Housing Inequality in the United States*

Gerard Domènech-Arumí† Lei Ma‡ Giovanni Paolo Mariani§

December 15, 2025

Abstract

We use county assessor and property deed records to construct, analyze, and disseminate contemporaneous and long-run estimates of housing inequality for the United States and its subnational levels. A dwelling in the top decile is eight times as valuable and four times as large as a dwelling in the bottom decile, and 12% of all housing assets are held by 1% of households. Housing value inequality has increased since the 1980s, while housing space inequality has remained relatively stable. Heterogeneity is substantial, with considerable variation between and within states, commuting zones, and census tracts. Income, land-use regulations, and housing supply and demand are associated with current and changing levels of housing inequality. This research highlights the importance of granularity when studying inequality and evaluating policies. We make all estimates publicly available to facilitate such work.

Keywords: Inequality, Gini, Housing
JEL Codes: D31, D63, O18, R21, R31

*We thank Elizabeth Cascio, Micael Castanheira, Paula Gobbi, Nathan Nunn, Gianluca Violante, Étienne Wasmer, and seminar and conference participants at the Workshop on Life-course Inequality Dynamics, and The Socio-Economics of Overcrowding Berlin Workshop for helpful comments and feedback. We thank Federico and Alessandro Gallina for logistical support developing the WHID website. Gerard Domènech-Arumí gratefully acknowledges funding from Fondation ULB. Giovanni Paolo Mariani gratefully acknowledges funding from the Fund for Scientific Research–FNRS (FC 44291). All errors are our own.

†Department of Economics, Vanderbilt University. E-mail: gerard.domenech.arumi@vanderbilt.edu
‡Terry College of Business, University of Georgia. E-mail: lei.ma@uga.edu
§ECARES, Université Libre de Bruxelles. E-mail: giovanni.paolo.mariani@ulb.be

1 Introduction

Housing is multifaceted. It is an asset – the main asset for most households – making it central to understanding wealth inequality (OECD 2024). It is a consumption good that provides shelter and an environment for raising a family, and it constitutes the largest expenditure category for most households, making it crucial to understanding income and consumption inequality (Diamond and Moretti 2021, Dustmann et al. 2022, Moretti 2013). It is also the "key to neighborhoods," shaping access to local labor markets, schools, and other public goods and amenities, and is therefore essential to understanding inequality of opportunity and social mobility (Chyn and Katz 2021, Ioannides and Ngai 2025). Housing is complex and central to explaining current and future inequality.

We know relatively little about housing inequality. There has been a myriad of studies on inequality measurement over the past two decades, and recent work recognizes the role of housing in explaining wealth inequality (Blanchet and Martínez-Toledano 2023), and accounts for imputed rents in income inequality (Piketty et al. 2018). However, with the notable exception of Aladangady et al. (2017), measuring housing inequality itself has rarely been the primary focus. In this paper, we use administrative data from county assessors and property deeds, covering nearly the universe of dwellings, to construct, analyze, and disseminate contemporaneous and long-run estimates of housing inequality for the United States and its subnational levels.

We obtain detailed property characteristics from CoreLogic's county assessor files. County assessor offices across the United States report extensive information on properties within their jurisdictions. These administrative records are typically richer and more accurate than survey-based census data, cover the full property universe, and are publicly available. However, in the absence of federal reporting standards, the same information is disseminated in many different formats across states and counties. CoreLogic, a private vendor, retrieves and harmonizes records from more than 3,000 counties across the United States at the parcel level. We then further process the data to harmonize it at the dwelling

level, ensuring that key attributes (e.g., square footage) are comparable across 91% of counties nationwide, covering 98% of the U.S. population.

CoreLogic also provides detailed transaction histories for individual properties, including sale prices and transaction dates. Because only a subset of homes are transacted in a given year, transaction data provide an incomplete picture of housing values. To address this limitation, we use a machine-learning model to impute the annual market values of all housing units for 1980-2020. We train the model on observed transaction prices and property characteristics, then predict market values for non-transacted properties. This approach allows us to construct a consistent, nationwide panel of housing values over time and across space.

We use the harmonized dwelling-level dataset to construct and analyze measures of housing attributes and their distributions. We examine three dimensions of housing: (1) value, defined by imputed market values for all residential units; (2) space, measured by each unit's square footage; and (3) assets, based on the distribution of housing assets across households (homeowners and renters). For each dimension, we compute a set of inequality metrics (e.g., the Gini coefficient) at multiple levels of aggregation, from the national level down to the census block. In this paper, we focus primarily on the Gini coefficient and selected distributional moments (e.g., the median and the 10th and 90th percentiles), and our descriptive analysis emphasizes levels and trends at the county level or above. Additional inequality estimates (e.g., the coefficient of variation and other moments) and aggregation levels (e.g., commuting zones, school districts) are publicly available as part of the World Housing Inequality Database (WHID) (www.worldhousingdata.com), released alongside this paper and two companion papers covering Belgium and Spain (Domènech-Arumí and Mariani 2025, Domènech-Arumí and Gobbi 2025).

Housing inequality in the U.S. is significant. A dwelling in the top decile is worth eight times as much ($612,000 vs. $76,000 in 2020) and is about four times as large (3,000 vs. 800 square feet) as one in the bottom decile. The corresponding Gini coefficients for value

and space are 0.50 and 0.35, respectively – both substantially higher than in Spain and especially Belgium (Domènech-Arumí and Mariani 2025, Domènech-Arumí and Gobbi 2025). Our estimates are comparable to Aladangady et al. (2017), who, using Census data, report a housing-value Gini just under 0.50 in 2012. Housing assets are even more concentrated. We estimate that 1% of households hold 12% of housing assets, implying a Gini of 0.68. Despite this concentration, housing assets are more evenly distributed than overall wealth (Chancel et al. 2022).

Geographic heterogeneity is substantial. As with individuals, dwellings are unevenly distributed across space. This implies that the distribution of housing values, space, and assets varies markedly between and within states, counties, and even census tracts. Our estimates confirm the high degree of heterogeneity and identify New York and, especially, California as disproportionately contributing to nationwide inequality. For example, a dwelling in the 99th percentile is worth 9.3 million (2020 USD) in California but only USD 320,000 in West Virginia. Although less stark, within-state variation is also large in California, where a dwelling in the Bay Area costs substantially more than one in a rural county in the Central Valley. When measuring housing inequality, aggregation matters.

Housing value inequality has increased since 1980. Using the Gini coefficient, we estimate that inequality rose from 0.34 in 1980 to 0.50 in 2020. Most states experienced increases over this period, but California stands out for a rapid rise in the early 2000s driven by high price growth across the distribution, especially at the top. By contrast, housing space inequality – which, with important caveats, we can measure since 1900 – has remained relatively stable nationwide and in most states throughout the period.

We examine several determinants of current and changing levels of housing inequality in the final sections of the paper. We find that income is strongly associated with housing prices and that income inequality moderately predicts housing value and space inequality, suggesting that housing prices in a geography are informative about income levels. Zoning and land-use regulations also matter. In an application, we find that more stringent

land-use regulations are associated with higher prices and greater inequality between cities, but lower inequality within cities, suggesting that zoning makes neighborhoods more homogeneous. Finally, we argue and provide descriptive evidence suggesting that changes in the distribution of housing characteristics and in housing supply and demand, particularly over business cycles, are central to the dynamics of housing inequality.

This research primarily contributes to the literature on inequality measurement. A large body of work has used survey and administrative tax data to measure inequality across the globe and over time (e.g., Chancel et al. 2022, Guvenen et al. 2022, Piketty and Saez 2003, Piketty and Zucman 2014, Saez and Zucman 2016). Despite recognizing housing as an important driver of (particularly wealth) inequality, the focus of this literature has rarely been on housing itself. That has changed in recent years. For example, Derenoncourt et al. (2024) identifies housing as a major determinant of racial disparities in wealth accumulation in the US. Kermani and Wong (2021) show that distressed sales affect racial minorities disproportionately, exacerbating racial inequalities. Knoll et al. (2017) and Lyons et al. (2025) use a variety of sources to measure long-run trends in housing prices across different countries, and Aladangady et al. (2017) – the paper closest to ours – uses historical census data to study housing value inequality in the US.

We contribute to this growing literature by providing novel housing estimates for the entire U.S., constructed from county assessor and property deed records. A key advantage of our approach relative to past work is that our estimates are based on geolocated, property-level microdata that are often more detailed than census data. For example, our data contains the names and addresses of owners. Relatedly, the nature of these data allows us to produce perfectly comparable distributions of housing values, space, and assets, at virtually any level of aggregation, from the nation down to the census block. This emphasis on granularity connects our work to the extensive literature on segregation measurement (e.g., Echenique and Fryer Jr 2007, Fogli and Guerrieri 2019, Logan and Parman 2025). After all, local inequality and segregation are opposite sides of the same

coin (Glaeser et al. 2009). Given the substantial geographic and temporal heterogeneity we document, the nature and granularity of our estimates – which we make publicly available through the WHID project – constitute our primary contribution. We hope our work will support future research and policy discussions on inequality and housing.

We also contribute to the literature on the determinants of inequality. We provide new evidence on several established facts. Consistent with Aladangady et al. (2017), we find that income is a major determinant of housing prices and inequality. Zoning and land-use regulations are crucial as they appear to raise housing prices (Büchler and Lutz 2024, Glaeser et al. 2005a;b), increase inequality between cities, and reduce it within them. As argued by Howard and Liebersohn (2025), housing prices have risen faster in high-income-growth cities because housing supply could not keep pace with faster demand growth in dense, tightly regulated environments – which may have had distributional consequences by hindering low-skilled migration to high-opportunity places (Ganong and Shoag 2017). Our results also indicate that housing space inequality is substantially lower than housing value inequality, consistent with the notion that *real* inequalities (accounting for cost-of-living differences) in America may not be as stark based on *nominal* inequalities (Albouy et al. 2016, Diamond and Moretti 2021, Moretti 2013). At the same time, land prices reflect local public goods and amenities (Couture et al. 2024, Diamond and Gaubert 2022, Gaubert and Robert-Nicoud 2025), and differential exposure to these across space is crucial for understanding inequality and segregation dynamics (Ioannides and Ngai 2025). Our findings on price dispersion per square foot are consistent with this perspective and confirm that prices have increased significantly faster in high-income cities (Davis et al. 2021, Van Nieuwerburgh and Weill 2010). While we cannot precisely pin down all key determinants of (housing) inequality, our work identifies several and sheds light on how different dimensions of housing inequality are linked to commonly studied measures such as income and wealth.

2 Data

2.1 Property data

County assessors collect information and assess the value of all properties in their counties for taxation purposes. These administrative records are rich and typically publicly available online, but their decentralized nature and the sheer number of counties (over 3,000) make country-level analysis challenging. CoreLogic, a private vendor, collected and partially harmonized these data. The preprocessed assessor data we obtained from them cover nearly the entire universe of U.S. properties as of 2020.

The unit of observation in the raw data is the parcel – a piece of land with legally defined boundaries used for taxation, ownership, and development. For this paper, the key variables we use are a unique parcel identifier, location, property type, the number of housing units, and the square footage of the residential area. The parcel identifier and location allow us to produce estimates at any level of aggregation – from the country down to the census block. Information on property type allows us to distinguish dwellings (single-family homes, condos, and apartments) from other types of real estate (e.g., office buildings). We keep all residential properties in our sample. The number of housing units enables us to construct a dataset in which the unit of observation is the dwelling. Residential space is both a variable of interest and a key input to the model we use to predict housing prices (see Section 2.2). We also observe owners' names and mailing addresses, which we use to determine owner-occupancy and to measure the distribution of housing assets.

Despite CoreLogic's preprocessing, the data still required substantial work on our end before becoming usable for this paper, as reporting standards and quality vary markedly across and within states. This is particularly true for multi-family homes and apartment buildings, which often lack information on dwelling-level characteristics. For instance, a large portion of the additional data processing work involved correctly inferring the

number of properties per parcel – a crucial input for transforming the raw data to the dwelling level. Appendix B discusses this process in great detail and provides a series of data validation checks. For example, Table B1 indicates that, based on our estimates, the United States had 144.4 million housing units in 2020, compared with 142.4 million in the 2020 Census – an overestimate of 2.0 million units (1.4%), suggesting that our approach captures the national housing stock reasonably well. Accuracy varies across states. While the margin of error is under 15% in most, it is sizable in others. Wisconsin is the worst-performing state, with an underestimation of 1.7 million units (a 66% error), driven by the lack of reported unit counts or sufficient information to impute them in some counties. In contrast, in places like Hawaii, our data matches the number of units perfectly. We believe variation in the reporting standards for multi-family buildings explains most of the cross-state variation in accuracy. After processing the data, our sample includes 91% of counties and covers 98% of the U.S. population.[1]

2.2 Real estate transactions and dwelling value estimations

County assessor data contain the assessed value of properties used to determine property tax liabilities. Ideally, assessed values track market prices. In practice, however, there is a disconnect between the two, which varies across and within jurisdictions due to differing appraisal methods (Avenancio-León and Howard 2022, Berry 2021).

We obtained real estate transaction data covering nearly the entire United States from 1980 to 2020. The data originate from county deed records, obtained via CoreLogic. The unit of observation is a transaction, which can be linked to the assessor data using the parcel identifier. We exclude non-arm's-length transactions and distressed sales, as these occur under special circumstances and their sale prices do not reflect market value. Our sample contains 149.6 million transactions.

[1]The 9% of counties missing from our data are excluded because county assessors do not report sufficient property information. At a minimum, we require housing space (e.g., square footage).

We use transaction data to estimate the market value of nearly all U.S. dwellings from 1980-2020. We train state-specific random forest models using transaction records and 12 key property features, including space (property square footage and acres in the lot), number of rooms, number of bathrooms, and year of construction and transaction, as well as location covariates (latitude and longitude) and categorical indicators for county and census tract.

Our models achieve an average prediction error of 0.18, measured by the out-of-bag root mean squared error (OOB-RMSE), indicating that dwelling value predictions are, on average, off by 18%. This performance is consistent with prior research in similar applications (Čeh et al. 2018, Mullainathan and Spiess 2017). The model slightly overestimates values at the bottom and slightly underestimates them at the top, implying that our housing-value inequality estimates may be slightly biased downward. In terms of predictive power, location and housing space are the most important predictors of dwelling value. Appendix B provides additional details and validation checks.

We use the trained model to impute a price for each of the over 140 million dwellings in the United States for each year between 1980 and 2020. We are unable to impute a price for 4.2 million dwellings (3% of the stock) because some county assessors do not provide sufficient information to estimate a dwelling's value with reasonable confidence.

2.3 Measuring housing inequality

We focus on three dimensions of housing: value, space, and assets. **Housing value** is the market value of a dwelling, estimated using transaction data, as described above. **Housing space** is the dwelling's floor area, measured in square feet. **Housing assets** are the dollar value of all residential properties owned by an individual. We can measure assets because the assessor data include owners' full names and mailing addresses, allowing us to link each dwelling's value to a specific owner-location. For example, if John Snow owns one dwelling worth $100,000 in 1000 Broadway, Nashville, TN (where he lives) and a second

dwelling worth $200,000 in 2000 W Jefferson St in Phoenix, AZ (which he rents), then he holds $300,000 in housing assets. If the Arizona property is occupied, we assume a renter (with $0 in housing assets) lives there. A caveat is that we do not distribute the value of properties owned by non-natural persons (e.g., real estate investors, limited liability companies, and trusts). Because wealthier individuals are more likely to hold real estate through trusts or LLCs, our estimates of housing assets and asset inequality are likely lower bounds.[2]

Inequality can be measured in many ways. Below, we focus primarily on the Gini index due to its widespread use in academic and policy settings. We also examine specific points in the distributions (e.g., the median) to better understand and complement inequality levels and trends. The public dataset released with this paper includes additional inequality metrics, such as the Theil index.

3 Housing inequality in the United States

3.1 Housing inequality in 2020

Housing value. Housing value inequality is high. Panel A of Table 1 shows that the value of a dwelling at the 10th percentile (P10) is $76,000 (2020 USD).[3] A dwelling at the median (P50) is worth $220,000, while dwellings at the 90th (P90) and 99th percentiles (P99) are worth $612,000 and $2.3 million, respectively. The distribution's corresponding Gini index is 0.50. For reference, this is significantly higher than the equivalent figure for Belgium in 2022 (Gini of 0.26) (Domènech-Arumí and Gobbi 2025), and the Catalonia region in 2019 (Gini of 0.37) (Domènech-Arumí and Mariani 2025). It is comparable to the estimate in Aladangady et al. (2017) for 2012 using Census data (approximately 0.47) and close to the

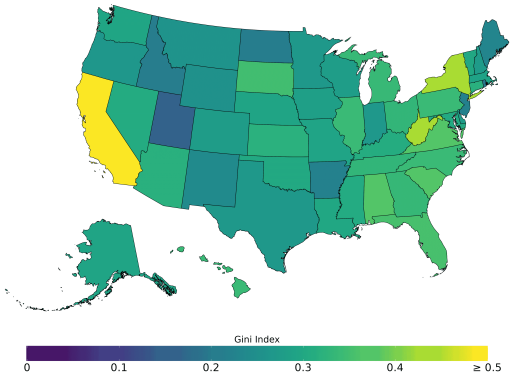
[2]Consistent with this, our estimated homeownership rate is 57%, whereas the official 2020 rate is 65% (U.S. Census Bureau n.d.). See Section B.5 in the Appendix for more details on our measurement of homeownership.

[3]This measure is calculated from the distribution of dwelling values in the United States. The 10th percentile (P10) of housing values is the value below which 10% of homes fall.

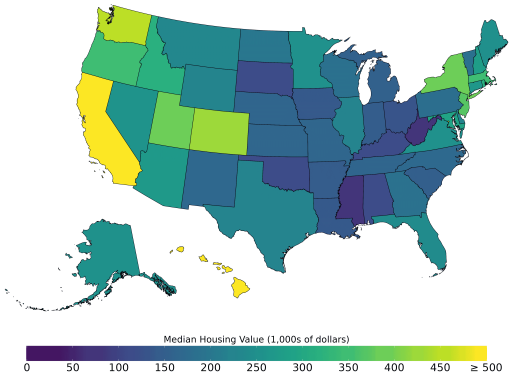
Table 1: Housing value inequality in the United States (2020)

	Gini	Top 1	Top 10	Bottom 50	P99	P90	P50	P10
A. Country								
United States	0.5	0.157	0.402	0.185	2299	612	220	76
B. State								
Alabama	0.365	0.04	0.264	0.249	496	275	113	44
Alaska	0.298	0.024	0.197	0.283	564	419	250	76
Arizona	0.319	0.03	0.229	0.28	768	483	239	96
Arkansas	0.222	0.025	0.186	0.344	333	222	136	79
California	0.554	0.119	0.474	0.163	9372	2226	595	225
Colorado	0.282	0.027	0.203	0.305	991	654	384	144
Connecticut	0.323	0.046	0.264	0.286	1327	540	254	140
Delaware	0.28	0.027	0.208	0.304	639	426	223	109
District of Columbia	0.236	0.023	0.182	0.334	1704	1180	763	359
Florida	0.361	0.037	0.249	0.251	841	463	224	63
Georgia	0.339	0.034	0.245	0.265	660	392	176	71
Hawaii	0.342	0.031	0.224	0.257	1653	1050	550	133
Idaho	0.208	0.02	0.173	0.351	555	437	272	160
Illinois	0.345	0.032	0.239	0.257	741	482	214	80
Indiana	0.318	0.034	0.237	0.28	449	258	125	55
Iowa	0.286	0.028	0.205	0.297	388	256	140	60
Kansas	0.318	0.031	0.23	0.275	541	337	160	71
Kentucky	0.316	0.033	0.233	0.282	385	231	112	47
Louisiana	0.304	0.034	0.232	0.291	496	278	139	67
Maine	0.228	0.024	0.187	0.34	571	408	241	142
Maryland	0.306	0.028	0.219	0.287	844	549	286	109
Massachusetts	0.293	0.027	0.217	0.294	1012	696	344	166
Michigan	0.341	0.03	0.235	0.257	493	333	150	59
Minnesota	0.269	0.025	0.196	0.31	600	410	243	99
Mississippi	0.308	0.036	0.234	0.286	292	170	79	41
Missouri	0.298	0.033	0.225	0.295	542	316	158	75
Montana	0.253	0.025	0.197	0.319	570	402	221	121
Nebraska	0.297	0.026	0.21	0.29	417	295	156	63
Nevada	0.304	0.031	0.223	0.289	773	462	241	102
New Hampshire	0.276	0.026	0.197	0.304	642	438	259	100
New Jersey	0.225	0.024	0.191	0.345	817	571	349	192
New Mexico	0.242	0.028	0.204	0.331	476	311	160	105
New York	0.442	0.039	0.272	0.177	1705	1036	387	82
North Carolina	0.341	0.035	0.246	0.264	623	364	162	69
North Dakota	0.209	0.023	0.178	0.354	450	310	195	117
Ohio	0.332	0.033	0.24	0.27	440	268	125	48
Oklahoma	0.305	0.033	0.233	0.288	397	234	110	58
Oregon	0.267	0.024	0.194	0.313	735	534	316	131
Pennsylvania	0.346	0.032	0.239	0.257	601	378	175	62
Rhode Island	0.207	0.025	0.191	0.361	759	493	295	198
South Carolina	0.371	0.043	0.274	0.248	711	349	145	62
South Dakota	0.349	0.032	0.244	0.25	403	254	109	45
Tennessee	0.334	0.035	0.248	0.267	608	351	154	73
Texas	0.263	0.025	0.202	0.316	540	372	210	101
Utah	0.168	0.019	0.161	0.382	671	520	354	236
Vermont	0.3	0.03	0.218	0.289	571	372	186	86
Virginia	0.373	0.032	0.253	0.24	912	605	242	79
Washington	0.302	0.025	0.209	0.286	1100	793	406	159
West Virginia	0.434	0.041	0.29	0.194	323	184	65	19
Wisconsin	0.281	0.027	0.206	0.303	483	336	180	81
Wyoming	0.253	0.033	0.219	0.329	786	401	224	135

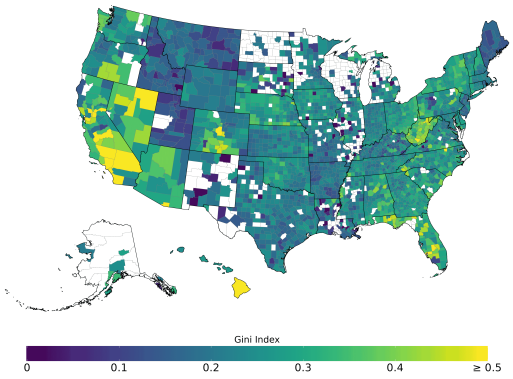
Notes: *Gini* is the Gini index. *Top 1* and *Top 10* measure the share of housing value that accrues to the top 1 or 10% most valuable dwellings in the geography, respectively. *Bottom 50* measure the share of housing value accrued to the 50% least valuable dwellings. *P99*, *P90*, *P50*, and *P10* indicate the value of the dwelling in the 99th, 90th, 50th, and 10th percentiles of the dwelling value distribution in the specific geography, respectively. Housing value is measured in thousands of 2020 USD.



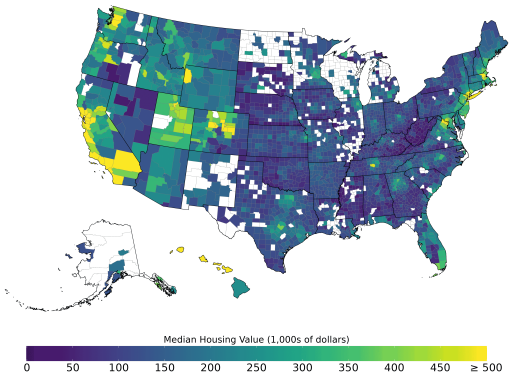
(a) Gini index by state



(b) Median housing value by state



(c) Gini index by county



(d) Median housing value by county

Figure 1: Housing value inequality in the United States

Notes: Top panels illustrate state-level housing value metrics; bottom panels illustrate county-level metrics. Left panels report the Gini index; right panels report the median housing value. Housing value is measured in thousands of 2020 USD.

U.S. pre-tax income Gini in 2022 (0.49) (Guzman and Kollar 2023).

There is significant variation across subnational levels. At the state level, California stands out as the most unequal (Gini of 0.55). In California, even dwellings at the bottom of the distribution are relatively expensive ($225,000 at the 10th percentile), while those at the top are extremely valuable ($2.2 million and $9.3 million at the 90th and 99th percentiles, respectively). New York (Gini of 0.44) and West Virginia (Gini of 0.43) follow as the second- and third-most unequal states, but for very different reasons. In New York,

the high Gini is driven by the right tail (e.g., a dwelling at the 99th percentile is worth $1.7 million), whereas in West Virginia, low values in the left tail are key (e.g., a dwelling at the 10th percentile is worth $20,000). Other states exhibit significantly lower inequality levels, with Gini coefficients mostly between 0.20 and 0.35.

Within-state variation is even greater. For example, in California, county-level Gini indices range from 0.11 (Mariposa County) to 0.63 (Sacramento County). Figure 1 illustrates this variation well while highlighting the role of high-value (mostly coastal) cities in driving inequality up (panel d). There is also substantial inequality within counties. Appendix Figure A1 focuses on Chicago and shows pronounced variation in median housing values across census tracts (right panel), with much higher prices downtown and in the northeastern areas near Lake Michigan, and very low values in the southern parts of the city. The left panel shows relatively low within-tract Gini coefficients, indicating small price dispersion within neighborhoods.

