Regional Science and Urban Economics 39 (2009) 542–552

Contents lists available at ScienceDirect

Regional Science and Urban Economics

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

Alternative measures of homeownership gaps across segregated neighborhoods☆

Paul Carrillo *, Anthony Yezer [1]

Department of Economics, The George Washington University, United States

ARTICLE INFO

Article history:
Received 19 September 2008
Received in revised form 5 March 2009
Accepted 9 March 2009
Available online 24 March 2009

JEL classification:
R21
J15
C15

Keywords:
Homeownership
Race
Quantile regression decomposition

ABSTRACT

The dramatic rise in the U.S. homeownership rate from 64% in 1996 to almost 70% in 2005 has prompted increased attention to the relation between homeownership and demographic characteristics of households. The recent rise and sharp decline of subprime lending will likely spur further interest in homeownership gaps. Statistical analysis of these differences or "gaps" in homeownership between white and minority households has evolved into a highly stylized comparison of differences in homeownership at the mean or the conditional mean. This study implements a quantile decomposition technique that identifies the unexplained portion of the gap not only at the mean, but at every percentile of the homeownership distribution. Results suggest that differences in homeownership gaps at the mean reflect a combination of small differences at the upper end and much larger gaps at the lowest end of the distribution of homeowners. This study also adds credit history to the factors that are used to explain homeownership gaps.

1. Introduction

The dramatic rise in the U.S. homeownership rate from 64% in 1996 to almost 70% in 2005 has prompted increased attention to the relation between homeownership and demographic characteristics of households. The homeownership gap between white non-Hispanic and minority households narrowed during this period. Special government efforts such as the American Dream Downpayment Act of 2003 may have an additional effect on these differences. The recent rise in foreclosures and changes in mortgage lending criteria will likely result in further movement in homeownership gaps.

Discussion of homeownership gaps in the literature has evolved over time and become highly stylized. First simple mean homeownership rates of different groups were compared. Second mean homeownership rates at various quintiles of the distribution of a single characteristic such as income or age were analyzed.[2] Third differences in the conditional mean homeownership rate adjusted for determinants of homeownership other than ethnicity were compared. In essence a tenure equation was estimated with the size and significance of dummy variables for minority status providing the basis for measuring gaps.[3] Fourth, the Oaxaca–Blinder technique, sometimes modified because probit models are non-linear using Fairlie's (2005) approach,[4] has been used to decompose differences in the conditional mean homeownership gap into a component that is due to differences in determinants of homeownership between white and minority groups and differences in the conditional mean that remain even when minority homeownership is evaluated using coefficients from a white tenure equation.[5] Fifth, Gabriel and Rosenthal (2005) modified the Oaxaca–Blinder technique to decompose changes in homeownership gaps over time into one component

☆ We appreciate helpful suggestions by two referees and the editor. An earlier version of this paper was presented at the ASSA meetings of AREUEA in New Orleans, Louisiana, January, 2008, and at the Southern Economic Association Annual Conference in Washington DC, November, 2008. We wish to thank Jim Albrecht, Nathan Anderson, Brian Bucks, Bryan Boulier, Tara Sinclair and the conference participants for helpful comments. Special thanks are due to Glenn Canner of the Board of Governors who played a vital role in arranging access to data on credit history. Contact either author with correspondence regarding this paper.

* Corresponding author. Monroe Hall, 2115 G Street NW Suite 364, Washington DC, 20052, United States. Tel.: +1 202 994 7608.
E-mail addresses: pcarrill@gwu.edu (P. Carrillo), yezer@gwu.edu (A. Yezer).
[1] Tel.: +1 202 994 6755.

[2] See particularly the approach in Christopher E. Herbert and Bulbul Kaul, "The Distribution of Homeownership Gains During the 1990's Across Neighborhoods," January 2005, U.S. Dept of HUD, Report. Based on their survey of the related literature, the authors conclude that differences in income, wealth, marital status, and age of the household are found to account for between 15 and 20 percentage points out of the total racial gap of roughly 25 percentage points.

[3] The pioneering studies using this technique are Kain and Quigley (1972), and Roistacher and Goodman (1976).

[4] More recently, Yun (2007) has proposed an extension of the Oaxaca decomposition using generalized residuals that can be implemented for linear or non-linear estimation techniques in the presence of endogeneity. Presumably this extension will be applied in the homeownership literature soon.

[5] See, for example, Coulson (1999), Painter et al. (2001), Wachter and Megbolugbe (1992), Myers and Chung (1996).

0166-0462/$ – see front matter
doi:10.1016/j.regsciurbeco.2009.03.002

due to changes in household characteristics and another due to structural parameter changes. Sixth, dynamic homeownership changes have been traced in studies that evaluate homeownership differentials over the life cycle.[6]

These six approaches have all proved very useful in advancing the understanding of the homeownership gap but all are based on differences at the mean or the conditional mean of the households being studied. While findings differ slightly, studies conducted using techniques three, four, and five generally conclude that 65–80% of the homeownership gap between white and black households is due to differences in endowments (income and wealth) and household characteristics (age, marital status, etc).

Overall, the current state of the homeownership gap literature appears similar to the male–female or white–minority wage gap literature before recent advances in statistical techniques. Labor economists have begun using quantile regression to identify the differences in the conditional mean of wage rates at different points in the distribution of wages for employed male and female workers. This approach identifies the wage gap for workers with different levels of human capital. Then they apply a technique proposed by Machado and Mata (2005) which essentially performs an Oaxaca–Blinder decomposition across the entire wage distribution and partitions the wage gap into a component due to differences in endowments evaluated based on the male return to human capital and an unexplained wage gap between males and females. The results obtained using the Machado–Mata technique rather than Oaxaca–Blinder decomposition for measuring wage gaps have been dramatic and led to the "glass ceiling" finding that small differences at the mean often conceal very large differences at the upper end of the wage distribution.[7] Specifically, studies find that the male–female wage gap in some countries is negligible at the lower end of the wage distribution and rises exponentially in the highest quintile resulting in a "glass ceiling" effect on the most skilled women.[8]

The goal of this research is to develop and illustrate the application of a method for adapting the Machado–Mata technique to the measurement of homeownership gaps. The principal challenge in this effort is that homeownership is a binary variable while wages and income are continuous – or at least they are continuous conditional on employment.[9] The requirement that the dependent variable be continuous, forces some changes in the structure of the test conducted here compared to those in the labor literature. Specifically, the dependent variable is the fraction of owner-occupants in highly segregated census block groups (CBG). The fraction of owner-occupants is continuous and confining the analysis to highly segregated CBGs, where the percentage white is either near to 100% or close to zero, allows us to characterize areas as essentially white or non-white. Working with highly segregated CBGs thus allows us to apply the Machado–Mata techniques to the homeownership gap question allowing the characterization and decomposition of racial gaps across the overall *distribution* of homeownership rates. This is the main advantage of the suggested approach.

Changing from probit estimates of individual tenure decisions to OLS estimates of the fraction owner-occupied from segregated CBGs has two principal effects on the test for homeownership gaps. First, a fraction of the population is excluded from the test data because they do not live in highly segregated CBGs. Second, the average characteristics of CBGs are used as arguments of the tenure equation rather than individual characteristics. Because information on credit scores is available at the CBG level, we are able to add variables reflecting credit history that have been previously neglected in the homeownership gap literature to the analysis.

Given that the test proposed here for CBGs appears to differ substantially from tests using individual tenure models, results may not be readily comparable with the previous literature. The first task thus will be to determine if the approach using segregated CBGs produces results comparable to the individual tenure models. This is done by testing to see if analysis of homeownership gaps using the CBG data on fraction owner-occupied produces estimates of the explained and unexplained gap using a standard Oaxaca–Blinder decomposition that are similar to the results found in the literature using individual tenure data and probit models. The results of this test for the size and decomposition of the white–minority homeownership gap are roughly comparable to those from the current literature using individual tenure models.

