# Housing Inequality over a Century

David Albouy[1], Mike Zabek*[2], and Aditya Aladangady[2]

[1] *University of Illinois and NBER*
[2] *Federal Reserve Board*

December 15 2023

Preliminary: numerical values subject to revision. Please do not cite or distribute.

**Abstract**

We document patterns of inequality in housing prices and rents housing since 1930, using Census data. First, housing inequality fell from 1930 to 1970 and rose thereafter, with a sharper fall and gentler rise than income inequality. Second, home-ownership rose considerably when inequality fell, but stayed steady when it grew. Third, the fall in housing inequality was most pronounced at the lower end, as the cheapest units without water or electricity disappeared. Fourth, the post-1970 rise was driven more by owner-occupied than rented units. Fifth, rising inequality is explained more by growing spatial differences across neighborhoods than by changing structural characteristics.

Keywords: Inequality, housing, neighborhoods, consumption, wealth
JEL Numbers: D63, R21

*Zabek: Corresponding author, Mike.Zabek@frb.gov; Alandangady: aditya.aladangady@frb.gov; Albouy: albouy@illinois.edu. Minh Joo Yi and Alok Ranjan provided excellent research assistance. We want to thank Hoyt Bleakley, John Bound, Charles Brown, Jan Brueckner, Gabriel Ehrlich, Raven Molloy, Enrico Moretti, Luigi Pistaferri, Matthew Shapiro, Jeff Smith, and Paul Willen for valuable comments and suggestions. Also, seminar participants at: the Federal Reserve System Committee on Regional Analysis; the Health, History, Demography and Development Conference at the University of Michigan; the Greater Boston Urban and Real Estate Economics Seminar; the AREUEA Annual Meetings; the Census Bureau; the International Institute of Public Finance; and the Urban Economics Association. Mike Zabek acknowledges prior funding from the National Institute on Aging and National Institute of Child Health and Human Development (NICHD) grants to the Population Studies Center at the University of Michigan (R24 HD041028 and P2CHD041028). The authors also acknowledge Ancestry.com for providing underlying data making this research possible and the BigTex High Performance Computing Group at the Federal Reserve Bank of Dallas for providing computational resources. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the NIH, the NICHD, the Board of Governors of the Federal Reserve System, or of any person associated with the Federal Reserve System.

Americans care tremendously about housing: it is the largest expenditure item for most households, and an almost defining component of middle-class wealth. A home and its location not only tell us much about a household's comfort indoors, but also much about the quality of life and opportunities their surroundings provide. Housing also provides access to utilities, such as running water and electricity, which provide incredible consumer surplus, but were not available in many places a century ago. To this day, neighborhood amenities such as safety, school quality, or high-paying jobs vary considerably across places.[1]

In this paper, we consider patterns in inequality in American housing rents and values since 1930 using data from the U.S. Census Bureau. Because housing is such a complex good, this analysis deepens understanding of how a principal component of consumption inequality manifests itself among upper, middle, and working classes. It may also shed additional light on debates over wealth and income inequality over the the last century. From far away, patterns in housing inequality vaguely track well-documented U-shaped patterns seen in income, wealth, and consumption inequality.[2]. In our analysis below, we highlight several patterns that seem worth noting.

First, the decline in housing inequality is more pronounced during the "Great Compression" of inequality prior to 1970, than during the later "Great Divergence" afterwards. Following the Great Depression in the 1930s, inequality in home prices fell between 1940 and 1970 at a rate faster than income inequality; thereafter it rose at a rate slower than for income.

Note, that the compression of home values, concentrated from 1940 to 1960, coincided with a well-documented 20 percentage-point rise in home-ownership (Fetter, 2016).[3] Yet, while years after 1970 saw home prices diverge, they did not see a symmetric decline in home-ownership. This lead to our second observation, that while the value of owner-occupied homes are roughly as unequal today as they were during the Great Depression, the share of home-owners is roughly half as large.

Third, during the Great Compression, inequality in rents fell even more strongly than housing

[1]Housing accounts for roughly one third of total household wealth and roughly 40 percent of the capital stock. Housing is almost two thirds of wealth for the middle three quintiles (Wolff, 2014). In 2004, 62 percent of housing wealth was held by the bottom 90 percent, and only 9.8 percent by the top 1 percent; for stocks and mutual funds, the comparable numbers are 14.6 and 44.8 percent (Wolff 2009, p. 160).

[2]Some particularly influential reviews include: Saez and Zucman (2020), Kopczuk (2015), and Aguiar and Bils (2015). Goldin and Margo (1992) highlight and examine the Great Compression in incomes in the 1930s and 1940s. Piketty and Saez (2003) examine that of top incomes. Kopczuk (2015) reviews some of the debate on measuring wealth inequality, as well as whether it has increased by as much as income inequality. E.g. Kopczuk and Saez (2004) find a great compression of wealth inequality but little divergence afterwards. Saez and Zucman (2016) find greater divergence, and Bricker et al. (2016) show something roughly in between.

[3]See also of the Census and Devaney (1994)

values. This decline was especially pronounced at the lower end of the distribution, where units with the very lowest rents disappeared. While the Census Bureau destroyed relevant micro-data on housing quality, surviving county-level records point to differences in access to public infrastructure. In 1940, almost a third of housing units were not connected to running water, while a quarter did not have electricity; by 1970 such units were rare. Rents rose the most in the largely rural counties where running water and electricity went from rare to ubiquitous.

Fourth, the rise in housing inequality post 1970 was much stronger for owner-occupied units than for rented units. Households at the bottom of the income distribution did not see their rents, and fell away from the median, rental expenditures did not follow suit. Almost all rental units units continued to provide basic comforts that were not guaranteed decades before. At the same time, rents have taken up a larger share of renters' incomes. Meanwhile, the value of homes diverged sharply.

The final patterns are those most related to the rise in housing inequality over the last fifty years. While Census measures of structural characteristics are limited, a standard re-weighting procedure suggests finds that changes in observable structural characteristics explain not even a little of the increases in home values. Put simply, houses don't seem to have become more unequal in size. Rather, inequality growth is seen more in the diverging value of neighborhoods themselves, as inequality between neighborhoods appears to have grown more strongly than inequality within them., complementing growing trends in income segregation (Massey et al., 2003; Reardon and Bischoff, 2011). Thus it appears that the growth in consumption inequality seen in Aguiar and Bils (2015) may manifest itself in the public goods different neighborhoods provide.

## 1 Objectives and Challenges

Like with other measures of expenditures, we do not observe quantities of housing consumed separately from its price. This lack of separation is common to inequality comparisons over time. Yet with housing the price of housing also varies considerably over space. Housing is especially heterogeneous, both in the characteristics of the dwelling itself and in its location. The particular value of those characteristics are often priced using hedonic methods, e.g. Lancaster, Rosen. Insofar as the price of dwelling characteristics are correlated, one may view these in terms of a composite commodity of "housing services," e.g. Muth (1969). This requires households to be indifferent across units costing the same rent within a location. It takes an additional leap to reduce locational characteristics into uniform "location services." which would require indifference across locations. Nonetheless, we can use observable locations to get some sense of inequality in location versus dwelling services. In later samples, we can use use dwelling characteristics to see how much inequality is explained by observable dwelling characteristics versus what may be due to changes in prices of those services.[4]

A second complication arises in our ability to measure housing consumption over time. Rents, in principle, offer a spot-price measure for dwelling and location services. Naturally, various frictions may mean that some households receive more for their rent payments. Less obvious is how to deal with owner-occupied homes. Here too, higher-price homes should reflect greater housing and location services. Nevertheless, a standard user-cost analysis implies that price-to-rent ratios are larger for units where greater rent appreciation is expected. Unfortunately, there is no unanimously agreed-upon method to estimate these expected user costs. Thus, we instead use a fixed percentage of housing prices as a rental equivalent and keep in mind an important caveat. Namely, the housing inequality we observe through prices has much to do with expected future conditions.

A great feature of the Census data is that it tells us whether a home is owned or rented, however it is extremely limited in telling us about home equity. Some samples do indicate whether a home is mortgaged or owned "free and clear," i.e. who has 100 percent equity. It also seems safe to assume that most renters have zero equity in their homes. However, the Census does not tell us who owns their homes. Rented and owned homes therefore have much of their equity in the hands of unidentified landlords and lenders, including the public at large through banks and other institutions. Thus our analysis can only provide us with a limited perspective on wealth inequality through the home-ownership rate, the free-and-clear rate, and an upper bound on home equity that households have.

.

[4]By providing access to employment, a home, like a car, may provide access to more gainful employment, and thus be partly a means of generating income, rather than providing pure consumption value. Our results on the spatial composition of housing inequality also mirror prior examinations of income inequality across neighborhoods by Bischoff and Reardon (2014) and Logan et al. (2020). In particular, both income and housing inequality have risen in large part because of differences across neighborhoods.

# Data

## Samples and Variables

Our data are drawn primarily from households in the Decennial Censuses from 1930 to 2000, and the American Community Survey (ACS) thereafter via Ruggles et al. (2010). These include self reports of rents, housing values, and other dwelling characteristics[5] To maximize comparability, we limit our analysis to the continental United States, and exclude farms and homes used for commercial activities, as well as condominiums. Appendix A provides more details, as well as analyses showing that are results are largely unaffected by these sample restrictions.

Dwelling characteristics include number of rooms and bedrooms, building age, plumbing, and what kind of heating system is installed. Locations are available by state, and may be assigned to 722 commuting zones with considerable accuracy, except in 1960.[6]

For greater geographic detail, we use summary tables for counties and census tracts going back to 1940.[7] These observations are harmonized to correspond to 2010 vintage counties and census tracts.[8]

## Interpolation and Normalization

In many years housing values and rents are only reported in discrete intervals, and thus use standard techniques to to infer continuous values more comparable over time.[9] These assume a smooth Pareto

[5]Renters from 1970 onward and home owners from 1980 onward also report their utility costs, which we include to increase the comparability of our measures of housing expenditures and rents across units.

[6]We use 1990 commuting zone definitions, which are more fully explained in Tolbert and Sizer (1996). They are made up of counties and designed to be places where people both live and work. For 1930 and 1940 we are able to assign every house to one commuting zone. In 1960 they are not available in public IPUMS samples. From 1970 onward some houses are identified as being in PUMA's or county groups in the IPUMS data that cut across multiple commuting zones. For these years we follow Dorn (2009), Autor, Dorn and Hanson (2013), and others in probabilistically assigning houses to commuting zones based on the proportion of responses in their most detailed geographic category that were in one commuting zone or the other. For most of our specifications we also report statistics using states. We are able to match each house to a state for each year.

[7]The data are compiled by NHGIS (Manson et al. (2020)) and several of the early files were digitized from Census publications by Elizabeth Mullen Bogue.

[8]The geographic harmonization follows the procedure used by Logan, Xu and Stults (2014) since 1970, and the extension by Lee and Lin (2017) for years prior to 1970.As the U.S. was only fully tracted starting in 1980, We vary the samples to discuss how this could impacts our main findings. Our preferred sample consists of neighborhoods that existed in 1940. Since these results are based on summary tables, there are slight differences in the universe of dwellings included in each year. However, Appendix A provides a robustness exercise, using the micro-data sample, that shows that the main results are robust to substantial changes in the sample in each year. These differences are also much smaller than differences due to changes in the universe of census tracts over the period.

[9]Data derived from summary tables are always intervaled, while data from the main micro data sample sometimes are not. Home values are recorded as intervals from 1960 through 2000 and monthly rents are reported as intervals

distribution within each interval, meaning the log survivor function (the log of one minus the CDF) is log linear within each interval, allowing them to vary by year and geography. [10] Appendix A.3 provides more detail, including the specific formulae.

While we show inequality in housing values and rents separately, we also sometimes convert housing values into a rental equivalent. To do so, we use a multiple of 0.0785 times the value of the house, plus utility payments when they are available. In principle, this figure includes foregone interest, depreciation, and property taxes, minus price appreciation. While user costs do vary over space and time, as mentioned earlier, creating such measures require data unavailable in many years and requires making debatable assumptions about how to treat expected capital gains and owners' valuations for local public goods.[11] Appendix B gives evidence that user costs have do not vary dramatically over extended periods of time, as rent-to-price ratios have not varied tremendously.[12]

Table 1 provides basic characteristics of the sample. It is worth noting that our basic unit,the household feel in size consistently, from 3.8 in 1930, to 3.3 in 1960, to 2.5 in 2010. Meanwhile, the number of rooms per house rose slowly after 1960 (from 5 to 5.75) along with the number of bedrooms.

The first two panels of 1 give important statistics when comparing rents, housing prices, and incomes. Panel A shows that how home-ownership, after a small dip in 1940, tremendously between up until 1960, largely stabilizing thereafter with the notable blip upwards during the Great Housing Bubble.[13]

The aggregate rent-to-price ratio saw a nearly opposite but less dramatic pattern, rising during the Great Depression, and thereafter falling. The drop in the mean ratio is greater than the median, which is indicative of the inequality measures we consider later. The ratio falls to a level of 4 to 5

for 1960, 1980, and 1990. In all other years the micro data sample contains a continuous measure with relatively little top coding. Note that for consistency, we also apply the procedure when the values are provided continuously, to provide a similar level of smoothing.

[10]Thus, data at the neighborhood level contains a different Pareto distribution for each bin within each neighborhood in each year. Fitting distributions in such small areas can lead to issues with interpolations of top bins in a few neighborhoods national distributions. But our results are robust to controlling for these issues in a number of ways.

[11]This commonly-used value is based on Peiser and Smith (1985). See Albouy and Hanson (2014) and Bishop et al. (2023) for examples using recent data. Moreover, there is little data to calculate year-by-commuting zone user costs prior to 1980, and even then, conceptual issues do arise, particularly with treatment of capital gains, ex post or ex ante. See also Verbrugge (2008) and Poterba and Sinai (2008).

[12]The scale invariance of the inequality measures also implies that mismeasurement of the constant will have no effect on inequality among home owners, though it would affect inequality between owners and renters.

[13]Note that Layton (2021) gives a longer and more continuous measure of home-ownership consistent with our findings. The period prior to 1930 was fairly stable over time.

Table 1: Summary statistics

Panel A: Means and medians

| | 1930 | 1940 | 1960 | 1980 | 2000 | 2019 |
|---|---|---|---|---|---|---|
| Sample size | 1,134,969.00 | 1,420,810.00 | 451,066.00 | 3,589,168.00 | 4,749,600.00 | 5,806,386.00 |
| Owner occupied | 0.45 | 0.41 | 0.61 | 0.64 | 0.65 | 0.63 |
| Home value median in thousands | 53.95 | 40.79 | 74.50 | 113.64 | 130.09 | 175.81 |
| Monthly rent median | 308.30 | 293.52 | 361.75 | 472.91 | 608.39 | 754.42 |
| Monthly rental equivalent median | 321.98 | 283.80 | 544.07 | 774.33 | 906.34 | 1,141.89 |
| Annual household wages median | . | 14.32 | 28.97 | 31.68 | 37.39 | 36.90 |
| Household size | 3.85 | 3.46 | 3.27 | 2.75 | 2.60 | 2.51 |
| Number of rooms | . | . | 4.92 | 5.31 | 5.45 | 5.75 |
| Plumbing facilities | . | . | 0.88 | 0.99 | 0.99 | 1.00 |

Panel B: Percentiles

| | 1930 | 1940 | 1960 | 1980 | 2000 | 2019 |
|---|---|---|---|---|---|---|
| Equivalent rent 10th percentile | 84.45 | 62.57 | 213.76 | 372.61 | 404.99 | 492.05 |
| Equivalent rent median | 321.98 | 283.80 | 544.07 | 774.33 | 906.34 | 1,141.89 |
| Equivalent rent 90th percentile | 802.85 | 661.03 | 1,051.65 | 1,573.36 | 2,082.08 | 2,854.03 |
| | | | | | | |
| Ratios | | | | | | |
| Equivalent rent 50 10 ratio | 3.81 | 4.54 | 2.55 | 2.08 | 2.24 | 2.32 |
| Equivalent rent 90 50 ratio | 2.49 | 2.33 | 1.93 | 2.03 | 2.30 | 2.50 |

NOTE: Households have grown larger in terms of people, rooms, median rents, median home values, and median rental equivalents. Ownership rates grew substantially from 1940 to 1960, at the same time as the ratio of the 10th and 50th percentiles of rental equivalents narrowed substantially. This table displays a series of descriptive statistics for the basic analysis sample in the Census and CPS IPUMS sample, which includes renters and owners. Blank cells indicate that the relevant statistic is not available. Panel A shows means and medians while Panel B shows percentiles of the rental equivalents. Dollar values are in 2007 dollars – deflated by the US Consumer Price Index excluding structures.

percent, which is a few percentage points lower than the number we use for turning housing values into rental equivalents. Since we are not comparing identical units, this difference may be due to rental properties simply providing less services than owned properties. Compositional issues aside, the facts that the ratio of owned to rented units grew, while their relative price also grew, suggests that the relative demand for owned units grew considerably between 1940 and 1960. It also grew briefly during the Housing Bubble.