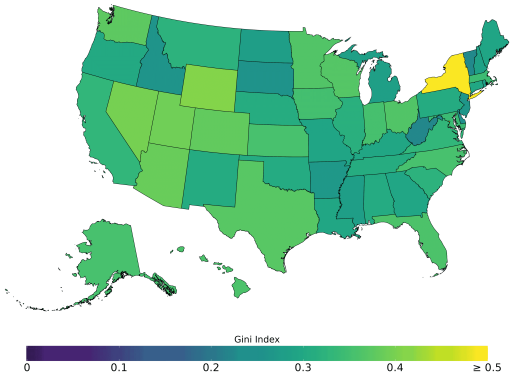
Table A1 quantitatively corroborates the patterns evident in the preceding maps. It decomposes the state-level Theil index into between- and within-county components. Columns 1-3 show that, with the notable exception of New York – where the between-county component accounts for 73% of the total – high inequality is predominantly driven by within-county variation, which generally accounts for 50-80% of the total.

Dispersion in the **price per square foot** (psqft) is even greater. Table A2 shows that, across the United States in 2020, psqft varied substantially: $52 at P10, $136 at the median, $451 at P90, and $3,211 at P99. As before, California is the most unequal state, driven by very large values at the top of the distribution (a square foot at P99 exceeds $13,000). Other states with significant psqft dispersion include New York (Gini of 0.55), Washington (Gini of 0.49), Pennsylvania (Gini of 0.44), and West Virginia (Gini of 0.41). Except for West Virginia, high-inequality states are characterized by elevated psqft in the right tail.

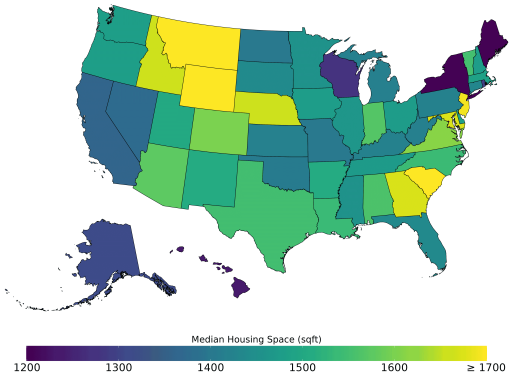
Table 2: Housing space inequality in the United States (2020)

	Gini	Top 1	Top 10	Bottom 50	P99	P90	P50	P10
A. Country								
United States	0.354	0.104	0.301	0.275	8368	2973	1475	804
B. State								
Alabama	0.305	0.081	0.27	0.308	7000	3016	1560	975
Alaska	0.354	0.086	0.283	0.269	6542	2656	1314	576
Arizona	0.386	0.138	0.339	0.263	15402	3098	1579	912
Arkansas	0.268	0.064	0.249	0.33	7167	2678	1512	996
California	0.329	0.07	0.265	0.283	6936	2691	1359	680
Colorado	0.381	0.084	0.295	0.248	8288	3576	1601	677
Connecticut	0.299	0.06	0.251	0.303	6038	2850	1431	837
Delaware	0.24	0.032	0.2	0.335	4224	2613	1485	882
District of Columbia	0.503	0.167	0.415	0.185	16928	3680	1184	620
Florida	0.361	0.122	0.314	0.274	10474	2669	1437	760
Georgia	0.296	0.075	0.255	0.307	6385	3198	1672	1000
Hawaii	0.36	0.094	0.288	0.267	5334	2538	1232	559
Idaho	0.261	0.048	0.221	0.326	5380	2932	1664	972
Illinois	0.318	0.095	0.266	0.29	6377	3233	1464	910
Indiana	0.353	0.104	0.307	0.278	8819	3200	1568	928
Iowa	0.35	0.126	0.311	0.28	9853	2816	1476	878
Kansas	0.356	0.114	0.301	0.273	6630	2973	1418	816
Kentucky	0.298	0.066	0.265	0.312	6992	2831	1419	924
Louisiana	0.299	0.082	0.264	0.309	6897	2700	1538	907
Maine	0.295	0.052	0.24	0.305	4065	2099	1092	576
Maryland	0.341	0.097	0.29	0.28	8529	3446	1676	999
Massachusetts	0.345	0.111	0.296	0.281	6883	2941	1488	825
Michigan	0.283	0.063	0.244	0.313	5854	2641	1416	871
Minnesota	0.366	0.14	0.313	0.268	6930	3056	1457	858
Mississippi	0.281	0.079	0.254	0.321	5968	2697	1456	980
Missouri	0.295	0.075	0.259	0.309	5940	2664	1400	882
Montana	0.324	0.076	0.261	0.288	8124	3266	1728	845
Nebraska	0.312	0.084	0.263	0.297	6948	3143	1662	936
Nevada	0.398	0.083	0.278	0.23	4972	2653	1375	253
New Hampshire	0.288	0.042	0.225	0.304	4838	2781	1476	751
New Jersey	0.265	0.053	0.225	0.322	5635	2995	1692	999
New Mexico	0.308	0.079	0.263	0.3	6217	2689	1507	780
New York	0.527	0.24	0.49	0.199	27958	2600	1200	662
North Carolina	0.354	0.114	0.316	0.283	12672	3053	1542	952
North Dakota	0.28	0.075	0.249	0.313	6348	2540	1400	987
Ohio	0.365	0.104	0.309	0.27	9058	3060	1475	818
Oklahoma	0.306	0.09	0.272	0.307	7930	2598	1408	840
Oregon	0.303	0.074	0.255	0.302	5500	2808	1480	846
Pennsylvania	0.308	0.073	0.257	0.3	5731	2688	1422	792
Rhode Island	0.282	0.056	0.237	0.313	4673	2408	1297	748
South Carolina	0.321	0.06	0.26	0.285	7598	3718	1722	999
South Dakota	0.259	0.052	0.225	0.327	4531	2550	1439	924
Tennessee	0.31	0.078	0.265	0.301	6269	2914	1472	851
Texas	0.371	0.13	0.307	0.261	8547	3430	1550	840
Utah	0.392	0.105	0.309	0.246	8459	3278	1501	594
Vermont	0.225	0.041	0.205	0.351	4560	2468	1554	924
Virginia	0.328	0.096	0.28	0.289	7139	3389	1612	986
Washington	0.377	0.131	0.305	0.255	8943	2978	1480	660
West Virginia	0.223	0.036	0.208	0.354	4367	2382	1415	942
Wisconsin	0.369	0.128	0.324	0.286	6415	2594	1272	660
Wyoming	0.408	0.119	0.348	0.246	10603	3991	1752	816

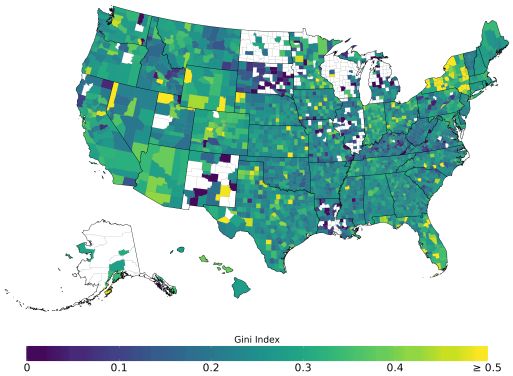
Notes: *Gini* is the Gini index. *Top 1* and *Top 10* measure the share of housing space that accrues to the top 1 or 10% most spacious dwellings in the specific geography, respectively. *Bottom 50* measure the share of housing space accrued to the 50% least spacious dwellings. *P99*, *P90*, *P50*, and *P10* indicate the space of the dwelling in the 99th, 90th, 50th, and 10th percentiles of the dwelling space distribution in the specific geography, respectively. Housing space is measured in square feet.



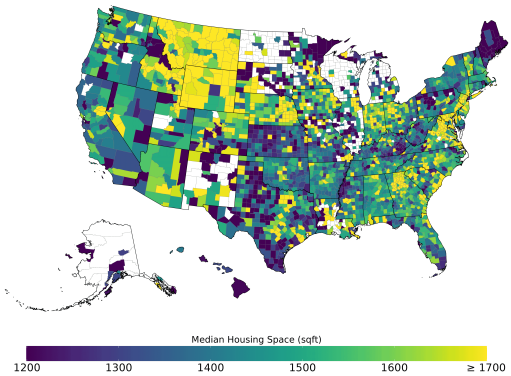
(a) Gini index by state



(b) Median housing space by state



(c) Gini index by county



(d) Median housing space by county

Figure 2: Housing space inequality in the United States

Notes: Top panels illustrate state-level housing space metrics; bottom panels illustrate county-level metrics. Left panels report the Gini index; right panels report the median housing space. Housing space is measured in square feet.

Housing space. Housing space inequality in the United States is relatively high (Gini of 0.35) but substantially lower than housing value inequality. A dwelling at the median contains roughly 1,500 sqft. At the extremes, dwellings range from 800 sqft at P10 to 3,000 sqft at P90, and exceed 8,300 sqft at P99.[4] The corresponding Gini index (0.35) is slightly higher than in Spain (0.31) and Belgium (0.25) (Domènech-Arumí and Mariani

[4]According to the American Housing Survey (U.S. Census Bureau 2021), in 2020 a dwelling at P10 contained between 750 and 999 sqft; at P50, between 1,500 and 1,999 sqft; at P90, between 3,000 and 3,999 sqft; and at P99, more than 4,000 sqft. This suggests we may slightly underestimate dwelling sizes.

2025, Domènech-Arumí and Gobbi 2025). The higher inequality in the United States is driven by the right tail (i.e., dwellings at the top are larger than in Belgium or Spain).

Subnational housing space inequality also varies substantially. New York and the District of Columbia are the two most unequal states, with Gini indices above 0.50, driven by very small dwellings at the bottom (600-700 sqft at P10) and very large dwellings at the top (e.g., 28,000 sqft at P99 in New York). The remaining states exhibit markedly lower inequality, with Gini indices between 0.25 and 0.40; California's Gini is 0.33. As Figure 2 illustrates, within-state inequality is substantial and appears to be driven by differences within counties. Columns 4-6 in Table A1 confirm that the within-county component accounts for over 90% of the variation in most states (based on the Theil index). Even in New York, the between-county component accounts for only 12% of the variation.

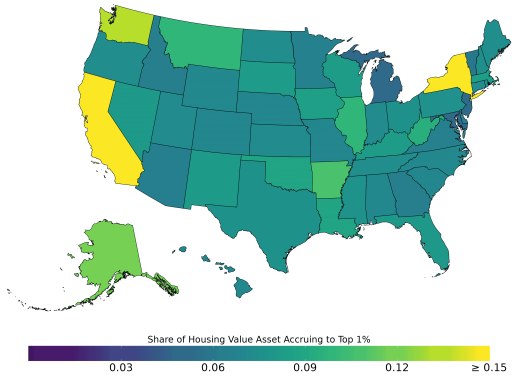
Housing assets. Housing asset concentration is extreme. Under the interpretation that one occupied dwelling corresponds to one household (homeowner or renter), Table 3 indicates that the top 10% of the households control 45% of the housing assets in the country, whereas the bottom 50% control only 1.7%.[5] The median household holds $100,000 in housing assets. Households at P90 and P99 hold $0.5 million and $1.3 million, respectively. The corresponding Gini index is 0.68. We lack strictly comparable figures for other contexts, and "housing assets" here do not equal housing wealth because liabilities are excluded. Nonetheless, OECD (2021) estimates that 50% of housing wealth is held by the top 10% of households, a figure close to ours (44-45%). Despite the magnitudes, housing assets (and housing wealth) are more evenly distributed than other forms of wealth: we find that the top 1% hold 12% of housing assets, whereas Chancel et al. (2022) and Zucman (2019) estimate that the top 1% in the United States hold about 35% of total wealth. Housing assets are highly concentrated, but less so than non-housing assets.

[5]This stark contrast is partly due to renters – over 35% of households – holding $0 in housing assets.

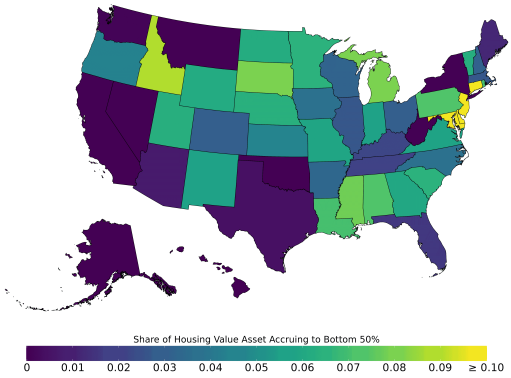
Table 3: Housing assets inequality in the United States (2020)

	Gini	Top 1	Top 10	Bottom 50	P99	P90	P50	P10
A. Country								
United States	0.679	0.123	0.45	0.017	1271	484	97	0
B. State								
Alabama	0.59	0.071	0.359	0.074	579	291	91	0
Alaska	0.719	0.12	0.459	0	1090	440	0	0
Arizona	0.629	0.067	0.366	0.007	816	446	113	0
Arkansas	0.635	0.107	0.405	0.033	691	264	83	0
California	0.708	0.158	0.477	0	2400	869	0	0
Colorado	0.601	0.072	0.346	0.03	1137	616	222	0
Connecticut	0.565	0.074	0.367	0.108	1427	538	208	0
Delaware	0.542	0.064	0.312	0.099	796	409	179	0
District of Columbia	0.764	0.118	0.524	0	2536	981	0	0
Florida	0.643	0.083	0.393	0.015	916	423	106	0
Georgia	0.595	0.065	0.356	0.06	717	388	122	0
Hawaii	0.651	0.072	0.375	0	1907	982	0	0
Idaho	0.541	0.065	0.303	0.089	881	445	213	0
Illinois	0.648	0.098	0.413	0.027	980	402	101	0
Indiana	0.607	0.072	0.362	0.047	542	293	87	0
Iowa	0.619	0.084	0.369	0.037	510	264	76	0
Kansas	0.613	0.073	0.366	0.046	619	345	97	0
Kentucky	0.641	0.083	0.389	0.022	463	239	57	0
Louisiana	0.596	0.087	0.373	0.069	692	275	99	0
Maine	0.623	0.074	0.358	0.012	882	463	138	0
Maryland	0.525	0.053	0.304	0.114	948	566	239	0
Massachusetts	0.633	0.084	0.391	0.027	1501	653	196	0
Michigan	0.562	0.053	0.321	0.08	533	334	116	0
Minnesota	0.566	0.067	0.32	0.065	754	427	177	0
Mississippi	0.584	0.077	0.355	0.078	377	179	60	0
Missouri	0.594	0.07	0.353	0.058	614	316	106	0
Montana	0.702	0.099	0.427	0	850	403	0	0
Nebraska	0.577	0.069	0.331	0.064	527	301	105	0
Nevada	0.687	0.084	0.407	0	833	422	0	0
New Hampshire	0.606	0.075	0.35	0.031	920	457	153	0
New Jersey	0.468	0.052	0.277	0.166	996	565	291	0
New Mexico	0.592	0.083	0.356	0.057	670	311	125	0
New York	0.814	0.171	0.636	0	2206	616	0	0
North Carolina	0.62	0.072	0.373	0.038	728	376	109	0
North Dakota	0.565	0.074	0.318	0.063	631	323	140	0
Ohio	0.624	0.076	0.371	0.029	526	296	84	0
Oklahoma	0.668	0.089	0.41	0	472	239	0	0
Oregon	0.59	0.075	0.339	0.044	1148	561	221	0
Pennsylvania	0.587	0.076	0.356	0.073	736	390	127	0
Rhode Island	0.584	0.076	0.349	0.068	1214	564	232	0
South Carolina	0.598	0.07	0.368	0.065	737	339	107	0
South Dakota	0.575	0.082	0.348	0.081	579	289	97	0
Tennessee	0.643	0.075	0.391	0.022	661	346	82	0
Texas	0.645	0.076	0.38	0.004	641	369	68	0
Utah	0.559	0.076	0.313	0.061	1084	523	275	0
Vermont	0.584	0.06	0.335	0.062	600	338	114	0
Virginia	0.603	0.066	0.362	0.059	1045	606	171	0
Washington	0.676	0.139	0.436	0.001	1548	730	116	0
West Virginia	0.708	0.094	0.45	0	417	213	0	0
Wisconsin	0.61	0.083	0.356	0.032	587	328	111	0
Wyoming	0.575	0.071	0.335	0.062	825	390	170	0

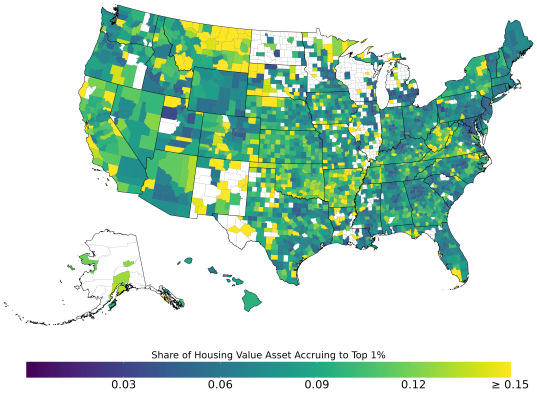
Notes: *Gini* is the Gini index. *Top 1* and *Top 10* measure the share of housing assets that accrues to the top 1 or 10% of the households (ranked by the value of their assets) in the geography, respectively. *Bottom 50* measures the share of housing assets accrued to the bottom 50% of the households. *P99*, *P90*, *P50*, and *P10* indicate the housing assets of the owners in the 99th, 90th, 50th, and 10th percentiles of the distribution in the geography, respectively. Households in non-owner-occupied properties are imputed $0 in housing assets. Housing assets are measured in thousands of 2020 USD.



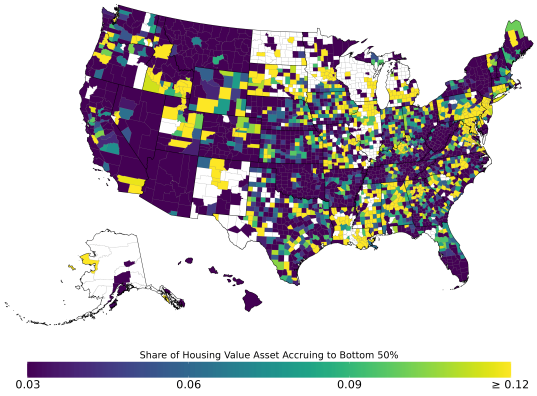
(a) Assets owned by the top 1% in the state



(b) Assets owned by the bottom 50% in the state



(c) Assets owned by the top 1% in the county



(d) Assets owned by the bottom 50% in the county

Figure 3: Housing asset inequality in the United States

Notes: The top panels show state-level housing asset metrics; the bottom panels show county-level metrics. The left panels report the share of housing assets held by the top 1% of households (ranked by the value of their housing assets), and the right panels report the share held by the bottom 50% of households (ranked likewise). Housing assets are measured in thousands of 2020 USD.

The distribution of housing assets is highly unequal across the country, and subnational variation is substantial. At the top, New York ranks as the most unequal state, with a Gini of 0.81 and the top 1% of households holding almost 17% of housing assets. At the bottom, Maryland is the most equal, with a Gini of 0.52 and the top 1% holding just over 5%. Figure 3 illustrates this variation and underscores the importance of cities. For example, housing asset inequality is exceptionally high in New York City due to very high dwelling

values, especially in the upper tail, and the relatively large share of non-owner-occupied properties (i.e., renters).

3.2 Trends in housing inequality

Housing value. The top-left panel of Figure 4 shows that housing value inequality has risen since 1980, with a pronounced acceleration in the early 2000s. The Gini coefficient increased from 0.34 in 1980 to 0.42 in 1990, 0.43 in 2000, 0.53 in 2010, and 0.50 in 2020.

As the other panels in the figure make clear, the nationwide increase in the early 2000s was predominantly driven by California, where the Gini rose by nearly 20 points (from 0.41 to 0.58) between 2001 and 2010. Although most states experienced increases in inequality from 1980 to 2020, none matched California's surge. For example, housing value inequality in New York remained relatively stable, fluctuating mostly between 0.4 and 0.5. Figures A2 and A3 show why: post-2000, prices at the top of the distribution increased especially rapidly in California.

Housing space. Housing space inequality has remained relatively stable since 1900. With the caveat that very-long run estimates must be interpreted with caution due to survivorship bias,[6] the top panel in Figure 5 indicates that housing space inequality in the United States (as measured by the Gini index) slowly and progressively decreased from 0.40 in 1900 to 0.35 in 2020.

This gradual decline appears in many – though not all – states, with a few notable exceptions. The most striking is Nevada, where housing space inequality rose sharply between 1940 (Gini of 0.20) and 1960 (Gini of 0.55), then declined over subsequent decades (Gini of 0.40 in 2020). These trends are closely tied to the growth of Las Vegas. Most other

[6]This is because dwellings that have remained in the stock (i.e., they have survived) for a long time are not a random sample of all dwellings that existed in the past. They are likely to be positively selected, for example, in the quality of the construction materials. Survivorship bias requires caution when interpreting the very long-run estimates. See Appendix B.1.7 for a more thorough discussion on the subject.

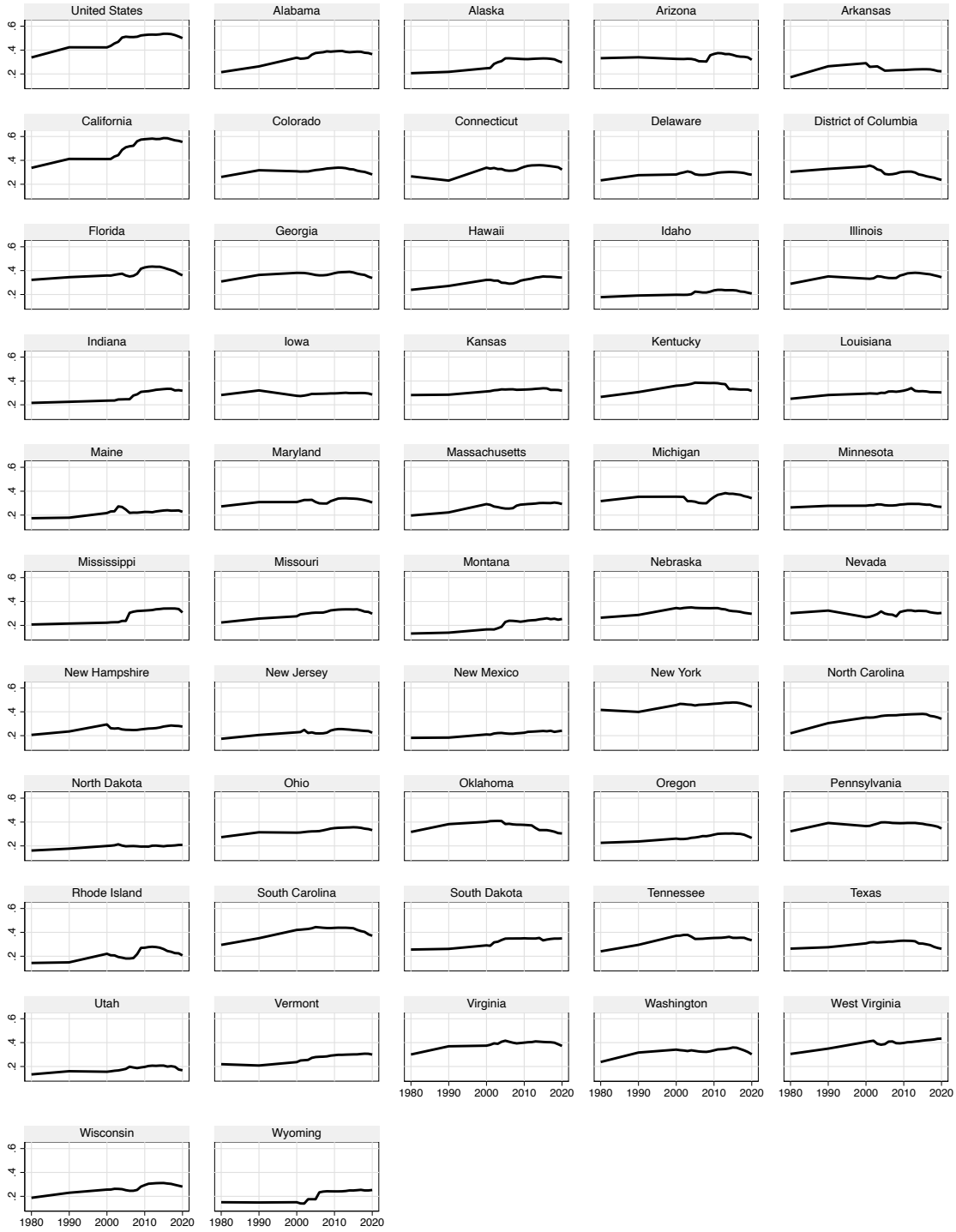


Figure 4: Trends in housing value inequality (Gini) across states

Notes: The top-left panel describes the evolution of housing value inequality (Gini) for the United States. The remaining panels report the evolution of housing value inequality at the state level. Housing value is measured in current USD (not adjusted for inflation). Inequality is measured with the Gini index.