Having reproduced results obtained elsewhere for the nature of the homeownership gap using the CBG data, the next step is to study the distribution of these gaps by estimating quantile regressions and implementing a Machado–Mata decomposition to determine if the gap found at the conditional mean of the sample is representative of the homeownership gap throughout the distribution. As expected based on the wage gap literature, the conditional homeownership gap varies dramatically across the distribution of census blocks which differ in the probability of homeownership. Specifically, differences in the unexplained portion of the conditional homeownership gap are large and positive at the lower end of the homeownership distribution, and the unexplained gap disappears entirely at the upper end of the distribution.

One could argue that results may depend on the specific definition of segregated areas. To test if this is the case, several robustness tests are performed to determine if the results vary with the criteria used to identify segregated census blocks. The findings appear quite robust leading to the conclusion that further investigation into and policy directed toward the white–black homeownership gap should be directed toward those areas where the overall likelihood of homeownership is lowest.

The rest of this paper is organized as follows. The next section introduces the definition of segregated areas, provides details about the data, and shows standard Oaxaca–Blinder mean decompositions. The third section computes the distribution of homeownership rates across white and minority CBGs and estimates the unconditional and conditional homeownership gap throughout the distribution. Robustness tests are reported in the fourth section and the last section concludes.

2. Data and measurement issues

The first step in analyzing differences in tenure rates between segregated areas is to provide a criterion for identifying a racially segregated neighborhood. While there may be several ways to define such an area, a simple rule is adopted here. "white" neighborhoods are CBGs[10] where the share of white population is above a threshold R, where $0<R<1$.[11] Similarly "non-white" neighborhoods are CBGs where the share of white population is below $1-R$. The focus here is on medium and large urban areas. Thus, data from all CBGs in

[6] See, for example, the application of this technique in Myers and Lee (1998), Myers et al. (1998), and Myers et al. (2005).

[7] Applications of the Machado–Mata technique include studies that explain wage (Albrecht et al., 2003, and Arulampalam et al., 2007), income (Nguyen et al., 2007) and housing price (McMillen, 2008) differentials.

[8] The male-female wage gap is relatively uniform for the U.S. but, for some European countries such as Sweden, it is negligible at the lower end and very large at the upper end of the wage distribution giving rise to what has been termed the "glass ceiling" effect on the most skilled women.

[9] In the earnings literature, the complication is that wages of those not currently employed must be imputed. For homeownership studies, households are either renters or owners. Studies routinely exclude individuals in institutionalized settings and this practice is followed here.

[10] The US 2000 Census divides the country in about 210,000 CBGs. A CBG is the smallest geographical area available for which a large set of demographic and income characteristics are available to the public.

[11] "whites" are defined as non-Hispanic white individuals; based on the variable P7.3 from the US Census.

Table 1
Homeownership rate in segregated neighborhoods.

R	Area type	Mean homeownership rate		Number CBGs	Total population (millions)
		Unweighted[a]	Weighted[b]		
$R=0.80$	White	0.773	0.782	77,498	107.3
	Non-white	0.479	0.472	26,839	36.5
$R=0.90$	White	0.806	0.816	50,240	67.2
	Non-white	0.474	0.463	19,770	25.7
$R=0.95$	White	0.823	0.832	28,495	35.7
	Non-white	0.476	0.469	14,900	18.4
$R=0.97$	White	0.827	0.836	17,520	20.6
	Non-white	0.483	0.481	12,017	14.3
$R=0.99$	White	0.828	0.838	7,650	8.4
	Non-white	0.485	0.493	6,647	6.6

Note: A segregated "white neighborhood" is defined as a Census Block Group(CBG) where the share of white population exceeds R.
Accordingly, a "non-white" neighborhood is a CBG where the share of white population is below $1-R$.
[a] Mean rates are average homeownership rates in each group of CBGs.
[b] Mean rates weighted by the total population of each CBG.

Metropolitan Statistical Areas (MSA) that have at least one hundred thousand residents was used to identify segregated areas.[12]

Clearly, the characteristics of segregated areas depend upon the chosen threshold. For high values of *R*, however, the average homeownership rate in segregated areas does not vary significantly as this threshold changes. For instance, Table 1 contains descriptive statistics of white and non-white areas for five different choices of *R*. The rows show characteristics of white and non-white neighborhoods when the threshold *R* equals 0.80, 0.90, 0.95, 0.97, and 0.99. In all specifications, the (simple) average homeownership rate in non-white CBGs is close to 48%. In white CBGs, average homeownership rates range between 81 and 82% when *R* exceeds 0.90. In addition, notice that the population-weighted mean homeownership rates are also invariant to the choice of *R* and that these are very close to the simple averages.

As the threshold rises, the number of CBGs and the population in the sample decrease significantly. For example, if *R* increases from 0.95 to 0.99, the sample of white CBGs is reduced by almost 75%. This provides an incentive to pick a low threshold. It is possible, however, that low values of *R* may not accurately describe a segregated neighborhood. For purposes of illustration, in the main portion of this paper, the intermediate threshold of 0.97 is chosen. Using this definition, there are 17,520 white and 12,017 minority CBGs. In the fourth section of the paper, robustness checks are performed using all five levels of *R*. The major findings of the paper regarding the distribution of the homeownership gap are unchanged across this wide range of choices for *R* and consequently across a wide range of sample sizes. Similar robustness checks should be performed whenever the methods for modeling ownership gaps proposed here are used.

Given the decision to set $R=0.97$, differences in homeownership between the segregated CBGs in this sample can now be analyzed and compared to the entire population. The average homeownership rate in the segregated white neighborhoods is about 34 percentage points higher than in the non-white CBGs. This estimate is somewhat larger than the 25 percentage point homeownership gap reported in studies using individual household data.[13]

Research has documented that homeownership gaps are largely explained by differences in the economic circumstances and structure of households. Typically, tenure equations have been estimated that incorporate several economic and household controls, and it has been established that differences in endowments and household structure account for a large portion of the gap.[14]

It is expected that white and non-white areas differ in other characteristics besides their racial composition and homeownership rates. The extent to which other characteristics of the neighborhood residents explain the mean homeownership gap can be explored using standard regression methods. The previous literature provides guidance for identifying the set of controls included in the specification. In particular, tenure equations in previous studies have included independent variables such as age, marital status, income and wealth, education, immigration status and access to credit.[15] Note that these are characteristics of households, not characteristics of the neighborhood or housing market. It may be that local school quality or rent-to-value ratios influence area homeownership but these are characteristics of the area not the household.[16]

The explanatory variables used in this study are gathered from two data sources. Demographic characteristics for each census block group, such as its population density, median income and unemployment rate, are collected from the US Decennial Census. To control for credit accessibility, information about the credit history of the residents in our selected areas has been provided by Equifax, one of the three national credit reporting agencies. Although it is aggregated at the census tract level, the credit data is notably rich. For instance, information about the share of consumers in credit files without a credit VantageScore as well as some information on the distribution of VantageScores in each census tract is available. Unfortunately, the credit data is not available for every census block group in the census. Thus, when the two datasets are matched, about 17% of the observations are dropped from the sample.[17] Descriptive statistics for both white and non-white areas as well as for those CBGs in non-segregated areas are found in Table 2.

White and non-white areas are substantially different. For instance, the mean population density of a white CBG is about one ninth than that of its counterpart in a non-white neighborhood. There are also important differences in income and employment between white and non-white areas. For example, the mean of the median household income in white CBGs is more than twice as large as the mean of the median income in non-white CBGs. Furthermore, white areas tend to have a higher proportion of people older than 65, and the mean unemployment rate is almost five times higher in minority neighborhoods. Finally, notice that the characteristics of both segregated groups differ from those of non-segregated areas. This raises the issue of potential sensitivity of the test results to the choice of $R=0.97$ that will be explored through robustness testing in the penultimate section of this paper.

Following the general practice in the literature, a linear model is used to explore how the characteristics of neighborhoods "affect" mean homeownership rates. The dependent variable is the aggregate homeownership rate in each CBG, and the explanatory variables include those described in Table 2. Table 3 contains estimation results.

The first four columns in Table 3 show estimates of pooled regressions that use all CBGs in white and non-white areas. In the first column, the set of explanatory variables includes only demographic

[12] There are 161,560 CBGs in these selected areas. Areas with missing homeownership rates were dropped (in other words, areas with no population were excluded).