Panel B displays household-level medians of how rents and housing prices have compared to rents. These numbers begin in 1940 when wage incomes become available. In both series, we see a bit of a trend downwards between 1940 and 1970, and a more pronounced trends upwards after 1970, with a respite in the 1980s and 1990s. The upward trend for housing prices to incomes is considerably stronger (note the relative equivalence of both axes), even after accounting for homeowners higher incomes.

## Empirical Methodology and Results

### General Trends over Time

In the following three panels of figure 1 we examine trends in housing inequality from 1930 onwards. To describe inequality in housing expenditures we use several scale and population invariant measures: the share accounted for by the top 10 percent, the Gini coefficient, and the Theil entropy index. Both the Gini coefficient and Theil better capture changes in the middle of the distribution, with the Theil being somewhat more sensitive to inequality at the bottom (Atkinson (1970))[14]

All three series describing inequality in housing values in Panel C show rather clear and almost

[14]The Gini coefficient is a measure of inequality that computes the area above the Lorenz Curve – the cumulative share of housing held by each quantile of the housing distribution. The Theil index is computed by taking a weighted average of the log of the (normalized) expenditure on housing by each household, and is given by:

$$T = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{y_i}{\bar{y}}\ln\left(\frac{y_i}{\bar{y}}\right)\right]$$

where $\bar{y} \equiv E[y_i]$. The Theil index satisfies the principle of transfers, so a transfer from a rich to a poor person always decreases inequality. Note that if was a monotonic relationship between income, $m$, and housing given by the income elasticity $\epsilon \equiv d\ln y/d\ln m$, then $y_i/\bar{y} = (m_i/\bar{m})^{\epsilon}$, and

$$T = \epsilon\frac{1}{N}\sum_{i=1}^{N}\left[\left(\frac{m_i}{\bar{m}}\right)^{\epsilon}\ln\left(\frac{m_i}{\bar{m}}\right)\right]$$

symmetric U-shaped patterns, with the bottoms centered at 1970. The Gini and Theil coefficients are roughly the same from 1980 onwards as they were from 1940 and before, although the Theil measure drops considerably more. The top 10 percent share falls from 34 percent in 1940 to 27 percent in 1970, before rising to 37 percent in 1990, falling slightly thereafter. The slight blips upwards do coincide with the Savings and Loan Crisis seen in 1990 and the Housing Bubble in 2009 (ACS years 2005 to 2009).

The series for rents in Panel D begin at roughly similar levels as the measures for inequality (note the equivalent axes), although the drop in inequality is greater, while the rise in 1990 is rather slight. Post 1960, the smaller number of renting households show lower and stable levels of inequality. One explanation is that owner-occupied housing is more affected by dispersion in the price of land, since it is more land-intensive than rental housing typically in multifamily structures. (Yao (2019)).[15] Another possibility is that shifts in ownership rates changed the composition of owners and renters, masking rising inequality in rents. In fact, we show in Appendix B that ownership rates rose among middle-class households and actually fell at lower values of equivalent rents, concentrating renters at lower rental rates. Moreover, shifts in housing expenditures are more pronounced in the upper tail of the distribution with few renters. As a result, composition changes seem to have held down inequality in rents.

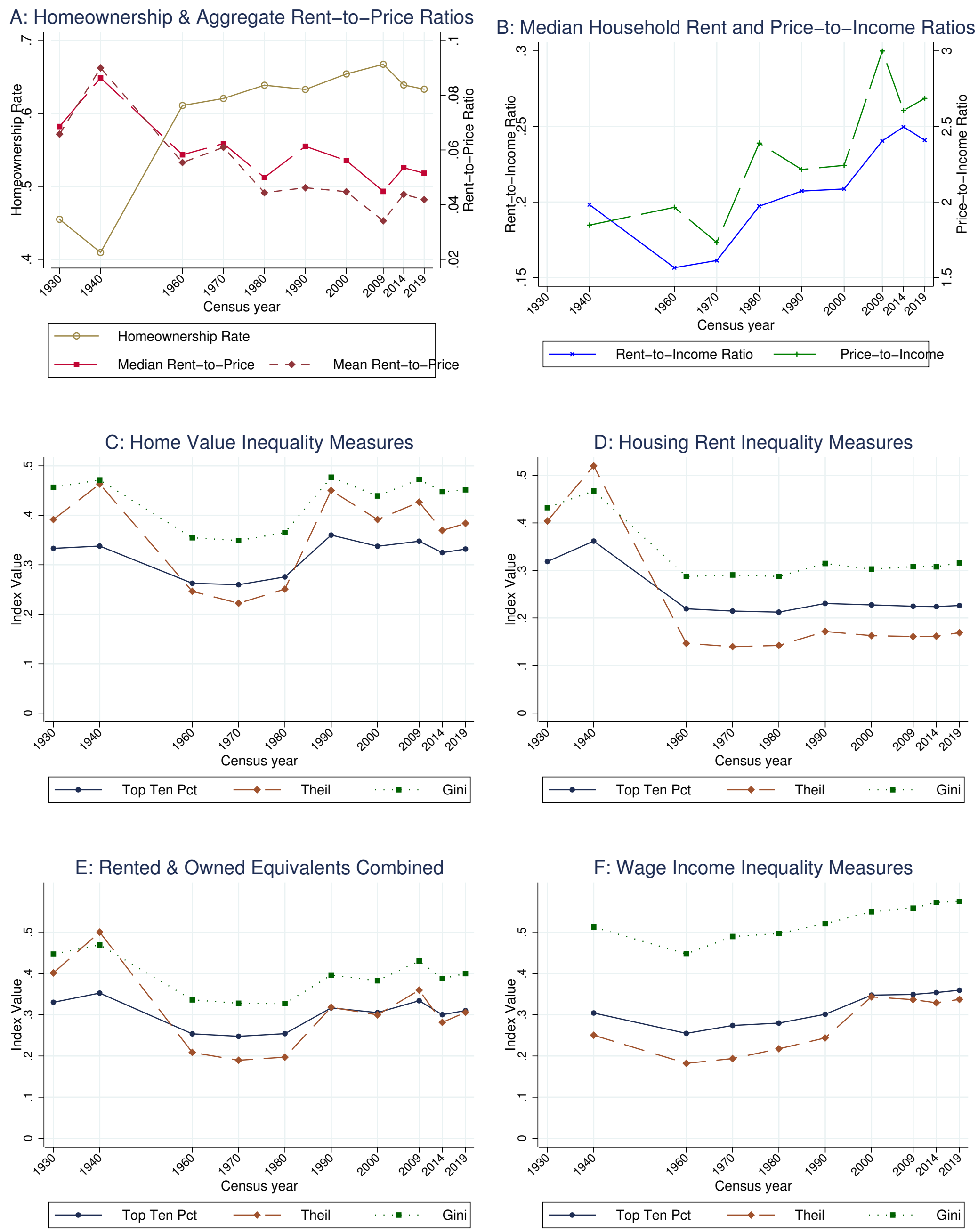
Panel E of Table 1 gets around these composition shifts by combining owners and renters using the rental equivalents described earlier. It reveals a U-shaped pattern, that looks much like an average of Panels A and B, despite shifts in home-ownership: the rise in mid-century inequality is in between that of rents and home values. However, later housing inequality is not as great as in the pre-War era.. [16]

Panel F presents inequality in household income using Census data. (Goldin and Margo (1992) cover decades up to 1990). The data that go as far back as 1940 only cover wages. What is is perhaps

[15] Appendix Figure A.12 also shows that inequality in rents also changed more at the top than at the bottom. Less rent dispersion at the lower end could reflect increased prevalence of market rate housing assistance, or enduring effects of rent control following World War II (Baum-Snow and Pavan (2013)).

[16] One potential concern is that shifts in household size and composition may have led to changes in housing services even if per-capita housing services do not exhibit rising inequality. In the appendix, we show that rental-equivalent adjusted for OECD adult-equivalent scale shows a similar pattern to that seen in Panel E. Another concern is that rising inequality in house prices is actually a product of rising inequality in the user cost of housing, even as rental equivalents have been stable. In Appendix B, we show that the distribution of price-rent ratios in the Consumer Expenditures Survey appear relatively stable over recent history, and flat rent inequality observed in Panel D is largely driven by selection.

Figure 1: Trends in Housing Inequality from 1930 to 2019




NOTE: Housing inequality had a U shape over the past 100 years. Overall inequality decreased following the Second World War, and inequality in the value of owner occupied houses has been increasing since 1970. The four panels in this figure display values of the a Theil inequality index and a Gini coefficient for the value of owner-occupied houses, reported gross rents among renters, a consumption equivalent measure that collapses house values and rents, and household wage income as recorded in the Decennial Census and ACS. The observation for 2012 comes from ACS for 2009-2012. See text and Data Appendix for more detail on data manipulations and sample construction.

most notweorthy here is that inequality after 1970 increases more gradually and consistently than our housing measures, to levels greater than we see in 1940. Thus, while wage income inequality appears to be greater now than in 1940, housing inequality appears to be slightly lower. This suggests that the link between wage inequality and housing inequality may be weaker than before.[17]

Compared to 1940, housing inequality is lower in the post-1980 period because of lower inequality among renters, despite the fact that the proportion of renters is lower. The larger reduction of inequality after 1940 among renters suggests a greater reduction in inequality at the bottom of the distribution. This is reinforced by the greater change in the Theil index, which is more sensitive to lower values, as well as the 50-10 ratio shown in Table 1.

### County-level changes during the Great Compression and Great Divergence

While the changing trends in inequality roughly coincide with similar findings in the inequality literature, they also coincide with two trends in urbanization noted by Boustan et al. (2018). The period from 1940 to 1980 saw declining urban wage and rent premia, despite the rate of urbanization growing from 57 percent to 74 percent. After 1980, those premia began rising again. while the rate of urbanization slowed. Nationwide, much of the urbanization came in the South, and sub-rbanization accounted for most urban growth.

In this section we examine housing changes across counties during periods of convergence, (1940 to 1970), and divergence (from 1970 to 2019). Figure 2 shows four related plots of how counties changed by their initial rental equivalent. While these growth patterns do not map neatly to the inequality data, they highlight important contrasting trends that likely contribute to it. Because of the incredible variation, all the graphs are in logarithmic scales.

Panel A examines how housing expenditure change by initial levels. Between 1940 and 1970, we see expenditures rise most in areas where they were once lowest. After 1970, we see this pattern reverse, with high expenditure counties seeing greater growth. Panel B shows a very similar set of patterns when we use initial density rather than rent. Thus, we see homes in the poorest, least expensive

[17]While we do not interpret our results as providing a causal estimate of the relation between housing and income, the temporal and cross-sectional correlations between housing and income in our data appear consistent with estimates of the income elasticity of housing demand in the literature at between 0.6 and 0.9. See Appendix D for additional details.

Growth patterns in the number of households, shown in Panel C, show an opposite pattern. From 1940 to 1970 households where expenditures are already high grew tremendously. After 1970, this pattern halts and reverse slightly. Thus, early on, many moved out of the least-valued homes into more valued spots. This trend then halted, as expensive areas become less accessible.

Panel D examines the role of access to public infrastructure, namely through running water. Not shown here is how strongly correlated this is with access to electricity. In 1940, there were still many counties without access to running water. That largely disappeared by 1970, after which there were fewer gains to be made. It appears that much of the compression at the lower end of the housing distribution is due to housing improvements in running water and electricity, considered by most as vital necessities today.

## Neighborhoods, Local Labor Markets, and Dwelling Characteristics

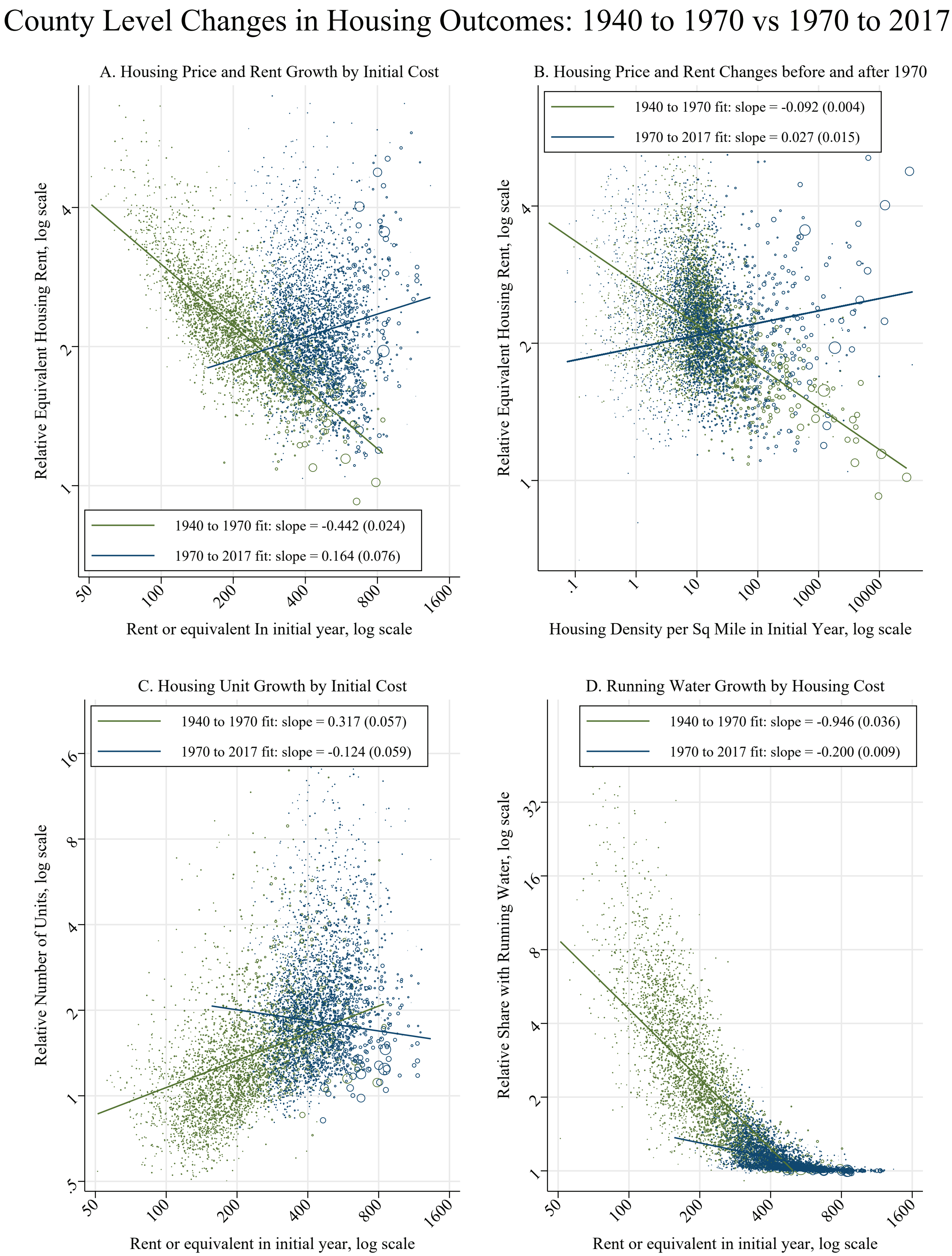
Although the share of income devoted to housing appears to have stayed relatively constant over the 20th century, patterns of housing consumption appears to have changed considerably. As pointed out in Table 1, while households shrank in human size, the units the live in have grown. They have have also moved to warmer and sunnier places in the South and West.

Below we use a non-parametric re-weighting procedure to account for changes in observable dwelling characteristics, as mentioned earlier, as well as household composition. The procedure allows us to hold these characteristics fixed to understand how much they account for inequality changes relative to other characteristics, such as location. We then further disentangle remaining inequality spatially across neighborhoods and local neighborhoods. This decomposition gives us some ability to see the relative contribution of local amenities and local labor markets in explaining housing inequality.

