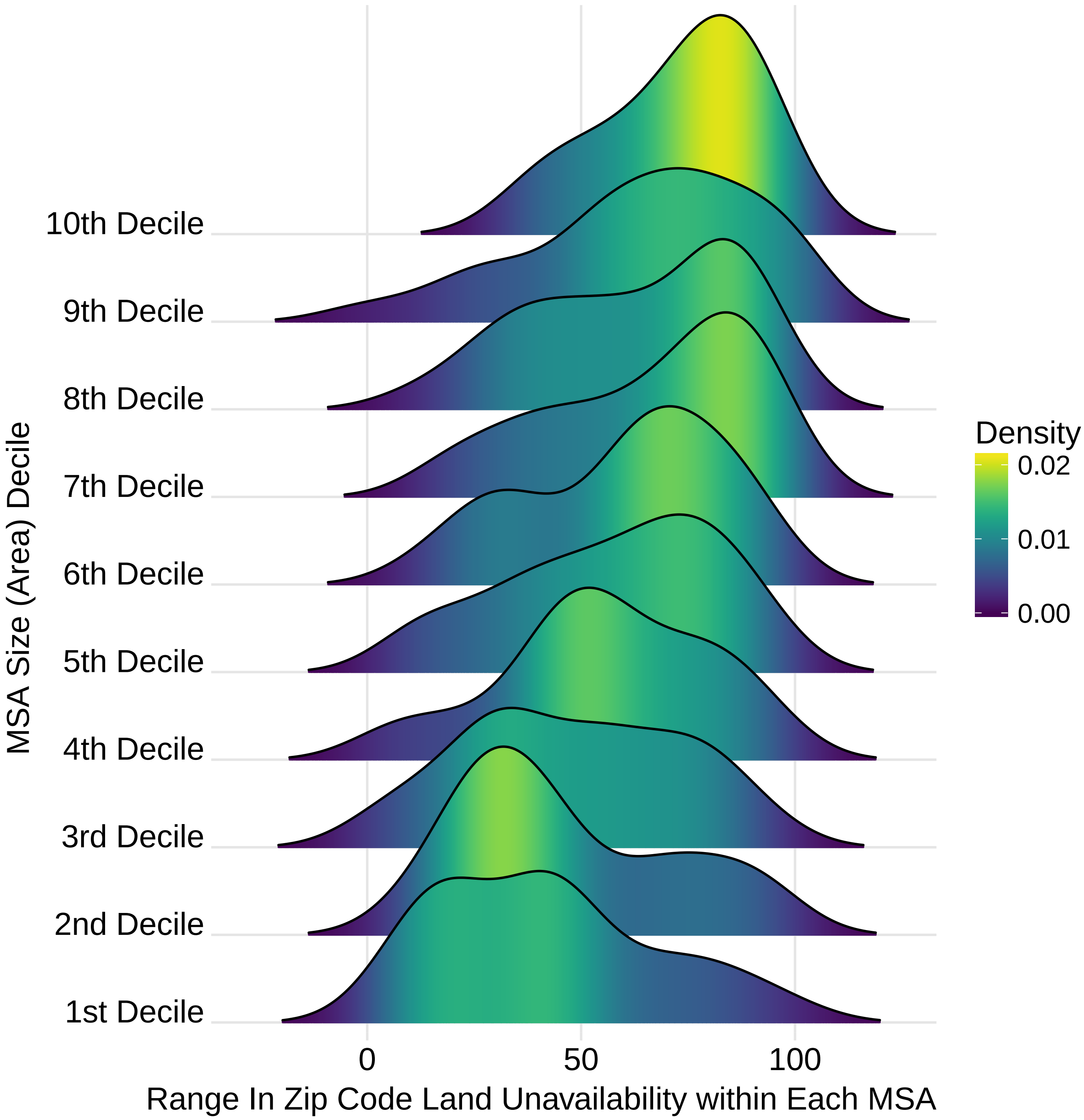
Notes: See the notes for figure 1. The yellow line is a 5 percent buffer around the Los Angeles-Long Beach MSA and represents the boundary used to calculate land unavailability in this paper.