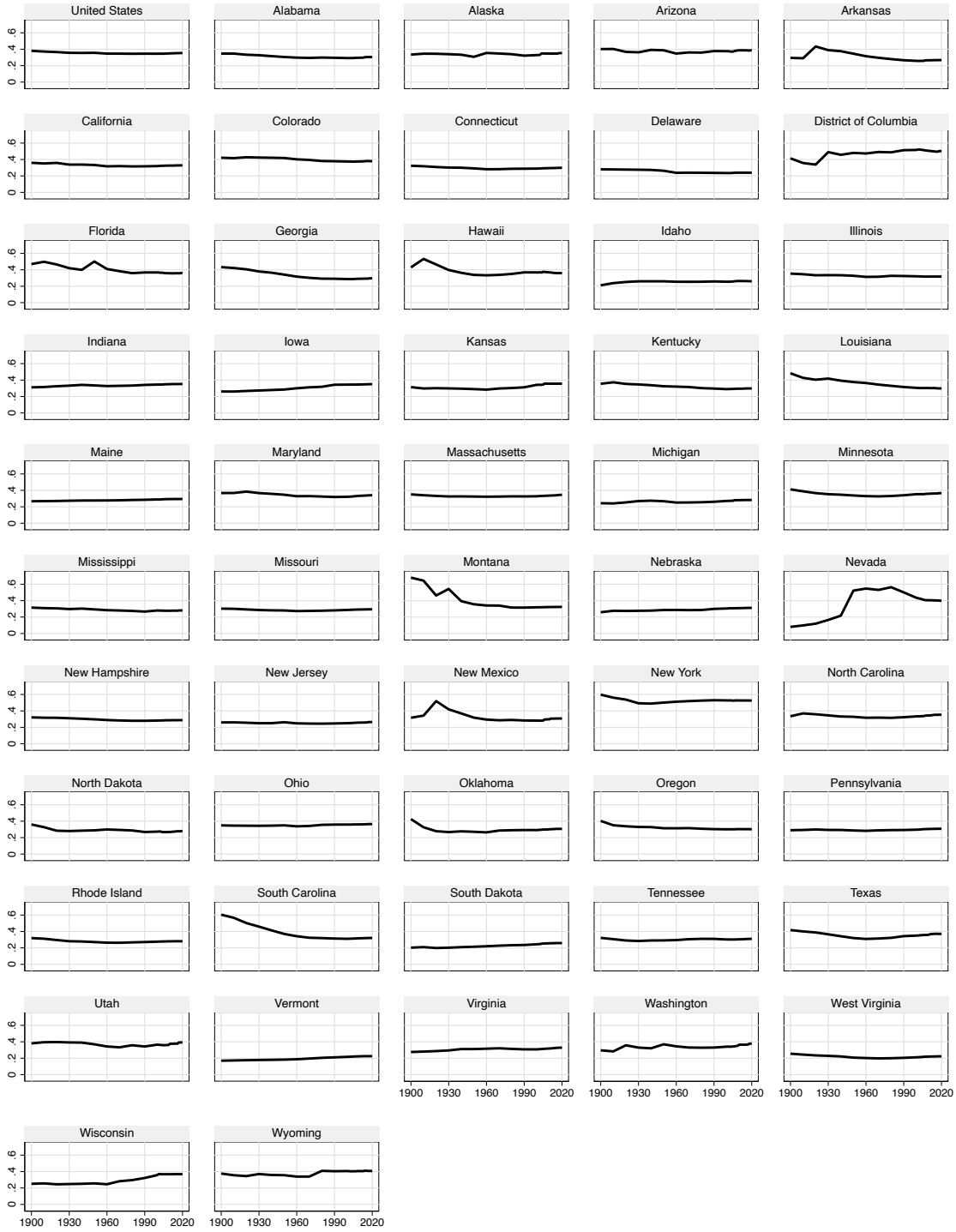


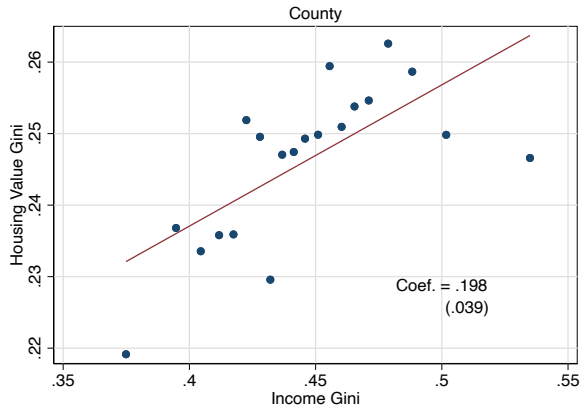
Figure 5: Trends in housing space inequality (Gini) across states

Notes: The top-left panel describes the evolution of housing space inequality for the United States. The remaining panels report the evolution of housing space inequality at the state level. Housing space is measured in square feet. Inequality is measured with the Gini index.

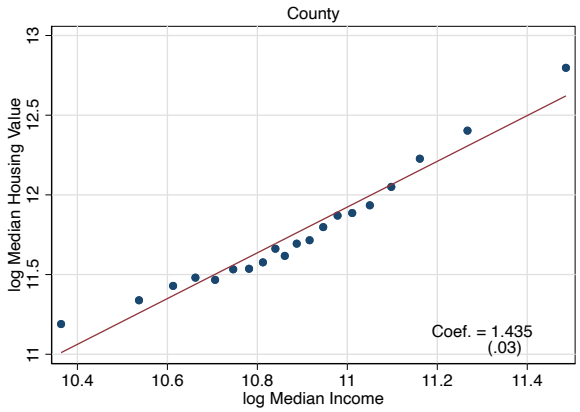
states exhibit modest declines in inequality.

The relatively stable levels of space inequality in Figure 5 might suggest that the distribution of space among newly built dwellings has also been stable, but Figure A4 shows otherwise. Patterns vary across states, but the figure shows that housing space at 90th percentile has generally decreased since 1900. For example, in Utah and Florida, a dwelling at P90 in 1900 had approximately 6,000 sqft; by 2020, a dwelling at the same percentile had under 4,000 sqft. This decline is less evident in most states, and in places like California, top-end dwellings have become slightly larger (about 2,100 sqft in 1900 vs. 2,700 sqft in 2020). At the same time, median dwelling size increased from roughly 1,350 sqft in 1900 to almost 1,500 sqft in 2020. Thus, in most states, the slight decline in space inequality reflects small decreases at the top and small increases at the bottom.

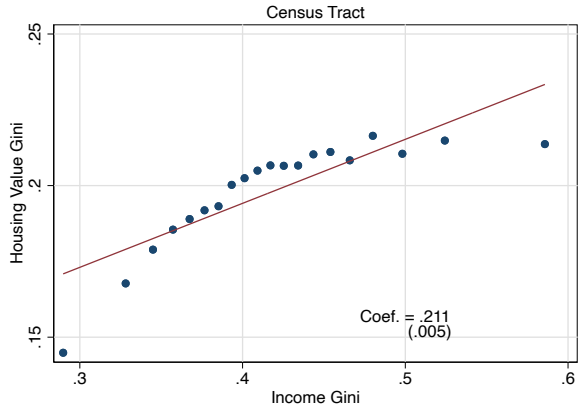
Housing assets. For completeness and transparency, we examine the evolution of housing asset inequality in Figure A5, but we strongly caution against overinterpretation. Because the series begins in the 1980s, survivorship bias is less central than before (though still relevant to some extent). Our primary concern is that, with our data, we cannot observe changes in the ownership structure of housing. Consequently, movements in inequality reflect only differential price changes across the distribution – a phenomenon known to affect wealth distribution (Martínez-Toledano 2020). With these caveats, Figure A5 suggests that housing asset inequality has remained very high and increased mildly over time: the Gini is 0.68 in 2020, up from 0.65 in 1980.



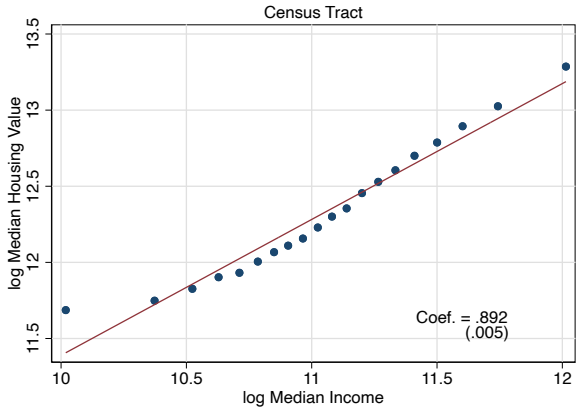
(a) Income and housing value Gini (county)



(b) Median income and housing value (county)



(c) Income and housing value Gini (tract)



(d) Median income and housing value (tract)

Figure 6: The association between housing value inequality and income

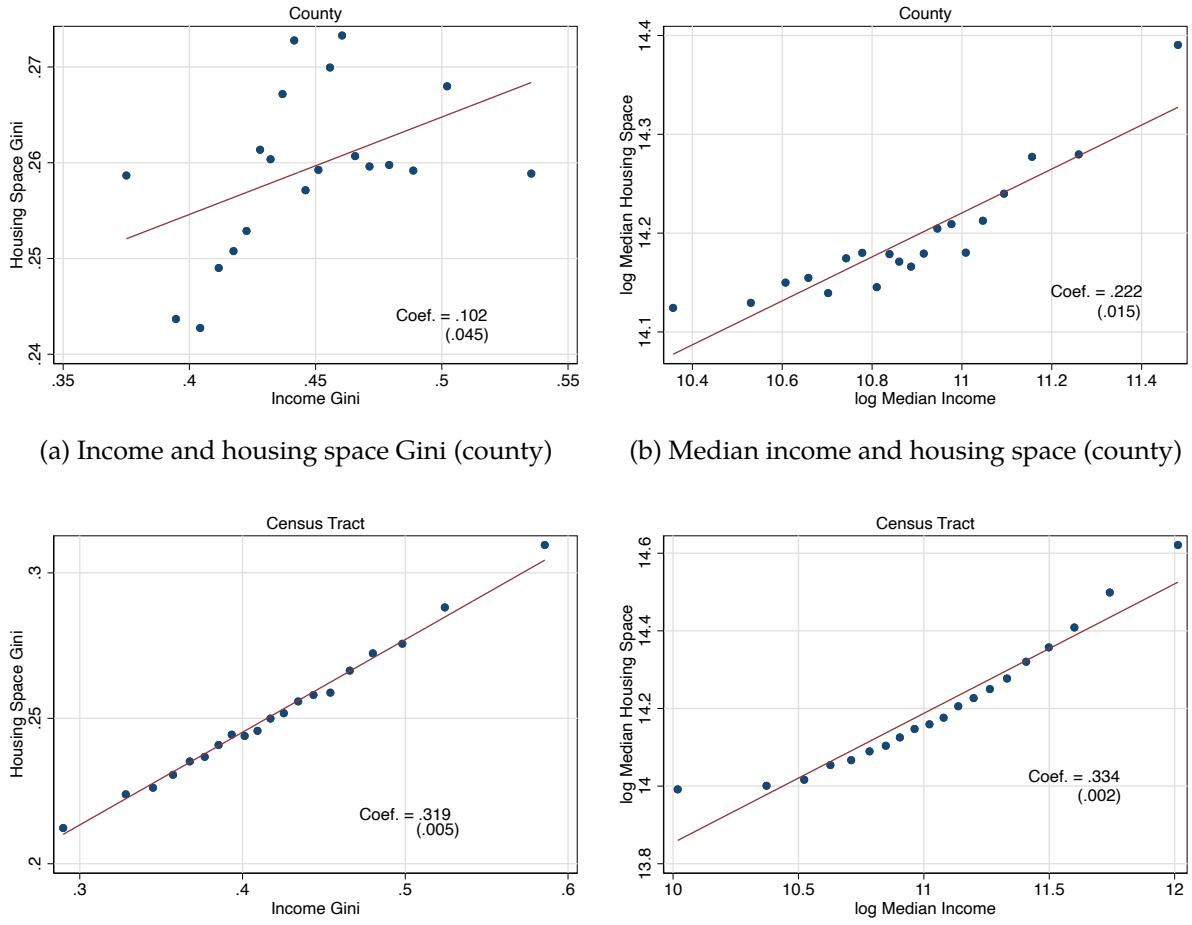
Notes: Binned scatterplots showing the association between housing value and income inequality. The top panels report county-level results; the bottom panels report census-tract results. Housing value and income are measured in 2020 USD. The text in the bottom right of each panel reports OLS estimates and standard errors from regressing the housing-inequality measure on the y-axis against the income-inequality measure on the x-axis.

4 The determinants of housing inequality

4.1 What explains housing inequality in 2020?

Income. We study the association between housing and income inequality by estimating:

$$HousIneq_i = \beta IncIneq_i + \gamma X_i^{'} + \epsilon_i \tag{1}$$



(a) Income and housing space Gini (county)

(b) Median income and housing space (county)

(c) Income and housing space Gini (tract)

(d) Median income and housing space (tract)

Figure 7: The association between housing space inequality and income

Notes: Binned scatterplots showing the association between housing space and income inequality. The top panels report county-level results; the bottom panels report census-tract results. Housing space is measured in square feet. Income is measured in 2020 USD. The text in the bottom right of each panel reports OLS estimates and standard errors from regressing the housing-inequality measure on the y-axis against the income-inequality measure on the x-axis.

Where $HousIneq_i$ denotes housing-space or housing-value inequality (Gini or log median) in 2020 for place i (county or census tract), and $IncIneq_i$ denotes income inequality (Gini or log median income) in 2020 for place i. X_i is a vector of controls that includes a constant and population counts; depending on the specification, it also includes log median income and state or county fixed effects. Figures 6 and 7 visually display key associations and report β estimates from regressions without controls. Appendix Table A3 reports the

same estimates, along with results from specifications that include controls.

The correlation between income and housing inequality is moderately high. At the county and census-tract levels (Panels a and c; Columns 1 and 3), a one-point increase in the income Gini is associated with roughly a 0.20-point increase in housing value inequality. After controlling for local income and population and exploiting within-state (Column 2) or within-county (Column 4) variation, the coefficient falls to about 0.17. Panel c of Figure 6 suggests a non-linear relationship: the association is relatively strong when income inequality is low (Gini < 0.4) and flattens thereafter. Figure 7 and Panel B of Appendix Table A3 show a similarly positive and significant association for housing space inequality. The relationship is stronger and appears linear at the census-tract level: a one-point increase in the income Gini is associated with a 0.30-point increase in the housing-space Gini. Accounting for local income and population barely changes the tract-level estimate and raises the county-level estimate to about 0.30. These magnitudes are comparable to those reported by Domènech-Arumí and Mariani (2025) for Spain.

Median income is strongly associated with housing value and size. Panels b and e in Figures 6 and 7, and Columns 5-8 in both panels of Appendix Table A3, examine correlations between the medians of the housing and income distributions. At the county level, a 1% increase in median income is associated with 1-1.4% higher housing values; at the census-tract level, the effect ranges from 0.5% to 0.9%. Panel B shows that, at the county level, a 1% higher income is associated with a dwelling 0.25% bigger; at the tract level, the association is 0.3-0.4%. Thus, the income-housing space association is statistically significant but more modest than for housing values.

Other factors. We examine raw correlations between housing inequality and other factors at the commuting-zone level in Figure 8, which replicates Figure VIII from Chetty et al. (2014) using their data.[7]

[7]We work at the commuting-zone level because the covariates in the replication package are defined at that level. Most covariates are measured around 2000.

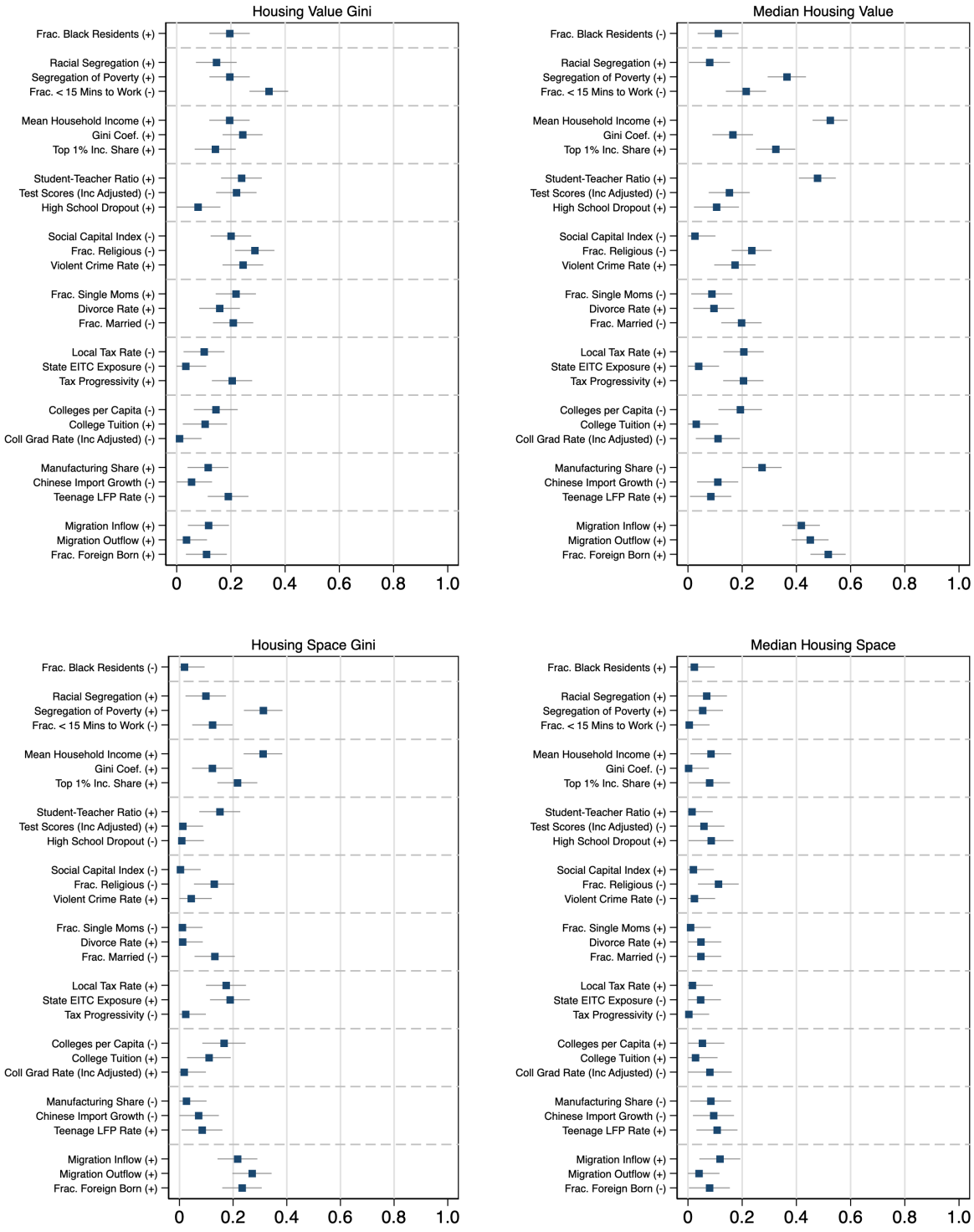


Figure 8: The determinants of housing inequality

Notes: This figure reports the raw correlations between housing inequality estimates (space and value) and several covariates at the 1990 Commuting Zone level and retrieved from Chetty et al. (2014). Most covariates are measured in 2000. See Section G in their online Appendix for more details. The + and − signs next to a covariate label denote a positive and negative correlation, respectively. Housing value is measured in thousands of 2020 USD. Housing space is measured in square feet.

A few factors are mildly correlated with housing value or space inequality, but none stands out (the largest correlations are below 0.3). Inequality in value and space is lower where more people commute less than 15 minutes, where students achieve higher test scores, and where a larger share of individuals are married; it is higher where racial segregation is greater and where migration inflows and the foreign-born share are larger. We find no clear association with industry composition or income-adjusted college graduation rates.

Zoning and land-use regulation may also matter. We analyze these in Section 5 as a separate application to showcase a use of our data.

Median housing values are associated with a clearer subset of covariates. The top-right panel shows a strong correlation (close to 0.5) between median housing value and household income, student-teacher ratios, and migration. It also shows mild correlations (around 0.2) with several other variables, including the manufacturing share, colleges per capita, and the fraction of married households. The correlation with racial segregation is positive but weak (around 0.1). In contrast, the bottom-right panel indicates that median housing space shows little to no correlation with any covariate. Appendix Figure A7 shows a high correlation between housing assets and income, but no correlation between income and housing asset inequality.

Income stands out as a key factor. The correlations between income and housing value inequality, and between income and housing space inequality, are positive and significant, with magnitudes closely aligned with the unconditional correlations in Figures 6 and 7 and Appendix Table A3. Likewise, mean household income is the variable most strongly associated with median housing value (top-right panel), with a correlation of roughly 0.5.

4.2 What explains the changes in housing inequality over time?

Changes in the characteristics of new housing inflows. Housing inequality changes mechanically when the characteristics of new housing inflows differ from those of the

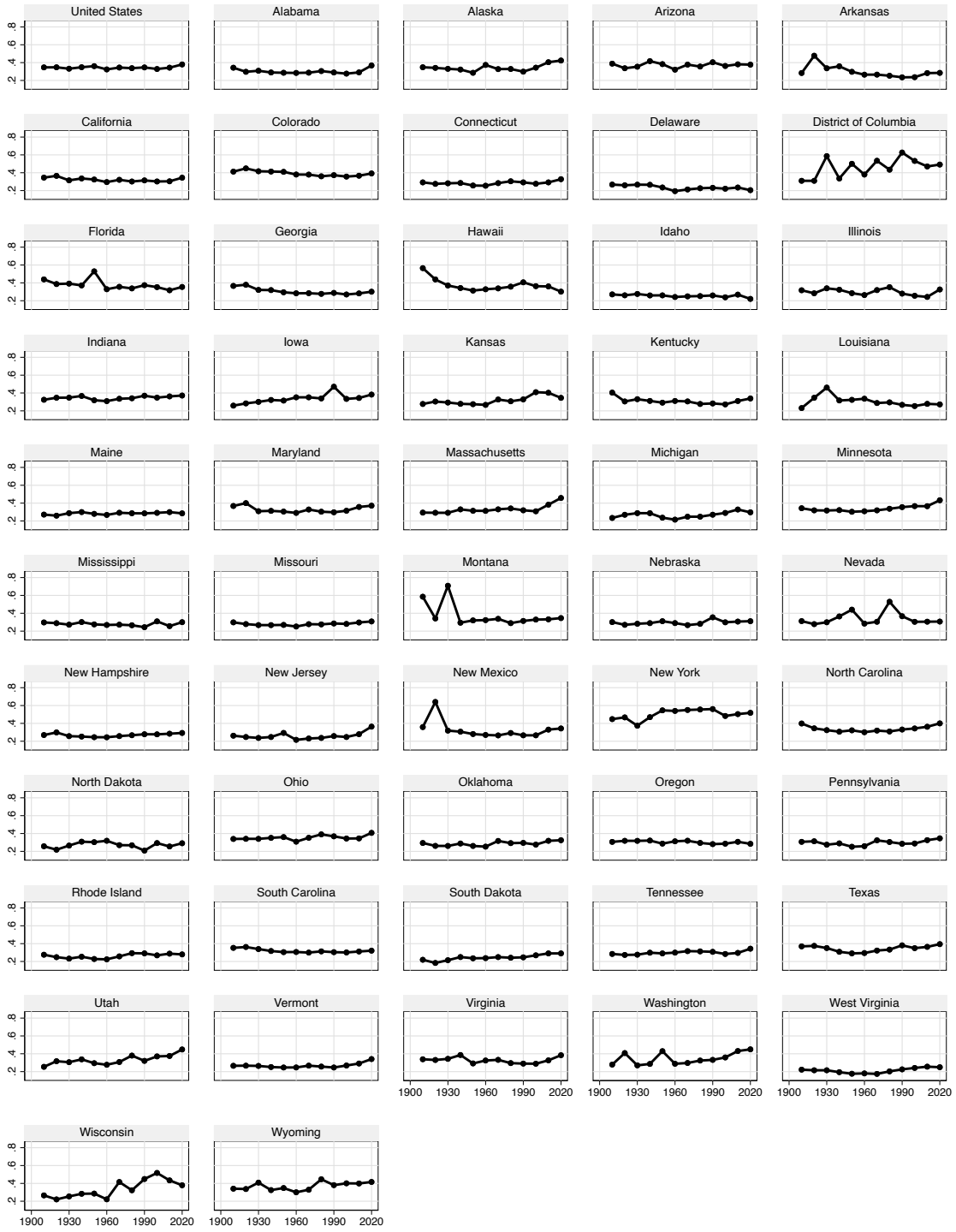


Figure 9: Housing space inequality (Gini) by decade-cohort across states

Notes: The top-left panel describes the evolution of housing space inequality (Gini) by decade-cohort for the United States. The remaining panels capture the evolution of the same outcome across the states. Estimates are constructed based on dwellings entering the housing stock in the previous decade (e.g., the 1910 estimate is based on dwellings constructed between 1901 and 1910). Housing space is measured in square feet.

existing stock. Figure 9 examines this channel by tracking the distribution of housing space (a major determinant of value) in new construction across decades. For instance, the last point in the top-left panel indicates that dwellings built between 2011 and 2020 have a Gini of 0.38 for housing space – close to, but slightly above, the stock-wide Gini of 0.35 in 2020. Overall, inequality in the space distribution of new inflows has been relatively stable over the past century, with Ginis fluctuating in the 0.30-0.40 range.

