Highly Disaggregated Land Unavailability*

Chandler Lutz
U.S. Securities and Exchange Commission

Ben Sand
York University

Land Unavailability Data:
https://github.com/ChandlerLutz/LandUnavailabilityData

October 31, 2019

Abstract

We use new large-scale data techniques and comprehensive high resolution satellite imagery for the contiguous United States to construct novel datasets that capture the geographic determinants of house prices and housing supply: The percentage of undevelopable land – Land Unavailability – and its complement, buildable land. 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. Using highly disaggregated data we show that Land Unavailability is uncorrelated with housing demand proxies, validating Land Unavailability as an instrument for house prices; that the geographic components of Land Unavailability, especially in combination with modern machine learning techniques, provide substantial incremental predictive power for house prices; and previous studies that employed limited land unavailability datasets underestimated the impact of house prices on unemployment during the Great Recession by 30%. With our buildable land dataset we then test the supply side speculation theory that aims to explain the previously puzzling large house price growth in traditionally elastic housing markets. In line with theory, results document that housing markets with intermediate amounts of buildable land, those that are elastic in the short run but plausibly inelastic in the long run, experienced abnormally large house price growth during the 2000s.

JEL Classification: R30, R31, R20;

Keywords: Land Unavailability, Buildable Land, Real Estate, Housing Market

*Lutz: lutzc@sec.gov. Sand: bmsand@yorku.ca. First Available Draft: August 10, 2017.

1 Introduction

Housing is the largest financial asset for the typical US household (Tracy and Schneider, 2001). Economists have thus naturally attempted to understand the variation in house prices and connect their fluctuations to the causes and consequences of the financial crisis.[1] A key insight emanating from the housing literature posits that given a demand shock, *ceteris paribus*, house price growth across housing markets will be linked to the elasticity of supply.[2] More bluntly, in the face of a common positive demand shock, cities with a more rigid housing supply will experience larger house price increases. Geographic impediments to construction, along with regulatory constraints, constitute key factors that contribute to housing supply inelasticity. In the context the nationwide boom during the 2000s, this theory predicts relatively damped house growth in areas with elastic housing supply (e.g. Wichita, Kansas) and substantially higher growth in so-called inelastic markets (Mian and Sufi, 2009).

With this framework in mind, economists sought to explain the financial crisis and the Great Recession through cross-sectional differences in house price growth. Noting that several potential sources of endogeneity obfuscate such economic relationships, researchers pursued 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), a key component in the determination of housing market elasticity.[3] 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.). As

[1] See Mian and Sufi (2009, 2011, 2014). Economists have also connected fluctuations in house prices to 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; Chetty et al., 2017).

[2] See Saiz (2010); Davidoff (2013, 2016) and the references therein.

[3] 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.

slope, water, and wetlands are determined by nature, this yields a plausibly exogenous instrument for house price changes. Using this instrument economists typically pursue a two-stage least squares (2SLS) 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.[4] If Land Unavailability is exogenous, 2SLS then yields the causal relationship between house prices and the outcome of interest. This approach, however, and the exogeneity of Land Unavailability more broadly, has recently been questioned (Davidoff, 2016), clouding our understanding of the financial crisis and the Great Recession as well as potentially rendering a plethora studies that measure the impacts of house prices during the 2000s invalid.[5] Further, the elasticity-house price relationship failed to explain the spectacular 2000s house price growth in many traditionally elastic Sand State markets.

In this paper, we address these issues through new, large-scale datasets based on precise satellite imagery that capture the geographic determinants of housing supply. Specifically, we first construct plausibly exogenous Land Unavailability due to water, wetlands, and steep sloped terrain as well as the amount of buildable land at the start of the housing boom (this proxy accounts for land unavailable due to water, wetlands, steep sloped terrain, previous construction, and parks). Our robust approach allows for these measures to be compiled at various levels of geographic disaggregation down to the zip code level. Broadly, our work extends the popular Land Unavailability proxy of (Saiz, 2010) in several directions. For example, we use more accurate satellite data that is now available from the United States Geographical Survey (USGS) and also exploit more precise and geo-spatially consistent polygon areas in our calculations. This easily allows us to extend our computational method to various units of geographic disaggregation.

We first examine the correlation between Land Unavailability and proxies for hous-

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

[5]Most of the studies listed in footnote 1 use Saiz elasticity to pursue causal inference.

ing demand as the use of Land Unavailability as an instrument depends on its exogeneity relative to demand factors. As noted above, recent research instead has suggested that Land Unavailability is not exogenous, potentially invalidating several studies that use it as an instrument for causal inference.[6] Criticisms of the use of Land Unavailability as an instrument contend households' demand with respect to unobserved demand factors related to amenities, the economics of agglomeration associated with higher education, and labor demand shocks have been increasing and thus cannot be accounted for through standard region fixed effects. Thus positive correlation between Land Unavailability and the foregoing housing demand factors would imply that Land Unavailability is correlated with these unobserved demand changes and subsequently is not a valid instrument for house prices. Yet previous attempts to assess the exogeneity of Land Unavailability suffer from an intrinsic sample selection issue: The only previously available Land Unavailability proxy from Saiz (2010) used MSAs, where MSA instantiation requires a population of 50,000 or more. MSAs thus do not provide complete coverage of the United States and are based on historical delineations of development. Using complete coverage of the United States and unique, highly local geographic data, we find that Land Unavailability is not positively correlated with amenities; the portion of people who are college educated or foreign born in 2000; or annual Bartik (1991) labor demand shocks over the 2000s. Together, these results validate the use of Land Unavailability as an instrument for house prices during the 2000s.

With our new dataset we also examine the predictive power of Land Unavailability. Not only do we re-consider the 2002 - 2006 boom and the 2006 - 2009 bust, but also the recent 2011 - 2017 post-Great Recession expansion. Establishing plausibly exogenous predictors for house price dynamics is increasingly important given the sustained volatility in local housing markets (Leamer, 2007; Ferreira and Gyourko, 2011; Sinai, 2012; Glaeser and Gyourko, 2018). Researchers clearly would like to employ such predictors in 2SLS designs, while policymakers need to predict housing market re-

[6] See Davidoff (2016) and the references in footnote 1.

sponses to economic shocks. We evaluate the predictive power of Land Unavailability, its components (e.g. slope, water, and wetlands), and various geographies, using both traditional and modern machine learning techniques. We find that the inclusion of the components of Land Unavailability increases predictive power and that combination of the Land Unavailability, its components, and the modern machine learning techniques produces substantially more accurate predictions than have previously been used in the literature.

Using our satellite imagery, we also construct an important new dataset that precisely measures the amount of buildable land in 2001, prior to the 2000s housing boom, within a geographic polygon. Buildable land is the amount of land available for development, after removing existing development, steep sloped terrain, water, wetlands, and parks. In a sense, buildable land is the complement to our Land Unavailability, but also accounts for previous development and parks. We then use this dataset to examine one of the largest puzzles in housing finance: Why did traditionally elastic housing markets, like Las Vegas, experience substantial house price growth even though these markets had room for housing construction expansion? In particular, we test the land supply side speculation theory of Nathanson and Zwick (2018). This theory posits that homebuilders during the 2000s boom viewed traditionally elastic housing markets with intermediate amounts of land available (e.g. Las Vegas and Phoenix) as potentially inelastic in the long run. Homebuilders then proceeded to bid up the prices of land in these intermediate markets and, as land is a key input for home construction, prices increased along with construction. While Nathanson and Zwick (2018) do provide anecdotal evidence in support of their theory, they were unable to perform formal statistical tests as no comprehensive buildable land dataset was previously available. In this paper, we undertake such tests and find that housing markets with intermediate amounts of buildable land experienced large house price booms during the 2000s, congruent with the supply side speculation theory.

Finally, we re-evaluate the impact of housing on unemployment during the Great Recession through a replication and extension of Mian and Sufi (2014). Mian and

Sufi find that large adverse housing net worth shocks negatively affected non-tradable employment, a key consequence of the economic malaise during the 2000s. We first replicate Mian and Sufi's key results which use the Saiz elasticity proxy as an instrument for house prices in their main causal regressions. We then extend their work using our Land Unavailability dataset and also employ a new machine learning approach to instrument variable (IV) estimation, the rigorous post-Lasso approach of Belloni et al. (2012), Belloni et al. (2014), and Chernozhukov et al. (2016). Using our expansive dataset in combination with the rigorous post-Lasso framework, we nearly double the sample size of Mian and Sufi's key regressions, likely reducing the 2SLS finite sample bias (Angrist and Krueger, 2001), and exploit multiple potential instruments generated by our new data. Moreover, Land Unavailability is more likely to be exogenous than the Saiz elasticity proxy used by Mian and Sufi (2014) as Saiz elasticity includes housing market regulatory constraints that are often a consequence of house prices (Davidoff, 2016; Wallace, 1988). Our results indicate that Mian and Sufi previously underestimated the impact of housing net worth shocks on unemployment by 30 over percent.

