Journal of Housing Economics 74 (2026) 102177

Contents lists available at ScienceDirect

Journal of Housing Economics

journal homepage: www.elsevier.com/locate/jhec

Race, rent, and homelessness: describing and explaining variation in racial disparities in US homelessness

Brendan O'Flaherty

Department of Economics, Columbia University in the City of New York, New York, United States

ARTICLE INFO

Keywords:
Homelessness
Race
Disparities
Community-level

ABSTRACT

Black poor people are much more likely to be homeless than white or Hispanic poor people. Why? The main correlates of disparities involving Black people are high rents, low rental vacancy rates, and segregation, with rents being by far the most important. I try to explain why with a model of selectivity on the part of landlords in a monopolistically competitive housing market, and with a model of informal housing markets. I also discuss why racial disparities are much greater in California and New York than in the rest of the country and why US homelessness is concentrated so heavily in those two states. All three questions are closely related.

In the United States, homelessness is racially concentrated. In 2019,[1] 39% of homeless Americans were Black, even though only 20% of Americans with income below the poverty level were Black. A policy that reduced the rate of homelessness among Blacks in poverty to the rate of homelessness among non-Blacks in poverty and did nothing else would reduce total homelessness by 19%.

Homelessness is also geographically concentrated. California and New York accounted for 43% of the nation's homeless population in 2019, but only 21% of the population with incomes below the poverty population. Reducing the rate of homelessness among poor people in California and New York to the rate of homelessness among poor people in the rest of the country would reduce total homelessness by 22%.

Not surprisingly, the racial and geographic concentrations are linked. Black people in California and New York account for 19% of the national homeless population but only 3% of national poverty population. A policy that only reduced the rate of homelessness among Black poor people in California and New York to the rate among the rest of the poor population and did nothing else would reduce total national homelessness by 16%. —almost as much as a policy that reduced the homelessness rate among all poor Black people.

Fig. 1 presents these numbers in a different format. The rate of homelessness among poor Black people in California and New York was staggeringly high: 108 per 1000 (over 10%). By contrast, the rate among poor non-Black people outside California and New York was 10 per 1000. The off-diagonal rates are in between but much closer to those of non-Black people outside California and New York than to Black people in California and New York: 17 per 1000 among poor Black people outside California and New York, and 22 per 1000 among non-Black poor people in California and New York. In California and New York, the difference between Black and non-Black homelessness rates is 86 per 1000; outside those states, the difference is 7 per 1000. I call Fig. 1 the "skyscraper picture."

This paper is about the racial concentration of homelessness, not explicitly about the geographic concentration. That it also sheds light on geographic concentration is an added bonus. Fig. 1, for instance, shows that the racial disparity in homelessness among poor people is greater in California and New York than in the rest of the country (86 per 1000 in California and New York, vs 7 per 1000 in the rest of the country). That is the focus of this paper. But arithmetic also guarantees that among Black poor people the disparity between rates in California and New York and those in the rest of the country (91 per 1000) is greater than the disparity between California and New York and the rest of the country among non-Black poor people (12 per 1000)—a fact relevant to the study of geographic concentration.

The first part of this paper provides a much richer description than Fig. 1 of the variation in racial disparities, particularly conditional variation. We outline the circumstances under which the disparities vanish or almost vanish, and the circumstances under which they are huge. Knowing how the disparities vary is necessary for understanding why they vary, and understanding why they vary is necessary for understanding why they exist. That's the first major part of the paper: how the disparities vary.

E-mail address: bo2@columbia.edu.

[1] 2019 was the last count where (1) all jurisdictions were required to count unsheltered homeless people, and (2) asylum seekers were not a significant share of the homeless population. On asylum-seekers and homeless population counts, see Meyer et al., 2025.

https://doi.org/10.1016/j.jhe.2026.102177
Received 26 October 2025; Received in revised form 30 August 2026; Accepted 21 September 2026
Available online 25 September 2026
1051-1377/© 2026 Elsevier Inc. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

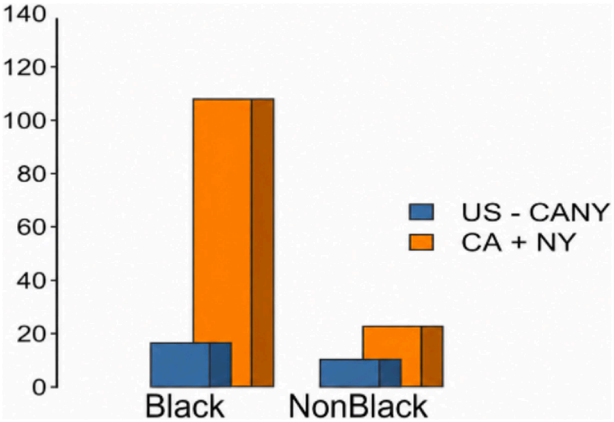


Fig. 1. 2019 rate of homelessness per thousand poverty population.

In this descriptive part, I use many different independent variables: for instance, region, weather conditions, type of area, vacancy rates, year, and segregation. My observations are administrative groups within areas—sheltered or unsheltered, families or individuals—and I treat the designations of these groups as additional independent variables.

My main approach in this section is non-parametric. Implicitly researchers often impose parametric restrictions by deciding *a priori* that they will report all results by shelter status, for instance, by family status or by the intersection of these statuses. In doing this they are following HUD administrative and reporting practices. They are treating these variables as special and privileged and thereby reducing the explanatory power that might be available to explore variables that could matter more.

In contrast, race is the only independent variable that I treat asymmetrically (except in the section about California and New York); that is because I decided to study race. Shelter status and family status are treated like rent and weather conditions: I let the data determine whether they matter enough to be included in the explanation. Sometimes they matter and sometimes they don't: in comparing Hispanics and Whites, for instance, family status is crucial, and shelter status is not. Wisdom is finite and so are data; their uses should be economized.

To summarize the results: rents are the greatest source of variation in the Black-white and Black-Hispanic disparities, while family status is the main source of variation in the Hispanic-white disparity. The importance of rents is consistent with the skyscraper picture, since rents in California and New York are higher than in the rest of the country.

The other main sources of variation in the Black-white and Black-Hispanic disparities are vacancy rates and segregation. The effect of segregation is not monotonic and is weak for the Black-Hispanic comparison, but generally greater segregation is associated with smaller disparities.

Family status, shelter status, minimum temperature around the time homeless people are counted, and Southern residency, the well-known correlates of unconditional variation, are also responsible for conditional variation but are not as powerful as rents.

In this descriptive part of the paper, I also examine more carefully why California and New York have such large racial disparities (and correspondingly why homelessness among poor Black Californians and New Yorkers is so staggeringly high). The main question is whether the different values of the covariates in California and New York—particularly higher rents and lower vacancy rates—are the only reason for the greater racial disparity. That is, whether an area in CANY (California and New York) with the same covariates as an area in the rest of the country would experience the same expected racial disparities.

Our tentative answer is yes on average for the disparity between Blacks and whites, and no for the disparity between Blacks and Hispanics. For Blacks and whites, areas with the same covariates in different regions have the same expected disparities. Los Angeles is just a big Omaha with high rents and low vacancy rates. But for Blacks and Hispanics, something is happening in California and New York that our covariates do not account for. Los Angeles is not simply a big Omaha with high rents and low vacancy rates.

The second major part of the paper explores why the disparities vary. The first part is description, and the second part is explanation. I draw on a variety of literature as well as the results from the first part to determine what explanations are plausible. Papers by (Meyer et al. (2024), Ellen et al. (2024), and Ananat and Washington (2009) are especially helpful for the explanatory task.

Why is high rent associated with large racial disparities? Probably because in tighter housing markets landlords can be more selective about whom they rent to, and this selectivity disproportionately harms low-income Blacks. We have no information about whether this discrimination is based on preferences, accurate stereotypes, inaccurate stereotypes, correlations with other characteristics like criminal records, or something else. No single landlord needs to discriminate either: for instance, landlords operating in mainly Black neighborhoods and seeing mainly Black tenants could operate differently than other landlords. Informal housing markets operate in tandem with formal housing markets in our model, and so we cannot sort out formal from informal effects.

Why is segregation correlated with less homelessness and smaller disparities when the index of dissimilarity is comparatively low? The most plausible story seems to be about Black political efficacy, which Ananat and Washington (2009) have demonstrated falls in more segregated metropolitan areas (the relationship they find is causal). More segregated cities are more repressive, offering harsher street conditions and fewer shelter opportunities. These forces are probably offset by the well-known causal relationship between segregation and many forms of relative Black distress, and when segregation is high enough, these forces associating greater segregation with greater Black homelessness sometimes prevail.

Richard (2025) is a recent paper that also contributes to our knowledge of racial differences in homelessness. Like this paper, it finds that variation in rents affects Black homelessness much more seriously than Hispanic or white homelessness. The analytical part of the paper, however, never considers disparities as dependent variables, instead presenting side-by-side linear regressions with race-specific homelessness rates as dependent variables. These equations cannot be easily translated as statements about disparities as I define them because the homelessness rates are logged. They could be translated into statements about the logs of ratios between homeless rates, but, as I will argue in the methods section of this paper, ratios are not useful measures to study. (Richard never uses ratio measures either, except for a passing reference to one ratio on the top of page 9.)

In the body of this paper, we concentrate on comparing Blacks with other groups: (non-Hispanic) whites and Hispanics. Comparisons between Hispanics and whites are in appendix.

The body of the paper presents unweighted GRF results, with occasional references to weighted GRF results. Appendices present regression results for these comparisons as well as GRF results for the HW comparison.

Before moving on, it is worthwhile to look at the skyscraper picture again briefly and note that it casts very serious doubt on two popular explanations for the Black-non-Black disparity in homelessness: that it is driven by white racism and that it is driven by Black culture. Both popular explanations have a hard time with the skyscraper picture: is it plausible that white racism is so much more virulent or Black culture so much more pathological in California and New York than in the rest of the country, and that nobody has ever noticed this before? Human ingenuity is amazing, and perhaps someone can develop a theory that rescues one or both of these explanations, but any imagined rescue is not likely to be a simple story.

HOW: DESCRIBING HOW RACIAL DISPARITIES VARY.

This part of the paper is almost entirely empirical. It starts with discussions of data and methods and then presents results.

1. Data

Dependent variables. Our dependent variables are functions of the numbers of homeless people in various categories in each administrative region in the annual point-in-time count. We observe aggregates, not information about individuals. The administrative regions are called Continua of Care (CoCs). They were set up by the Department of Housing and Urban Development (HUD) to administer homeless programs and collect data. There were between 381 and 386 CoCs during the years that we study. Some of them are recognizable units like cities or counties, but others are portions of counties, sets of counties, or other combinations.

Our data span the period from 2014 through 2022. During that period, 29 new CoCs were created from old ones, and 32 CoCs were merged into others (we do not count the merger of three CoCs into the Maryland Balance of State because it appears to have occurred before 2014). We drop the CoCs in the island territories (4 before 2017 and 5 after that).

Since 2007, every CoC has been required to conduct a point-in-time (PIT) count of sheltered homeless people in the last week of January every year, and a point-in-time count of unsheltered homeless people at the same time every odd-numbered year. Many CoCs also conduct and report unsheltered counts in even-numbered years. However, because of COVID, hardly any unsheltered counts were conducted in 2021,

For each year and each CoC, HUD divides the homeless population into $3 \times 2 \times 2 = 12$ cells formed by the intersection of three race/ethnicity groups (Blacks, Hispanics, and whites), two shelter statuses (sheltered or unsheltered), and two relationship statuses (families or individuals). "Families" means households that include at least one person 18 or older and one person younger than 18. "Individuals" means people who are not members of families. HUD does not specify definitions of Black, Hispanic, and white.

An observation in our data set is a combination of race, date, location, shelter status, and family status. For instance, sheltered Black families in Bergen County, NJ in 2018 is a typical observation.

We use point-in-time counts for 2014 through 2022 from the website https://www.huduser.gov/portal/datasets/ahar/2024-ahar-part-1-pit-estimates-of-homelessness-in-the-us.html . Before 2014, the detail we require is not available. We omit unsheltered data for even-numbered years when it is the same as the data for the previous odd-numbered year because this indicates that an unsheltered count was not conducted that year, and we omit all unsheltered data for 2021.

We also omit all counts for unsheltered families. Unsheltered families account for only about 2% of total homeless population, and the counts are heavily influenced by variation in state child protection laws, since many states will remove children from an unsheltered family—rendering the family no longer a family for HUD purposes—or place them in shelters—rendering them no longer unsheltered. For instance, New York City, with at least 30,000 people in families in shelters every year of our study period, reported precisely zero unsheltered families for this period. Because of our decision, each CoC-year has potentially 9 observations: 3 races/ethnicities x 3 statuses—sheltered individuals, sheltered families, and unsheltered individuals).