### Shifts in Structural Characteristics

While Table 1 shows that houses have been getting larger, on average, the size and structural characteristics of houses have not become appreciably more unequal. To see this, Figure 3 plots inequality in the hedonic value of structural characteristics over the half century. The predicted housing values are based on regressing the characteristics – the number of rooms, number of bedrooms, the presence and type of heating system, and the presence of indoor plumbing – on the

Figure 2: County Level Changes in Housing Outcomes: 1940 to 1970 vs 1970 to 2019





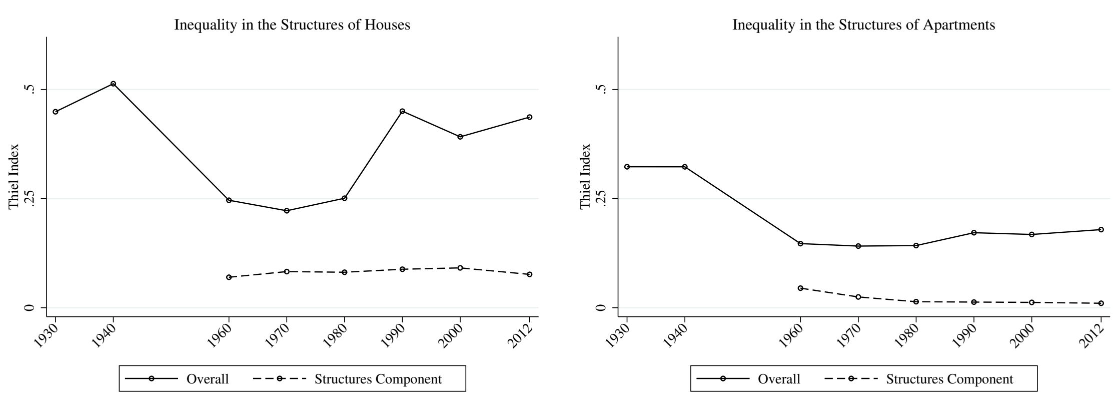

NOTE:

log housing value – housing value and rents. Under conditional independence the figure plots how unequal structural characteristics contribute to overall inequality. Even if conditional independence does not hold, this exercise measures how well observable characteristics predict differences in housing values.

The dashed line in Figure 3 shows that the Theil value of predicted housing values based on structural characteristics has remained stable for home owners, even while overall inequality rose. The pattern for renters shows a contrast in sign, with inequality in structural characteristics falling, even as overall inequality edges upwards. Unless unobserved structural characteristics behave very differently than observed characteristics, it appears that structural characteristics are unlikely to explain the rise in housing inequality seen in the post-War period.

Two robustness exercises support the notion that limited contribution of structures does not depend on the specifics of the hedonic pricing equation. First, inequality in the number of rooms or bedrooms does not change substantially, independently of how without attaching specific valuations, as shown in Appendix Figure A.7. Second, an analysis that re-weights housing in 2012 to match characteristics of housing in earlier years using propensity scores (DiNardo, Fortin and Lemieux, 1996), Appendix C, produces similar results.

Figure 3: Impacts of Observable Structural Characteristics on Inequality




NOTE: Inequality in the structural values of owner occupied houses and rental units has not appreciably risen since 1960, unlike overall housing inequality. Plotted in the solid dashed line is inequality in the structural value of the dwellings specified. The structural value is the predicted value from a hedonic regression of housing characteristics – the number of rooms, number of bedrooms, presence of plumbing, and the presence and type of heating unit – on the log home value or rent. For robustness, Figure A.6 performs a similar exercise and obtains similar results by adjusting the structural characteristics of houses in 2012 to prior years using propensity score re-weights. Each panel refers to the different labeled variables and sub-samples. The solid lines are the same Theil index values as in Figure 1 and we plot them to show scale.

Our results on dwelling characteristics support previous reflections on this issue. For example, Glaeser and Gyourko (2008) find that per-capita square footage consumed by rich and poor households has become more equal over time. Since construction costs vary little within cities, much of the growing inequality in housing value seems to be due to growing inequality in land values, or the right to build on such land.

### Neighborhoods versus Local Labor Markets (under revisions

The service flow from housing combines the benefits from the structure itself with the land and amenities that come with it. Land values could potentially diverge due to differences in access to local labor markets, as some cities increase in productivity relative to others (Van Nieuwerburgh and Weill (2010), Diamond (2016), Gyourko, Mayer and Sinai (2013), and Moretti (2013)). Additionally, local amenities such as schools, crime, and other features of neighborhoods impact values of land (Wilson (2012), Chetty, Hendren and Katz (2016), Rossi-Hansberg, Sarte and Owens (2010), Guerrieri, Hartley and Hurst (2013), Autor, Palmer and Pathak (2014), and Couture et al. (2019)).

In this section, we attempt to disentangle the impact of local labor markets and neighborhoods by decomposing inequality metrics across neighborhoods and commuting zones. In particular, the Theil index–like all generalized entropy indexes–admits a decomposition across $J$ areas as follows:

$$T = \sum_{j=1}^{J} s_j T_j + \frac{1}{J} \sum_{j=1}^{J} \left[ \frac{\bar{y}_j}{\bar{y}} \ln \left( \frac{\bar{y}_j}{\bar{y}} \right) \right]$$

The first term gives the weighted sum of within-area variances (using shares of $y_i$ in region $j$ as weights), and the second term gives the Theil index across area means $\bar{y}_j$.

Panel A of Figure 4 shows this decomposition by grouping Census tracts, a proxy for neighborhoods. Specifically, the dashed line in Panel A of Figure 4 plots the second term in the decomposition – interpretable as the contribution of across neighborhood inequality to overall Theil inequality.[18]

[18] To make across-neighborhood values comparable across time, the statistics are based only on areas where the census provided information about neighborhoods from 1940 onwards, with historical census tract boundaries adjusted to match current definitions. Doing so omits newer tracts within cities, so may provide an incomplete sample of neighborhoods. To address this, Panel B presents results for a sample of tracts in each Census year. These show the same overall qualitative pattern, though there can be some differences in magnitudes. See the Appendix for more

Panel A shows that inequality across neighborhoods makes up a substantial component of the time trend of overall inequality. Inequality across neighborhoods declines by roughly a third from 1940 to 1960. Then from 1960 to 2017 inequality across neighborhoods increases more than three-fold. And the overall time pattern matches trends in overall housing inequality.

The proportion of overall inequality due to differences across neighborhoods (the solid and dashed lines) also grew over the period we study – from around 35 to over 45 percent . Panel B of Figure 4 decomposes overall housing inequality, as measured by the Theil index, into contributions from cross-city differences and cross-neighborhood differences. Neighborhoods are making up an increasing share of overall inequality, with an notable increase in 1990. And the same general trend applies regardless of when they are mapped.

Panel B also shows that the proportion of overall inequality attributable to commuting zones declined from 1930 to 1980, but then increased from 1980 onwards (Gyourko, Mayer and Sinai (2013) and Moretti (2013)). However, cities' contributions to overall housing inequality are less than half the contribution of neighborhoods.

Overall, Figure 4 emphasizes neighborhoods (Wilson (1987) and Chetty, Hendren and Katz (2016)) and local amenities (Rossi-Hansberg, Sarte and Owens (2010)). Neighborhoods have always led to a meaningful level of inequality and a drop in across neighborhood inequality matches drops overall inequality. And more of housing inequality can be explained by neighborhoods than ever.

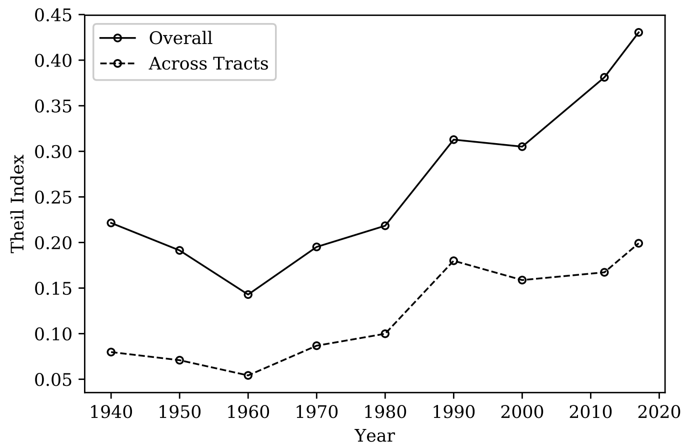
Our results are consistent with the idea that income inequality has led to growing inequality in the enjoyment of of amenities are opportunities, in line with results on consumption inequality more broadly Aguiar and Bils (2015).[19] However, it could also be that the price of fixed amenities is growing instead: for instance the premium for having a property on the beach may have simply grown as wealthier households bid over them.

Several studies (e.g. Van Nieuwerburgh and Weill (2010), Moretti (2013), and Gyourko, Mayer and Sinai (2013)) have examined increases in the inequality of aggregate housing price indices across cities in the post-war period. In comparison, our analysis examines statistics at the household level

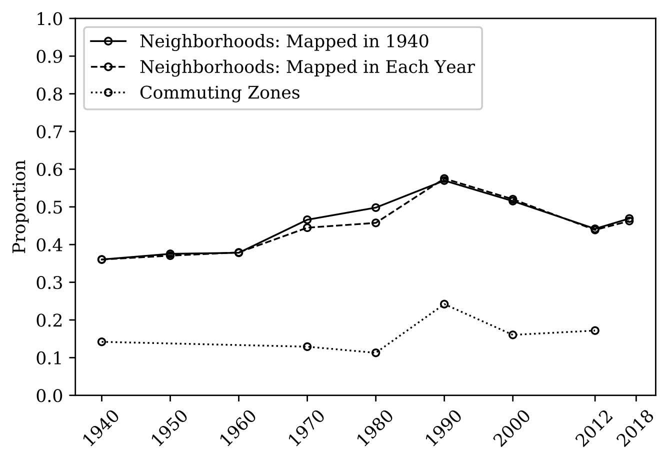
details.

[19] Aguiar and Bils (2015) find inceases in both income and consumption inequality after correcting for measurement error in the Consumer Expenditures Survey (CEX). Their result stands in contrast to prior studies using the CEX which find only a modest increase in consumption inequality given rises in income inequality Krueger and Perri (2006) Meyer and Sullivan (2017). Work by Attanasio and Pistaferri (2014) and earlier evidence by Cutler and Katz (1992) also suggest that consumption inequality has increased recently.

Figure 4: Inequality Across Neighborhoods and Commuting Zones
Panel A: Theil Values




Panel B: Proportions of Inequality Across Geographies




Note: Most inequality is due to differences across neighborhoods (census tracts). Inequality across neighborhoods is a large and growing proportion of overall theil inequality, more so than inequality across commuting zones. This figure shows a decomposition of the Theil inequality index of rents and equivalents to differences across census tracts and commuting zones. Panel A plots a decomposition of the level of Theil Inequality in areas that had tracts in 1940, using time-consistent tract identifiers. Panel B plots the proportion of overall Theil inequality due to differences across neighborhoods and commuting zones. Since many areas only had tract boundaries late in the century, the panel plots both statistics for areas where tracts contemporaneously existed (so additional neighborhoods are included each year) as well as areas that had tracts mapped beginning in 1940 (so the areas are the same for all years). The entire Continental US is always included in the commuting zone sample. All tract boundaries are harmonized to match 2010 boundaries using techniques from Logan, Xu and Stults (2014) and Lee and Lin (2017). Data are via NHGIS (Manson et al. (2020)) and IPUMS (Ruggles et al. (2010)). House prices are interpolated as described in the text.

and includes the pre-war period, besides also covering rental units, and the separate role of dwelling and location.

## Conclusion

We document patterns of inequality in housing prices and rents housing since 1930, using Census data. First, housing inequality fell from 1930 to 1970 and rose thereafter, with a sharper fall and gentler rise than income inequality. Second, home-ownership rose considerably when inequality fell, but stayed largely steady thereafter. Third, the fall in housing inequality was most pronounced at the lower end, as the cheapest units without water or electricity disappeared. Fourth, the post-1970 rise was driven more by owner-occupied than rented units. Fifth, rising inequality is explained more by growing spatial differences across neighborhoods than by changing structural characteristics.

While the descriptive analysis here does not permit one to draw causal inferences, the role of federal housing policy is rather conspicuous: see Snowden, Fetter and Rose (2018) for a more detailed analysis. [20]. The Great Depression saw the creation of the Federal Housing Administration and government sponsored enterprises (viz. Fannie Man and Freddie Mac) to guarantee mortgages for moderately priced homes. Veterans' Assistance did much to help first-time home buyers, often in suburban homes near federally-subsidized highways. Meanwhile, income taxes rates grew, greatly increasing the value of the home mortgage interest deduction, particularly for high earners. Housing policy has continued to evolve through the Savings and Loan Crisis in 1990 and the great housing bubble and the subprime mortgage crisis.

In urban areas, wartime rent controls in the 1940s may have helped compress rents, even though these were eventually eliminated. Various slum clearance and renewal programs affected as well as relocated many of the low-quality housing. In rural areas, the Rural Electrification Administration played a hand in helping the vast majority of housing units have electricity and telephone service by 1970. Massive infrastructure investments made it so that much smaller communities had access to running water and sewage than in 1930. Meanwhile, continuing rural-to-urban migration made it so that Americans were more geographically concentrated, even if they found other reasons to value their neighborhoods differently.

[20] most notably the National Housing Act of 1934, Fair Housing Act of 1968

# References


**Aguiar, Mark, and Mark Bils.** 2015. "Has Consumption Inequality Mirrored Income Inequality?" American Economic Review, 105(9): 2725–56.

**Atkinson, Anthony B.** 1970. "On the measurement of inequality." Journal of Economic Theory, 2: 244–263.

**Attanasio, Orazio P., and Luigi Pistaferri.** 2014. "Consumption inequality over the last half century: Some evidence using the new PSID consumption measure." American Economic Review, 104(5): 122–126.

**Autor, David H., Christopher J. Palmer, and Parag A. Pathak.** 2014. "Housing Market Spillovers: Evidence from the End of Rent Control in Cambridge, Massachusetts." Journal of Political Economy, 122(3): 661–717.

**Autor, David H., David Dorn, and Gordon Hanson.** 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." American Economic Review, 103(6): 2121–2168.

**Baum-Snow, Nathaniel, and Ronni Pavan.** 2013. "Inequality and City Size." Review of Economics and Statistics, 95(5): 1535–1548.

**Bischoff, Kendra, and Sean F. Reardon.** 2014. "Residential Segregation by Income, 1970-2009." In Diversity and Disparities: America Enters a New Century. , ed. John Logan, Chapter 7, 208–233. Russel Sage Foundation.

**Bricker, Jesse, Alice Henriques, Jacob Krimmel, and John Sabelhaus.** 2016. "Measuring income and wealth at the top using administrative and survey data." Brookings Papers on Economic Activity, 2016(1): 261–331.

**Chetty, Raj, Nathaniel Hendren, and Lawrence F. Katz.** 2016. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." American Economic Review, 106(4): 855–902.

**Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst.** 2019. "Income growth and the distributional effects of urban spatial sorting." NBER Working Paper 26142.

**Cutler, David M., and L. F. Katz.** 1992. "Rising inequality? Changes in the distribution of income and consumption in the 1980's." American Economic Review, 82(2): 546–551.

**Davis, Morris A., and Jonathan Heathcote.** 2007. "The price and quantity of residential land in the United States." Journal of Monetary Economics, 54(8): 2595–2620.

**Diamond, Rebecca.** 2016. "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000." American Economic Review, 106(3): 479–524.

**DiNardo, John, Nicole M Fortin, and Thomas Lemieux.** 1996. "Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach." Econometrica, 64(5): 1001–1044.

**Dorn, David.** 2009. "Essays on Inequality, Spatial Interaction, and the Demand for Skills." PhD diss. University of St. Gallen.

**Fetter, Daniel K.** 2016. "The Home Front: Rent Control and the Rapid Wartime Increase in Home Ownership." The Journal of Economic History, 76(4): 10011043.

**Firpo, Sergio.** 2007. "Efficient semiparametric estimation of quantile treatment effects." Econometrica, 75(1): 259–276.

**Fortin, Nicole, Thomas Lemieux, and Sergio Firpo.** 2011. "Chapter 1 Decomposition Methods in Economics." In Handbook of Labor Economics. Vol. 4, 1–102. Elsevier.

**Glaeser, E.L., and J.E. Gyourko.** 2008. Rethinking federal housing policy: how to make housing plentiful and affordable. Vol. 76, AEI Press.

**Goldin, Claudia, and Robert A Margo.** 1992. "The great compression: The wage structure in the United States at mid-century." The Quarterly Journal of Economics, 107(1): 1–34.

**Guerrieri, Veronica, Daniel Hartley, and Erik Hurst.** 2013. "Endogenous gentrification and housing price dynamics." Journal of Public Economics, 100: 45–60.

**Gyourko, Joseph, Christopher Mayer, and Todd Sinai.** 2013. "Superstar Cities." American Economic Journal: Economic Policy, 5(4): 167–199.

**Hanushek, Eric A, and John M Quigley.** 1980. "What is the Price Elasticity of Housing Demand?" Source: The Review of Economics and Statistics, 62(3): 449–454.

**Hirano, Keisuke, Guido W Imbens, and Geert Ridder.** 2003. "Efficient estimation of average treatment effects using the estimated propensity score." Econometrica, 71(4): 1161–1189.

**Kopczuk, Wojciech.** 2015. "What Do We Know about the Evolution of Top Wealth Shares in the United States?" Journal of Economic Perspectives, 29(1): 47–66.

**Kopczuk, Wojciech, and Emmanuel Saez.** 2004. "Top Wealth Shares in the United States: 1916-2000: Evidence from Estate Tax Returns." National Tax Journal, 57(2): 445–488.

**Krueger, Dirk, and Fabrizio Perri.** 2006. "Does income inequality lead to consumption inequality? Evidence and theory." Review of Economic Studies, 73(1): 163–193.