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; section 4 outlines the construction of our Land Unavailability dataset; section 5 presents correlations between Land Unavailability and housing demand factors during the 2000s; the predictive power of Land Unavailability with regard to house prices is in section 6; section 7 develops our buildable land dataset and tests the supply side speculation theory; our replication and extension of Mian and Sufi's analysis of employment during the Great Recession is in section 8; and section 9 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.[7] The first is the

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

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.[8] The original Saiz dataset 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.[9] These data use Landsat imagery to classify land use in the US. The relevant categories for Land Unavailability 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 to 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 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 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

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

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

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 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 MSA (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.

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 2 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.

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 1. 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 is able to compute Land

Unavailability at levels of aggregation used by economists and researchers.

Despite the differences in computational methods, our proxy for Land Unavailability is highly correlated with that from Saiz (2010). Figure 3 shows a scatter plot of our land unavailability measure compared with Saiz. The slope of 0.80 and and R^2 of 0.70 highlight the similar nature of our two measures.

5 The Validity of Land Unavailability as an Instrument during the 2000s

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 (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 building and construction, MSAs are unlikely to be located in areas with high Land Unavailability. To see this, consider figure 4 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 between Land Unavailability and proxies of demand with near complete national coverage. The proxies of demand that we consider include a zip code amenities index compiled from a large internet aggregator of such information, the college share in 2000, the foreign share in 2000, and housing density in 2000. We consider these variables at the zip code and three digit zip code levels. The output of regressions of these variables on Land Unavailability is in table 1 where panel A shows the results using zip code data and zip code Land Unavailability, while panel B measures all variables at the three digit zip code level. Regressions are weighted by the number of households in 2000 and robust standard errors are clustered at the county level (panel A) or commuting zone level (panel B). As suggested by Davidoff (2016), in order for Land Unavailability to fail the exclusion restriction, Land Unavailability must be positively correlated with other housing demand factors as households' demand for amenities, the economics of agglomoration, etc. has been increasing over time and cannot be accounted using standard region fixed effects. Instead, our results show the opposite. At the zip code level, Land Unavailability is negatively correlated the amenities index, foreign share, and housing density, while being uncorrelated with college share. These results are not surprising as increased Land Unavailability makes the construction of housing and amenities more expensive. At the three digit zip code level, Land Unavailability is slightly negatively correlated with amenities but uncorrelated with the other housing demand factors.

Another key determinant of housing demand are changes to labor demand within

a city. We follow the labor literature and Davidoff (2016) and employ Bartik (1991) Labor demand shocks at the county level. We assess the correlation between Land Unavailability and the Labor demand shocks through the following specification:

$$Bartik_{it} = \alpha + \delta_t + \sum_{y=2002}^{2010} \theta_y LandUnavailability_i \mathbf{1}\{y = t\} + \varepsilon_{it} \tag{1}$$

$\mathbf{1}\{y = t\}$ is an indicator function that takes a value of 1 for year y and the coefficients of interest, θ_y, $y = 2002, \ldots, 2010$, measure the annual correlations between Land Unavailability and the Bartik Labor Demand shocks. The regressions are weighted by the number of households in 2000 and robust standard errors are clustered the county level. The regression output for θ_y is in figure 5 where the shaded bands are ± 2 standard errors. Clearly, the Bartik Labor Demand shocks are uncorrelated with Land Unavailability during the 2000s boom and bust.

Overall, the results in this section show that Land Unavailability is uncorrelated with key housing demand factors during the 2000s and thus suggest that Land Unavailability does not violate the exclusion restriction as an instrument for 2000s house price growth.

6 The Predictive Power of Disaggregated Land Unavailability

We assess the predictive power of our Land Unavailability data for house price growth during the 2000s boom and bust, as well as during the recent, post-Great Recession expansion. The ability to predict house price growth using Land Unavailability is important for both researchers and policymakers. Researchers often use Land Unavailability as an instrument for house price growth and instrument relevance is a necessary condition, while volatile housing markets contribute to economic booms and busts (Leamer, 2007; Sinai, 2012; Glaeser and Gyourko, 2018) and predicting cross-sectional house price growth is a key issue for policymakers.

We compare the predictive power of our Land Unavailability measures relative to the Saiz's proxy out-of-sample using 10 repeats of 10-fold cross-validation via OLS,

Lasso, and Random Forest learners. We use the rigorous Lasso of Belloni et al. (2012) where we conduct variable selection via Lasso and then model estimation using OLS (Belloni et al., 2014). The Random Forest uses optimal tuning parameters as suggested by Breiman (2001). Our Land Unavailability predictors exploit different levels of geographic disaggregation as well as the components of Land Unavailability (e.g. slope unavailability, water unavailability, and wetlands unavailability). The holdout sample within each fold is constructed using only regions available in the Saiz dataset. All available data is used for model training and the mean-squared error (MSE) is used to evaluate model performance.

We examine the predictive power of Land Unavailability at the county level, a level of disaggregation often used for house prices. Note that the Saiz dataset only has information available for 269 MSAs (with population of 50,000 or more) and the map of the Saiz MSA data to the county level is from Mian and Sufi (2014). In contrast, our Land Unavailability data comprises all counties in the contiguous United States. Thus in model training, our holdout sample will consist of a biased sample of counties that are associated with large MSAs and the inclusion of all counties may adversely affect predictive performance for linear models, like OLS and Lasso. The Random Forest model, however, which is apt at discovering non-linearities in the data, may be able to exploit this information to improve predictive performance. In the end, we retain all available data to avoid data snooping.

The results are in table 2. The table reports the average MSE for each specification across the different holdout samples with bootstrapped 95 percent confidence intervals in parentheses. The results for the 2002 - 2006, 2006 - 2009, and 2011 - 2017 time periods are in panels A, B, and C, respectively. Column (1) shows the results using Saiz Land Unavailability only, where the first row within each panel displays the results from OLS; this is the model previously used in literature to predict house price growth. The average MSE using Saiz Land Unavailability during the 2002 - 2006 period is 373.46. Note that the t-statistic (using robust standard errors clustered at the state level) from a full-sample OLS regression of house price growth on Saiz Unavailability

during the 2002 - 2006 period is 7.62, indicating that Saiz Unavailability has strong predictive power for house price growth and more accurate predictions from other models accuracy are notable.

When using Saiz Unavailability, the MSE falls for the 2006 - 2009 and 2011 - 2017 time periods, meaning that OLS predictions were substantially more accurate during the two latter time periods. The Lasso predictions in column (1) match the OLS findings, meaning the Lasso considers Saiz Unavailability as relevant predictor for house price growth in all periods (Lasso sets the coefficients of non-relevant predictors to zero; in this case if the coefficient on Saiz Unavailability is set to zero, the prediction would have been based on the sample mean). Finally, the Random Forest model in the last row each panel in column (1) produces substantially more accurate predictions than the OLS estimates, indicating that unspecified non-linearities in the Saiz Unavailability are important for house price growth predictions. Overall, this latter result is not surprising as Random Forests generally provide more accurate predictions compared to traditional techniques (Mullainathan and Spiess, 2017; Athey, 2018).

Column (2) shows the predictions using our county Land Unavailability measure only, while column (3) presents the predictive results from Land Unavailability at different geographic aggregates and for adjacent, touching counties (county, commuting zone, adjacent mean, adjacent min, and adjacent max Land Unavailability). These predictions are slightly less accurate than those in column (1), possibly due to the bias in the Saiz holdout samples.

Columns (4) and (5) document that including more geographic information improves predictive performance. Column (4) includes all components of Land Unavailability (total, water, slope, and wetlands) available at all of the geographies employed in column (3) for a total of 15 predictors. Column (5) explicitly forces non-linearities in each model by including all the predictors from column (4) and their squares. Overall, the predictive power of these models, especially the Random Forest, is noteworthy. Specifically, column (5) suggests that the MSE from using all Land Unavailability components and their squares during 2002 - 2006 is 199.74, corresponding a decrease of 46

percent relative to the Saiz Unavailability OLS estimates in the top row of column (1) in Panel A. The Random Forest models in column (5) also produce similar increases in predictive accuracy during the 2006 - 2009 and 2009 - 2012 periods.