Housing space has increased slightly over the past 100 years, but the overall inequality in the distribution has remained relatively stable. For example, the underlying data for the top-left panel in Figure A8 reveals that a dwelling at P10 built between 1901 and 1910 contained almost 700 sqft. Housing sizes slowly increased in the subsequent cohorts of dwellings and, in the most recent complete decade (2011-2020), a dwelling in the P10 contained 930 sqft. Similar growth occurred elsewhere in the distribution: at P50, from 1,300 sqft (1901-1910) to 1,850 sqft (2011-2020); at P90, from 2,500 to 3,750 sqft. Thus, overall space inequality has changed little because inequality in the inflows has fluctuated within a narrow range, and dwellings across most parts of the distribution experienced growth.

At the state level, space inequality in inflows has varied more. The rest of the panels in Figures 9 and A8 reveal that, while the majority of states have followed relatively similar trends (i.e., inequality in the flows has remained stable), a few experienced significant change over the decades. For example, in Hawaii, space inequality declined in recent cohorts due to compression at the top. The opposite occurred in Wisconsin, Wyoming, and Utah, where widening gaps between the smallest and largest dwellings increased inequality. Finally, in places like the District of Columbia and Nevada, space inequality fluctuated across cohorts – driven by shrinking bottom-end sizes in DC and broad distributional shifts in Nevada.

Size of new inflows (housing supply). Growth rates in the housing stock have slowed in

Table 4: Changes in housing inequality and housing stock

A. Housing Value

	Δ Housing Value Gini							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Housing Stock	2.237***	2.575***	0.570	0.620	0.002***	0.002***	0.108	0.108
	(0.575)	(0.660)	(0.566)	(0.604)	(0.000)	(0.000)	(0.112)	(0.112)
Δ Population		-1.175		-0.854***		2.321**		0.300
		(1.036)		(0.240)		(0.776)		(0.370)
DV Mean	6.789	6.789	9.417	9.417	13.012	13.012	-8.453	-8.453
DV SD	14.084	14.084	22.745	22.745	80.963	80.963	91.035	91.035
Adj-R^2	0.071	0.082	0.009	0.014	-0.000	0.001	0.004	0.005
N	203	203	1978	1978	7842	7842	48213	48213
Aggregation	State	State	CZ	CZ	County	County	Tract	Tract

B. Housing Space

	Δ Housing Space Gini							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Δ Housing Stock	0.273	0.351	0.037	0.046	0.001	0.001	0.021*	0.021*
	(0.499)	(0.542)	(0.147)	(0.150)	(0.001)	(0.002)	(0.008)	(0.008)
Δ Population		-0.661		-0.168		-5.296		-0.003
		(0.500)		(0.097)		(5.010)		(0.025)
DV Mean	0.199	0.199	1.065	1.065	5.910	5.910	2.432	2.432
DV SD	8.713	8.713	8.144	8.144	353.588	353.588	105.914	105.914
Adj-R^2	0.024	0.034	-0.000	0.001	-0.000	0.000	0.000	0.000
N	560	560	1999	1999	8143	8143	48772	48772
Aggregation	State	State	CZ	CZ	County	County	Tract	Tract

Notes: Δ*Housing Value Gini* and Δ*Housing Space Gini* characterize the decennial percentage change in housing value or space inequality based on the Gini index. Housing value is measured in 1,000s of 2020 USD (inflation-adjusted). Housing space is measured in square feet. Δ*Housing Stock* is the decennial percentage change in the housing stock based on our estimates. Δ*Population* is the decennial percentage change in population based on census data. Independent variables are rescaled so that the coefficients can be interpreted as the effect of a 10pp increase in the variable. Heteroskedasticity-robust standard errors in parentheses. $^*p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

recent decades. Since 1900, the United States expanded its housing stock by roughly 25% per decade (see Appendix Section B.1.7). Growth rates declined after the 1980s and fell sharply in the most recent decade, with only 6.5% growth between 2011 and 2020. Because housing supply is a key determinant of prices, changes in construction pace likely affected both prices (see Appendix Figure A9) and inequality.

We assess this relationship more formally by estimating:

$$\Delta HousIneq_{it} = \alpha + \beta \Delta H_{it} + \gamma \Delta Pop_{it} + \epsilon_{it} \tag{2}$$

Where $\Delta HousIneq_{it}$ is the 10-year percentage change in housing inequality (Gini) in place i during decade t. ΔH_{it} and ΔPop_{it} are the 10-year percentage changes in the housing stock and population, respectively. Table 4 reports the results.

Growth in the housing stock is associated with higher value inequality. Columns 1-2 in Panel A indicate that a 10 percentage point (pp) higher stock growth is associated with 2.2-2.5pp larger increase in housing value inequality (Gini). This positive relationship is present at all levels of aggregation considered (state, commuting zone, county, and census tract), but is economically meaningful only at the state level. Panel B shows a positive but statistically insignificant association between stock growth and housing space inequality.

Expanded housing supply appears to affect prices differently across the distribution. Appendix Table A4 relates stock growth to housing prices at various points in the value distribution. Columns 1, 5, and 9 in Panel A show that a 10pp increase in housing supply growth is associated with 5, 4, and 2.5pp lower housing price growth at P10, P50, and P90, respectively. Thus, stock growth seems to raise housing value inequality by reducing prices relatively more at the bottom. This mechanism is evident only at the state level.

These results contrast with Domènech-Arumí and Mariani (2025), who find a negative relationship between stock growth and both housing space and value inequality in Spain. Neither set of estimates is causal, so differences may reflect multiple factors, including regulatory environments (e.g., Spain's emphasis on multifamily housing in cities) and time periods (the Spain study covers 2009–2019, the aftermath of a housing bubble).

Business cycles (and housing demand). Housing demand co-moves with the business

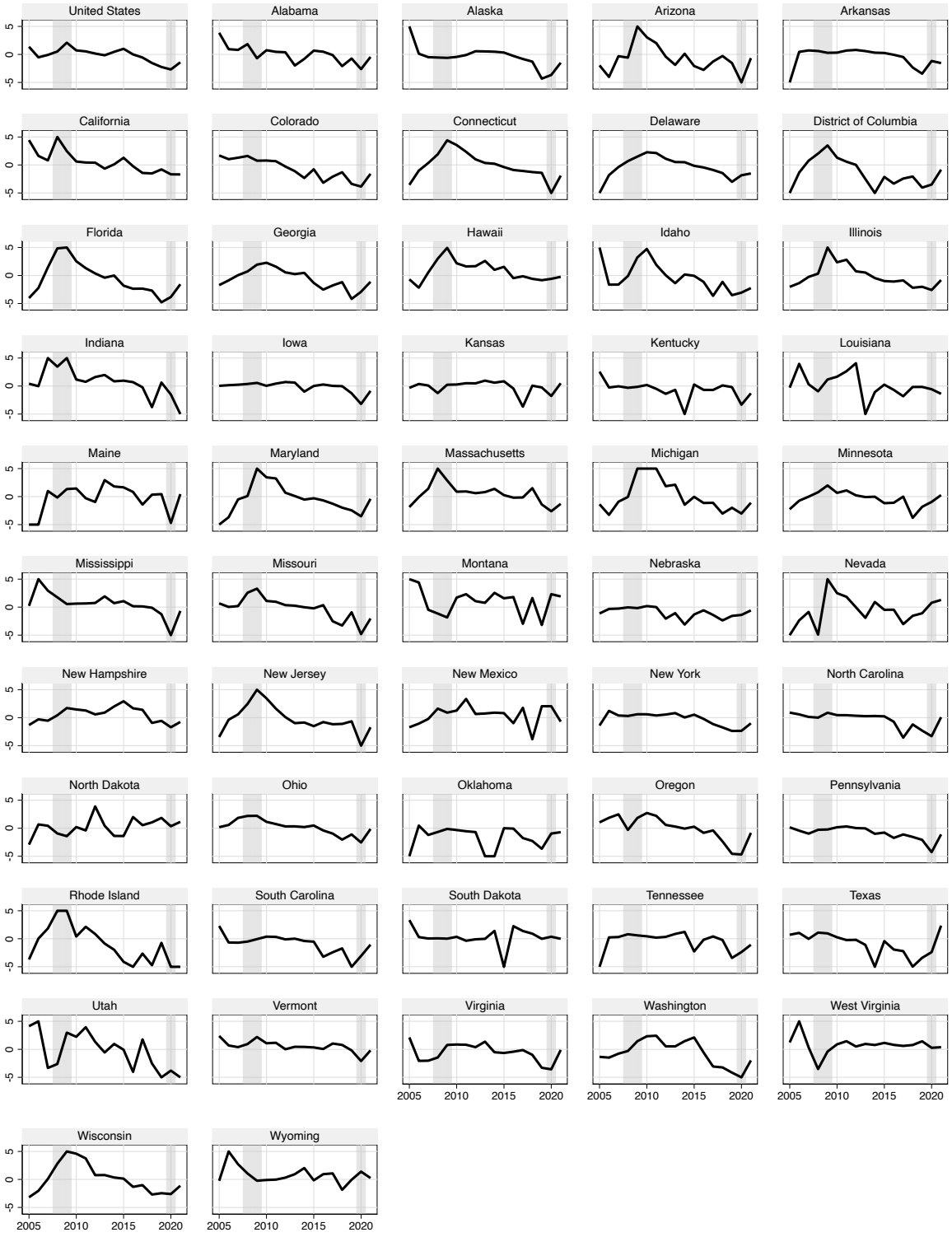


Figure 10: Trends in housing value Gini growth rates (2005-2020)

Notes: The top-left panel shows the evolution in the growth rate (%) of the housing value Gini index for the United States. The remaining panels describe the evolution of growth rates at the state level. Growth rates in the figure are capped at $+/-5\%$ for illustration purposes. Vertical gray bars indicate recession years.

cycle (Davis and Heathcote 2005, Iacoviello and Pavan 2013). Housing prices therefore rise faster during booms and slow (or fall) during recessions. Appendix Figure A10 shows house-price growth across states and confirms this pattern in our data.[8]

To the extent that housing demand may change differently at different segments of the housing value distribution, housing value inequality is also affected by the business cycle. Figure 10 examines this by showing the annual growth rates in the housing value Gini index from 2005 to 2020. The top-left panel suggests that value inequality increases immediately after recessions. Based on Appendix Figure A10, this appears driven by slower post-recession recovery of prices at the bottom of the distribution – a result consistent with low-income households spending a relatively higher share of income on housing (Dustmann et al. 2022) and with findings for Catalonia in Domènech-Arumí and Mariani (2025).

The distribution of housing assets is also influenced by the business cycle. Since renters hold no housing assets, downturns reduce asset inequality through the price mechanism (property values fall), as seen when focusing on the Great Recession years in Appendix Figure A6. However, business cycles also affect homeownership rates (e.g., low-income households may lose market access) and portfolio composition (Martínez-Toledano 2020), thereby reshaping the asset distribution. Our estimates cannot speak to the latter channel, given that we do not observe changes in the homeownership structure in the data.

All results in this subsection are descriptive, but they underscore an obvious point: business cycles (and housing demand more broadly) affect housing prices and inequality.

Other factors. Many other factors shape levels and trends in housing prices and inequality. We next examine zoning and land-use regulations as a standalone application for the estimates available in the WHID.

[8]Our estimates suggest that housing price growth slowed down before the Great Recession, a result consistent with the idea that housing could be a leading indicator of the business cycle (Leamer 2007; 2015). This did not occur during the COVID-19 recession, perhaps because of the very special nature of that recession or because housing is not as good a leading indicator as it used to be (Green 2022).

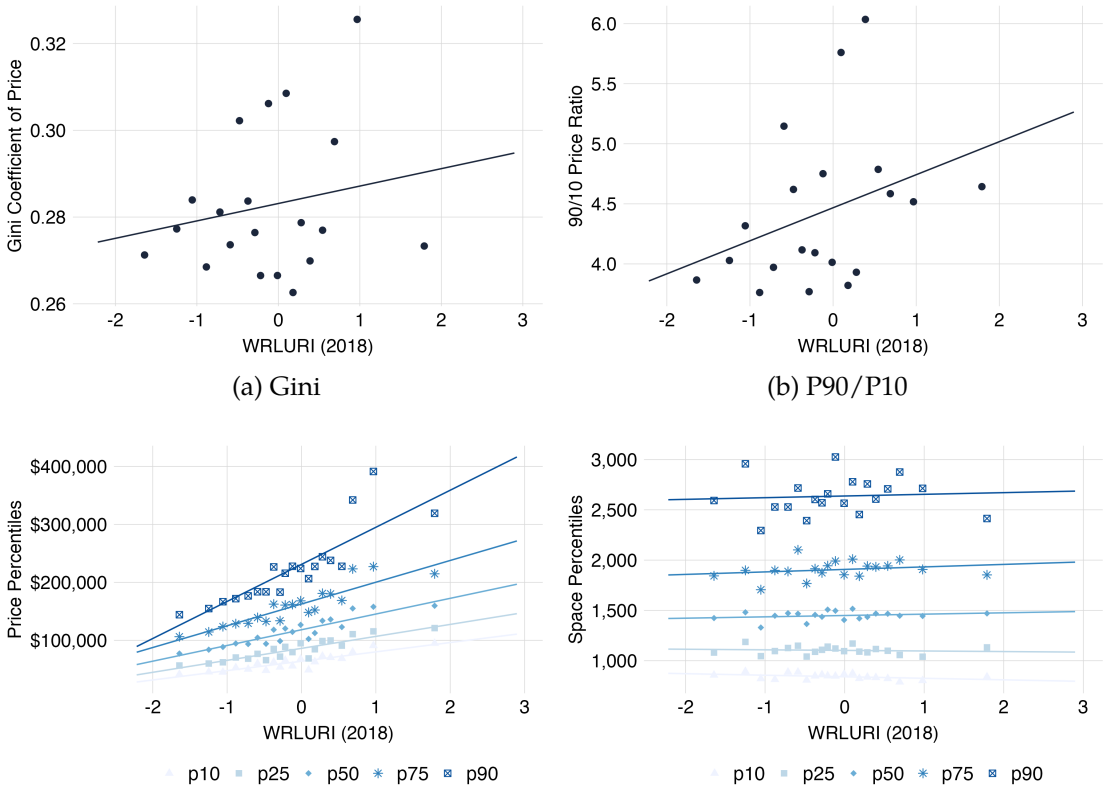
5 Application: zoning and housing prices and inequality

Zoning and land-use regulations in the U.S. are pervasive and highly localized, with municipalities exercising substantial control over land use. Originally introduced in the 1910s as a tool to protect home values from devaluation due to exposure to industry (Fischel 2004), these regulations determine where and what can be built, as well as the approval process. By determining the quantity, type, and location of housing, these regulations can affect housing inequality. We measure regulatory stringency using the updated 2018 Wharton Land Use Regulation Index (WRLURI), which summarizes each municipality's regulatory environment into a single composite measure (Gyourko et al. 2021).

Before presenting results, we caution against causal interpretations. Land-use policies are endogenous outcomes shaped by local demographic, economic, and political forces that likely influence both regulatory choices and housing market outcomes. Recent studies use exogenous zoning reforms and border discontinuity designs to identify causal effects (Büchler and Lutz 2024, Freemark 2020, Song 2025). In contrast, our analyses are descriptive: they provide motivating examples of how WHID inequality measures can be used in empirical work and highlight empirical patterns to motivate future research employing more rigorous identification strategies.

Our empirical analysis begins with a cross-sectional examination of the association between regulatory stringency and housing inequality at the Core-Based Statistical Area (CBSA) level as of 2018. While the WRLURI is measured at the zoning municipality level (typically cities, towns, or counties), we construct a CBSA-level index by averaging the WRLURI scores of all surveyed municipalities within each CBSA.

Figure 11 presents a binned scatter plot of the correlation between housing value inequality and regulatory stringency. The top panels plot the housing value Gini index and the 90/10 ratio. Both indicate that metropolitan areas with more stringent land-use regulations tend to exhibit greater disparities in housing values.



(a) Gini

(b) P90/P10

(c) Housing Value

(d) Housing Space

Figure 11: Zoning (WRLURI) and housing value and space inequality (CBSA level)

Notes: Binned scatterplots describing the correlation between the 2018 WRLURI and different measures of housing value and space inequality at the CBSA level. Panels (a) and (b) investigate the housing value Gini and 90/10 ratio. The 90/10 ratio is defined as the ratio of housing values at the 90th and 10th percentiles. Panel (c) displays a binned scatter plot showing how WRLURI correlates with different percentiles of housing values across CBSAs, while Panel (d) examines the corresponding relationship for housing space percentiles. Housing value is measured in current USD. Housing space is measured in square feet.

The bottom panels in Figure 11 explore whether the previous associations between regulation and housing value inequality are driven by dwellings at the top or bottom of the value distributions. Panel (c) shows that in metropolitan areas with more stringent zoning regulations, housing value inequality is more pronounced and driven primarily by higher values at the top. Specifically, the 10th percentile of housing values remains relatively similar across CBSAs, suggesting relatively little variation in lower-end housing. In contrast, the 90th percentile exhibits substantial divergence, with more regulated areas showing significantly higher housing values. Panel (d) examines

regulatory stringency and housing space (square footage), suggesting little variation in the distribution of housing space across CBSAs with differing levels of regulation; home sizes at both the lower and upper ends appear similar regardless of regulatory stringency. This implies that the observed housing value inequality is unlikely to be driven by differences in housing size.

Table 5: Correlation between WRLURI and Tract-Level Housing Inequality

	A. Housing value				B. Housing Space			
	Log(Mean)	Gini	Theil	P90/P10	Log(Mean)	Gini	Theil	P90/P10
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WRLURI	0.0725***	-0.0046***	-0.0047*	-0.0281	0.0159*	0.0069	0.0113	0.1490
	(0.0192)	(0.00118)	(0.0018)	(0.0393)	(0.0074)	(0.0036)	(0.0070)	(0.3079)
State FE	X	X	X	X	X	X	X	X
Adj-R^2	0.71	0.32	0.33	0.15	0.23	0.17	0.14	0.02
N	9,000	9,000	9,000	9,000	9,051	9,051	9,051	9,051

Notes: This table reports regression estimates examining the relationship between tract-level measures of housing value and housing space inequality and the 2018 municipality-level WRLURI. Standard errors are clustered at the state level. All regressions control for tract-level population and median income. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Our second part of the empirical analysis focuses on within-neighborhood variation in housing inequality. We test whether census tracts located in municipalities with more stringent land-use regulations exhibit different levels of within-neighborhood housing inequality than tracts in less regulated municipalities within the same state. We use census-tract housing inequality measures and merge each tract to its corresponding census place, which typically aligns with the municipality that holds zoning authority. We then assign municipality-level WRLURI values. Specifically, we estimate:

$$Ineq_{c,m,s} = \beta Zoning_m + \gamma X_{c,m,s} + \alpha_s + \varepsilon_{c.m,s} \tag{3}$$

where c indicates the census tract, m indicates the municipality, and $X_{c,m,s}$ includes tract-level population and median income, and α_s represents state fixed effects. Table 5 reports

regression estimates for various measures of housing value (Panel A) and space (Panel B) inequality at the tract level.

Panel A shows that tracts in municipalities with more stringent land-use regulations tend to have higher average housing values, consistent with previous research (Glaeser et al. 2005a, Kok et al. 2014). Moreover, these areas exhibit lower within-neighborhood inequality, suggesting that housing values within each tract are more homogeneous. In other words, stricter regulatory environments are associated with reduced dispersion in housing values at the neighborhood level.

Panel B indicates that tracts in more regulated municipalities tend to have larger homes, but we find no statistically or economically significant relationship between regulatory stringency and housing space inequality, consistent with the CBSA-level analysis. This indicates that variation in land-use regulation does not appear to meaningfully affect dispersion in housing sizes within tracts, even though stricter regulations are associated with larger home sizes.

Taken together, the results suggest that stricter land-use regulation is associated with less housing value dispersion within neighborhoods but greater disparities across neighborhoods. In other words, zoning tends to make individual neighborhoods more internally homogeneous while widening differences between them.

6 Discussion: what is housing inequality?

This final section discusses our perspective and interpretation of the housing inequality estimates described and analyzed throughout the paper. We follow the arguments outlined in Domènech-Arumí and Mariani (2025) closely.

Housing is multidimensional, and so is housing inequality. Housing space inequality concerns the distribution of space (square footage). Housing value inequality concerns the distribution of prices. Housing asset inequality concerns the distribution of housing

assets. Each is important to study on its own. For example, lack of housing space (overcrowding) is linked to lower test scores, reduced well-being, and lower fertility, among other outcomes (Goux and Maurin 2005, Kohl et al. 2024, Kulu and Vikat 2007, Solari and Mare 2012). Moreover, housing affordability is a pressing issue in many places, and the housing market is frequently targeted by government policy; mortgage interest deductions, rental subsidies, and rent control are examples. These policies shape the distributions of housing prices and assets (Braakmann and McDonald 2020, Glaeser and Shapiro 2003), so tracking them is essential for evaluating their impact.

Housing is also informative about other forms of inequality. In this work, we find a relatively strong association between housing value and income inequality, consistent with findings for Belgium (Domènech-Arumí and Gobbi 2025), Spain (Domènech-Arumí and Mariani 2025), and the United States (Aladangady et al. 2017). Because housing expenditures are a major component of total consumption, a high correlation between housing values and income is expected. Households at the top of the income distribution spend a smaller share on housing (non-homothethic preferences) (Albouy et al. 2016, Couture et al. 2024, Gaubert and Robert-Nicoud 2025), implying that the correlation between the two distributions is less than one. Even so, housing prices should contain meaningful information about the income distribution (which may be the primary object of interest for some). Future work should more formally characterize this relationship, taking into account that the level of geographic aggregation at which these distributions are measured is crucial given spatial sorting dynamics (Diamond and Gaubert 2022).

Beyond income, dwelling values capitalize local amenities and public goods (Baum-Snow and Hartley 2020, Bilal and Rossi-Hansberg 2021, Mariani 2025, De Moura et al. 2025), which is why a square foot of floor space is far more valuable in New York City than in upstate New York. Price dispersion across and within places (states, commuting zones, neighborhoods) shapes endogenous sorting, and through sorting, individuals can in turn influence price distributions (Diamond and Gaubert 2022). In short, housing prices both

reflect – and serve as a natural proxy for – the value of location (amenities, labor markets, public goods).

Housing space inequality is informative about *real* consumption inequalities. Moretti (2013) and Diamond and Moretti (2021) argue that, net of the cost of living (i.e., housing), income differences across U.S. cities are smaller than they appear, because high-income cities are less affordable. If housing expenditures could be cleanly decomposed into two components – location (land) and "actual" housing (e.g., floor space) – much of the observed consumption differences would be attributable to the location component. Recent evidence shows that land prices have risen faster in large metropolitan areas (Davis et al. 2021). Consistent with this view, we find that housing space inequality in the United States is substantially lower than housing value inequality, suggesting that net-of-housing consumption differences are indeed less stark.

Housing asset inequality is informative about housing wealth inequality and, to a lesser extent, overall wealth inequality. Housing wealth is housing assets net of liabilities. Lacking mortgage data, we cannot precisely measure housing wealth. However, this caveat does not apply to renters, who hold no housing assets or liabilities, and it is limited for homeowners (except in cases of default), who eventually pay off mortgages. Therefore, we can sign the bias: by ignoring leverage, our estimates overstate the true extent of housing wealth inequality.

Our results also shed light on overall wealth inequality. Housing is the most evenly distributed asset, and housing assets constitute a smaller share of the portfolios of the wealthiest – who are more heavily invested in financial assets (Kuhn et al. 2020, Zucman 2019). Consequently, our estimates of housing asset inequality likely understate the true extent of wealth inequality.

7 Conclusions

We constructed and analyzed novel estimates of housing inequality for the United States (1900-2020). Using county assessors' harmonized data on the full stock of dwellings (over 140 million), and property deed records to estimate market prices from 1980 to 2020, we examined the distributions of housing values, space, and assets. We found that housing inequality in the United States is high compared to Spain and, especially, Belgium. For example, in 2020, the Gini index for the distribution of dwelling values was 0.50 in the US, 0.37 in Catalonia, and 0.25 in Belgium. Housing space is more evenly distributed (Gini of 0.35), but housing assets are not (Gini of 0.68). Subnational variation is substantial, with California a major driver of high inequality due to elevated prices across the distribution, particularly at the top. We found that income and housing inequality are associated, as places with higher housing prices tend to have richer populations. Land-use regulations, changes in housing characteristics, and shifts in housing supply and demand are key factors behind cross-sectional and temporal variation in housing inequality.

Because households and dwellings are non-randomly distributed across space, aggregation masks substantial heterogeneity, implying that policies are likely to have heterogeneous geographic effects. Therefore, the estimates constructed here and made publicly available through the WHID project can help researchers and policymakers rigorously evaluate the impacts of policies (e.g., housing policy, land-use regulation, and monetary policy) on local and aggregate housing markets.

Beyond policy evaluation, future research should more formally define housing inequality by establishing explicit links between its dimensions (value, space, assets, and wealth) and concepts emphasized in other literatures (e.g., nominal and real income inequality, and overall wealth inequality).

References

Aladangady, Aditya, David Albouy, and Mike Zabek, "Housing inequality," *NBER*

Working Paper Series, 2017, *21916.*

Albouy, David, Gabriel Ehrlich, and Yingyi Liu, "Housing demand, cost-of-living inequality, and the affordability crisis," Technical Report, National Bureau of Economic Research 2016.