To find rates of homelessness, we link CoCs to census tract data using Byrne's HUD-CoC Geography Crosswalk available at https://github.com/tomhbyrne/HUD-CoC-Geography-Crosswalk/blob/master/output/tractoc_match.csv . From the census we use tract-level data from the American Community Survey, 5-year estimates for 2017 for non-Hispanic whites (alone), Blacks (alone) and Hispanics (alone), since we believe HUD uses "white" to mean "non-Hispanic white" (because the sum of the three groups is often the total for entire population or close to it).

Table 1
GRF dataset derivation.

Year	Non-territory CoCs	Nine times CoCs	Other deliberate exclusions	Missing dependent variables	GRF data set
2014	376	3384	249	6	3129
2015	379	3411	0	6	3405
2016	378	3402	171	6	3225
2017	379	3411	0	24	3387
2018	379	3411	159	20	3232
2019	378	3402	0	0	3402
2020	379	3411	138	90	3183
2021	380	3420	1140	2	2278
2022	380	3420	69	3	3348

We then divide the homeless populations by the corresponding census population below the poverty level to obtain rates of homelessness per 10,000 poverty population. Notice that we do not use separate estimates of individual and family poverty population—the rates we use are, for instance, homeless Hispanic individuals per 10,000 Hispanic poverty population, not homeless Hispanic individuals per 10,000 Hispanic individuals in the poverty population. This is mainly because the Census definition of family is different from the HUD definition.[2]

Notice also that the races/ethnicities we use for the denominator are not precisely the same as those we use for the numerator. We make no apologies for this and attempt to make no correction because we do not know what the races/ethnicities that appear in PIT counts mean, and neither does anyone else. The thousands of enumerators working independently, often in the dark of night, did not have uniform instructions, if they had any instructions at all. Since no one in the US has an official permanent race (much less a "true" race), our relaxed attitude is not inappropriate. For many Americans (but not all), almost everyone in the US agrees on what their race/ethnicity is.

For some observations, we were unable to create the dependent variable because the poverty data were missing (most commonly), or HUD did not publish an appropriate homeless count. We dropped these observations. We use the remaining observations for our GRF analysis and so we call this data set the GRF data set. Table 1 shows the derivation of the GRF data set. For each year it shows the number of non-territory CoCs that HUD lists, nine times that number to account for the exclusion of unsheltered families, the number of exclusions because of other deliberate exclusions (COVID and even year unsheltered count skips) and the number of observations for which we do not have dependent variables. This constitutes our GRF data set.

Independent variables. Our independent variables come from several sources. We have already discussed the race/ethnicity variable, which comes from HUD's compilation of PIT counts. The dummy variables for sheltered or unsheltered status, and family or individual status come from the same source. We also use a dummy variable for whether a CoC is in a state that is part of the region the Census calls South, and dummy variables for whether HUD considers a CoC a major city, other largely urban, largely suburban, or largely rural.

We use two covariates to reflect weather conditions close to the date of the PIT: the minimum temperature on January 25 each year, and a

[2] For PIT counts, HUD defines a family as any group of people experiencing homelessness together that includes at least one person under 18 and one person 18 or over (AHAR 2024, p. viii). The census defines a family as a group of people living in a household that includes at least one person 15 or over and who are related by birth, marriage, or adoption. Trivially, at any one time, it is impossible to be both a HUD family and Census family, because all HUD families are homeless and no Census families are homeless. A HUD family that moved into a house would not necessarily become a Census family because they might not be related by birth, marriage, or adoption. A Census family that became homeless would not necessarily become a HUD family because it might not include any member under 18 or 18 or over.

dummy variable for precipitation that day. The data are from the National Oceanic and Atmosphere Administration (NOAA) https://www.ncei.noaa.gov/pub/data/ghcn/daily/by_year/ . These data are for weather stations.

We match the weather stations to CoCs by the following process. First we assign the weather stations to counties by matching latitude and longitude from NOAA https://www.ncei.noaa.gov/pub/data/ghcn/daily/ghcnd-stations.txt . Then we take a weighted average over the counties in the CoC. The weights are given by the share of CoC poverty population in each county. Poverty population in each county for each year is estimated using county population from the Census at https://www2.census.gov/programs-surveys/popest/technical-documentation/file-layouts/2010–2020/ multiplied by the poverty rate from Data USA, 2022 https://datausa.io/map?measure=2fQpSH&groups%5B0%5D=ZFt1KK%7C0&time=2022.

Housing market variables. For rents, we took the Fair Market Rent (FMR) that HUD uses for various housing programs. This is approximately the 40th percentile of the rent distribution. We use county-level data for two-bedroom apartments from https://www.huduser.gov/portal/datasets/fmr.html. For CoCs with more than one county, we use the same weighting process that we used for the weather variables.

For rental vacancy rates, we use data for individual census tracts and years from census table B25002 https://data.census.gov/table?q=B25002:%20Occupancy%20Status&g = 010XX00US$1400000

These were aggregated to the CoC level using the Byrne crosswalk and weighting by census tract poverty population from table S1701 https://data.census.gov/table/ACSST1Y2022.S1701?q=S1701:%20Poverty%20Status%20in%20the%20Past%2012%20Months.

Segregation. Our data on residential segregation originate with Diversity and Disparities, a project of the Russell Sage Foundation and Brown University, 2021a, at https://s4.ad.brown.edu/projects/diversity/Data/data.htm. We use data for 2020. This project has dissimilarity indices and isolation indices for many types of geographical region and several races and ethnicities.

Isolation and dissimilarity are both commonly used measures of segregation, and they are correlated in our data set. (For 2019, the weighted correlation at the CoC level between the Black-white dissimilarity index and Black isolation is 0.78; between the Black-Hispanic dissimilarity index and Black isolation, 0.68; and between the Hispanic-white dissimilarity index and Hispanic isolation, 0.62.)

However, the two measures aim to measure distinct notions. Dissimilarity is a binary (and symmetric) measure of how different one group's location is from another group's location. It tells what proportion of group A would have to move in order for group A to be distributed across neighborhoods in the same way that group B is actually distributed. Dissimilarity is not strongly related to the size of a group in the following sense: if a new member of a group is added in a randomly selected neighborhood, the dissimilarity index could go either up or down.

Isolation is a one-group measure of how little exposure the average member of a group has to non-members. For group A, it tells the proportion of group A members in the typical neighborhood that a group A member lives in Isolation is positively related to the size of a group in the following sense: if a new member of the group is added in a randomly selected neighborhood, the index of isolation always goes up.

The differing reactions to an additional member suggest that the isolation index for a group reflects in part the size of a group within a metropolitan region, not just its separateness from other groups. For some purposes, this dual nature of the isolation index may not be a problem, but our purposes are not one of those. The small Black-white disparity in the South is a puzzle that we want to study, and the proportion of Blacks in the South is considerably greater than the proportion in other regions. High isolation indices may be correlated with Southern location because of the large size of Black populations in Southern metropolitan regions. Because this correlation may make the interpretation of results about the isolation index difficult, we have opted to use the dissimilarity index in this paper. Using the isolation index produces many similar results, but these results cannot be so cleanly interpreted.

Matching dissimilarity indices to CoCs took work. We started in Diversity and Disparities with indices for metropolitan divisions. There are two kinds of metropolitan divisions: those that stand alone like the Albuquerque metropolitan division (Bernalillo, Torrance, Valencia, and Sandoval counties in New Mexico), and those that are parts of larger areas called metropolitan statistical areas (like the Newark metropolitan division (Essex, Morris, Union, Sussex, and Hunterdon counties in New Jersey and Pike County in Pennsylvania); which is part of the larger New York-Newark-Jersey City metropolitan statistical area). We use both kinds of metropolitan divisions, but we do not use the larger metropolitan statistical areas.

To assign dissimilarity indices to CoCs, we use a process like that which we used for weather variables. Metropolitan divisions consist of sets of counties, and so we assign dissimilarity measures to each county for which we have a metropolitan division. Using Byrne's crosswalk, we assign CoCs to counties and thus assign dissimilarity measures to CoCs. When a CoC contains counties that are parts of more than one metropolitan division, we take the poverty-population-weighted average of the county dissimilarity indices. For some CoCs we find no county in a metropolitan division, and so we drop those CoCs.

Some CoCs include more than one metropolitan division, and some metropolitan divisions contain more than one CoC. For clustering purposes, therefore, we created "super-metro regions." The super-metro for a CoC that includes only one metropolitan division is that metropolitan division (no matter how many other CoCs are also in that metropolitan division). The super-metro for a CoC that includes parts of more than one metropolitan division has a name that is a conjunction of the names of those metropolitan divisions.

Missing independent variables. In the BW data set, 1701 observations were missing both weather covariates, minimum temperature and precipitation. In addition, 1340 observations lacked a segregation measure and 6 lacked a vacancy rate. Hence 4748 covariates were missing. The BW data set contains 20,051 observations, each with 12 potential covariates, not counting year dummies, for a total of 240,612 possible covariate values. Thus 2% of covariate values were missing. For the BH and HW data sets, the numbers are slightly different because of different missing dependent variables, but the proportion of missing covariate values still round to 2%.

Poverty and deep poverty. Homelessness may be more associated with deep poverty than simple poverty and so perhaps we should either use deep poverty as the denominator in our homelessness rates or as a control variable. By deep poverty I mean the proportion of the relevant population in households with income less than half of the relevant poverty threshold. However, measures of deep poverty are not available for racial and ethnic groups in geographies as fine as those we use for CoCs.

Moreover, using deep poverty rather than poverty would probably make little difference to our analysis because deep poverty rates (relative to population) are highly correlated with poverty rates in the data where both numbers are available.

Readers who think that deep poverty is the appropriate measure should therefore be content that poverty is an excellent proxy.

2. Methods

This paper focuses on disparities between races in rates of homelessness. I define the disparity between two races as the difference between the two rates of homelessness. Disparity has a natural unit—people per 10,000 people in poverty—and answers the question: How many more or fewer people would be homeless if poor people of this race were equally likely to be homeless as poor people of the other race?

Sometimes "disparity" is defined as the ratio of homeless rates, rather than the difference. I do not use this definition because it has no natural

unit and is not tied to the experience of either group. Mendez (1983), p. 13) makes the same point in the discussion of police homicides: "This difference of rates measure reflects the excess [of] black deaths over white deaths and hence reveals the adverse impact upon the black population of police use of deadly force. Comparing rates as a ratio of black rate to white rate is less desirable."

We use two techniques: generalized random forests (GRF), and regression. In each case we examine three different comparisons: Blacks vs whites (BW), Blacks vs Hispanics (BH), and Hispanics vs whites (HW). Disparities—the difference between the homelessness rate of one race/ethnicity and that of another—are our chief concern, and so it is most natural to operate in dyadic environments. GRFs, moreover, require that disparities be the dependent variable. In all our analyses, we cluster standard errors by CoC and by super-metro.

For our main results we do not weight CoCs by population or poverty population. The results are thus about how disparities differ among CoCs on average, not among people. This procedure has several advantages: it means that we do not have to worry about either overlap or outsize influence for atypical CoCs, especially New York City and Los Angeles City/County. See Athey and Wager, 2019 for an example of a similar approach.

We provide complete weighted GRF and weighted regression results in the appendix for the chief specifications. But we include weighted GRF results in the text when they differ from the unweighted results in ways that add to our analysis.

This part of the paper makes no claims about causality; it is purely descriptive. GRFs were devised by Athey, Tibshirani and Wager (2019) to analyze heterogeneous treatment effects in randomized control trials, and so the results had causal interpretations in that context. They noted, however, that the same calculations could be applied to observational data (although without the causal interpretation). They applied GRF to observational data in Athey and Wager, 2019. For our purposes, race plays the role of treatment in the original papers (as in Montiel Olea et al., 2023). Of course, we don't believe that race gets assigned to people randomly or is a treatment in any meaningful sense. However, the conditional disparities are calculated the same way that treatment effects of race would be calculated if race were assigned randomly.

For missing independent variable values, we used the "missing incorporated attributes" (MIA) criterion (Twala et al., 2008; Naf, 2023). GRF involves tree-splitting, MIA is a tree-splitting method that bypasses the need to impute data (Tang and Ishwara, 2017). It is incorporated in the GRF algorithm package. Consider some covariate X, ordered or numeric, and suppose that for some observations the value of X is missing. Normally the splitting rule would search over values of s and for each s consider the split into two nodes

$$\{X \leq s\} \textit{ versus } \{X > s\}.$$

Table 2
Unweighted summary statistics by race (standard deviations in parentheses) for the GRF data set.