7 Buildable Land and Supply Side Speculation

Nathanson and Zwick (2018) develop a theoretical model that documents how disagreement and supply side speculation in housing markets can produce house price booms in traditionally supply elastic areas. Specifically, the model posits that homebuilders may view housing markets with intermediate amounts of land available for development (buildable land) as supply elastic in the short run, but inelastic in the long run. When these homebuilders are optimistic about future prices (e.g. in the midst of a national housing boom during the 2000s), they acquire and subsequently bid up the prices of available land. As land is a key factor in housing production, this raises house prices in markets with intermediate amounts of buildable land even in the face of large scale construction and generates house price booms in traditionally supply elastic housing markets. This theory aims to explain the previously puzzling house price booms in areas like Phoenix, Las Vegas, Florida, and inland California.

Nathanson and Zwick (2018) provide several pieces of empirical evidence in support of their theory. For example, they cite a Polte homes investor presentation that stated that the traditionally elastic markets of West Palm Beach, Orlando, Tampa, Ft. Myers, Sarasota, Las Vegas, and Chicago were surprisingly constrained. A more formal test of the supply side speculation theory would require precise data on the amount of buildable land within housing markets. To our knowledge, no such dataset exists.

In this section, we exploit detailed satellite Land Cover and Slope image files to construct a new, unique dataset that precisely measures amount of buildable land across the contiguous United States.

The basis of our computation of buildable land is the 2001 USGS LandSat Land Cover Dataset. The LandSat Land Cover data classifies land use in the United States at a spacial resolution of 30 meters. Figure 6 plots the LandSat Land Cover data for

Florida. In the satellite image, red pixels correspond to developed land, where darker red pixels represent more dense development. Similarly, blue areas represent water and wetlands. The most developed area is downtown Miami (dark red in southeast Florida) and the map clearly shows how water and wetlands restrict housing expansion in that housing market. Oppositely, other areas along coastal and in central Florida are comparatively at intermediate stages of development with lower density and surrounding areas that appear to be available for development.

We compute land area available for development within each housing market by first removing developed land (e.g. red pixels on the Florida map) and water and wetlands (blue pixels). We also remove steep sloped terrain measured using USGS 1 arc-second DEM slope files (using no buffer for polygons in the shapefiles) and exclude regions designated as parks using a shapefile from data.gov. We then calculate the land area of the remaining, buildable land.

In a sense, buildable land is the complement to our Land Unavailability proxy constructed above, but additionally classifies start of period developed land (2001) and parks as unavailable as well.

We compute buildable land within three digit US zip codes. As US zip codes were developed in the 1960s they better reflect pre-2000s housing boom US populations and geographies, especially in the Western US, compared to counties or MSAs counties which are based on geographic definitions dating back to the 1800s.[10]

To test the relationship between buildable land and 2002 - 2006 house price growth, we group three-digit US zip codes into 2001 buildable land deciles. Summary statistics for buildable land deciles are in table 3. Column (1) shows the buildable land decile and column (2) displays the average amount of buildable land for three digit US zip codes in that buildable land decile (thousands of square kilometers). As expected in column (2), buildable land is monotonically increasing in the over buildable land deciles. Notice also, however, that there is very little available buildable land in deciles 1 and 2. Three

[10]For example, the land area of the Riverside-San Bernardino MSA is 260 percent larger than the land area of the entire state of Massachusetts.

digit zip codes in these deciles are likely the "inelastic" housing markets characterized by Nathanson and Zwick that likely have both high Land Unavailability and regulatory supply restrictions.[11] Similarly, column (3) shows the mean percentage of land that is buildable (relative to all available land) within each buildable land decile. Again, the percentage of buildable land is monotonically increasing over buildable land deciles. A potential concern when using buildable land defined within three digit zip codes, which can vary in size, is that buildable may simply be a function of available land. We partially address this in column (4) which shows the correlation between available and buildable land by buildable land decile. The correlations are wide ranging and only in buildable land decile 10 is the correlation with available land over 0.5. We return to this issue below.

Figure 7 maps three-digit US zip codes where red areas correspond to buildable land decile 1 (least amount of buildable land), blue areas represent buildable land decile 5 (intermediate amount of buildable land), and yellow areas are buildable land decile 10 (largest amount of buildable land). Buildable land decile 1 indeed corresponds to housing markets that would traditionally be considered "inelastic" due to density, Land Unavailability, and regulatory constraints. These housing markets in include New York City, Boston, Miami, Downtown Tampa, New Orleans, Downtown Chicago, Downtown Milwaukee, Coastal Los Angeles, and areas adjacent the San Francisco Bay. Three digit zip codes in buildable land decile 5 (intermediate amounts of buildable land) consist of suburban areas in inland southern California, central California, and northern California. Buildable land decile 5 also includes Las Vegas, Phoenix, Colorado Springs, several suburban regions in central and coastal Florida, suburban Chicago, and several suburban housing markets in the northeast. Finally, yellow areas showing buildable land decile 10 are largely rural areas in the Midwest and Texas.

Nathanson and Zwick's supply side speculation theory aims to explain housing markets with intermediate land supply. Note also that they concede that supply inelastic markets should also experience a large house price growth during a boom

[11] See also the references in Nathanson and Zwick (2018).

(e.g. Saiz (2010)) and that the house price growth in inelastic markets is not the focus of their theory. Thus, the null hypothesis of interest is that house price growth in traditionally supply elastic areas with intermediate amounts of buildable land is equal to house price growth in areas with relatively smaller or relatively larger amounts of buildable land. A rejection of this null supports the supply side speculation theory and would yield a hump-shaped, non-monotonic relationship between buildable land decile and house price growth.

We evaluate the supply side speculation theory in table 4 by regressing three digit zip code 2002 - 2006 house price growth on 2001 buildable land decile indicators. Robust standard errors clustered at the state level are in parentheses. Column (1) shows the mean house price growth within each buildable land decile. Not surprisingly, house price growth is largest in areas with the least amount of buildable land (buildable land decile 1, likely inelastic markets), at 58.6 percent. Yet the second highest mean house price growth is in buildable land decile 5 at 44.1 percent followed closely by buildable land decile 4 at 42.9 percent. House price growth in buildable deciles 2 and 3 is substantially smaller at 35 and 27 percent, respectively (buildable land decile 2 also likely contains inelastic housing markets, accounting for its slightly higher house price growth relative to decile 3). Similarly, house price growth is markedly lower for buildable land deciles 6 through 10. Note also that the R-squared is 25 percent and thus suggests that the buildable land deciles explain a large portion of the cross-sectional variation in house price growth during the 2000s. Together, this evidence suggests that inelastic and housing markets with intermediate amounts of buildable land experienced the largest house price growth during the 2000s.

Columns (2) and (3) statistically test the supply side speculation theory. Here we exclude the indicator for buildable land decile 5, but retain the intercept. Thus the intercept is the house price growth for buildable land decile 5 and the regression coefficients are the difference in mean house price growth relative to decile 5. The coefficients on the indicators for buildable deciles 2 and 3 as well as deciles 6 through 10 are all negative and statistically significant at the 1 percent in column (2). Hence three

digit zip codes in buildable land deciles 2, 3, and 6 - 10 experienced noticeably lower house price growth than three digit zip codes with intermediate amounts of buildable land. Similarly, column (3) shows that controlling for plausibly exogenous Bartik Labor Demand Shocks does not affect our results (the Bartik is demeaned relative to the entire sample so the intercept can be interpreted as the mean house price growth in decile 5 in a three digit zip code with an average Bartik shock). Together, these regressions document that three digit zip codes with intermediate amounts of buildable land experienced statistically larger house price growth from 2002 - 2006, congruent with the supply side speculation theory.

As noted above, a potential concern with the construction of buildable land within three digit zip codes is that buildable land may be a function of available land and that the amount of available land within a three digit zip code may be driving our results. We address this concern with a falsification test. Specifically, we retain all three digit zip codes outside of buildable land deciles 1 (plausibly inelastic areas) and 5 (intermediate buildable land areas). Of these remaining regions, we then collect the three digit zip codes whose available land is within the range of available land for the original buildable land decile 5. This yields 294 (out of 607) three digit zip code regions whose available land is within the range of buildable 5. The mean house price growth for these regions is 25.1 percent. All other three digit zip codes outside of our original deciles 1 and 5 have a mean house price growth 31.9 percent. The difference of -6.8 percentage points is statistically significant at the 1 percent level (robust t-stat $= -2.6$). Hence, other three digit zip codes whose available land is within the range of the available land for regions in buildable land decile 5 actually have *lower* house price growth. The results from this falsification test thus suggest that buildable land, and not available land, drive the above relationship between buildable land and house price growth.