**Lee, Sanghoon, and Jeffrey Lin.** 2017. "Natural Amenities, Neighbourhood Dynamics, and Persistence in the Spatial Distribution of Income." The Review of Economic Studies, 85(1): 663–694.

**Lerman, Robert, and Shlomo Yitzhaki.** 1985. "Income Inequality Effects by Income." The Review of Economics and Statistics, 67(1): 151–56.

**Logan, John R, Andrew Foster, Hongwei Xu, and Wenquan Zhang.** 2020. "Income Segregation: Up or Down, and for Whom?" Demography, 57(5): 1951–1974.

**Logan, John R., Zengwang Xu, and Brian J. Stults.** 2014. "Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database." The Professional Geographer, 66(3): 412–420. PMID: 25140068.

**Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles.** 2020. "IPUMS National Historical Geographic Information System: Version 15.0."

**Massey, Douglas S., Mary J. Fischer, William T. Dickens, and Frank Levy.** 2003. "The Geography of Inequality in the United States, 1950-2000 [with Comments]." Brookings-Wharton Papers on Urban Affairs, 1–40.

**Matlack, Janna L., and Jacob L. Vigdor.** 2008. "Do rising tides lift all prices? Income inequality and housing affordability." Journal of Housing Economics, 17(3): 212–224.

**Mayo, Stephen K.** 1981. "Theory and estimation in the economics of housing demand." Journal of Urban Economics, 10(1): 95–116.

**Meyer, Bruce D, and James X Sullivan.** 2017. "Consumption and Income Inequality in the U.S. Since the 1960s." National Bureau of Economic Research Working Paper 23655.

**Moretti, Enrico.** 2013. "Real wage inequality." American Economic Journal: Applied Economics, 5(1): 65–103.

**of the Census, Bureau, and F. John Devaney.** 1994. Tracking the American Dream: 50 Years of Housing History from Census Bureau: 1940 to 1990. Vol. H121 of Current Housing

Reports. 94-1 ed., U.S. Government Printing Office.

**Pareto, Vilfredo.** 1896. Oeuvres complètes: Tome 3, Ecrits sur la courbe de la répartition de la richesse. Librairie Droz.

**Peiser, Richard B., and Lawrence B. Smith.** 1985. "Homeownership Returns, Tenure Choice and Inflation." Real Estate Economics, 13: 343–360.

**Piketty, Thomas, and Emmanuel Saez.** 2003. "Income inequality in the United States, 1913–1998." The Quarterly journal of economics, 118(1): 1–41.

**Polinsky, A Mitchell, and David T Ellwood.** 1979. "An empirical reconciliation of micro and grouped estimates of the demand for housing." The Review of Economics and Statistics, 199–205.

**Poterba, James, and Todd Sinai.** 2008. "Tax Expenditures for Owner-Occupied Housing: Deductions for Property Taxes and Mortgage Interest and the Exclusion of Imputed Rental Income." American Economic Review, 98(2): 84–89.

**Reardon, Sean F., and Kendra Bischoff.** 2011. "Income Inequality and Income Segregation." American Journal of Sociology, 116(4): 1092–1153.

**Rossi-Hansberg, Esteban, PierreDaniel Sarte, and Raymond Owens.** 2010. "Housing Externalities." Journal of Political Economy, 118(3): pp. 485–535.

**Ruggles, Steven, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek.** 2010. "Integrated Public Use Microdata Series: Version 5.0."

**Saez, Emmanuel, and Gabriel Zucman.** 2016. "Wealth inequality in the United States since 1913: Evidence from capitalized income tax data." The Quarterly Journal of Economics, 131(2): 519–578.

**Saez, Emmanuel, and Gabriel Zucman.** 2020. "The rise of income and wealth inequality in America: Evidence from distributional macroeconomic accounts." Journal of Economic Perspectives, 34(4): 3–26.

**Snowden, Kenneth, Daniel Fetter, and Jonathan Rose.** 2018. "Housing in American Economic History." The Oxford handbook of American economic history, , ed. Louis P. Cain, Price V. Fishback and Paul W.Editors Rhode. Oxford University Press.

**Tolbert, Charles M., and Molly Sizer.** 1996. "U.S. Commuting Zones and Labor Market Areas: A 1990 Update." US Department of Agriculture: Economic Research Service Staff Working Papers Number 9614.

**Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill.** 2010. "Why Has House Price Dispersion Gone Up?" Review of Economic Studies, 77(4): 1567–1606.

**Verbrugge, Randal.** 2008. "The puzzling divergence of rents and user costs, 1980-2004." Review of Income and Wealth, 54(4): 671–699.

**Wilson, William Julius.** 1987. The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy. Chicago:University of Chicago Press.

**Wilson, William Julius.** 2012. The truly disadvantaged: The inner city, the underclass, and public policy. University of Chicago Press.

**Yao, Yuxi.** 2019. "Land and the Rise in the Dispersion of House Prices and Rents across U.S. Cities." Unpublished manuscript.

# Appendix - For Online Publication

## A Data

### A.1 Integrated Public Use Microdata Series sample restrictions

Our Census data are provided by the Integrated Public Use Microdata Series, from usa.ipums.org detailed in Ruggles et al. (2010). We use the following samples.

- 1930: 5 percent sample
- 1940: 1 percent sample
- 1960: 1 percent
- 1970: 1 percent state and metro (depending on state or commuting zone) and format 1 or 2 (depending on the questions needed).
- 1980: 5 percent state sample
- 1990: 5 percent sample
- 2000: 5 percent sample
- 2009-12 ACS five year 2012 sample with interviews in 2008 dropped

As mentioned in the main text, we include households in the United States with three restrictions: We omit houses outside the continental US, we exclude houses used to generate revenue, and we similarly exclude owner occupied multiple family residences. Appendix table A.1 shows the percentage of houses we exclude for each year due to being used to generate revenue or encompassing multiple units. We present these separately for both owners and renters.

The first row for each panel shows the number of houses in the continental US that are included in our sample. this is quite high for each year, and in later years it is over 90 percent. The next few rows show reasons why houses are excluded. By far the most important reason is because they are farms. For example, in 1930 25 percent of households lived on a farm. Multiple family houses and houses used for commercial purposes are also excluded in our analysis for the years where the census identifies them. In 1960, the first year we can identify them, these housing categories made up roughly 6 percent of owner occupied housing.

Additionally, the 1930 and 1940 data also do not include dwelling characteristics, nor do they indicate the presence of businesses, so we can neither measure nor exclude businesses and multi-family structures in these years. These two categories account for only 6 percent of owner-occupied homes in 1960. The 1940 enumerator instructions also noted that respondents should exclude the value of units rented out and to also exclude the area of the house that is used for business purposes if it is a “considerable portion” of the house, however. So values 1940 may be more comparable to our later sample than 1930 in terms of methodology. This difference in methodology limits the comparability of 1930 and 1940, however.

The final few rows show categories of houses that were excluded from questions in some years. From 1960-1980 mobile homes and trailers were excluded and in 1970 condos were as well. In other years they are included. These categories, again, make up a small proportion of housing units in these years, They never reach six percent, and often are well below that.

Table A.1: Exclusions imposed in sample selection

Panel A: Home Owners

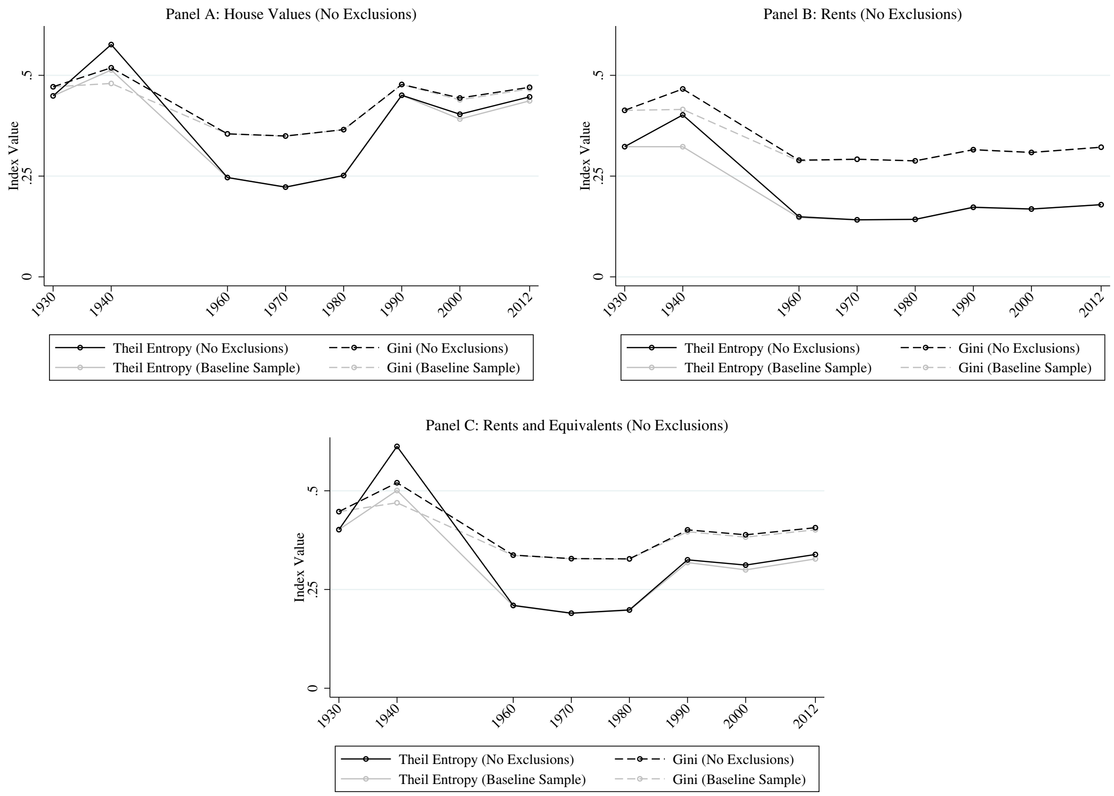
| | 1930 | 1940 | 1960 | 1970 | 1980 | 1990 | 2000 | 2012 |
|---|---|---|---|---|---|---|---|---|
| Included | 0.748 | 0.758 | 0.843 | 0.873 | 0.881 | 0.910 | 0.900 | 0.922 |
| Farm | 0.252 | 0.242 | 0.093 | 0.051 | 0.029 | 0.023 | 0.019 | 0.017 |
| Commercial use | | | 0.013 | 0.010 | 0.015 | 0.020 | 0.033 | 0.015 |
| Multiple families | | | 0.051 | 0.065 | 0.075 | 0.047 | 0.048 | 0.046 |
| Mobile home or similar | | | 0.015 | 0.038 | 0.055 | 0.082 | 0.083 | 0.067 |
| Condominium | | | | 0.002 | 0.007 | 0.015 | 0.018 | 0.032 |
| Sample size | 516,941 | 585,459 | 275,602 | 1,388,542 | 2,292,767 | 2,826,214 | 3,244,508 | 3,128,981 |

Panel B: Renters

| | 1930 | 1940 | 1960 | 1970 | 1980 | 1990 | 2000 | 2012 |
|---|---|---|---|---|---|---|---|---|
| Included | 0.798 | 0.837 | 0.977 | 0.986 | 0.988 | 0.989 | 0.986 | 0.993 |
| Farm | 0.202 | 0.163 | 0.023 | 0.008 | 0.005 | 0.004 | 0.003 | 0.002 |
| Commercial use | | | | 0.007 | 0.007 | 0.007 | 0.011 | 0.005 |
| Mobile home or similar | | | 0.003 | 0.011 | 0.023 | 0.050 | 0.046 | 0.044 |
| Condominium | | | | | 0.017 | 0.045 | | |
| Sample size | 629,593 | 850,940 | 175,464 | 848,785 | 1,296,401 | 1,346,603 | 1,505,211 | 1,281,471 |

NOTE: The gives the (unweighted) numbers of houses falling into categories that either merit exclusion or fall out of the universe of houses we have data for in some years. Farms, units used for commercial purposes, and multiple family owner occupied structures are excluded from our statistics. Mobile homes (as well as tents, vans, trailers, and boats) and condominiums are not asked about in certain years, so they are absent in those years but included when they are asked about.

Figure A.1: Housing Inequality Without Sample Exclusions




Our decisions about sample comparability do not appear to drive our results, since the same statistics calculated among all dwellings surveyed in each year show very similar patterns. Appendix Figure A.1 plots in black the gini and theil values for all dwellings available in the data in each year. The grey lines represents the values for the restricted sample of owner and renter occupied dwellings in each year, the same as in Figure 1. The statistics show a U shaped pattern regardless of which sample we use, and they are quite similar in the later period, when the number of exclusions is much less. If anything, the inclusion of farms in 1940 tends to increase inequality in that year, heightening the size of the initial decline.[21]

## A.2 Variable harmonizations

Several variables are harmonized over time, which occasionally leads us to intentionally ignore information in some years to prevent our results from being driven by the more detailed information (generally) available in later years. Most notably, data about the structural characteristics of houses is only available for 1960 onwards. Data on room numbers are also not available in the same detail for each year, and so we collapse categories of the number of rooms in houses to allow us to cover each year.

Additionally, the number of rooms is available in 1960, but not the number of bedrooms. Since we use both the number of rooms and the number of bedrooms in our calculation of the structural

[21]Enumerators were instructed to ask about the value of buildings on farms used as dwellings separately from the rest of the farm in 1940. So the values of farms in this year may be closer to our preferred concept of separating out residential from commercial value.

value of a house for other years, this leads to an inconsistency in 1960. We resolve this by ignoring variation in the number of bedrooms – we set the number of bedrooms to zero in this year and introducing a fixed effect for the year. So structural values are based on variation in other variables – most notably the number of rooms, but not bedrooms.

Total household incomes are also measured inconsistently across years. So we focus on household wage income only, because it is more consistently measured.

## A.3 Pareto Interpolation

To account for issues with rounding, top coding, questions asking about bins, and data being reported in intervals, we use a pareto interpolation procedure. We use a pareto distribution to interpolate values both in years where data are intervalled and where they are continuous. Where data are intervalled, we interpolate within the specified intervals. Where data are continuous, we divide the data into equally size intervals. We start with 25, but decrease the number until each interval has a reasonable "width," or distance for its minimum to maximum value.[22]

The pareto distribution's cumulative distribution function (CDF) has the from $F(y) = 1 - (k/y)^\alpha$ for $y \geq k$. It follows that the logarithm of the complementary CDF (or "tail" function) has a slope of $-\alpha$ with respect to $\ln(y)$. The greater the slope $\alpha$, the faster observations "die off," and the less disperse the distribution. For any value $y_i$, define

$$p_i \equiv ln[1 - F(y)] = \alpha \ln(k) - \alpha \ln(y) \tag{A.1}$$

The interpolation method (Pareto (1896)) estimates $\alpha$ by differencing expression A.1 at the values for the two endpoints of a bracket $[y_i, y_{i+1}]$, where $i = 1, ..., I$ indexes the intervals, namely

$$\hat{\alpha}_i = \frac{\ln(y_{i+1}) - \ln(y_i)}{p_i - p_{i+1}}, i = 1, ..., I - 1, \tag{A.2}$$

and $\hat{k}_i = y_i p_i^{1/\alpha}$. Even when the distribution may not be considered Pareto globally, this procedure fits a simple line segment to the log survivor function in terms of $\ln(y)$.

### Extrapolations for top and bottom bins

Since the Pareto distribution is unbounded, determining a value of $\alpha$ beyond the top code, $y_I$ in the data can be difficult. This problem is largely remedied if the mean is available, as it is for housing values in 1930, 1940, and 2012. There, the property that the mean value of $\bar{y}_I \equiv E[y|y \geq y_I] = \alpha y_I/(\alpha - 1)$, imples a value of

$$\hat{\alpha}_I = \frac{\bar{y}_I}{\bar{y}_I - y_I} \tag{A.3}$$

In situations where we do not have a mean above a given cutoff we extrapolate $\alpha_J$ for the topmost bin from either other values of $\alpha_j$ or an assumed hard code. Depending on the relative sample sizes, the values come from the next interval down, other intervals in the top 10 percent of

[22] The precise definition is that the procedure will decrease the number of intervals until each interval has a max value that is 1.01 times its min value or greater. We do this because the procedure fails for intervals that are excessively "narrow" as sometimes occurs where respondents round values.

the distribution, the average of topmost parameters in other areas, or the hard code.[23][24]

For the bottom interval we extrapolate values using a uniform distribution to deal with bottom coding and problems associated with zero values. We set the minimum of this interval as one one-hundredth of the top value.

