

Highly Disaggregated Topological Land Unavailability*

Chandler Lutz
Copenhagen Business School

Ben Sand
York University

First Available Draft: August 10, 2017

August 31, 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 disaggregated data in the construction of land unavailability as 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, we show that Land Unavailability is uncorrelated with housing demand proxies, validating disaggregated Land Unavailability as an instrument for house prices.

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. 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), 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. 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. First, we review the Saiz methodology and how he constructed Land Unavailability. Saiz constructs his measure of Land Unavailability by computing the percentage of undevelopable land within a 50 km radius circle from an MSA centroid. Yet MSAs are small in the Northeast, where house price growth was relatively muted during the 2000s, but large in California and the Southwest, areas that experienced high house price volatility before and during the crisis. Hence,

¹Saiz’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 is likely endogenous leaving topographical land unavailability as the candidate instrument.

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 introducing bias into two-stage least squares (2SLS) estimates that use Land Unavailability as an instrument.

In the construction of our Land Unavailability proxy, we use more accurate satellite data that is now available from the United States Geological 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 typically the geographic aggregation for Land Unavailability used by researchers 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 Land Unavailability is measured 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.² Using zip code level data, we find that Land Unavailability is not 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 an evaluation. We find that the Land Unavailability in the areas surrounding the

²See, for example, [Mian and Sufi \(2011, 2014\)](#) and [Davidoff et al. \(2016\)](#).

a given zip code is a key predictor of house prices, highlighting how households move within cities in response to a neighborhood price shock.

2 Data Sources

The United States Geographical Survey (USGS) provides the two main datasets that we use to measure slope and water land unavailability.³ 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 approximately a resolution of 30 meters.⁴ These 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.⁵ 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, and 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<https://viewer.nationalmap.gov/launch/>

⁴For a sample file, see <https://www.sciencebase.gov/catalog/item/5903e5b0e4b022cee40c773d>. The Coordinate Reference System (CRS) used for these data is GRS80.

⁵For a sample, see <https://www.sciencebase.gov/catalog/item/581d5a13e4b0dee4cc8e5120>. The CRS used for these data are NAD83.

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 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 US metropolitan area.

As an example of the geographies that Saiz uses within each MSA, we plot Google satellite imagery for the Los Angeles-Long Beach and Riverside-San Bernardino MSAs in figure 1. Here, the blue outlined areas in the map represent the polygon boundaries for the Los Angeles and Riverside MSAs, respectively. The red dots are the centroids of each polygon, and the red circles represent a 50 km radius around the MSA centroid. The 50 km circle around the centroid of the Los Angeles-Long Beach MSA captures most of the Los Angeles area, but does not cover important geographical areas in Southern Los Angeles such as Torrance or Long Beach. In the polygon for the Riverside-San Bernardino MSA (right polygon on the plot), the circle around the centroid is in rather sparsely populated and flat area between the Mojave National Preserve, Joshua Tree National Park, and the San Bernardino National Forest and thus misses the key population areas in Riverside, Ontario, San Bernardino, and Palm Springs. Further, as the key population centers around Riverside are surrounded by the San Gabriel and San Bernardino Mountains, measuring land unavailability within a 50 km of the MSA centroid likely understates the percentage of undevelopable land facing Riverside inhabitants. As a second example, consider the Las Vegas MSA shown in figure 2. Again the circle with a 50 km radius around the city centroid does not overlap with the key population areas or major freeways in Las Vegas or Henderson. Figures 1 and 2 also show the disparity of geographic land area within MSAs and, along with differing housing market and income dynam-

ics within cities, suggesting that more disaggregated measures of land unavailability may be of use to researchers.