8 Land Unavailability, Housing Markets, and Unemployment During the Great Recession

In this section, we replicate and extend Mian and Sufi (2014) who examine the impact of housing net worth shocks on employment during the Great Recession. Mian and Sufi find that adverse housing net worth shocks adversely impacted non-tradable employment during the 2000s bust. Specifically, they construct non-tradable employment based on retail and restaurant employment (Rest. and Retail) or geographic concentration (Geog. Concen.) where more non-tradable employment sectors are assumed to be more geographically disperse. See Mian and Sufi (2014) for a more detailed description of this data. The key regression of interest is

$$\Delta \ln E_i^{\text{NT}} = \alpha + \eta \cdot \Delta HNW_i + \gamma \mathbf{X}_i + \varepsilon_i \tag{2}$$

where $\Delta \ln E_i^{\text{NT}}$ is the log change in non-tradable employment for county i from 2007 to 2009, ΔHNW_i is the change in housing net worth from 2006 to 2009, and $\mathbf{X}_i$ is a vector of industry controls. The coefficient of interest, η, measures the elasticity between housing net worth and non-tradable employment.

The results are in table 5. Columns (1) - (4) replicate the key results from Mian and Sufi (2014). Columns (1) and (2) show the OLS estimates while (3) and (4) present the 2SLS estimates using Saiz elasticity (both Land Unavailability and regulatory constraints) as an instrument. There are several things to notice in columns (1) - (4). First, the change in housing net worth is positive and statistically significant, indicating that a decline in housing net worth is associated with a drop in non-tradable employment. The IV estimates in columns (3) and (4) are larger and thus imply that the OLS estimates are biased towards zero. Yet as we move from the OLS estimates in columns (1) and (2) to the IV estimates that employ Saiz elasticity in columns (3) and (4), the numbers of observations falls by nearly half. This loss of observations is not ideal given the finite sample bias of 2SLS (Angrist and Krueger, 2001). In

columns (5) through (8), we employ our Land Unavailability proxy that is more likely to be exogenous than Saiz elasticity and was not correlated with housing demand factors during the 2000s. In columns (4) and (5), we retain the Saiz sample and the 2SLS coefficients increase slightly, but are less precise. When we use all available data in columns (7) and (8) (936 total observations), the coefficients increase further. In fact, the coefficient estimate in column (7) is 37 percent larger than Mian and Sufi's corresponding estimate in column (3), while our estimate of the elasticity when non-tradable employment based geographic concentration is the dependent variable is twice as that from Mian and Sufi.

Finally, in columns (9) and (10) we employ the rigorous post-Lasso approach of Belloni et al. (2012), Belloni et al. (2014), and Chernozhukov et al. (2016) for instrument selection. The advantage of this approach is that we can consider Land Unavailability and its components along with their interactions at multiple levels of geographic disaggregation. This yields 230 candidate instruments. The rigorous lasso selects the mean Land Unavailability in adjacent (touching) counties as the only instrument in the first stage, implying that this is the most relevant predictor for housing net worth from 2006 - 2009. From there, we perform 2SLS relying Chernozhukov et al. (2016) and the references therein for theoretical justification for the standard errors. The results again are larger than those from Mian and Sufi, but also smaller and slightly more precise than our previous estimates in columns (7) and (8).

Altogether, the results in this section document the robustness of the relationship between housing net worth and non-tradable employment during the Great Recession to the use of different instruments and techniques, but also note that previous studies slightly underestimated the magnitude of the effects.

9 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 geographic disaggregation.

Using our new data, we re-examine the predictive power of Land Unavailability, its correlation housing demand proxies during the 2000s, exploit satellite imagery to construct a novel dataset of buildable land, and extend our understanding of the relationship between housing net worth and employment during the Great Recession.

Specifically, results indicate that Land Unavailability is a key predictor of housing markets during the 2000s boom, the 2000s bust, and during the recent post-Great Recession expansion. The geographic components of Land Unavailability and modern machine learning techniques also improve predictive performance.

Further, recent work posited that Land Unavailability is not an exogenous predictor of house prices, clouding several studies that connect the fall of housing markets to the downturn of economic activity during the Great Recession (Davidoff, 2016). We first note that that previous attempts to examine the correlations between Saiz MSA-level Land Unavailability and housing demand proxies suffer from sample selection bias related to MSA instantiation. Then we use our disaggregated Land Unavailability dataset with near complete coverage of the United States to assess the correlation between Land Unavailability and housing demand proxies. Findings indicate that there little evidence of correlation between Land Unavailability and proxies of hosing demand.

Using the complement of Land Unavailability, we construct a new, comprehensive dataset for buildable land to test the recent supply side speculation theory that contends that homebuilders perceive housing markets with intermediate amounts of buildable land as elastic in the short run, but inelastic in the long run. The empirical tests are consistent with theory and help explain the previously puzzling 2000s boom in traditionally elastic Sand State housing markets.

Future research may employ our Land Unavailability or buildable land datasets in the prediction of housing markets for policy purposes, to test housing market theories, or to compute causal estimates through instrumental variable techniques. Our robust method for the compilation of the unavailable or buildable land also allows researchers

to conduct tests using comprehensive data that span the United States at all levels of geographic disaggregation commonly employed in the literature.

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.

J. D. Angrist and A. B. Krueger. Instrumental variables and the search for identification: From supply and demand to natural experiments. *Journal of Economic perspectives*, 15(4):69–85, 2001.

S. Athey. The impact of machine learning on economics. *Working Paper*, 2018.

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.

A. Belloni, D. Chen, V. Chernozhukov, and C. Hansen. Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80 (6):2369–2429, 2012.

A. Belloni, V. Chernozhukov, and C. Hansen. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2):29–50, 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.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

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.

V. Chernozhukov, C. Hansen, and M. Spindler. High-dimensional metrics in r. *arXiv preprint arXiv:1603.01700*, 2016.

R. Chetty, L. Sándor, and A. Szeidl. The effect of housing on portfolio choice. *The Journal of Finance*, 72(3):1171–1212, 2017.

T. Davidoff. Supply elasticity and the housing cycle of the 2000s. *Real Estate Economics*, 41(4):793–813, 2013.

T. Davidoff. 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.

F. Ferreira and J. Gyourko. Anatomy of the beginning of the housing boom: Us neighborhoods and metropolitan areas, 1993-2009. 2011.

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. Glaeser and J. Gyourko. The economic implications of housing supply. *Journal of Economic Perspectives*, 32(1):3–30, 2018.

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.

S. Mullainathan and J. Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

C. G. Nathanson and E. Zwick. Arrested development: Theory and evidence of supply-side speculation in the housing market. 2018.

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

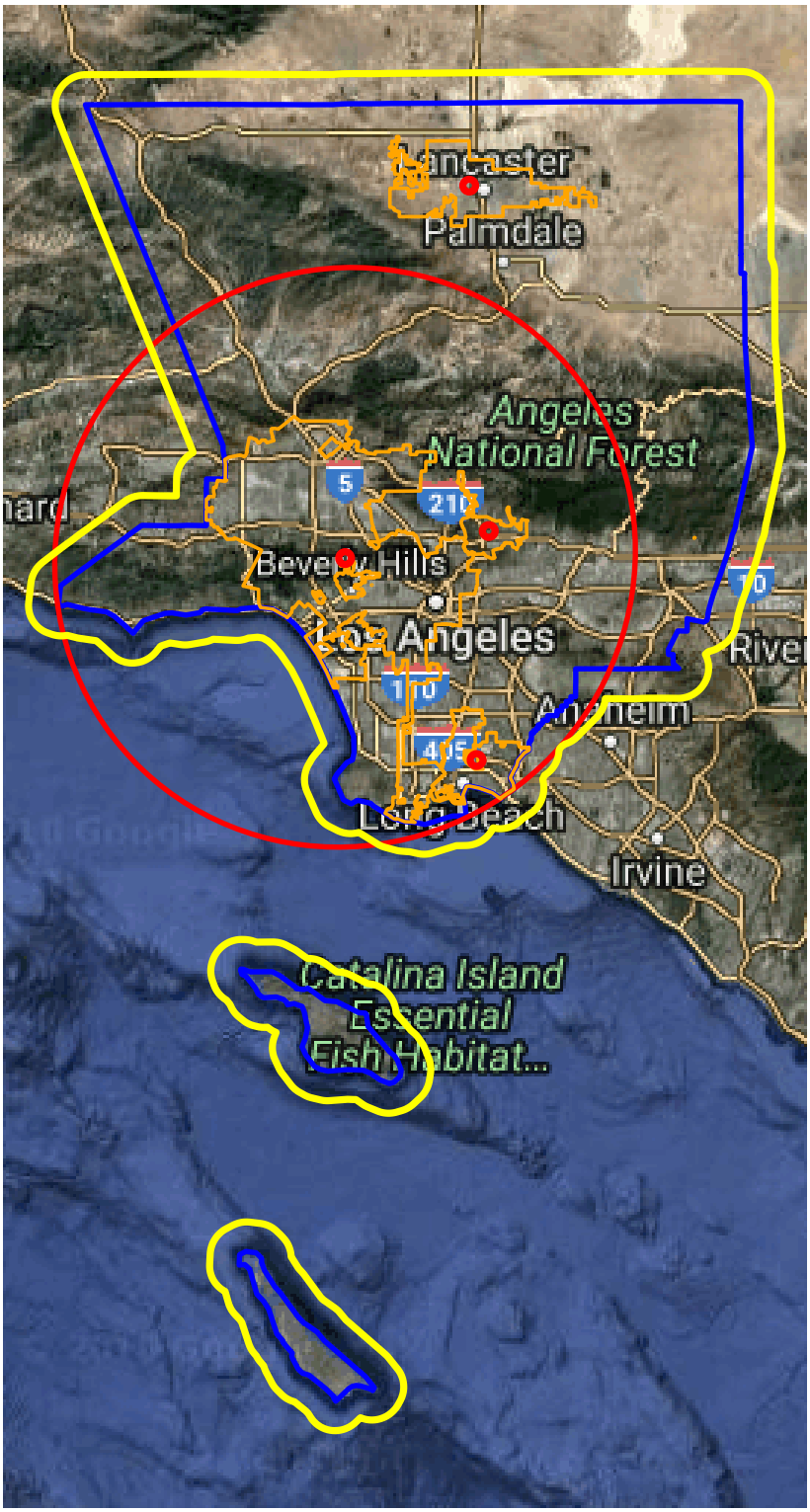
T. M. Sinai. House price moments in boom-bust cycles. 2012.