Figure 9: Comparison of Land Unavailability Measures



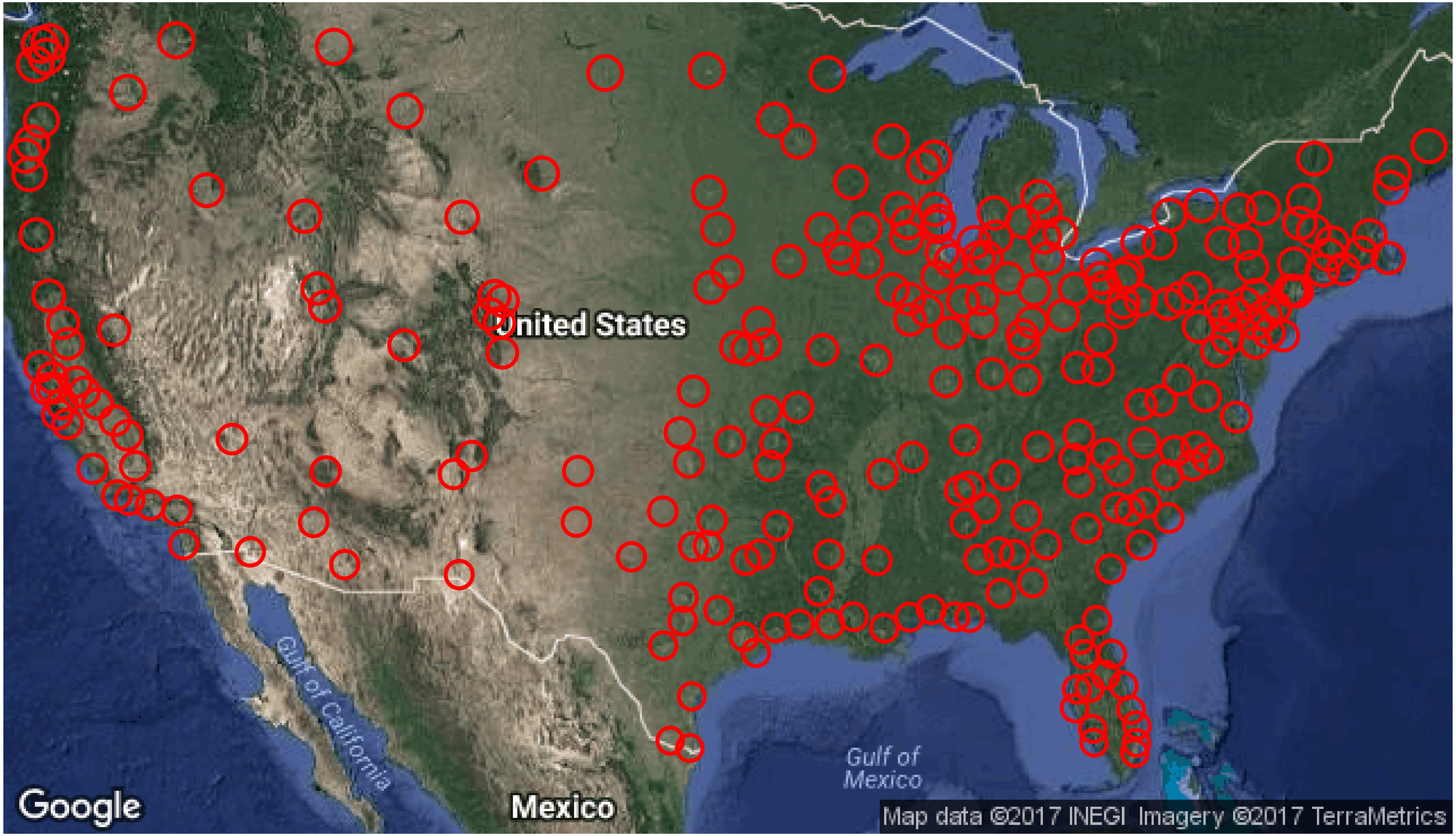
Notes: The Saiz (2010) proxy for the percentage of unavailable land is on the horizontal axis; the vertical axis shows the measure of land unavailability constructed in this paper. Points correspond to MSAs.