VARIABLES	Hispanic	Black	White
N	9365	9455	9595
Homeless per 10k poverty	35.31 (0.69)	107.40 (2.57)	64.84 (1.55)
Sheltered	0.722 (0.005)	0.715 (0.005)	0.710 (0.005)
Family	0.361 (0.005)	0.357 (0.005)	0.355 (0.005)
Largely rural	0.280 (0.005)	0.280 (0.005)	0.283 (0.005)
Largely suburban	0.433 (0.005)	0.433 (0.005)	0.431 (0.005)
Major city	0.126 (0.003)	0.125 (0.003)	0.126 (0.003)
Other largely urban	0.161 (0.004)	0.161 (0.004)	0.161 (0.004)
Temperature min (°C)	−2.554 (0.082)	−2.556 (0.082)	−2.578 (0.081)
Precipitation (mm)	2.439 (0.065)	2.436 (0.065)	2.433 (0.065)
Rent (dollars)	1059 (3.9)	1057 (3.9)	1056 (3.8)
South	0.370 (0.005)	0.372 (0.005)	0.371 (0.005)
Vacancy per 100	12.34 (0.071)	12.35 (0.071)	12.40 (0.071)
Dissimilarity Index (BH)	34.22 (0.12)	34.25 (0.12)	–
Dissimilarity Index (BW)	–	48.26 (0.13)	48.23 (0.13)
Dissimilarity Index (HW)	38.02 (0.11)	–	37.94 (0.11)

The MIA criterion considers a larger set of possible splits for each s:

$$\textit{Split A}: \ \{X \leq s \textit{ or } X \textit{ missing}\} \textit{ versus } \{X > s\}$$

$$\textit{Split B}: \ \{X \leq s\} \textit{ versus } \{X > s \textit{ or } X \textit{ missing}\}$$

$$\textit{Split C}: \ \{X \textit{ missing}\} \textit{ versus } \{X \textit{ not missing}\}$$

For each s and split type (split A, B, or C) that maximizes the objective criterion, and then finds the *optimum optimorum* over values of X.

We redid our GRFs with the regression data set (which drops entire observations if any covariate value is missing) instead of using MIA. The results were not substantially changed.

The weaknesses of linear regressions are well-known, especially their dependence on a priori functional form. The GRF results are therefore more reliable, but we will present regression results in the appendix. As far as the paper's main results are concerned, the two methods are reasonably well aligned, with some exceptions that we note.

3. Summary statistics by race

Table 2 below presents summary statistics by race for the GRF data set. Since we have three observations for each CoC-category pair, averages on a CoC level should be approximately the same for all races, with the only differences being due to missing dependent variables. When an independent variable was missing, we did not include anything in the average (implicitly imputing the average of the non-missing values). Note that in the GRF sample, the unconditional unweighted Black-white homelessness disparity is 42.6; the Black-Hispanic disparity is 72.1; and the average Hispanic-white disparity is −29.5.

Recall that these are unweighted averages over CoCs. Summary statistics for the regression data set are only slightly different.

4. GRF results

Our chief results are for the Black-white and Black-Hispanic comparisons. The Hispanic population in poverty is large relative to the white population in poverty: in 2019, the Hispanic poverty population was 67% of the non-Hispanic white poverty population in the nation and so it is important to present results for both groups. Results for the HW comparison are in the appendix.

Because the covariates associated with all three racial groups are almost completely the same, the estimated average treatment effects should be almost the same as the unconditional differences. In our results, they are, and there is no need to publish them.

The key results from GRF are about observation-specific estimated unexplained disparities. We call these CATEs: conditional average treatment effects. This follows our practice of using the nomenclature associated with causality without implying cause. Sometimes we abbreviate the CATE as tau or tauhat. For each observation, the CATE means the estimated unexplained disparity between a Black and a white observation, each with the same non-racial characteristics as the observation in question. These are calculated as "out-of-bag" (OOB) estimates: they are the predicted values for trees that do not include the observation. (The sample for each of the many trees is drawn randomly.)

Figs. 2 and 3 present the distribution of CATEs for the BW and BH comparisons, with Fig. 3 being a close-up of Fig. 2 for the smaller values. Figs. 2a and 3a for BW and 2b and 3b for BH.

Only a small fraction of CATEs are negative. The modal value is small, less than ten, but the right-hand tail is very long.

Table 3 provides a measure of variable importance. For each independent variable, this table gives the proportion of times that a tree splits on that variable, relative to the number of times a tree could have split on that variable (remember that the variables used in each tree are randomly determined). All variables that split >1% of the times they could have been used are listed in order.

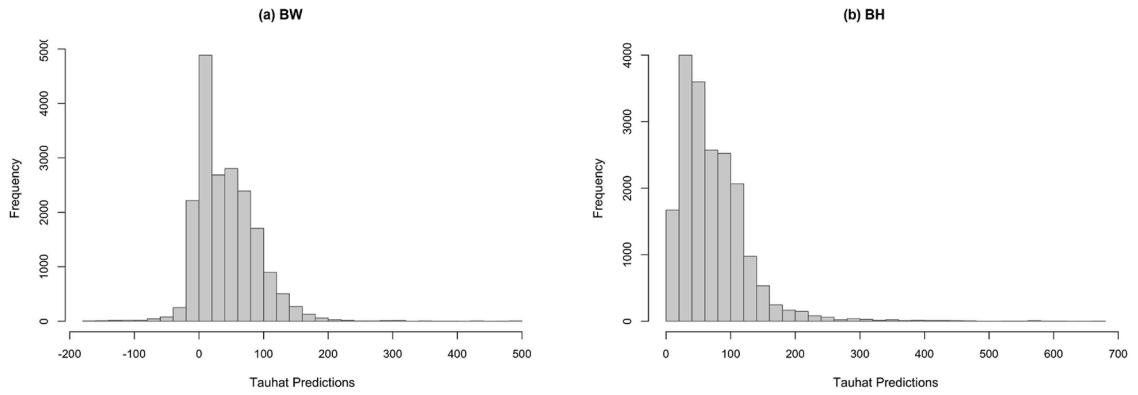


Fig. 2. Histogram of CATE predictions.

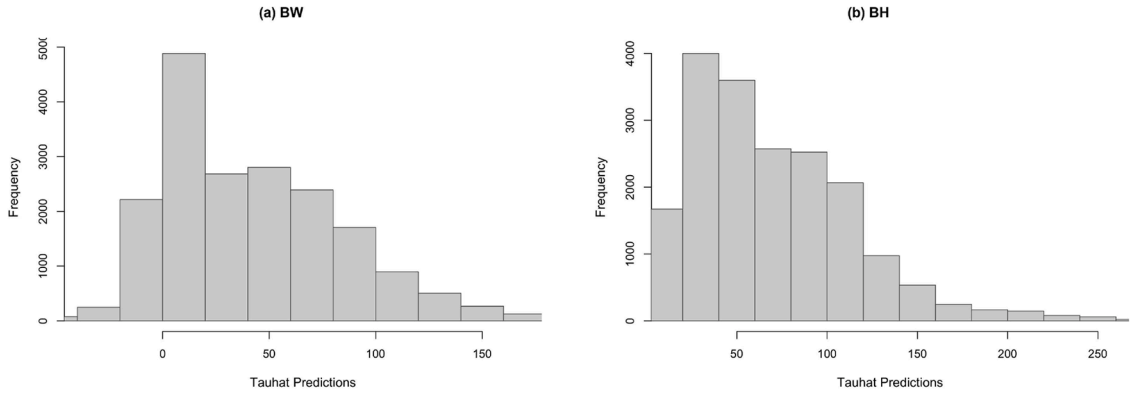


Fig. 3. Histogram of CATE predictions: central 98%.

Table 3
Top covariates by probability of split (all covariates with >1% probability).

Black–white comparison			Black–Hispanic comparison		
Rank	**Variable**	**Probability of split**	**Rank**	**Variable**	**Probability of split**
1	Rent	34.1%	1	Rent	46.2%
2	Dissimilarity index (BW)	17.2%	2	Vacancy rate	14.9%
3	Sheltered	13.5%	3	South	11.2%
4	South	12.0%	4	Dissimilarity index (BH)	10.4%
5	Minimum temperature	7.84%	5	Minimum temperature	7.13%
6	Vacancy rate	7.21%	6	Sheltered	3.55%
7	Family	5.88%	7	Major city	1.06%

The overwhelming message from Table 3 is that rent is the most important variable for understanding the two disparities. For BW, 34% of the time that a tree could have split on rent, it did so. For BH, a tree split on rent 46% of the time it could have done so. For BW, the next two most important variables are segregation (17%) and shelter (13%) in the unweighted data, but with the weighted data, vacancy rate is in third place (15%). For BH in the unweighted data, vacancy rate is in second place (15%) and South is third (11%), but with weighted data, segregation is third (12%). The message is that housing market tightness (rent and vacancy rate) matters a great deal.

Table 4 examines heterogeneity in the CATEs. Consider a specific covariate, say rent. We array observations by the size of CATE. For each interval of observations arrayed by CATE, we find the average rent. The purpose is to see how values of the variable are associated with values of CATEs. These interval averages are shown in Table 4. Table 4 shows results for rent, the variable we are most interested in, as well as for vacancy rates, segregation, and South, the other variables that seem to have some influence. I add shelter and family because these are binary variables that are hard to illustrate in some contexts.

In this table, in panel a, the average CATE in the first decile of the Black–White CATE distribution is −17.6, and the average FMR for observations in that decile is 979.2. In panel b, the average CATE in the first decile of the Black–Hispanic CATE distribution is 16.7, and the average FMR for observations in that decile is 759.2.

Fig. 4, Fig. 5, and Fig. 6 are scatter-plots that repeat the information in this table in different form for the continuous covariates. Consider rent. For each value of rent, we show the CATEs of the observations where rent has that value. We fit a LOWESS line (locally weighted scatter-plot smoothing) to summarize the scatter-plot (Cleveland, 1979). We do not present scatter-plots for binary variables.

Both Table 4 and Fig. 4 show that higher rent is correlated with greater disparity. The greatest effect is in the top 20%. Note also that CATE for very low observed values of rent in figure, the average CATE is close to zero: the disparity almost disappears when rent is very low. Similarly, Table 4 and Fig. 5 show that lower vacancy rates are correlated with greater disparity. Together, we see that tighter housing markets are correlated with greater disparity.

Similar strong correlations can be seen for shelter, even though trees do not often split on shelter. <5% of observations in the bottom decile by tauhat are sheltered, <5% in the second decile. But about 90% of the observations in the top two quintiles are sheltered.

For the South and dissimilarity, Table 3 does not show a monotonic relationship. In the bottom half of the distribution of CATEs, dissimilarity is negatively correlated with CATE, and Southern location is positively correlated. The reverse occurs in the top half of the distribution.

Table 4
Average covariate values by interval of estimated CATE.

PANEL a BW **Variable**	**0–10%**	**10–20%**	**20–40%**	**40–50%**	**50–60%**	**60–80%**	**80–100%**
$\hat{\tau}$ (tauhat)	−17.621	0.482	10.960	27.626	42.273	62.955	111.568
Rent	979.151	930.048	884.373	945.688	1034.037	1070.926	1382.720
Vacancy per 100	15.614	11.523	14.148	12.818	11.216	11.719	10.420
Sheltered	0.071	0.151	0.619	0.894	0.937	0.975	0.943
Family	0.000	0.000	0.260	0.349	0.283	0.441	0.765
Dissimilarity index (BW)	48.665	51.274	47.364	50.154	50.810	46.351	46.835
South	0.560	0.471	0.741	0.583	0.319	0.096	0.053
PANEL b BH **Variable**	**0–10%**	**10–20%**	**20–40%**	**40–50%**	**50–60%**	**60–80%**	**80–100%**
$\hat{\tau}$ (tauhat)	16.719	24.541	39.424	54.982	67.639	90.363	149.487
Rent	759.189	833.330	918.241	928.985	949.083	1079.986	1558.282
Vacancy per 100	14.064	14.799	14.513	13.282	12.247	11.428	8.574
Sheltered	0.236	0.497	0.647	0.812	0.878	0.908	0.828
Family	0.222	0.269	0.344	0.368	0.439	0.433	0.371
Dissimilarity index (BH)	33.875	34.377	35.781	35.508	32.843	31.234	35.660
South	0.539	0.733	0.631	0.405	0.332	0.134	0.086

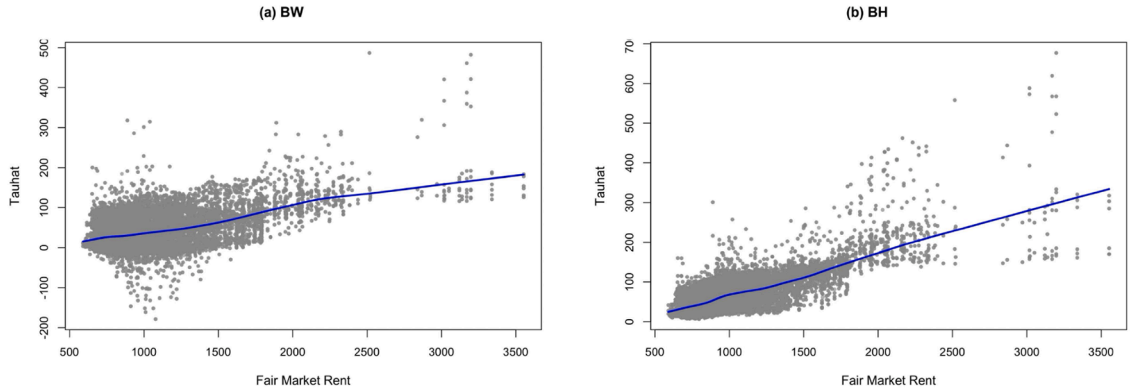


Fig. 4. Treatment effect vs. fair market rent.