Together, figures 1 and 2 highlight instances where a circle with a 50 km radius around the city centroid may produce land unavailable calculation irregularities. Generally, MSAs that span large geographic areas are prone to larger estimation errors in the aforementioned calculation of land unavailability, while the circles with a 50 km radius circle cover more land area than comparatively smaller polygons. Obviously if this error is 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 3 extends figures 1 and 2 and plots all MSA polygons for the Saiz dataset in blue and the corresponding circles with a 50 km radius centered around the centroid 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 4 further highlights the geographic coverage of the circles relative to the MSA polygons. Here, yellow areas correspond to MSAs where the circle with a 50 km radius completely covers the MSA, while blue represents MSAs whose polygons extend beyond the circle. Colors in the map are the absolute difference in area between the MSA and the polygon where more color for a given polygon represents a larger absolute difference. The difference in coverage of the 50 km radius circles across US geographies is stark: In the Northeast the circles are typically larger than the MSAs (and thus are in yellow), while MSAs in the Southwest and California, cities that experienced large housing variance in the 2000s, are notably larger than their corresponding circles.

For all of the MSAs in the Saiz dataset, the correlation between FHFA house price growth from 2002 - 2005 and the difference between the MSA polygons and their corre-

sponding circles (in square km) is 0.32 (t-stat = 5.12).⁶ Similarly, when aggregating the data up to the state level, the correlation between the average difference in size between the polygons and their circles within each state and 2002-2005 FHFA house price is 0.43 (t-stat = 3.05). Hence, house price growth is highly correlated with the difference in the area between the MSA polygon and the circle with a 50 km radius centered at the centroid.

Typically in housing analyses using Land Unavailability, researchers employ the following first stage regression:

$$\Delta \ln HousePrice = \beta_0 + \beta_1 Unavailability + \varepsilon \quad (1)$$

We run this regression using 2002-2005 FHFA house prices and the Saiz Unavailability proxy. As expected, $\hat{\beta}_1$ is positive and significant (White t-stat = 9.89), 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 0.34 (t-stat = 5.60), meaning that areas where the MSA polygon is large relative to its 50 km radius circle have a larger residual for the regression in equation 1. These findings, together with those in figures 3 and 4, show that the Saiz proxy for Land Unavailability is highly correlated with MSA size and house price growth and thus that Land Unavailability measures need to account for polygon size.

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

⁶We multiply areas where the circle covers the polygon by -1 .

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 5. 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 MSA 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 easily 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 6 shows a scatter plot of our land unavailability measure compared with Saiz. The slope of 0.82 and R^2 of 0.71 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. We explore this possibility in table 1 and figure 7. First in table 1, 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 7 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, 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 [Davidoff et al. \(2016\)](#) contends that Land Unavailability is positively correlated with housing demand. We examine the correlations of between Land Unavailability and proxies of demand at the zip code level. The proxies of demand that we consider at the zip code level include (1) the county Bartik shock mapped to the zip code level; (2) a zip code level amenities index; and (3) 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 2. 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 the amenities index and Land Unavailability are negative and significant. This is not a concern as these correlations would need to be positive to invalidate Land Unavailability as an instrument. Further, these negative correlations are not surprising as higher land unavailability increases the marginal cost of amenity construction and immigrant enclaves are highly related to the distance from the US boarder or traditional port cities.

Table 3 shows zip code level regressions of 2002 - 2005 Zillow house price growth on Land Unavailability and proxies of demand. These regressions 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 0.225. Column (2) also includes the Bartik labor demand shock, the college share, and the foreign born share. The coefficient on Land Unavailability jumps to 0.304 as Land Unavailability is negatively correlated with the foreign born share. The standard error also falls to 0.033 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. Last, column (3) adds the amenities index and the housing density. The coefficient on Land Unavailability increases further the amenities index is also negatively correlated with Land Unavailability. 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.

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 Land Unavailability at the 4 and 3 digit level (e.g. the first four and three digits of the zip code). The 4 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. We report the predictive effects of Land Unavailability at different levels of aggregation in table 4. Column (1) is identical to column (1) 3 and show that a one percent increase in Land Unavailability is associated with a 0.225 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 0.344. Notice that the R^2 in column (2) doubles to 0.06 relative to column (1). The coefficient on the 3 digit zip code unavailability in column is even larger at 0.523 and the R^2 doubles again to 0.124. Clearly, the areas surrounding a zip code are important for Land Unavailability and 3 digit zip code provides strong predictive power for highly disaggregated house prices.