### Geographic breakdowns

To perform the procedure, we aggregate observations in each year by geography. Our choice of geography balances the additional resolution that comes from finer geographies with the possibility of spurious parameter values, particularly for the extrapolated portions. It also is dependent on having a consistent measure of geography, so we choose a different geography based on the geographic specificity of our data source. For data on individual houses from IPUMS, we perform the interpolation at the state by year level, since we have consistent information about what state a house is in for all years. For the data drawn from neighborhood summary tables posted in NHGIS, we perform the imputation at the level of the neighborhood in each year.[25]

### Illustrations

Figure A.2 shows several illustrations of our imputation procedure for different types of situations. We plot both the CDF and the tail function, which is linear in the pareto distribution. The second plot is especially relevant for determining the fit in the very top of the distribution.

Panel A of Figure A.2 shows both the empirical cumulative density function for home prices in 2012 and the pareto CDF used in the paper. Note that this is a year where respondents reported a continuous quantity, though the distribution plot shows that many did round. The dashed lines show intervals in each bin used for rounding. As in other years, we divided the data into 25 bins to perform the CDF interpolation. The dashed lines show the min and max proportions for each bin, as well as its placement along the scale.

Figure Panel B shows how the Pareto distribution matches the top of the distribution. Overall the fit is quite good, with the line tracking the distribution until it encounters the state by year topcodes for the top 0.5 percent of the distribution.

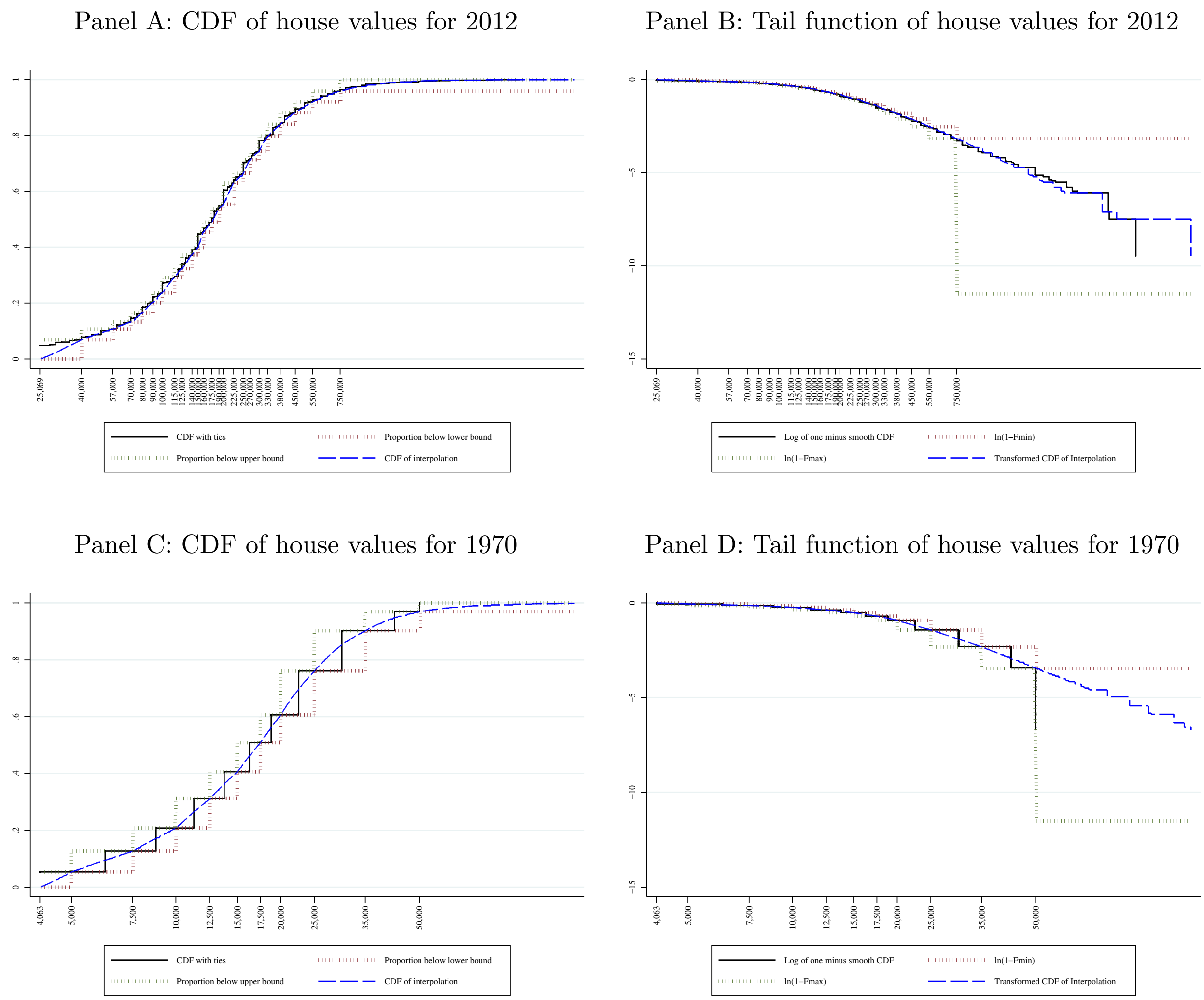
Similarly Panel C and Panel D show the same for the 1970 distribution where the variable is divided into 11 intervals. Here the distribution meets the interval boundaries exactly and the interpolation simply smooths the area between them. The top code extrapolation has the same slope as the one for 2012, but this slope matches the slope of the previous interval rather well, implying the extrapolation is reasonable for this year as well.

[23] Rent values of the topmost bins in 1930 and 1940 are always hard coded to be 2.5. In these years we have a continuous distribution of rents (and home values) with very few top codes; parameters estimated from the mean procedure lead to suspiciously high inequality statistics.

[24] In the extremely rare case where we attempt to use paramters in the top decile, but all values are in the topmost bin, we check the top 11, 12, 13, etc. percent, stopping at the first percentile where an estimate exists.

[25] To make our analyses less susceptible spuriously high dispersion in the top bin for small geographies, we windsorize the topmost values of the imputed distribution for individual neighborhoods at 1,000 times the overall median value. Since the procedure is done separately by neighborhood and since the top tail is larger in later years, this decision should decrease the growth of between tract inequality in later years. In practice, it rules out to relatively few observations and results are fairly similar without the windsorization.

Figure A.2: Example CDF plots for the pareto interpolation




NOTE: The term "tail function" corresponds to the logarithm of one minus the the CDF.

## A.4 Assigning Commuting Zones in Micro Data Samples

We keep a consistent geographic sample across all years in the micro data sample by primarily using 1990 local labor markets, or commuting zones, as described by Tolbert and Sizer (1996). Local labor markets are a natural geography for this analysis since housing values are closely related to local income levels. For 1970-2011 we match geographically using the probabilistic mapping made available by David Dorn via his website and used in, for example, Dorn (2009). For 2012 we replicate the probabilities matching using updated PUMA definitions for the 2012 census and for 1930 and 1940 we use the direct county to commuting zone mapping in Tolbert and Sizer (1996) since counties are available. Unfortunately it is not possible to match to commuting zones in for 1960 based on the publicly available samples, so we instead use states as our geographic entity where we use data from 1960 (this is generally marked in each table).

This mapping uses the most disaggregated geography available in each census to project this geography onto counties and then project these counties onto commuting zones. Unfortunately, the

PUMAS and county groups available often contain multiple counties and these counties sometimes fall into separate commuting zones. In these cases we compute a probability that a given person resides in one commuting zone based on the proportion of people in her identified geography that in fact live in the given commuting zone based on non-restricted census summary tables. Where these probabilistic matches occur, we create multiple observations – one for each possible commuting zone – and weight them by the probability the original observation is in that commuting zone, times their initial weight if applicable. Samples for 1930 and 1940 contain county information so the match is exact and we avoid the first projection.

### A.5 Deflator

Since the inequality measures we use are scale invariant, our deflator is generally unimportant. We deflate values using the Consumer Price Index excluding structures published by the BLS, from the Federal Reserve Bank of St.Louis' FRED system. The series mnemonic is: `CUUR0000SA0L2`. Before 1935, when structures were not identified separately, we combine the data series from the general CPI (`CPIAUCNS`) and the CPI for rent of primary residence (`CUUR0000SEHA`) by using an OLS regression to predict the shelters excluded series for the years it is available prior to 1941 based on the general CPI and the CPI for rent of primary residence. This regression has good out of sample prediction powers for years up to 1960, so we believe it is a reasonable approximation for the initial five years.

## B Robustness of User Cost Assumptions

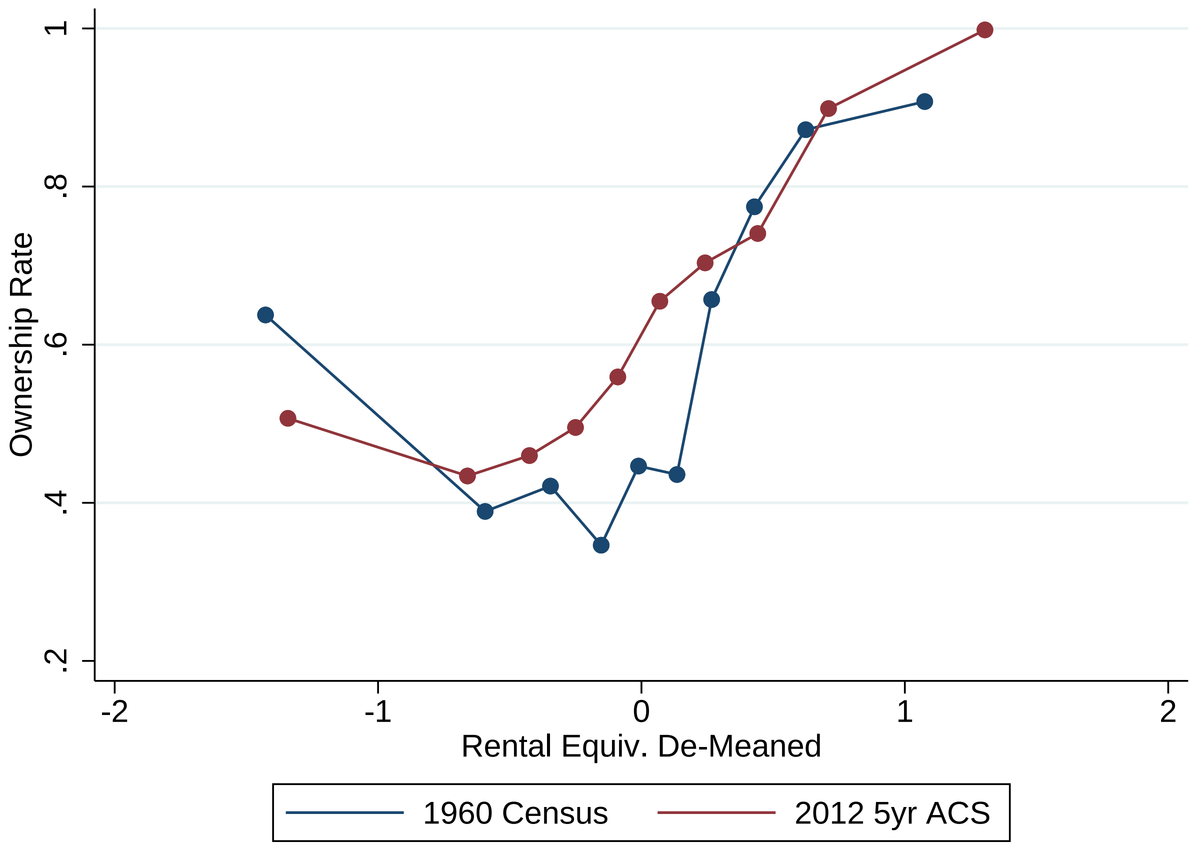
We show that inequality in house prices for owners has risen considerably since 1960, though inequality in rents for renters has remained relatively stable. In the baseline results, we combine these two facts into single housing services measure by converting home values into rental equivalents using a common user cost over time. Doing so leads to a conclusion that that inequality in housing services has risen over time. In this appendix we explore how selection into ownership and renting impact the observed distributions of rents and home prices, and examine our assumptions about user costs used in our baseline results.

### B.1 Composition shifts in ownership

As we have showed, inequality in rents among renters remained relatively stable over time since the 1960s while inequality in housing values for owners rose. The two facts paint very different pictures for housing inequality over the last 50 years. The differences reflect shifts in ownership rates at various points in the distribution of equivalent rents, hence generating differential selection into owning or renting at various points. This section analyzes how this selection interacts with shifts in the underlying distribution of equivalent rent.

Figure A.3 shows ownership rates by demeaned equivalent rents in 1960 and 2012. Each dot denotes the average within a decile of the equivalent rent distribution. In the second half of the last century, home ownership rates rose at middle and high values of equivalent rents and actually fell some at lower equivalent rent values. (Ownership has been persistently high at very high levels of equivalent rents.) As such, renters became concentrated towards the bottom of the distribution, reducing inequality within renters due to selection. At the same time tail values of housing services moved outward (as seen by the rightward shift of upper deciles), raising inequality among owners due to widening gaps in home values. This meant shifts in the distribution of equivalent rents showed through primarily to home owners, while selection offset increasing inequality in rents

Figure A.3: Home Ownership vs Equivalent Rent



among renters. As we see in Figure 1, the full distribution of rental equivalents–which is not subject to selection into ownership–combines these two stories, showing an overall fanning out of housing services across the distribution.

### B.2 Stability of Price-Rent Ratios

In our baseline results, we assume the user cost of housing (or the price-rent ratio) is stable over time, such that the increase in house price inequality maps to an increase in owner-equivalent rents. When combined with rents among renters, we see an increase in inequality of housing services over time. Of course, an alternative story is that rising home price inequality is simply driven by rising inequality in user costs over time, while inequality in housing services has remained rather stable. If this were true, combined housing services inequality would have been largely flat since 1960.

To allay concerns about shifts in the distribution of user costs driving our results, we provide evidence from the Consumer Expenditure Survey (CEX) which has collected data from owners on both their self-reported home value and their perceived equivalent rents since 1984, providing a measure of price-rent ratios at the household level over time.[26] By asking owners to report both prices and equivalent rents directly, we can abstract from model-based measures of user costs that attempt to add up interest costs, maintenance, depreciation, and (notably) expected capital gains, which are not observed. Instead, our approach provides a direct view of how households perceive the value of their home as an asset and the value of service flows it produces.

Figure A.4 shows the distribution of price-rent ratios reported by owners in the CEX is stable over the observed period from 1984 onward, and roughly centered at our assumed price-rent conversion factor. Price-rent ratios did rise some during the housing bubble period, but did so across the distribution of price-rent ratios. Moreover, the spread of the distribution appears largely stable across Census years covered by the CEX sample period.

The figure suggests that while user costs are unlikely to be constant across households and time, our constant user cost assumption provides a reasonable baseline. Importantly, assuming a different level of constant user cost would impact results little, as doing so simply scales how house price inequality maps into housing services, but does not alter its shape. The CEX data support the fact that user costs do not show a clear temporal trend that would cause this scaling to differ by year, reducing the increase we see in house price inequality when translated into housing services.

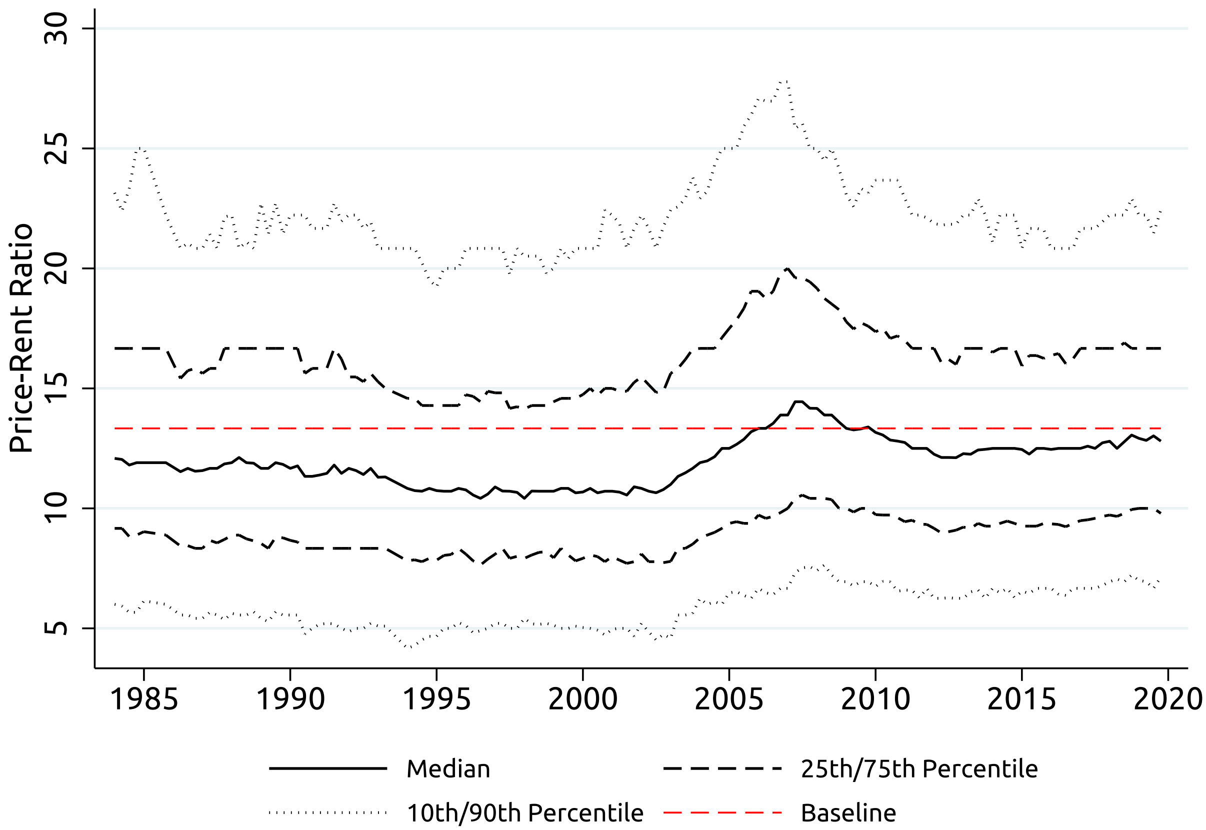
The spread of price to rent ratios in the CEX could reflect cross sectional variation in user costs that could influence our results, though it is unlikely to overturn our main takeaways for three reasons.[27] The first and most direct is simply that the spread of price to rent ratios is remaining relatively constant over time. Assuming no change in the correlation of user costs with housing values, a constant variance of user costs would imply no impact on overall inequality through changes in user costs. The second is our finding that differences across neighborhoods are driving rising housing inequality. For the increasing differences across neighborhoods to be driven by user costs, user costs would have to be changing across neighborhoods within cities.[28] The third is

[26]The CEX began collecting owner equivalent rent measures in 1984 to assist with better measurement of the CPI rents series. Home values are also collected in the first survey, and also the outgoing survey in more recent years. CEX microdata from BLS and ICPSR provide household-level responses for both variables in most quarters since 1984q1, though owners' rents are missing in 1985q2-3 and 1986q1.