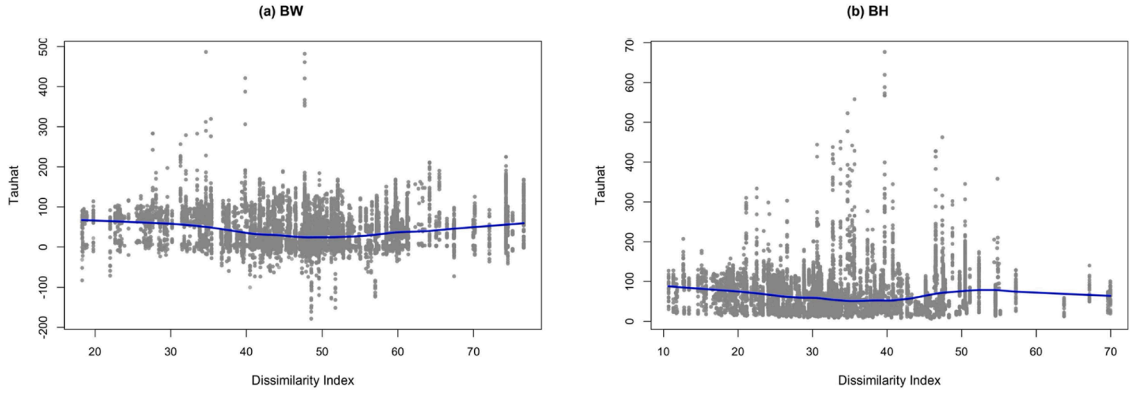


Fig. 5. Treatment effect vs. dissimilarity index.

5. Why are racial disparities so much greater in California and New York?

Since the racial disparity is greater where housing markets are tighter and California and New York (CANY) have tighter housing markets than the rest of the country, a natural question is whether we can attribute the difference in racial disparities completely or mostly to tighter housing markets. A slightly more general question is whether we can attribute the entire difference in racial disparities to our measured independent variables, or whether something unobserved sets those two states apart from the rest of the country: Do observations otherwise identical in our data set have systematically larger racial disparities if they are in CANY rather than in the rest of the country?

In this section we use a causal GRF to answer the general question, once for the Black-white disparity and once for the Black-Hispanic disparity. (We derive a similar result from a Blinder-Oaxaca decomposition in the appendix.[3])

To do this, we first transform our dependent variable for each CoC-year-shelter status-family status in the data set to a homeless disparity: the difference between the Black homeless rate and white (or Hispanic)

[3] GRF avoids the linearity restriction that limits standard Blinder-Oaxaca analysis.

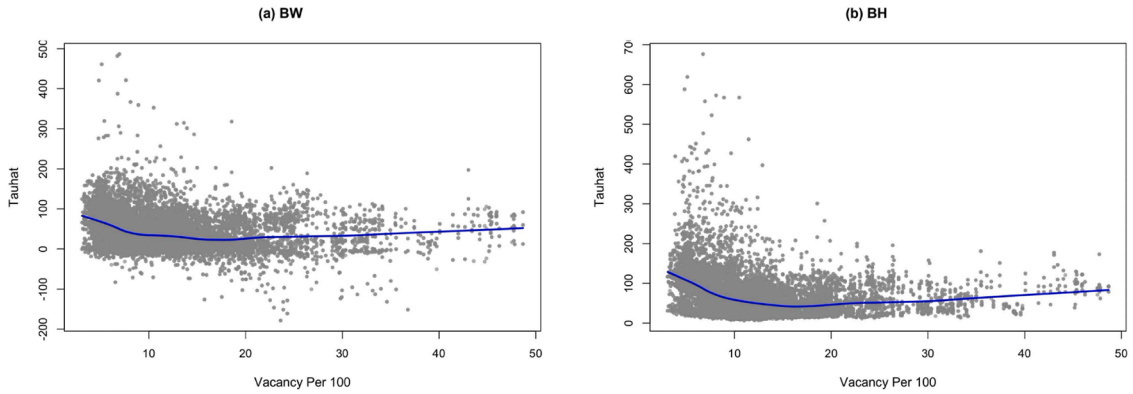


Fig. 6. Treatment effect vs. vacancy per 100.

homeless rate (per 10,000 people in poverty). This cuts the number of observations in half, of course. Then we consider CANY as the treatment group and the rest of the country (RoC) as the control group and estimate a causal GRF. (Of course, we don't interpret CANY as a Neyman-Rubins treatment any more than we interpreted race as a Neyman-Rubins treatment). The CATE for a CoC-year-shelter status-family status observation is then an estimate of how much greater the racial disparity is in CANY (for that CoC-year-shelter status-family status) than the racial disparity in the RoC. In rough terms, it is the effect of being in CANY rather than RoC, conditional on CoC-year-shelter status-family status. Or the "unexplained difference" or "residual" in Blinder-Oaxaca terms.

If all relevant differences between CANY and RoC are captured by the variables in our data set, then CANY has a greater average racial disparity than RoC only because it has different average characteristics: all the CATEs are zero, and so are the average treatment on the treated (ATT) and average treatment on the controls (ATC). In that case, there is no room for an explanation based on misguided policies, woke culture, glitter, or any other aspect of CANY life we have not studied (except as these explanations affect the independent variables in our data set). Significantly positive CATEs indicate that CANY exceptionalism in some yet unstudied dimension is at least part of the story.

Before running the GRF, however, we must eliminate observations in CANY that have no or few counterparts in RoC, and observations in RoC that have no or few counterparts in CANY. One cannot meaningfully compare an observation in CANY to a similar observation in the RoC when there is no such observation, or when such observations are very rare. And vice versa. Hence, we use propensity scores (calculated by random forest) that estimate the probability that an observation is in CANY as a function of the independent variables. Following Crump et al. (2009) we drop all observations with a propensity score <0.10 or greater than 0.90. The resulting data set (the observations with propensity scores in [0.10, 0.90]) is called the overlap (sometimes also called the common support).

This process resulted in dropping a large proportion of observations from the control group and a small proportion of observations from the treatment group. Thus, a large proportion of RoC observations have no counterparts in CANY. Most CANY observations have counterparts in the RoC, however. Since we are testing for CANY exceptionalism, not RoC exceptionalism, this is reasonably good news (the results would probably not change, for instance, if the US annexed Greenland or Oman, since they would be in the RoC but probably not in the overlap).

For the BW comparison, CANY originally had 1633 observations, of which 1495 were retained in the overlap. The rest of the country (RoC) originally had 7801 observations, of which 4847 were retained in the overlap. For the BH comparison, CANY originally had 1621 observations, of which 1486 were retained in the overlap. The rest of the country (RoC) originally had 7689 observations, of which 4022 were retained in the overlap.

In the overlap, as shown in Table 5, the ATT and ATC for the BH comparison are statistically significant and positive. The ATT and ATC for the BW comparison are both statistically insignificant.

For the BH comparison, we can reject the hypothesis that the variables in our data set fully explain why the disparity in CANY is so large (although the variables explain a large portion of the difference). We cannot reject this hypothesis for the BW comparison. The regression results in appendix 1 support these conclusions.

Examining the GRF results for the BH comparison can give us more information about where the unexplained differences are greatest or least. The three most important covariates in the GRF are segregation (splitting 17.7% of possible nodes), rent (17.6%), and minimum temperature (16.2%). The scatter-plots for these three covariates are below in Figs. 7–9.

For all three covariates, greater values are associated with greater CATEs in a fairly smooth fashion.

The minimum temperature result is probably easiest to understand. Lower minimum temperatures in CANY are mainly in New York and those observations are being compared with observations in cold states outside CANY like Michigan and Illinois. Higher minimum temperatures in CANY are mainly in California and those observations are being compared with observations in warm states like Florida and Texas. The scatter-plot indicates that California is more different (greater racial disparity) from Florida and Texas than New York is from Michigan and Illinois.

On rent, the finding is that moderate rent CoCs in CANY are pretty similar to moderate rent CoCs in RoC in racial disparity, but the highest rent CoCs have considerably greater disparities. The picture may be that smaller cities in CANY—Buffalo and Stockton, for instance–are somewhat similar to their counterparts in RoC—Pittsburgh and Reno, for instance—while the highest rent cities are more different from their counterparts. The Los Angeles, New York, and San Francisco metropolitan areas are more different from high rent metropolitan areas like Honolulu and Boston than smaller cities in CANY are from their RoC counterparts. How they are more different is an open question that we cannot address in this paper.

On the dissimilarity index, I have no reasonable explanation at this time.

Table 5
Estimated effects of being in CANY, conditional on all covariates (GRF).

	Black–white comparison	Black–Hispanic comparison
Unconditional average disparity	42.8	73.6
Average treatment effect on the treated (ATT)	16.8 (31.6)	67.0 (24.2)
Average treatment effect on the controls (ATC)	−16.3 (36.2)	46.9 (14.2)

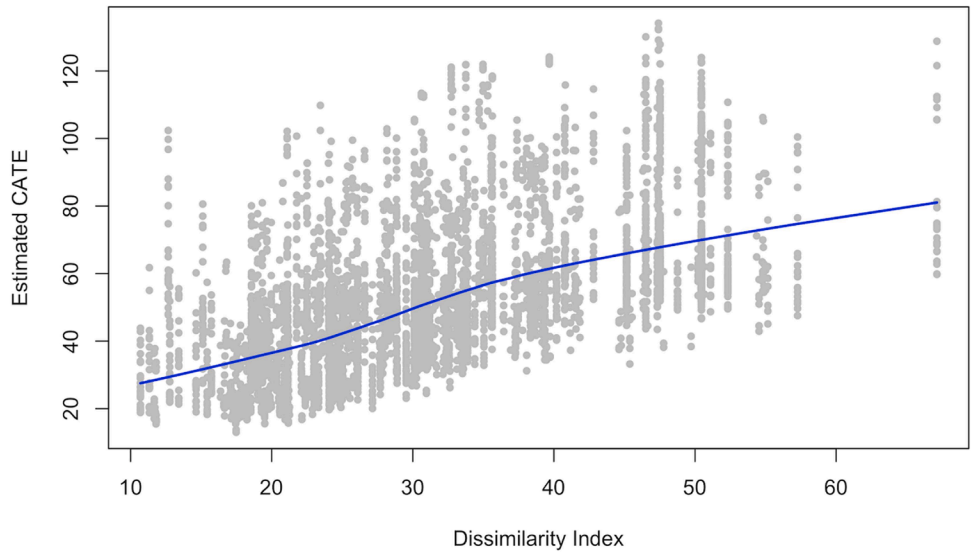


Fig. 7. Treatment effect vs. dissimilarity index.

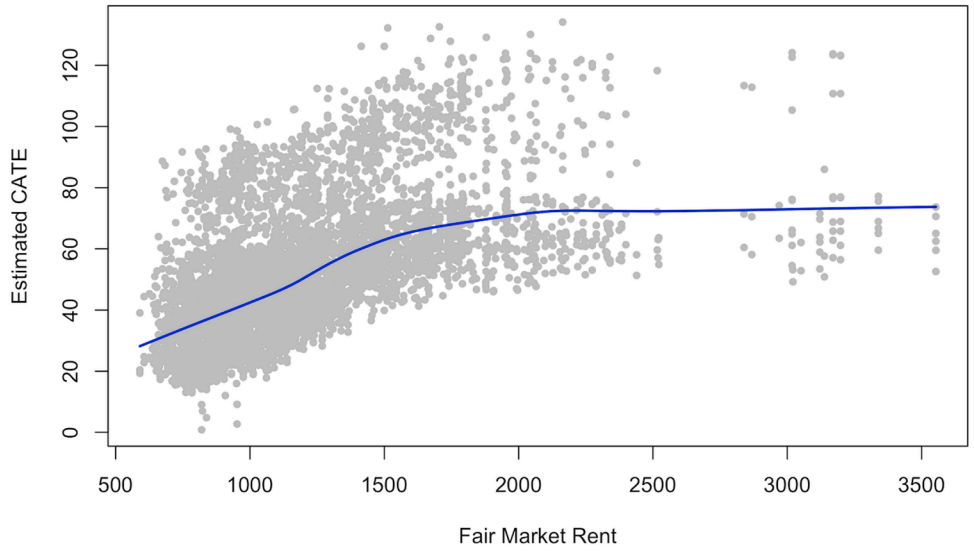


Fig. 8. Treatment effect vs. fair market rent.