[13] For example, see Coulson (1999), Painter et al. (2001), Deng et al. (2003), and Gabriel and Rosenthal (2005), among others. Notice, however that our estimate of the gap is not fully comparable with the previous literature for several reasons. First, rather than using individual level surveys, this analysis relies on aggregate data. More importantly, the sample of CBGs is not representative of the US population because of the ad-hoc definitions of white and non-white areas.

[14] Haurin et al. (2007) provide a comprehensive survey of the related literature. They report that differences in income, wealth, marital status, and age of the household are found to account for between 15 and 20 percentage points out of the total racial gap of roughly 25 percentage points.

[15] See, for example, Coulson (1999), Painter et al. (2001), Wachter and Megbolugbe (1992), and Gabriel and Rosenthal (2005).

[16] To the extent that discrimination influences homeownership gaps through its effect on residential location including segregation, it is important that the conditional estimation of tenure not include such neighborhood characteristics and, instead, be based on the physical, educational, and economic circumstances of the household.

[17] Out of the 161,560 CBGs in large-and-medium MSAs, 125,121 can be matched with credit score data. When the 29,537 CBGs in the sample of segregated areas are matched with the credit data, 4921 (about 17%) observations are dropped.

Table 2
Descriptive statistics (mean and standard deviation).

Variables	Segregated CBG		Non-segregated CBG
	White	Non-White	
HO rate: Number of owner-occupied housing units divided by the total occupied units in Census Block Group (CBG)	0.827 (0.145)	0.446 (0.267)	0.653 (0.2261)
Density: Population density (total population per hectare)	8.552 (15.534)	79.783 (116.205)	27.897 (57.537)
Older 65: Share of population older than 65	0.163 (0.110)	0.116 (0.075)	0.127 (0.084)
Family: Share of family households	0.733 (0.123)	0.709 (0.137)	0.679 (0.163)
Married: Proportion of population (15 and older) who is married	0.631 (0.099)	0.373 (0.116)	0.545 (0.137)
HS dropout: Share of population (25 and older) that does not have a High School diploma	0.142 (0.097)	0.392 (0.155)	0.190 (0.141)
Median income: Median household income in CBG ($ thousands) in 1999	52.597 (21.806)	26.070 (12.697)	47.930 (23.717)
Unemployment rate: Share of population (16 and older) in the labor force that are unemployed	0.038 (0.036)	0.162 (0.103)	0.058 (0.059)
Bad English: Share of population (between 18 and 64 years old) that does not speak English well	0.004 (0.012)	0.067 (0.122)	0.044 (0.079)
Low credit score: Share of consumers in credit files in census tract with a VantageScore below 640 points	0.163 (0.075)	0.436 (0.072)	0.237 (0.117)
No credit score: Share of consumers in credit files in census tract without a VantageScore	0.061 (0.035)	0.223 (0.086)	0.106 (0.075)
Nobs: Number of observations (CBG)[a]	16,433	8,183	100,505

[a] We drop CBGs with missing credit score data from our sample. Thus, the number of CBGs shown in this Table is smaller than the ones reported on Table 1.

characteristics. The second and third specifications include variables that explain economic and migration status, respectively.[18] In all equations most of the coefficients are significant and have the expected sign. For example, areas with a higher share of married individuals and family households have larger homeownership rates. In addition, household income is positively correlated with homeownership rate, and areas with large shares of high school dropouts have lower ownership rates.

An innovation of this study is the addition of credit history information to the factors that explain homeownership gaps. In the fourth column of Table 3, credit history variables are added to the previous specification. Estimates suggest that the fraction of households lacking sufficient information to formulate a score is associated with lower homeownership rates. The effect is both statistically and economically important. This is an unsurprising result to those familiar with the previous homeownership literature.[19] The share of individuals with a VantageScore below 640 has an uneven relation to homeownership with an estimated coefficient that varies in both sign and statistical significance. This result is surprising and might be interpreted as an indication that lower credit scores were not a significant impediment to homeownership. Alternatively, there may be an issue of timing because current credit score may not reflect credit score at the time when the individual became a homeowner. The variable "white" equals one if the CBG is a white neighborhood and measures the unexplained portion of the conditional mean homeownership gap. As with studies using individual household data, the unexplained portion of the gap decreases as relevant explanatory variables are added. When demographic covariates are included the average (conditional) gap decreases to 15%. When variables that account for economic and migration status are added, the average unexplained gap decreases to 7 and 5%, respectively. Moreover, once the full set of controls is included, the conditional mean gap virtually disappears (4%). This result is consistent with other findings in the literature using individual tenure equations and suggests that the unexplained portion of the homeownership gap is very small.

The pooled regressions shown on the first four columns of Table 3 assume that the marginal effects of neighborhood characteristics on homeownership are the same in white and non-white neighborhoods. In the fifth and sixth columns this assumption is relaxed and separate regressions for each group are estimated. Results suggest that marginal effects of neighborhood characteristics on homeownership rates are substantially different for each area.[20] For this reason, the Oaxaca–Blinder decomposition is used to estimate the unexplained portion of the homeownership gap. That is, the estimated coefficients in column (5) are used to predict the mean homeownership rate that would prevail in white areas if they had average non-white endowments and household structure. The predicted rate is 0.56 and suggests that differences in endowments and household characteristics can explain a large fraction, $0.69 = (0.827 - 0.56)/(0.827 - 0.447)$, of the mean differences in homeownership rates between white and non-white CBGs. This result is within the 65–80% range reported in a recent literature review of Oaxaca–Blinder studies of tenure choice equations by Haurin et al. (2007). This demonstrates that the difference between the simple mean of a homeownership gap and the same gap measured using an Oaxaca–Blinder decomposition of a tenure equation based on individual household data is similar to the difference between the mean homeownership gap and an Oaxaca–Blinder decomposition with the CBG data used in this study. Accordingly, it appears that this examination of the uniformity of the homeownership gap across the distribution of gaps using the CBG data provides useful information about the general issue of the distribution of the homeownership gap in the literature using individual tenure models.

[18] To account for migration status one could add the percentage of recent migrants to our specification. However, there is a high level of correlation between the percentage of residents who do not speak English well and the share of recent immigrants. To avoid multicollinearity problems the latter variable was not included. Main results of this paper are virtually unchanged if this variable is added.

[19] Current credit score is presumably the cumulative result of past behavior that may have occurred some time ago and even in a different location. The lack of sufficient history to formulate a score presumably reflects longer run characteristics of the individual that should be a significant impediment to homeownership.

[20] With the exception of the credit, high school dropouts, and bad English variables, the differences between the coefficients in columns (5) and (6) are statistically significant. The null joint hypothesis that all coefficients in columns (5) and (6) are equal can be rejected at virtually any significance level (the value of the F-test is 69.87).

3. Differences in the distribution of homeownership gaps

The previous section used CBG data to construct measures of the homeownership gap based on the mean, or the conditional mean adjusted using an Oaxaca–Blinder decomposition. With all this as background, it is appropriate to begin the promised innovation in this study, the determination and decomposition of the *distribution* of the homeownership gap.

3.1. Unconditional differences[21]

Fig. 1 displays the distribution of homeownership rates across CBGs for both white and non-white neighborhoods.[22] Note that, in the bottom 10% of white neighborhoods homeownership rates are below 63%, while in the bottom 10% of non-white CBGs homeownership rates are below 7%. Thus, the homeownership rate gap at the 10th percentile is 56 percentage points. Fig. 2 displays the corresponding gap at each percentile of these distributions. Interestingly, the gap reaches its maximum at about the 10th percentile and decreases monotonically at a constant rate across the higher percentiles. For instance, it decreases to 40, 27 and 17 percentage points at the 50th, 75th, and 90th percentiles, respectively.

These findings suggest that a large portion of the average racial homeownership gap between segregated neighborhoods is driven by differences in CBGs at the left tails of the distributions.[23]

Fig. 2 measures the unconditional homeownership gap. To assess what fraction of the gap can be explained by differences in endowments across these segregated areas, quantile regressions are used.