[27]Impacts on inequality would depend on how they relate to housing values. For example if user costs were negatively associated with housing prices, they could decrease inequality.

[28]It would be difficult to imagine that changes in tax treatment, maintenance, or interest expenses could vary so dramatically across neighborhoods. Expectations of future price increases could vary by neighborhood and affect user costs, but even if these differences were driving our results, the takeaway would still be related to (expected future) valuations of housing services.

Figure A.4: Distribution of Price-Rent Ratio over Time



that it is possible that the observed variance of price to rent ratios in the CEX is largely due to mismeasurement of either quantity in the survey, and not variances in underlying user costs.

### B.3 Per-Capita Housing Services

Our results suggest that housing services have become more disparate across households, but this fact may simply reflect differences in household size and composition over time. In particular, if household sizes have become increasingly different, households may demand different amounts of housing. While this may lead to heterogeneity at the household level, the distribution of per-capita housing services may not be altered.

To check whether this is driving our result, we adjust our user cost conversion to allow for differences based on household size. Specifically, rather than assuming a constant user cost of 0.0785, we divide this figure by the OECD equivalent-scaled household size given by $1 + 0.7(A - 1) + 0.5C$ where $A$ is the total number of adults in the household and $C$ is the number of children (14 years old or below). The OECD equivalent scaling is a commonly used method of accounting for economies of scale in consumption, such as sharing of common spaces, kitchens, etc. This measure provides a view housing services on a per-person basis.

Broad trends in housing inequality are similar with and without including adjustments for household size with an equivalence scale. Figure A.5 shows that the same U shape, with sharp declines in the mid 20th century, is apparent after adjusting for changes in household size. Additionally, the difference between the dark lines that include the adjustment and the grey lines that do not shows a narrowing gap after including the adjustment. This could this narrowing gap makes the declines in mid century somewhat less dramatic, though they still are large. The declining impact of differences in household size could reflect the trend of households tending to be smaller and more homogeneous in terms of size throughout the 20th and early 21st century.

## C Reweighting

### C.1 Procedure

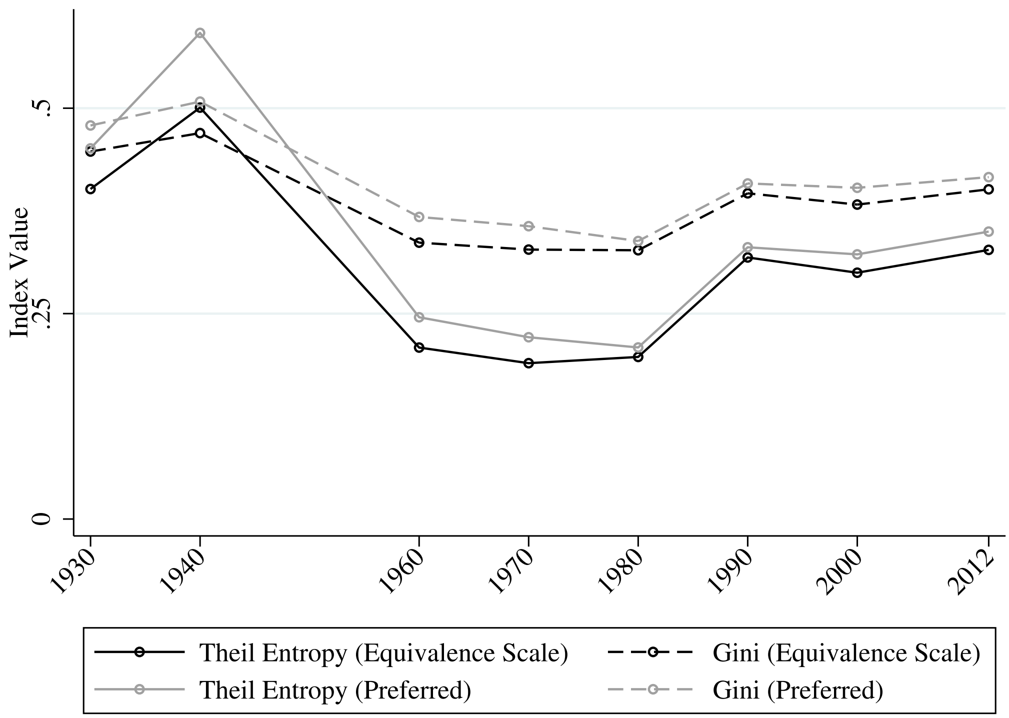
Formally, using notation similar to Fortin, Lemieux and Firpo (2011), we have two groups of houses, one group of houses observed in $t$ and another in $t'$. We have prices for each group and some houses may appear in both groups. We want to compute a statistic, $\nu(\mathcal{P})$, that is a function of a distribution of house prices, $\mathcal{P} \to \mathbb{R}$. With the above caveats in mind, we denote the distribution $\mathcal{P}_{t,t'}$ where the first subscript refers to the year we would like to apply prices from and the second year is the year we would like to match in terms of observables. So we re-weight observed houses in $t$ so that they corresponded to the distribution of characteristics of houses in $t'$. Note that in this notation $\mathcal{P}_{t,t}$ is simply the distribution of houses in period $t$. Putting this together, we are seeking to recover:

$$\nu\left(\mathcal{P}_{t,t'}\right) \tag{A.4}$$

Conditional independence, or ignorability, in this context means that once we condition on all observable characteristics ($S$ and $L$), the unobservable characteristics ($\epsilon$) are independent of whether we observe the house in $t$ or $t'$. Here this would imply that once we take into account each observable detail about the house there is no special trait which we omitted that differs systematically between $t$ and $t'$.

To actually estimate $\nu\left(\mathcal{P}_{t,t'}\right)$, we compute re weighting factors following DiNardo, Fortin and Lemieux (1996) which Hirano, Imbens and Ridder (2003) and Firpo (2007) formally establish are

Figure A.5: Inequality in Rents and Equivalents (Equivalence Scale)




Note: Trends in housing inequality are broadly similar with and without including adjustments for household size with an equivalence scale. Adjusting for changes in household sizes makes inequality smaller in early years however. Plotted are Gini and Theil indexes of inequality for rents and rental equivalents divided by an equivalence scale and without adjustment (the same as Panel C of Figure 1).

efficient. $\Phi(S, L)$ with the following form:

$$\Phi(S, L) = \frac{\Pr(\tau = t'|S, L)/\Pr(\tau = t')}{\Pr(\tau = t|S, L)/\Pr(\tau = t)} \tag{A.5}$$

Where $\tau$ represents the year the house was observed in (either $t'$ or $t$). The intuition is that we would like to heavily weight the observations from year $t$ that look very much like the observations in year $t'$. We do this by using the odds that an observation came from $t'$ given its observables.

We compute the probabilities using logit regressions that pool houses observed in both $t$ and $t'$. By regressing a dummy indicator of a house being observed in $t'$ on a flexible functional form of $S$ and $L$ In this context we mostly use dummy variables for different categories. we are able to compute predicted probabilities $\hat{\Pr}(\tau = t'|S, L)$ and $\hat{\Pr}(\tau = t|S, L)$. To compute $\hat{\Pr}(\tau = t')$ and $\hat{\Pr}(\tau = t)$ we simply use the sample proportions (appropriately weighted). Applying equation A.5 then gives a value of $\hat{\Phi}(S, L)$ that we use to re-weight each statistic.

We then multiply $\hat{\Phi}(S, L)$ by the weights already applied to the houses observed in $t$ to simulate the density that would exist under the counter-factual that houses had identical characteristics to houses in $t'$ but were priced in terms of the prices that prevailed in $t$.

### C.2 Results

To account for how observable changes in housing characteristics have affected inequality, we use the propensity-score re-weighting approach of DiNardo, Fortin and Lemieux (1996). This approach estimates what the distribution of housing prices would be if the observed characteristics of houses had remained identical to earlier years, and those characteristics had been priced as in later years. Specifically, the re-weighting makes the observable characteristics of a sample of houses in a later "current" year, $t$ (e.g. 2010), resemble characteristics from a prior "target" year, $t'$ (e.g. 1970). [29] By doing so, we can infer the distribution of house prices holding fixed observed characteristics at a base year. Such a decomposition does rule out general-equilibrium effects (such as changes in demand) affecting prices. Nonetheless, it is a useful accounting method, as changes from re-weighting tell us how well observable characteristics could affect inequality. Residual variation is informative of how *other* factors such as local amenities impact inequality.

Results in Figure A.6 and Table A.2 show that if households lived in housing units with dwelling characteristics, locations, and household composition observably identical to those in 1970, housing values would be only slightly more equal. Relative to overall changes in housing inequality, changes attributable to movements in these observable characteristics appear rather small.[30]

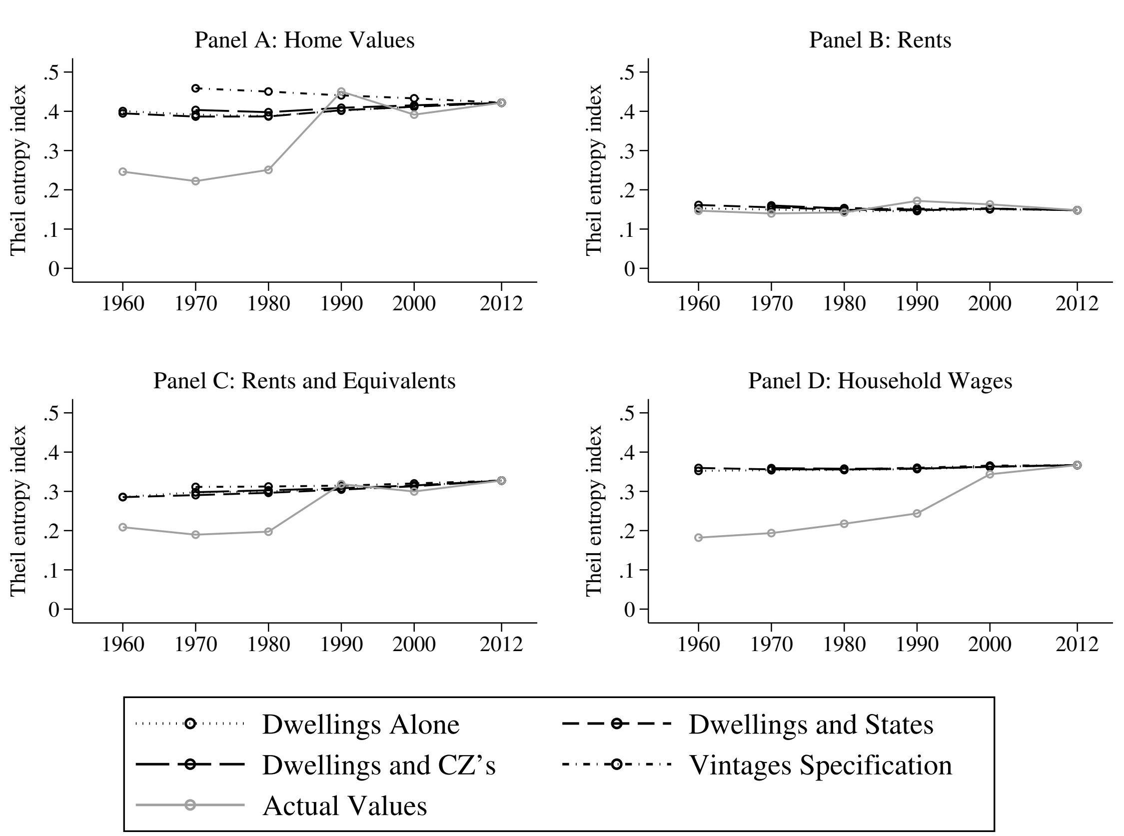
Table A.2 shows that dwelling characteristics can explain, at most, less than a third of the change in the two inequality statistics for rents and rent equivalents. The direction also varies between home owners and renters, and reweighting based on previous characteristics actually increases inequality among renters, suggesting that the rental housing stock has actually become more equal

[29] Fortin, Lemieux and Firpo (2011) provides detailed conditions for identification and implementation details. The procedure re-weights observations using the relative odds of a house being observed in the target year relative to the current year, which we determine using a logit regression on the pooled sample in years $t$ and $t'$.

[30] The re-weighting specification we use allows for depreciation by including their age as an explanatory variable. This rules out controlling for "vintages" of houses with features particular to their year of construction. The "Vintages" specification includes a house's vintage, but not its age (implicitly assuming there is no depreciation net of usual maintenance). Davis and Heathcote (2007) find that spending on structural improvements in the US (roughly 0.8 percent) is very similar to a reasonable estimate of structural depreciation (roughly 1 percent). Our conclusions using vintages reinforce the idea that changes in dwelling characteristics have not driven the changes in housing inequality that we document.

in terms of structural characteristics. Additionally, neither location nor dwelling characteristics do much to explain changes in household wage inequality (shown in the last column).

Figure A.6: Impacts of Observable Structural Characteristics on Inequality




NOTE: Measures of inequality are mostly unchanged by changes in the observed structural characteristics of houses included in the Decennial Census. Lines in black refer to re-weightings of the 2012 data, following DiNardo, Fortin and Lemieux (1996) and Fortin, Lemieux and Firpo (2011), to match the distribution of structural characteristics observed in each Decennial Census. The line in grey shows the values of the statistic in the data for each year. Each panel refers to the different labeled variables and sub-samples.

# D The Relationship between Income and Housing Inequality

Our results suggest a link between local income and housing inequality. Studies of housing demand (e.g. Polinsky and Ellwood (1979), Hanushek and Quigley (1980), and Mayo (1981)) have generally concluded that housing is a necessity, a regularity sometimes known as "Schwabe's Law." Most estimates of the income elasticity of demand for housing are between 0.3 and 1.0, so variations in income should be reflected in housing consumption. In fact, other things equal, the variance of log housing consumption should equal the square of the income elasticity of housing times the variance of log income.[31]

[31] This exercise carries several caveats: General equilibrium effects may interact with consumer preferences to either dampen (Matlack and Vigdor (2008)) or amplify (Van Nieuwerburgh and Weill (2010)) the direct effects of changes in incomes. Additionally, since housing is so heterogeneous, it is difficult to quantify how much "housing service" a house of a given size or in a given neighborhood provide.