Figure 10: Densities of the Range in Land Unavailability by MSA Size



Notes: The horizontal axis is the range in zip code Land Unavailability within each MSA. The vertical axis are MSA deciles by size.

Figure 11: Saiz Land Unavailability Coverage for United States



Notes: Red circles have a radius of 50 kilometers and are centered around polygon centroid in the Saiz (2010) dataset.

Table 1: Consumption, Housing Prices, and Land Unavailability 2002-06

	Saiz Unavailability			Land Unavailability		
	Δ ln Zillow HP 2002-06 (OLS)	Δ ln Auto Sales 2002-06 (IV)	Saiz IV Residuals (OLS)	Δ ln Zillow HP 2002-06 (OLS)	Δ ln Auto Sales 2002-06 (IV)	LU IV Residuals (OLS)
	(1)	(2)	(3)	(4)	(5)	(6)
Saiz Unavailability	0.6017***					
	(0.1016)					
Δ ln Zillow HP 2002-06		0.2256**			0.3175***	
		(0.1149)			(0.1213)	
Saiz Unavailability Error			−0.0019*			−0.0010
			(0.0010)			(0.0010)
Land Unavailability				0.6911***		
				(0.1110)		
Constant	0.2253***	−0.0917*	−0.0155	0.2022***	−0.1291**	−0.0079
	(0.0480)	(0.0547)	(0.0139)	(0.0443)	(0.0530)	(0.0135)
Observations	542	542	542	542	542	542
R^2	0.255		0.016	0.298		0.004

Notes: This table presents county-level regressions for 2002-06 auto sales growth, 2002-06 house price growth, and Land Unavailability. Columns (1) - (3) show regressions using Saiz Unavailability, while columns (4) - (6) use Land Unavailability developed in this paper. Columns (1) and (4) are first stage regressions of 2002-066 Zillow county house price growth on Saiz Unavailability and Land Unavailability, respectively; columns (2) and (5) present the IV regressions of 2002-06 auto sales growth on 2002-06 house price growth, instrumented using either Saiz Unavailability or Land Unavailability; and columns (3) and (6) regress the residuals from the IV regressions on the Saiz Land Unavailabilty measurement error. All regressions are weighted by the number of households in 2000. In columns, (1), (2), (4), and (5) heteroskedasticity robust standard errors are clustered at the state level; standard errors are bootstrapped in columns (3) and (6).

Table 2: Consumption, Housing Prices, and Land Unavailability 2006-09

	Saiz Unavailability			Land Unavailability		
	Δ ln Zillow HP 2006-09 (OLS)	Δ ln Auto Sales 2006-09 (IV)	Saiz IV Residuals (OLS)	Δ ln Zillow HP 2006-09 (OLS)	Δ ln Auto Sales 2006-09 (IV)	LU IV Residuals (OLS)
	(1)	(2)	(3)	(4)	(5)	(6)
Saiz Unavailability	-0.4073^{***}					
	(0.1455)					
Δ ln Zillow HP 2006-09		0.7364^{***}			1.0912^{***}	
		(0.1803)			(0.3807)	
Saiz Unavailability Error			0.0011			-0.0017
			(0.0010)			(0.0013)
Land Unavailability				-0.3553^{**}		
				(0.1650)		
Constant	-0.0622	-0.3728^{***}	0.0086	-0.0813^{*}	-0.3070^{***}	-0.0135
	(0.0454)	(0.0369)	(0.0127)	(0.0416)	(0.0648)	(0.0148)
Observations	557	557	557	557	557	557
R^2	0.138		0.006	0.093		0.011

Notes: See the notes for table 1

Table 3: Non-tradable Employment, Housing Net Worth, and Land Unavailability, 2006-09