WHY: EXPLAINING WHY THE DISPARITIES VARY

In this part of the paper, we concentrate on the three most important variables—rent (and housing market tightness generally), vacancy rates, and segregation. We ask why variation in each of these three variables is associated with variation in the Black-white and Black-Hispanic homelessness disparity. Since rent and vacancy rates are negatively correlated in the data and theory also suggests they will be correlated, we consider them together.

5.1. Rent and vacancies

The strongest result from the descriptive section was that disparities involving Black people varied with housing market tightness: high rent and low vacancy CoCs had much greater disparities than moderate and low rent CoCs. This section sets out our most plausible theory of why this should be the case and then provides evidence that supports this theory. Finally, we discuss why several alternative stories do not appear plausible.

5.1.1. The most plausible story

The preferred explanation is that landlords treat Blacks and non-Blacks with the same observable non-racial characteristics differently in offering and continuing rental contracts, not in the rents they collect from those who have contracts. They offer and continue contracts with some whites who have the same observable non-racial characteristics as some Blacks whom they turn down. We call this behavior quantity discrimination, in contrast with the price discrimination that appears not to be occurring.

We start this section with a note about findings from individual-level data that any explanation must be consistent with.

We continue with a model of a monopolistically competitive housing market and show how higher rents in certain circumstances can be correlated with greater racial disparities in apartment-finding. This is a model of a formal market in the sense that participants use written leases. We then extend this model to encompass informal markets (living with friends or relatives, for instance) as well.

Next, we present evidence from the existing literature that quantity discrimination does occur in private housing markets similar to the ones in which people at risk of homelessness operate. This evidence involves

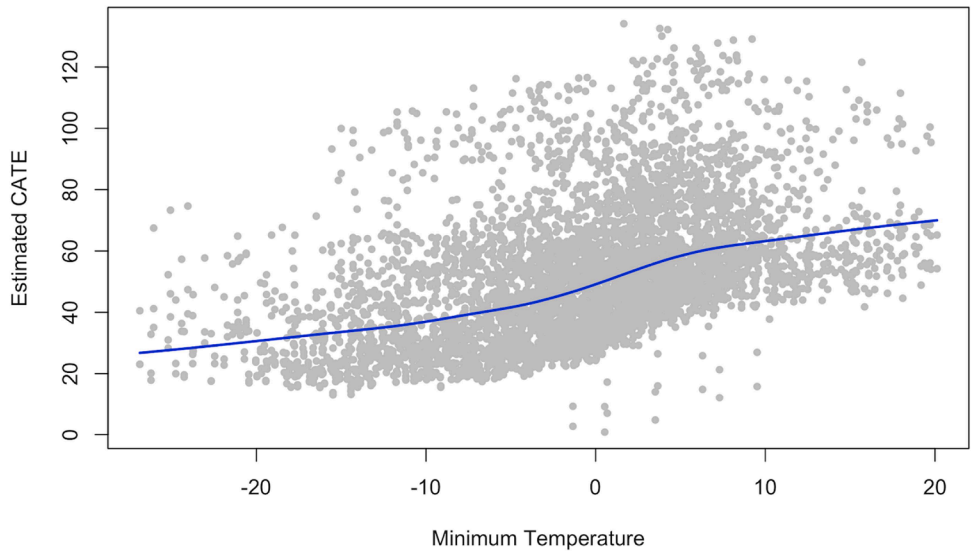


Fig. 9. Treatment effect vs. minimum temperature.

both audits and administrative records. Another subsection describes explicit quantity discrimination on the basis of characteristics correlated with race in public housing markets. The literature provides little insight into the consequences of this explicit quantity discrimination but demonstrates that the behavior we are concerned with occurs in many places.

Finally, we consider some other possible explanations for the correlation between rents and the Black-white and Black-Hispanic homelessness disparity and show that they have serious problems.

5.1.2. What individual-level data say about the black-white homelessness disparity

Meyer et al. (2024) is the best study of the Black-white homelessness disparity with individual-level data. They find that on average Black homeless people have greater resources (of many different kinds) than white homeless people. That is, some Black people are not as successful at avoiding homelessness as white people with the same resources. Both our formal and informal models of the housing market imply this result.

5.1.3. A model of tenant selection and retention stringency

Why should we expect tenant selection and retention stringency to be correlated with measures of housing market tightness? In this section we present an atemporal model to answer this question. Because the model is atemporal, it covers both selection and retention. To save space, we will use words only about selection, but they can be easily translated into language about retention.

We are thinking of two rental housing markets. In the formal market, which is the one we usually study, tenants enter into written contracts with landlords, pay monetary rent every month, have explicit legal rights and can usually exclude the landlord from their possessions and sleeping location, and tenants do not have access to landlords' homes. The informal market is apartment-sharing and couch-surfing. "Tenants" or guests do not have written leases or many legal rights, they may or may not pay rent, and they cannot exclude the person who has legal right to the dwelling (the "landlord" or the host) from their possessions and sleeping location. Nor can the landlord exclude the tenant from their living quarters and possessions as long as the relationship exists.

Consider how formal market landlords will respond to a positive demand shock. If the market is originally loose and landlords have many vacancies, they may simply fill vacancies with minimal changes in rent. In tighter markets, they have two possible responses: raising rents or becoming more selective in whom they permit to rent apartments. Usually, they will do some of both (so changes in rents will be negatively correlated with changes in vacancy rates, ceteris paribus). If some landlords bear animosity toward Blacks or believe, rightly or wrongly, that Blacks are more likely to be troublesome tenants, greater selectivity will disproportionately affect Blacks. Unless the informal housing market is large enough, homelessness will rise more among Blacks than non-Blacks.

What about informal markets? Because of security concerns, a person's alternatives for sharing in the informal market are limited to family and friends. Circles of family and friends are generally largely racially segregated. Legal, *de facto* racial discrimination is therefore usually rife in informal markets. Since non-Blacks in general are wealthier, live in larger homes, and have larger family circles, non-Black people at risk of homelessness may have better options than Black.

Even though formal rents may not be recorded in the informal market, greater demand in the formal market can change the availability of informal apartments. First, greater rents may force some formal market tenants to give up their apartments and enter the informal market. This has two effects on homelessness: it reduces the number of potential hosts in the informal market, and it increases the number of aspiring guests. Second, greater selectivity in the formal market may force some formal market participants to switch to the informal market and this again increases homelessness.

Since greater selectivity in the formal market has a larger effect on Blacks than non-Blacks, it will also have a larger effect in the informal market. Moreover, since the informal market for non-Blacks may have greater slack when rents are low than the informal market for Blacks—more willing potential hosts that no potential guests are asking for help, the non-Black informal market may absorb a larger shock from the formal market than the Black informal market would and so cause less homelessness.

Sometimes analysts portray the housing market discrimination and family assistance explanations for the racial disparity in homelessness as competing (e.g., Meyer et al., 2024; O'Flaherty, 2019). Instead, it appears that they are complementary. Our preferred explanation here incorporates both. The data that we have cannot distinguish between them any more than data on automobile travel can distinguish between the theory that the front wheels are more important and the theory that the back wheels are more important.

5.1.4. Evidence of quantity discrimination from existing literature

A cursory review of the literature might suggest that quantity discrimination in the rental housing market—greater selectivity for Blacks than non-Blacks—is not an important phenomenon. The most

recent large government-sponsored audit study of rental housing market discrimination—the Housing Discrimination Study summarized in Turner et al. (2013) —found that "Minority renters who call to inquire about recently advertised homes or apartments are rarely denied appointments that their white counterparts are able to make" (p. xiii). Blacks are shown fewer units on average than whites, but the difference is only about one unit on average for every 25 visits (p. xiv).

For three reasons, however, all of which Turner et al. (2013) acknowledge, these results may not be dispositive for our purposes. The first reason is that the results are only about the initiation of tenancies, not their continuation: "[D]iscrimination against established tenants … (such as in lease renewal or property maintenance cannot be captured through paired testing because the housing provider already knows the details of consumers' actual characteristics" (p. 3).

Second, audits can reveal only what happens as the parties begin exploring a potential transaction; it cannot reveal whether they actually go through with the transaction: "[T]hird-party testing protocols cannot legitimately involve the formal submission of fraudulent information in a signed rental … application, so it is not possible to capture discrimination that might occur at the final stage of a rental … transaction" (p. 3).

Finally, testers were not representative of people at risk of homelessness: "All testers were assigned financial characteristics that qualified them for the housing units about which they were inquiring," (fn 4, p. xiii) and none presented themselves as having criminal records or histories of drug abuse, for instance. Note that the first two reasons are endemic to audit studies, but the third is not.

Chan and Fan (2023) is an audit study that concentrates on a low-income housing market and finds considerable discrimination against applicants with Black-sounding names. Thus, they study the third HDS weakness. They conduct a correspondence audit on a website, GoSection8, that landlords use to advertise to low-income renters. With a sample of about 1200 inquiries, they found discrimination in both obvious and subtle ways. Inquiries by prospective tenants with Black-sounding names were 17% less likely to be answered than inquiries by prospective tenants with white-sounding names, and when they were answered the answers were less likely to contain the word "please." Discrimination was greater in neighborhoods with better schools and larger white populations.

Ellen et al. (2024) addresses both the second and third weaknesses of Turner et al. (2013): they study actual lease-ups rather than just the beginnings of conversations and they focus on a low-income population. Because they use nationwide administrative data over several years, they have many more observations than audit studies usually do—over 408,000 searchers.

About 61 percent of voucher holders leased up successfully eventually. Blacks were 9% less likely than whites to lease up in the first 60 days, and the average wait of those who were successful was 13 days longer for Blacks. Hispanics were 4 percent less likely to lease up in the first 60 days, and those who eventually succeeded searched on average four more days (Table 3). All these differences were significant at conventional levels and hold many market-level variables constant.

(Part, but not all, of these disparities is due to minorities being more likely to search in neighborhoods with older housing that more often fails to meet HUD structural standards. Households tend to concentrate their searches close to where they were living when they received the voucher, and minorities tended to live near older housing. The entirety of the disparity is not due to taste or statistical discrimination on the part of landlords. For our purposes in this paper, it is immaterial what the source of the contracting disparity is, as we have stated earlier.)

Are disparities greater in tighter housing markets? The obvious approach to answering this question—comparing markets based on rent—will not work with this data because the value of the voucher adjusts to FMR: HUD tries to make every housing market equally tight for voucher holders.

Ellen et al., however, devise a measure that they claim proxies the tightness of the market that voucher holders face. This is rent dispersion (specifically the "90:10 SAFMR ratio[4]").

The claim is that since voucher holders can pay roughly the FMR, their search is easier if more neighborhoods have rent that is close to the 40th or 50th percentile of the rent distribution. Voucher holders cannot afford to pay rents much above these percentiles and apartments that rent for considerably lower rent are not likely to meet HUD structural standards. So, voucher holders face a better housing market when rents are tightly bunched around the 40th and 50th percentile, and a worse housing market when the tails of the rent distribution are heavy.

Ellen et al. find that racial disparities are greater in metropolitan areas where rent dispersion is greater. This holds both for the probability of leasing up and the average search time for successful searchers. If we interpret rent dispersion as a measure of market tightness for voucher holders, this result supports our finding of greater disparities in homelessness in areas with higher rents.

Finally, data also indicate that Blacks in general have more difficulty searching for rental housing than whites do, unconditionally. Zillow Research (2024) has a nationally representative sample of renters who moved in 2021 and found that Black and Hispanic apartment seekers submitted more applications than white or Asian: the median number of applications for Blacks and Hispanics was 3, compared with 2 for whites and Asians, and 38 percent of Blacks and Hispanics filed five or more applications, compared with 33 percent for Asians and 21 percent for whites. The median application fee for Hispanics ($80) and Blacks ($65) was also higher than the median application fee for whites ($50). Blacks and Hispanics were also more likely to pay security deposits than whites (93 percent and 91 percent, respectively, as compared with 85 percent), and to pay larger deposits when they paid them (the Black median security deposit was $700, the Hispanic was $650, and the white was $600). We will see below that unconditionally Blacks pay lower rents—at least among all renters, not necessarily recent movers.

Results on racial differences in the ability to continue in rental relationships are much more limited. Thomas (2017) found that Black households in King County WA were roughly four times as likely as whites to face eviction. Whether this result holds nationally and whether it holds conditional on the usual covariates is unknown. It is also difficult to discern whether Blacks are more likely to experience involuntary exits from tenancy other than eviction. The distribution of tenancies by ongoing duration in 2019 for Blacks was very similar to that for whites (American Housing Survey), but a large proportion of exits from tenancy are voluntary.

Notice that Blacks could still be more likely to enter homelessness than whites even if they were equally likely to have tenancies involuntarily terminated. This could occur if whites were more likely to follow terminations of tenancies with initiations of new tenancies than Blacks.