3.2. Quantile regressions

Quantile regressions can be used to assess what fraction of the homeownership gap observed in Fig. 2 remains after adjusting for the effects of differences in endowments across these segregated areas.[24] Quantile regression is a method to estimate the conditional quantile of a variable. Traditional quantile regression models assume that the conditional quantile of a random variable y is linear in the regressors X

$$Q_\theta[y|X] = X\delta_\theta, \tag{1}$$

where $Q_\theta\,[y|X]$ is the θth conditional quantile of y, and the coefficients δ_θ measure the effects of the covariates at the θth conditional quantile.[25] Estimation of the quantile parameters, the δ_θ, is performed as the solution to

$$\underset{\delta_\theta}{arg\,min}\left\{\sum_{i:y_i > X_i\delta_\theta}\theta|y_i - X_i\delta_\theta| + \sum_{i:y_i \le X_i\delta_\theta}(1-\theta)|y_i - X_i\delta_\theta|\right\}. \tag{2}$$

Quantile regression models were introduced by Koenker and Bassett (1978). There have been many applications of quantile regression models in the literature including recent studies such as McMillen and Thorsnes (2006), Nguyen et al. (2007), Albrecht et al. (2003), Bassett and Chen (2001), and Gyourko and Tracy (1999). Buchinsky (1998) and Koenker and Hallock (2001) present useful surveys.

The main advantage of the quantile regression model is that each point (quantile) of a conditional distribution can be characterized. More importantly, a set of quantile regressions can provide a more complete description of the underlying conditional distribution compared to other mean-based estimators (for example, OLS). For this reason, these models are particularly useful when the conditional distribution does not have a "standard" symmetric shape as suggested by the distribution of homeownership gaps displayed in Fig. 2.

The form of the estimated quantile regressions is:

$$Q_\theta[y|Z,W] = \alpha_\theta + \beta_\theta W + Z\gamma_\theta, \tag{3}$$

where y is the homeownership rate, W is an indicator for a white neighborhood, Z is a vector of neighborhood characteristics described in Table 2, and $Q_\theta\,[y|Z,W]$ is the θth conditional quantile of y. The estimated coefficients γ_θ measure the effects of the neighborhood's characteristics at the θth conditional quantile and estimates of the parameter β_θ represent the homeownership gap at the corresponding quantile.

Table 4 contains estimates of β_θ for different quantiles and specifications of the homeownership equation. Each column represents a particular quantile and each row a different specification. The first row of this table includes only a constant term in addition to the "white" indicator W. In the second row, several demographic variables have been added to the previous basic specification. The third, fourth, fifth and sixth rows incorporate education, income, immigration and credit variables, respectively.

The estimates of β_θ in the first row are, by construction, equivalent to those depicted in Fig. 2. This coefficient decreases significantly for every quantile as explanatory variables are added. For example, it diminishes from 0.41 to 0.043 for the median case ($\theta = 0.5$) suggesting that the median homeownership gap between white and non-white CBGs can be almost fully explained by differences in their observed characteristics. The gap also disappears for higher quantiles as the set of controls increases. However, there remains a sizable and statistically significant difference at the left tail of the distribution. Presumably this reflects the effects of factors other than the measured differences in endowments used as explanatory variables in this study.

Table 5 displays estimates of every coefficient in Eq. (3). Notice that at every quantile the sign of the parameters is similar to the OLS estimates. In particular, the estimates for the median regression are remarkably close suggesting that there may be little difference between the (pooled) conditional mean and conditional median homeownership rate.

So far, the discussion has assumed that the relationship between homeownership rates and CBG characteristics is the same in both white and non-white neighborhoods so that the conditional difference was captured by the estimated coefficient of the white dummy variable. To

[21] McMillen and Singell (2008) use a similar approach to compare changes in the distribution of district-level real expenditures per student and class sizes over time.

[22] Given the large sample size, standard errors for the empirical cdfs are substantially small and, thus, not reported.

[23] One may argue that the estimated gap displayed in Fig. 2, does not take into consideration the population of each CBG. For instance, if the population of whites and non-white areas was not uniformly distributed across their CBGs, Fig. 2 may be an inaccurate representation of the overall homeownership gap between these two groups. To assess if this is the case, a population-adjusted homeownership gap was computed as follows. For both white and non-white areas, CBGs were ranked according to their homeownership rate and the cumulative share of the population who reside in them was computed. Then a population-adjusted homeownership gap was computed and compared with the one displayed in Fig. 2. No significant differences were found. For details, please contact either author.

[24] Homeownership is a variable that is naturally bounded between 0 and 1. Thus, differences in the distribution of homeownership rates between two groups should be close to zero at either very low or very high quantiles. This is evidenced in Fig. 2. Notice, however, that even at the first and 99th percentiles, these differences are positive. Quantile regressions allow one to decompose this unconditional gap into a portion that can be "explained" by observed characteristics and other "unexplained" factors at any chosen quantile. In our application, this means that quantile regression methods could be used to decompose gaps at percentiles even as small (big) as one (99).

[25] A detailed introduction to quantile regression models can be found in Koenker (2005).

Table 3
Determinants of mean homeownership rates.

Independent variables	(1) Pooled	(2) Pooled	(3) Pooled	(4) Pooled	(5) White	(6) Non-white
Constant	−0.058***	−1.208***	−1.315***	−1.282***	−0.671***	−1.688***
	(0.011)	(0.046)	(0.046)	(0.051)	(0.056)	(0.084)
White	0.156***	0.067***	0.049***	0.039***		
	(0.006)	(0.005)	(0.006)	(0.007)		
Density	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Older 65	0.366***	0.409***	0.385***	0.388***	0.256***	1.025***
	(0.015)	(0.015)	(0.015)	(0.015)	(0.016)	(0.047)
Family	0.514***	0.437***	0.415***	0.409***	0.544***	0.456***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.021)	(0.026)
Married	0.478***	0.261***	0.355***	0.334***	0.198***	0.400***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.025)	(0.033)
HS dropout		−0.116***	0.011	0.031**	−0.042**	−0.025
		(0.014)	(0.015)	(0.016)	(0.018)	(0.026)
Log median household income		0.134***	0.141***	0.140***	0.088***	0.170***
		(0.005)	(0.005)	(0.005)	(0.006)	(0.007)
Unemployment rate		−0.168***	−0.198***	−0.169***	0.066	−0.110***
		(0.026)	(0.026)	(0.026)	(0.050)	(0.029)
Bad English			−0.440***	−0.437***	−0.491***	−0.378***
			(0.024)	(0.024)	(0.101)	(0.032)
Low credit score				0.107***	0.010	−0.001
				(0.018)	(0.024)	(0.038)
No credit score				−0.315***	−0.126**	−0.214***
				(0.023)	(0.051)	(0.030)
Observations	24,616	24,616	24,616	24,616	16,433	8183
R-squared	0.673	0.727	0.736	0.740	0.508	0.575

Note: The dependent variable in each OLS regression is the homeownership rate in a CBG. Definitions and descriptive statistics of the explanatory variables are found on Table 2. Standard errors are in parentheses. **, and *** denote significance at the 5, and 1 percent level, respectively. With the exception of the credit score, high school dropouts, and bad-English variables, the differences between the coefficients in columns (5) and (6) are statistically significant. The null joint hypothesis that all coefficients in columns (5) and (6) are equal can be rejected at virtually any significance level (the value of the *F*-test is 69.87).

test for heterogeneous effects, separate quantile regressions are estimated for each group and results are shown in Table 6. Clearly, there are important differences.[26] For example, at every considered quantile, the share of married households in a CBG "explains" a larger portion of homeownership rates in non-white neighborhoods. In addition, the coefficient on income in the white areas is significantly smaller than in the non-white counterparts.

Because all the quantile coefficients differ between white and non-white areas, the estimates of β_θ in Eq. (2) are sum of two different effects, one due to unexplained differences associated with race and the other due to differences in the effects of endowments on homeownership. Put, another way, this is the same issue that prompted use of the Oaxaca–Blinder decomposition in studies of differences in the conditional mean. To address this problem in the case of comparisons across the distribution, the decomposition suggested by Machado and Mata (2005) is employed.