	Saiz Unavailability			Land Unavailability		
	Δ Housing Net Worth, 2006-09 (OLS)	Δ ln Non-Tradable Emp, 2007-09 (IV)	Saiz IV Residuals (OLS)	Δ Housing Net Worth, 2006-09 (OLS)	Δ ln Non-Tradable Emp, 2007-09 (IV)	LU IV Residuals (OLS)
	(1)	(2)	(3)	(4)	(5)	(6)
Saiz Unavailability	-0.1843^{***}					
	(0.0358)					
Δ Housing Net Worth, 2006-09		0.3913^{***}			0.3028^{***}	
		(0.1023)			(0.1137)	
Saiz Unavailability Error			-0.0005^{*}			-0.0001
			(0.0003)			(0.0002)
Land Unavailability				-0.1830^{***}		
				(0.0538)		
Constant	-0.0428^{***}	-0.0017	-0.0036	-0.0448^{***}	-0.0103	-0.0011
	(0.0156)	(0.0093)	(0.0042)	(0.0140)	(0.0132)	(0.0038)
Observations	567	567	567	567	567	567
R^2	0.143		0.010	0.122		0.001

Notes: See the notes for table 1.

Table 4: Regressions of MSA Area on Zip Code Land Unavailability Spread Proxies

	Dependent variable:		
	MSA Area (Sq KM, 000s)		
	(1)	(2)	(3)
Land Unavailability Zip Code Variance	6.172***		
	(2.231)		
Land Unavailability Zip Code Abs Range		82.392***	
		(20.825)	
Land Unavailability Zip Code Interquartile Range			132.068***
			(50.325)
Constant	4,695.527***	1,533.291*	3,792.827***
	(462.615)	(844.672)	(611.503)
Observations	269	269	269
R^2	0.030	0.052	0.042

Notes: Regressions of MSA Area in thousands of square KM on the spread in within-MSA Land Unavailability measured at the zip code level. Land Unavailability spread proxies include (1) the variance, (2) the range, and (3) the interquartile range. Heteroskedasticity robust standard errors are in parentheses.

Table 5: Correlations between Zip Code Land Unavailability and Demand Proxies

	Corr Coef (1)	Reg p-value (2)	Reg R^2 (3)
County Bartik	0.050	0.821	0.000
Zip Code Amenities Index	-0.246	0.000	0.081
Zip Code College Share 2000	-0.003	0.734	0.000
Zip Code Foreign Share 2000	-0.123	0.000	0.039
Zip Code Housing Density 2000	-0.015	0.720	0.000
Δ ln Zip Code Foriegn Share, 2000-2011	-0.037	0.430	0.000

Notes: Correlations between zip code Land Unavailability and housing demand proxies. The Bartik is computed from 2000-2006 at the county level and then mapped to all zip codes within that county. Column (1) displays the correlation coefficient with Land Unavailability. Columns (2) and (3) display the p-value and R^2 from individual regressions of Land Unavailability on each variable, weighted by the number of households. Heteroskedasticity robust standard errors are clustered at the three-digit zip code level.

Table 6: Zip Code Regressions of House Price Growth on Land Unavailability and Demand Proxies

	Dependent variable:							
	Zillow 2002-06 HP Growth							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Land Unavailability	0.224***	0.307***	0.323***	0.323***				
	(0.040)	(0.032)	(0.033)	(0.033)				
Slope Land Unavailability					0.195***			0.281***
					(0.039)			(0.038)
Water Land Unavailability						0.382***		0.354***
						(0.051)		(0.049)
Wetlands Land Unavailability							0.407***	0.486***
							(0.069)	(0.066)
County Bartik		4.760***	4.671***	4.675***	5.194***	4.711***	4.597***	4.439***
		(0.435)	(0.435)	(0.435)	(0.471)	(0.461)	(0.464)	(0.438)
College Share 2000		−0.324***	−0.328***	−0.328***	−0.325***	−0.313***	−0.285***	−0.316***
		(0.034)	(0.034)	(0.034)	(0.035)	(0.034)	(0.035)	(0.035)
Foreign Share 2000		1.053***	1.017***	1.017***	0.967***	1.009***	1.019***	1.039***
		(0.086)	(0.083)	(0.083)	(0.084)	(0.087)	(0.086)	(0.084)
Amenities Index			1.144**	1.152**	0.310	−0.298	0.258	1.242***
			(0.453)	(0.450)	(0.453)	(0.441)	(0.429)	(0.441)
Housing Density 2000				−0.00000	0.00000	−0.00000	0.00000	−0.00000
				(0.00000)	(0.00000)	(0.00000)	(0.00000)	(0.00000)
Observations	11,062	10,959	10,959	10,959	10,959	10,959	10,959	10,959
R^2	0.029	0.376	0.378	0.378	0.338	0.340	0.339	0.383