5.1.5. *Federally subsidized housing*

Access to federal housing subsidies is an important resource for families at risk of homelessness. In many data sets this resource is unobservable, but in some it is observable. This resource may or must be denied to certain people with histories of alcohol abuse, drug abuse, or criminal activity. These denials affect both who can enter these programs and the housing that comes with them, and who can continue in them.

Because some of the personal histories that trigger denial of service are correlated with race, either in fact or in popular stereotype, these policies could increase racial disparities in homelessness. Popular stereotypes matter independent of fact because although these bans on federal assistance to households that include members with certain criminal records are legislatively mandated, they are administered by workers with considerable discretion applying local guidelines that have

[4] SAFMR stands for "small area fair market rent;" roughly speaking, the FMR calculated for a zip code area rather than a metropolitan area.

been developed with considerable discretion (Curtis et al., 2013). Standards of proof are also low: "When making decisions about whether to evict tenants or deny applicants for criminal activity, PHAs [public housing authorities] can use evidence that is significantly weaker than that which would be required to convict in a criminal court. Further, PHAs may evict a household for criminal activity regardless of whether an arrest or conviction has occurred" (Curtis et al., 2013, p.47).

Could PHAs vary how they use their discretion with housing market conditions? There is reason to think (but no evidence yet) that they might tighten standards in a way that would increase disparities when housing markets tighten. Any speculation like this about how they would act has to be based on a theory of what their objectives are. Tighter housing market conditions make their subsidies more valuable and so make it easier for them to fill their buildings or use all their vouchers with less troublesome tenants. If their objective is to collect as much rent as possible with as little trouble as possible, then tighter market conditions would induce them to tighten their interpretations.

The effect of these bans on certain prospective tenants, however, and how they might contribute to racial disparities in homelessness (or reduce crime, even) have been little studied and need more study. Their existence shows that quantity constraints are present in low-income housing markets and probably matter.

5.1.6. Two alternative explanations that do not have a lot of support

Family and friends' financial support. Since white families on average have more income and wealth than Black families, they may be able to provide more money and non-housing resources to their needy members and so keep those members out of homelessness longer when those members lack their own resources. (We have already discussed provision of housing resources in informal markets above and it does not present the problems raised here.) This difference in family background would result in both a lower homelessness rate for whites, and greater average (visible, non-family) resources for Black homeless people. This explanation is thus formally the same as the explanation that posits that whites have more "off-the-books" income than Blacks.

Note that this explanation does not require that all white people have greater hidden resources than all comparable Black people—only that white people who are not homeless with the same visible resources as Black people who are homeless have these hidden resources. White homeless people can have the same hidden resources as Black homeless people do.

The main weakness in this explanation is that it does not imply that the disparity should be greater in tighter housing markets—that tight housing markets somehow make white families relatively more generous. (This is the key difference between financial support and informal housing support.)

Since Hispanic households on average have less wealth that white households—indeed wealth much closer to that of Blacks than of whites—this explanation implies that the disparity between Blacks and Hispanics should be smaller than the disparity between Blacks and whites; increasing rents should cause a smaller increase in the Black-Hispanic disparity than the Black-white disparity; and the Hispanic-white disparity should be positive and increase with rent. The data support none of these three inferences.

Another problem with this explanation is that it requires some unusual assumptions about the joint distribution of hidden and visible resources to reconcile with the finding that white homeless people have a higher mortality rate than Black homeless people, conditional on observed resources (Meyer et al., 2023; Metraux et al., 2016). This mortality rate differential probably indicates that white homeless people do not have greater hidden resources than Black homeless people with the same observed resources. If this is the case, then the hidden resources explanation implies a sudden jump in the white relative hidden resources at the level of observable resources of the white homeless person with greatest observed resources. Essentially, white families determine a particular level of observable resources: they provide relatively great resources to white relatives with observed resources above that level and relatively meager resources to relatives with visible resources below that level. This behavior on the part of white families is not impossible to imagine, but it does seem unusual.

Alternatively, one could reconcile greater non-housing generosity on the part of white families with higher mortality if most or all of those greater resources were devoted to consumption that increased mortality, like tobacco, alcohol, or other drugs. But if those non-financial resources were devoted to these purchases, they would not be reducing the white homelessness rate.

Housing price discrimination against Black people. Black people with the same observed resources as housed non-Black people might be homeless if landlords would charge them higher rents. While we cannot observe the rents that homeless people would pay if they were housed, there is little evidence to indicate that among people similar to homeless people, Blacks pay noticeably higher rents than whites.

Unconditionally, Black renters have lower mean and median rents than white or Hispanic renters. Table 6 is from the American Housing Survey illustrates for 2021.

If we restrict attention to renters in households with income below the poverty level or half the poverty level, the same pattern persists as shown in Table 7,

These tables compare costs of housing, not prices of housing. That is, they ignore comparability in the quality and size of apartments. Some researchers have tried to discern whether Blacks pay more for comparable apartments. The most recent answers are yes, but the premium is small, under 3% (Myers, 2004; Early et al., 2019). Early et al. (2019) found that the premium was close to non-existent in heavily Black neighborhoods but rose close to 3% in heavily white neighborhoods. Myers (2004) has similar point estimates but much larger standard errors, partly because she has a considerably smaller data set. Early et al. (2019) is particularly appropriate for our purposes because its data are mainly low-income households. Ihlanfeldt and Mayock (2009) and Bayer et al. (2017) find similar magnitudes for the Black premium in the owner-occupied housing market.

All papers in this literature include both structural and neighborhood characteristics as controls. Including neighborhood characteristics is appropriate in searching for discrimination, but probably not appropriate in looking for differing incentives to remain housed. Becoming homeless means losing access to a structure, not losing access to a neighborhood. Everyone must be somewhere always; that is a fundamental physical law. If the alternative to being housed is being homeless in the same neighborhood, then the cost of not being housed is loss of structure, not of neighborhood. In that case, the appropriate controls are only structural characteristics, which would probably show that Blacks pay less for comparable structure. If the alternative to being housed is being homeless in some other neighborhood, then the cost of not being housed includes the cost or benefit of relocating to another neighborhood.

Even if Blacks paid substantially more than non-Blacks for comparable apartments, this circumstance alone would not cause the disparity to rise in tighter housing markets. Some other mechanism would be needed.

5.1.7. Segregation

First, consider the effect of Black-white segregation on the Black-white homelessness disparity. Recall that the effect was negative at comparatively low levels of segregation and positive at high levels of

Table 6
Monthly housing cost for all renters, 2021.

	Hispanic	Non-Hispanic white	Non-Hispanic Black
Median housing cost	$1275	$1180	$1033
Mean housing cost	1451	1460	1182

Source: American Housing Survey detail.

Table 7
Monthly housing costs for renters with income below the poverty level or half the poverty level, 2021.

	Hispanic, <50% P.L.	NHW, <50% P.L.	NHB, <50% P.L.	Hispanic, 50–99% P.L	NHW, 50–99% P.L	NHB, 50–99% P.L
Median housing cost	$1016	$965	$818	$976	$792	$735
Mean housing cost	1176	1196	948	1090	971	828

Source: American Housing Survey detail.

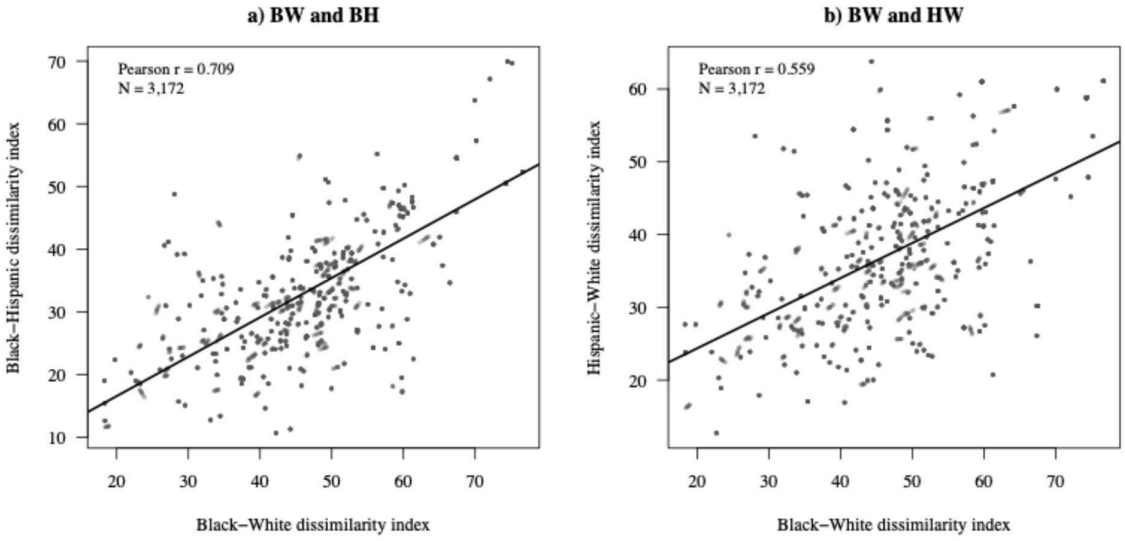


Fig. 10. Scatter-plots of the relationships among dissimilarity indices.

segregation. The explanation may be that two different effects are at work. I will call them the Black personal outcome effect, which raises the disparity, and the political and attitudinal effect, which reduces the disparity.

The Black personal outcome effect is well known to economists. Many studies of segregation find that increasing segregation in a metropolitan area increases personal outcomes that are correlated with homelessness: for instance, Cox et al. (2026), Shihadeh and Flynn (1996), Cutler and Glaeser (1997), Ellen (2000), Card and Rothstein (2007), and Ananat (2011). Almost all these studies find that segregation produces worse personal outcomes for Black people and does not affect personal outcomes for white people to any considerable extent. Thus, the effect of increasing segregation should be increases both in homelessness and the Black-white disparity.

The political and attitudinal explanation emphasizes the effect of segregation on political outcomes and white attitudes. Ananat and Washington (2009) show that segregation reduces "political and civic efficacy" for Blacks. They define political efficacy as "the ability to elect Representatives (in Congress) who vote in a manner that the Black electorate favors." The main reason why this occurs is that "in more segregated areas non-Blacks are more likely to hold negative views of Blacks and of policies that aid Blacks."[5]

For identification reasons, their data do not include the South, but when we run the basic GRF without the South the result on segregation is almost the same.

If their arguments are correct, political efficacy on a state or local level should be affected the same way as political efficacy on the federal level. Since Blacks are more likely to be homeless unconditionally than other groups, the Black electorate probably favors policies that help homeless people relatively more than non-Blacks, and non-Blacks are likely to consider programs that aid homeless people as "policies that aid Blacks." Hence, ceteris paribus, more segregated metropolitan regions would be less generous toward homeless people in general and Black homeless people in particular.

Greater segregation should reduce Black unsheltered homelessness through harsher policing and more aggressive attitudes of the non-Black public. Support services like "midnight runs" and medical outreach might also be less. Some of these characteristics of segregated metropolitan regions may affect only Black people while others may be race-blind in application, and so both Black and non-Black unsheltered homelessness would go down in more segregated metropolitan regions, but Black unsheltered homelessness would go down more.

Similarly, local governments in more segregated metropolitan areas may be less willing to support shelters with their own funds. NIMBYism may be stronger, and governments may be less likely to push back against it. Moreover, Cox et al. (2026) find that local governments in more segregated metropolitan areas raise less revenue and spend a higher proportion of the revenue they do raise on policing. There is less money available for shelters even if governments wanted to spend it. As with unsheltered homelessness, potentially homeless people of all races would be affected, but the Blacks would be affected more.

Our empirical results then suggest that the political and attitudinal effect is usually stronger, especially at low levels of segregation, but the Black personal outcome effect is sometimes stronger at high levels of segregation.

As far as the effect of Black-Hispanic segregation on the Black-Hispanic disparity, I have much less to say. I'm not aware of any study of the effect of Black-Hispanic segregation on anything, much less something correlated with homelessness. A search by Claude Opus 5 on August 13, 2026 did not find anything either. The most promising explanation is that Black-Hispanic segregation is positively correlated with Black-white segregation and Black-white segregation affects mainly Black homelessness. In our data, the correlation between the

[5] The results of Ananat and Washington are causal, and so the anti-Black attitudes could not be causing the variation in segregation that they are studying.

Black-white dissimilarity index and the Black-Hispanic dissimilarity index is 0.71; between the Black-white dissimilarity index and the Hispanic-white dissimilarity index the correlation is 0.56. Fig. 10 shows scatter-plots of these relationships.

6. Conclusion

What explains variation in Black-white disparities in homelessness? The simple answer is rent, vacancy rates, and segregation, with an emphasis on rent. Racial disparities are quite small among unsheltered people in CoCs with modest rent, high vacancies, and moderate segregation. While this paper makes no claims about causality, Maddock et al. (2026) have shown that higher rents cause greater homelessness. Since a disproportionately large part of the rise in homelessness that higher rents cause is among Black people, it seems likely that the rise in Black homelessness that accompanies higher rents is caused by those higher rents. But rigorous work on this question has not been done yet.