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 more disaggregated Land Unavailability data is preferred. 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 find that Land Unavailability, especially in surrounding areas, 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.
- 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.
- 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.

A Tables

Table 1: Regressions of MSA Area on Zip Code Land Unavailability Spread Proxies

	<i>Dependent variable:</i>		
	MSA Area (Sq KM, 000s)		
	(1)	(2)	(3)
Land Unavailability Zip Code Variance	6.194*** (2.240)		
Land Unavailability Zip Code Abs Range		84.566*** (21.453)	
Land Unavailability Zip Code Interquartile Range			131.562*** (50.333)
Constant	4,754.817*** (474.811)	1,523.305* (857.338)	3,865.551*** (608.005)
Observations	257	257	257
R ²	0.030	0.054	0.041

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 2: 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

Notes: Correlations between zip code Land Unavailability and housing demand proxies. The Bartik is computed 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 3: Zip Code Regressions of House Price Growth on Land Unavailability and Demand Proxies

	<i>Dependent variable:</i>		
	Zillow 2002-05 HP Growth		
	(1)	(2)	(3)
Land Unavailability	0.225*** (0.040)	0.304*** (0.033)	0.318*** (0.033)
Bartik		4.704*** (0.450)	4.627*** (0.454)
College Share 2000		−0.275*** (0.034)	−0.279*** (0.034)
Foreign Share 2000		1.043*** (0.081)	1.012*** (0.078)
Amenities Indiex			1.011** (0.432)
Housing Density 2000			−0.00001 (0.0001)
Constant	34.478*** (1.755)	8.060*** (1.800)	8.240*** (1.798)
Observations	11,062	10,959	10,959
R ²	0.030	0.373	0.375

Notes: Regressions of 2002-2005 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.