3.3. Machado–Mata decomposition

The Machado–Mata decomposition is used here to identify the fraction of the homeownership gap that remains unexplained at several quantiles of the homeownership distribution. This decomposition is based on quantile regression techniques and is similar in spirit to the Oaxaca–Blinder technique (e.g., Oaxaca, 1973; Blinder, 1973) which identifies the sources of the differences between the means of two distributions. The advantage of the Machado–Mata method is that it allows us to evaluate the sources of the differences between the white and the non-white homeownership distributions at each quantile.

As noted above, applications of the Machado–Mata technique include studies that explain wage (Albrecht et al., 2003, and Arulampalam et al., 2007), income (Nguyen et al., 2007) and housing price (McMillen, 2008) differentials. The method generates a counterfactual distribution, for example, the distribution of homeownership rates in white neighborhoods if they had the observed characteristics of non-white areas, and compares it with the actual distribution, that is, with the observed distribution of homeownership rates in non-white areas. The differences between the counterfactual and the actual distribution may be computed at every quantile and used to identify the fraction of the homeownership gap that cannot be explained by differences in endowments.

Application of the Machado–Mata decomposition to the homeownership gap measurement problem proceeds as follows. Define Z^W and Z^{NW} as the observed characteristics of white and non-white CBGs, respectively. Furthermore, let γ_θ^W be the coefficient of the θth conditional quantile regression of homeownership rates in white

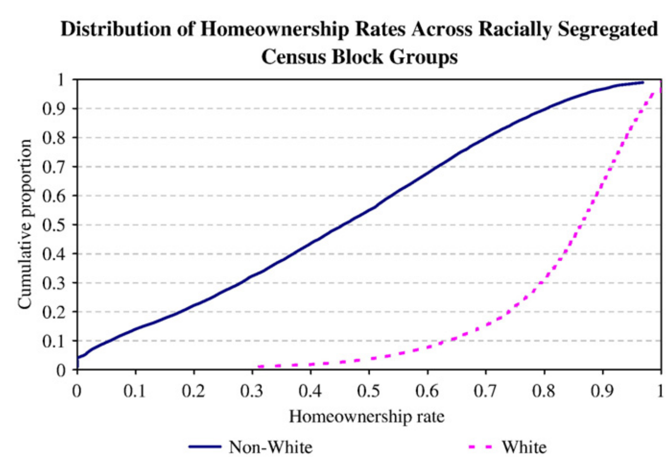

Fig. 1. Distribution of homeownership rates across racially segregated census block groups.

[26] In all cases, Wald tests reject the null hypothesis that the coefficients are equal.

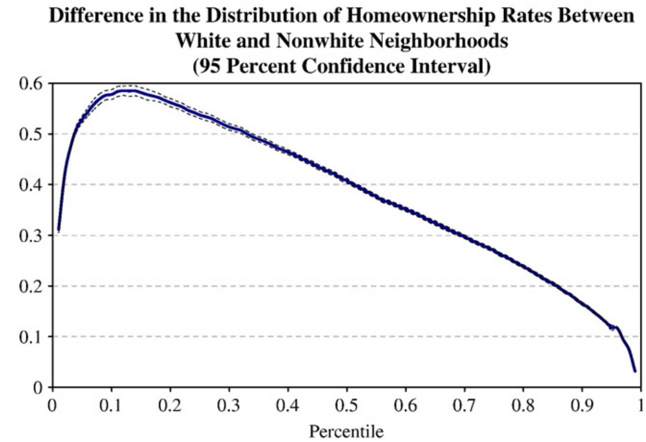

Fig. 2. Difference in the distribution of homeownership rates between white and non-white neighborhoods (95% confidence interval).

neighborhoods; that is, $Q_\theta\ [y^W|Z^W]=Z^W\ \gamma_\theta^W$. The counterfactual distribution is generated as follows:

1. Pick n equally spaced quantiles $\{\theta_i^*\}$, $i=1,..,n$; for example: $\{\theta_i^*\}=\{0.01,0.02,\ldots,0.99\}$.[27]
2. Use the sample of white neighborhoods to estimate $\gamma_{\theta\ i}^{W}{}^*$, $i=1,..,n$.
3. For each quantile, randomly select M draws (with replacement) from the non-white sample denoted z_{ij}^{NW}, where $i=1,\ldots,n$, and $j=1,...,M$.
4. Compute the counterfactual as $\{y_{ij}^*=z_{ij}^{NW}\gamma_{\theta\ i}^{W}\}$, $i=1,\ldots,n$, $j=1,...,M$.

The decomposition can be done for any quantile as follows. Let Q_θ $[y^W]$, Q_θ $[y^{NW}]$, and Q_θ $[y^*]$ be the θth quantile of the white, non-white, and counterfactual distributions, respectively. Then,

$$Q_\theta\left[y^W\right]-Q_\theta\left[y^{NW}\right]=\left(Q_\theta\left[y^W\right]-Q_\theta\left[y^*\right]\right)+\left(Q_\theta\left[y^*\right]-Q_\theta\left[y^{NW}\right]\right).$$

The first term in parenthesis is the component of the gap that can be explained by differences in endowments. The second term measures the "unexplained" portion of the homeownership gap. Albrecht et al. (2009) have shown that the decomposition is consistent and asymptotically normal.[28]

The Machado–Mata method is used to estimate Q_θ $[y^*]$, the counterfactual distribution of homeownership rates that would exist if white neighborhoods had non-white endowments. All the variables previously considered in the OLS and quantile models are included. The predicted mean homeownership rate of this counterfactual distribution is close to 0.61 which is 5 percentage points higher than the one obtained using the Oaxaca–Blinder decomposition (0.56). The differences in these estimates can be explained by the assumptions of the two methods. For instance, the traditional Oaxaca–Blinder decomposition assumes that the conditional expectation is linear. On the other hand, the Machado–Mata method makes no explicit assumption about the functional form of the conditional expectation but assumes that the conditional quantiles are linear, instead. The choice of either method depends on the assumptions that the researcher is willing to make. Given the shape of the homeownership gap illustrated in Fig. 2, quantile decompositions are more appropriate to estimate a decomposition of the homeownership gap.

The estimated Q_θ $[y^*]$ is used to compute the difference between the counterfactual distribution of homeownership rates that would prevail if white neighborhoods had non-white endowments and the actual distribution of homeownership rates in non-white areas. That is, a measure of Q_θ $[y^*]-Q_\theta$ $[y^{NW}]$ is estimated. This is the "unexplained" homeownership gap at each quantile of the distribution.

The results are illustrated in Fig. 3 and suggest that the unexplained portion of the gap is much larger at the left tail of the distribution.[29] For instance, unexplained factors account for about 26 percentage points or about 49% of the total gap at the 10th percentile of the distribution. As higher percentiles are reached, the unexplained portion of the homeownership gap decreases steadily to 19 and 12% at the 50th and 75th percentile, respectively.

4. Testing *R* for robustness

Results presented in Sections 3 and 4 of this paper rely on a single definition of the segregated areas (namely, that $R=0.03$). This section shows that the conclusions of these sections are robust to alternative definitions and the substantially different sample sizes implied by the definition of R. In particular, the finding that most of the unexplained portion of the homeownership racial gap is located at the left tail of the distribution is notably strong.

As alternative definitions for *white* and *non-white* areas, the four other values of the threshold R described in Table 1, ranging from 0.80 to 0.99, were considered. As it was previously discussed, changing R has a direct effect on the sample size. This point is illustrated on the second and third columns of Table 7.[30] For instance, when $R=0.99$, the sample of white and non-white CBG drops to 6231 and 5148 respectively. When $R=0.80$, on the other hand, the total sample in both groups exceeds 86,000 CBGs.