Table A.2: Re-weighting

Panel A: Theil entropy

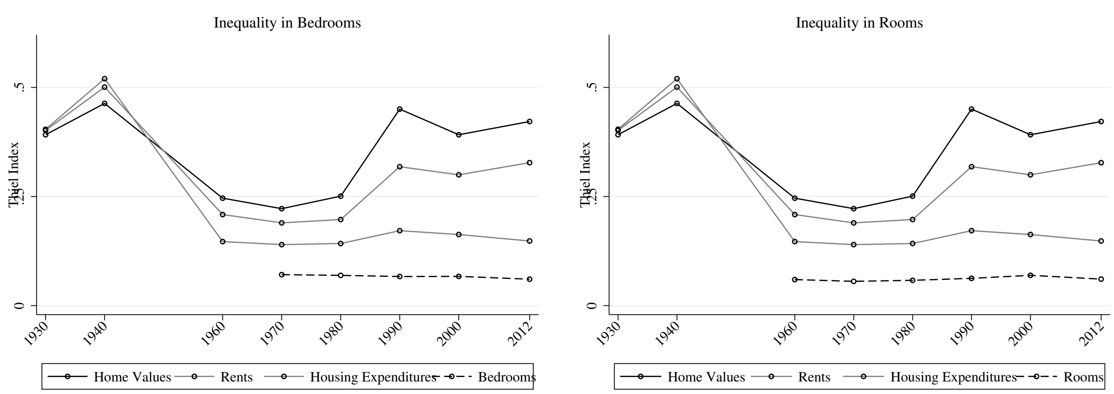
| | Rent and Equiv | Rents | Home Values | HH Wages |
|---|---|---|---|---|
| 2012 | 0.328 | 0.148 | 0.422 | 0.367 |
| Dwellings | 0.298 | 0.150 | 0.391 | 0.354 |
| Dwellings and State | 0.290 | 0.155 | 0.387 | 0.356 |
| Dwellings and CZ | 0.298 | 0.160 | 0.403 | 0.359 |
| Vintages Specification | 0.311 | 0.156 | 0.459 | 0.356 |
| 1970 | 0.190 | 0.140 | 0.222 | 0.194 |

Panel B: Variance of logs

| | Rent and Equiv | Rents | Home Values | HH Wages |
|---|---|---|---|---|
| 2012 | 0.540 | 0.507 | 0.977 | 1.189 |
| Dwellings | 0.535 | 0.540 | 1.001 | 1.177 |
| Dwellings and State | 0.535 | 0.550 | 1.046 | 1.184 |
| Dwellings and CZ | 0.549 | 0.569 | 1.097 | 1.197 |
| Vintages Specification | 0.503 | 0.521 | 0.990 | 1.194 |
| 1970 | 0.440 | 0.367 | 0.609 | 0.857 |

NOTE: Changes in observable dwelling characteristics cannot explain increases in housing inequality. Statistics are taken of 2012 data, re-weighted to emulate the distribution of observable dwelling and location characteristics of previous years. The top and bottom row refers to the actual values in the specified years. Dwellings refers to: Indicators for the number of rooms, bedrooms, the decade of construction, plumbing facilities, and the heating system type. States and CZs include indicators for the geographic entities, and demographics is the interactions of the household type (single, a couple, a single parent, or non-related individuals) with indicators for the number of children and adults (separately). The solid line contains weights accounting for all of these which were computed as noted in the text, following DiNardo, Fortin and Lemieux (1996) and Fortin, Lemieux and Firpo (2011).

Figure A.7: Inequality in the Number of Rooms and Bedrooms Compared to Values



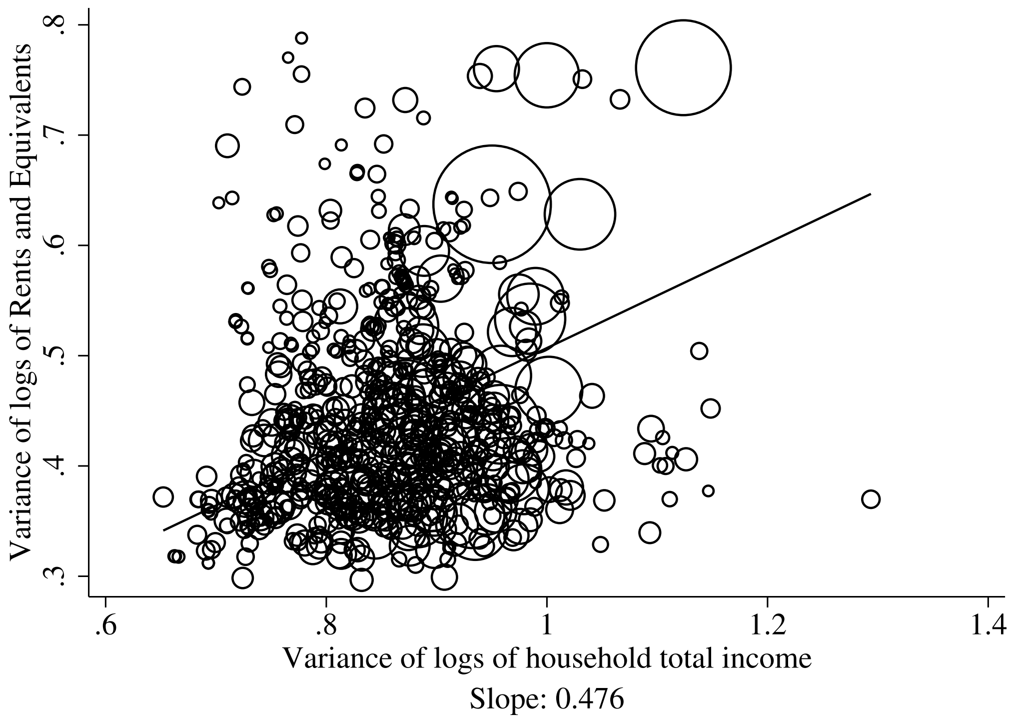

NOTE: Inequality in the size of houses has not meaningfully increased since 1960. The Dashed line plots a Theil Index of Inequality in the number of rooms or bedrooms, while the solid lines plot inequality in housing values and rents. Inequality in the number of rooms and bedrooms is for the whole sample of houses – among both owners and renters. The solid lines are the same as the Theil indexes plotted in Figure 1.

Patterns of income and housing inequality over time and space suggest a considerable relationship between the two. Between 1970 and 2012, the variance of log income increased by 0.26, while the variance of log housing consumption increased by 0.09, consistent with an income elasticity of 0.59. As with the tract-level results, we explore the cross-sectional relationship between housing and income across cities in Table A.3 and Figure A.8.

Figure A.8 shows a positive relationship between demeaned log real household-level income and rents and equivalents, in keeping with an elasticity of roughly 0.8 in 2012. The horizontal axis shows the variance of the log of household total incomes and the vertical axis shows the variance of logged rents and their equivalent. Each dot represents a commuting zone and the radius is proportionate to the population of the commuting zone. Overall there is a strong positive correlation of 0.48, which implies an elasticity of 0.69 – matching previous micro estimates.

Additionally, Table A.3 shows that the relationship plotted in Figure A.8 is stable over time. The table shows implied elasticities taken by computing the square root of the coefficient of the relationship between the variance of the log of total household incomes and the variance of the log of rents and their equivalent across commuting zones. Going back to 1970, the first year where we have a similar measure of total household income, the relationship continues to match previous estimates of the elasticity of housing demand with respect to income since it remains between 0.5 and 1.

Figure A.8: Income and Housing Inequality across Commuting Zones



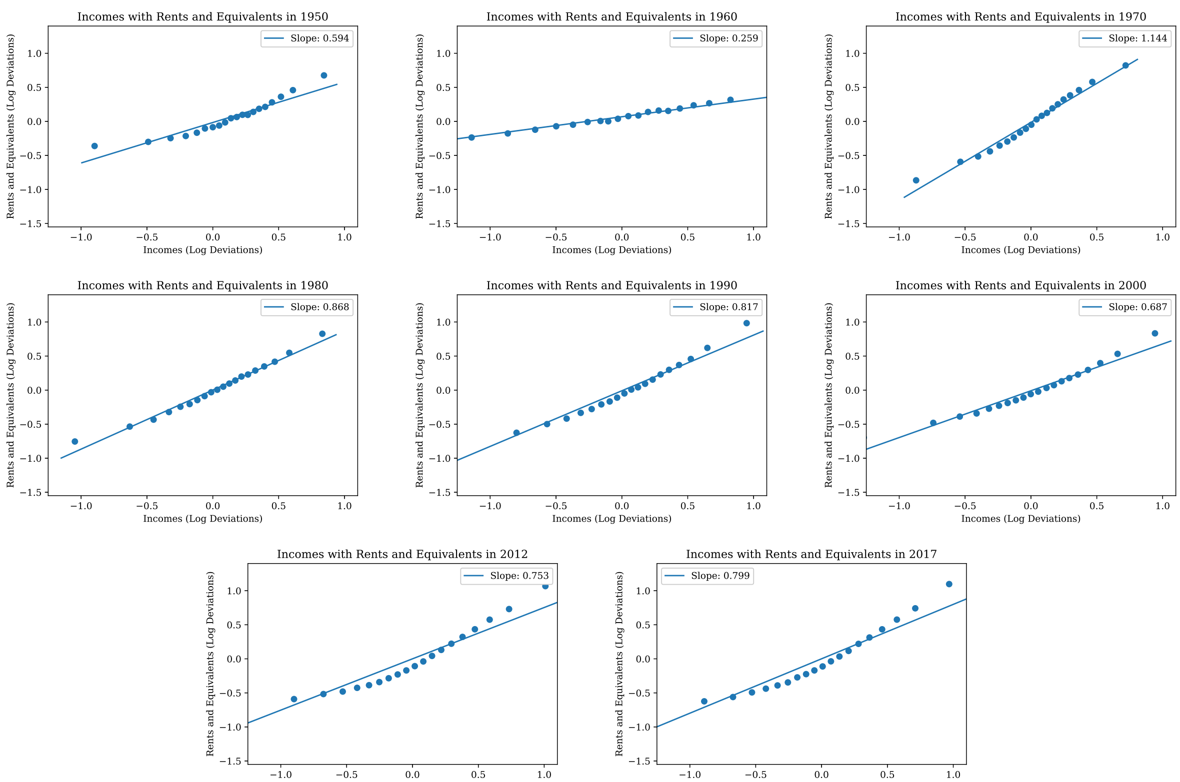

NOTE: Inequality in incomes accompanies inequality in rents and their equivalent across commuting zones. The horizontal axis shows the variance of the log of household total incomes and the vertical axis shows the variance of logged rents and their equivalent. Each dot represents a commuting zone and the radius is proportionate to the number of (weighted) households in each commuting zone. The OLS regression line plotted incorporate these weights.

Table A.3: Implied Household Income Elasticities

| | Rents and Equivalents |
|---|---|
| 1970 | 0.95 |
| 1980 | 0.66 |
| 1990 | 0.54 |
| 2000 | 0.61 |
| 2012 | 0.69 |

NOTE: Each coefficient is the (re-signed) square root of the absolute value of the regression coefficient relating the variance of the log of household wage income with the variance of the log of rents and their equivalent within each commuting zone.

Figure A.9: Neighborhood Incomes and Housing Prices: All Available Years

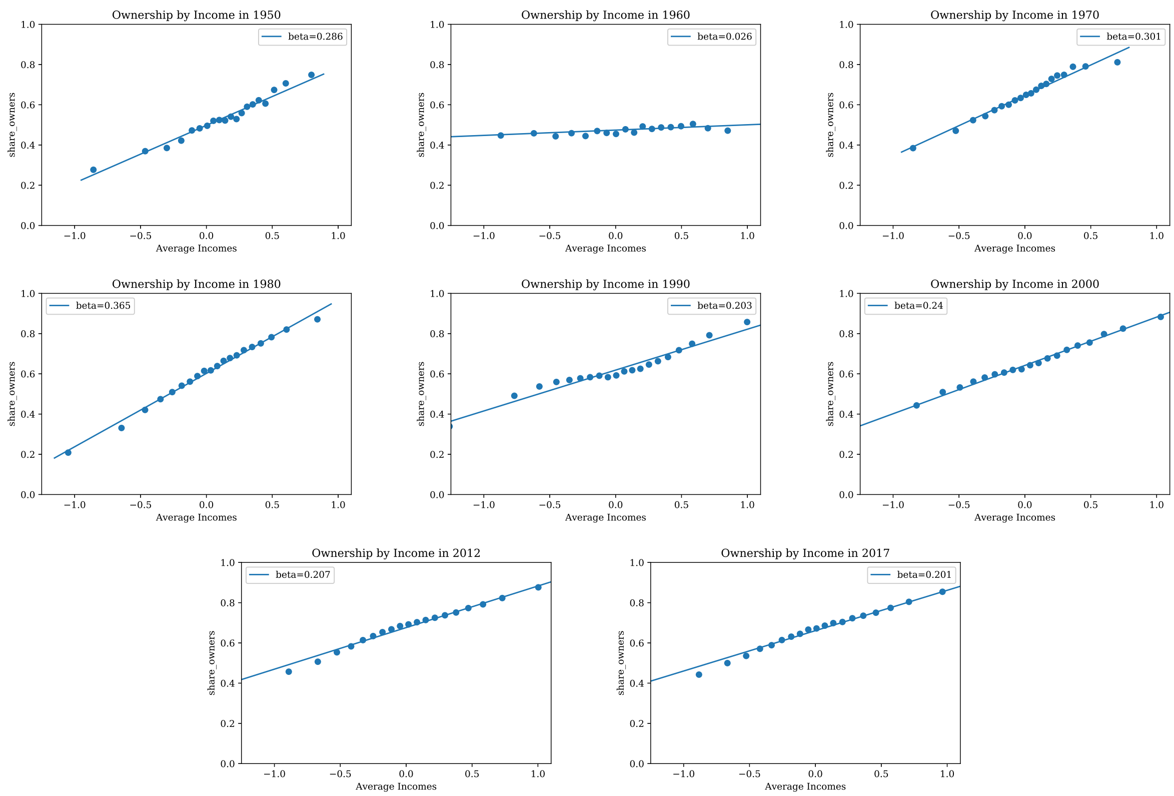



NOTE: The figure shows the robustness of Figure **??** by including additional years.

# E Neighborhood Home Ownership Rates

One important factor the keep in mind for the analysis of neighborhoods is that owner-occupied housing tends to be in higher income neighborhoods. The fact is apparent in Figure A.10, which plots the a binned scatterplot of share of houses that are owner-occupied (as opposed to rented) in a neighborhood compared with the average of logged incomes in the neighborhood (in terms of deviations from the national average). The relationship between income and ownership is generally stable. However, the spreading of the distribution of income across neighborhoods leads to very high ownership rates in the highest income neighborhoods in later years.

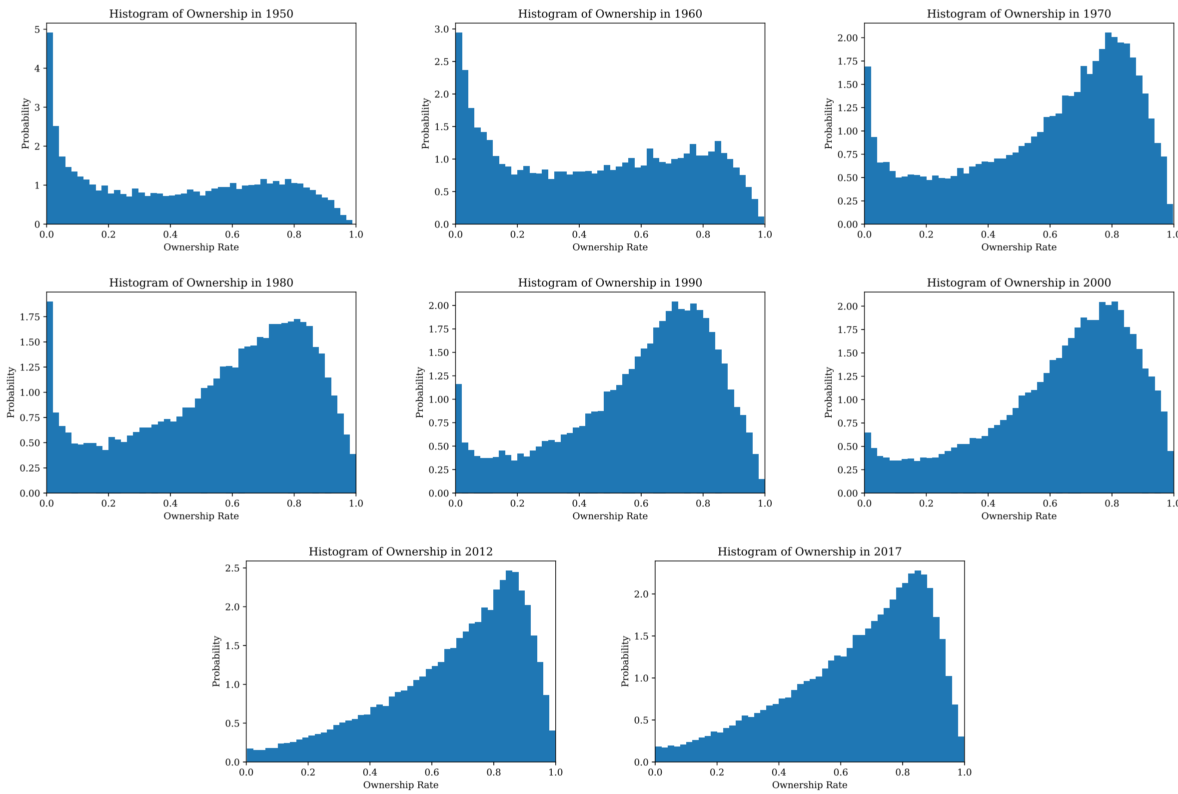
Figure A.10: Neighborhood Incomes and Home Ownership




NOTE: Home ownership rates are higher in higher income areas, and the relationship is relatively stable over time. These panels display binned scatter plots of neighborhood level average incomes and home ownership rates, based on the number of rental and owner occupied units listed in Census summary tables. Note that there are compositional changes in each year in terms of both the neighborhoods involved and the housing units covered (e.g. condos are excluded in 1970). Incomes are logged, averaged, and put into deviations by subtracting the average of all logged incomes. The reported beta is for the line of best fit in the underlying data.

