# Racial Differences in the Total Rate of Return on Owner-Occupied Housing


Rebecca Diamond — Stanford GSB and NBER

William Diamond — Wharton


March 22, 2024


**Abstract**

We quantify racial differences in the total rate of return on a representative sample of the owner-occupied housing stock from 1975-2021. The total rate of return of buying a house equals the price appreciation plus the rental value of its housing services, minus taxes, maintenance, and home improvement. To measure the total return, we develop a new method to estimate the rental value of each owner-occupied house, using houses that switch between the rental and owner-occupied market. We then use this method to predict the rental value of the entire owner-occupied housing stock and find this prediction out-performs standard hedonic techniques. We document across multiple datasets that Black homeowners earn a 1.5 percentage points higher rental yield on housing than white homeowners, consistent with our method's estimates. This gap largely explains why minority homeowners earn 1.0 percentage points higher total returns on housing. Minority homeowners' total returns are also more volatile and sensitive to the business cycle. These racial differences can be fully explained by other observables, with household income and education differences playing the largest role. Our findings are broadly consistent with a model with a more severe credit constraint for minorities, which bids up rents, lowers house prices, and makes house prices sensitive to credit supply.



*First draft: November 18, 2023. We thank our research assistants Alex Almeida, Sam Asher, Marie Kaul and Weldon Lin for outstanding work. Rebecca Diamond acknowledges support from the Stanford Graduate School of Business, the National Science Foundation (CAREER Grant 1848036), and the Sloan Foundation.

One of the largest racial disparities in the United States is the racial wealth gap. In 2019, Blacks had roughly 1/6 the wealth of Whites, while earning 1/2 as much income [Derenoncourt et al., 2023]. A common policy proposed to narrow this gap has been to promote minority homeownership. The largest single component of most household's wealth is their home [Campbell, 2006]. Racial minorities have a disproportionate share of their wealth (59% for Black vs 38% for White ), so the return on housing places a key role in the evolution of the wealth gap [Derenoncourt et al., 2023]. In this paper, we quantify racial differences in the total rate of return earned by homeowners. We find that Black and Hispanic homeowners earn higher but more volatile returns on housing than White homeowners, suggesting that minority homeownership may narrow the wealth gap but also comes with greater financial risk.

Like most assets, the return on owning a house includes both a dividend payment (rental value of housing services net of costs) and a capital gain (house price appreciation). We develop new methods to infer the rental value of the housing services provided by owner occupied housing. Standard methods project a house's rent onto observable characteristics and use this to infer the rental value of observably similar homes. However, owner-occupied and rented houses can differ in their unobservable quality, leading to bias in this imputation. We present a conceptual framework that relates the rent $r$ of a home to its sales price $p$ and observable characteristics $X$. If unobservable home quality is reflected in both rent and prices $E(r|X,p)$ should be increasing in the price $p$. As a result an observable-based rent imputation, $E(r|X)$, should overestimate the rent $r$ when the price $p$ is low and underestimate $r$ when $p$ is high $p$.

Using data on a panel of homes from the American Housing Survey (AHS), we observe a subset of homes switching between the rented and owner-occupied segments of the housing market. This allows us to see the rental value and sale price of the same home and estimate an empirical relationship between rents and house prices. We therefore use the house's resale price as an additional predictor of its rental value beyond the information contained in the house's observable characteristics. We develop a method that uses the information in house prices to accurately predict the rents of both low-value and high-value homes. In contrast, we show that standard methods overestimate the rent of low-value homes and underestimate the rent of high-value homes in the AHS.

We apply our method to measure racial differences in the total rate of return on owner-occupied housing. Existing literature on racial and gender disparities in housing returns [Kermani and Wong, 2022, Wolff, 2022, Goldsmith-Pinkham and Shue, 2023] measures the price appreciation of individual homes accurately and considers property taxes [Avenancio-Leon and Howard, 2022] but either ignores the rental value of housing or imputes it at an aggregated level. As we document both in raw data from the Consumer Expenditure Survey and in our imputed rent for the AHS, Black and Hispanic homeowners have rental yields that are 1.5% to 2% higher than White homeowners. These rental yield disparities are primarily due to low price homes having higher rental yields, and they are not entirely accounted for by local geography. As a result, is is crucial to measure rental yields and other costs at the level of an individual home to accurately quantify racial differences in the total return on housing. Our result that the total return on housing is higher for minority homeowners is new in the literature.

We find in our benchmark result with AHS data that the average log return on owner-occupied housing is 8.1 percent for Black homeowners, 8.5 percent for Hispanic homeowners, and 7.0 percent for White homeowners from 1975 to 2021. Even though Black and White homeowners see their house prices grow at similar average rates, the overall rate of return for Black homeowners is higher due to a 1 percent higher dividend yield. For Hispanic homeowners, their high rate of return depends both on a 0.7 percentage point higher dividend yield and a 1.1% higher price appreciation rate than earned by White homeowners. Our average rate of return is similar to the housing return in Jorda et al. [2019]. Our results on racial return disparities at an individual level are similar to region-level evidence in Demers and Eisfeldt [2001], who shows that measuring rental yields across cities is crucial to explain spatial variation in housing returns.

We use our rich microdata to analyze the determinants of these rate of return differences at the household level. We regress the rate of return on each house on the homeowner's race together with a growing list of control variables. After controlling for homeowner education and income, we are able to fully explain the average racial return differences with observable characteristics. Using a Gelbach [2016] decomposition, we find that income/education is the most useful explainer of both the Black-White total return gap and Hispanic-White total return gap. Our evidence is consistent with lower income homeowners buying lower value homes, which in turn have higher average rates of return due to their higher rental yields.

A second key difference in the total return on housing for Black and Hispanic homeowners relative to White homeowners is greater volatility and greater sensitivity to the business cycle. We analyze these return differences both for a portfolio composed of the entire housing wealth of each group as well as at an individual homeowner level. For the group-level portfolios, we find that Hispanic homeowners have the highest total return of 5.57 % but also the greatest losses during recessions. In contrast, the aggregate portfolio of composed of all Black owner-occupied homes has a 4.31 % average return (near the White average return of 4.3 %) and similar volatility and cyclicality as that of White homeowners. At the level of an individual homeowner, however, we find that both Black and Hispanic homeowners have particularly volatile and cyclical returns. This suggests that Black homeowners with the lowest price homes are exposed to the most risk, which is not reflected in the return on the group-level portfolios that primarily reflect the returns on more expensive homes.

Unlike the disparities in average returns, we find that observable differences between racial groups do not entirely explain the excess exposure of minority housing returns to the business cycle. Controlling for income/education largely explains the White-Hispanic difference in return cyclicality, but observables poorly account for the more cyclical returns experienced by the average Black homeowner. This is broadly consistent with the finding [Kermani and Wong, 2022] that the risk in minority housing returns is largely due to distressed sales during downturns, based on more severe financial constraints [Gupta et al., 2023, Bhutta and Hizmo, 2021] for Black and Hispanic homeowners.

We complement our AHS results by analyzing the total return on housing in the larger Corelogic database. Despite a shorter time series, Corelogic provides a much larger sample size of housing transactions. We impute rental yields in Corelogic based on those in our AHS data and similarly find higher rental yields for Black and Hispanic homeowners. This again results in higher total returns for Black and Hispanic homeowners in our benchmark specification. Controlling for zip codes explains the vast majority of the return differential we document [Higgins, 2023], and controlling for a house fixed effect actually results in a higher rate of return for White homeowners of roughly .5 %. We argue that this may be because White homeowners are more likely to buy a given house when Black and Hispanic homeowners are excluded by tight credit. Periods of tight credit are in turn times when rates of return are the highest.

Similar to Kermani and Wong [2022], we find in our Corelogic data that Hispanic and especially Black homeowners have the largest downside risk in the return on housing. If we accounted only for the changes in house prices while ignoring rental yields, we would confirm the finding in Kermani and Wong [2022] that Black homeowners earn the lowest rates of return. However, the lower expected price appreciation earned by Black homeowners is either mostly or entirely offset by the their higher rental yields relative to White homeowners.

Our paper contributes to the literature on disparities in housing returns by the use of nationally representative, household-level data that allows us to measure the key determinants of housing returns. Much of the previous literature on racial disparities in housing returns uses geographically aggregated data, with Kermani and Wong [2022] as a notable exception. Our AHS data has costs and benefits compared to the large transaction-level data used in Kermani and Wong [2022]. Their baseline analysis uses realized housing transactions, which more accurately measures the returns on transacted homes but requires a correction for the selection bias of choosing to sell a house. Our data provides a housing value even if there is not a transaction, but may miss some of the financial distress they observe in sales prices. Another benefit of the AHS is that we observe rents, maintenance, and property taxes at a property level, as well as household characteristics for studying the determinants of returns. As work in progress, we are working to combine transaction-level Corelogic data (including house level rental yields from the MLS) with the AHS to improve our rate of return estimates.

In total, our paper demonstrates that it is crucial to account for the total return on housing, not just changes in house prices, when measuring racial disparities in housing returns. Our new method for imputing the rental value of owner-occupied housing may therefore be a broadly useful tool for this literature. Because lower value homes tend to have the highest rental yields, any method of measuring rates of return that ignores rents will understate the returns on low value homes and overstate the returns on high value homes. Because the average value of homeowners' residences varies by racial group, an accurate rental yield calculation is crucial for understanding racial disparities in the rate of return on housing. Together with other evidence on racial disparities in liquidity constraints [Ganong et al., 2024], home-ownership Charles and Hurst [2002], and credit market outcomes [Indarte et al., 2023], an accurate total rate of return measure for housing is key for understanding the future of the racial wealth gap.

The rest of the paper is organized as follows. Section 1 defines the total return on housing and presents a method for imputing the rental value of owner-occupied housing in subsection 1.1. Section

2 demonstrates the success of our rent imputation method for both low-value and high-value homes, unlike traditional hedonic methods. Section 3 then provides direct evidence of differing rental yields on housing across racial groups in raw CEX and AHS data and shows that our imputation method captures the key features of the raw data. Section 4 then presents our baseline results and examines the ability of observable characteristics to account for racial disparities in the total rate of return on housing. Finally, section 5 confirms the importance of accounting for rental yields when measuring racial disparities in the total return on housing in Corelogic.

# 1 Measuring the Return on Housing

Unlike financial assets such as stocks or bonds, housing is both an investment good and a consumption good. Because a homeowner can live in their house, investing in housing generates a flow of consumption services as well as a possible increase in the value of the investment. To correctly measure the return on housing, it is therefore crucial to quantify the market value of the consumption services a house generates. If a house is worth $P_t$ at time t and $P_{t+1}$ at time t+1, we measure the log return on housing as

$$log(P_{t+1} + r_{t+1} - tax_{t+1} - maint_{t+1}) - log(P_t). \tag{1}$$

In this expression, $r_{t+1}$ is the rental value of getting to live in the house at time t+1, $tax_{t+1}$ is property taxes paid at time t+1, and $maint_{t+1}$ is maintenance expenditures at time t+1. Taxes and maintenance (for which we include both routine maintenance and home improvement) are cashflows that can directly be observed in our AHS data. However, the rent $r_{t+1}$ for an owner-occupied house is never explicitly observed. As a result, much of the literature on inequality in returns has simply used the price appreciation $\frac{P_{t+1}}{P_t}$ to measure housing returns. To include all of the relevant cashflows to measure housing returns, we develop and apply a procedure to impute the rental value of owner-occupied housing.

## 1.1 Rent Imputation Procedure

One of the most important components of the return on investing in housing is the rental income that a house generates. For a landlord, this is a cash flow paid explicitly by the tenant to the landlord. For a homeowner, the "owner's equivalent rent" measures the consumption value of getting to live in that home for a period of time. Because homeowners do not explicitly pay or receive rent, it is necessary to impute this owner's equivalent rent to accurately measure their investment returns.

The rent $r_h$ of a house $h$ should be determined entirely by the house's characteristics $r_h = r(X_h, X_h^u)$ [Bajari and Benkard, 2005, Bajari et al., 2005], with more desirable houses earning higher rent. However, the characteristics of a house that impact its rent include both observable variables $X_h$ such as square footage and location as well as unobservable variables $X_h^u$, such as the subjective quality of the house's design. As a result, the expectation of a house's rent given its observable characteristics $E(r_{h,t}|X_{h,t})$ faces an omitted variables problem. Houses with rent above $E(r_{h,t}|X_{h,t})$ are likely to have better unobservable characteristics than houses with rent below.