For each of these five definitions, the differences in the distribution of homeownership rates between white and non-white areas are computed and displayed in Fig. 4. When R equals 0.99, 0.97, and 0.95, the estimate of the unconditional gap is almost identical. As this threshold decreases, the gap is less pronounced at the lower tail of the distribution. Nonetheless, in all specifications the general pattern of the gap remains.[31]

The robustness testing of the conditional gap is accomplished by estimating Eq. (3) using the complete set of covariates including demographic, education, income, immigration and credit variables (the same set of controls used in Table 5). The estimate of β_θ, which measures the unexplained portion of the homeownership gap, is reported on Table 7. Each column represents a chosen quantile and each row corresponds to a different sample. All estimates of β at lower quantiles are positive, statistically significant, and substantially higher than those at higher quantiles. These results suggest that the unexplained portion of the gap is significantly higher at lower percentiles despite the choice of R. Finally, the Machado–Mata decomposition is performed with each sample independently using the full set of covariates. Fig. 5 shows the "unexplained" portion of the gap in each case.[32] While the actual size of the unexplained gap depends upon R, results strongly suggest that unexplained factors are important determinants of the gap particularly at the low tail of the distribution.

The main purpose of this paper is to develop and illustrate the application of a method for adapting the Machado–Mata technique to the measurement of homeownership gaps. The particularly illustration

[27] The method described here is the one used by Albrecht et al. (2003) which is a modification of the method originally proposed by Machado and Mata (2005). Machado and Mata randomly select quantiles in this step.

[28] Albrecht et al. (2009) describe the asymptotic properties of the original method used by Machado and Mata (2005).

[29] Standard errors are computed using bootstrapping.

[30] Notice that the number of CBG in these samples does not match those displayed on Table 1. Table 1 includes all CBGs in the selected MSAs while the samples in Table 7 exclude CBGs with missing credit scores.

[31] To avoid cluttering and enhance readability, standard errors are not reported in Fig. 4. All reported differences, however, are statistically different than zero at the 95% level.

[32] Again, to avoid cluttering standard errors are not reported.

Table 4
Quantile regression estimates of the racial homeownership gap.

Independent variables included in the equation	Quantile regression (percentile)				
	(10)	(25)	(50)	(75)	(90)
Constant	0.578***	0.537***	0.408***	0.270***	0.165***
	(0.005)	(0.004)	(0.002)	(0.002)	(0.002)
Constant and basic demographics (density, older 65, family, married)	0.304***	0.228***	0.157***	0.099***	0.068***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.004)
Constant, basic demographics, and education (high school dropouts)	0.215***	0.154***	0.099***	0.046***	0.022***
	(0.007)	(0.004)	(0.003)	(0.004)	(0.005)
Constant, basic demographics, education, and income (median H. income, unemployed)	0.191***	0.131***	0.082***	0.031***	0.008*
	(0.007)	(0.004)	(0.003)	(0.003)	(0.005)
Constant, basic demographics, education, income, and immigration (bad English)	0.158***	0.098***	0.066***	0.021***	0.004
	(0.008)	(0.006)	(0.005)	(0.004)	(0.005)
Constant, basic demographics, education, income, immigration, and credit score	0.138***	0.080***	0.043***	0.008*	0.002
	(0.008)	(0.006)	(0.005)	(0.004)	(0.005)

Note: The table displays estimates for the parameter "beta" in Eq. (3) for several specifications and quantiles. The dependents variables in each quantile regression are the homeownership rate in a CBG. Each column represents a particular quantile and each row a different specification. The number of observations in each equation is 24,616. Standard errors are in parenthesis. *, and *** denote significance at the 10, and 1% level, respectively.

chosen has involved a comparison between white and non-white areas without regard to the ethnic-racial composition within non-white areas. Because much of the previous literature has focused on the black–white homeownership gap, however, a decomposition between white and "black" areas is discussed in the rest of this section.

To analyze differences in homeownership rates between white and "black" neighborhoods the same framework is used. First, a definition for a "black" area is provided. To be consistent with the previous sections, a "black" neighborhood is defined as a CBG where the share of black population is higher than R. The third column of Table 8 shows how the number of black CBGs changes with the choice of R. The samples range from 1637 to 9599 observations for thresholds between 0.99 and 0.80. Second, the difference in the unconditional distribution of homeownership rates between white and black neighborhoods is computed. As it was the case with non-white areas, the difference in the unconditional gap is substantial at the lower end of the distribution and steadily decreases after the tenth percentile (see solid line in Fig. 6).[33] Third, the conditional gap at several percentiles of the distribution is computed using pooled quantile regressions. Results are displayed on Table 8 and suggest that the unexplained portion of the gap is larger at lower percentiles despite the choice of R. Finally, the Machado–Mata decomposition is performed. The dashed line in Fig. 6 presents estimates of the "unexplained" gap, the difference between the counterfactual distribution of homeownership rates if white CBGs had black characteristics and the actual distribution of homeownership rates in black areas.[34] These results provide robust evidence that unexplained factors account for a larger portion of the white–black homeownership gap at lower quantiles.

5. Summary and conclusions

The premise of this paper is that current approaches to the analysis of homeownership gaps at the conditional mean could benefit from disaggregation to consider the distribution of gaps. This was motivated by analogy with the literature on wage and earnings gaps where differences at the mean have been shown to conceal very different pattern of differences across the wage and earnings distribution. The techniques proposed here should be differentiated from the current practice of analyzing differences in the gap across the distribution of some independent variables like income, education, or age reflecting differences in endowments or household characteristics. The advantage of using quantile regression and the Machado–Mata method is that they expose differences in the gap across the distribution of the dependent variable itself.

In analogy with various tests of the mean gap, dummy variable in a tenure choice equation, and the Oaxaca–Blinder decomposition, alternative approaches including descriptive measures of the distribution of the gap, quantile regression, and the Machado–Mata decomposition to study the distribution of the homeownership gap are considered. Because this must be done with a continuous variable, a test based on the fraction of homeowners in segregated CBGs was devised. The qualitative results are consistent with the literature, although the simple descriptive gap is larger for the measure used here. However, the effect on the conditional mean of the endowment variables is comparable to that found in the homeownership gap literature. More importantly, there is a substantial gain in insight provided because it is clear that the homeownership gap arises primarily at the lower end of the distribution. Indeed, the unexplained portion of the homeownership gap at the upper end of the distribution, once adjusted by the Machado–Mata decomposition, is

Table 5
Determinants of homeownership rates: quantile regression estimates (pooled sample).

Independent variables	Quantile regression (percentile)				
	(10)	(25)	(50)	(75)	(90)
Constant	−1.215***	−1.230***	−1.128***	−0.873***	−0.559***
	(0.073)	(0.047)	(0.034)	(0.031)	(0.036)
White	0.138***	0.080***	0.043***	0.008*	0.002
	(0.008)	(0.006)	(0.005)	(0.004)	(0.005)
Density	−0.001***	−0.001***	−0.001***	−0.001***	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Older 65	0.262***	0.339***	0.392***	0.378***	0.308***
	(0.020)	(0.013)	(0.010)	(0.008)	(0.010)
Family	0.575***	0.496***	0.443***	0.335***	0.208***
	(0.020)	(0.013)	(0.009)	(0.008)	(0.011)
Married	0.456***	0.414***	0.286***	0.194***	0.120***
	(0.023)	(0.015)	(0.011)	(0.010)	(0.013)
HS dropout	−0.002	0.041***	0.029***	0.009	−0.005
	(0.020)	(0.013)	(0.010)	(0.009)	(0.011)
Log median household income	0.098***	0.117***	0.127***	0.124***	0.113***
	(0.007)	(0.004)	(0.003)	(0.003)	(0.003)
Unemployment rate	−0.199***	−0.234***	−0.219***	−0.183***	−0.092***
	(0.026)	(0.017)	(0.014)	(0.013)	(0.016)
Bad English	−0.564***	−0.532***	−0.435***	−0.320***	−0.141***
	(0.027)	(0.019)	(0.015)	(0.013)	(0.016)
Low credit score	0.001	0.071***	0.153***	0.187***	0.191***
	(0.025)	(0.017)	(0.013)	(0.012)	(0.014)
No credit score	−0.182***	−0.298***	−0.416***	−0.485***	−0.514***
	(0.029)	(0.021)	(0.016)	(0.015)	(0.018)

Note: The dependent variable in each quantile regression is the homeownership rate in a CBG. Definitions and descriptive statistics of the explanatory variables are found on Table 2. The number of observations in each equation is 24,616. Standard errors are in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

[33] For the sake of brevity, Fig. 6 displays results when $R = 0.97$ only.