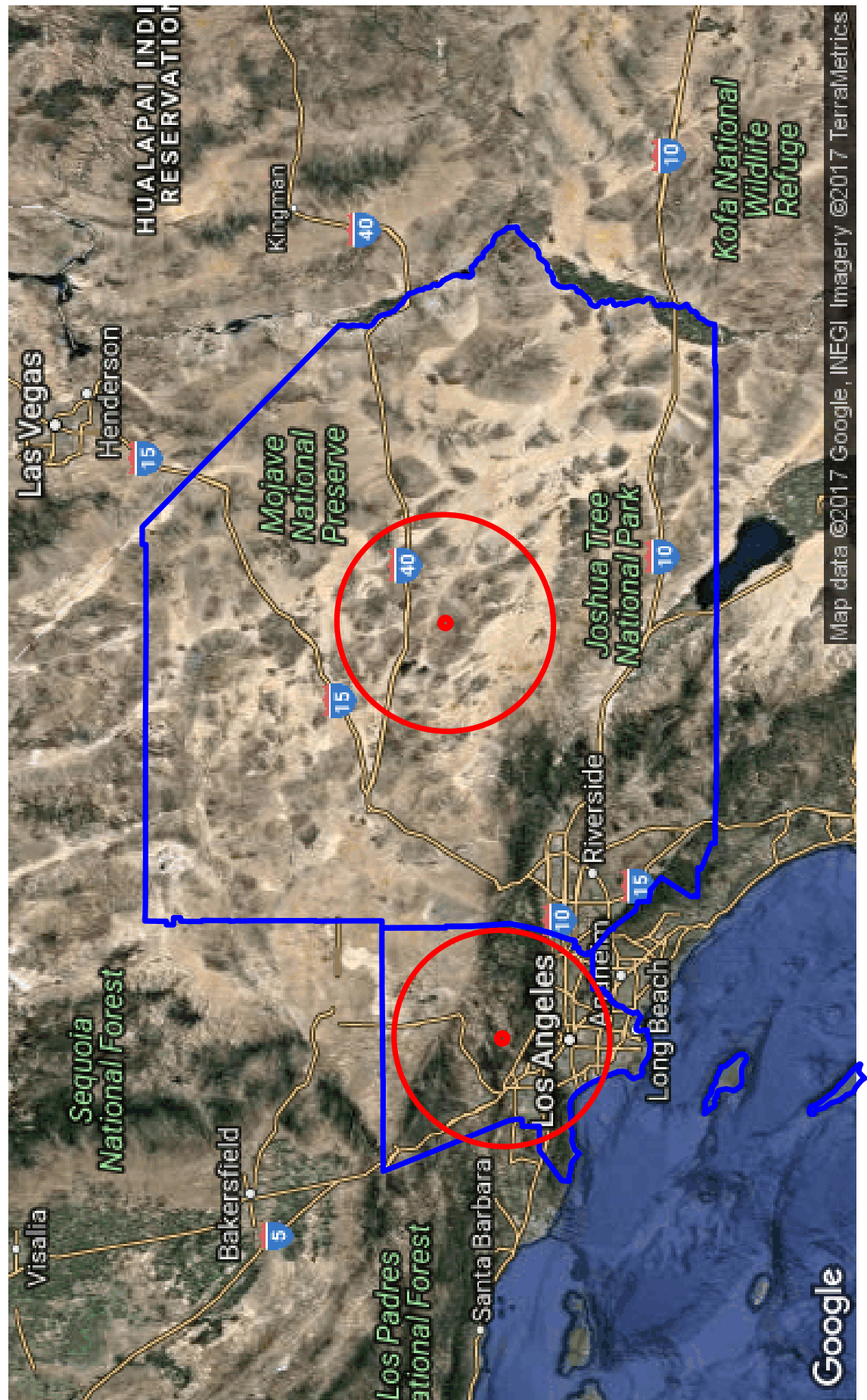
Table 4: Zip Code Land Unavailability Regressions

	<i>Dependent variable:</i>		
	2002-2005 Zillow HP Growth		
	(1)	(2)	(3)
Land Unavailability	0.225*** (0.040)	-0.005 (0.037)	-0.057 (0.040)
4 Digit Zip Code Land Unavailability		0.344*** (0.053)	
3 Digit Zip Code Land Unavailability			0.523*** (0.065)
Constant	34.478*** (1.755)	31.169*** (1.908)	26.718*** (1.973)
Observations	11,062	11,062	11,062
R ²	0.030	0.061	0.124

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 are in parentheses.

B Figures

Figure 1: Saiz Land Unavailability Coverage for the Los Angeles and Riverside MSAs



