Highly Disaggregated Topological Land Unavailability*

Chandler Lutz
Copenhagen Business School

Ben Sand
York University

First Available Draft: August 10, 2017

August 10, 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.

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), firm formation (Adelino et al., 2015), 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 from 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 an instrumental variable 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 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 and calculate land unavailability at various units of economic geographic aggregation including MSAs, counties, and zip codes. Our zip code proxy for land unavailability may be especially pertinent for researchers with the proliferation of zip code level housing market data and analyses.

2 Data Sources

The United States Geographical Survey (USGS) provides the two main datasets that we use to measure slope and water land unavailability.[2]. The first is the USGS National Elevation

[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 is likely endogenous leaving topographical land unavailability as the candidate instrument.

[2] https://viewer.nationalmap.gov/launch/

Dataset (NED) 3DEP 1 arc-second Digital Elevation Model. The 1 arc-second DEM data provide continuous coverage of the United states at approximately a resolution of 30 meters.[3] 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.[4] These data use Landsat to classify land use in the US. The relevant categories for this paper are water (oceans, lakes, rivers, etc.) and wetlands.

In addition our study also includes the following data: Shapefiles for various geographies from the US Census Bureau, satellite imagery from Google Maps, and house prices from the FHFA.

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 unavailability 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 Torrence 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

[3]For a sample file, see https://www.sciencebase.gov/catalog/item/5903e5b0e4b022cee40c773d.
[4]For a sample, see https://www.sciencebase.gov/catalog/item/581d5a13e4b0dee4cc8e5120.

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.

Together, figures 1 and 2 highlight cases where a circle with a 50 km radius around the city centroid may produce land unavailable calculation irregularities. Generally, MSAs that span large geographic are prone to larger estimation errors in the aforementioned calculation of land unavailability. 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. For all of the MSAs in the Saiz dataset, the correlation between FHFA house price growth from 2002-2005 and land area (square km) is 0.28 (t = 4.58). Similarly, when aggregating the data up to the state level, the correlation between the average MSA size within each state and 2002-2005 FHFA house price is 0.44 (t = 3.02).

Figures 1 and 2 also show the disparity of geographic land area within MSAs and, and along with differing housing market and income dynamics within cities, suggest that more disaggregated measures of land unavailability would be of use for researchers.

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 or 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 3. 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.

Despite the differences in computation method, our proxy for Land Unavailability is highly correlated with that from Saiz (2010). Figure 4 shows a scatter plot of our land unavailability measure compared with Saiz. The slope of 0.82 and and R^2 of 0.71 highlights the similar nature of our two measures.

Next, using the method described above, we compute the Land Unavailability at various other levels of geographic aggregation including at the county and zip code levels. Please contract the authors regarding Land Unavailability data inquiries.

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.

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.

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

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.