Why do the disparities vary this way? For rents and vacancies, the best candidate explanation is quantity discrimination: landlords become more selective in tight housing markets, and their selectivity induces them to become more hesitant about renting to poor Black tenants. The basic structure is the same whether the "landlords"—the people controlling the apartments—are formal third-party owners or family and friends in the informal housing market. Explanations for segregation are more speculative.

The skyscraper picture, which began the paper, showed that the outstanding fact about homelessness in the United States is that it is heavily concentrated among Black people in the high-rent states of California and New York. I've tried to explain why. Other people should start trying too.

AI statement

AI statement: I used Claude Opus to do a literature review on for results on the effects of Black-Hispanic residential segregation on any outcomes. This is detailed in the paper.

CRediT authorship contribution statement

Brendan O'Flaherty: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Acknowledgments

Theo Burden, Chiara Lane Delgado Chanoi, Jinhao Liu, Francesca Malandruccolo, Ren Su, Jinxin Tian and Siqi Yang provided excellent research assistance. Ingrid Gould Ellen, Marah Curtis, Ayse Imrohoroglu, Bruce Meyer, Ann Elizabeth Montgomery, Ebonya Washington, and Angela Wyse provided many helpful comments on earlier drafts; as did the workshop audiences at UC Santa Barbara, the CUNY Graduate Center, the New York Federal Reserve Bank, and the urban homelessness class at CCNY. While writing this paper, I also benefited from the hospitality of TMJ4 Milwaukee.

Appendix 1. Results with regressions for the Black-white and Black-Hispanic comparisons

Results on the correlates of the national disparity. In this section, we present the results of regressions where the dependent variable is the rate of homelessness, and the independent variables are our full list and a constant, once entered simply, and once entered interacted with race (a dummy that equals one when the observation is about Black people).Table A-1 below presents both the unweighted results and the results for observations weighted by population in poverty.

Table A1
Fully interacted regression results for Black-white and Black-Hispanic comparisons.

Variable	BW unweighted	BW weighted	BH unweighted	BH weighted
Race	−72.166[c] (26.444)	−147.328 (95.453)	−59.101 (36.956)	−90.645[b] (35.434)
South	−36.392[b] (17.286)	−12.387[c] (4.439)	−18.587[c] (2.860)	−8.511[b] (3.998)
Race x South	−10.784 (16.144)	−27.588[c] (9.766)	−31.553[c] (5.587)	−29.381[c] (11.190)
Sheltered	0.752 (10.956)	3.145 (8.992)	6.975[b] (2.785)	−5.920 (12.706)
Race x Sheltered	32.979[c] (10.830)	36.629[a] (20.916)	27.179[c] (8.253)	45.838[c] (15.386)
Family	−34.875[c] (4.306)	−23.640[c] (2.649)	9.559[c] (2.893)	16.154[a] (8.600)
Race x Family	25.082[c] (5.127)	29.494[b] (13.416)	−19.351[c] (5.403)	−10.300 (6.732)
Largely rural	−17.998 (11.193)	−9.069[a] (5.486)	−9.218** (4.603)	−0.316 (7.273)
Race x Largely rural	13.544 (10.943)	17.236[b] (8.001)	10.598 (9.926)	7.825 (10.180)
Largely suburban	−9.080 (8.769)	−20.776[c] (5.532)	−14.447[c] (4.891)	−16.581[b] (8.246)
Race x Largely suburban	−9.348 (11.573)	−20.299[a] (12.255)	−14.648 (9.984)	−23.036[b] (11.060)
Major city	30.345[c] (10.829)	23.820[b] (9.964)	0.347 (6.524)	5.852 (8.706)
Race x Major city	−21.742[a] (12.955)	−8.615 (12.584)	0.915 (13.165)	11.782 (11.092)
Minimum temperature	2.111[a] (1.102)	0.715 (0.440)	−0.430 (0.324)	−1.659[b] (0.695)
Race x Minimum temperature	−2.413[b] (0.993)	−1.707[c] (0.659)	0.722[a] (0.395)	0.480 (0.856)
Precipitation	−1.298[b] (0.502)	−0.673[a] (0.373)	−0.082 (0.152)	0.516 (0.379)
Race x Precipitation	0.448 (0.491)	0.334 (0.575)	−1.000[c] (0.372)	−0.760 (0.705)
Fair Market Rent	0.097[c] (0.015)	0.123[c] (0.016)	0.076[c] (0.011)	0.089[c] (0.017)
Race x Fair Market Rent	0.129[c] (0.033)	0.139[c] (0.032)	0.151[c] (0.030)	0.175[c] (0.028)
Vacancy rate	0.972 (1.401)	−0.186 (0.328)	1.054[c] (0.287)	1.072 (0.759)
Race x Vacancy rate	−0.669 (1.325)	−0.139 (0.959)	−0.802 (0.561)	−1.157 (1.061)
Segregation index	−1.285[c] (0.280)	−0.760[c] (0.207)	−0.735[c] (0.161)	−0.134 (0.356)
Race x Segregation index	−0.992[b] (0.493)	0.365 (0.948)	−0.378 (0.383)	−0.528 (0.399)
Constant	57.922[c] (18.552)	−1.610 (15.104)	−21.558* (12.496)	−65.197 (41.014)

(*continued on next page*)

Table A1 (*continued*)

Variable	BW unweighted	BW weighted	BH unweighted	BH weighted
Observations	16,323	16,323	16,177	16,177
Number of CoC clusters	344	344	344	344
R-squared	0.193	0.398	0.243	0.434
Adjusted R-squared	0.192	0.396	0.242	0.433
Year fixed effects	Yes	Yes	Yes	Yes
CoC-clustered standard errors	Yes	Yes	Yes	Yes
Weights	No	Race-specific poverty population	No	Race-specific poverty population

Notes: Cells report coefficients with CoC-clustered standard errors in parentheses. All models include year fixed effects. Weighted regressions use analytic weights based on the race-specific population below poverty.

[a] $p < 0.10$.

[b] $p < 0.05$.

[c] $p < 0.01$.

The coefficients on the independent variables other than race in Table A-1 are the coefficients that would appear in a regression where the data set was restricted to white observations. The coefficients on the independent variables interacted with race are the coefficients that would appear in a regression on the full data set where the dependent variable for each observation was the racial disparity for that observation. Accordingly, these coefficients are our chief concern. If the coefficients on each independent variable separately were added to the coefficient on the interacted version of that variable, the resulting sums would be the coefficients on a regression where the dependent variable was the homelessness rate and the data set was restricted to Black observations.

Several important results derive from this table. First, the coefficient on race in the interacted regression is negative. This indicates that if all the relationships were in fact linear, the regression would more than "explain" the disparity. Of course, there is no reason to believe that all relationships are linear.

Second, the interacted coefficient on fair market rent in the BW unweighted regression is 0.129. Start with an observation where all independent variables are at their mean value, and so the disparity is 40.2, the mean disparity in the regression data set. Holding every other independent variable the same, reduce rent for this observation by one standard deviation, or $381.4 in the regression data set. This reduces the expected disparity by $381.4 \times 0.129 = 49.2$, which is greater than the observed disparity. The disparity is predicted to go zero when rent is a little bit more than one standard deviation below the mean.

For the BH comparison, the relevant coefficient on the interacted term in the equivalent of column 3 is .151 and the mean disparity is 69.7. Reducing rent by one standard deviation reduces the disparity by $.151 \times 382.4 = 57.8$, somewhat less than the mean disparity. Reducing rent by 1.21 standard deviation below the mean makes the predicted disparity go to zero. Rent clearly has a very strong influence on the disparity in this regression, just as it did in the GRF.

Results on segregation are more mixed. The coefficient on dissimilarity interacted with race is negative and significant in the BW unweighted regression, but is not significant in any other regression (and the coefficient is positive albeit insignificant in the weighted BW equation). Considering the non-monotonic scatter-plots for the GRF results, these differing regression

The disparity falls with greater vacancy rates, even holding rent constant, in accordance with the theory in this paper, but are not significantly different from zero.

Single-equation and Blinder-Oaxaca analysis: Why is the average racial disparity in California and New York so much bigger than the average racial disparity in the rest of the country?

We can also use regressions to try to explain the difference in average racial disparity between California and New York (CANY) and the rest of the country (RoC).

To put the regression data set into a form suitable for Blinder-Oaxaca analysis, we follow essentially the same steps we followed for the GRF data set in the body of the paper. First we transform the dependent variable into a racial disparity instead of a race-specific homelessness rate and in doing so cut the number of observations in half. Second, we divide the data set into observations from CANY (the treatment group) and observations from RoC (the control group). Third, we drop the South dummy as an independent variable. Finally, we construct the overlap and drop all observations outside the overlap.

To construct the overlap, we use a probit equation to estimate propensity scores and retain the observations with propensity scores in the range [0.10, 0.90].

The regression data set has 8535 observations. Of these 4619 retained in the overlap and 3916 were outside the overlap. For CANY, 986 observations were in the overlap and 3633 observations were in the overlap for rest of the country.

With this data set in hand, we perform two analyses that are common in regression studies of discrimination: a single-equation regression and a Blinder-Oaxaca decomposition with two equations. In the single-equation regression, we regress disparities on the full set of covariates (minus South) and a dummy variable for CANY. The coefficient for CANY is −8.52 (23.21) for BW and 50.41 (13.35) for BH. As in the GRF analysis, we reject the hypothesis that the CANY coefficient is zero for BH and not for BW. The point estimates are within one standard error of the GRF point estimates for ATT and ATC.

We also perform a Blinder-Oaxaca analysis to explain the difference in (unweighted) average racial disparity between CANY and RoC. (Geographic regions play the same role in this analysis as races or genders play in the common Blinder-Oaxaca exercises.) We run two (unweighted) ordinary-least-squares regressions with racial disparity as the dependent variables and the usual independent variables (including a constant and omitting South). Algebraically, these are

$$d_i = \beta^{CANY} x_i + \varepsilon_i, \quad \mathrm{i} \in CANY$$

$$d_i = \beta^{RoC} x_i + \varepsilon_i, \quad i \in RoC.$$

Here i denotes an observation, d_i is the racial disparity for that observation, β^g is a row vector of coefficients for geographic region g, x_i is a column vector of independent variables for that observation, and ε_i is an error term.

Let $\overline{d}^g$ denote the mean disparity in region g and let $\overline{x}^g$ denote the column vector of means of independent variables in region g. By the properties of OLS,

$$\overline{d}^g = \beta^g \overline{x}^g$$

Consider the RoC. At mean RoC independent variables, CANY has an expected disparity of $\beta^{CANY}\overline{x}^{RoC}$ while the expected value of racial disparity is

$$\overline{d}^{RoC} = \beta^{RoC}\overline{x}^{RoC}$$

So the expected difference at this point is

$$\beta^{CANY}\overline{x}^{RoC} - \overline{d}^{RoC} = \beta^{CANY}\overline{x}^{RoC} - \beta^{RoC}\overline{x}^{RoC} = \left(\beta^{CANY} - \beta^{RoC}\right)\overline{x}^{RoC}.$$

This expression is the analog of the ATC in the GRF and is one measure of discrimination ("the residual") when Blinder-Oaxaca methods are used to study race or gender.

Similarly, if we start with mean CANY independent variables, we obtain the analog of the ATT:

$$\overline{d}^{CANY} - \beta^{RoC}\overline{x}^{CANY} = \beta^{CANY}\overline{x}^{CANY} - \beta^{RoC}\overline{x}^{CANY} = \left(\beta^{CANY} - \beta^{RoC}\right)\overline{x}^{CANY}.$$

These values (ATC and ATT) are shown in Table A-2

Table A2
Blinder-oaxaca estimates of the effect of CANY status on racial disparities.

	ATT	ATC
Black-white	−2.7	−19.6
Black-Hispanic	55.9	46.6

The point estimates are within one standard error of the corresponding GRF estimates. Both types of regression results are consistent with the GRF results.

Appendix 2. Results for the Hispanic-white comparison

In the unweighted GRF, the most important variables are family status (splits in 32% of the opportunities), minimum temperature (24%), and rent (12%). In the weighted GRF, the most important variables are quite different: segregation (29%), family status (22%) and major city (12%).

In the unweighted GRF, the disparity increases strongly with family status. No families were observed in the lowest CATE quintile, but 99% of the observations in the highest CATE quintile were families. Poor Hispanic families are much more likely to be homeless than poor whitefamilies, while poor Hispanic individuals are much less likely to be homeless than poor white individuals.