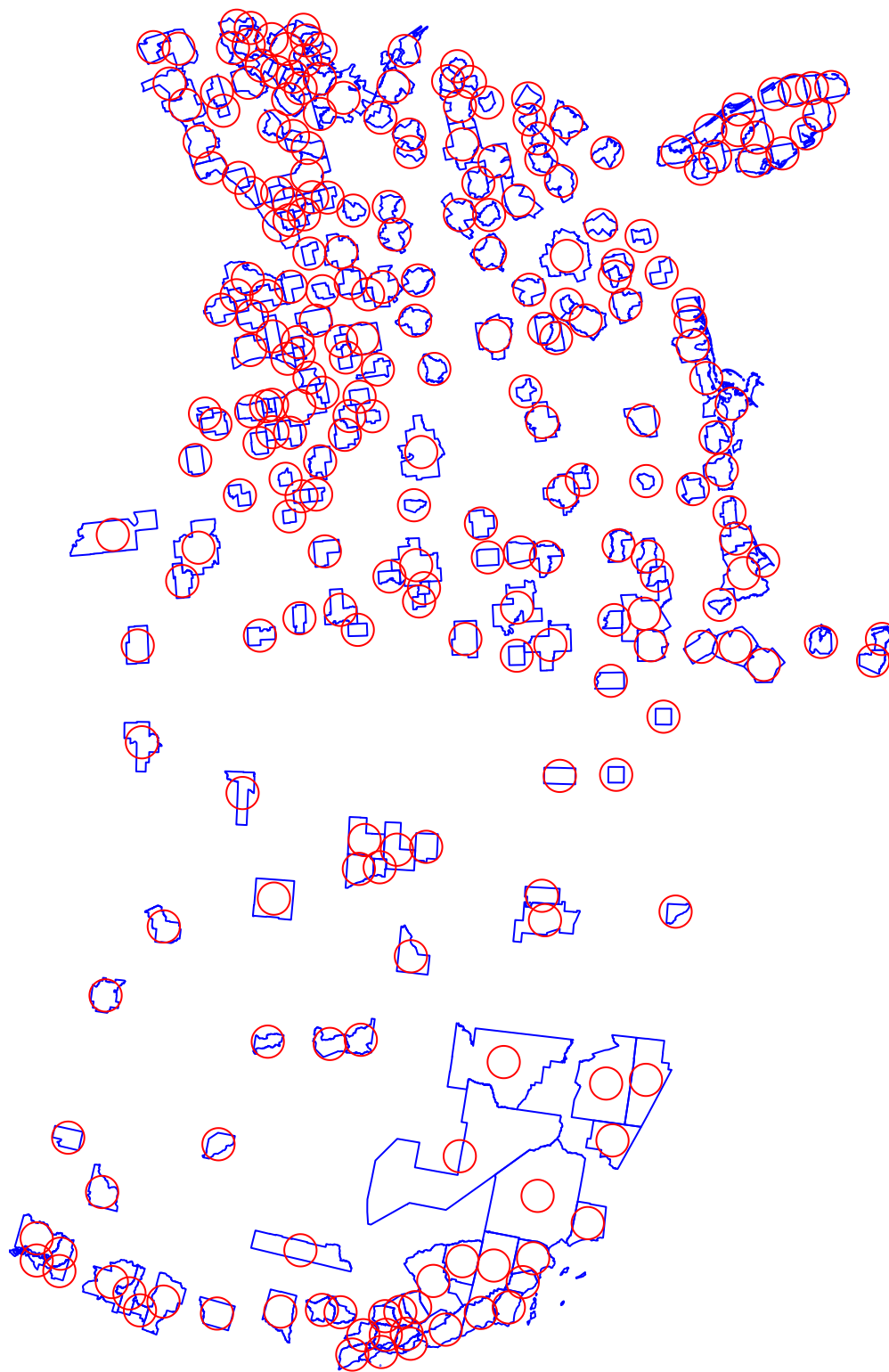
Notes: The blue lines represent the MSA polygons for the Los Angeles-Long Beach and Riverside-San Bernardino MSAs, respectively. The red dots are the centroids for the polygons, and the red circles have a radius of 50 kilometers and are centered around polygon centroid.