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.

N. E. Wallace. The market effects of zoning undeveloped land: Does zoning follow the market? *Journal of Urban Economics*, 23(3):307–326, 1988.

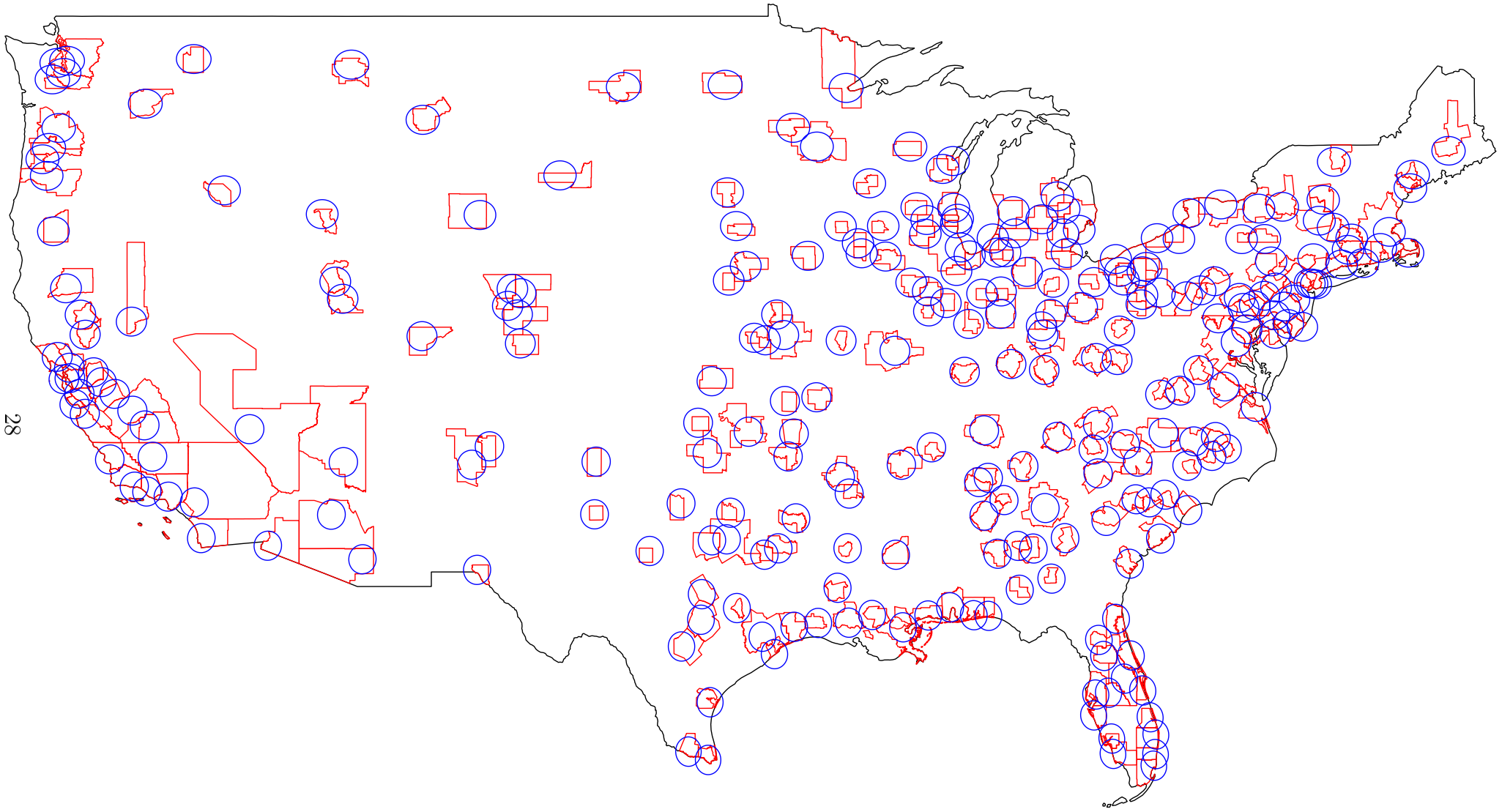
A Tables & Figures

Figure 1: Saiz and Buffered Land Unavailabilty Coverage for the Los Angeles MSA



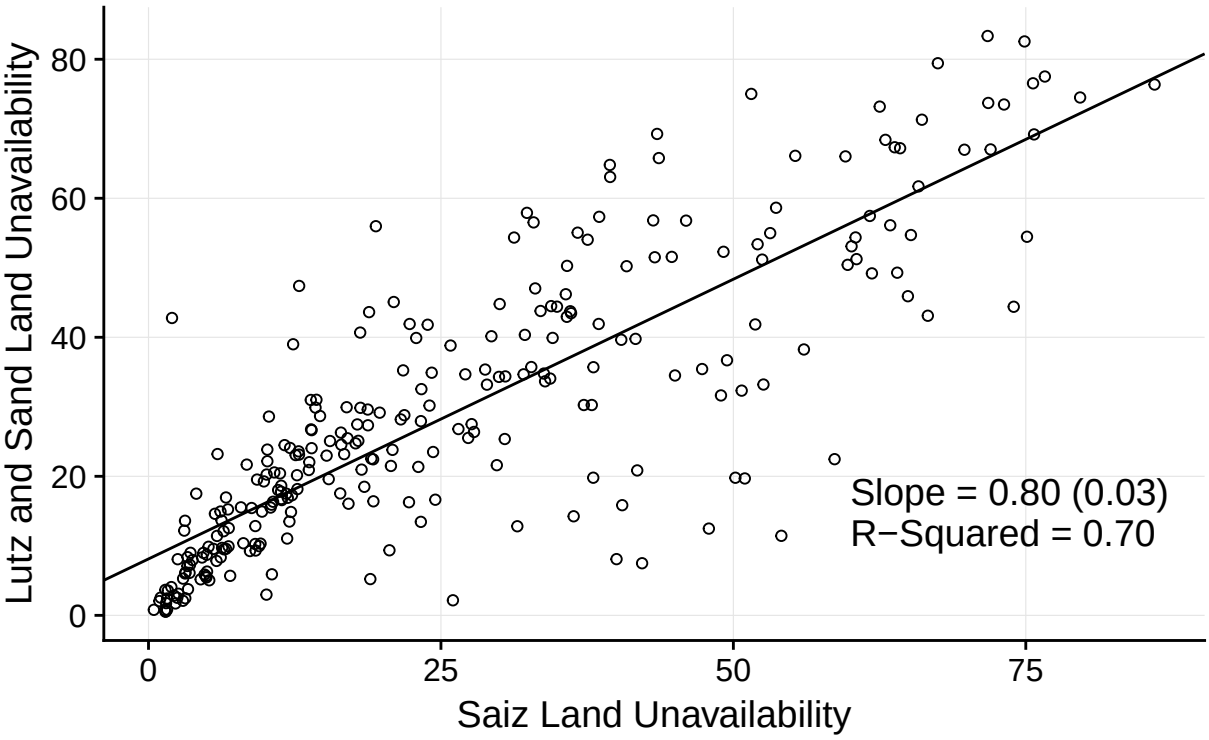
Notes: The blue lines represent the MSA polygons for the Los Angeles-Long Beach MSA. The orange lines signify the central cities within the Los Angeles MSA and the red dots are the centroids for the central cities. The red circle is has a radius of 50 kilometers and is centered around polygon centroid for the first Los Angeles central city (Los Angeles). 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 2: Saiz Land Unavailability Coverage Across US MSAs



