What Drives Rent?

A deep dive into people, place and price.

Shengmin Xiao Alex TenPas Denvir Higgins

April 17, 2025

1. Introduction

As young professionals navigating an increasingly unaffordable housing market, we are directly affected by the structural inequalities shaping where and how people live. The soaring cost of rent in urban and suburban areas is not just a reflection of supply and demand – it reveals deeper patterns tied to income, education, and social status.

This study investigates how demographic and socioeconomic disparities influence rental prices across the United States. Specifically, we ask: What patterns and associations can be observed between rent prices and the demographics of the people across different regions? This question emerges from the recognition that where someone lives – and what they pay for it – is often shaped by more than location alone. It's shaped by who they are.

While previous research has explored predictors of rent and housing affordability, our work contributes by applying multiple regression models and integrating geographic data to highlight the statistical linkages between rent and variables like race, income, education, and marital status. By identifying these linkages, we aim to provide a data-driven baseline for policymakers, urban planners, advocacy groups, and particularly for renters like ourselves to make informed decisions, advocate for fairer housing policies, and navigate the rental market with greater awareness of systemic inequalities.

2. Description of the Data Source

To investigate how socioeconomic factors relate to rental prices, we integrated data from three publicly available sources.

1. American Community Survey (ACS)[1]: The U.S. Census Bureau conducts the ACS annually to provide detailed demographic, social, economic, and housing data. Approximately 3.5 million addresses are randomly sampled each year, with participation required by law, ensuring a high response rate and robust data quality. The unit of observation in this dataset is the ZIP Code. We used the 2023 ACS S0601 dataset, which includes variables such as median income, educational attainment, racial composition, and housing characteristics, which are core indicators for examining disparities across regions.

2. Fair Market Rent (FMR)[2]: The U.S. Department of Housing and Urban Development (HUD), through its Office of Policy Development and Research, publishes annual FMR for use in federal housing programs such as Section 8. These rent estimates are calculated using a combination of the Census Bureau's five-year ACS data (as a base), HUD-defined geographic areas, and adjustments based on the Bureau of Labor Statistics' Consumer Price Index (CPI). The unit of observation here is also at the ZIP-level, aligning with the ACS data for comparison. FMR serves as a standardized estimate of local rental costs, allowing us to assess rent relative to regional socioeconomic conditions.
3. HUD-USPS ZIP-TRACT Crosswalk File[3]: To ensure geographic precision and data relevance, we used the 2023 Q2 ZIP-TRACT Crosswalk file developed by HUD in collaboration with the U.S. Postal Service. This dataset links USPS ZIP Codes with Census geographic areas and includes quarterly USPS residential vacancy indicators. We used it to filter out ZIP Codes with low residential ratios or primarily commercial/industrial land use, focusing our analysis on areas where rental data reflects meaningful residential patterns.

By combining demographic, economic, and housing price data at the ZIP Code level, we are able to explore nuanced associations with strong statistical and geographic resolution.

3. Data Wrangling

Our data wrangling process was designed to be modular, efficient, and transparent, using Python with the Pandas and NumPy libraries.

The 2023 ACS dataset contained 33,773 rows and 427 columns, with ZIP Code information embedded in a Geography column. Using the official Census data dictionary, column codes were mapped to human-readable names. Irrelevant columns were dropped, and variables representing demographic characteristics such as race, income, education, gender, age, marital status, and language were retained.

The 2023 HUD Fair Market Rent dataset included 27,331 rows and 18 columns, providing ZIP Code-level rent data across five housing types (defined by bedroom count), as well as 90th and 110th percentile rent estimates. This was joined with the ACS data on ZIP Code, resulting in a merged dataframe of 20,855 rows and 445 columns. At this stage, a focused subset of 28 columns was retained—demographics from the ACS and rent estimates from HUD.

The Q2 2023 ZIP-TRACT Crosswalk dataset originally contained 188,563 rows across 8 columns, representing ZIP Code-to-tract relationships with quarterly USPS vacancy data. We filtered it to include only tracts with a residential ratio of 80% or higher—defined as predominantly residential areas—producing a filtered subset of 35,434 rows. This dataset was aggregated to the ZIP Code level, retaining only ZIP Codes with sufficient residential population, and merged with the existing dataset. The final unified dataframe contained 18,988 rows and 36 columns.

To ensure balanced sampling across geographic regions, we stratified the dataset into five regions (Midwest, Northeast, Southeast, Southwest, and Northwest) and randomly selected 1,000 ZIP Codes per region, resulting in a final dataset of 5,000 rows and 37 columns. The additional column, region, was derived based on state-to-region mappings and appended for regional analysis. This consolidated dataframe served as the single source of truth for all subsequent modeling and visualization steps. See Appendix C for a complete list of columns in the final dataset.