Our solution to this omitted variables problem relies on observing the price $p_h$ of a house. Suppose there are two functions $r_h = r(X_h, X_h^u)$ and $p_h = p(X_h, X_h^u)$ that respectively give a house's rent and price given all of its observable and unobservable characteristics. Without loss of generality, we can define two scalars "rent quality" $r_{h,u}$ and "price quality" $p_{h,u}$ so that rents and prices can be written as functions

$$r_h = r^*(X_h, r_{h,u}) \tag{2}$$

$$p_h = p^*(X_h, p_{h,u}). \tag{3}$$

To most accurately infer the rent $r_h$ of an owner occupied house, we condition both on the house's price and its observable characteristics to obtain

$$E(r_h|X_h, p_h) = E(r_h|X_h, p_{h,u}). \tag{4}$$

By conditioning on a house's price in addition to observable characteristics, we can include its price quality to improve the accuracy of our predicted rent. If a house's rent and sales price are determined by similar unobservable characteristics, conditioning on $p_{h,u}$ should significantly improve the accuracy of our imputed rent.

If the distribution of unobserved rent quality $r_{h,u}$ conditional on a level of unobserved price quality $p_{h,u}$ is increasing[1] in $p_{h,u}$, a clear bias appears in any procedure that only uses observables $X_h$ to predict rent. For houses with sufficiently low prices, their rent will generally be below $E(r_h|X_h)$, so the procedure overestimates the house's rent. Similarly, for houses with high enough prices, $E(r_h|X_h)$ is an underestimate of the houses rent.

Because minority homeowners tend to own lower price houses than White homeowners, this argument suggests that using only observables to predict a house's rent would induce bias into our calculation of race-specific rates of return on housing. In particular, the rental yield earned by minority homeowners would be artificially inflated compared to the rental yield earned by White homeowners. As we show empirically in section 2, the procedure we develop next accurately measures average rents for both low and high price homes and is therefore ideal for our application to comparing housing rates of return across racial groups.

**Inferring rental yields using switchers in the AHS**

Our baseline rent imputation procedure requires data on a panel of houses, as is provided by the AHS. Each house $h$ at time $t$ is either vacant, rented, or owner-occupied. For rented houses, we observe the rent $r_{h,t}$. For owned houses, we observed the price $p_{h,t}$. The key difficulty we face is that in each period at most one of $r_{h,t}$ and $p_{h,t}$ is observed. In addition, we see a vector of house characteristics $X_{h,t}$. Our goal is to produce an estimate $\hat{r}(p_{h,t}, X_{h,t})$ of the rent that could have been obtained for owner-occupied house $h$ at time $t$. We implement this procedure separately by racial group to account for unobservable differences in the housing stock owned by different racial groups.

[1] That is, for any number n, $Pr(r_{h,u} > n|p_{h,u})$ is increasing in $p_{h,u}$.

Our rent imputation procedure relies on the population of "switcher" houses that switch between owned and rented status between times $t$ and $t+1$. We only can see data for both the price and rent of a single house if that house is a switcher. However, only a small fraction of the houses in our sample at any given time are switchers. We therefore use the switchers from all years of our data to try to estimate a relationship between rents and prices that is stable over time. We work under the following rank stability assumption, for which we provide empirical evidence in appendix 8.

**Assumption 1** *(Rank stability). Let $q_{r,h,t}$ be the quantile of house h's rent $r_{h,t}$ at time t within the distribution of all observed rents at time t. Similarly, let $q_{ph,t}$ be the quantile of house h's price $p_{h,t}$ at time t within the distribution of all observed prices at time t.*

*1. Let $F_{switcher}(q_{r,h,t}, q_{p,h,t+1})$ be the joint distribution of time t rent quantiles and time t+1 price quantiles for all houses that switch from being rented to owner occupied at time t+1. We assume that this distribution is the same for all times t.*

*2. The joint distribution $F^*_{switcher}(q_{r,h,t+1}, q_{p,h,t})$ between rent and price quantiles for those that switch to being rented at time t+1 is the same as $F_{switcher}(q_{r,h,t}, q_{p,h,t+1})$.*

Under this assumption, we can pool all of the quantile pairs $(q_{r,h}, q_{p,h})$ observed for switchers across our years into a single joint distribution. From observing this distribution, we can see the conditional distribution of rent quantiles given a price quantile $F_{q,rent}(q_{r,h}|q_{p,h})$. This function tells us that a house of price quantile $q_{p,h}$ has a probability $F_{q,rent}(q_{r,h}|q_{p,h})$ of having a rent quantiles of $q_{r,h}$ or less. The conditional distribution of rent quantiles should be increasing in a house's price quantile. We are particularly interested in the inverse function $F^{-1}_{q,rent}(q|q_{p,h})$ which tells us given a price quantile $q_{p,h}$ what rent quantile $q_{r,h}$ satisfies $q = F_{q,rent}(q_{r,h}|q_{p,h})$. That is, the conditional probability of drawing a rent of unconditional quantile of $q_{r,h}$ or below is equal to q given a house of price quantile $q_{p,h}$.

We also need to use information about the marginal distributions $F_{p,t}(p_{h,t})$ of prices and of rents $F_{r,t}(r_{h,t})$ for houses at time t. Because we have a large population of homeowners at time t and of renters at time t, we can estimate these marginal distributions separately year by year. Given all of this information, we can now compute the conditional distribution of rents a house at time t can have given that its price is $p_{h,t}$. Given a house of price $p_{h,t}$, let $r_{h,t}(q|p_{h,t})$ be the level of rent that the conditional probability the house's rent is weakly less than $r_{h,t}$ is equal to q. This rent level is given by the expression

$$r_{h,t}(q|p_{h,t}) = F^{-1}_{r,t}(F^{-1}_{q,rent}(q|F_{p,t}(p_{h,t}))). \tag{5}$$

In particular, the expected rent level at time t given a house of price $p_{h,t}$ is[2]

$$E(r_{h,t}|p_{h,t}) = \int_0^1 r_{h,t}(q|p_{h,t})dq = \int_0^1 F^{-1}_{r,t}(F^{-1}_{q,rent}(q|F_{p,t}(p_{h,t})))dq. \tag{6}$$

Our main rent imputation is an estimate of the median rent $r_{h,t}(q|p_{h,t})$ given $p_{h,t}$, though an estimate of the expected rent $E(r_{h,t}|p_{h,t})$ is also feasible. For both, we would use quantile regressions to estimate

[2] We use the result that if $F^{-1}$ is the inverse CDF or quantile function of a random variable $X$, $E(X) = \int_0^1 F^{-1}(q)dq$.

$F_{q,rent}^{-1}(q|q_{p,h})$. For each q, we run the quantile regression

$$q_{r,h,i} = \alpha_q + \beta_{q1} q_{p,h,i} + +\beta_{q2} q_{p,h,i}^2 + +\beta_{q1} q_{p,h,i}^3 + fe_j + \epsilon_i. \tag{7}$$

The fixed effects $fe_j$ are at the metro-census region level to account for the finding in Demers and Eisfeldt [2001] that rent to price ratios vary systematically across regions at a given time. This quantile regression yields a consistent estimate of $F_{q,rent}^{-1}(q|q_{p,h})$.

Finally, we use a non-parametric estimate of the inverse CDF $\hat{F}_{r,t}^{-1}(q)$, which we obtain by local linear regression. We then simply use the empirical CDF for $p_{h,t}$ as our estimate $\hat{F}_{p,t}(p_{h,t})$. Our imputed rent, using our median estimator, is then for each owner-occupied house i

$$\hat{r}_{h,t,i,med} = \hat{F}_{r,t}^{-1}(\alpha_{.5} + \beta_{.5} \hat{F}_{p,t}(p_{h,t,i})), \tag{8}$$

where $\alpha_{.5}$ and $\beta_{.5}$ are the coefficients from estimating our quantile regression for q=.5. Using a mean estimator, the imputed rent would be

$$\hat{r}_{h,t,i,mean} = \int_0^1 \hat{F}_{r,t}^{-1}(\alpha_q + \beta_q \hat{F}_{p,t}(p_{h,t,i}))dq. \tag{9}$$

Even for a mean estimator, note that we never directly plug in an estimate of the conditional expectation $E(q_{r,h}|q_{p,h}))$. This is $F_{r,t}^{-1}$ is a nonlinear function, so $F_{r,t}^{-1}(E(q_{r,h}|q_{p,h})))$ need not be equal to a house's expected rent conditional on $q_{p,h}$. Indeed, because of the convexity of our quantile function $F_{r,t}^{-1}$, with $\frac{d^2 F_{r,t}^{-1}(q)}{dq^2} > 0$, such a procedure would underestimate rent levels on average, by Jensen's inequality.

# 2 Comparing Rent imputation methods

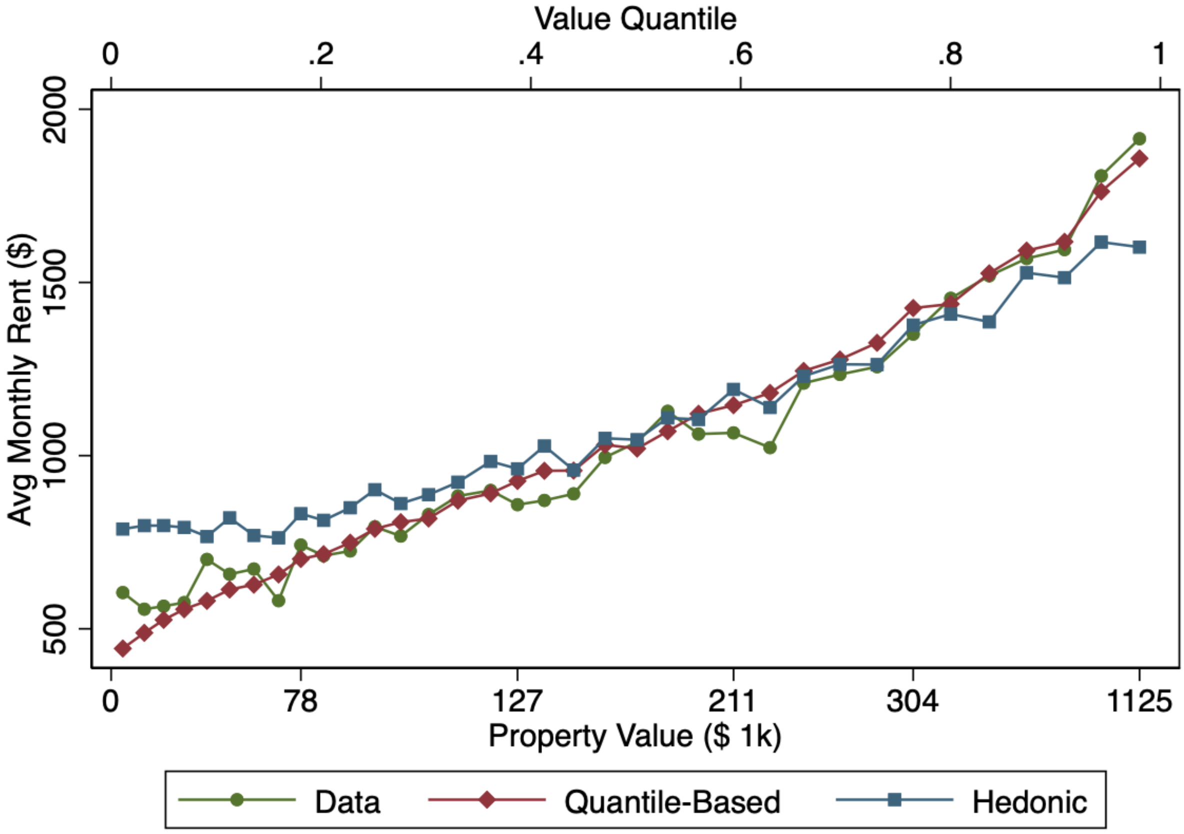
This section compares our rent imputation method to simpler methods that only use the observable characteristics of a property. If a property has a rent $r$, observable characteristics $X$ and a market price $p$, then we expect $E(r|X,p)$ to be increasing in the price p. As a result, $E(r|X)$ should overestimate the rent of low-price properties and underestimate the rent of high-price properties. Using our AHS data on houses that switch between being owned and rented to observe prices and rents together, we show that a rent imputation method using only observable characteristics has this bias. Our quantile-based method, however, predicts the average rent of both low- and high-price properties correctly.

To impute a property's right using only observable characteristics, we run the regression in each year for properties i in regions j

$$yij = \beta_0 + \beta_1 sqft + \beta_2 sqft^2 it + \beta_3 Rooms_i + \beta_4 Bathroomsi + \beta_5 bld.age_i + \beta_6 bld.age_i^2 it + \tag{10}$$

$$\beta_7 bld.age_i^3 it + \beta_8 SingleFamilyResidence_i + RegionalFEs_j + \epsilon_{ij}. \tag{11}$$

Figure 2.1: Performance of Quantile-based and Hedonic Rent Imputation Methods on AHS Data




*Note:* This figure illustrates the results in rent imputation methods for a 20% holdout sample of "switcher" houses with both rent and price data in the AHS from 1980 to 2019. All values are in 2019 dollars. The scale on the x-axis is equally spaced by quantiles of property value.