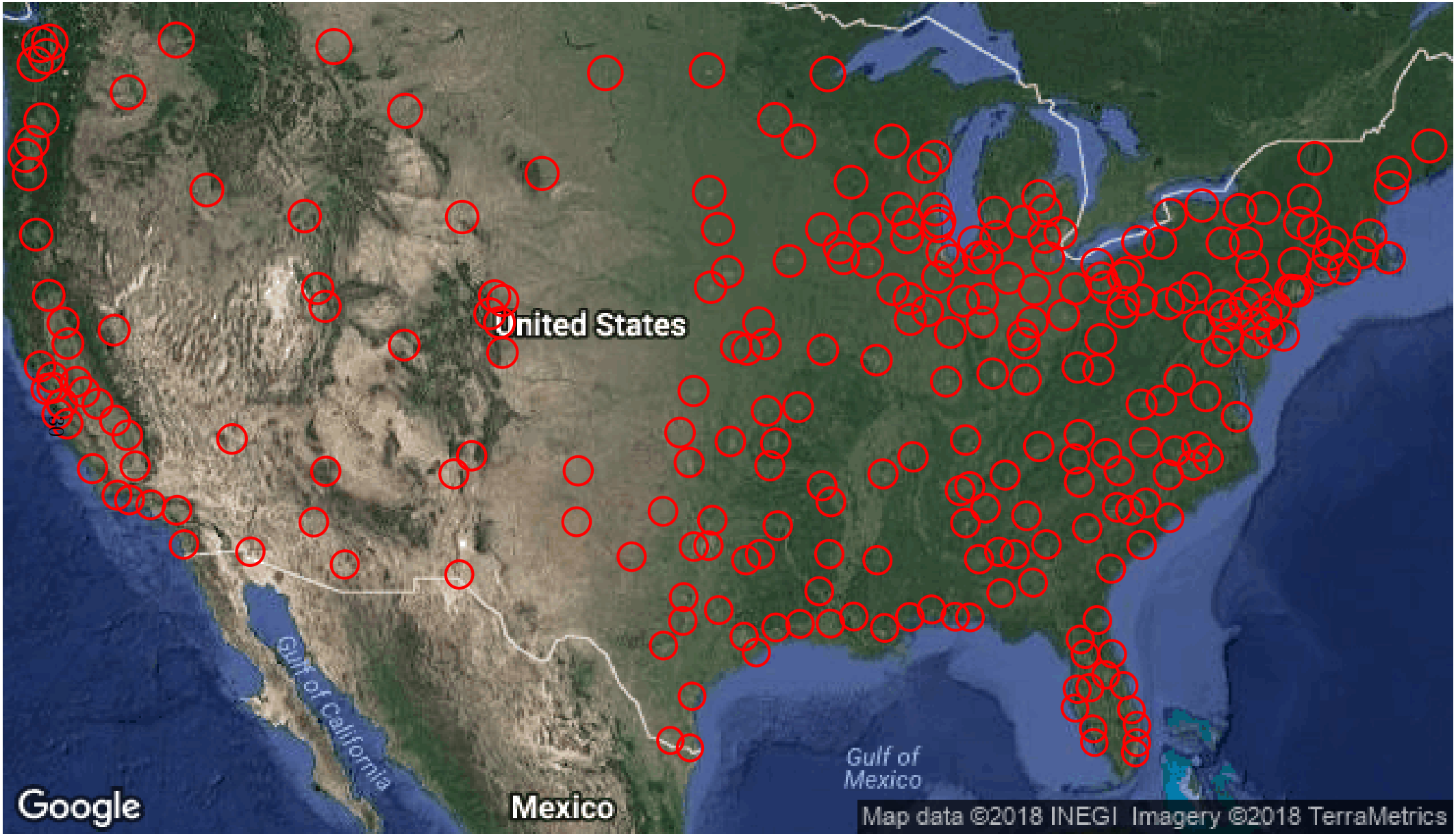
Notes: The blue lines represent MSA polygons. The red circles have a radius of 50 kilometers and are centered around first central city polygon centroid.

Figure 3: Comparison of Land Unavailability Measures



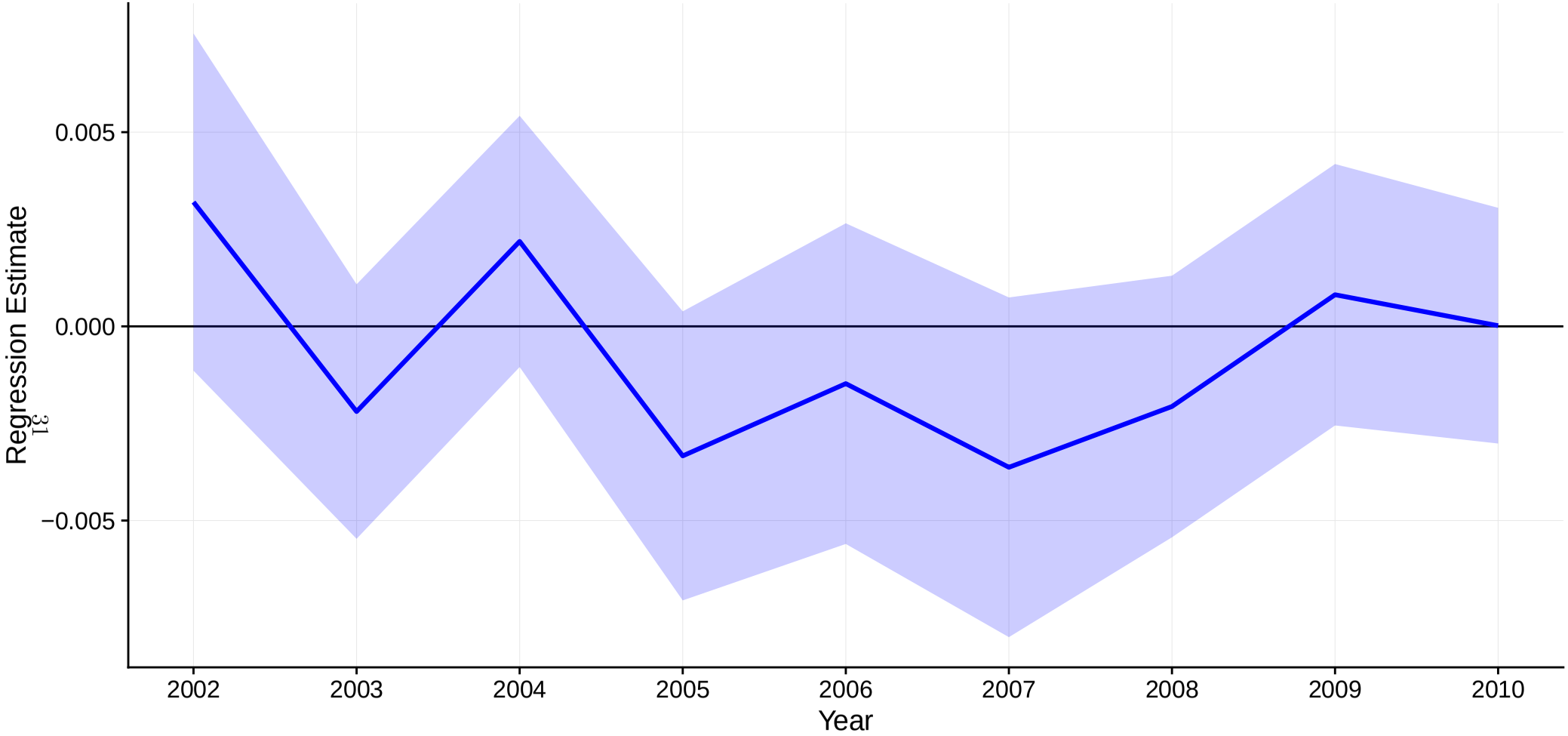
Notes: The Saiz (2010) proxy for the percentage of unavailable land is on the horizontal axis; the vertical axis shows the measure of land unavailability constructed in this paper. Points correspond to MSAs.

Figure 4: 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.

Figure 5: 2000s Annual Correlations between Bartik Shocks and Land Unavailability



Notes: Correlations between Land Unavailability and (Bartik, 1991) Labor Demand Shocks in the 2000s from the model $Bartik_{it} = \alpha + \delta_t + \sum_{y=2002}^{2010} \theta_y LandUnavailability_i \mathbf{1}\{y = t\} + \varepsilon_{it}$. The blue line is θ_y, the correlation between Land Unavailability and the Bartik Labor Demand Shock for year y, and the shaded bands are ± 2 standard errors. Standard errors are clustered at the county level and the regression is weighted by the number of households in 2000.