Figure 2: Saiz Land Unavailability Coverage for the Las Vegas MSA



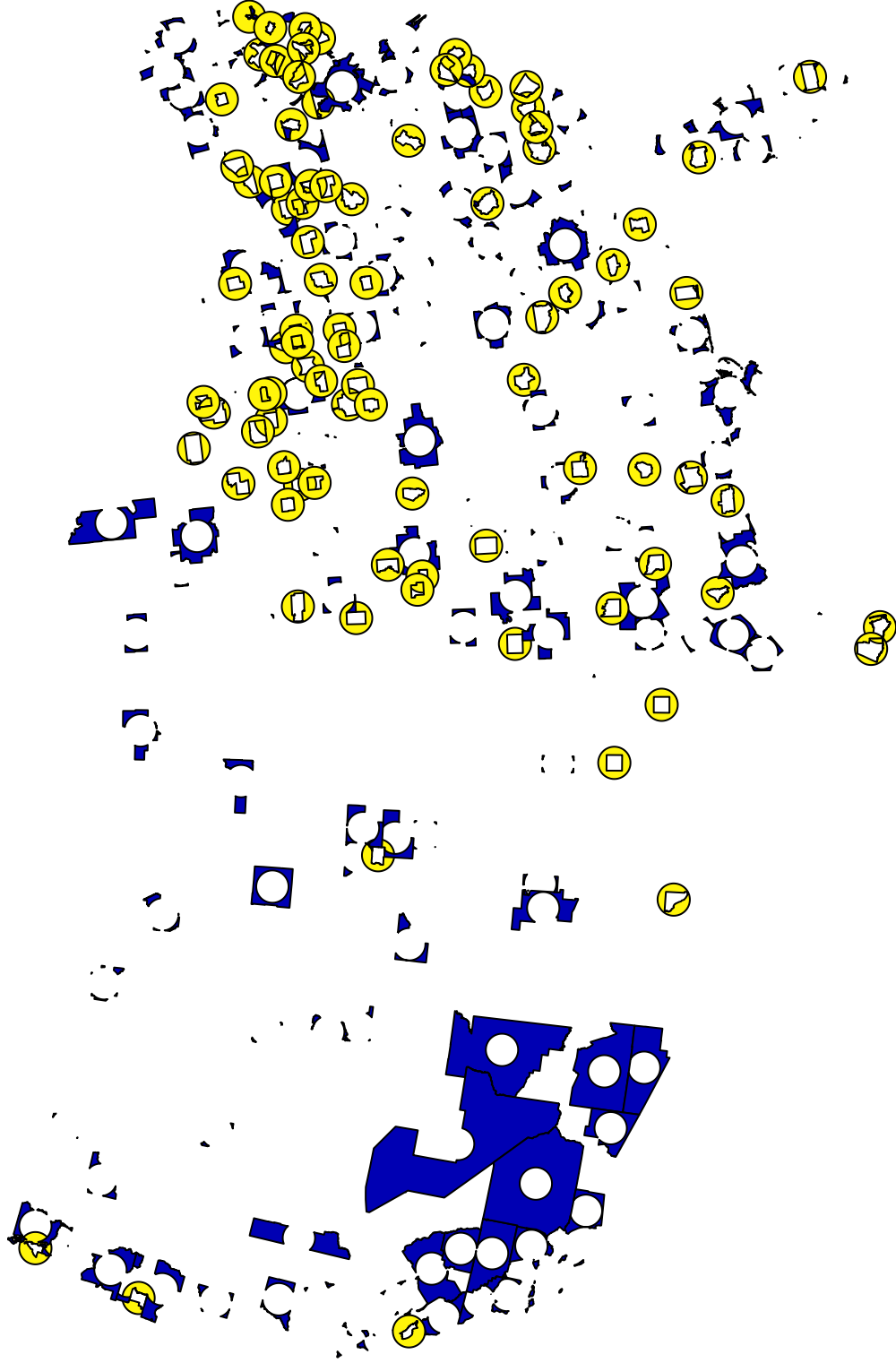
Notes: See the notes for figure 1.