Table 2.1: Summary statistics for rent imputation comparison

| | Mean | Median | Std. Dev. | 25th Pct. | 75th Pct. | Num. Obs. |
|---|---|---|---|---|---|---|
| *Property Value: Q1* | | | | | | |
| Actual Monthly Rent ($) | 649.86 | 562.50 | 459.94 | 412.35 | 782.24 | 1,726 |
| Fitted Monthly Rent (Quantile Reg, $) | 613.65 | 622.85 | 145.77 | 526.03 | 723.45 | 1,726 |
| Fitted Monthly Rent (Hedonic Reg, $) | 802.93 | 744.63 | 317.19 | 576.24 | 972.47 | 1,506 |
| *Property Value: Q2* | | | | | | |
| Actual Monthly Rent ($) | 878.71 | 876.08 | 394.77 | 620.53 | 1,097.44 | 1,372 |
| Fitted Monthly Rent (Quantile Reg, $) | 906.13 | 905.51 | 124.70 | 825.35 | 978.77 | 1,372 |
| Fitted Monthly Rent (Hedonic Reg, $) | 956.30 | 922.24 | 311.18 | 727.57 | 1,147.37 | 1,206 |
| *Property Value: Q3* | | | | | | |
| Actual Monthly Rent ($) | 1,136.71 | 1,109.72 | 584.34 | 750.83 | 1,430.00 | 1,155 |
| Fitted Monthly Rent (Quantile Reg, $) | 1,183.64 | 1,162.72 | 184.16 | 1,052.99 | 1,309.09 | 1,155 |
| Fitted Monthly Rent (Hedonic Reg, $) | 1,177.52 | 1,143.78 | 371.24 | 924.95 | 1,384.90 | 1,024 |
| *Property Value: Q4* | | | | | | |
| Actual Monthly Rent ($) | 1,601.90 | 1,479.66 | 1,056.81 | 938.32 | 1,858.30 | 1,047 |
| Fitted Monthly Rent (Quantile Reg, $) | 1,605.10 | 1,571.05 | 325.69 | 1,387.69 | 1,803.46 | 1,047 |
| Fitted Monthly Rent (Hedonic Reg, $) | 1,491.76 | 1,456.16 | 470.54 | 1,179.84 | 1,785.19 | 925 |

*Note:* The table documents the differences in rent imputation methods for a 20% holdout sample of "switcher" houses by quartiles of the value of underlying properties. Data are limited to AHS from 1980 to 2019. All values are in 2019 USD.

In each year, we use all rented properties in the AHS as our sample except for a random holdout sample of 20% of those properties that switch between being owned and rented. To test our paper's quantile-based method which uses a property's price to help predict its rent, we run our specification on the 80% of switcher properties that are not included in our holdout sample. We then test its predictions on our 20% holdout sample and present the results in figure 2.1 and table 2.1.

Unlike our quantile-based procedure, the hedonic rent imputation over-predicts the rents on low price properties and under-predicts the rent of high price properties. The orange line representing the average predicted rent from our quantile-based procedure stays close to the green line that presents the average rent in the raw data of our holdout sample. In contrast, the lowest price properties have a rent that is over 100 dollars too high using the hedonic method. Similarly, the hedonic method predicts a rent that is over 200 dollars too low for the highest price properties. In table, 2.1 we present summary statistics about our predicted rents by quartiles of the value of underlying properties. For the lowest quarter of property values, the quantile-based method has an average rent that is 36 dollars above the true data. Meanwhile, the hedonic method is 151 dollars above on average. For the top quarter of property values, our quantile-based method is only 3 dollars above the true data on average. Meanwhile, the hedonic regression's average predicted rent is 110 dollars lower than the data.

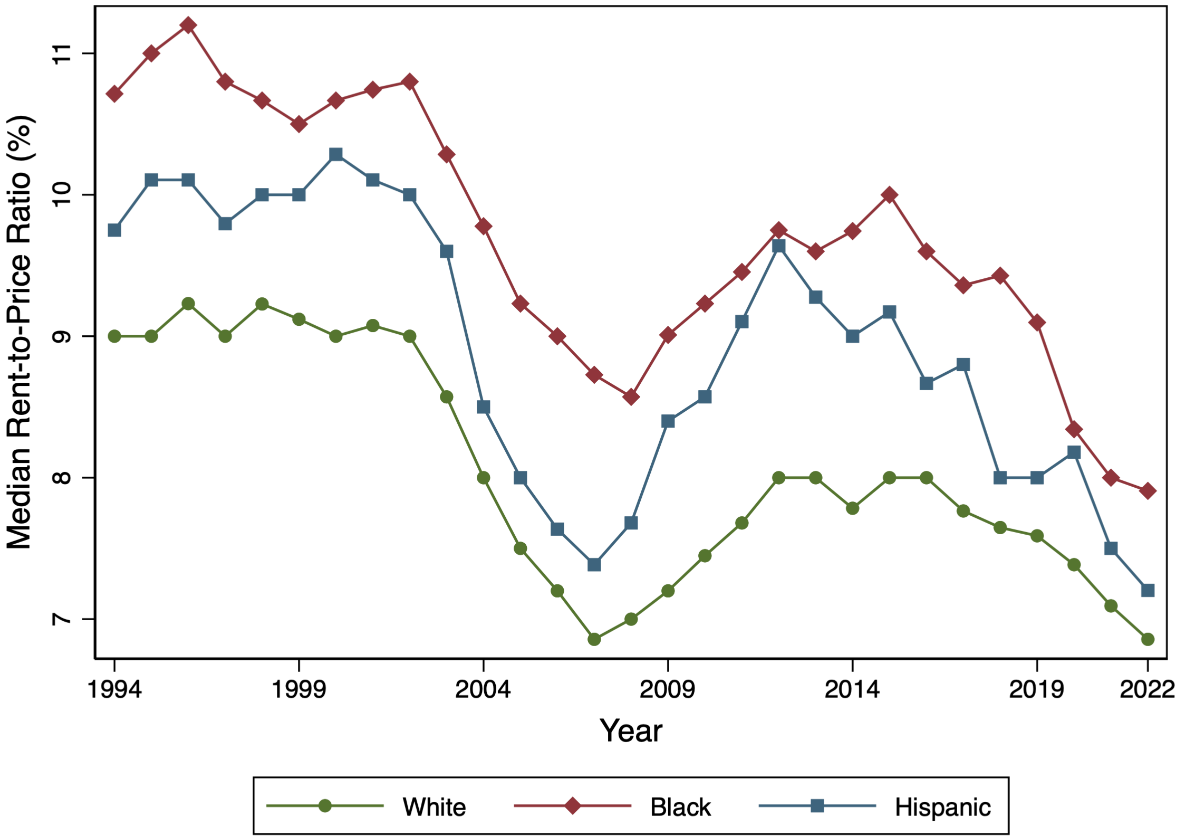
Using a method that accurately estimates the expected rent $E(r|p)$ of a property with a given price $p$ is crucial for our application to understanding racial differences in housing returns. Because minority homeowners tend to live in lower price houses on average than White homeowners, the hedonic method would artificially increase the rental yield of minority homeowners relative to White homeowners. As a result, the return on housing investment for minority homeowners would be artificially increased relative to that of White homeowners. Even with our quantile-based method that accurately predicts the average rent of houses across the property value distribution, we find that including rental yields in our housing return calculation results in a relatively higher return for minorities than White homeowners. If we used a hedonic method of imputing rents, we would mistakenly infer that this return difference is even larger than we find in favor of minorities.

# 3 Direct evidence of rent-to-price ratio differences across racial groups

While much of this paper uses a statistically imputed measure of rent, this section shows directly in raw data from the Consumer Expenditure Survey(CEX) that rent-to-price ratios vary by racial group. Because of this variation in rent-to-price ratios across racial groups, any measure of the return on housing that ignores rents will not accurately measure racial disparities in house investment returns. The CEX includes a question where homeowners are asked the amount they could earn by renting out their home. As a result, this data uniquely includes both a rent and price number for the same house at a given time. Second, we examine summary statistics for the subset of houses that switch between being owner-occupied and rented in the AHS.

Figure 3.1 presents a time series plot of the median rent-to-price ratio in the CEX over time by racial group. There are two key points. First, Black homeowners consistently have the highest ratio, with Hispanic homeowners second highest and White homeowners third. This results in table 3.1 in a median rental yield that is 1.4% higher for Black and .8% for Hispanic homeowners than for White homeowners. Second, rent-to-price ratios fall during housing booms and rise during housing busts for all racial groups. From 2003 until the 2008 financial crisis, rental yields decreased sharply. Then, rental yields rose for all groups until 2012, which was the bottom of the housing market. Then, rental yields fell again for all groups as the housing market recovered. This evidence against stability of the distribution of rent-to-price ratio over time provides additional motivation for our quantile-based approach to inferring rents, which allows for such temporal instability.

Figure 3.1: Median Rent-to-Price Ratios for Homeowners by Racial Group over Time, 1994-2022 in the Consumer Expenditure Survey




While the CEX is the only dataset that directly allows us to see a rent measure and price measure for the same house at the same time, we find consistent evidence from houses in the AHS that switch between owned and rented in figure 3.2 and table 3.3. Among these "switchers," the table shows that Black homeowners have a 1% higher average and .4% higher median rent-to-price ratio than White homeowners. Hispanic switcher homeowners have a 1.1% higher average and .2% lower median rent-to-price ratio than White homeowners. Although qualitatively consistent with the CEX evidence, these differences are likely smaller than those in the CEX because houses that switch between being rented and owned are a selected sub-sample. As a result, this sub-population of homeowners likely has less difference across racial groups than the overall sample. In addition, the sample sizes here are roughly 1/10 of those in the CEX, which adds statistical noise. This noise is visible in the year-to-year volatility in figure 3.2, even though the broad time-series movements in rent-to-price ratios over the business cycle are similar to those in figure 3.1.

After using our quantile-based rent imputation method on AHS data, we obtain rent-to-price ratios for the overall population of homeowners of each race that are similar to the raw CEX data. We plot the median rent-to-price ratio by racial group for all homeowners in the AHS in figure 3.3. Here, the median Black rent-to-price ratio is consistently above that for Hispanics, which is consistently above that for Whites. The fall in these ratios before the 2008 crisis and increase after also closely tracks the CES data.

Table 3.2 compares our CEX rent variables to predictions based on our quantile-based method. Based on our regressions run on the AHS, we use variables in the CEX to construct fitted values we compare to reported rents. The difference between rent and imputed rent is roughly constant across racial groups. Homeowners likely are over-optimistic about the rent their homes could earn, but this bias seems unlikely to explain the disparity in rent-to-price ratios we observe across racial groups.

Table 3.1: Summary Statistics on Rent-to-Price Ratios by Race for Homeowners in the Consumer Expenditure Survey, 2003-2022

| Race | Mean | Median | Std. Dev. | 25th Pct. | 75th Pct. | Num. Obs. |
|---|---|---|---|---|---|---|
| White | 9.29% | 7.50% | 8.62% | 5.71% | 10.00% | 256,268 |
| Black | 11.50% | 9.09% | 9.77% | 6.88% | 12.00% | 26,526 |
| Hispanic | 11.07% | 8.33% | 11.47% | 6.01% | 11.33% | 29,049 |

*Note:* The sample includes all homeowners and is winsorized at the 1st and 99th percentiles.

Table 3.2: Comparison between imputed and reported rents: by race

| | White<br>Mean | Black<br>Mean | Hispanic<br>Mean |
|---|---|---|---|
| Reported | 19021.84 | 16489.37 | 18648.4 |
| Imputed rent | 14703.46 | 12221.18 | 14087.65 |
| Reported - imputed | 4318.377 | 4268.183 | 4560.748 |
| Observations | 288349 | 24869 | 26482 |

Sample: 1994 to 2021.

Table 3.3: Rent-to-Price Ratio for Homes that Switch Between Owned and Rented in the AHS by Owner Race, 1975-2019

| **Rent-to-price ratio** | Num. Obs. | Mean | Median | SD |
|---|---|---|---|---|
| **Race of owner-occupant** | | | | |
| White | 22,062 | 10.6 % | 7.7 % | 11.9 % |
| Black | 3,847 | 11.5 % | 8.1 % | 12.5 % |
| Hispanic | 4,406 | 11.6 % | 7.5 % | 13.9 % |
| Other | 1,769 | 9.4 % | 6.2 % | 12.1 % |
| Total | 32,084 | 10.8 % | 7.6 % | 12.2 % |

*Note:* Winsorized at 5% level.