4. Operationalization

To explore how demographic characteristics influence rent prices across U.S. regions, we operationalized our outcome variable as the Small Area Fair Market Rent for one-bedroom units (SAFMR.1BR). Our predictor variables were selected to represent core dimensions of socioeconomic status, educational opportunity, geographic mobility, and social structure. Specifically, we included:

- Median household income (economic access),
- Percent of adults without a high school degree (educational disadvantage),
- Racial composition (structural racial disparities),
- Percent born in-state (community rootedness vs. mobility),
- Marital status (percent married) (often correlated with household stability and dual-income likelihood),
- Median age.

The final data frame focused on isolating residential areas and balanced regional representation to ensure compatibility across diverse geographic contexts. While various scaling techniques were explored to potentially improve model performance, they did not yield substantial improvements and, in some instances, complicated interpretation. Therefore, the unscaled variables were retained for the primary analysis. During statistical analysis, VIF values were used to assess multicollinearity, of which several variables were part of a closed system that added up to 100% and skyrocketed the VIF values to 40,000. Upon removing portions of these closed systems based on very and extremely significant p-values, VIF values ranged from 1.3 to 2.5. This methodological design supports our central inquiry into how demographic disparities shape rent prices, reflecting the lived realities of renters navigation affordability challenges across the U.S..

5. Model Specification

To evaluate the relationship between demographic characteristics and rent prices, we create a series of linear regression models using the one-bedroom SAFMR rent as the outcome variable. Each model iteration progressively refined the set of predictors to maximize explanatory power. All model specifications and results, including coefficients, standard errors, and model statistics, are displayed using the stargazer package (see Appendix A).

Model 4 emerged as the final, optimized specification. It excludes two variables, percent race, two or more races, and percent below poverty line, due to their high multicollinearity and weak predictive contributions.

The regression coefficients in Model 4 reveal several key relationships:

- Median age and percent married also show positive coefficients, suggesting that older and more family-oriented populations are linked to higher-rent areas.
- Educational attainment (percent less than high school) and percent born in-state have negative coefficients, indicating lower rents in areas with lower educational achievement and more local-born residents.
- Language diversity is positively associated with rent, potentially reflecting urban and immigrant-dense areas with higher demand.

Interestingly, some variables such as percent race: Black and median income in the past year were retained for theoretical relevance despite low statistical significance, reflecting our intent to balance statistical and social interpretation.

To better understand the interplay between predictors, we also generated a correlation heatmap (see Appendix B):

- Median income is moderately correlated with percent Asian and language diversity, suggesting cultural and economic clustering.
- Median age and percent married are positively correlated, consistent with demographic life-stage patterns.
- In contrast, percent less than high school is negatively correlated with both income and marriage rates, highlighting the compounding effects of educational and economic disadvantage.

Together, these models and visual diagnostics support our central hypothesis: that demographic disparities—particularly around income, education, and race—are strongly associated with differences in local rental prices.

6. Model Assumptions

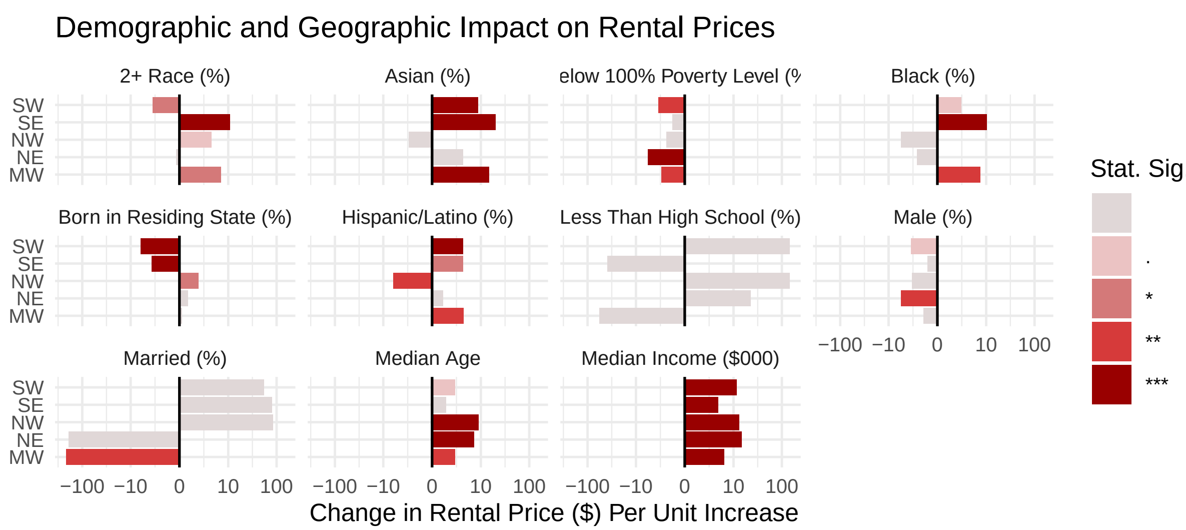
#	Assumption	Description
1	IID	Data points (Zip Codes) are assumed to be independent and drawn from similar distributions across U.S. regions.
2	Constant Error Variance	Breusch-Pagan test rejects homoskedasticity (BP = 189.84, $p < 2.2e\text{-}16$).
3	Normal Distribution	Errors have a heavy-tailed (leptokurtic) distribution, kurtosis = 3.3619.
4	Linear Conditional Expectation	Heavy tailed distribution rejects linear conditional expectation.
5	No Perfect Collinearity	low Variance Inflation Factor of each predictor fails to reject no perfect collinearity.