Figure 3: Saiz Land Unavailability Coverage Across US MSAs



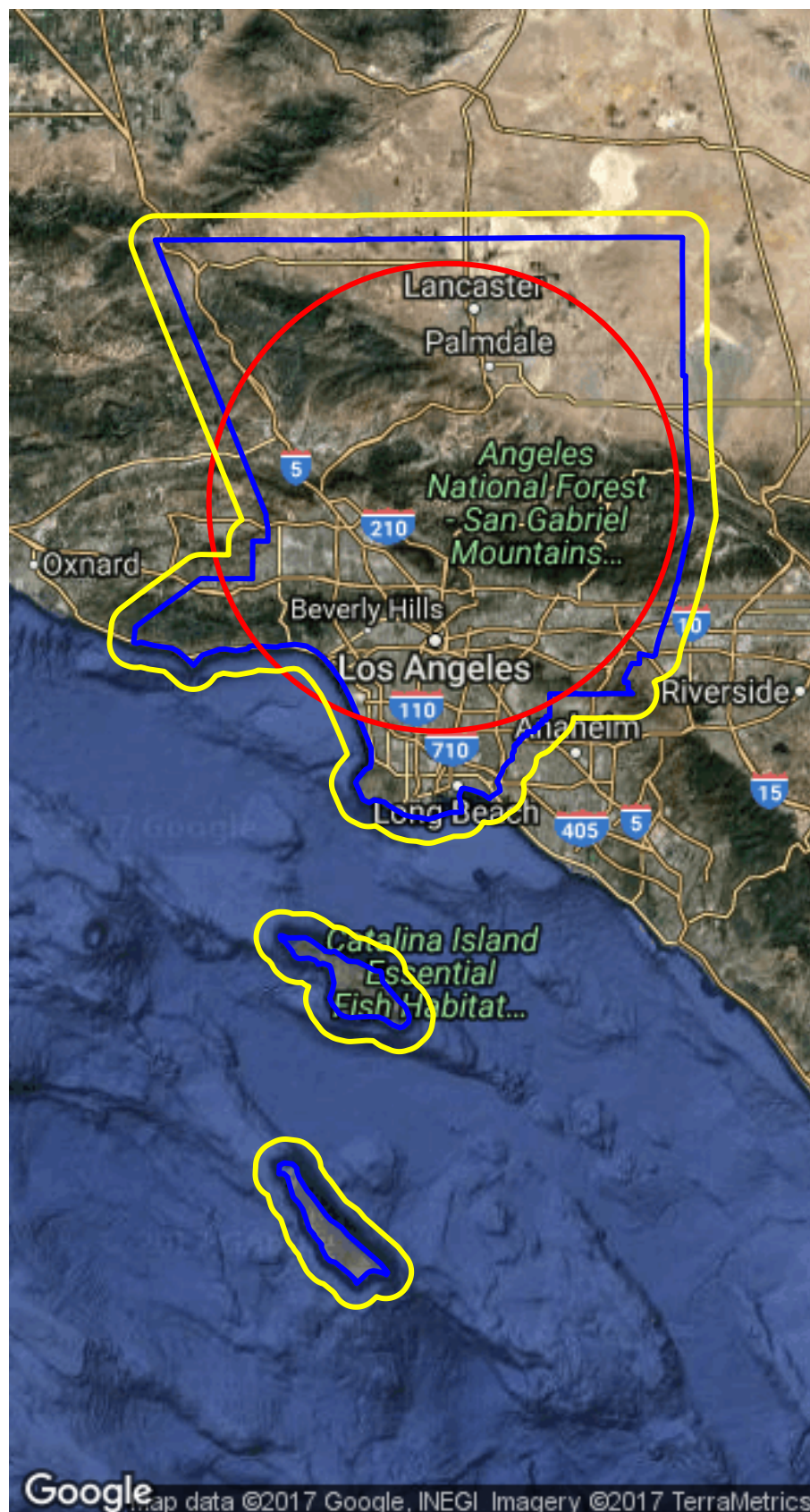
Notes: The blue lines represent MSA polygons. The red circles have a radius of 50 kilometers and are centered around polygon centroid.

Figure 4: Difference between MSA polygons and the Saiz Land Unavailability Coverage



Notes: Yellow areas coincide with areas where a circle with a 50 km radius centered at the polygon centroid covers the polygon MSA. Blue areas correspond to instances where the circle with a 50 km radius centered at the centroid does not cover the polygon.