Notes: Regressions of 2002-2006 Zillow house price growth on Land Unavailability and housing demand proxies. All regressions are weighted by the number of households in 2000. Heteroskedasticity robust standard are clustered at the three-digit zip code level and are in parentheses.

Table 7: Zip Code Land Unavailability Regressions 2002 - 2006

	Dependent variable:					
	2002 - 2006 Zillow HP Growth					
	(1)	(2)	(3)	(4)	(5)	(6)
Land Unavailability	0.224***	−0.032	−0.084**	−0.213***	−0.198***	−0.035**
	(0.040)	(0.033)	(0.037)	(0.043)	(0.047)	(0.015)
4 Digit Zip Code Land Unavailability		0.383***				
		(0.049)				
3 Digit Zip Code Land Unavailability			0.570***			
			(0.062)			
County Land Unavailability				0.754***		
				(0.056)		
Commuting Zone Land Unavailability					0.877***	
					(0.071)	
3 Digit Zip FE	No	No	No	No	No	Yes
Observations	11,062	11,062	11,062	11,062	7,221	11,062
R^2	0.029	0.065	0.136	0.288	0.351	0.908

Notes: Regressions of zip code zillow house price growth on zip code Land Unavailability, 4 Zip Code Land Unavailability, and 3 Digit Zip Code Land Unavailability. Regressions are weighted by the number of households in 2000. Heteroskedasticity robust standard errors clustered at the three digit zip code level are in parentheses.

Table 8: Land Unavailability and House Prices in California

	$\Delta \ln$ HP 2002-06			ΔHP 2002-06		
	(1)	(2)	(3)	(4)	(5)	(6)
Land Unavailability	−0.076***	−0.062**	−0.069**	809.280***	760.345***	895.215***
	(0.022)	(0.027)	(0.032)	(300.512)	(284.916)	(340.218)
FE	County	Zip3	Zip4	County	Zip3	Zip4
Observations	1,053	1,053	1,053	1,053	1,053	1,053
R^2	0.757	0.768	0.849	0.480	0.571	0.718

Notes: Regressions are weighted by the number of households in 2000 and heteroskedasticity robust standard errors clustered at the three digit zip code level are in parentheses. Controls include the zip code level amenities index and the share of foreign born in 2000.

Table 9: Slope Unavailability and House Prices in California

	$\Delta \ln$ HP 2002-06			ΔHP 2002-06		
	(1)	(2)	(3)	(4)	(5)	(6)
Slope Unavailability	-0.066^{***}	-0.059^{**}	-0.060^{**}	632.302^{**}	659.879^{**}	816.342^{**}
	(0.020)	(0.025)	(0.030)	(270.425)	(278.607)	(339.848)
FE	County	Zip3	Zip4	County	Zip3	Zip4
Observations	1,053	1,053	1,053	1,053	1,053	1,053
R^2	0.756	0.767	0.849	0.473	0.567	0.715

Notes: Regressions are weighted by the number of households in 2000 and heteroskedasticity robust standard errors clustered at the three digit zip code level are in parentheses. Controls include the zip code level amenities index and the share of foreign born in 2000.