Figure 6: Florida 2001 LandCover Dataset

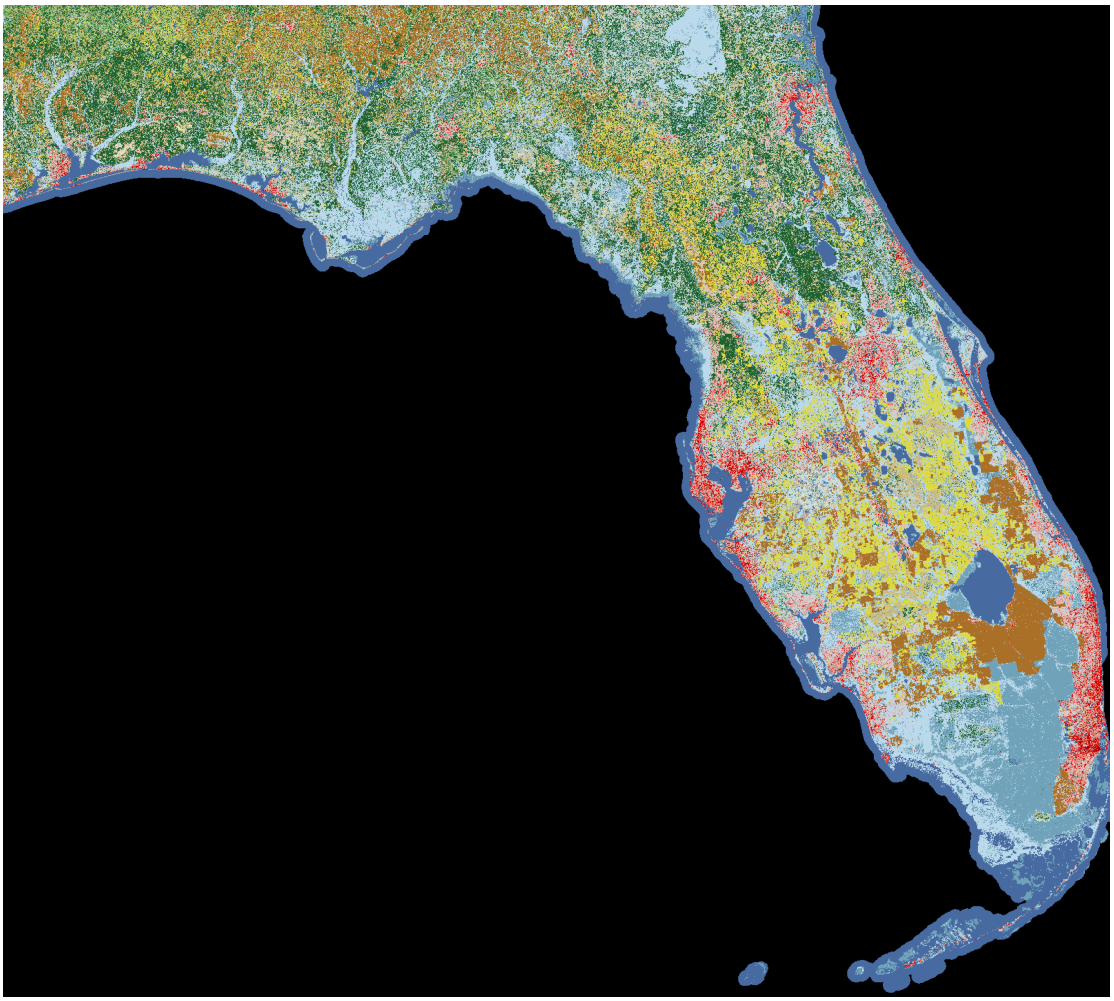
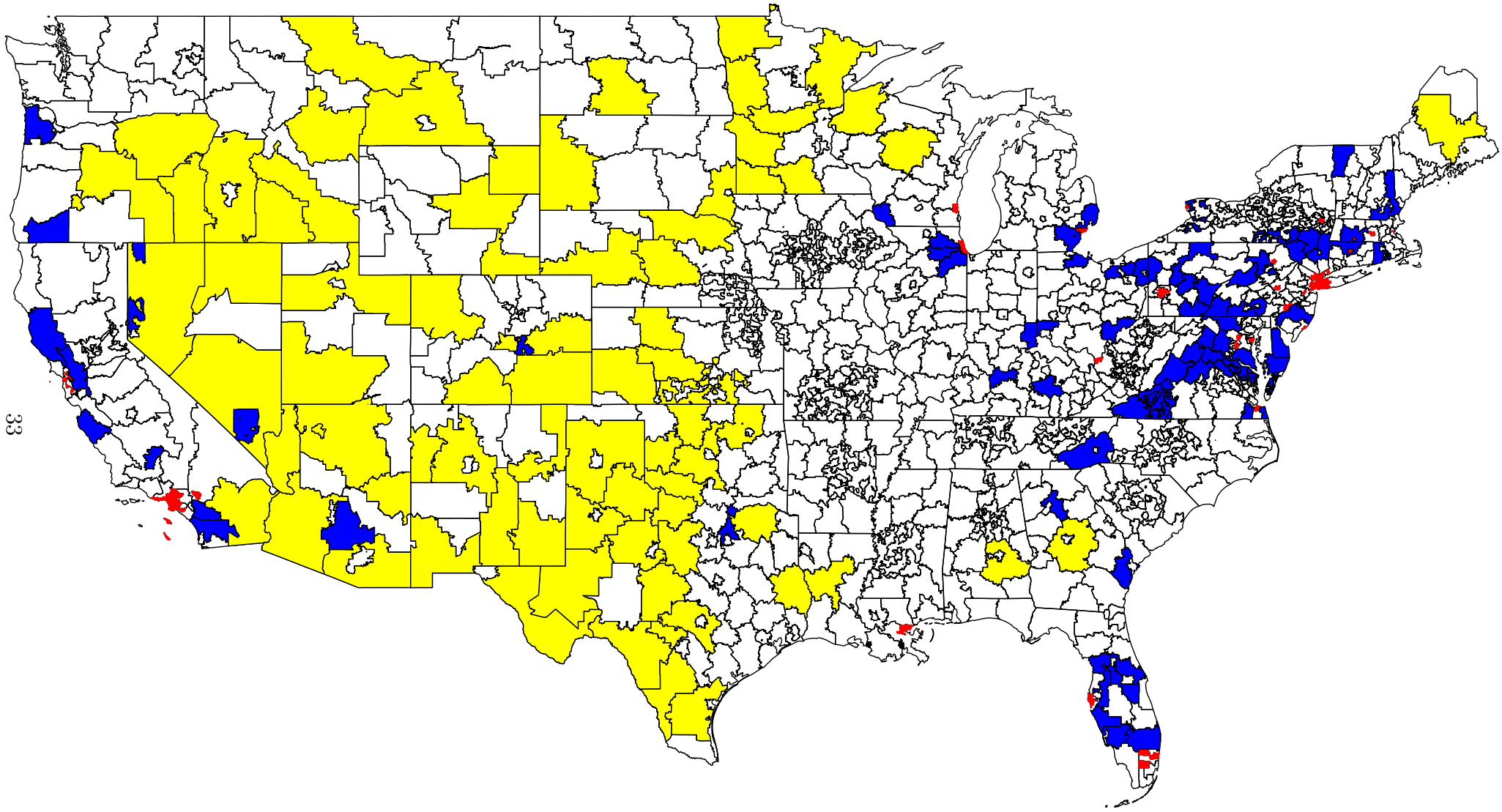


Figure 7: Three Digit Zip Codes and Buildable Land Deciles, 1, 5, and 10



Notes: Three digit zip codes. Red areas are Buildable Land Decile 1; Blue areas are Buildable Land Decile 5; and yellow areas are Buildable Land Decile 10.