Figure 5: Saiz and Buffered Land Unavailability Coverage for the Los Angeles MSA



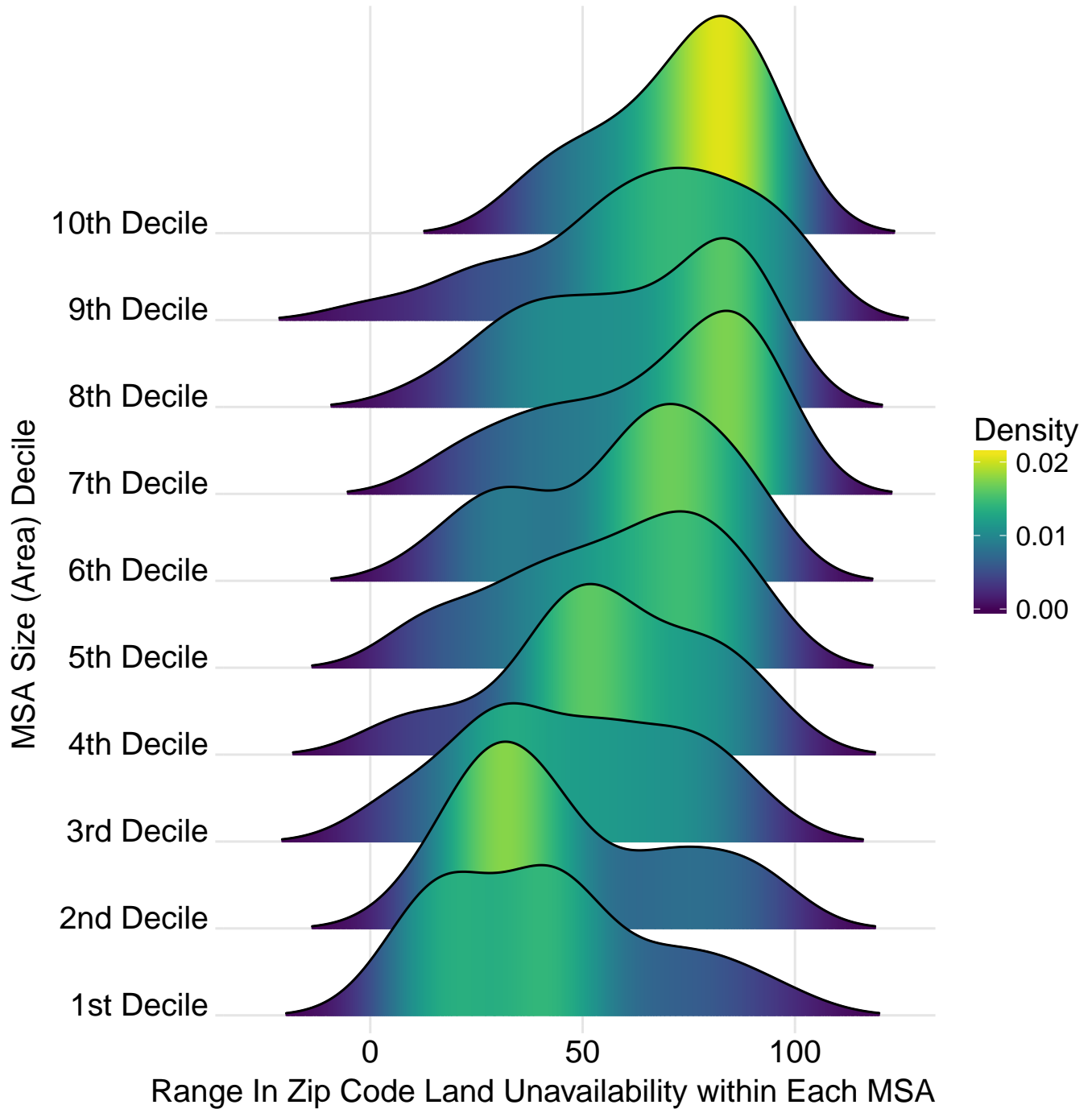
Notes: See the notes for figure 1. The yellow line is a 5 percent buffer around the Los Angeles MSA and represents the boundary used to calculate land unavailability in this paper.

Figure 6: 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 7: 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.