[34] To avoid cluttering and enhance readability, standard errors are not reported in Fig. 6. Most reported differences are statistically different than zero at the 95% level.

Table 6
Determinants of homeownership rates: quantile regression estimates in white and non-white CBGs.

Independent Variables	Quantile regression (percentile)					
	(25)		(50)		(75)	
	White	Non-white	White	Non-white	White	Non-white
Constant	−0.539***	−1.982***	−0.472***	−2.013***	−0.244***	−1.647***
	(0.051)	(0.086)	(0.031)	(0.079)	(0.036)	(0.093)
Density	−0.001***	−0.001***	−0.001***	−0.001***	0.000	−0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Older 65	0.227***	1.275***	0.266***	1.217***	0.235***	1.009***
	(0.011)	(0.045)	(0.007)	(0.036)	(0.009)	(0.040)
Family	0.703***	0.541***	0.532***	0.511***	0.352***	0.418***
	(0.015)	(0.024)	(0.010)	(0.022)	(0.012)	(0.025)
Married	0.250***	0.395***	0.191***	0.426***	0.134***	0.436***
	(0.018)	(0.030)	(0.011)	(0.027)	(0.014)	(0.032)
HS dropout	−0.069***	0.013	−0.048***	0.026	−0.034***	−0.020
	(0.015)	(0.025)	(0.009)	(0.023)	(0.011)	(0.027)
Log median income	0.059***	0.178***	0.072***	0.194***	0.071***	0.177***
	(0.005)	(0.008)	(0.003)	(0.007)	(0.003)	(0.008)
Unemployment rate	0.049*	−0.119***	0.055***	−0.109***	0.012	−0.092***
	(0.029)	(0.029)	(0.019)	(0.026)	(0.024)	(0.031)
Bad English	−0.542***	−0.437***	−0.303***	−0.429***	−0.203***	−0.307***
	(0.086)	(0.034)	(0.054)	(0.029)	(0.057)	(0.032)
Low credit score	−0.007	−0.065	0.016	−0.009	−0.009	0.096**
	(0.021)	(0.041)	(0.014)	(0.036)	(0.016)	(0.041)
No credit score	−0.197***	−0.100***	−0.194***	−0.207***	−0.109***	−0.332***
	(0.039)	(0.036)	(0.025)	(0.033)	(0.030)	(0.038)
Observations	16,433	8,183	16,433	8,183	16,433	8,183

Note: The dependent variable in each quantile regression is the homeownership rate in a CBG. Definitions and descriptive statistics of the explanatory variables are found on Table 2. Standard errors are in parentheses. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively. In all cases, Wald tests reject the joint null hypothesis that coefficients in White and Non-white areas are equal.

non-significant while that at the lower end of the distribution is statistically significant and substantial. Robustness tests are recommended to insure that the findings are not an artifact of the R-value used to define segregated neighborhoods. Addition of information on credit score to the traditional homeownership gap variables had a modest effect on the results. Lack of a sufficient history to form a score proved quite important as opposed to simply currently having a low score.

The main advantage of the approach suggested in this paper is the characterization and decomposition of racial gaps across the overall *distribution* of homeownership rates. It is important, however, to be aware of certain limitations. First results may depend on the definition of segregated areas. Robustness checks with alternative specifications should alleviate this concern. Second results may not be fully comparable with the previous literature. The focus here is on segregated areas while previous studies have analyzed the population.

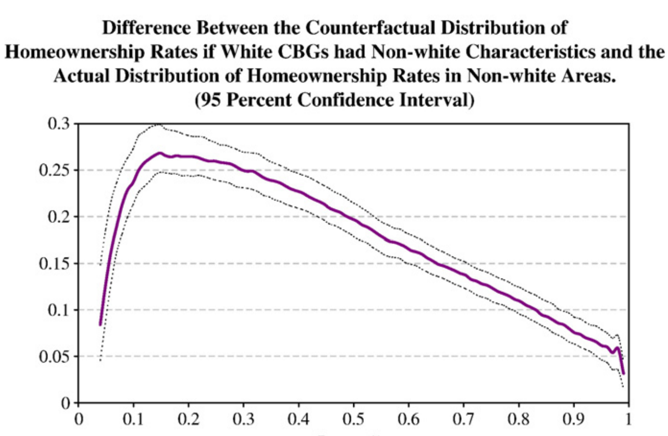

Fig. 3. Difference between the counterfactual distribution of homeownership rates if white CBGs had non-white characteristics and the actual distribution of homeownership rates in non-white areas. (95% confidence interval).

The fact that the nature of the mean decomposition of the gap using aggregate data from segregated neighborhoods is similar to previous decompositions that use individual level data provides initial insights about the relationship between the conventional approach and the alternative measures of homeownership gaps suggested in this study. Further research is needed to understand other relationships between these two methods. It should be emphasized that the method suggested in this paper should be viewed as a complement rather than a substitute of the conventional one.

Given that the product of homeownership gap analysis is either a further research challenge to find omitted variables that explain the unexplained portion of the gap or to identify policies that can act selectively and efficiency to close the gap, there should be a priority on correct decomposition of the problem into a portion explained by

Table 7
Quantile regression estimates of the racial homeownership gap for alternative definitions of segregated areas.

Value of threshold "R" for area selection	Observation (CBG)		Quantile regression (percentile)				
	White	Non-White	(10)	(25)	(50)	(75)	(90)
$R=0.80$	67,731	18,485	0.101***	0.078***	0.049***	0.019***	0.010***
			(0.004)	(0.003)	(0.002)	(0.002)	(0.002)
$R=0.90$	45,781	13,653	0.129***	0.097***	0.061***	0.027***	0.016***
			(0.005)	(0.003)	(0.003)	(0.003)	(0.003)
$R=0.95$	26,553	10,222	0.146***	0.090***	0.053***	0.019***	0.008**
			(0.006)	(0.004)	(0.003)	(0.004)	(0.004)
$R=0.97$	16,433	8,183	0.138***	0.080***	0.043***	0.008*	0.002
			(0.008)	(0.006)	(0.005)	(0.004)	(0.005)
$R=0.99$	6,231	5,148	0.146***	0.088***	0.030***	0.001	−0.002
			(0.013)	(0.009)	(0.007)	(0.006)	(0.008)

Note: This Table displays estimates for the parameter "beta" in Eq. (3) for several definitions of the segregated areas and quantiles. The dependent variable in each quantile regression is the homeownership rate in a CBG. Covariates include demographic, education, income, immigration and credit variables (the same set of control variables shown in Tables 5 and 6). Each column represents one particular quantile and each row a different specification. Standard errors are in parenthesis. *, **, and *** denote significance at the 10, 5, and 1% level, respectively.

Table 8
Quantile regression estimates of the racial homeownership gap between white and black areas.

Value of threshold (percentile) "R" for area selection	Observations (CBG)		Quantile regression				
	White	Black	(10)	(25)	(50)	(75)	(90)
$R=0.80$	67,731	9,599	0.088 *** (0.004)	0.063 *** (0.003)	0.039 *** (0.002)	0.019 *** (0.002)	0.012 *** (0.003)
$R=0.90$	45,781	6,733	0.097 *** (0.005)	0.070 *** (0.003)	0.049 *** (0.003)	0.028 *** (0.003)	0.023 *** (0.003)
$R=0.95$	26,553	4,436	0.091 *** (0.006)	0.062 *** (0.005)	0.042 *** (0.004)	0.025 *** (0.004)	0.024 *** (0.005)
$R=0.97$	16,433	3,137	0.086 *** (0.009)	0.063 *** (0.006)	0.043 *** (0.005)	0.028 *** (0.005)	0.034 *** (0.006)
$R=0.99$	6,231	1,637	0.094 *** (0.015)	0.080 *** (0.010)	0.046 *** (0.006)	0.037 *** (0.008)	0.038 *** (0.009)

Note: This Table displays estimates for the parameter "beta" in Eq. (3) for several definitions of "black" segregated areas and quantiles. The dependent variable in each quantile regression is the homeownership rate in a CBG. Covariates include demographic, education, income, immigration and credit variables (the same set of control variables shown in Tables 5 and 6). Each column represents one particular quantile and each row a different specification. Standard errors are in parenthesis. *** denotes significance at the 1% level.