Avenancio-León, Carlos F and Troup Howard, "The assessment gap: Racial inequalities in property taxation," *The Quarterly Journal of Economics*, 2022, *137* (3), 1383–1434.

Baum-Snow, Nathaniel and Daniel Hartley, "Accounting for central neighborhood change, 1980–2010," *Journal of urban economics*, 2020, *117*, 103228.

Berry, Christopher R, "Reassessing the property tax," *Available at SSRN 3800536*, 2021.

Bilal, Adrien and Esteban Rossi-Hansberg, "Location as an Asset," *Econometrica*, 2021, *89* (5), 2459–2495.

Blanchet, Thomas and Clara Martínez-Toledano, "Wealth inequality dynamics in Europe and the United States: Understanding the determinants," *Journal of Monetary Economics*, 2023, *133*, 25–43.

Braakmann, Nils and Stephen McDonald, "Housing subsidies and property prices: Evidence from England," *Regional science and urban economics*, 2020, *80*, 103374.

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

Büchler, Simon and Elena Lutz, "Making housing affordable? The local effects of relaxing land-use regulation," *Journal of Urban Economics*, 2024, *143*, 103689.

Čeh, Marjan, Milan Kilibarda, Anka Lisec, and Branislav Bajat, "Estimating the performance of random forest versus multiple regression for predicting prices of the apartments," *ISPRS international journal of geo-information*, 2018, *7* (5), 168.

Chancel, Lucas, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman, *World inequality report 2022*, Harvard University Press, 2022.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Chyn, Eric and Lawrence F Katz, "Neighborhoods matter: Assessing the evidence for place effects," *Journal of Economic Perspectives*, 2021, *35* (4), 197–222.

Couture, Victor, Cecile Gaubert, Jessie Handbury, and Erik Hurst, "Income growth and the distributional effects of urban spatial sorting," *Review of Economic Studies*, 2024, *91* (2), 858–898.

Davis, Morris A and Jonathan Heathcote, "Housing and the business cycle," *International Economic Review*, 2005, *46* (3), 751–784.

—— , **William D Larson, Stephen D Oliner, and Jessica Shui**, "The price of residential land for counties, ZIP codes, and census tracts in the United States," *Journal of Monetary Economics*, 2021, *118*, 413–431.

Derenoncourt, Ellora, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick, "Wealth of two nations: The US racial wealth gap, 1860–2020," *The Quarterly Journal of Economics*, 2024, *139* (2), 693–750.

Diamond, Rebecca and Cecile Gaubert, "Spatial sorting and inequality," *Annual Review of Economics*, 2022, *14* (1), 795–819.

—— **and Enrico Moretti**, "Where is standard of living the highest? Local prices and the geography of consumption," Technical Report, National Bureau of Economic Research 2021.

Domènech-Arumí, Gerard and Giovanni Paolo Mariani, "Housing Inequality in Spain," *Available at SSRN 5688222*, 2025.

—— **and Paula E Gobbi**, "Housing inequality in Belgium: Novel estimates and a policy application," *Available at SSRN*, 2025.

Dustmann, Christian, Bernd Fitzenberger, and Markus Zimmermann, "Housing expenditure and income inequality," *The Economic Journal*, 2022, *132* (645), 1709–1736.

Echenique, Federico and Roland G Fryer Jr, "A measure of segregation based on social interactions," *The Quarterly Journal of Economics*, 2007, *122* (2), 441–485.

Fischel, William A, "An economic history of zoning and a cure for its exclusionary effects," *Urban Studies*, 2004, *41* (2), 317–340.

Fogli, Alessandra and Veronica Guerrieri, "The end of the american dream? inequality and segregation in us cities," Technical Report, National Bureau of Economic Research 2019.

Freemark, Yonah, "Upzoning Chicago: Impacts of a zoning reform on property values and housing construction," *Urban affairs review*, 2020, *56* (3), 758–789.

Ganong, Peter and Daniel Shoag, "Why has regional income convergence in the US declined?," *Journal of Urban Economics*, 2017, *102*, 76–90.

Gaubert, Cécile and Frédéric Robert-Nicoud, "Sorting to expensive cities," Technical Report, National Bureau of Economic Research 2025.

Glaeser, Edward L and Jesse M Shapiro, "The benefits of the home mortgage interest deduction," *Tax policy and the economy*, 2003, *17*, 37–82.

—— **, Joseph Gyourko, and Raven E Saks**, "Why have housing prices gone up?," *American Economic Review*, 2005, *95* (2), 329–333.

___ , ___ , **and Raven Saks**, "Why is Manhattan so expensive? Regulation and the rise in housing prices," *The Journal of Law and Economics*, 2005, *48* (2), 331–369.

___ , **Matt Resseger, and Kristina Tobio**, "Inequality in cities," *Journal of regional science*, 2009, *49* (4), 617–646.

Goux, Dominique and Eric Maurin, "The effect of overcrowded housing on children's performance at school," *Journal of Public economics*, 2005, *89* (5-6), 797–819.

Green, Richard K, "Is housing still the business cycle? Perhaps not.," in "Handbook of real estate and macroeconomics," Edward Elgar Publishing, 2022, pp. 269–283.

Guvenen, Fatih, Luigi Pistaferri, and Giovanni L Violante, "Global trends in income inequality and income dynamics: New insights from GRID," *Quantitative Economics*, 2022, *13* (4), 1321–1360.

Guzman, Gloria and Melissa Kollar, "Income in the United States: 2022," *United States Census Bureau. https://www. census. gov/content/dam/Census/library/publications/2023/demo/p60-279. pdf*, 2023.

Gyourko, Joseph, Jonathan S Hartley, and Jacob Krimmel, "The local residential land use regulatory environment across US housing markets: Evidence from a new Wharton index," *Journal of Urban Economics*, 2021, *124*, 103337.

Howard, Greg and Jack Liebersohn, "How regional inequality and migration drive housing prices and rents," *Journal of Economic Perspectives*, 2025, *39* (3), 3–26.

Iacoviello, Matteo and Marina Pavan, "Housing and debt over the life cycle and over the business cycle," *Journal of Monetary Economics*, 2013, *60* (2), 221–238.

Ioannides, Yannis M and L Rachel Ngai, "Housing and inequality," *Journal of Economic Literature*, 2025, *63* (3), 916–963.

Kermani, Amir and Francis Wong, "Racial disparities in housing returns," Technical Report, National Bureau of Economic Research 2021.

Knoll, Katharina, Moritz Schularick, and Thomas Steger, "No price like home: Global house prices, 1870–2012," *American Economic Review*, 2017, *107* (2), 331–353.

Kohl, Sebastian, Max Friedrich Steinhardt, Luca Stella, and Simon Voss, "Crowding (at) the margins: Investigating the unequal distribution of housing space in Germany," Technical Report, Discussion Paper 2024.

Kok, Nils, Paavo Monkkonen, and John M Quigley, "Land use regulations and the value of land and housing: An intra-metropolitan analysis," *Journal of Urban Economics*, 2014, *81*, 136–148.

Kuhn, Moritz, Moritz Schularick, and Ulrike I Steins, "Income and wealth inequality in America, 1949–2016," *Journal of Political Economy*, 2020, *128* (9), 3469–3519.

Kulu, Hill and Andres Vikat, "Fertility differences by housing type: The effect of housing conditions or of selective moves?," *Demographic research*, 2007, *17*, 775–802.

Leamer, Edward E, "Housing is the business cycle," 2007.

___ , "Housing really is the business cycle: what survives the lessons of 2008–09?," *Journal of Money, Credit and Banking*, 2015, *47* (S1), 43–50.

Logan, Trevon D and John M Parman, "Racial residential segregation in the United States," *Journal of Economic Literature*, 2025, *63* (3), 964–1010.

Lyons, Ronan C, Allison Shertzer, Rowena Gray, and David Agorastos, "The price of housing in the United States, 1890–2006," *The Quarterly Journal of Economics*, 2025, p. qjaf047.

Mariani, Giovanni Paolo, "The Value of Local Public Goods: Evidence from Massachusetts' Property Tax Limits," *Job Market Paper*, 2025.

Martínez-Toledano, Clara, "House price cycles, wealth inequality and portfolio reshuffling," *WID. World Working Paper*, 2020, *2*.

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

Moura, Micael Castanheira De, Giovanni Paolo Mariani, and Clemence Tricaud, "Do Public Goods Actually Reduce Inequality?," *CEPR Discussion Paper No. 20446*, 2025.

Mullainathan, Sendhil and Jann Spiess, "Machine Learning: An Applied Econometric Approach," *Journal of Economic Perspectives*, May 2017, *31* (2), 87–106.

Nieuwerburgh, Stijn Van and Pierre-Olivier Weill, "Why has house price dispersion gone up?," *The Review of Economic Studies*, 2010, *77* (4), 1567–1606.

OECD, *Brick by Brick* 2021.

___ , "OECD Affordable Housing Database," 2024. https://www.oecd.org/housing/data/affordable-housing-database/. Last accessed May 2024.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States, 1913–1998*," *The Quarterly Journal of Economics*, 02 2003, *118* (1), 1–41.

___ **and Gabriel Zucman**, "Capital is back: Wealth-income ratios in rich countries 1700–2010," *The Quarterly Journal of Economics*, 2014, *129* (3), 1255–1310.

___ **, Emmanuel Saez, and Gabriel Zucman**, "Distributional national accounts: methods and estimates for the United States," *The Quarterly Journal of Economics*, 2018, *133* (2), 553–609.

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

Solari, Claudia D and Robert D Mare, "Housing crowding effects on children's wellbeing," *Social science research*, 2012, *41* (2), 464–476.

Song, Jaehee, "The effects of residential zoning in US housing markets," *Available at SSRN 3996483*, 2025.

U.S. Census Bureau, "American Housing Survey," 2021. Access date: 2025-11-14.

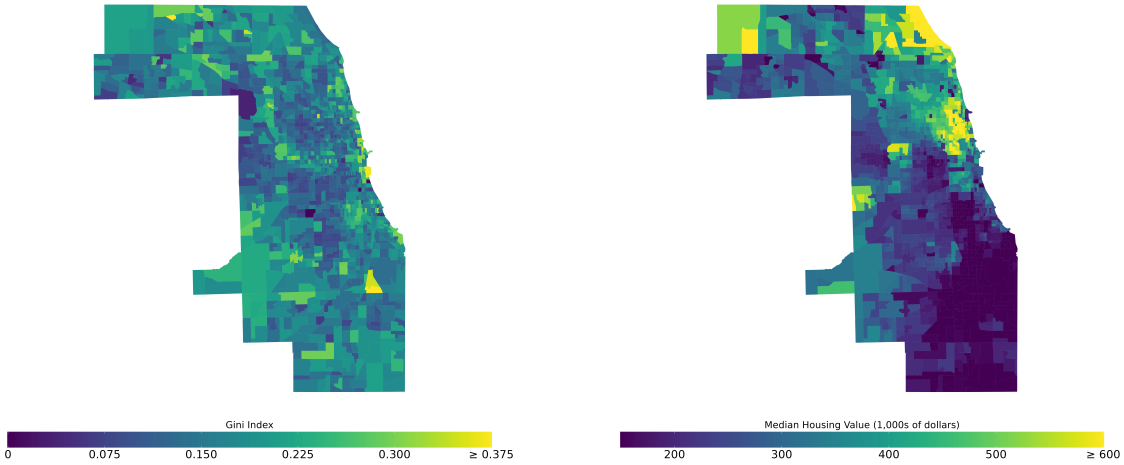
___ , "Homeownership Rate in the United States," FRED, Federal Reserve Bank of St. Louis. Series ID: RHORUSQ156N. Retrieved November 21, 2025.

Zucman, Gabriel, "Global wealth inequality," *Annual Review of Economics*, 2019, *11* (1), 109–138.

Web Appendix of
Housing Inequality in the United States

Gerard Domènech-Arumí, Lei Ma, and Giovanni Paolo Mariani

A Supporting figures and tables



(a) Gini index by census tract (b) Median housing value by census tract

Figure A1: Housing value inequality in the Chicago

Notes: Each polygon represents a census tract in Cook County, IL. The left panel reports the Gini index. The right panel reports the median housing value. Housing value is measured in thousands of 2020 USD.

Table A1: Housing inequality in the United States: Theil decomposition

	Housing Value			Housing Space		
	Theil	Theil-Between	Theil-Within	Theil	Theil-Between	Theil-Within
Alabama	0.219	0.045	0.174	0.248	0.022	0.226
Alaska	0.144	0.043	0.101	0.271	0.008	0.263
Arizona	0.166	0.038	0.128	0.483	0.007	0.476
Arkansas	0.079	0.012	0.067	0.189	0.006	0.182
California	0.64	0.125	0.515	0.233	0.007	0.226
Colorado	0.137	0.059	0.078	0.323	0.029	0.294
Connecticut	0.186	0.051	0.134	0.196	0.004	0.192
Delaware	0.125	0.004	0.12	0.096	0.003	0.093
District of Columbia	0.091	0	0.091	0.625	0	0.625
Florida	0.22	0.052	0.168	0.386	0.018	0.368
Georgia	0.185	0.074	0.111	0.252	0.009	0.243
Hawaii	0.2	0.048	0.152	0.322	0.001	0.32
Idaho	0.067	0.029	0.039	0.136	0.011	0.125
Illinois	0.19	0.055	0.135	0.195	0.011	0.184
Indiana	0.164	0.038	0.126	0.35	0.011	0.338
Iowa	0.13	0.038	0.092	0.324	0.022	0.302
Kansas	0.161	0.076	0.085	0.356	0.026	0.33
Kentucky	0.164	0.078	0.085	0.225	0.034	0.191
Louisiana	0.151	0.036	0.115	0.301	0.005	0.295
Maine	0.082	0.042	0.04	0.171	0.016	0.155
Maryland	0.152	0.054	0.098	0.349	0.01	0.339
Massachusetts	0.137	0.059	0.078	0.373	0.003	0.37
Michigan	0.184	0.051	0.133	0.202	0.01	0.192
Minnesota	0.116	0.053	0.063	0.425	0.021	0.405
Mississippi	0.155	0.049	0.106	0.319	0.031	0.288
Missouri	0.145	0.03	0.115	0.24	0.013	0.227
Montana	0.101	0.046	0.055	0.237	0.005	0.232
Nebraska	0.141	0.049	0.092	0.284	0.006	0.277
Nevada	0.153	0.016	0.136	0.366	0.014	0.352
New Hampshire	0.125	0.037	0.088	0.15	0.002	0.148
New Jersey	0.082	0.021	0.061	0.178	0.005	0.173
New Mexico	0.095	0.009	0.085	0.267	0.011	0.256
New York	0.319	0.233	0.085	0.876	0.098	0.778
North Carolina	0.189	0.061	0.128	0.339	0.038	0.301
North Dakota	0.07	0.012	0.058	0.203	0.01	0.193
Ohio	0.179	0.041	0.137	0.359	0.022	0.337
Oklahoma	0.151	0.027	0.124	0.257	0.006	0.251
Oregon	0.119	0.049	0.07	0.248	0.01	0.238
Pennsylvania	0.193	0.064	0.129	0.249	0.021	0.228
Rhode Island	0.072	0.013	0.059	0.17	0.003	0.167
South Carolina	0.231	0.064	0.166	0.241	0.044	0.197
South Dakota	0.193	0.071	0.122	0.162	0.017	0.145
Tennessee	0.181	0.066	0.115	0.261	0.008	0.253
Texas	0.11	0.044	0.066	0.32	0.061	0.259
Utah	0.045	0.006	0.038	0.337	0.027	0.31
Vermont	0.144	0.038	0.106	0.106	0.003	0.102
Virginia	0.223	0.139	0.084	0.291	0.026	0.265
Washington	0.147	0.067	0.08	0.325	0.009	0.316
West Virginia	0.305	0.093	0.212	0.096	0.006	0.089
Wisconsin	0.126	0.027	0.099	0.396	0.093	0.303
Wyoming	0.109	0.049	0.06	0.408	0.052	0.356

Notes: *Theil* is the Theil index. *Theil-Between* and *Theil-Within* decompose the Theil index into a between and within (county) component. Housing space is measured in square feet. Housing value is measured in thousands of 2020 USD.

Table A2: Housing value inequality (price per square foot) in the United States (2020)

	Gini	Top 1	Top 10	Bottom 50	P99	P90	P50	P10
A. Country								
United States	0.613	0.26	0.546	0.143	3211	451	136	52
B. State								
Alabama	0.292	0.052	0.23	0.304	300	122	74	31
Alaska	0.319	0.046	0.256	0.293	661	348	175	66
Arizona	0.369	0.111	0.311	0.272	1525	244	155	48
Arkansas	0.152	0.024	0.153	0.396	163	115	90	59
California	0.689	0.172	0.618	0.1	13153	2869	378	158
Colorado	0.391	0.061	0.312	0.249	1361	518	211	87
Connecticut	0.322	0.085	0.289	0.299	995	342	170	104
Delaware	0.278	0.054	0.246	0.321	621	255	147	82
District of Columbia	0.388	0.068	0.276	0.231	3441	1156	570	186
Florida	0.401	0.069	0.316	0.245	959	340	147	38
Georgia	0.278	0.042	0.227	0.315	349	176	103	50
Hawaii	0.367	0.076	0.28	0.262	2683	773	450	115
Idaho	0.209	0.034	0.187	0.358	432	242	163	99
Illinois	0.373	0.072	0.293	0.254	660	287	127	57
Indiana	0.309	0.051	0.229	0.287	298	142	83	33
Iowa	0.317	0.043	0.234	0.282	311	184	91	39
Kansas	0.329	0.074	0.283	0.288	804	203	103	58
Kentucky	0.325	0.079	0.257	0.286	411	135	75	32
Louisiana	0.298	0.059	0.283	0.315	456	151	85	53
Maine	0.286	0.051	0.239	0.31	774	399	214	117
Maryland	0.262	0.032	0.197	0.32	419	277	168	71
Massachusetts	0.37	0.065	0.294	0.258	1148	514	220	100
Michigan	0.26	0.033	0.194	0.316	276	169	109	47
Minnesota	0.306	0.056	0.244	0.3	703	278	152	68
Mississippi	0.233	0.032	0.192	0.339	143	88	54	30
Missouri	0.255	0.044	0.209	0.327	352	179	109	59
Montana	0.384	0.097	0.315	0.253	1022	264	133	58
Nebraska	0.279	0.051	0.22	0.312	360	146	95	40
Nevada	0.387	0.048	0.299	0.25	1093	627	192	113
New Hampshire	0.314	0.067	0.257	0.296	743	307	170	77
New Jersey	0.222	0.033	0.203	0.35	543	340	185	133
New Mexico	0.168	0.028	0.173	0.388	250	175	112	82
New York	0.548	0.076	0.36	0.118	2316	1166	292	47
North Carolina	0.315	0.067	0.25	0.293	472	177	106	43
North Dakota	0.191	0.02	0.163	0.365	241	206	133	82
Ohio	0.425	0.096	0.349	0.23	836	163	92	27
Oklahoma	0.245	0.048	0.207	0.336	275	122	82	44
Oregon	0.293	0.045	0.229	0.306	706	364	206	83
Pennsylvania	0.44	0.133	0.372	0.229	1460	239	121	45
Rhode Island	0.261	0.046	0.228	0.33	796	420	224	138
South Carolina	0.31	0.052	0.236	0.29	310	149	86	34
South Dakota	0.267	0.033	0.204	0.313	231	129	81	37
Tennessee	0.291	0.044	0.239	0.305	401	192	102	56
Texas	0.358	0.084	0.313	0.278	897	249	120	64
Utah	0.383	0.059	0.356	0.265	1748	514	224	135
Vermont	0.286	0.044	0.223	0.303	419	227	123	62
Virginia	0.361	0.061	0.282	0.266	648	318	140	53
Washington	0.486	0.112	0.399	0.198	3176	775	252	98
West Virginia	0.406	0.069	0.272	0.211	185	108	48	14
Wisconsin	0.343	0.089	0.295	0.282	770	255	130	68
Wyoming	0.279	0.049	0.233	0.316	545	238	136	73

Notes: *Gini* is the Gini index. *Top 1* and *Top 10* measure the share of housing value (psqft) that accrues to the top 1 or 10% most valuable dwellings in the geography, respectively. *Bottom 50* measure the share of housing value (psqft) accrued to the 50% least valuable dwellings. *P99*, *P90*, *P50*, and *P10* indicate the value (psqft) of the dwelling in the 99th, 90th, 50th, and 10th percentiles of the dwelling value (psqft) distribution in the geography, respectively. Housing value (psqft) is measured in 2020 USD.



Figure A2: Trends in housing value inequality (P10, P50, P90) across states

Notes: The top-left panel describes the evolution of housing value inequality (P10, P50, P90) for the United States. The remaining panels report the evolution of housing value inequality at the state level. Housing value is measured in millions of 2020 USD (adjusted for inflation).



Figure A3: Trends in housing value (psqft) inequality (P10, P50, P90) across states

Notes: The top-left panel describes the evolution of housing value inequality (P10, P50, P90) for the United States. The remaining panels report the evolution of housing value inequality at the state level. Housing price per square foot is measured in 1,000s of 2020 USD (adjusted for inflation).



Figure A4: Evolution of housing space inequality (P10, P50, and P90) across states

Notes: The top-left panel describes the evolution of housing space inequality (P10, P50, and P90) for the United States. The remaining panels report the evolution of housing space inequality at the state level. Housing space is measured in 1,000s of square feet.

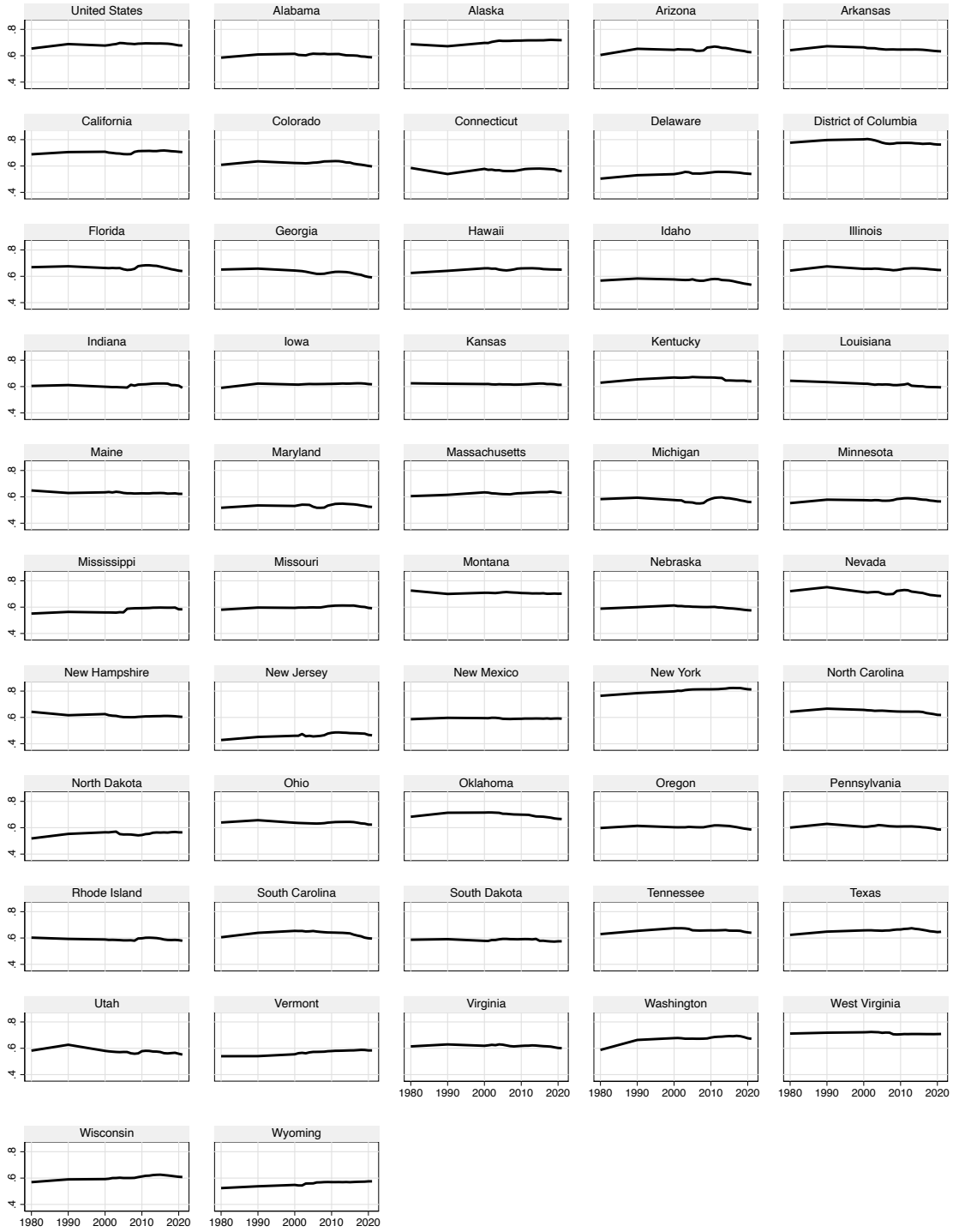


Figure A5: Trends in housing asset inequality (Gini) across states

Notes: Inequality is measured with the Gini index. The top-left panel describes the evolution of housing space inequality (Gini) for the United States. The remaining panels report the evolution of housing space inequality at the state level. Housing assets are measured in millions of current USD.