So many neighborhoods are predominantly composed of owner occupied housing, particularly in later years. Figure A.11 plots average home ownership rates by neighborhood for each year in the sample. In early years home ownership rates are relatively evenly distributed, with a large spike at zero. This could be due to the prevalence of city centers in the sample, and also the lower rate of home ownership. In later years, culminating in 2017, however, there is an especially large concentration of neighborhoods with home ownership rates very close to one.

Figure A.11: Home Ownership Histogram




NOTE: The sample of neighborhoods and housing units covered by tract level summary tables has become more focused on owner occupied housing, and with a higher concentration of neighborhoods with very high home ownership rates in later years. This may be due to increasing coverage of rural areas and to changes in the sample, to a lesser degree. These panels display a histogram of home ownership rates across neighborhoods covered by census tract level summary files across the years in the analysis. it is weighted by the number of housing units (houses and apartments) in each neighborhood. Note that there are also compositional changes in terms of the housing units covered (e.g. condos are excluded in 1970).

# F Role of Housing in Driving Wealth Inequality

Growing inequality in house prices from 1970 onwards, documented in the previous sections, likely maps into growth in inequality in overall wealth studied in the literature. Moreover, since housing assets are used as collateral for mortgages, home equity loans, and lines of credit, households with housing assets may purchase them by increasing leverage. As such, inequality in net positions in housing may show larger inequality than those in gross positions.[32] To provide insight into the quantitative importance of housing and housing-related debt in driving wealth inequality, we use household-level data from the SCF which provides a comprehensive picture of household balance sheets throughout the wealth distribution. Specifically, the data allow us to both net out housing debt held against housing assets and compute the fraction of overall net worth comprised of housing

[32] When a household owns a home with a mortgage, the mortgage issuer has a claim on the gross housing wealth. These claims may, in turn, be sold or further collateralized and eventually held by other household, likely higher up the wealth distribution. We focus here on home equity rather than indirect claims on housing assets for two reasons. First, data on indirect claims are not readily available. Second, since houses are primarily purchased using debt, capital gains always accrue to homeowner alone.

wealth at various points in the distribution.

Using data from the SCF, Figure A.13 plots the Gini coefficient, Thiel index, and the standard deviation of logs for gross housing wealth, home equity, and net worth.[33] As with the Census data, results from the SCF suggest inequality in gross housing wealth has risen since 1983. The increase in inequality based on home equity appears to have been somewhat larger than the increase based on gross housing wealth, suggesting that homeowners with lower overall wealth also have higher leverage. While inequality in both net and gross housing wealth has risen, the change in inequality in overall net worth appears somewhat larger.

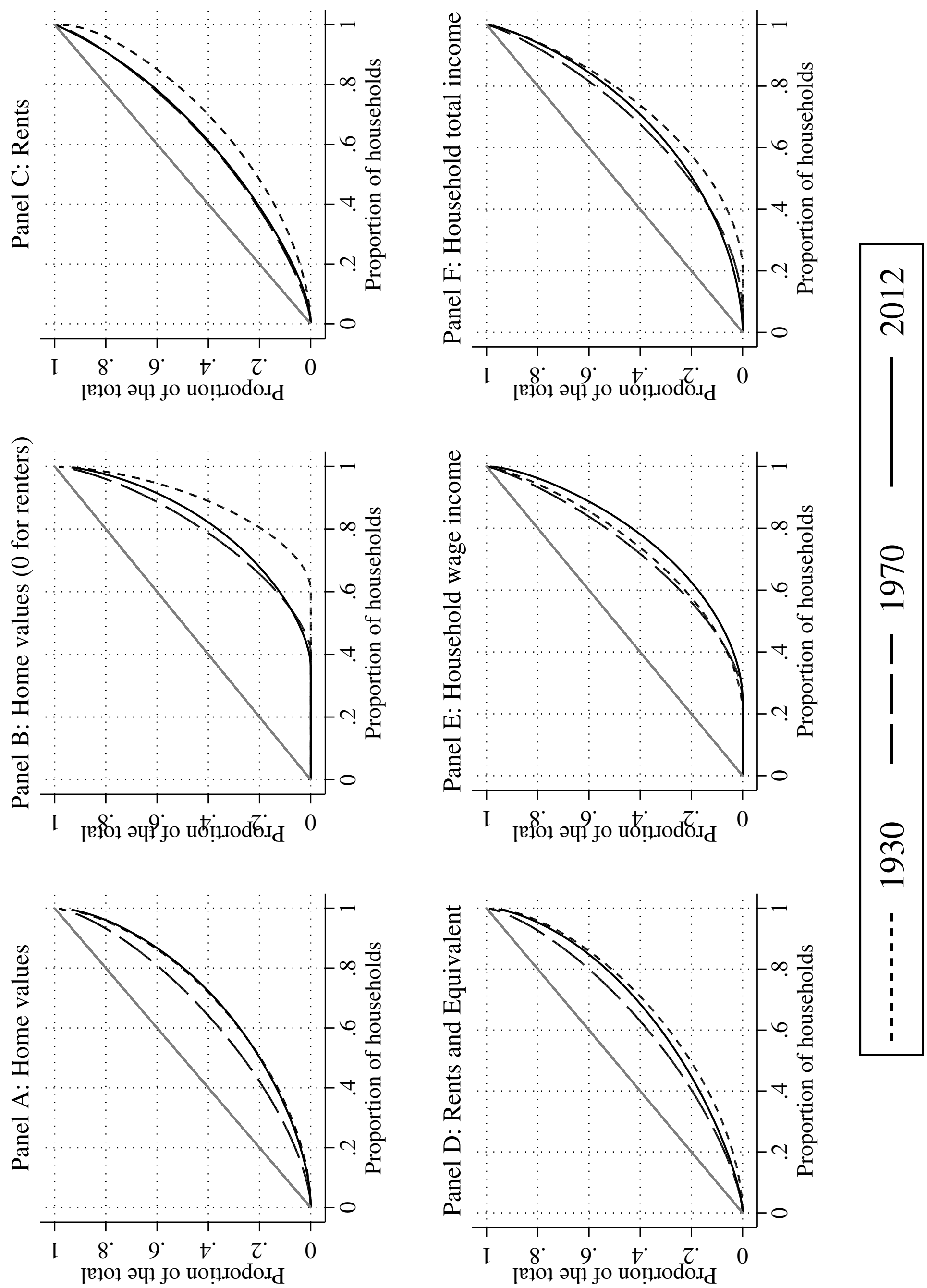
To study the role of housing in driving rising wealth inequality, Figure A.14 plots Lorenz Curves for overall net worth along with the contributions from net equity in the household's primary home and net equity in other real estate. The overall Lorenz curves in each year (solid lines) are computed as the fraction of aggregate net worth held by the lowest *pth* percentile of the net worth distribution. The shaded regions show the fraction of this group's net worth comprised of home equity in their primary residence and other real estate respectively. At the 100th percentile, the charts show home equity makes up just over 20 percent of overall net worth in the SCF, broadly consistent with aggregate national statistics such as the Flow of Funds Accounts.

In the cross-section, home equity explains much of the inequality in housing wealth at the middle of the wealth distribution, where housing wealth comprises a larger fraction of overall net worth. Households at the top of the distribution are more likely to hold larger shares of their overall wealth in other non-real-estate assets. Looking over time, housing wealth's contribution to overall wealth inequality appears to have diminished somewhat since 1983. This is evident from the fact that the shaded region below the dashed line has fallen by more than the solid line at each quantile. Moreover, the Gini coefficient for overall net worth can be decomposed into its contribution from home equity as in Lerman and Yitzhaki (1985). Doing so suggests that home equity explained roughly 20 percent of inequality in net worth in 1983, but has fallen to explaining only 16 percent in 2013. The decline occurs because the decline in the share of net worth coming from home equity outweighs the increase in inequality within home equity over this period.

[33] All housing statistics apply to primary residence only. Gross housing wealth is the self-reported house price for owner-occupied homes. Home equity is gross housing wealth less mortgages, home equity loans, and home equity lines of credit secured with the respondent's primary residence. Net worth is home equity plus all financial and nonfinancial assets, less all household debts. Imputed values of pensions are included in all waves except 1983 and 1986.

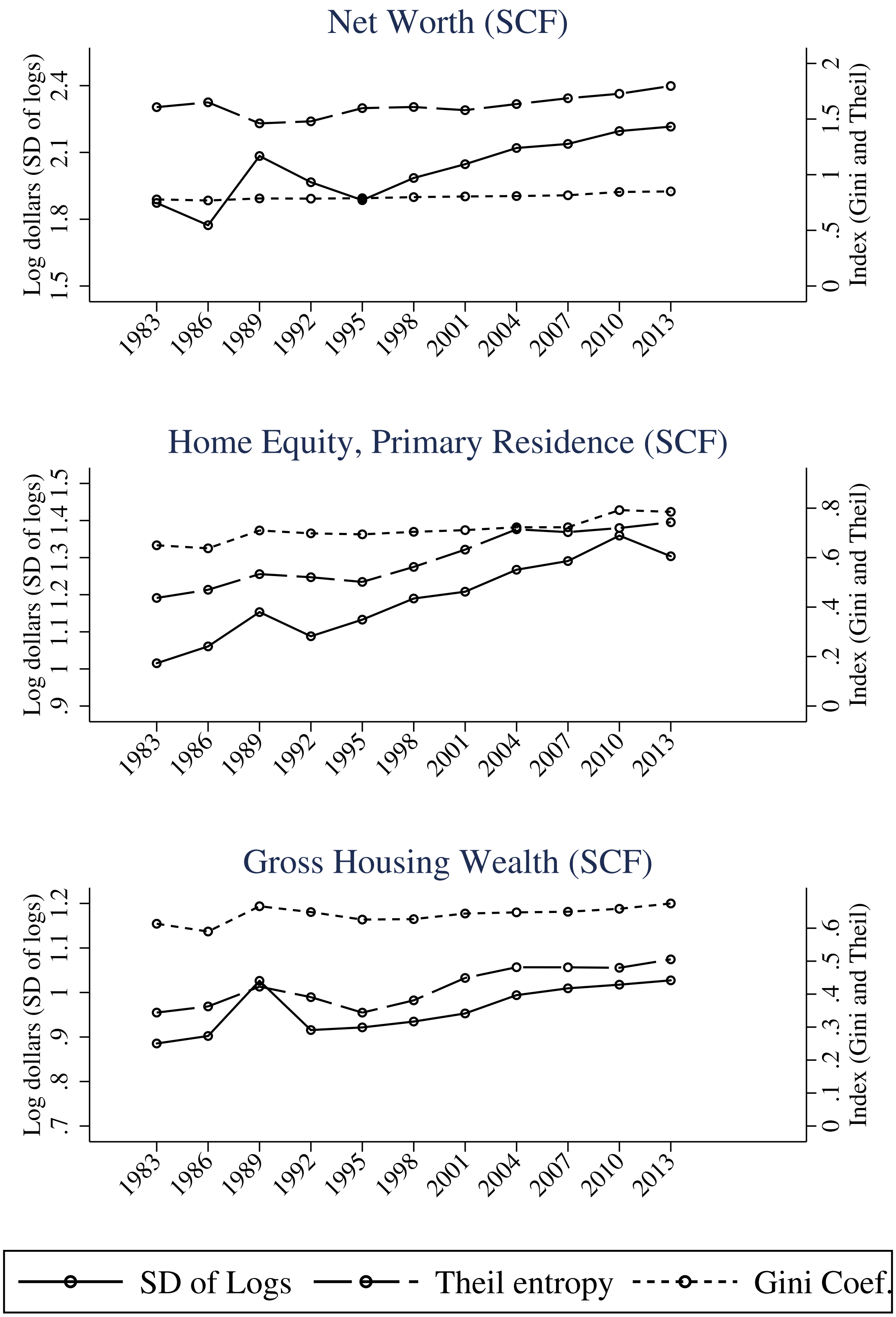
# G Additional Appendix Figures

Figure A.12: Lorenz Curves for Home Values, Rents, Housing Consumption, and Household Income




NOTE: These curves graph the cumulative percentage of housing expenditures (from lowest to highest), against the cumulative percentage of households. Panel A, is for owning households; Panel C is for renters. All others are for the sample of home owners and renters combined. Panel B includes renters with an implied home value of zero. Data are interpolated within intervals as described in the text. Data are from the Decennial Census and the 2009-2012 ACS. See the text and data appendix for more detail.

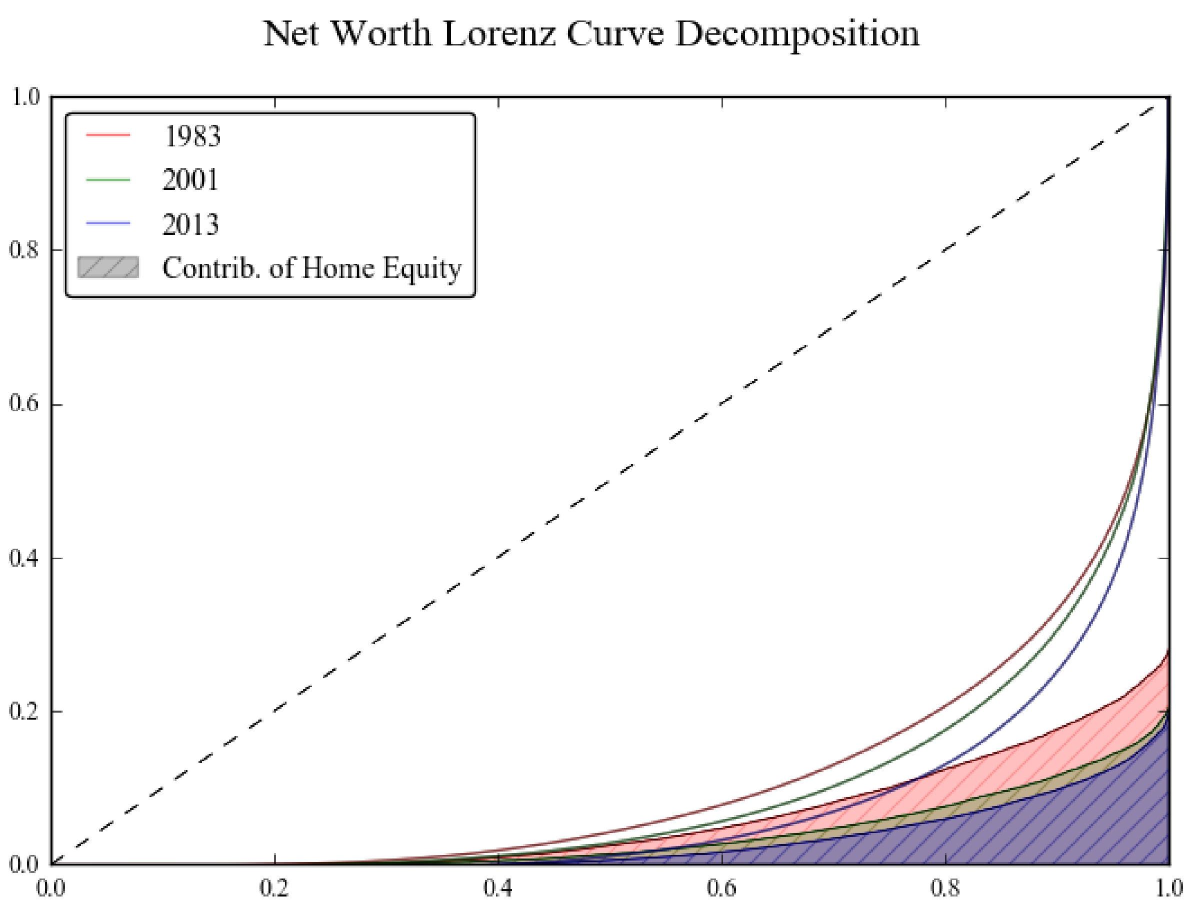
Figure A.13: Inequality in Home Equity






NOTE: Each panel presents inequality measures for overall net worth, home equity (net housing wealth), and gross housing wealth in the full SCF sample by wave.

Figure A.14: Contribution of Home Equity to Lorenz Curves for Net Worth




NOTE: These curves graph the cumulative percentage of total net worth held by households in various percentiles of the wealth distribution. The area under the curve is further decomposed into the contribution from home equity in primary homes (shaded hatched region) for each of the three waves drawn.