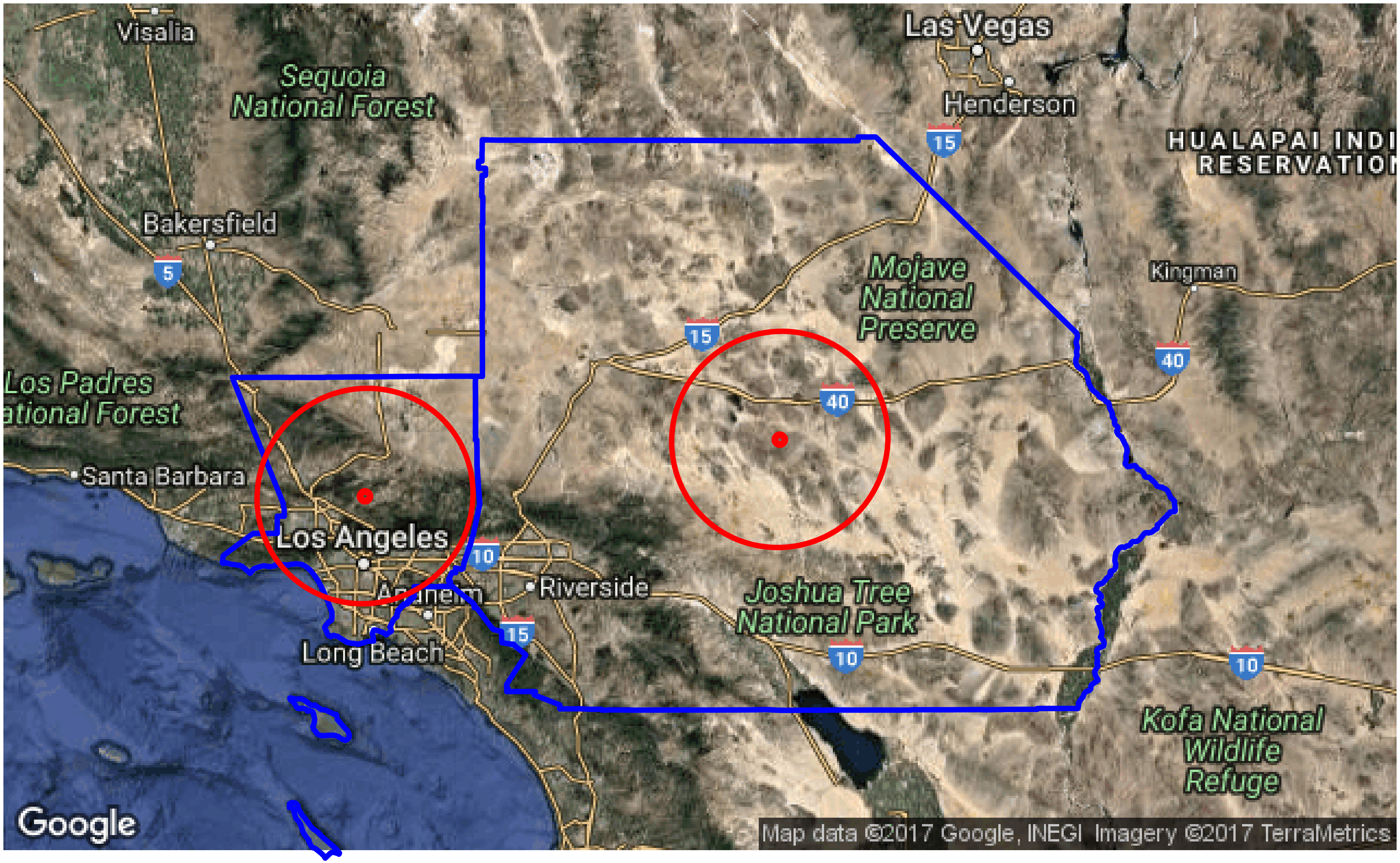
A. Mian and A. Sufi. What explains the 2007–2009 drop in employment? *Econometrica*, 82 (6):2197–2223, 2014.

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.

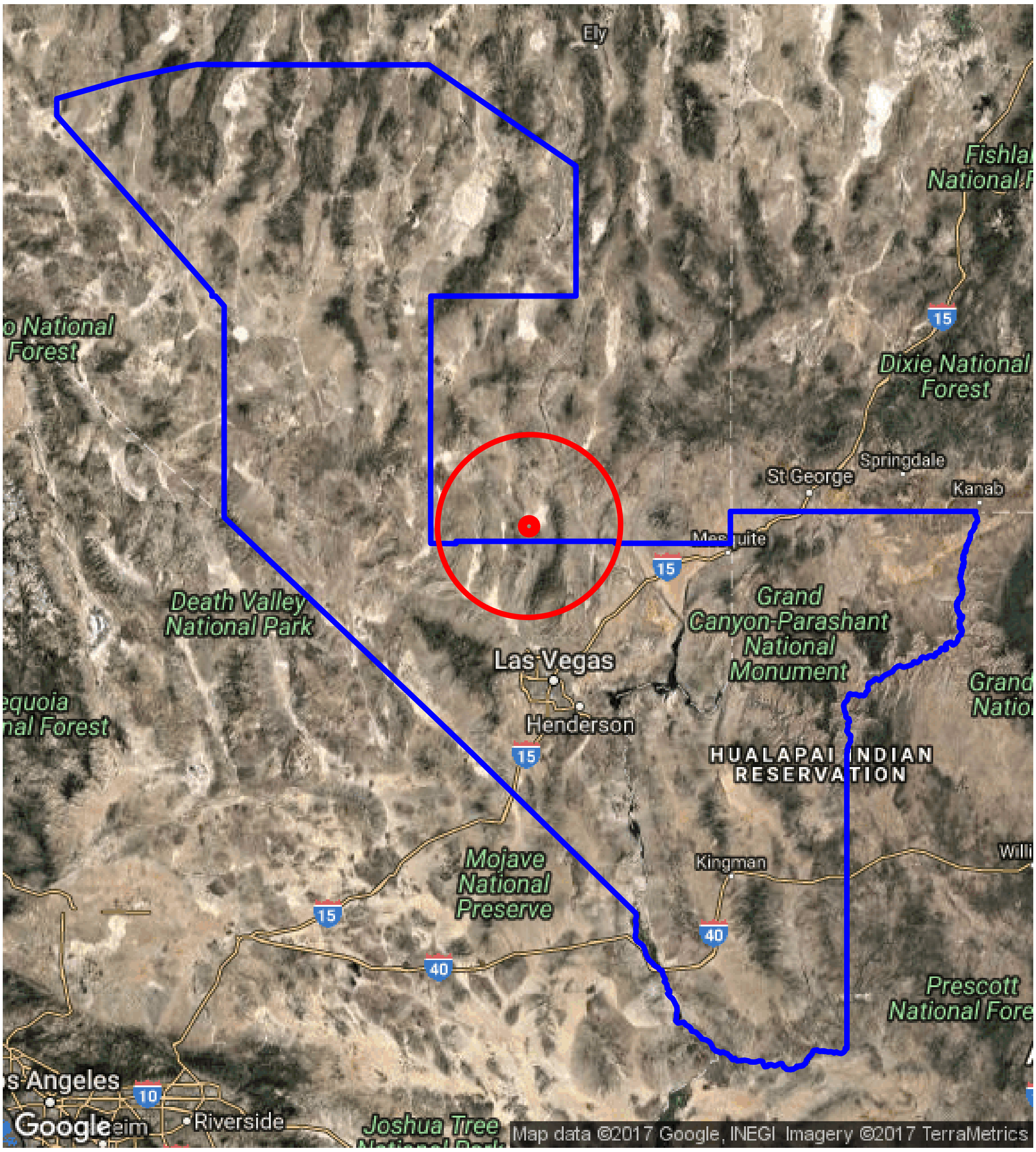
A. Saiz. The geographic determinants of housing supply. *Quarterly Journal of Economics*, 125 (3), 2010.