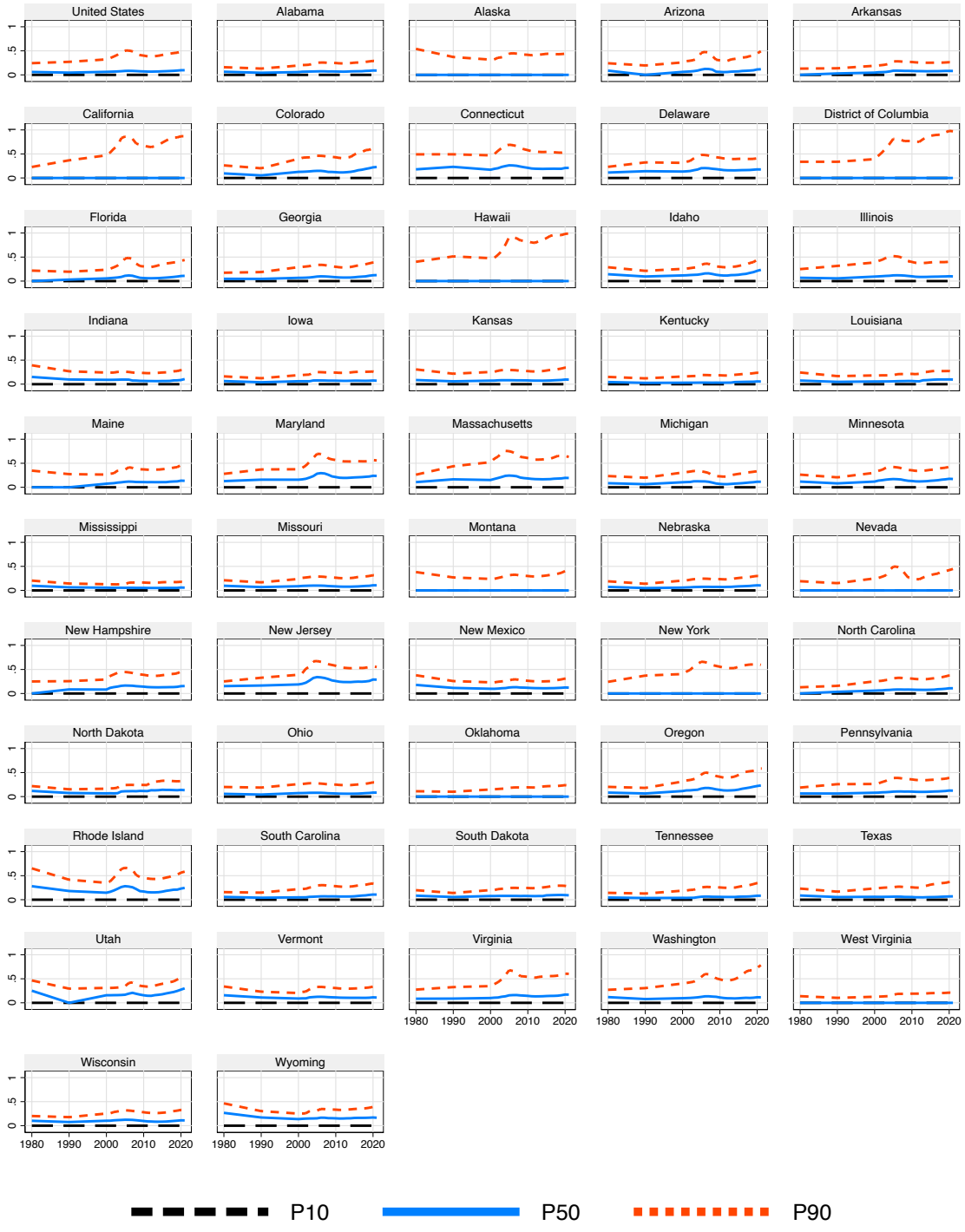


Figure A6: Trends in housing assets inequality (P10, P50, P90) across states

Notes: The top-left panel describes the evolution of housing value inequality (P10, P50, P90) for the United States. The remaining panels report the evolution of housing value inequality at the state level. Housing assets are measured in millions of 2020 USD (adjusted for inflation).

Table A3: The association between housing inequality and income

A. Housing Value

	Housing Value Gini				log Median Housing Value			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income Gini	0.196***	0.170***	0.215***	0.172***		1.399***		1.077***
	(0.040)	(0.039)	(0.005)	(0.004)		(0.201)		(0.020)
log Median Income		-0.023**		-0.005***	1.435***	1.149***	0.892***	0.498***
		(0.007)		(0.001)	(0.032)	(0.039)	(0.005)	(0.004)
log Population		0.004**		0.005***		0.077***		0.044***
		(0.001)		(0.001)		(0.005)		(0.003)
DV Mean	0.246	0.246	0.198	0.198	11.760	11.760	12.330	12.330
DV SD	0.078	0.078	0.096	0.096	0.542	0.542	0.758	0.758
Adj-R^2	0.009	0.402	0.024	0.472	0.456	0.744	0.313	0.839
N	2759	2759	77405	77405	2759	2759	77405	77405
FE		State		County		State		County
Aggregation	County	County	Tract	Tract	County	County	Tract	Tract

B. Housing Space

	Housing Space Gini				log Median Housing Space			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Income Gini	0.108*	0.283***	0.321***	0.296***		0.177		0.435***
	(0.046)	(0.046)	(0.006)	(0.006)		(0.128)		(0.021)
log Median Income		0.034***		-0.030***	0.222***	0.272***	0.334***	0.423***
		(0.009)		(0.001)	(0.016)	(0.025)	(0.003)	(0.004)
log Population		0.015***		-0.009***		0.013**		0.022***
		(0.001)		(0.001)		(0.004)		(0.004)
DV Mean	0.259	0.259	0.251	0.251	14.194	14.194	14.205	14.205
DV SD	0.090	0.090	0.105	0.105	0.218	0.218	0.357	0.357
Adj-R^2	0.002	0.333	0.045	0.295	0.068	0.222	0.198	0.441
N	2847	2847	77980	77980	2847	2847	77980	77980
FE		State		County		State		County
Aggregation	County	County	Tract	Tract	County	County	Tract	Tract

Notes: Housing value is measured in 2020 USD. Housing space is measured in square feet. *Income Gini*, *Income Gini*, *log Median Income*, and *log Population* are retrieved from the ACS and measured in 2020. Heteroskedasticity-robust standard errors in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

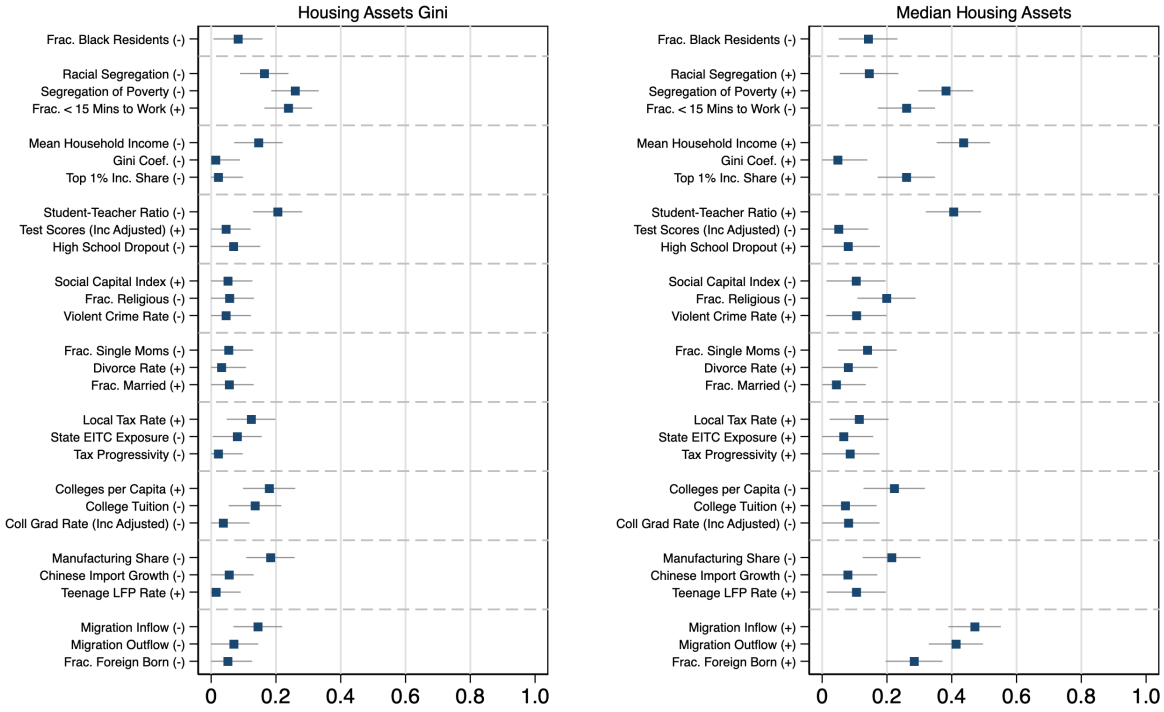


Figure A7: The determinants of contemporaneous housing asset inequality

Notes: This figure reports the raw correlations between housing asset inequality estimates (Gini and median) and several covariates at the 1990 Commuting Zone level and retrieved from Chetty et al. (2014). Most covariates are measured in 2000. See Section G in their online Appendix for more details. The + and − signs next to a covariate label denote a positive and negative correlation, respectively. Housing assets are measured in 2020 1,000s of USD.



Figure A8: Housing space inequality (P10, P50, P90) by decade-cohort across states

Notes: The top-left panel describes the evolution of housing space inequality (Gini) by decade-cohort for the United States. The remaining panels capture the evolution of the same outcome across the states. Estimates are constructed based on dwellings entering the housing stock in the previous decade (e.g., the 1910 estimate is based on dwellings built between 1901 and 1910). Housing space is measured in square feet.

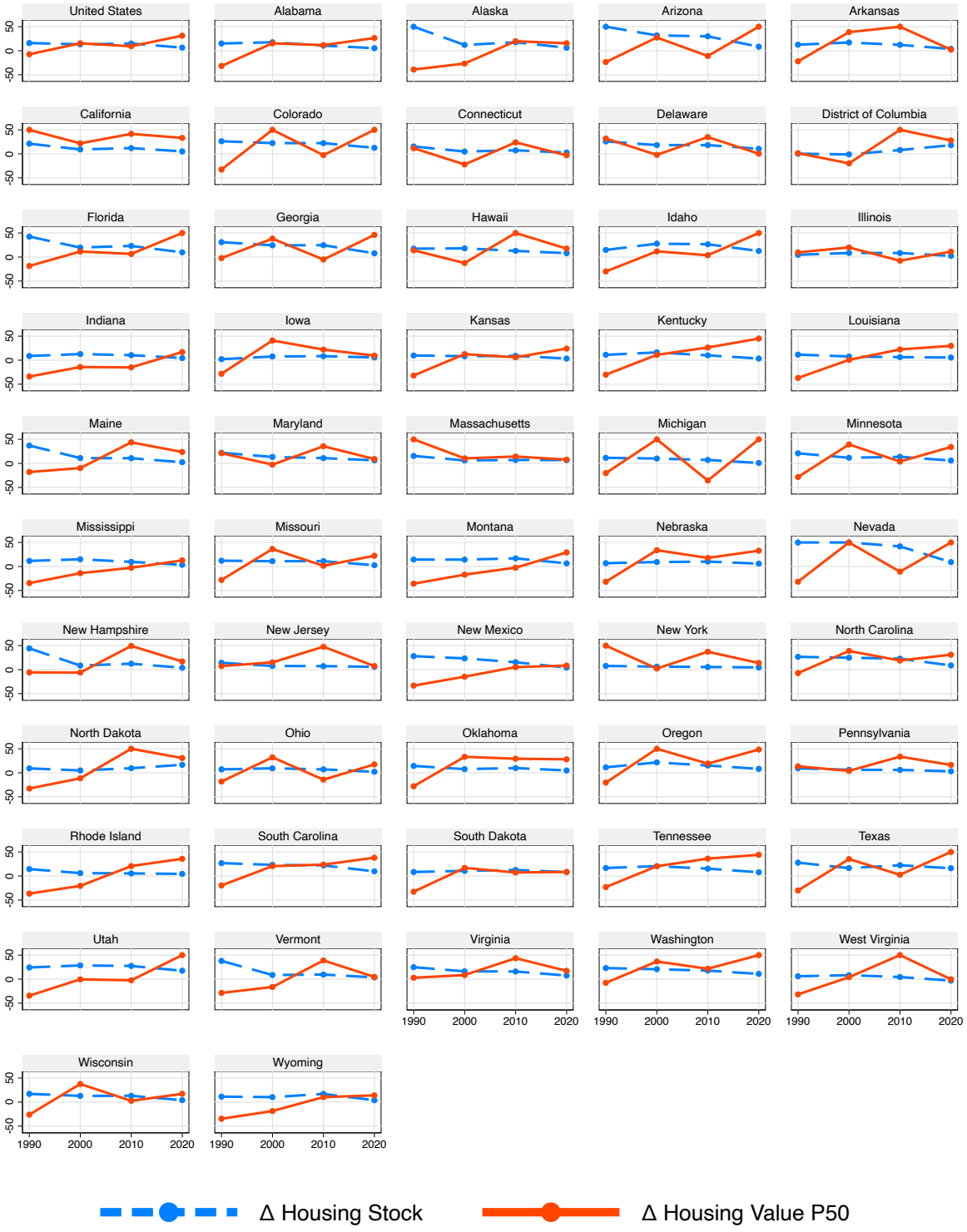


Figure A9: Growth in the housing stock and prices over time across states

Notes: Δ *Housing Stock* is the decennial percentage change in the housing stock based on Decennial Census data. Δ *Housing Value P50* is the decennial percentage change in inflation-adjusted median housing price based on our estimates. Growth rates are censored at 50% for illustration purposes.

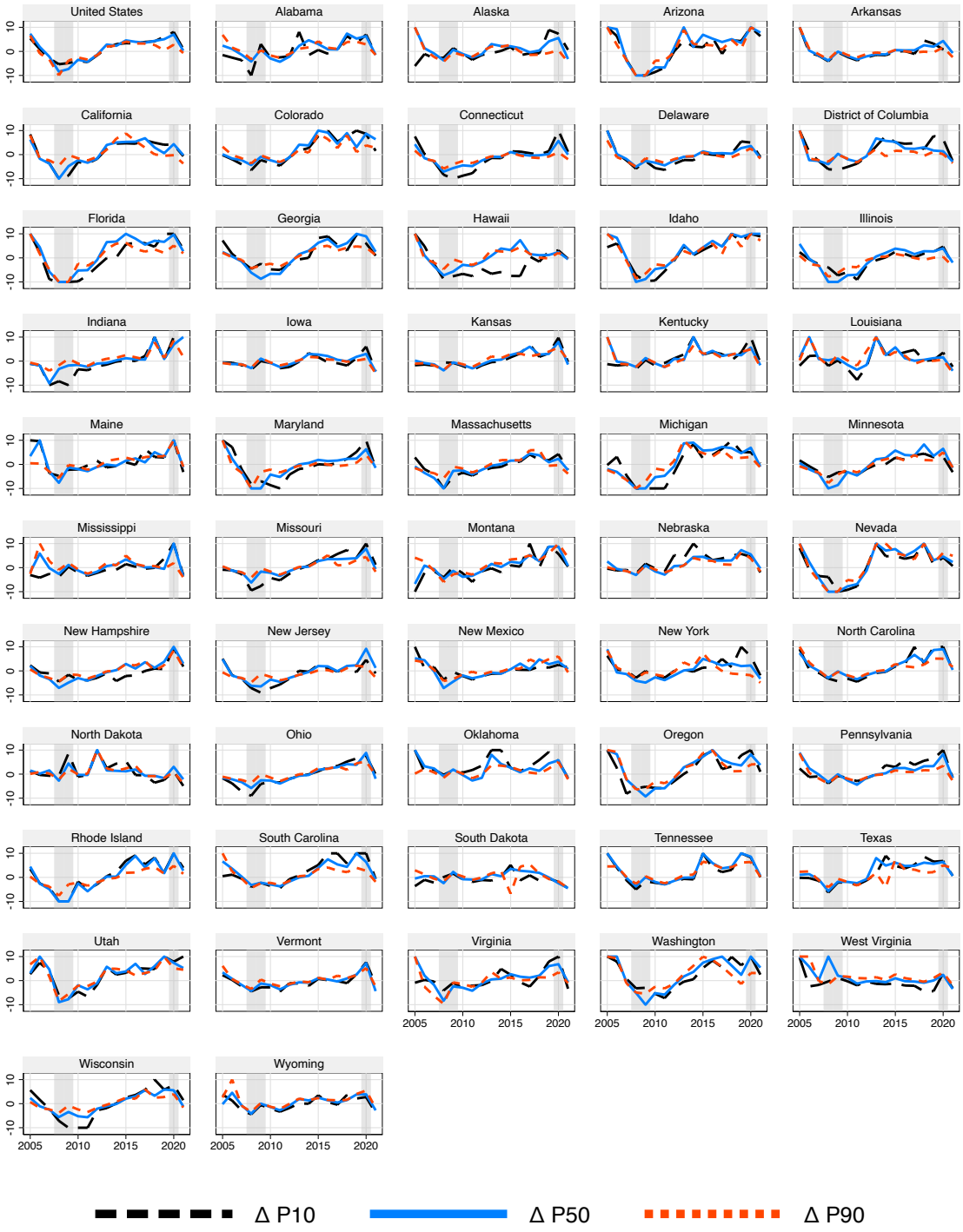


Figure A10: Trends in housing value (P10, P50, P90) growth rates (2005-2020)

Notes: The top-left panel shows the evolution in the growth rate (%) of the housing value at the percentiles 10, 50, and 90 for the United States (2005-21). The rest of the panels describe the evolution in the growth rates at the state level over the same period. Housing values are measured in 1,000s of 2020 USD (inflation-adjusted). Growth rates in the figure are capped at $+/-10\%$ for illustration purposes. Vertical gray bars indicate recession years.

Table A4: Changes in housing inequality and housing stock

A. Housing Value	Δ Housing Value P10				Δ Housing Value P50				Δ Housing Value P90			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Δ Housing Stock	-5.382***	-0.188	-0.000	-0.001	-4.157**	0.106	0.000	0.000	-2.586*	0.731	0.003	0.003
	(1.279)	(0.366)	(0.001)	(0.002)	(1.522)	(0.092)	(0.002)	(0.002)	(1.057)	(0.380)	(0.003)	(0.004)
Δ Population	0.409	1.647***	1.164***	0.367**	-0.918	1.560***	2.673***	0.214**	-1.786	0.907*	3.904***	0.160**
	(2.534)	(0.410)	(0.320)	(0.139)	(3.374)	(0.413)	(0.281)	(0.080)	(1.560)	(0.398)	(0.289)	(0.061)
DV Mean	3.135	11.085	12.603	34.825	11.931	18.091	18.510	31.232	14.196	21.091	21.673	20.683
DV SD	28.971	31.859	37.187	44.159	29.697	25.714	30.480	33.965	26.632	26.269	28.840	29.462
Adj-R^2	0.090	0.011	0.002	0.002	0.059	0.016	0.015	0.001	0.041	0.018	0.040	0.001
N	203	1978	7846	48222	203	1978	7846	48222	203	1978	7846	48222
Aggregation	State	CZ	County	Tract	State	CZ	County	Tract	State	CZ	County	Tract
B. Housing Space	Δ Housing Space P10				Δ Housing Space P50				Δ Housing Space P90			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Δ Housing Stock	-0.259	0.111	-0.000	-0.002	-0.698*	0.084	0.000*	-0.001	-0.575***	0.049	0.001	-0.000
	(0.275)	(0.110)	(0.000)	(0.002)	(0.337)	(0.068)	(0.000)	(0.001)	(0.082)	(0.051)	(0.001)	(0.001)
Δ Population	0.715	0.280**	1.174***	0.291	2.884	0.372*	1.178***	0.004	0.520	0.266***	0.958***	0.041
	(0.397)	(0.088)	(0.085)	(0.205)	(1.971)	(0.147)	(0.057)	(0.037)	(0.279)	(0.075)	(0.129)	(0.077)
DV Mean	1.308	1.536	1.706	0.335	1.456	1.920	1.592	0.473	0.424	2.546	2.343	1.674
DV SD	7.542	5.893	7.253	9.227	19.320	8.580	5.445	9.904	6.291	4.751	8.516	41.928
Adj-R^2	0.028	0.015	0.051	0.027	0.049	0.009	0.089	-0.000	0.187	0.015	0.027	-0.000
N	560	1999	8155	48809	560	1999	8155	48809	560	1999	8155	48809
Aggregation	State	CZ	County	Tract	State	CZ	County	Tract	State	CZ	County	Tract

Notes: Dependent variables characterize the decennial percentage change in a percentile of the distribution of housing values or space. Δ*Housing Stock* is the decennial percentage change in the housing stock based on our estimates. Δ*Population* is the decennial percentage change in population based on census data. Independent variables are rescaled so that the coefficients can be interpreted as the effect of a 10pp increase in the variable. Housing value is measured in 1,000s of 2020 USD (inflation-adjusted). Housing space is measured in square feet. Heteroskedasticity-robust standard errors in parentheses. *$p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

B Data Appendix

B.1 County assessor data

B.1.1 Data description

All counties in the United States have a county assessor, who is responsible for determining the value of real and personal property (*assessed value*) within their jurisdictions for property taxation. Thus, county assessor registries contain detailed information on all existing properties in a county, typically including a precise geolocation, several property characteristics (e.g., size and year of construction), the name(s) of the owner(s), and a property and/or parcel identifier.

County assessor data is extremely rich and (generally) easily accessible on county government websites, but its decentralized nature makes it challenging to use for cross-county studies. Most county assessors report the same information across the board, but counties do not use a standardized format, and the data may be hard to find. CoreLogic, a private vendor, has collected property records from all over 3,000 existing U.S. counties and has put in substantial effort into harmonizing the data into a format usable by private companies and researchers. The data we obtained from CoreLogic reflects the stock of properties in the United States as of 2020.

Despite CoreLogic's data collection and harmonization efforts, the data still requires substantial cleaning before becoming usable for our research.

B.1.2 Data cleaning: geocoding

We geocode each property in the data to all relevant administrative and census boundaries within the US. We first retrieve the 2020 Census block for the property (the smallest administrative unit in the US). From there, we create a geodatabase that maps each census block to all relevant administrative boundaries (census block groups, census tracts, school districts, electoral districts, etc.).

Geocoding is performed by loading the appropriate state-level census border dataset for each county-assessor file. If available, the property's latitude and longitude are used. If not, the latitude and longitude of the property's census block (as reported by CoreLogic) are used. Properties that lack geolocation data for both the parcel and census block are not used to compute estimate below the county level.[1]

[1]The county is always available since the origin of the data are county assessors.

B.1.3 Data cleaning: number of units

CoreLogic provides a variable that describes the number of units (*CLUnits*) within a property (e.g., the number of dwellings in an apartment building). However, we noticed that their variable does not accurately reflect the actual number of units in many instances (particularly in multi-family housing). We create our own version of that variable (*Units*), applying the following imputation algorithm:

Single-family housing. If the property is identified as "single-family" in the data:

1. Set $Units = 1$.
2. Replace $Units = CLUnits$, if *CLUnits* is non-missing and greater than 1, and $LivingSpace > 10{,}000$ (square feet).

The implicit assumption here is that a "single-family" property identifies a single unit (generally true in the data). In some instances, however, the CoreLogic number of units *CLUnits* is greater than 1. We chose to trust CoreLogic, but only when the living area associated with the single-family dwelling is large enough to plausibly contain more than one unit (i.e., $LivingSpace > 10{,}000$).

Condos. If the property is identified as a "condo" in the data:

1. Set $Units = 1$.
2. Replace $Units = Bathrooms$, if *CLUnits* and *Bathrooms* are non-missing and $CLUnits > (Bathrooms \times 10)$.
3. Replace $Units = CLUnits$, if *CLUnits* is non-missing, *Bathrooms* is missing, and $LivingSpace > 15{,}000$.

The working assumption here is that a "condo" identifies one single property (generally true in the data). Rules #2 and #3 account for the instances in which the "condo" identifies buildings with several condos contained within, and CoreLogic tracks the number of units in *CLUnits*. Rule #2 accounts for instances in which *CLUnits* is implausibly large given the other information that is available and uses the assumption that condos will have, on average, one bathroom per unit to impute a value to *Units*.

Apartments. If the property is identified as an "apartment" in the data:
If *CLUnits* is missing:

1. Set *Units* based on county use codes when number of units can be inferred with "reasonable confidence."

   (a) Example 1: set *Units* $= 2$ when use code is "2 APTS. BLDG.", "APARTMENTS (2)", or "DUPLEX".

   (b) Example 2: set *Units* $= 3$ when use code is "2-4 FAMILY UNITS", "MULTI-FAM 3 UNITS", or "THREE-FAMILY".

   (c) Example 3: set *Units* $= 30$ when use code is "APT LOW-RISE 23-39 UNITS", "APT 21-40 UNITS", or "RES MULTI-UNIT 30 UNITS".

   (d) Example 4: set *Units* $= 750$ when use code is "APT 700-799 UNITS".

2. Set *Units* $= 1$, if the address contains "UNIT", "APT", or "APARTMENT" followed by a number (with or without a space in between) or "#" followed by a number (with or without space) or letter (without space).