Figure 3.2: Plot of Median Rent-to-Price Ratio by Owner Race and Year, Owned/Rented Switcher Units

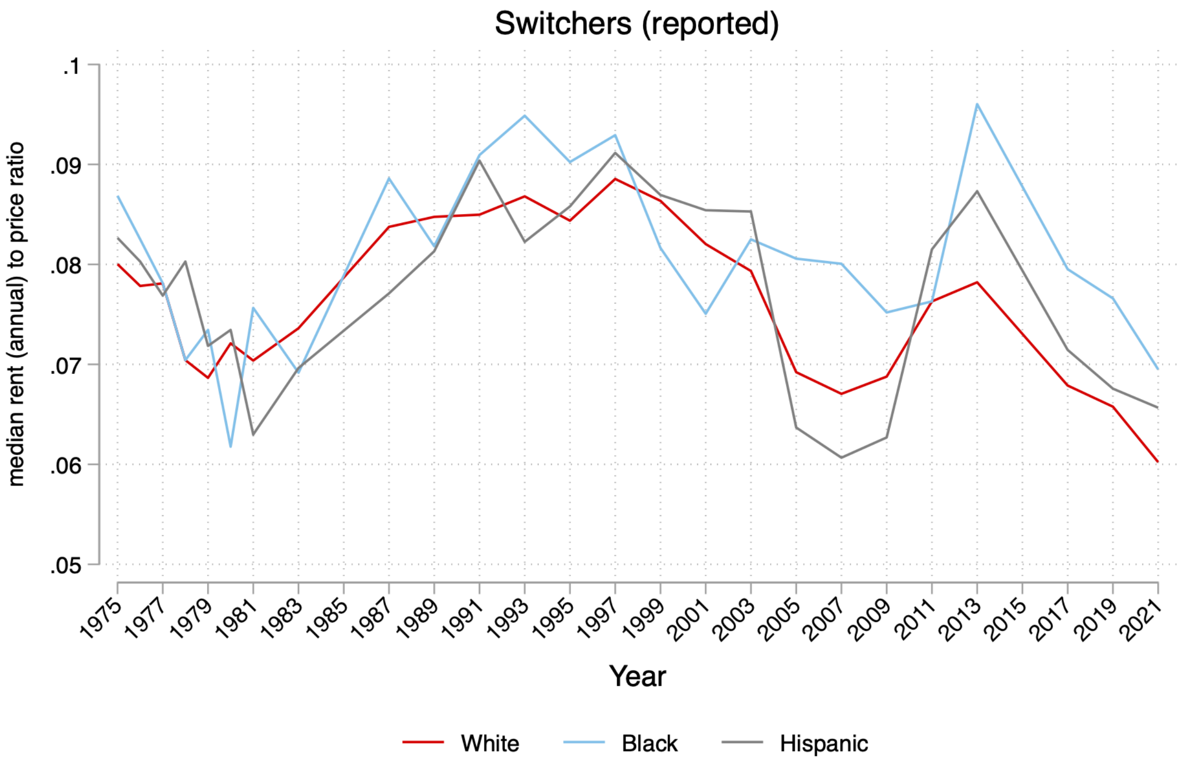
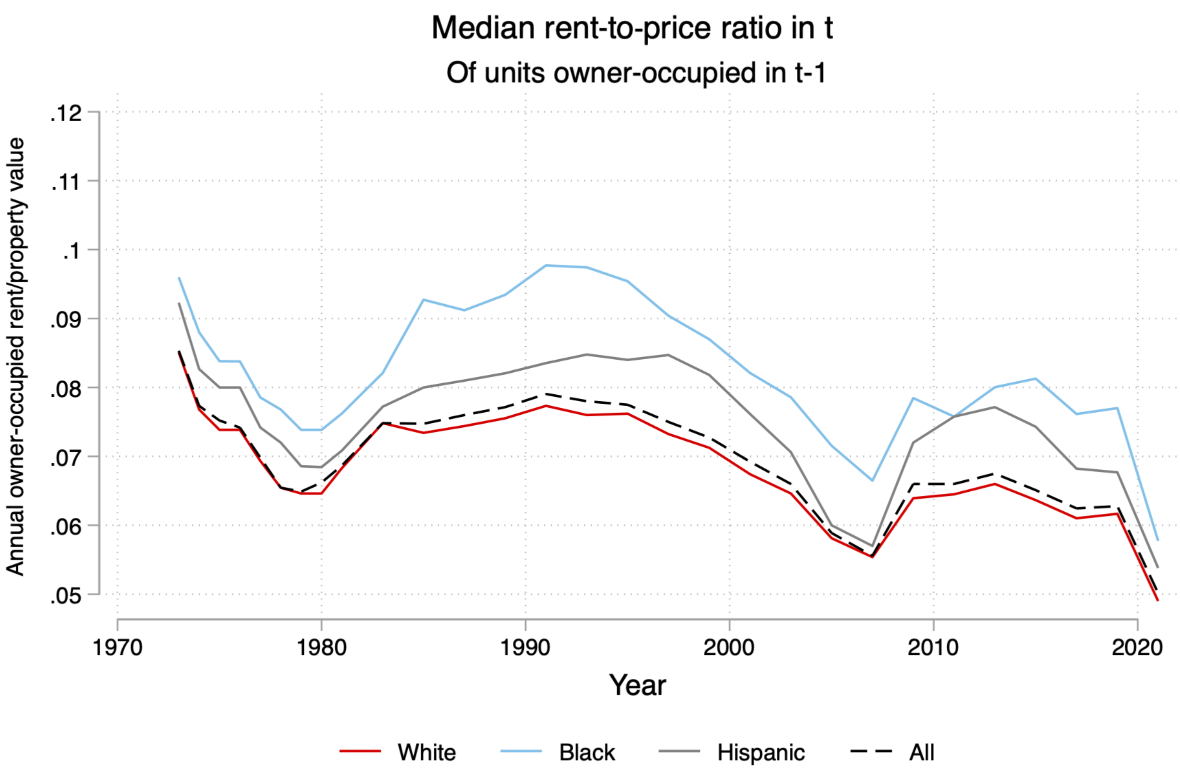



Figure 3.3: Plot of Median Rent-to-Price Ratio by Owner Race and Year, Imputed Rent for all Homeowners



# 4 Racial Disparities in the Return on Owner-Occupied Housing in the AHS

This section applies our quantile-based rent imputation to compare the rates of return on owner-occupied housing earned by different racial groups. Our analysis so far has shown that our rent imputation method provides accurate rent predictions both for low- and high- price homes and that rental yields vary significantly between racial groups. This section now shows that the higher rent-to-price ratios for minority homeowners are reflected in a higher average return on housing investment as well.

We begin in section 4.1 by comparing the returns on portfolios composed of all owner-occupied housing of a given racial group. We find that the Black and Hispanic portfolios have higher returns than the White portfolio but also more exposure to the business cycle. Next, we analyze the return on housing at an individual homeowner level in section 4.2. We find that the average rate of return for Black and Hispanic homeowners is higher than for White homeowners, due in large part to a higher rental yield for minority homeowners. However, we are able to account for most of this rate-of-return difference with differences in homeowner characteristics across racial groups. Following Cook et al. [2021], we find using a Gelbach [2016] decomposition that controlling for the income and education of homeowners largely explains racial differences in average rates of return. We also find that minority homeowners have worse average returns during recessions, and we are able to account for this disparity by controlling for observable characteristics for Hispanic homeowners but not for Black homeowners.

**AHS Summary Statistics by Racial Group**

Table 4.1 reports summary statistics in the AHS for across the White, Black, and Hispanic racial groups. As can be seen in the number of observations that report property value versus rent, the White homeownership rate is considerably higher than the Black or Hispanic homeownership rates. However, White and Hispanic homeowners have very similar average property values ($ 233,504 versus $ 234,500). In contrast, the average value of $163,995 for Black homeowners is considerably lower. For each racial group, the average owner's equivalent rent we impute is higher than the average rent paid by renters, reflecting the fact that homeowners tend to be wealthier and live in more expensive homes than renters. We also find that for each racial group, average annual property taxes are 2-3 times average monthly owner's equivalent rent. Routine maintenance is somewhat below one month of owner's equivalent rent. Average home improvement is 4-5 times routine maintenance, median far below its mean reflecting rare but large improvements. As a result, the average dividend yield earned by homeowners in our return calculation is roughly 6 months worth of owner's equivalent rent on average. We find however, that the ratio of average owner's equivalent to average property value is slightly higher for Hispanic than for White homeowners and is the highest for Black homeowners. This is consistent with the time-series plot in figure 3.3.

Table 4.1: Summary Statistics for Owners and Renters in the AHS by Racial Group, 1975-2021

| | Obs. | Mean | Median | Std. Dev. |
|---|---|---|---|---|
| **White** | | | | |
| Property value | 709,212 | 233,504 | 173,260 | 246,399 |
| Rent (monthly) | 298,008 | 877 | 754 | 646 |
| Owner's equivalent rent (monthly) | 709,212 | 1,033 | 976 | 411 |
| Routine maintenance (annual) | 409,624 | 793 | 358 | 1,357 |
| Home improvement (annual) | 391,671 | 5353 | 568 | 19,797 |
| Property tax (annual) | 715,586 | 2,777 | 1,944 | 3439 |
| **Black** | | | | |
| Property value | 67,914 | 163,995 | 118,797 | 191,319 |
| Rent (monthly) | 89,847 | 722 | 656 | 496 |
| Owner's equivalent rent (monthly) | 67,914 | 896 | 843 | 374 |
| Routine maintenance (annual) | 41,137 | 774 | 329 | 1,374 |
| Home Improvement (annual) | 41,443 | 3724 | 196 | 21386 |
| Property tax (annual) | 68,350 | 1,958 | 1,262 | 2,762 |
| **Hispanic** | | | | |
| Property value | 52,972 | 234,500 | 170,484 | 247,137 |
| Rent (monthly) | 67,086 | 925 | 825 | 611 |
| Owner's equivalent rent (monthly) | 52,972 | 1,106 | 1,037 | |
| Routine maintenance (annual) | 39,672 | 895 | 430 | 1,479 |
| Home improvement (annual) | 41, 810 | 4495 | 319 | 16,548 |
| Property tax (annual) | 53,084 | 2,647 | 1,814 | 3,398 |

Summary statistics for all data except rent are from owner-occupied homes. Rent is from rented homes. Owner's equivalent rent is imputed based on the method in section 1.1.

## 4.1 Housing Rates of Return:Group-Level Portfolios

This section analyzes the returns on group-level housing portfolios owned by white, black, and Hispanic homeowners. For each racial group, we construct a portfolio composed of all owner-occupied housing owned by members of that racial group. We then compute the rate of return an investor would earn if they bought this entire portfolio, accounting both for appreciation in house prices as well as rental income, property taxes and maintenance. The returns on such a group-level portfolio may be more precisely estimated than the returns earned by individual homeowners. This is because there is a possibility of noise in each individual homeowner's reported data, which averages out as part of a portfolio.

Let $W_t^g = \sum_i P_t^i h_t^{i,g}$ be the housing wealth of group g at time t. In this expression, i indexes a house and $h_t^{i,g} = 1$ if a member of group g owns the house, otherwise it equals 0. $P_t^i$ is the market price of house i at time t. and house i at time t provides a "rent value" of $r_t^i$ to its owner. In addition, there is a property tax $\pi_t^i$ on house i and maintenance expenditures $m_t^i$ on house i. The return $R_t^g$ on group g's housing investment from time t-1 to t is

$$R_t^g = \frac{\sum_i [P_t^i - P_{t-1}]h_{t-1}^{i,g}}{\sum_i P_{t-1}^i h_{t-1}^{i,g}} + \frac{\sum_i [r_t^i - \pi_t^i - m_t^i]h_{t-1}^{i,g}}{\sum_i P_{t-1}^i h_{t-1}^{i,g}}. \tag{12}$$

The first term in equation (12) is the "price appreciation" $\frac{\sum_i [P_t^i - P_{t-1}]h_{t-1}^{i,g}}{\sum_i P_{t-1}^i h_{t-1}^{i,g}}$ that measures the growth rate for the value of housing the group owns starting in time t-1. In addition, we have the "dividend yield" $\frac{\sum_i [r_t^i - \pi_t^i - m_t^i]h_{t-1}^{i,g}}{\sum_i P_{t-1}^i h_{t-1}^{i,g}}$ that accounts for all cash flows generated by the houses. This return equals what a landlord would get who earned rent (and also paid taxes and maintenance) or can be interpreted as the return of a homeowner who would otherwise have to pay rent to consume the same quality of housing.

We present summary statistics on the real rates of return of our group-level housing portfolios in table 4.2. The average rate of return is 5.2% for the White housing portfolio, strictly below the 6.6% average rate of return for the Black portfolio and 7.0% for the Hispanic portfolio. We also find that the White and Black portfolios have similar return standard deviations (4.2% and 4.4%), but the Hispanic portfolio has a considerably higher return standard deviation (5.7%). Because such a large share of the population is white, particularly in the earliest years of our data, the summary statistics for the entire stock of housing wealth is similar to those for the White portfolio.

Table 4.2: Rate of return on group-level housing portfolios, 1975-2021

| | White | Black | Hispanic | All groups |
|---|---|---|---|---|
| Mean | 4.38 % | 4.31 % | 5.57 % | 4.45 % |
| Median | 4.92 % | 4.32 % | 6.77 % | 4.96 % |
| Std. Dev. | 3.90 % | 3.92 % | 5.53 % | 3.98 % |

Each racial group's portfolio is composed of all owner-occupied housing owned by members of that group. The summary statistics of each portfolio reflect time-series variation in the performance of the portfolio.