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



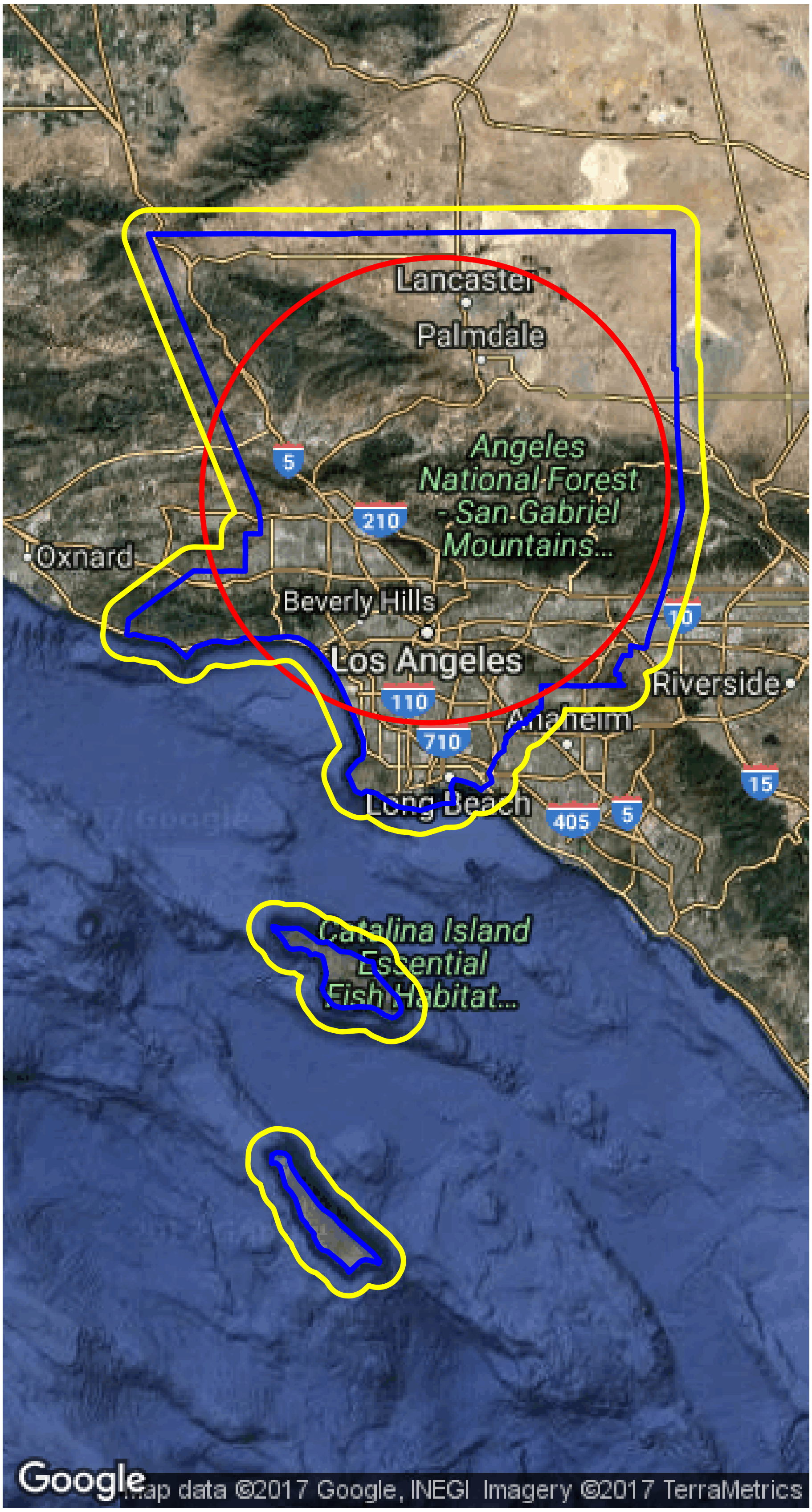
Notes: The blue lines represent the MSA polygons for the Los Angeles-Long Beach and Riverside-San Bernardino MSAs, respectively. The red dots are the centroids for the polygons, and the red circles have a radius of 50 kilometers and are centered around polygon centroid.