7. Model Results and Interpretation

The results of our multiple regression analysis reveal significant relationships between demographic factors and one-bedroom rent prices across U.S. regions. Model 4 explains approximately 63% of variation in one-bedroom rent prices (adjusted R^2 of 0.630) with high statistical significance ($F = 319.893$, $p < 0.01$). Demographic factors show varied relationships with rent prices. Median age ($\beta = 3.245$, $p < 0.05$) and percentage of Asian residents ($\beta = 16.393$, $p < 0.01$) both positively correlate with higher rents. Conversely, percentage married ($\beta = -6.135$, $p < 0.01$), percentage with less than high school education ($\beta = -4.228$, $p < 0.01$), and percentage born in-state ($\beta = -4.437$, $p < 0.01$) all demonstrate significant negative relationships with rent prices. Language diversity ($\beta = 7.149$, $p < 0.01$), measured by the percentage speaking another language at home, shows a strong positive association with rent, suggesting that multicultural areas command higher housing costs. Median income shows a modest effect increasing rent prices \$0.017 for every dollar earned.

To uncover regional differences, the data was further segmented into five regions: Northeast, Southeast, Southwest, Northwest, and Midwest. Across all regions, median income and age were consistently among the most statistically significant predictors of rent prices. In all five regions, median income had a strong positive association with rent (e.g., $\beta = 14.93$ in NE, $\beta = 6.33$ in MW, all $p < 0.001$), reflecting that higher income levels are linked to higher rents. Similarly, median age showed a significant positive association in most regions, especially NE ($\beta = 7.16$, $p < 0.001$) and NW ($\beta = 8.94$, $p < 0.001$), indicating that areas with older populations tend to have higher rental prices.

In summary, rental prices are significantly associated with income, age, the proportion of Asian residents, marital status, educational attainment (particularly less than a high school diploma), multilingual households, and the percentage of individuals residing in their birth state. These findings reveal that rental markets reflect complex socioeconomic patterns beyond income alone, with implications for housing policies aimed at addressing affordability challenges across demographic groups.



References

1. U.S. Census Bureau, "S0601 Selected Characteristics of the Total and Native Populations in the United States", <data.census.gov/table?q=s0601>, accessed on March 25, 2025.
2. HUD User's Office of Policy Development and Research, "Fair Market Rents", <www.huduser.gov/portal/datasets/fmr.html>, accessed on March 25, 2025.
3. HUD User's Office of Policy Development and Research, "HUD-USPS ZIP Crosswalk Files" <www.huduser.gov/apps/public/uspscrosswalk/home>, accessed on March 25, 2025.

Appendix

A - Stargazer Table

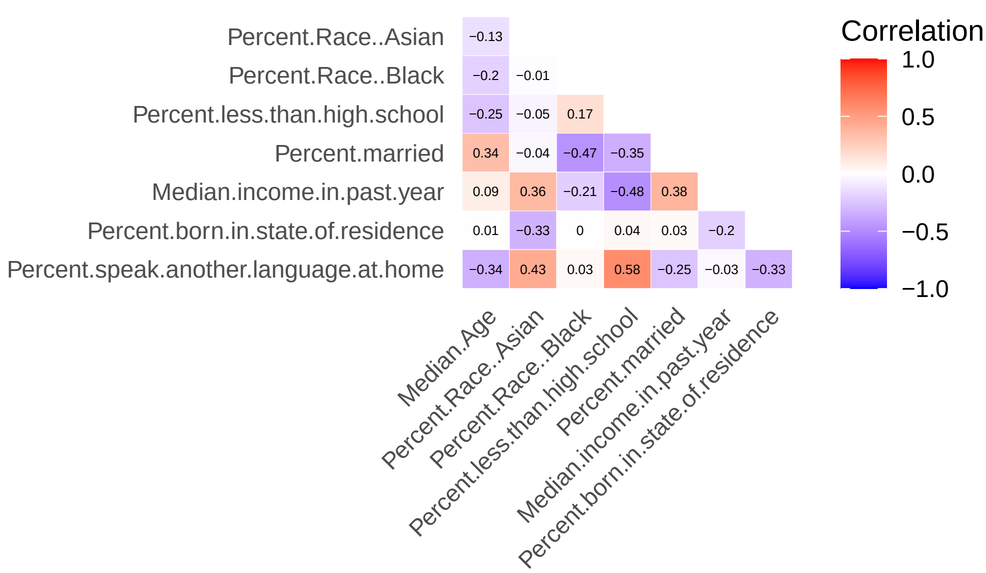
% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Thu, Apr 17, 2025 - 11:14:47 PM