In tables 4.3 and 4.4, we decompose each portfolio's rate of return into a dividend yield term and a capital gains term. Table 4.3 shows that the rate at the White housing portfolio and Black housing portfolio appreciate in value is quite similar (1.34% versus 1.26%). However, the Hispanic housing portfolio appreciates almost twice as fast (2.42% per year). In addition, the White and Black portfolio's have similar standard deviations for their appreciation rates over time (3.82% versus 4.08%), while that for the Hispanic portfolio is considerably higher (5.15%).

Table 4.4 compares the dividend yields of the housing portfolios of our 3 racial groups. The dividend on a portfolio is its imputed rental income estimates minus property taxes and maintenance. The average dividend yield is similar for the Black and White portfolios and slightly higher for the Hispanic portfolio. While the dividends yields are slightly more volatile for the two minority portfolios, the vast majority of each portfolio's return volatility is due the the capital gains term in table 4.3. Black and particularly Hispanic homeowners have more volatile rates of return on housing because the market value of the houses they buy are more volatile.

Rate of return decomposition for group-level housing portfolios, 1975-2021

Table 4.3: Value appreciation rate (excl. dividends)

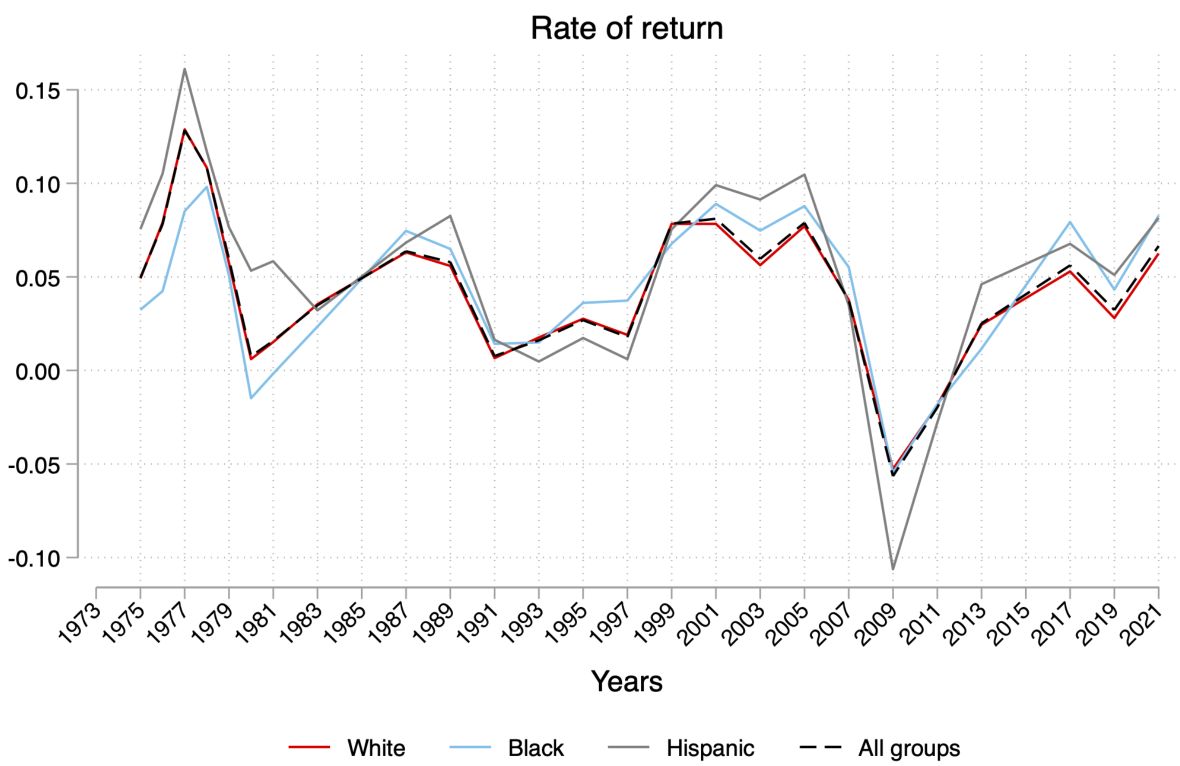
| | White | Black | Hispanic | All groups |
|---|---|---|---|---|
| Mean | 1.34 % | 1.26 % | 2.42 % | 1.39 % |
| Median | 1.19 % | 2.04 % | 3.19 % | 1.54 % |
| Std. Dev. | 3.82 % | 4.08 % | 5.15% | 3.91 % |

Table 4.4: Dividend rate

| | White | Black | Hispanic | All groups |
|---|---|---|---|---|
| Mean | 3.04 % | 3.06 % | 3.17 % | 3.07 % |
| Median | 2.84 % | 3.07 % | 3.14 % | 2.90 % |
| Std. Dev. | 1.29% | 0.73 % | 1.0 % | 1.2 % |

Each racial group's portfolio is composed of all owner-occupied housing owned by members of that group.
The summary statistics of each portfolio reflect time-series variation in the performance of the portfolio.

Figure 4.1: Rate of Return on Total Housing Wealth Portfolio of Each Racial Group, 1975-2021




To complement our summary statistics, we present in figure 4.1 the returns on each racial group's housing wealth portfolio year. We find sharp declines in housing returns around recessions in 1980, 1991, and especially 2008. We find rising returns through out the 1980s credit boom, the 2000s housing boom (preceded by a late 1990s expansion of GSE credit [Bhutta, 2012]), and in the recovery after the 2008 financial crisis. Moreover, in periods of high returns, we tend to find that Black and Hispanic investment returns are particularly high. In contrast, during the low return periods in the 1990s and especially the housing crash in 2008-2012, Hispanic returns tend to be the lowest. Hispanic housing returns therefore seem to have the most sensitivity to the business cycle, and their high average returns could be compensation for this risk. In contrast, we find that the return on the Black housing portfolio is either equal to or greater than the White housing return in every year except 1975-1976.

## 4.2 Housing Rates of Return for Individual Households

This section analyzes the rates of return individual households earn on their housing investments to complement the group-level evidence above. Like above, we find that minority homeowners earn higher average rates of return than White homeowners, due in large part to a higher rental yield for minority-owned homes. In addition, we find that both Black and Hispanic homeowners have housing returns that are considerably more sensitive to the business cycle than White homeowners. Using a Gelbach [2016] decomposition, we show that observable covariates are largely able to explain the difference in average rates of return between racial groups. However, we find that observables are unable to account for the higher cyclicality of returns for particularly for Black homeowners.

For each homeowner, we calculate the log-return on its housing to account for their home's rental value, property taxes and maintenance. First, we compute the total "dividend" of house $i$ at time $t$ as

$$dividend_{it} = rent_{it} - proptax_{it} - maint_{it}, \tag{13}$$

where the rent is imputed using our quantile-based methodology and property taxes and maintenance costs are directly reported in the AHS. For years when the AHS is reported annually, we measure log-returns as

$$log(P_{it+1} + dividend_{it}) - log(P_{it}). \tag{14}$$

After 1981, when the AHS becomes only bi-annual, we annualize our rate our rate of return measure by dividing the expression in equation 14 by 2. This yields a house-time level panel of rates of returns.

Table 4.5 presents summary statistics on our return measure by race. The average return on housing is 6.1% for White homeowners, 6.7% for Black homeowners, and 7.9% for Hispanic homeowners. These average returns are higher than the return on the group-level portfolio returns in table 4.2, but the differences in average returns across racial groups are similar. Unlike our group-level returns above, we now find that the standard deviation of returns is highest for Black, not Hispanic homeowners. This suggests that Black homeowners with the lowest property values have particularly volatile returns. Such homeowners would have little impact on the volatility of the return on the aggregate housing stock of Black homeowners (which is impacted more by the return on higher value homes) but would contribute to a high standard deviation in table 4.5 (which weighs all homeowners equally).

We decompose racial rate of return differences into property appreciation rates and dividend yields in tables 4.6 and 4.7. Like for the group-level portfolios, we find similar appreciation rates for White (2.1%) and Black (2.4%) homeowners but a higher appreciation rate for Hispanic (3.9%) homeowners. In addition, we find that Hispanic and particularly Black homeowners have a higher average dividend yield (5.1% and 5.0%, respectively) than White homeowners (4.6%). The high 37.9% standard deviation of returns for Black homeowners is explained primarily by their property appreciation rate being 38.8%, the highest of all racial groups. This is broadly consistent with the Kermani and Wong [2022] finding that a disproportionate share of distressed sales for minority homeowners raises their probability of a large negative investment return.

Table 4.5: Return on housing investments for individual homeowners by race

| | White | Black | Hispanic | All groups |
|---|---|---|---|---|
| Mean | 6.1 % | 6.7 % | 7.9 % | 6.3 % |
| Median | 4.0 % | 4.7 % | 5.3 % | 4.1 % |
| Std. Dev. | 30.8 % | 37.9 % | 33.3 % | 31.7 % |

Return on housing computed using equation 14.

Table 4.6: Property appreciation rate (no dividends)

| | White | Black | Hispanic | All groups |
|---|---|---|---|---|
| Mean | 2.1 % | 2.4 % | 3.9 % | 2.3 % |
| Median | 0.3 % | 0.6 % | 1.76 % | 0.4 % |
| Std. Dev. | 31.4 % | 38.8 % | 33.9 % | 32.3 % |

Appreciation rate computed as $log(P_{it+1}) - log(P_{it})$.

Table 4.7: Dividend yield

| | White | Black | Hispanic | All groups |
|---|---|---|---|---|
| Mean | 4.6 % | 5.1 % | 5.0 % | 4.7 % |
| Median | 4.2 % | 4.4 % | 4.3 % | 4.2 % |
| Std. Dev. | 9.6 % | 10.2 % | 11.8 % | 9.8 % |

Dividend yield computed as $log(1 + \frac{dividend_{it+1}}{P_{it}})$.

We examine in table 4.8 the extent to which we can account for disparities in housing rates of returns between racial groups with observable characteristics. To do so, we regress our individual-level rates of return on a dummy variable for eac racial group together with a growing list of control variables. Controlling for year fixed effects results in a remaining .66% higher rate of return for Blacks homeowners and 1.5% higher rate of return for Hispanic homeowners relative to White homeowners. This is somewhat close to the difference in average returns in table 4.5. Controlling for income, education, and years since most recent move more than entirely accounts for the Black-White difference in returns and the Hispanic-White difference in returns. This suggests that houses bought by lower income homeowners tend to have higher average returns and that these houses are disproportionately bought by minority homeowners. Adding an additional control for housing characteristics results in effectively zero difference in returns across groups. One explanation is that Black and White homeowners of similar income earn similar rates of return but that White homeowners of a given income tend to buy smaller homes of higher unobservable quality. Controlling for observable housing characteristics therefore might compare White homeowners in a lower return segment of the market to observably similar Black homeowners.

We use a Gelbach [2016] decomposition in figure 4.2 to quantify our ability to explain differences in average returns across homeowners of difference races. This decomposition compares a regression with all covariates column 5 of table 4.8 to the one in the first column. The decomposition uses a multivariate version of the formula for omitted variable bias to account for how much each variable helps to explain the change in our regression coefficient when all are included jointly. Unlike the standard approach of adding regressors one-by-one, this decomposition does not require us to arbitrarily choose the order in which regressors are added. We find consistent with table 4.8 that the income/education controls are by far the most important for explaining racial differences in rates of return. However, several other covariates also help for explaining the Hispanic-White return gap. After conditioning on income/education, observable housing characteristics provide minimal help to account for residual differences in rates of return.