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



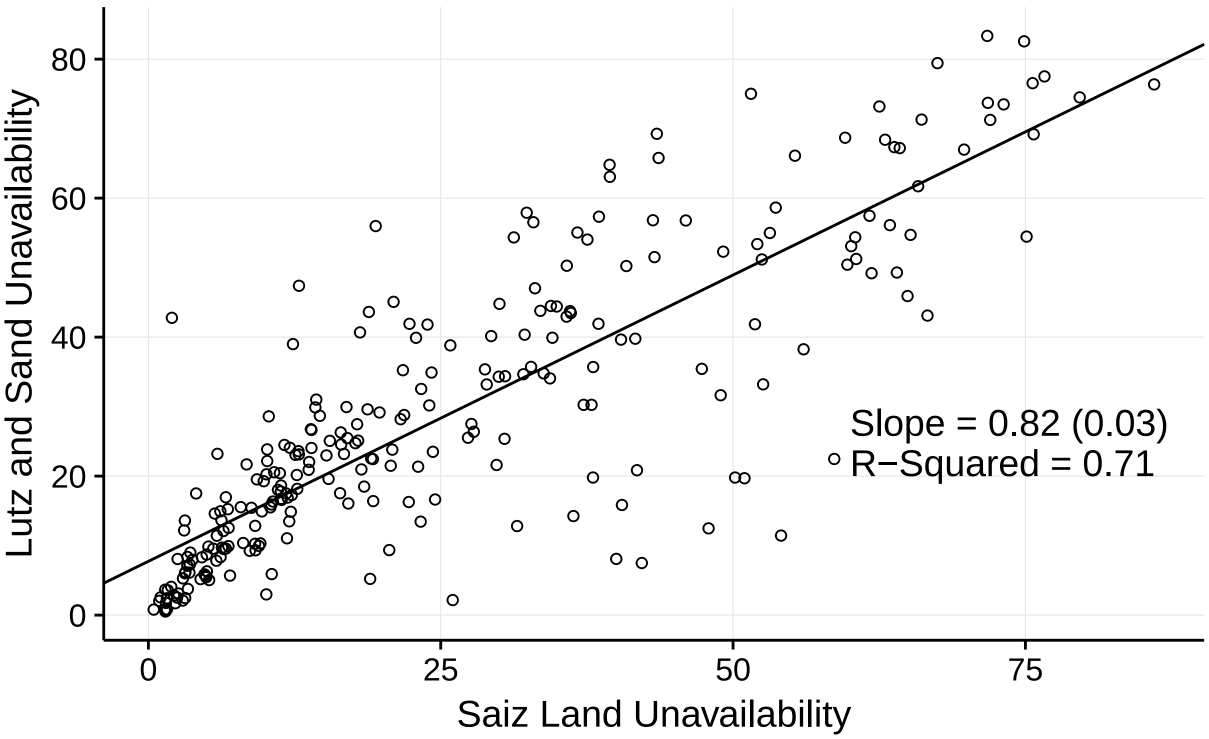
Notes: See the notes for figure 1.

Figure 3: 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 MSA and represents the boundary used to calculate land area in this paper.

Figure 4: 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.