B Appendix: Tables

Table B1: Land Unavailability and House Prices in the US

	$\Delta \ln$ HP 2002-06			ΔHP 2002-06		
	(1)	(2)	(3)	(4)	(5)	(6)
Land Unavailability	-0.022	-0.001	-0.010	769.463***	761.810***	819.627***
	(0.015)	(0.015)	(0.018)	(127.092)	(116.437)	(138.062)
FE	County	Zip3	Zip4	County	Zip3	Zip4
Observations	11,060	11,060	11,060	11,060	11,060	11,060
R^2	0.927	0.913	0.946	0.843	0.846	0.901

Notes: Regressions are weighted by the number of households in 2000 and heteroskedasticity robust standard errors clustered at the three digit zip code level are in parentheses. Controls include the zip code level amenities index and the share of foreign born in 2000.

Table B2: Slope Unavailability and House Prices in the US

	$\Delta \ln$ HP 2002-06			ΔHP 2002-06		
	(1)	(2)	(3)	(4)	(5)	(6)
Slope Unavailability	-0.042***	-0.028	-0.029	577.006***	533.825***	637.139***
	(0.015)	(0.017)	(0.020)	(169.908)	(165.330)	(198.944)
FE	County	Zip3	Zip4	County	Zip3	Zip4
Observations	11,060	11,060	11,060	11,060	11,060	11,060
R^2	0.927	0.913	0.946	0.838	0.840	0.897

Notes: Regressions are weighted by the number of households in 2000 and heteroskedasticity robust standard errors clustered at the three digit zip code level are in parentheses. Controls include the zip code level amenities index and the share of foreign born in 2000.

Table B3: Zip Code Land Unavailability Regressions 2007 - 2010

	Dependent variable:					
	2007 - 2010 Zillow HP Growth					
	(1)	(2)	(3)	(4)	(5)	(6)
Land Unavailability	0.020	0.160^{***}	0.233^{***}	0.324^{***}	0.321^{***}	0.131^{***}
	(0.043)	(0.037)	(0.049)	(0.055)	(0.055)	(0.019)
4 Digit Zip Code Land Unavailability		-0.207^{***}				
		(0.053)				
3 Digit Zip Code Land Unavailability			-0.389^{***}			
			(0.079)			
County Land Unavailability				-0.527^{***}		
				(0.071)		
Commuting Zone Land Unavailability					-0.647^{***}	
					(0.082)	
3 Digit Zip FE	No	No	No	No	No	Yes
Observations	11,520	11,520	11,520	11,520	7,571	11,520
R^2	0.0002	0.011	0.051	0.129	0.173	0.822

Notes: Regressions of zip code zillow house price growth on zip code Land Unavailability, 4 Zip Code Land Unavailability, and 3 Digit Zip Code Land Unavailability. Regressions are weighted by the number of households in 2000. Heteroskedasticity robust standard errors clustered at the three digit zip code level are in parentheses.

Table B4: Zip Code Land Unavailability Regressions 2012 - 2016

	Dependent variable:					
	2012 - 2016 Zillow HP Growth					
	(1)	(2)	(3)	(4)	(5)	(6)
Land Unavailability	-0.042	-0.209^{***}	-0.268^{***}	-0.350^{***}	-0.322^{***}	-0.122^{***}
	(0.035)	(0.026)	(0.028)	(0.032)	(0.032)	(0.012)
4 Digit Zip Code Land Unavailability		0.247^{***}				
		(0.047)				
3 Digit Zip Code Land Unavailability			0.410^{***}			
			(0.052)			
County Land Unavailability				0.530^{***}		
				(0.040)		
Commuting Zone Land Unavailability					0.675^{***}	
					(0.045)	
3 Digit Zip FE	No	No	No	No	No	Yes
Observations	12,907	12,907	12,907	12,907	8,475	12,907
R^2	0.002	0.024	0.088	0.203	0.301	0.839

Notes: Regressions of zip code zillow house price growth on zip code Land Unavailability, 4 Zip Code Land Unavailability, and 3 Digit Zip Code Land Unavailability. Regressions are weighted by the number of households in 2000. Heteroskedasticity robust standard errors clustered at the three digit zip code level are in parentheses.