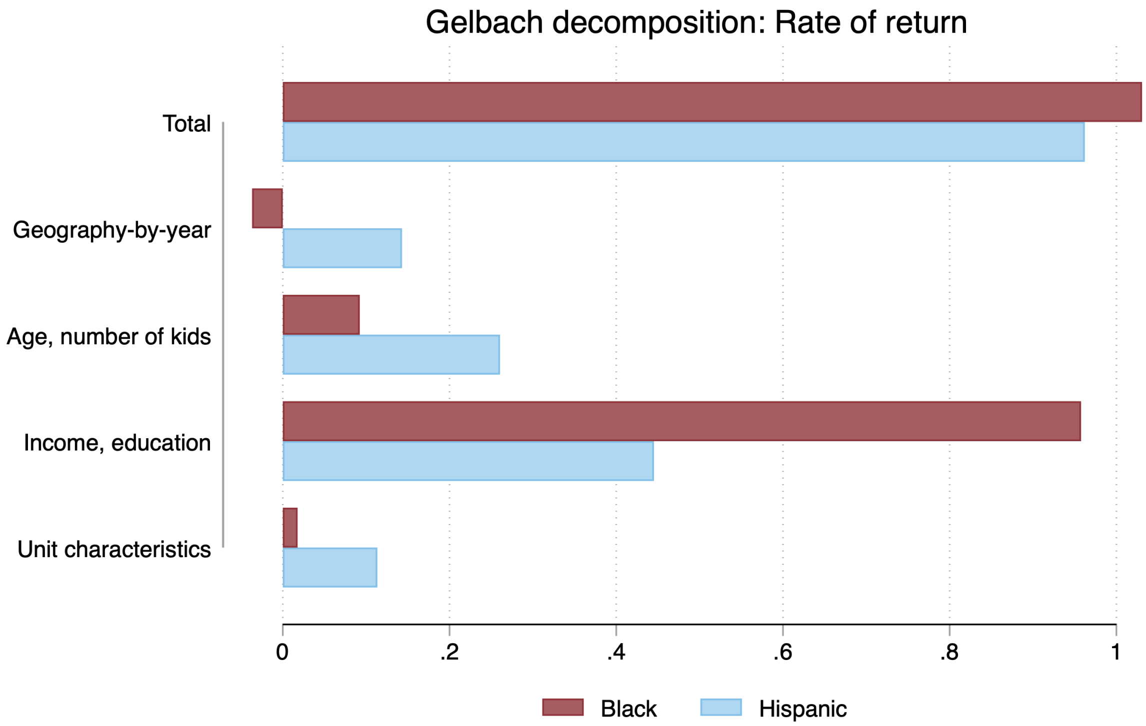
Table 4.8: Log rate of return

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Black owner-occupant=1 | 0.00665*** | 0.0113*** | 0.00776*** | 0.00893*** | -0.00641*** | -0.00263 |
| | (0.00141) | (0.00142) | (0.00142) | (0.00142) | (0.00141) | (0.00138) |
| Hispanic owner-occupant=1 | 0.0150*** | 0.0184*** | 0.0135*** | 0.0156*** | -0.00391* | -0.00187 |
| | (0.00157) | (0.00166) | (0.00165) | (0.00165) | (0.00164) | (0.00158) |
| Observations | 739127 | 739127 | 739127 | 739127 | 739127 | 739127 |
| Dummy for other racial group | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE (not interacted) | Yes | - | - | - | - | - |
| Region/Division-by-year FE | No | No | No | No | No | No |
| MSA-by-year FE | No | Yes | No | No | No | No |
| Geography-by-year FE | No | No | Yes | Yes | Yes | Yes |
| Demographics (age, kids) | No | No | No | Yes | Yes | Yes |
| Inc.,educ., yrs since move | No | No | No | No | Yes | Yes |
| Housing characteristics | No | No | No | No | No | Yes |

Standard errors are clustered at the unit level. The excluded race group are white owner-occupants.
Geography groups region, metro, degree, division, urban, msa. The imputation for owner-occupied rent ran separately by racial group.

We attempt to account for racial disparities in housing returns over the business cycle in table 4.9. In all of our specifications, an interaction between a recession indicator and an indicator for either a Black or Hispanic homeowner is negative. This implies that Black and Hispanic homeowners have returns on housing that are disproportionately sensitive to the business cycle. However, column 6 shows that controlling for homeowner income/education largely accounts for this cyclicality for Hispanics. As shown in Landvoigt et al. [2014], the low price end of the housing market is more sensitive to business cycles, and lower income homeowners are likely to have lower value homes. As a result, since Black and Hispanic homeowners have lower average income, they tend to buy homes that provide investment returns that are more sensitive to the business cycle. Once education/income is accounted for, some of the racial difference in return cyclicality is explained. However, we find that these observables have almost no ability to account for the excess cyclicality of Black homeowners' returns. One possibility is that segregation between Black and White homeowners results in Black homeowners living in more cyclical areas. Even after conditioning on income and housing characteristics, this excess cyclically remains.

Figure 4.2: Gelbach [2016] Decomposition of Explainable Share of Racial Disparities in Returns




Fraction of Return Disparity Explained By Each Characteristic

Table 4.9: Log rate of return (incl. recession interaction)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Black owner-occupant=1 | 0.00957*** | 0.00925*** | 0.0147*** | 0.0116*** | 0.0129*** | -0.00330 | 0.000788 |
| | (0.00181) | (0.00182) | (0.00185) | (0.00186) | (0.00187) | (0.00185) | (0.00180) |
| Black owner-occupant=1 × Recession=1 | -0.0103** | -0.01000** | -0.0120*** | -0.0138*** | -0.0141*** | -0.0110** | -0.0121*** |
| | (0.00340) | (0.00341) | (0.00347) | (0.00361) | (0.00362) | (0.00363) | (0.00362) |
| Hispanic owner-occupant=1 | 0.0202*** | 0.0147*** | 0.0230*** | 0.0164*** | 0.0187*** | -0.00255 | -0.000466 |
| | (0.00192) | (0.00191) | (0.00204) | (0.00204) | (0.00204) | (0.00204) | (0.00197) |
| Hispanic owner-occupant=1 × Recession=1 | -0.0215*** | -0.0160*** | -0.0188*** | -0.0117** | -0.0122** | -0.00610 | -0.00625 |
| | (0.00372) | (0.00372) | (0.00385) | (0.00386) | (0.00389) | (0.00394) | (0.00389) |
| Observations | 739127 | 739127 | 739127 | 739127 | 739127 | 739127 | 739127 |
| Recession-by-race FE (fully interacted) | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Dummy for other racial group | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year FE (not interacted) | Yes | - | - | - | - | - | - |
| Region/Division-by-year FE | No | Yes | No | No | No | No | No |
| MSA-by-year FE | No | No | Yes | No | No | No | No |
| Geography-by-year FE | No | No | No | Yes | Yes | Yes | Yes |
| Demographics (age, kids) | No | No | No | No | Yes | Yes | Yes |
| Inc., educ., yrs since move | No | No | No | No | No | Yes | Yes |
| Housing characteristics | No | No | No | No | No | No | Yes |

Standard errors are clustered at the unit-level (CONTROL). The excluded race group are white owner-occupants. Recession year dummies (1975,1979,1980,1981,1989, 1991,2007,2009) come from FRED: https://fred.stlouisfed.org/series/JHDUSRGDPBR. Geography groups region, metro, degree, division, urban, msa.
The imputation for owner-occupied rent ran separately for each racial group. In absorbing FE, singelton observations are dropped.

# 5 Racial Disparities in Housing Returns in Corelogic

This section analyzes racial disparities in the return on owner-occupied housing in Corelogic data. Corelogic's data attempts to track all individual purchases and sales of housing, providing considerably more data than the AHS for studying housing returns. In addition, Corelogic data has the benefit of featuring actual housing transactions rather than survey responses, potentially reducing measurement error relative to the AHS. A downside of this data is that homeowner race and housing rents, and the data in most counties begins only in the late 1990s, providing a shorter time series than the AHS. Our analysis currently covers the 50 most populous counties in the United States, with a population of just over 100 million. For each resident of owner-occupied housing, we infer the price and rental value of their home in each year to construct a panel of housing investment returns.

**Data Processing**

We impute homeowner race following [Ye et al., 2017, Diamond et al., 2019]. The first step uses the NamePrism machine learning algorithm to predict race given a homeowner's name. Next, the racial demographics of each census block are used to reweight the probabilistic race predictions of NamePrism using Bayes' rule. We then classify any homeowner with greater than a .8 chance of being White, Black or Hispanic as part of that racial group. This leads to a modest overestimate of the White share of the population, since the larger White population in the US will be reflected in the Bayesian weights coming from population shares. Moreover, Black and Hispanic homeowners are more likely to be identified in minority-dominant areas, leading to a greater degree of geographic segregation than in the true data. That said, name and geography are jointly a highly accurate predictor of race, and our large sample size allows for precise statistical inference even if we underestimate the size of the minority population.

We use a hedonic model for house prices to compute annual rates of return on each house despite the fact that houses only trade infrequently. We use all transactions $i$in a given county $j$ and year $t$ where an owner-occupant either buys and/or sells their home to run the regression[3]

$$P_i = \beta_1 sqft_i + \beta_2 sqft_i^2 + \beta_3 rooms_i + \beta_4 bedrooms_i + \beta_5 bathrooms_i + \beta_6 bld.age_i \tag{15}$$

$$+\beta_7 (bld.age_i)^2 + \beta_8 (bld.age_i)^3 + \alpha_{proptype_i} + \alpha_{zip_i} + \epsilon_i. \tag{16}$$

Here, $P_i$ is the transaction price of house i, $sqft$ and $bld.age$ are its square footage and building age, and rooms/bedrooms/bathrooms are the number of total rooms, bedrooms, and bathrooms in the property. We also include a fixed effect $\alpha_{proptype_i}$ for houses and apartments and a fixed effect $\alpha_{zip_i}$. We then use our estimated regression model to get predicted property values for homes, including those that did not transact in year t.

Next, we use a procedure to impute the rental value of each house, because our transaction data only reports house prices. We use our rent estimates from the AHS to impute rents on observably similar properties in Corelogic. First, we use the procedure described above to compute the rental yield for each owner-occupied property in the AHS. We then compute the rent to price ratio $\frac{rent}{price}_i$ for each AHS

[3]We also run specifications without bedrooms and/or bathrooms and without building age to get a price prediction for properties that are missing some of the covariates in our benchmark regression.

Table 5.1: Corelogic Data Count by Decade

| Decade | # Obs |
|---|---|
| 1980s | 42,707,063 |
| 1990s | 120,669,848 |
| 2000s | 186,933,864 |
| 2010s | 180,270,334 |

Each observation is a property-year with a non-missing annual return and non-missing resident race.

property and run separately for each year t the regression [4]

$$(\frac{rent}{price})_i = \beta_1 sqft_i + \beta_2 sqft_i^2 + \beta_3 rooms_i + \beta_4 bedrooms_i + \beta_5 bathrooms_i + \beta_6 bld.age_i \quad (17)$$

$$+\beta_7(bld.age_i)^2 + \beta_8(bld.age_i)^3 + \alpha_{region_i} + \epsilon_i, \quad (18)$$

where $\alpha_{region_i}$ is a geographical fixed effect at the same AHS region-metro level described above. Once we have predictions for the rent of a property and its rent to price ratio, we multiply the two together to get a prediction of the property's rent.

Using our imputed price and rent series, we measure the annual rate of return on a housing investment as

$$log(P_{t+1} + r_{t+1}) - log(P_t) \quad (19)$$

using our imputed price and rent data. This approach has the benefit of providing a consistent year-by-year return on housing for all homeowners. The downside is that imputed rather than true transaction prices are used in the calculation. We additionally compute an annualized return measure

$$(\frac{P_{sell}}{P_{buy}})^{\frac{1}{years}} - 1 \quad (20)$$

for a property that a homeowner first buys and then sells. Years is measured as a non-integer number based on the difference between the date of purchase and date of sale. Table 5.1 reports the number of observations in our panel of imputed rents and prices by decade. We count any observation where resident race is observable as well as the log-return on housing from the previous year. After the 1980s, with only 42.7 million observables, we see 120.7 million observations in the 1990s, 187.0 million in the 2000s, and 180.2 million in the 2010s. We therefore have roughly 30 years during which the data is particularly well populated. The overall time series goes back to 1980, somewhat short of the 1975 start date in the AHS. Nevertheless, there are 5 recessions (1980, 1981-1982, 1990-1991, 2000-2001, 2007-2009) in the data, allowing us to consider business cycle variation in the return on housing investments.

**Results**

We present summary statistics on these rate of return estimates in table 5.2. Like in the AHS, we find that minorities earn higher average housing returns than White homeowners but that they face higher volatility. The White average return of 9.56% is slightly below that of Black homeowners at 9.67% and even further below the 11.32% earned by Hispanics. However, the standard deviation of White returns at

[4]Like in our price regression, we also run this without rooms, without bedrooms, and again without bathrooms to construct a fitted value for observations missing one of these variables. We also include a dummy for a variable being top-coded in AHS data.

13.51% is well below the 23.56% standard deviation for Black returns and 20.86% for Hispanic returns. As in the AHS, our benchmark result is that Black and Hispanic homeowners have higher but more volatile rates of return for housing investment than White homeowners. Average returns are higher here than in our AHS results, likely due to the fact that AHS returns account not only for rent but also for property taxes and maintenance.