3. Set *Units* $=$ *Bathrooms*, if *Bathrooms* ≥ 5.

4. Set *Units* $=$ *Bedrooms*, if *Bedrooms* ≥ 5.

5. Set *Units* $=$ *LivingSpace/1,000*, if *LivingSpace* $\geq$ 5,000.

6. Set *Units* $= 1$, if *LivingSpace* $\leq$ 2,000, or *Bathrooms* ≤ 2, or *Bedrooms* ≤ 2.

7. Set *Units* $= 1$, *LivingSpace* $\in$ (2000, 5000) or *Bathrooms* $\in$ (2, 5) or *Bedrooms* $\in$ (2, 5).

If *CLUnits* is non-missing (and *Units* has not been imputed using the previous rules yet):

8. Set *Units* $=$ *Bathrooms*, if *CLUnits* $>$ *Bathrooms* $\times$ 10 and *LivingSpace*/*Bathrooms* $<$ 10,000

9. Apply county-specific rules (see below).

   - Cook County, IL. Set Units $= 200$, if property is an "APARTMENT 3+ STORIES" or "APARTMENT MULTI STRY".
   - Oklahoma County, OK. Set Units $=$ *CLUnits*/*CLBuildings* (CoreLogic number of units divided by CoreLogic number of buildings in the parcel).
   - New York County, NY, Sacramento County, CA, and Westchester County, NY. County assessor data contains duplicated observations. Identify and drop them based on parcel number and CoreLogic number of units.

10. Set *Units* based on county or state zoning codes when number of units can be inferred with "some confidence."

    (a) Example 1: Set *Units* = 4 when use code is "2+ DWELLINGS", or "SFR DET 2+ DWELLINGS".

    (b) Example 2: Set *Units* = 6 when use code is "3+ DWELLINGS", or SINGLE APARTMENT BLDG 3+ UNITS".

    (c) Example 3: Set *Units* = 7 when use code is "APARTMENTS (UNDER 14 UNITS)".

    (d) Example 4: Set *Units* = 20 when use code is "MULTI-FAM 10+", or "APARTMENT 10+ UNITS".

    (e) Example 5: Set *Units* = 200 when use code is "CONDO APT 99+ UNITS", "HI-RISE 100+ APTS".

11. Set *Units* = NA (missing) if county population (in 2019) is under 100,000, and imputed number of Units is greater than 80, and *Units* was imputed using one of the two previous rules.

12. Set *Units* = 1 if *LivingSpace* $\in$ (2,000, 5,000).

B.1.4 Data cleaning: ownership

Owners. Owners are uniquely identified by the *Owner Name* and *Owner Mailing Address* fields.

Owner-occupied. We create a variable identifying owner-occupied properties using property address (*PropertyAddress*) and owner's mailing address (*MailingAddress*).

1. We set *OwnerOccupied* = 1 if *PropertyAddress*=*MailingAddress*.

2. We also set *OwnerOccupied* = 1 if *PropertyAddress*$\neq$*MailingAddress*, but we believe that the two fields not matching is due to a typo in the string.

    - Example: "123 Mian Street, Springfield" is the same address as "123 Main Street, Springfield". Typo in the word "Main".

Company. We create an indicator variable identifying properties owned by legal entities. We classify a property as company-owned if:

1. The owner owns more than 50 properties.

2. The string *OwnerName* contains:

    - a series of isolated keywords, such as LLC, INC, LTD, or CORP.
    - only alphabetic characters with no spaces.
    - at least three numeric characters.
    - only individual letters separated by spaces,

B.1.5 Data cleaning: additional imputations and winsorization

Dwelling characteristics (e.g., dwelling space, year of construction) are sometimes missing. Whenever possible, we impute these characteristics, as they are key inputs in the dwelling price prediction model. We follow this algorithm:

Dwelling size:

1. Impute using linear regression. We regress dwelling size on all available covariates (e.g., number of rooms, bedrooms, bathrooms, stories, acreage, year built, location), and use the regression estimates to impute size. We repeat this process multiple times to impute a value in the instances in which one of the covariates is missing (by re-running a linear regression excluding the missing covariate).
2. Winsorize imputed values at the 99th percentile. Imputations by linear regression sometimes yield implausibly large values.
3. Impute the median value based on administrative level and property type, in the following order:
    (a) Census block and property type.
    (b) Census tract and property type.
    (c) County and property type.
    (d) Census block only.

Other covariates:

1. Apply the same algorithm described above.
2. For **dwelling space** in apartments, impute using residential space in the building when space in the unit is not observed:

$$\text{Dwelling Space} = \frac{\text{Residential Space in the Building}}{\text{Number of Dwellings in the Building}} \tag{B1}$$

3. For **year of construction**, impute the median value in the census tract based on ACS data.

B.1.6 Data validation: CoreLogic vs Census

Number of dwellings. Table B1 compares the number of dwellings in the WHID (CoreLogic-based) database with the 2020 Census.

Our estimates align closely with official figures. We count 144.4 million housing units in 2020, versus 142.4 million in the Census – an overestimate of about 2.0 million, or roughly 1%.

Accuracy varies more at the state level. In some states (e.g., New York), discrepancies are under 1%. Panel B of Table B1 shows margins of error below 15% in over 31 states and below 25% in 42 states. However, our estimates are substantially off in Wisconsin (-64%), North Dakota (-47%), Louisiana (-44%), Michigan (-43%), Arkansas ($+41\%$), and West Virginia ($+39\%$).

Discrepancies are driven primarily by properties classified as apartments. The "number of units" field is frequently missing for apartments. We impute this field using use-code descriptions and other property characteristics (see Section B.1.3), but when imputation is not feasible the observation is dropped. Consequently, apartment units may be undercounted when data cannot be imputed, and under- or overcounted when imputation relies on limited information. We conducted manual checks for states and counties with large divergences and added county-specific rules where assessor or CoreLogic data issues were clear. Ultimately, however, accuracy cannot be fully improved where county assessor records lack sufficient detail.

Housing space. According to the American Housing Survey (U.S. Census Bureau 2021), in 2020 a dwelling at P10 contained between 750 and 999 sqft; at P50, between 1,500 and 1,999 sqft; at P90, between 3,000 and 3,999 sqft; and at P99, more than 4,000 sqft. Comparing these figures with Table 2 suggests that we slightly underestimate dwelling sizes.

Homeownership. Table B2 compares our estimated state-level homeownership rates with the Census's American Community Survey (ACS). Our estimates are generally close to, but below, the ACS figures. The national average homeownership rate in our data is 57%, versus 65% in the official data (U.S. Census Bureau n.d.). The underestimation is most pronounced in Alaska (48 vs. 67%), Nevada (46 vs. 59%), and Maine (54 vs. 75%). We slightly overestimate rates in a few states. We suspect the overall downward bias stems

Table B1: Data validation: number of housing units in 2020 (WHID vs. Census)

	WHID	Census	Difference	Gap
A. Country				
United States	144.43	142.41	2.02	1.01
B. State				
Alabama	2.42	2.28	0.13	1.05
Alaska	0.36	0.32	0.03	1.10
Arizona	3.53	3.08	0.45	1.14
Arkansas	1.93	1.36	0.56	1.41
California	14.5	14.3	0.16	1.01
Colorado	2.87	2.49	0.38	1.15
Connecticut	1.47	1.53	-0.05	0.96
Delaware	0.45	0.44	0	1
District of Columbia	0.34	0.35	-0.00	0.97
Florida	11.1	9.86	1.24	1.12
Georgia	4.51	4.41	0.10	1.02
Hawaii	0.56	0.56	0.00	1.00
Idaho	0.74	0.75	-0.01	0.98
Illinois	4.92	5.42	-0.50	0.90
Indiana	3.58	2.92	0.66	1.22
Iowa	1.64	1.41	0.22	1.16
Kansas	1.23	1.27	-0.04	0.96
Kentucky	1.58	1.99	-0.41	0.79
Louisiana	1.18	2.07	-0.89	0.56
Maine	0.69	0.73	-0.04	0.93
Maryland	2.41	2.53	-0.11	0.95
Massachusetts	3.07	2.99	0.07	1.02
Michigan	2.64	4.57	-1.9	0.57
Minnesota	2.47	2.48	-0.01	0.99
Mississippi	1.49	1.32	0.17	1.13
Missouri	2.70	2.78	-0.07	0.97
Montana	0.68	0.51	0.16	1.32
Nebraska	0.83	0.84	-0.01	0.98
Nevada	1.66	1.28	0.38	1.3
New Hampshire	0.71	0.63	0.07	1.12
New Jersey	3.01	3.76	-0.74	0.80
New Mexico	0.73	0.94	-0.21	0.77
New York	8.52	8.48	0.04	1
North Carolina	5.37	4.70	0.66	1.14
North Dakota	0.19	0.37	-0.17	0.53
Ohio	6.48	5.24	1.23	1.23
Oklahoma	2.00	1.74	0.25	1.14
Oregon	1.95	1.81	0.13	1.07
Pennsylvania	5.30	5.74	-0.43	0.92
Rhode Island	0.49	0.48	0.00	1.01
South Carolina	2.42	2.34	0.07	1.03
South Dakota	0.29	0.39	-0.1	0.74
Tennessee	3.34	3.03	0.30	1.10
Texas	14.1	11.5	2.53	1.21
Utah	1.35	1.15	0.20	1.17
Vermont	0.28	0.33	-0.05	0.83
Virginia	3.73	3.61	0.11	1.03
Washington	3.85	3.20	0.65	1.20
West Virginia	1.19	0.85	0.34	1.39
Wisconsin	1	2.72	-1.7	0.36
Wyoming	0.29	0.27	0.02	1.07

Notes: *WHID* refers to the number of housing units (millions) estimated in the WHID ($Units_{WHID}$). *Census* refers to the number of housing units (millions) according to the 2020 Census ($Units_{Census}$). *Difference* describes the raw difference between $Units_{WHID}$ and $Units_{Census}$. *Gap* is defined as $Units_{WHID}/Units_{Census}$.

Table B2: Data validation: homeownership rates by state (WHID vs. Census)

State	WHID	Census	State	WHID	Census
Alabama	0.69	0.70	Montana	0.62	0.69
Alaska	0.48	0.67	Nebraska	0.73	0.68
Arizona	0.58	0.68	Nevada	0.46	0.59
Arkansas	0.66	0.67	New Hampshire	0.63	0.72
California	0.55	0.56	New Jersey	0.78	0.64
Colorado	0.63	0.67	New Mexico	0.66	0.70
Connecticut	0.69	0.67	New York	0.40	0.55
Delaware	0.71	0.73	North Carolina	0.60	0.67
Florida	0.57	0.67	North Dakota	0.71	0.63
Georgia	0.67	0.66	Ohio	0.62	0.67
Hawaii	0.55	0.63	Oklahoma	0.65	0.65
Idaho	0.73	0.72	Oregon	0.66	0.64
Illinois	0.66	0.68	Pennsylvania	0.68	0.70
Indiana	0.68	0.71	Rhode Island	0.62	0.63
Iowa	0.68	0.72	South Carolina	0.70	0.72
Kansas	0.65	0.67	South Dakota	0.75	0.69
Kentucky	0.64	0.69	Tennessee	0.60	0.67
Louisiana	0.70	0.67	Texas	0.60	0.63
Maine	0.54	0.75	Utah	0.68	0.70
Maryland	0.74	0.68	Vermont	0.65	0.73
Massachusetts	0.64	0.63	Virginia	0.69	0.68
Michigan	0.70	0.73	Washington	0.53	0.64
Minnesota	0.72	0.73	West Virginia	0.54	0.75
Mississippi	0.70	0.70	Wisconsin	0.70	0.68
Missouri	0.66	0.69	Wyoming	0.64	0.71

Notes: The table reports owner-occupancy rates across states, comparing our WHID estimates based on CoreLogic data with the Census (ACS, 2020).

from homeowners holding properties in trusts: because trusts are legal entities rather than natural persons, we classify them with companies and real estate investors, which leads us to treat households residing in trust-owned properties as renters.

B.1.7 Discussion: accuracy of long-run estimates

Cross-sectional property data can be used to analyze housing over time because (1) the year of construction is observed and (2) housing, unlike people, is immobile. The main caveat – emphasized throughout the paper – is that dwellings exiting the stock (e.g., demolitions) are not observed. This section discusses the reliability and potential bias of our long-run estimates.

Figure B1 shows the evolution of the U.S. housing stock based on our cross-sectional CoreLogic data (characterizing the 2020 stock; black solid line) and Census data (1900–2020). We make three observations.

First, the two lines closely track each other. The two series exhibit similar slopes, suggesting that our estimates capture the real trends remarkably well throughout the whole period.

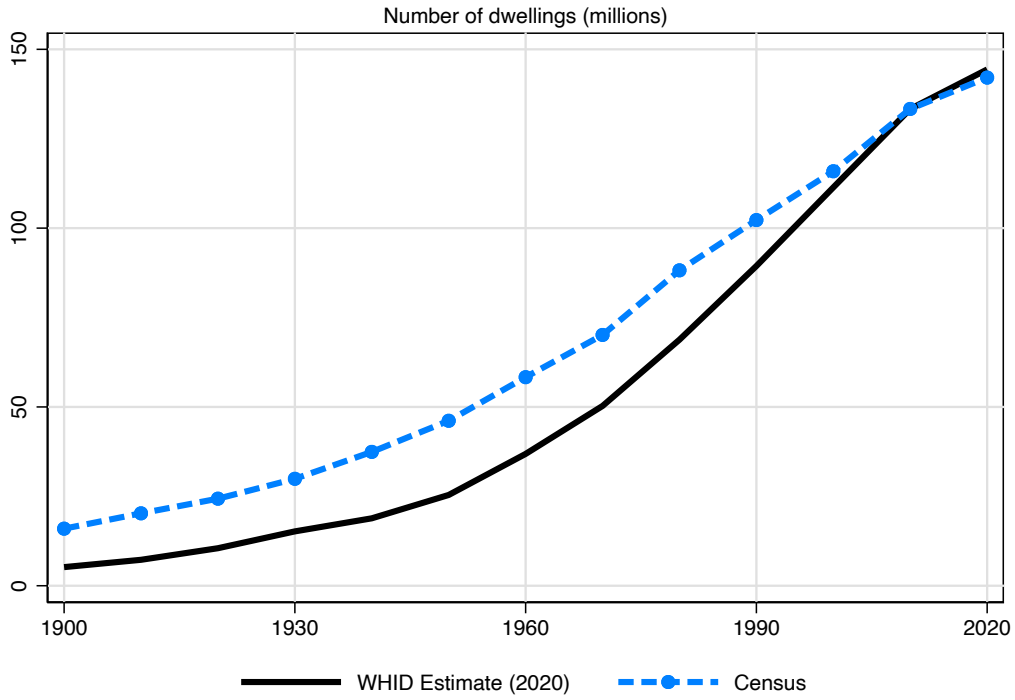


Figure B1: Evolution of the housing stock in the United States (1900-2020)

Notes: The black solid line describes the evolution of the housing stock based on our cross-section of CoreLogic data as of 2023. We use the year of construction to infer the number of dwellings present at a given point in time. This series primarily misses demolished dwellings, which are not observed in the data. The dashed blue line shows the evolution of the dwelling stock based on Census data (1930-2020). Census estimates for 1900, 1910, and 1920 were retrieved from a summary table in the 1930 Census.

Second, the gap between the two series is narrow in recent years and wider early in the period, as expected. A dwelling constructed in 1900 is more likely to have been demolished by 2020 than one constructed in 2010, so accuracy declines the farther back in time we look.

Third, we capture a substantial fraction of the historical stock, extending back to 1900. Our data indicate 5.2 million dwellings built in 1900 or earlier, while Census estimates suggest 15.9 million – roughly one-third coverage at that date. Coverage rises to about two-thirds by 1960 (36.8 vs. 58.3 million), about 90% by 1990 (89.4 vs. 102.3 million), and nearly 100% post-2000.

Based on the previous observations, we make the following remarks:

First, the primary concern for the early estimates is selection. Even if our data contain only a subset of all dwellings that existed at a point in time (e.g., 33% in 1900), we can still make inferences about population objects (e.g., housing space Gini in 1900). However,

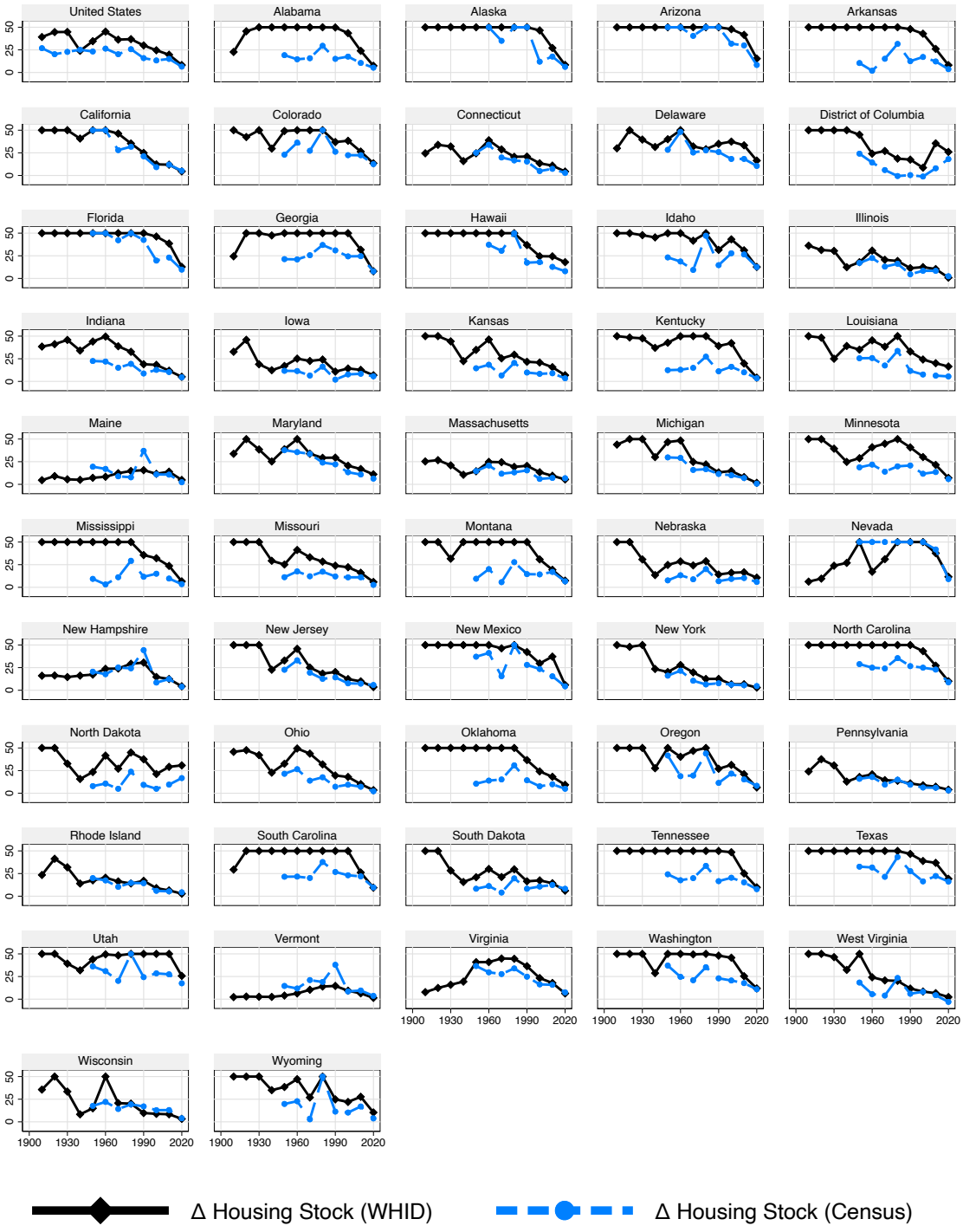


Figure B2: Growth in the housing stock over time across states (WHID vs Census)

Notes: Δ*Housing Stock (WHID)* and Δ*Housing Stock (Census)* are the decennial percentage change in the housing stock based on the WHID and the Decennial Census, respectively. The Δ*Housing Stock (Census)* series starts in 1940 for the panels describing the evolution of the stock in the state because we were unable to collect information on the number of units at the state level before 1930. Growth rates are censored at 50% for illustration purposes.

selection is a major concern in this exercise as the subsample of dwellings built before 1900 that survived until 2020 is certainly not random. At a minimum, these units are likely positively selected on quality – they survived, so they probably had better structures than those that did not.

Second, our long-run estimates are likely to underestimate housing inequality. We cannot determine the precise direction of bias without reliable data, but if the units that did not survive were disproportionately in the lower tail of the distribution (in space or value), then our long-run estimates should understate inequality.

Third, we are substantially more confident in the post-1940 estimates for two reasons. On the one hand, we capture more than 50% of the stock after 1940, mitigating selection concerns. On the other hand, the close alignment of trends in Figure B1 suggests that we capture most of the dwellings entering the stock in the second half of the 20th Century.

Fourth, we are considerably less confident about the pre-1940 period for three reasons. First, based on available estimates, we capture less than 50% of the stock at the time, exacerbating selection concerns. Second, sample representativeness and selection likely vary across locations, complicating interpretation below the national level. Figure B2 illustrates this point. The figures compare decennial percentage changes in the housing stock based on WHID estimates (solid black line) and the Decennial Census (dashed blue line). The top panel indicates that, despite overestimating growth in earlier decades (due to survivorship bias), we capture nationwide trends reasonably well. Moreover, growth trends match almost one-for-one in several states – including California, Illinois, Maryland, Massachusetts, New Hampshire, New Jersey, Michigan, Pennsylvania, and Rhode Island – and are reasonably approximated in most remaining states. However, our data tend to substantially overestimate growth in places like Alabama, Arkansas, Georgia, Kentucky, Montana, and Tennessee, particularly in the earlier years.

Finally, despite these concerns, we view this as a valuable exercise. With appropriate caveats, our results yield important insights, and future research may be able to correct some of the biases in our data.

B.2 Transaction data

B.2.1 Data description

We use real estate transactions covering nearly the entire US, beginning in the 1980s. The data originate from county deed records obtained from CoreLogic. We exclude non-arm's-length transactions and distressed sales, as these occur under special circumstances and their sale prices do not reflect market value.

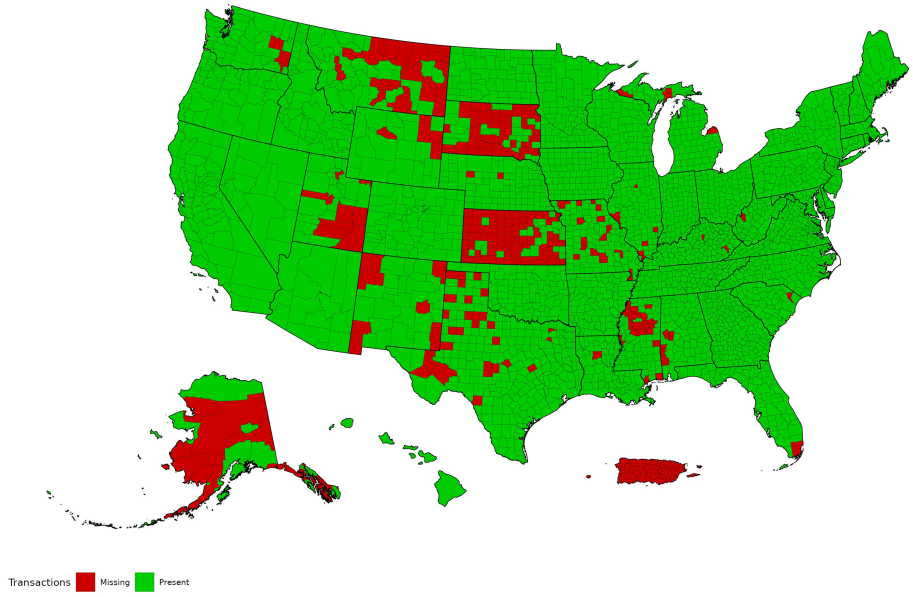


Figure B3: Real estate transactions data coverage

Notes: Each polygon represents a county. Counties colored in green are present in the transactions data. Counties in red are not observed.

Figure B3 that the geographic coverage of the data is good, though we lack data for some primarily rural counties in South Dakota, Montana, Kansas, Alaska, and for Puerto Rico. These transactions allow us to impute market prices for all dwellings – whether on or off the market – within the covered period and geography.

B.2.2 Data cleaning

Number of units. We apply the same algorithm outlined in Section B.1. If *Units* is still missing after the imputation algorithm, we assign a value equal to the number of bedrooms divided by three and rounded to the nearest integer. If the value is still missing or rounded to 0, we assign a default value of 1.

Dwelling space. We divide total space in the building by the number of units.

$$\text{Dwelling Space} = \frac{\text{Residential Space in the Building}}{\text{Number of Dwellings in the Building}} \tag{B2}$$

If *Dwelling Space* is below 120 square feet and the number of units in the building is greater than 1, we replace *Dwelling Space* with *Residential Space in the Building*. We apply this latter rule because we realized that some county assessors report the actual unit

space in an apartment building or condo in the data field intended to describe the total building space.

Age. We define building age as follows:

$$\text{Age} = \text{Transaction Year} - \text{Construction Year} \tag{B3}$$

Additional imputations. Whenever a dwelling characteristic used in the prediction algorithm (see next section) is still missing, we impute it using the median value of the census tract. If there are still missing values, we impute using the median value in the county.

Sample trimming. We keep only observations with a sale price under 150 million USD.