Table 2: Stargazer table for model comparison

	Dependent variable:			
	SAFMR 1BR			
	(1)	(2)	(3)	(4)
Median.Age	3.392^{**} (1.383)			3.245^{**} (1.347)
Percent.married	-6.014^{***} (1.010)			-6.135^{***} (0.958)
Percent.Race..Black	0.313 (0.616)	1.589^{***} (0.549)		
Percent.Race..Asian	18.482^{***} (2.138)	30.281^{***} (2.404)		18.285^{***} (2.100)
Percent.Race..Two.or.more.races	0.783 (1.458)			
Percent.less.than.high.school	-4.134^{**} (1.967)	5.931^{***} (1.472)		-4.228^{**} (1.972)
Median.income.in.past.year	0.017^{***} (0.001)	0.017^{***} (0.001)	0.020^{***} (0.001)	0.017^{***} (0.001)
Percent.born.in.state.of.residence	-4.402^{***} (0.497)			-4.437^{***} (0.489)
Percent..100.poverty.status	−1.946 (1.383)			−1.835 (1.350)
Percent.speak.another.language.at.home	6.914^{***} (0.973)			7.149^{***} (0.871)
Constant	746.608^{***} (102.466)	217.120^{***} (59.039)	295.636^{***} (42.368)	767.694^{***} (96.921)
Observations	1,500	1,500	1,500	1,500
R^2	0.632	0.543	0.353	0.632
Adjusted R^2	0.630	0.541	0.352	0.630
Residual Std. Error	302.997 (df = 1489)	337.132 (df = 1495)	400.658 (df = 1498)	302.852 (df = 1491)
F Statistic	255.728^{***} (df = 10; 1489)	443.342^{***} (df = 4; 1495)	816.107^{***} (df = 1; 1498)	319.893^{***} (df = 8; 1491)
Note:				*p<0.1; **p<0.05; ***p<0.01

Appendix

B - Correlation Matrix



Appendix

C - Variable Table

	Variable Name	Data Type	Data Source
1	ZIP.Code	double	All
2	SAFMR.0BR	double	HUD FMR
3	SAFMR.1BR	double	HUD FMR
4	SAFMR.2BR	double	HUD FMR
5	SAFMR.3BR	double	HUD FMR
6	SAFMR.4BR	double	HUD FMR
7	Total.Population	double	Census ACS
8	Median.Age	double	Census ACS
9	Percent.Males	double	Census ACS
10	Percent.Females	double	Census ACS
11	Percent.Race..White	double	Census ACS
12	Percent.Race..Black	double	Census ACS
13	Percent.Race..Native.or.Alaskan	double	Census ACS
14	Percent.Race..Asian	double	Census ACS
15	Percent.Race..Native.Hawaiian.or.Pacific.Islander	double	Census ACS
16	Percent.Race..Two.or.more.races	double	Census ACS
17	Percent.Race..Hispanic.or.Latino	double	Census ACS
18	Percent.speak.another.language.at.home	double	Census ACS
19	Percent.never.married	double	Census ACS
20	Percent.married	double	Census ACS
21	Percent.divorced	double	Census ACS
22	Percent.widowed	double	Census ACS
23	Percent.less.than.high.school	double	Census ACS
24	Percent.high.school.graduate	double	Census ACS
25	Percent.some.college.or.associates	double	Census ACS
26	Percent.college.graduate	double	Census ACS
27	Percent.graduate.degree	double	Census ACS
28	Median.income.in.past.year	double	Census ACS
29	Percent..100.poverty.status	double	Census ACS
30	Percent..100.TO.149.poverty.status	double	Census ACS
31	Percent..150.poverty.status	double	Census ACS
32	Total.population..born.in.state.of.residence	double	Census ACS
33	Percent.born.in.state.of.residence	double	Census ACS
34	RES_RATIO	double	Census ACS
35	State	character	HUD FRM
36	Total.Gender	double	Census ACS
37	Region	character	Derived From HUD