Table 5.2: Rate of Return Summary Statistics by Racial Group

| | # Obs | Mean | Median | 25th Pct. | 75th Pct. | Std. Dev. |
|---|---|---|---|---|---|---|
| White | 316,222,955 | 9.56 % | 9.75 % | 3.87 % | 15.39 % | 13.51 % |
| Black | 21,987,465 | 9.67 % | 10.48 % | 3.03 % | 10.52 % | 23.56 % |
| Hispanic | 64,372,476 | 11.32 % | 12.01 % | 3.98 % | 19.75 % | 20.86 % |

Tables 5.3 and 5.4 decompose racial disparities in the return on owner-occupied housing into a rental yield and appreciation in house prices. Table 5.3 shows that racial differences in the rental yields on owner-occupied properties explain a large share of the difference in average returns. Black homeowners get a 1.21% higher average rental yield than White homeowners. This difference is larger than the Black-White difference in expected returns, and table 5.4 shows that Black homeowners would have a 1.15% lower average return than White homeowners without accounting for differences in rental yields. In contrast, even without their .72% higher rental yield, Hispanic homeowners would still get a 1.1% higher average return on housing than housing from price appreciation alone. The greater volatility of housing returns for Black and Hispanic homeowners than White homeowners is due to greater volatility in house price appreciation. The 23.43% Black and 20.65% Hispanic standard deviations are far larger than the 13.22% White standard deviation. In addition the 25th percentile of each racial group's return distribution shows that Black and Hispanic homeowners are exposed to larger potential losses, with a 1.17% loss for White homeowners, 3.51% loss for Black homeowners, and 2.1% loss for Hispanic homeowners.

Table 5.3: Rent-to-Price Ratios Summary Statistics by Racial Group

| | # Obs | Mean | Median | 25th Pct. | 75th Pct. | Std. Dev. |
|---|---|---|---|---|---|---|
| White | 454,275,392 | 6.15 % | 5.88 % | 3.93 % | 7.84 % | 3.05 % |
| Black | 40,080,825 | 7.36 % | 7.44 % | 5.89 % | 8.82 % | 2.37 % |
| Hispanic | 106,802,838 | 6.87 % | 6.70 % | 5.15 % | 8.47 % | 2.58 % |

Table 5.4: House Price Appreciation Rate by Racial Group

| | # Obs | Mean | Median | 25th Pct. | 75th Pct. | Std. Dev. |
|---|---|---|---|---|---|---|
| White | 316,222,910 | 3.93 % | 4.14 % | -1.17 % | 9.32 % | 13.22 % |
| Black | 21,987,464 | 2.78 % | 3.75 % | -3.51 % | 10.52 % | 23.43 % |
| Hispanic | 64,372,476 | 5.03 % | 5.65 % | -2.14 % | 13.27 % | 20.65 % |

We examine the determinants of the racial gap in housing investment returns in table 5.5. Controlling for zip code fixed effects flips the average Black homeowner return from .13% above White homeowners to .12% below. The Hispanic return advantage over White homeowners is also sharply reduced from 1.77% to only .33%. This suggests that minorities tend to own housing in areas with higher returns but that they obtain little or no advantage in investment returns over White homeowners in the same zip code. When we use even stricter house fixed effects instead of zip code fixed effects, White homeowners get a .27% higher return than Black homeowners and .56% higher than Hispanic homeowners. This implies that for a given house, its rate of return tends to be higher when the homeowner is White than Black or Hispanic. However, minority homeowners tend to own homes whose average return is higher when averaged over time. This suggests that for a given house White homeowners have a "market timing" advantage over minority homeowners. To the extent that minority homeowners are credit constrained and buy when credit is easily available and house prices are high, we should expect to see this market timing advantage for White homeowners who may be more likely to buy when credit conditions are tight. In addition, we find like in the AHS that Black and particularly Hispanic homeowners have investment returns that are sensitive to business cycles. Including recession dummies also significantly increases the R-squared of our regression, suggesting that the business cycle is a key determinant of housing returns.

We examine the house price appreciation earned by homeowners who buy and then sell homes based on equation 20 in table 5.6. This table presents summary statistics where each observation is weighted in proportion to the holding period of the homeowner. This makes the results more consistent with our annual rates of return above, since in the property-year panel houses held for several years result in multiple observations. We find lower house price appreciation here for all racial groups than in table 5.4. We also find that White homeowners have a 2% higher price appreciation rate for house purchases than Black homeowners, slightly larger than the 1.21% difference in mean and 1.56% difference in median rental yields between White and Black homeowners. At the very least, including rental yields sharply reduces the difference between White and Black returns on housing. This table also is consistent with our other results in suggesting that Hispanic homeowners have the highest rates of return once rental yields and house price appreciation are both included.

Table 5.5: Determinants of Disparities in Housing Returns across Racial Groups 1980-2019

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Constant | 9.56 % (0) | 9.81 % (0) | 9.92 % (.001) | |
| Black | 0.13 % (.003) | -0.12 (.004)% | -0.27 % (.007) | .45 % (.0036) |
| Hispanic | 1.77 % (.002) | 0.33% (.003) | -0.56 % (.005) | 3.32 % (.0022) |
| Black* Recession | | | | -1.53 % (.0085) |
| Hispanic * Recession | | | | -8.15 % (.0052) |
| Fixed effects | none | zip code | house | none |
| R-squared | .002 | .023 | .060 | .1028 |

The dependent variable of each regression is the annual log-return on housing in equation 19.

Table 5.6: Average Annualized Price Appreciation Rate for Housing Transactions by Race

| Race | # Obs | Mean | Median | Std. Dev. | 25th Pct. | 75th Pct. |
|---|---|---|---|---|---|---|
| White | 65,573,411 | 4.487 % | 3.766 % | 6.841 % | 1.634 % | 6.398 % |
| Black | 3,166,885 | 2.475 % | 2.698 % | 7.718 % | -0.339 & | 5.466 % |
| Hispanic | 5,269,717 | 4.214 % | 4.226 % | 7.988 % | 1.218 % | 7.251 % |

Summary statistics on realized property price appreciation rates using transactions prices as in equation 20. Statistics are weighted by the number of years each property is held.

Broadly, these results are consistent with the idea that financial constraints explain some of the differing experiences of White and minority homeowners. Financially constrained residents with less access to credit are likely to bid up rents but remain unable to purchase large quantities of housing. Because this increases rents and lowers house prices, equilibrium rates of return in places with constrained residents are likely to be higher. Similarly, an expansion of credit supply may lead to a growth in the relative share of minority homeowners while bidding up current prices and lowering future returns. This is consistent with White homeowners tending to own a given house in periods when its return tends to be highest.

# 6 Conclusion

This paper quantifies racial differences in the total rate of return on owner-occupied housing. Our results rely crucially on a new method we develop to impute the rental value of owner-occupied housing that performs well for both low-value and high-value homes. The total rate of return of buying a house equals the appreciation in its price plus the rental value of its housing services minus taxes and costs. To measure the total return, we develop a new estimator of the rental value of each individual owner-occupied house. Unlike existing hedonic methods, our estimator accurately predicts rents of both low and high value homes by using price information to account for house quality. This accuracy across the value distribution is crucial for our application to racial inequality, since White, Black, and Hispanic homeowners vary in the average value of their homes.

With our total return measure, we document a new stylized fact that Black and Hispanic homeowners earn higher average total return on housing than White homeowners. This is largely explained by their higher rental yields as measured by our imputation method. This rental yield disparity is largely driven by a higher rent-to-price ratio for lower value homes, and controlling for either homeowner income or a house fixed effect largely explains the return disparity it creates. However, we also find that Black and Hispanic homeowners face riskier and more cyclical returns than White homeowners. Going forward, we hope to quantify more explicitly the implications of our return measurements for the evolution of racial wealth disparities. Increasing the availability of credit would allow Black and Hispanic homeowners to access a high-risk high-return investment in housing, and our data can help us say more about the costs and benefits of such a credit expansion.

# References


Ellora Derenoncourt, Chi Hyun Kim, Moritz Kuhn, and Moritz Schularick. Wealth of two nations: The U.S. racial wealth gap, 1860-2020. *Quarterly Journal of Economics, forthcoming*, 2023.

John Y. Campbell. Household Finance. *Journal of Finance*, 61(4):1553–1604, 2006.

Amir Kermani and Francis Wong. Racial Disparities in Housing Returns. *working paper*, 2022.

Edward N. Wolff. Heterogenous rates of return on homes and other real estate: Do the rich do better? Do Black households do worse? *working paper*, 2022.

Paul Goldsmith-Pinkham and Kelly Shue. The Gender Gap in Housing Returns . *Journal of Finance*, 78(2):1097–11459, 2023.

Carlos Avenancio-Leon and Troup Howard. The Assessment Gap: Racial Inequalities in Property Taxation . *Quarterly Journal of Economics*, 137(3):1383–1434, 2022.

Oscar Jorda, Katharina Knoll, Dmitry Kuvshinov, Moritz Schularick, and Alan Taylor. The Rate of Return on Everything, 1870–2015. *Quarterly Journal of Economics*, 2019.

Andrew Demers and Andrea Eisfeldt. Total Returns to Single Family Rentals. *working paper*, 2001.

Jonah B. Gelbach. When Do Covariates Matter? And Which Ones, and How Much? *Journal of Labor Economics*, 34(2), 2016.

Arpit Gupta, Christopher Hansman, and Pierre Mabille. Financial Constraints and the Racial Housing Gap. *working paper*, 2023.

Neil Bhutta and Aurel Hizmo. Do Minorities Pay More for Mortgages? . *Review of Financial Studies*, 34(2):763–789, 2021.

Brian Higgins. Racial Segmentation in the US Housing Market. *working paper*, 2023.

Peter Ganong, Damon Jones, Pascal J Noel, Fiona E Greig, Diana Farrell, and Chris Wheat. Wealth, race, and consumption smoothing of typical labor income shocks. *working paper*, 2024.

Kerwin Kofi Charles and Erik Hurst. The Transition to Home Ownership and the Black/White Wealth Gap . *Review of Economics and Statistics*, 84(2):281–297, 2002.

Sasha Indarte, Bronsyn Argyle, Ben Iverson, and Christopher Palmer. Explaining Racial Disparities in Personal Bankruptcy Outcomes . *working paper*, 2023.

Patrick Bajari and C. Lanier Benkard. Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach. *Journal of Political Economy*, 113(6), 2005.

Patrick Bajari, C. Lanier Benkard, and John Krainer. House Prices and Consumer Welfare. *Journal of Urban Economics*, 58(3), 2005.

Cody Cook, Rebecca Diamond, Jonathan V. Hall, John A. List, and Paul Oyer. The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers . *Review of Economic Studies*, 88(5):2210–2223, 2021.

Neil Bhutta. GSE Activity and Mortgage Supply in Lower-Income and Minority Neighborhoods: The Effect of the Affordable Housing Goals. *The Journal of Real Estate Finance and Economics*, 45: 238–261, 2012.

Tim Landvoigt, Monika Piazzesi, and Martin Schneider. The Housing Market(s) of San Diego. *American Economic Review*, 105(4):1371–1407, 2014.

Junting Ye, Shuchu Han, Yifan Hu, Baris Coskun, Meizhu Liu, Hong Qin, and Steven Skiena. Nationality Classification Using Name Embeddings. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, page 1897–1906, 2017.

Rebecca Diamond, Tim McQuade, and Franklin Qian. The effects of rent control expansion on tenants, landlords, and inequality: Evidence from san francisco. *American Economic Review*, 109(9):3365–3394, 2019.

## 7 Data Section: AHS

The data come from the American Housing Survey (AHS). The AHS is a nationally representative panel survey that has been conducted by the U.S. Census Bureau since 1973. The latest survey we use comes from 2021. The AHS follows housing units, rather than individuals or households, and collects information on housing characteristics, housing costs, as well as housing tenure, and demographics of the occupants.

### Construction of dataset

We construct the panel of housing units from the national AHS files by extracting the relevant variables from each national file and linking units by the unique identifier (*CONTROL*) across survey waves. The initial sample was drawn in 1973 and the sample was redrawn in 1974, 1985, and 2015. From 1973 to 1981 the AHS was conducted annually and it has been conducted bi-annually since 1983. Between any adjacent survey waves, excluding those in which samples were redrawn, we distinguish between three cases:

1. The housing unit is present in the sample in both survey waves.
2. The housing unit is not present in one of the two periods but appears in previous (subsequent) periods.
3. The housing unit is not present in one of the two periods and has never appeared before (sample entry) or never appears again (sample exit).

In Case (1) we link the unit across waves using CONTROL. For Case (2) and Case (3) we use information from the Sample Case History File (1985-2013, 2015-2021) to determine why the unit was added or (temporarily) removed from the sample. For 51% of units in Case (2) we have information on why these units were not interviewed, the remaining observations are labelled as "unclear" temporary exits. In contrast to Case 2, Case 3 covers units that permanently exit the sample; for 9% we know the exit reason, the remaining units are labelled as "unclear" permanent exits. The AHS sample is constantly in flux and in each survey wave new units are added to reflect new construction and improvements in sample coverage (more details for each wave can be found here: "Accuracy of Data"). Units that enter are classified as new construction – 9% of entries outside resampling periods – if declared as such by the Sample Case History File, and classified as unclear sample entries otherwise. Note that the Sample Case History Files are only available after 1985. While there is some, albeit limited, information on whether sample additions are new construction, all sample exits prior to 1985 are classified as "unclear" exits.

The AHS variable *TENURE* records the occupancy status of housing units that are in-sample. We consolidate the information into one of three categories: owner-occupied, rented, or vacant.