Our baseline sample contains 149.6 million transactions, covering all 50 states from 1980 to 2020.

B.3 Housing value imputations

Table B3: Variables used in the Random Forest model

Variable	Description
tract_id	2020 census tract identifier
county_id	2020 county identifier
lat	Latitude coordinate
lon	Longitude coordinate
property_type	Type of property (single family, apartment, etc.)
bedroom	Total number of bedrooms in the property
bathroom	Total number of bathrooms in the property
log_sqft	Log of residential area in the property (ft^2)
log_lot_acre	Log of lot area (acres)
bldg_age	Age of the property at transaction date (years)
year_construction	Year of construction
year	Year of the transaction

We use CoreLogic transaction data – comprising almost 150 million sales across the United States from 1980 to 2020 – to train a predictive model of dwelling values. Specifically, we implement Breiman's random forest algorithm (Breiman 2001) using the Ranger package in R. Prior research shows that random forests perform especially well in predicting housing prices (Mullainathan and Spiess 2017).

Our training set includes 12 primary dwelling-related predictors, covering key housing characteristics (e.g., building age, size). To capture location effects precisely, these include

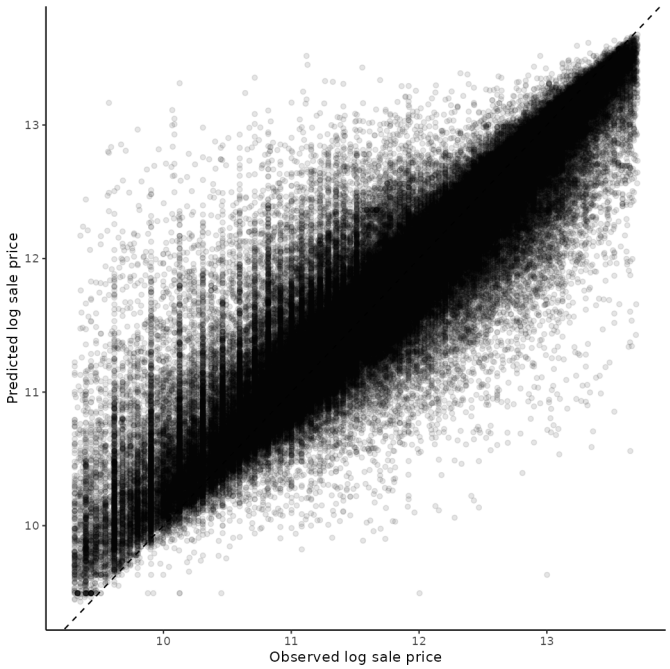


Figure B4: Prediction validation: model in-sample fit in North Carolina

Notes: Each dot represents a real estate transaction. The x-axis reports the actual transaction price observed in the data. The y-axis reports the model predicted price.

factor variables identifying the administrative levels of the parcel and latitude/longitude coordinates. Table B3 describes the main predictors used in the model.

To improve the predictive performance of the random forest model, we tuned three core hyperparameters. First, we adjusted the number of predictors considered at each split, which controls model complexity and variable selection. Second, we tuned the minimum number of observations per terminal node to regulate tree depth and reduce overfitting. Third, we optimized the subsample proportion used to grow each tree, affecting variance and computational efficiency. We conducted a grid search over plausible ranges, evaluating configurations using the out-of-bag root mean squared error (OOB-RMSE). Specifically, we tested four levels for predictors-per-split, three levels for minimum terminal node size, and two levels for subsample proportion, yielding 24 model combinations. The selected configuration samples 4 predictors at each split, uses a minimum terminal node size of 9, and trains each tree on 75% of the available data.

Figure B4 presents the in-sample fit of our predictions for North Carolina. The high correlation between predicted and actual prices provides reassurance about model performance. However, the figure also suggests slight overestimation at the lower end and underestimation at the upper end of the distribution, implying that we may be understating housing value inequality.

Table B4: Random forest prediction performance by state

State	OOB-RMSE	R^2	State	OOB-RMSE	R^2
Alabama	0.23	0.94	Missouri	0.16	0.95
Alaska	0.14	0.94	Montana	0.17	0.95
Arizona	0.14	0.97	Nebraska	0.21	0.93
Arkansas	0.22	0.94	Nevada	0.14	0.97
California (North)	0.13	0.96	New Hampshire	0.18	0.93
California (South)	0.13	0.96	New Jersey	0.14	0.97
Colorado	0.15	0.97	New Mexico	0.14	0.93
Connecticut	0.14	0.96	New York	0.23	0.96
Delaware	0.18	0.95	North Carolina	0.22	0.94
District of Columbia	0.14	0.98	North Dakota	0.18	0.91
Florida	0.19	0.96	Ohio	0.19	0.95
Georgia	0.21	0.94	Oklahoma	0.25	0.94
Hawaii	0.26	0.95	Oregon	0.18	0.94
Idaho	0.13	0.95	Pennsylvania	0.18	0.96
Illinois	0.17	0.95	Rhode Island	0.13	0.95
Indiana	0.21	0.93	South Carolina	0.24	0.94
Iowa	0.20	0.94	South Dakota	0.23	0.93
Kansas	0.14	0.96	Tennessee	0.22	0.95
Kentucky	0.28	0.93	Texas	0.14	0.96
Louisiana	0.23	0.94	Utah	0.12	0.95
Maine	0.18	0.94	Vermont	0.30	0.90
Maryland	0.15	0.96	Virginia	0.16	0.96
Massachusetts	0.16	0.96	Washington	0.15	0.97
Michigan	0.19	0.94	West Virginia	0.29	0.94
Minnesota	0.18	0.94	Wisconsin	0.15	0.95
Mississippi	0.24	0.94	Wyoming	0.14	0.94

Notes: The table reports the prediction performance of state-specific random forest models in terms of out-of-bag root mean squared error (OOB-RMSE) and R^2. All models use the hyperparameter configuration that achieved the best performance in North Carolina.

We use the hyperparameters that minimized OOB-RMSE in North Carolina to train state-specific models. A hyperparameter grid search – requiring multiple random forest fits – is extremely time-consuming (especially in large states) and yields only marginal gains in accuracy (OOB-RMSE range from 0.1789 to 0.1836). Accordingly, we adopt the North Carolina–optimal settings for all states.[2] Because of its size, we train separate models for Northern and Southern California. Table B4 reports OOB-RMSE for each state. Performance is best in California, Idaho, and Rhode Island (0.13) and worst in Vermont (0.30). The average prediction error across states is 0.18.

In terms of variable importance, location-related predictors dominate: latitude, longitude, and census block are in the top five. Transaction year and residential space (square feet) complete the top five. Other important predictors include the number of bathrooms, the parcel's acreage, and the building age.

We then use the trained model to predict the value of all dwellings in the United States

[2]County assessors in Maryland and Arizona do not report the number of bedrooms, so this variable is excluded from the prediction models in these states.

(over 100 million), regardless of whether they transacted between 1980 and 2020. For each dwelling, we generate annual predictions for every year from 1980 through 2020, matching the span of the transaction data.

B.4 Other data

We use several additional datasets throughout the paper. These include:

- **American Community Survey**. Population counts, income (median and Gini), and poverty rates at multiple geographic levels (country, state, county, census tract).
- **American Housing Survey**. Characteristics of the housing stock in 2020.
- **Decennial Census**. Population and housing counts at the national level since 1900. Dwelling counts at the state level since 1930.
- **FRED**. Inflation and housing price data.

B.5 Housing inequality estimates

Geographies. We consider the following geographic aggregations: country, region, division, state, 2000 commuting zone, ZCTA, CSA, CBSA, 2020 county, 2020 school district, 2020 congressional district, 2020 electoral precinct, 2020 census tract, 2020 census block group, and 2020 census block.

Metrics. For each geography, we estimate the following metrics for the distributions of housing space, housing value, and housing price per square foot: Gini index, Share top 0.1%, Share top 1%, Share top 10%, Share top 20%, Share bottom 50%, Share bottom 40%, Share bottom 20%, Share bottom 10%, Theil index, Theil between and within components (state-county and commuting zone-tract), coefficient of variation, standard deviation, mean, percentiles P1, P5, P10, P25, P50, P75, P90, P95, P99, P99.9 (99.9th percentile), 90-10 ratio, 90-50 ratio, and number of dwellings.

The "share" metrics are defined as follows:

$$\text{Share Top } k = \frac{\text{Total space/value in the } k\% \text{ most valuable/spacious dwellings}}{\text{Total space/value contained in the housing stock}} \tag{B4}$$

$$\text{Share Bottom } k = \frac{\text{Total space/value in the } k\% \text{ least valuable/spacious dwellings}}{\text{Total space/value contained in the housing stock}} \tag{B5}$$

The "ratios" are defined as follows:

$$k\text{-}y \text{ Ratio} = \frac{\text{Value/Space in } k\text{-th percentile}}{\text{Value/Space in } y\text{-th percentile}} \tag{B6}$$

Housing assets. In addition to distributions of housing value and space across dwellings, we also estimate the distribution of housing assets across households within each geography. We make the following assumptions:

- Non-owner-occupied properties are occupied by renters, who hold no housing assets (all renters have $0 in housing assets).
- Owner-occupied properties are occupied by a household with positive housing assets. If a household owns multiple properties (identified by owner name and mailing address), we assign the household's housing assets to a geography based on the owner's mailing address.

Example: John Snow owns five dwellings across the United States and resides in Arizona. We assign the total value and square footage of all five dwellings to him and attribute that value to the parcel he occupies. Thus, if each dwelling is valued at $100,000, we assign $500,000 to the Arizona property and zero to the other four. The rationale is that those other units are likely rented out and therefore occupied by individuals who do not hold housing assets.[3]

When computing the distribution of housing assets, we consider only natural persons. Assets owned by legal entities (including trusts) are left undistributed. Because these assets are ultimately owned by natural persons in a nonrandom manner, our results likely underestimate housing asset inequality in the population.

We use the term "housing assets" instead of "housing wealth" because we do not have information on housing liabilities (i.e., mortgages).

Time. We compute all of these statistics at all geographies (except census block, block group, and precinct level) for every year between 2000 and 2020, and for every decade between 1900 and 1990. For value-related estimates, we only compute the statistics post-1980. Estimates at the census block, block group, and precinct levels are only computed for 2020.

Stocks and flows. We compute all of these statistics separately for the housing stock at a given point in time and for the housing flows over a 10-year period. For example,

[3]If an individual owns multiple units within the same parcel, we assign the total value of these units to one unit only, assuming renters occupy the remaining units.

we compute all the statistics outlined above by restricting attention to all dwellings that entered the stock between 2011 and 2020.

B.6 WHID - United States

All housing inequality estimates are publicly available at https://doi.org/10.7910/DVN/5C9AJD as part of the World Housing Inequality Database (WHID), accessible at www.worldhousingdata.com.

Tables B5 and B6 provide a list of names and a description for all the variables present in the dataset.

Table B5: WHID - United States: variables description

Variable name	Variable description
A. Year, Variable type, and geography identifiers	
year	Calendar year of the estimate
agglevel	Aggregation level of the estimate (string)
agg	Aggregation level of the estimate (numeric)
ineqvar	Inequality dimension
var_type	Variable type
ctry_id	Country ID (ISO)
ctry_name	Country name
reg_id	Census Region ID
reg_name	Census Region name
div_id	Census Division ID
div_name	Census Division name
state_id	State FIPS code
state_name	State name
cz_id	2000 Commuting Zone ID
county_id	2020 County FIPS code
county_name	2020 County name
tract_id	2020 Census tract ID
cbg_id	2020 Census Block Group ID
block_id	2020 Census Block ID
schooldis_id	2020 School District ID
schooldis_name	2020 School District name
congdis_id	2020 Congress District ID
congdis_name	2020 Congress District name
precinct_id	2020 Electoral Precinct ID
precinct_name	2020 Electoral Precinct name
zcta_id	2020 Census Zip Code Tabulated Area ID
cbsa_id	2020 Census Core Based Statistical Area ID
cbsa_name	2020 Census Core Based Statistical Area name
csa_id	2020 Census Combined Statistical Area ID
csa_name	2020 Census Combined Statistical Area name
msa_id	2020 Metropolitan Statistical Area ID
msa_name	2020 Metropolitan Statistical Area name

Table B5: WHID - United States: variables description (continued)

Variable name	Variable description
B. Metrics	
gini	Gini index
share_top_01	Share of total dwelling value/space contained in the top 0.1% of most valuable/spacious dwellings (available at county level or above when N_dwell>1000)
share_top_1	Share of total dwelling value/space contained in the top 1% of most valuable/spacious dwellings
share_top_10	Share of total dwelling value/space contained in the top 10% of most valuable/spacious dwellings
share_top_20	Share of total dwelling value/space contained in the top 20% of most valuable/spacious dwellings
share_bottom_50	Share of total dwelling value/space contained in the bottom 50% of least valuable/spacious dwellings
share_bottom_40	Share of total dwelling value/space contained in the bottom 40% of least valuable/spacious dwellings
share_bottom_20	Share of total dwelling value/space contained in the bottom 20% of least valuable/spacious dwellings
share_bottom_10	Share of total dwelling value/space contained in the bottom 10% of least valuable/spacious dwellings
theil	Theil index
theil_state_b_county	Between-Theil index (variation between counties) (available at state level only)
theil_state_w_county	Within-Theil index (variation within counties) (available at state level only)
theil_cz_b_tract	Between-Theil index (variation between census tracts) (available at CZ level only)
theil_cz_w_tract	Within-Theil index (variation within census tracts) (available at CZ level only)
coef_var	Coefficient of variation
sd	Standard deviation
mean	Mean
p1	Value/Space of the dwelling in the 1st percentile
p5	Value/Space of the dwelling in the 5th percentile
p10	Value/Space of the dwelling in the 10th percentile
p25	Value/Space of the dwelling in the 25th percentile
p50	Value/Space of the dwelling in the 50th percentile (median)
p75	Value/Space of the dwelling in the 75th percentile
p90	Value/Space of the dwelling in the 90th percentile
p95	Value/Space of the dwelling in the 95th percentile
p99	Value/Space of the dwelling in the 99th percentile
p999	Value/Space of the dwelling in the 99.9th percentile (available at county level or above when N_dwell>1000)
ratio9010	p90/p10
ratio9050	p90/p50
N_dwell	Number of dwellings in the geography/time period
censored	=1 if estimates are censored (less than 25 dwellings in the geography-year)

Table B6: WHID - United States: geographic and time coverage

Variable name	Type	Values
year	Numeric	1900-2020
agglevel	String	ctry; region; division; state; cz; county; tract; cbg; block; schooldis; congrdis; precinct; zcta; cbsa; csa; msa
agg	Numeric	1 (country); 2 (region); 3 (division); 4 (state); 5 (2000 commuting zone); 6 (county); 7 (census tract); 8 (CBG); 9 (Census block); 10 (school district); 11 (congress district); 12 (precinct); 13 (ZCTA) 14 (CBSA); 15 (CSA); 16 (MSA)
ineqvar	String	space; value; value_sqft; asset_space; asset_value
var_type	String	stock; flow
ctry_id	String	840
ctry_name	String	United States
reg_id	String	1, ..., 4
reg_name	String	Northeast, ..., West
div_id	String	1, ..., 9
div_name	String	New England, ..., Pacific
state_id	String	01, ..., 56
state_name	String	Alabama, ..., Wyoming
cz_id	String	001, ..., 903
county_id	String	01001, ..., 56045
county_name	String	Autauga County, ..., Weston County
tract_id	String	01001020100, ..., 56045951300
cbg_id	String	010010201001, ..., 560459513003
block_id	String	010010201001000, ..., 560459513003062
schooldis_id	String	0100001, ..., 5606240
schooldis_name	String	Fort Rucker School District, ..., Washakie County School District 1
congdis_id	String	0101, ..., 5600
congdis_name	String	Congressional District 1, ..., Congressional District (at Large)
precinct_id	String	01001000001, ..., 56045000503
precinct_name	String	Jones Community Ctr, ..., Newcastle South Voting District
zcta_id	String	02134, . . . , 30318
cbsa_id	String	10100, ..., 49820
cbsa_name	String	Aberdeen, SD Micro Area, ..., Zapata, TX Micro Area
csa_id	String	104, ..., 566
csa_name	String	Albany-Schenectady, NY CSA, ..., Youngstown-Warren, OH-PA CSA
msa_id	String	11244, ..., 48864
msa_name	String	Anaheim-Santa Ana-Irvine CA, ..., Wilmington DE-MD-NJ

C Supplementary results

This Appendix contains additional results not discussed in the main text, but that might be of interest to some readers.

Figure C1 shows the evolution of housing value (price per square foot, psqft) inequality based on the Gini index across states since 1980.

Figure C2 shows the evolution of housing value inequality (Gini) in new inflows across states since 1980. Figure C3 shows the evolution of housing value inequality (P10, P50, P90) in new inflows across states since 1980. Housing values are measured in millions of 2020 USD.

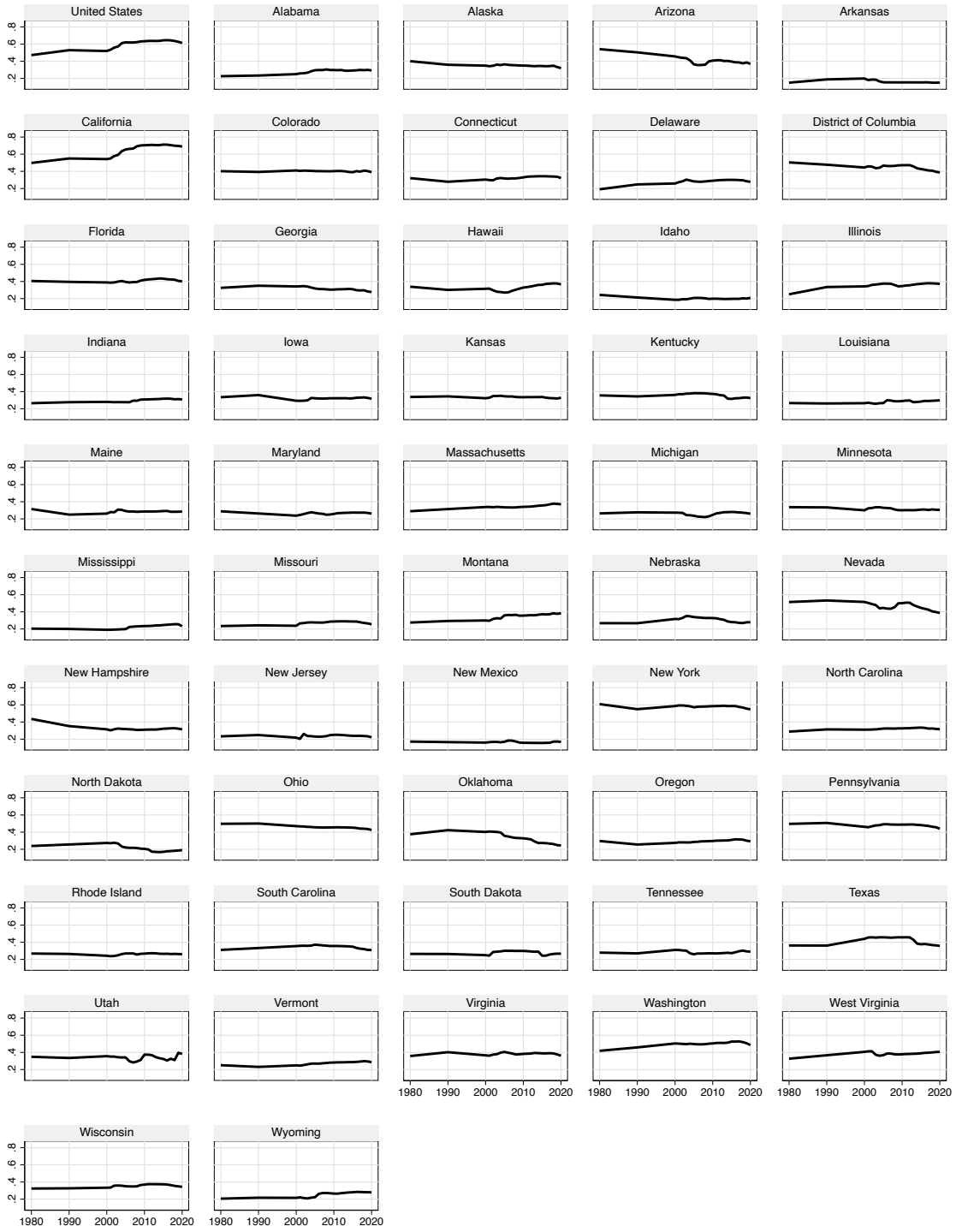


Figure C1: Trends in housing value inequality (psqft) (Gini) across states

Notes: Inequality is measured with the Gini index. The top-left panel describes the evolution of housing value inequality (Gini) for the United States. The remaining panels report the evolution of housing value inequality at the state level. Housing price per square foot is measured in current USD (not adjusted for inflation).

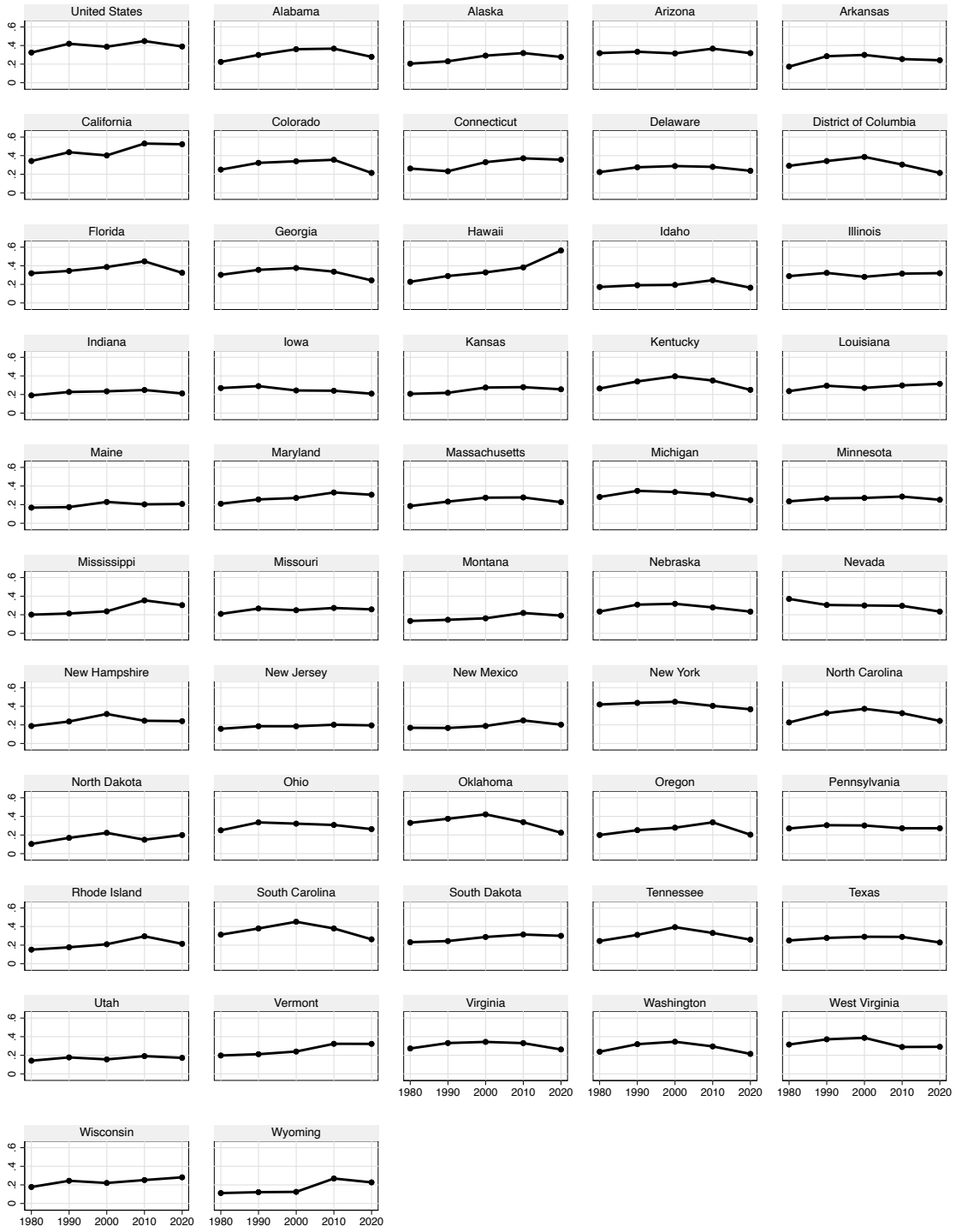


Figure C2: Housing value inequality (Gini) by decade-cohort across states

Notes: The top-left panel describes the evolution of housing value inequality (Gini) by decade-cohort for the United States. The remaining panels describe trends in the same outcomes across the states. Estimates are constructed based on dwellings entering the housing stock in the previous decade (e.g., the 1910 estimate is based on dwellings built between 1901 and 1910). Housing value is measured in 2020 USD.



Figure C3: Housing value inequality (P10, P50, P90) by decade-cohort across states

Notes: The top-left panel describes the evolution of housing value inequality (P10, P50, and P90) by decade-cohort for the United States. The remaining panels describe trends in the same outcomes across the states. Estimates are constructed based on dwellings entering the housing stock in the previous decade (e.g., the 1910 estimate is based on dwellings built between 1901 and 1910). Housing value is measured in millions of 2020 USD (inflation-adjusted).