Table 1: Zip and Zip3 Land Unavailability Correlations with Housing Demand

	Dependent variable:			
	Amenities Index	College Share in 2000	Foreign Share in 2000	Housing Density in 2000
	(1)	(2)	(3)	(4)
Panel A: Zip Code				
Land Unavailability	-0.018^{***}	0.006	-0.120^{***}	-0.434^{***}
	(0.002)	(0.019)	(0.020)	(0.133)
Constant	0.692^{***}	24.272^{***}	12.711^{***}	27.309^{***}
	(0.082)	(0.625)	(1.218)	(6.912)
Observations	13,044	30,606	30,614	30,645
R^2	0.081	0.0001	0.039	0.018
Panel B: Zip3				
Land Unavailability	-0.005^{**}	-0.009	0.053	-0.008
	(0.002)	(0.025)	(0.038)	(0.071)
Constant	0.422^{***}	24.391^{***}	8.012^{***}	18.694^{***}
	(0.119)	(0.918)	(1.399)	(5.250)
Observations	727	832	840	867
R^2	0.014	0.0004	0.012	0.00001

Notes: Zip code and three-digit zip code (Zip3) regressions of housing demand proxies on Land Uanvailability. Robust standard errors are clustered at the county level (panel A) or the commuting zone level (panel B). Regressions are weighted by the number of households in 2000.

Table 2: Predicting House Price Growth with Land Unavailability – Average MSEs

Prediction Model	Saiz LU Only (1)	County LU Only (2)	LU All Geog. (3)	LU All Geog. + Components (4)	LU All Geog. + Components + Squares (5)
Panel A: 2002 - 2006					
OLS	373.46	414.81	386.25	355.32	335.13
	(359.05, 387.88)	(397.66, 431.96)	(369.70, 402.81)	(338.37, 372.27)	(319.63, 350.63)
Lasso	373.46	414.81	386.75	353.44	353.47
	(359.05, 387.88)	(397.66, 431.96)	(370.17, 403.33)	(338.07, 368.81)	(338.08, 368.86)
Random Forest	234.21	523.67	293.25	207.43	199.74
	(220.48, 247.94)	(504.72, 542.62)	(281.59, 304.91)	(196.65, 218.20)	(189.25, 210.23)
Panel B: 2006 - 2009					
OLS	285.56	304.01	297.82	287.22	263.17
	(268.38, 302.74)	(285.11, 322.91)	(280.37, 315.26)	(270.33, 304.11)	(247.97, 278.38)
Lasso	285.56	304.01	300.82	291.11	291.08
	(268.38, 302.74)	(285.11, 322.91)	(282.97, 318.67)	(273.86, 308.37)	(273.83, 308.34)
Random Forest	206.86	421.01	241.71	179.32	175.75
	(190.55, 223.17)	(400.22, 441.80)	(227.38, 256.04)	(168.28, 190.37)	(164.94, 186.57)
Panel C: 2011 - 2017					
OLS	273.20	281.34	269.06	268.63	239.46
	(264.66, 281.73)	(272.20, 290.48)	(260.75, 277.36)	(260.22, 277.04)	(232.23, 246.70)
Lasso	273.20	281.34	269.65	269.85	268.42
	(264.66, 281.73)	(272.20, 290.48)	(261.21, 278.08)	(261.37, 278.33)	(259.97, 276.87)
Random Forest	172.05	337.31	212.62	161.62	159.90
	(162.53, 181.58)	(326.94, 347.68)	(205.35, 219.90)	(155.69, 167.55)	(153.89, 165.91)

Notes: Average mean-squared errors of Land Unavailability predictions for house price growth from 10 repeats of 10-fold cross-validation for each time period. Bootstarpped 95 percent confidence intervals are in parentheses. The holdout sample within each fold is constructed only using counties in the Saiz dataset. All available data is used for training for each model. Column (1) uses only the Saiz Land Unavailability proxy (1 predictor); column (2) employs the the county Land Unavailabilty proxy developed in this paper (1 predictor); in column (3) we use total Land Unavailabilty for all geographies at the commuting zone and county levels as well as as the maximum, minimum, and mean Land Unavailability in adjacent (touching) counties (5 predictors); column (4) uses Land Unavailability and all of its components (e.g. water, slope, wetlands, and total Land Unavailability) at all geographies (15 predictors); and column (5) employs all Land Unavailability predictors and their squares (30 predictors).

Table 3: Buildable Land (BL) Summary Statistics by Decile

BL Decile	BL Mean (km^2, 000s)	BL Percent	Corr with Available Land
(1)	(2)	(3)	(4)
1	12.18	0.07	0.46
2	113.06	0.20	0.34
3	355.81	0.33	0.34
4	1013.92	0.41	0.34
5	2058.12	0.48	0.27
6	3516.77	0.58	0.46
7	5084.49	0.61	0.15
8	6883.83	0.67	0.26
9	9535.30	0.71	0.48
10	20188.21	0.75	0.83

Notes: Summary Statistics for Buildable Land (BL) Deciles based on three-digit zip codes. The computation of Buildable Land (BL) for each three digit zip code is described in the text

Table 4: 2002 - 2006 House Price Growth by Buildable Land Decile

	Dependent variable:		
	$\Delta(\ln \text{HP})_{\text{2002-06}}$		
	(1)	(2)	(3)
Buildable Land Decile 1	58.597***	14.518***	13.283**
	(4.887)	(4.611)	(5.322)
Buildable Land Decile 2	35.124***	−8.955***	−9.166***
	(4.717)	(3.276)	(3.484)
Buildable Land Decile 3	27.319***	−16.760***	−18.068***
	(3.164)	(4.628)	(4.419)
Buildable Land Decile 4	42.914***	−1.165	−1.819
	(3.664)	(2.918)	(3.166)
Buildable Land Decile 5	44.079***		
	(4.833)		
Buildable Land Decile 6	29.078***	−15.001***	−13.624***
	(3.060)	(3.781)	(3.755)
Buildable Land Decile 7	21.289***	−22.790***	−20.849***
	(2.356)	(3.971)	(3.906)
Buildable Land Decile 8	23.240***	−20.839***	−20.751***
	(3.066)	(3.798)	(3.742)
Buildable Land Decile 9	24.520***	−19.559***	−19.910***
	(3.819)	(4.662)	(4.667)
Buildable Land Decile 10	25.174***	−18.905***	−22.593***
	(4.171)	(5.802)	(5.862)
Bartik Labor Demand Shock$_{\textit{2002-06}}$			2.074***
			(0.637)
Constant		44.079***	44.481***
		(4.833)	(4.631)
Observations	757	757	757
R^2	0.250	0.250	0.280

Notes: 2002 - 2006 house price growth means by Buildable Land Decile. In column (1) the intercept is exluded and each coefficient represents the mean house price growth for the given Buildable Land decile. The excluded dummy in column (2) is Buildable Land decile 5 and thus coefficients represent the difference in means relative to decile 5. Robust standard errors are clustered at the state level.

Table 5: Housing Net Worth and Non-Tradable Employment Growth During the Great Recession

	Dependent variable:									
	Rest. & Retail	Geog. Concen.	Rest. & Retail	Geog. Concen.	Rest. & Retail	Geog. Concen.	Rest. & Retail	Geog. Concen.	Rest. & Retail	Geog. Concen.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Δ Housing Net Worth, 2006-09	0.174***	0.166***	0.374***	0.208**	0.466*	0.356*	0.515**	0.432**	0.437**	0.372**
	(0.043)	(0.046)	(0.132)	(0.086)	(0.249)	(0.191)	(0.211)	(0.173)	(0.207)	(0.145)
Specification	OLS	OLS	IV	IV	IV	IV	IV	IV	IV Lasso	IV Lasso
Instrument			Elasticity	Elasticity	LU	LU	LU	LU	All LU	All LU
Observations	944	944	540	540	540	540	936	936	936	936
Sample	All	All	Saiz Counties	Saiz Counties	Saiz Counties	Saiz Counties	LU: No AK or HI	LU: No AK or HI	LU: No AK or HI	LU: No AK or HI
R-squared	0.175	0.236								

Notes: County Level Regressions of non-tradable Employment Growth from 2007-2009 on the Housing Net Worth Shock from 2006-2009. Columns (1) - (4) replicate columns (1) - (4) of Mian and Sufi (2014), table III p. 2208. Non-tradable employment is measured using the restaurant and retail industries (Rest. & Retail) or through geographic concentration (Geog. Concen.). See Mian and Sufi (2014) for more details regarding industry definitions. Columns (5) - (6) use the same sample as Mian and Sufi in columns (3) - (4) (counties for which the Saiz elasticity is available), but use Land Unavailabiliy (LU) as an instrument. Columns (7) - (10) expand the sample to use all counties with avialable Land Unavailability and Housing Net Worth data (counties in Alaska and Hawaii are missing as they do not have available Land Unavailability). Columns (9) - (10) use the Rigurous Post-Lasso appraoch (Belloni et al., 2012, 2014; Chernozhukov et al., 2016) to choose relevant instruments from 230 slope, water, wetlands, and total Land Unavailability proxies at various levels of disaggregation, their squares, and their interactions. This approach selects the mean Land Unavailability in adjacent (touching) counties as the only relevant instrument. Regressions are weighted by the number of households in 2000 and robust standard errors clustered at the state level are in parentheses.