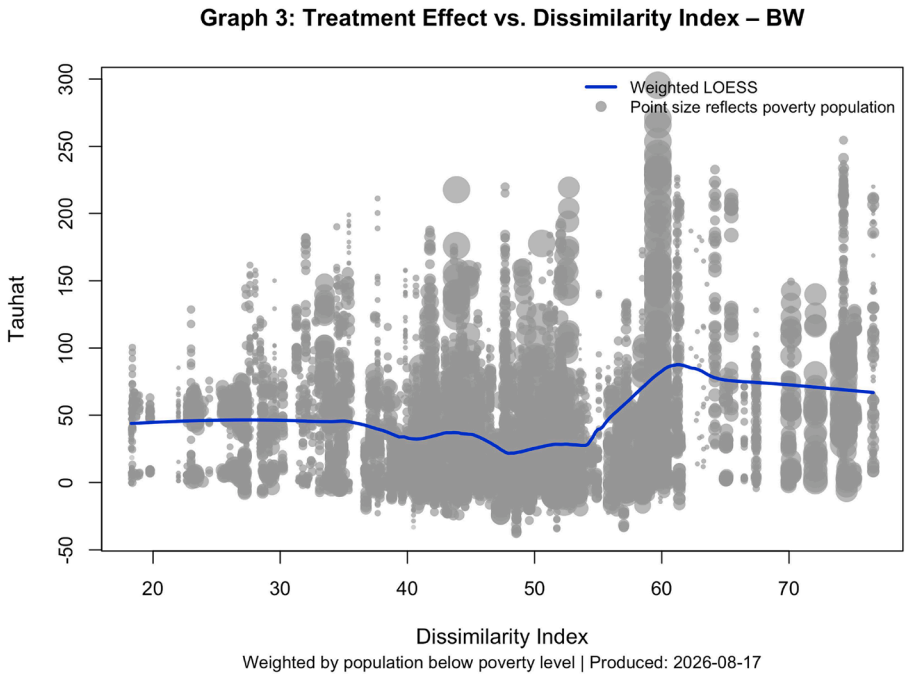
The disparity decreases as temperature rises. Rising temperatures are associated with greater increases in white homelessness than in Hispanic homelessness. The disparity also decreases as rents rise. Rising rents are associated with greater increases in white homelessness than in Hispanic homelessness—consistent with the faster growth of the Black-Hispanic disparity with rent than the Black-white disparity. Figures A-1 and A-2 illustrate these relationships.

In the fully-interacted unweighted regression, the coefficient on family interacted with race is large, negative, and very significant: −44.4. Other variables with significant negative coefficients on the interaction with race are largely suburban, major city, minimum temperature and segregation. These coefficients imply white homelessness increasing faster than Hispanic when the covariate increases, or Hispanic homelessness decreasing slower than white. The coefficient on the interaction of race and precipitation is positive and significant. The interaction terms with rent and vacancy rates are insignificant.

Appendix 3. Weighted GRFs for the Black-white and Black-Hispanic comparisons

With weighted GRFs, for both BW and BH, the most important three independent variables are rent, vacancy rate, and segregation, with rent as the most important. Thus, weighting does not seriously change the order of importance. Importance, though, is considerably less concentrated: the proportion of times that rent is split on falls, and the number of covariates on which splits occur at least 1% of the time grows considerably. Table A-3 shows the importance of all the variables on which splits occur at least 1% of the time.

The weighted scatter-plots for rent and vacancy rates are similar to the unweighted scatter-plots. The weighted scatter-plots for segregation are different, however. The LOWESS line rises sharply in the middle of the range, and then falls sharply at slightly higher values. Fig. A-3 illustrates. This suggests that the negative effect of segregation that we found in the unweighted data may not be robust.



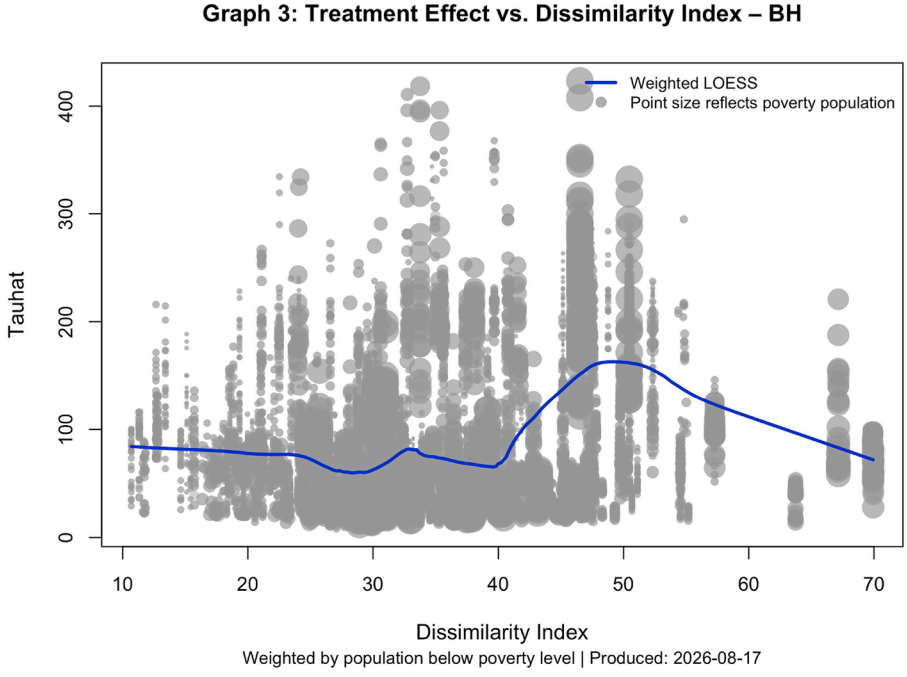


Fig. A3. Treatment effect vs. dissimilarity index in weighted BW and BH GRFs.

Table A3
Variable importance in the weighted BW and BH GRFs (greater than 1%).

Black–white comparison			Black–Hispanic comparison		
Rank	**Variable**	**Probability of split**	**Rank**	**Variable**	**Probability of split**
1	Rent	23.4%	1	Rent	32.2%
2	Dissimilarity index (BW)	16.3%	2	Vacancy rate	19.1%
3	Vacancy rate	14.9%	3	Dissimilarity index (BH)	13.2%
4	Family	8.18%	4	South	6.58%

(*continued on next page*)

Table A3 (*continued*)

Black–white comparison			Black–Hispanic comparison		
Rank	**Variable**	**Probability of split**	**Rank**	**Variable**	**Probability of split**
5	South	7.32%	5	Minimum temperature	6.27%
6	Minimum temperature	7.31%	6	Sheltered	4.93%
7	Sheltered	7.25%	7	Major city	4.01%
8	Major city	4.11%	8	Precipitation	3.12%
9	Precipitation	3.37%	9	Other largely urban	2.63%
10	Other largely urban	3.22%	10	Largely rural	2.35%
11	Largely suburban	1.48%	11	Family	2.02%
			12	Largely suburban	1.42%

Data availability

Data will be made available on request.

References

Ananat, E.O., 2011. The wrong side of the track(s): estimating the causal effects of racial segregation on city outcomes. Am. Econ. J. Appl. Econ. 3 (1), 34–66.

Ananat, E.O., Washington, E., 2009. Segregation and black political efficacy. J. Public Econ. 93 (5–6), 807–822.

Athey, S., Tibshirani, J., Wager, S., 2019. Generalized random forests. Ann. Stat. 47 (2), 1148–1178.

Athey, S., Wager, S., 2019. Estimating treatment effects with causal forests: an application. Obs. Stud. 5 (2), 37–51.

Bayer, P., Casey, M., Ferreira, F., McMillan, R., 2017. Racial and ethnic price differentials in the housing market. J. Urban. Econ. 102, 91–105.

Brown University, 2021a. Diversity Project. "U.S. Segregation data." https://s4.ad.brown.edu/projects/diversity/Data/data.htm.

Byrne, T.. 2021b. "HUD CoC Geography Crosswalk." GitHub Repository. https://github.com/tomhbyrne/HUD-CoC-Geography-Crosswalk/blob/master/output/tractoc_match.csv.

Card, D., Rothstein, J., 2007. Racial segregation and the black-white test score gap. J. Public Econ. 91 (11–12), 2158–2184.

Chan, E.W., Fan, Y., 2023. Housing discrimination in the low-income market: evidence from a correspondence audit. J. Hous. Econ. 59, 101889.

Cleveland, W.S., 1979. Robust locally weighted regression and smoothing scatterplots. J. Am. Stat. Assoc. 74 (368), 829–836.

Cox, R., Cunningham, J.P., Ortega, A., Whaley, K., 2026. Black lives: the high cost of segregation. Am. Econ. J. Econ. Policy 18 (1), 1–34.

Crump, R.K., Hotz, V.J., Imbens, G.W., Mitnik, O.A., 2009. Dealing with limited overlap in estimation of average treatment effects. Biometrika 96 (1), 187–199.

Curtis, M.A., Garlington, S., Schottenfeld, L.S., 2013. Alcohol, drug, and criminal history restrictions in public housing. Cityscape. J. Policy Dev. Res. 15 (3), 37–52.

Cutler, D.M., Glaeser, E.L., 1997. Are ghettos good or bad? Q. J. Econ. 112 (3), 827–872.

Data USA. 2022. "Map Tool." https://datausa.io/map?measure=2fQpSH&groups5B05D=ZFt1KK7C0&time=2022.

Early, D.W., Carrillo, P.E., Olsen, E.O., 2019. Racial rent differences in U.S. Housing markets: evidence from the housing voucher program. J. Reg. Sci. 59 (4), 669–700.

Ellen, I.G, 2000. Is segregation bad for your health? The case of low birth weight. Brookings-Wharton Papers on Urban Affairs, pp. 203–229.

Ellen, I.G, O'Regan, K., Strochak, S., 2024. Race, space, and take-up: explaining housing voucher lease-up rates. J. Hous. Econ. 63, 101980.

Ihlanfeldt, K., Mayock, T., 2009. Housing price differentials in U.S. Cities: household and neighborhood racial effects. J. Urban. Econ. 66, 125–140.

Maddock, L., Pena, A.A, Petach, L., 2026. Housing affordability and homelessness: evidence from commuting zoners. Economic Development Quarterly (forthcoming).

Mendez Jr., G.A, 1983. The Role of Race and Ethnicity in the Incidence of Police Use of Deadly Force. National Urban League National Criminal Justce Reference Service document, Mimeo, New York, 97379.

Metraux, S., Manjelievskaia, J., Treglia, D., Hoffman, R., Culhane, D.P., Ku, B.S., 2016. Posthumously assessing a homeless population: services use and characteristics. Psychiatr. Serv. 67 (12), 1334–1339.

Meyer, B.D., Wyse A., and Logani I. 2023. Life and death at the margins of society: the mortality of US homeless population." NBER working paper 31843.

Meyer, B.D., Wyse A., Meyer G., Grunwaldt A., and Wu D. 2024. "Homelessness and the persistence of deprivation: income, employment, and safety net participation." NBER working paper 32323.

Meyer, B.D., Wyse A. and Williams D. 2025. "Asylum seekers and the rise in homelessness." NBER Working Paper 33655.

Montiel Olea, J., O'Flaherty B., and Sethi R. 2023. "Police homicide and racial injustice." Working Paper.

Myers, C.K., 2004. Discrimination and neighborhood effects: understanding racial differentials in U.S. Housing prices. J. Urban. Econ. 56, 279–302.

Naf, J. 2023. "Random forests and missing values." Accessed 19 March 2026 at https://medium.com/data-science/random-forests-and-missing-values-3daaea103db0.

O'Flaherty, B., 2019. Homelessness research: a guide for economists (and Friends). J. Hous. Econ.

Richard, M.K., 2025. Homelessness and race: the impact of structural conditions on black, white, and Latine Homelessness. Soc. Probl.

Shihadeh, E.S., Flynn, N., 1996. Segregation and crime: the effect of black social isolation on rates of black urban violence. Soc. Forces 74 (4), 1325.

Tang, F., Ishwaran, H., 2017. Random forest missing data algorithms. Stat. Anal. Data Min. 10 (6), 363–377.

Thomas, T, 2017. Forced Out: Race, Market, and Neighborhood Dynamics of Evictions. Thesis. University of Washington, Seattle, WA. https://https://lib.uw.edu/research-services/publishing/.

Turner, M.A, Rob Santos, Levy, D.K., Doug, W., Claudia, A., Pitingolo, R., the Urban Institute, 2013. Housing Discrimination Against Racial and Ethnic Minorities 2012. Office of Policy Development and Research U.S. Department of Housing and Urban Development, Washington, DC.

Twala, B.E.T.H., Jones, M.C., Hand, D., 2008. Good methods for coping with missing data in decision trees. Pattern. Recognit. Lett. 29, 950–956.

Zillow Research. n.d. Renters: Results from Zillow Consumer Housing Trends Report 2021. Accessed 10 August 2024. https://www.zillow.com/multifamily-knowledge-center/renter-insights/rentals-consumer-housing-trends-report-2021/.