### Households

While the AHS follows housing units over time, we can infer whether the same group of people, that is, the same household, occupies the housing unit in subsequent waves. This improves our accuracy in imputing missing information on variables that likely change with the occupants such as demographics, and later allows us to cluster standard errors at the household level. We use information from two variables to infer whether a new household moved in since the last survey wave: *SAMEHH* and *MOVE*. *SAMEHH* asks directly whether the unit is occupied by the same household members now as in the previous survey period, but this variable is only available after 1987. To supplement *SAMEHH* and infer household changes prior to 1987, we use information on the reported move-in year. Last, if the information for both variables is missing, we use changes in tenure, race, and/or reported purchase year to determine whether a new household moved in.

| **New houshold identifier** | No. | Col% | Cum% |
|---|---|---|---|
| Same houshold | 922,244 | 59.15 | 59.15 |
| New houshold | 637,006 | 40.85 | 100.00 |
| Total | 1,559,250 | 100.00 | |

| **Flag assignment of new household identifier** | No. | Col% | Cum% |
|---|---|---|---|
| Variables agree | 634,469 | 32.26 | 32.26 |
| Same hh variable | 412,383 | 20.97 | 53.23 |
| Move-in year variable | 570,198 | 28.99 | 82.22 |
| First observation | 343,953 | 17.49 | 99.71 |
| Ownership change | 5,738 | 0.29 | 100.00 |
| Total | 1,966,741 | 100.00 | |

## Race

We combine information from the variables RACE and SPAN to assign the respondent to one of four racial groups: White (non-Hispanic), Black (non-Hispanic), Hispanic, and Other. The same household may have different respondents in different survey waves and the different respondents, in turn, may belong to different racial groups. We replace the racial group of all observations for the particular household, not housing unit, with the majority report for the household.

| **Racial group of HH** | No. | Col% | Cum% |
|---|---|---|---|
| White | 1,261,950 | 76.17 | 76.17 |
| Black | 190,707 | 11.51 | 87.68 |
| Hispanic | 141,925 | 8.57 | 96.25 |
| Other | 62,184 | 3.75 | 100.00 |
| Total | 1,656,766 | 100.00 | |

| **Flag for re-coding of racial group** | No. | Col$_{\%}$ | Cum$_{\%}$ |
|---|---|---|---|
| Reported | 1,835,507 | 93.24 | 93.24 |
| Race replaced with majority household response | 48,114 | 2.44 | 95.68 |
| Unit exited: carried forward racial group | 85,019 | 4.32 | 100.00 |
| Total | 1,968,640 | 100.00 | |

## Housing costs

All monetary values are CPI-adjusted to 2019.

### Property values

Property values in the AHS are directly reported by respondents, but the way they are recorded changes over time.

- Between 1973 and 1983, property values are categorical with bracket sizes that increase from $2,500 for low-value properties to $50,000 for high-value properties. We assign the mean value in each bracket to all properties in the bracket. To determine which value to assign to the top-coded bracket we utilise information on the distribution of property values in 1985 – when property values are first reported on a continuous scale. Specifically, we determine the percentile of properties in the top-bracket in each year before 1985 and compare this percentile to the distribution of property values in 1985. We then adjust the value for the top-coded bracket in each year before 1985 to match the 1985 distribution. Since property values in 1985 are top-coded, in some cases, i.e. when the percentile of properties prior to 1985 surpasses the top-coded percentile in 1985, there is no adjustment and the value assigned to units in the top-bracket remains unchanged.
- After 1985, property values are reported on a continuous scale and top-coded at levels that vary by metropolitan area.

There are some property values that are most likely misreported; take for example a property with reported values of $200,000 in 1985, $20,000 in 1987, and $210,000 in 1989. To correct misreported values we follow a similar strategy as the one used, also for the AHS, by Harding et al. (2022). First, a report is flagged as "unreliable" if two out of following four conditions are met and condition 5 is met:

1. The value exceeds four times the reported mortgage amount or the value is less than 25% of the reported mortgage amount.
2. The value exceeds four times the reported purchase price or the value is less than 25% of the reported purchase price.
3. The purchase price was missing.
4. The value exceeds 1,000,000 or is less than 1,000.
5. The appreciation rate of the unit in the adjacent time periods is higher than the 95th percentile of appreciation rates for reliable reports across the entire sample.

In our computation of the rate of return we assume that units that are owner-occupied in t and rented in t+1 are still owner-occupied at the beginning of t+1. For these units we have to impute property values. We separate units that are owner-occupied in t but not in t+1 into three groups: own to rent, own to sample exit, and own to vacant. For each group we regress property values on year, race, building year cohort, geographic region, and the age of the householder for properties that transition from owner-occupied to the relevant group in the following time period.

### Property taxes

Property taxes are reported on a continuous scale prior to 1985 and reported in steps of $50-$100 thereafter. For reports after 1985, we assign the mean value of each bracket to all observations in the bracket. Property taxes are also top-coded. We predict the property tax in the top-coded bracket using a regression of property taxes on property values and region-by-metro fixed effects. To impute missing reports and impute property taxes for units that are owner-occupied in t but not t+1 we regress property taxes on property value, race, year, building year cohort, and region-by-metro fixed effects.

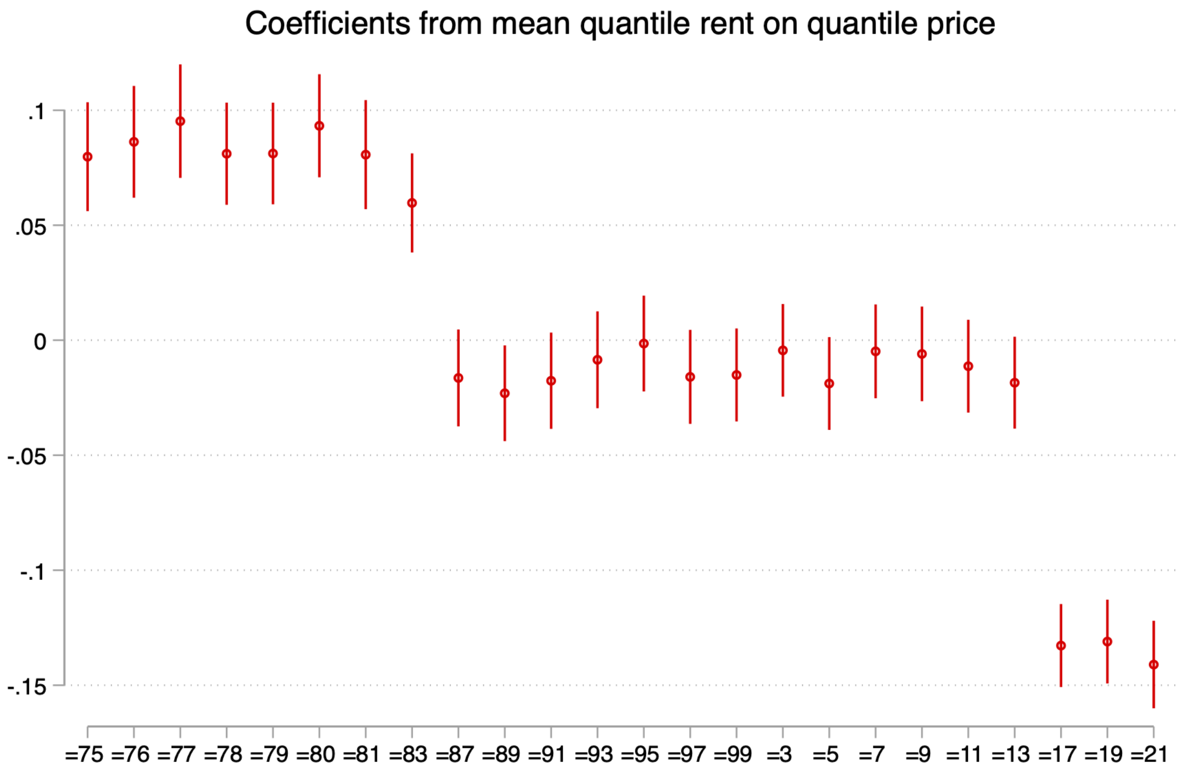
### Maintenance cost

Our variable for maintenance cost includes routine maintenance cost as well as home improvement projects. The AHS contains information on maintenance costs after 1985. Routine maintenance is asked for directly and we determine home improvement costs by aggregating information from a number of variables on cost for home improvement jobs (e.g. for kitchen, bath, etc.), alterations, and additions, over the past two years and halving the aggregate to obtain annual maintenance cost.

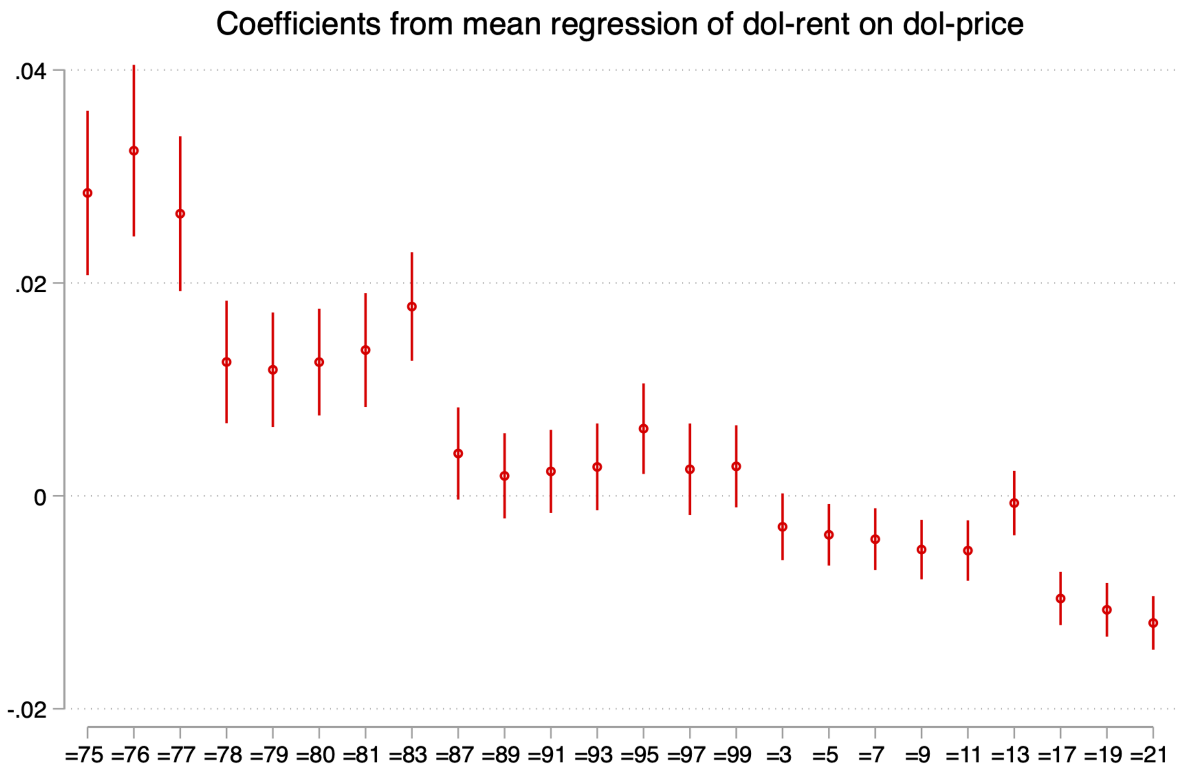
To impute missing maintenance cost for units that are owned in t, we regress maintenance cost on property value, property tax, building age (current year - built year), building age squared, unit size (in square feet), geographic region, and race dummies. The regression equation is also fully interacted with the race dummies. The maintenance cost variable is only available after 1985 so we also impute maintenance cost prior to 1985. To do so, we predict maintenance cost in the last period the unit is owned, using the same regression as described above, and linearly extrapolate backwards using the average change in maintenance cost across the reporting horizon (1985 to 2021). Last, we predict maintenance costs for units that are owner-occupied in t but no longer in t+1 using the same regression of maintenance cost on observables as above.

# 8 Evidence of Copula Stability

This appendix presents plots that analyze how the relationship between rents and prices varies over time. We break our sample at 1985 and 2015, which are times when the data collection process in the AHS changes. We run year-by-year regressions of rent on price in dollar terms. Then, we present regressions of the quantile of rent with that year's marginal distribution on the quantile of price within that year's price marginal distribution. As is clearly visible, the quantile-quantile relationship is stable




within wave of the AHS, while the dollar-dollar relationship reflects a downward trend over time. As a result, our quantile-based method is a more appropriate way to pool data across years when estimating a rent imputation model.


Coefficients from mean regression of dol-rent on dol-price
.04
.02
0
-.02
=75 =76 =77 =78 =79 =80 =81 =83 =87 =89 =91 =93 =95 =97 =99 =3 =5 =7 =9 =11 =13 =17 =19 =21