endowments and an unexplained portion. Analysis of the distribution of the unexplained gap should help in both the research and policy tasks. For researchers, it points the way to potential omitted variables

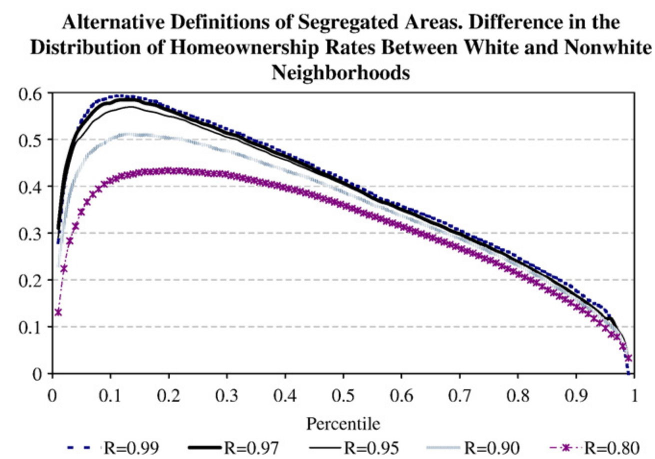

Fig. 4. Alternative definitions of segregated areas. Difference in the distribution of homeownership rates between white and non-white neighborhoods.

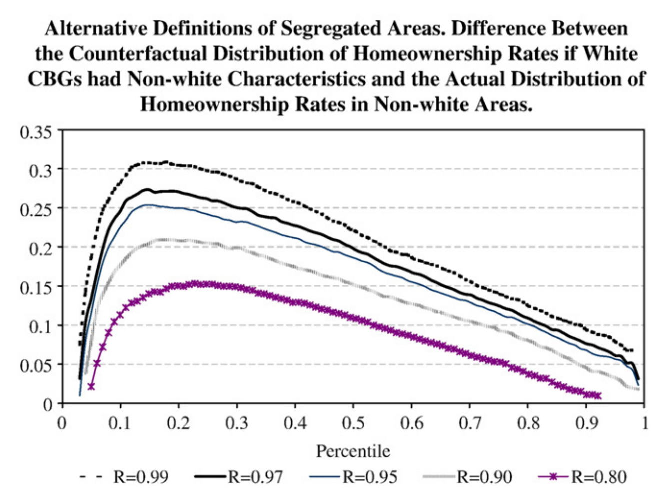

Fig. 5. Alternative definitions of segregated areas. Difference between the counterfactual distribution of homeownership rates if white CBGs had non-white characteristics and the actual distribution of homeownership rates in non-white areas.

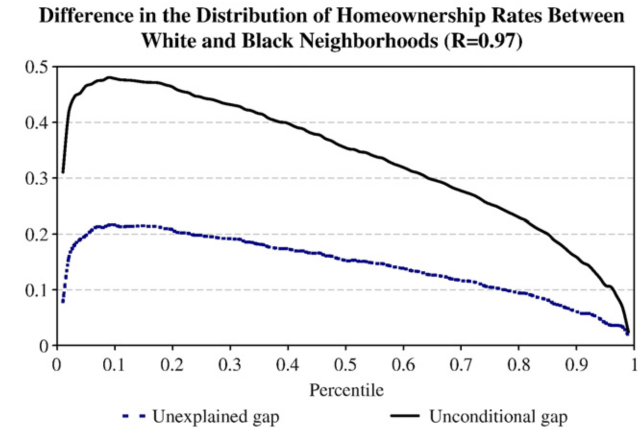

Fig. 6. Difference in the distribution of homeownership rates between white and black neighborhoods ($R=0.97$).

and to policy makers it indicates areas where the justification for action may be greatest.

References

Albrecht, J., Bjorklund, A., Vroman, S., 2003. Is there a glass ceiling in Sweden? Journal of Labor Economics 21 (1), 145–177.

Albrecht, J., Vuuren, A., Vroman, S., 2009. Counterfactual distributions with sample selection adjustments: econometric theory and an application to the Netherlands. Labor Economics 16 (4), 383–396.

Arulampalam, W., Booth, A., Bryan, M., 2007. Is there a glass ceiling over Europe? Exploring the gender pay gap across the wage distribution. Industrial and Labor Relations Review 60 (2), 163–186.

Bassett, G., Chen, H., 2001. Portfolio style: return-based attribution using quantile regression. Empirical Economics 26, 293–305.

Blinder, A., 1973. Wage discrimination: reduced form and structural estimates. Journal of Human Resources 8, 436–455.

Buchinsky, M., 1998. Recent advances in quantile regression models: a practical guideline for empirical research. Journal of Human Resources 33, 88–126.

Coulson, E., 1999. Why are Hispanic- and Asian-American homeownership rates so low?: immigration and other factors. Journal of Urban Economics 45, 209–227.

Deng, Y., Ross, S., Wachter, S., 2003. Racial differences in homeownership: the effect of residential location. Regional Science and Urban Economics 33, 517–556.

Fairlie, R.W., 2005. An extension of the Blinder–Oaxaca decomposition technique to logit and probit models. Journal of Economic and Social Measurement 305–316.

Gabriel, S., Rosenthal, S., 2005. Homeownership in the 1980s and 1990s: aggregate trends and racial gaps. Journal of Urban Economics 57, 101–127.

Gyourko, J., Tracy, J., 1999. A look at real housing prices and incomes: some implications for housing affordability and quality. Federal Reserve Bank of New York Economic Policy Review 63–77 (September).

Haurin, D., Hebert, C., Rosenthal, S., 2007. Homeownership gaps among low-income and minority households. Cityscape 9 (2), 5–52.

Kain, J.F., Quigley, J.M., 1972. Housing market discrimination, homeownership, and savings behavior. American Economic Review 263–277.

Koenker, R., 2005. Quantile regression. Econometric Society Monograph. Cambridge University Press.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.

Koenker, R., Hallock, K., 2001. Quantile regression. Journal of Economic Perspectives 15, 143–156.

Machado, J., Mata, J., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of Applied Econometrics 20, 445–465.

McMillen, D., 2008. Changes in the distribution of house prices over time: Structural characteristics, neighborhood, or coefficients? Journal of Urban Economics, 64 3, 573–589.

McMillen, D., Thorsnes, P., 2006. Housing renovations and the quantile repeat sales price index. Real Estate Economics 34 (4), 567–584.

McMillen, D., Singell, L., 2008. School finance reforms and property tax limitation measures. In: Ingram, Gregory K., Hond, Yu-Hung (Eds.), Fiscal Decentralization and Land Policies. Lincoln Institute of Land Policy, Cambridge Massachusetts, pp. 141–166.

Myers, S.L., Chung, C., 1996. Racial differences in home ownership and home equity among preretirement-aged households. The Gerontologist 350–360.

Myers, D., Lee, S.W., 1998. Immigrant trajectories into homeownership: a temporal analysis of residential assimilation. International Migration Review 593–624.

Myers, D., Megbolugbe, I., Lee, S.W., 1998. Cohort estimation of homeownership attainment among native-born and immigrant populations. Journal of Housing Research 237–269.

Myers, D., Painter, G., Yu, Z., Ryu, S.H., Wei, L., 2005. Regional disparities in homeownership trajectories: impacts of affordability, new construction, and immigration. Housing Policy Debate 53–83.

Nguyen, B., Albrecht, J., Vroman, S., Westbrook, D., 2007. A quantile regression decomposition of urban–rural inequality in Vietnam. Journal of Development Economics 83 (2), 466–490.
Oaxaca, R.L., 1973. Male female wage differentials in urban labor markets. International Economic Review 14 (3), 693–709.
Painter, G., Gabriel, S., Myers, D., 2001. Race, immigrant status, and housing tenure choice. Journal of Urban Economics 49, 150–167.
Roistacher, E.A., Goodman, J.L., 1976. Race and home ownership: is discrimination disappearing? Economic Inquiry 59–70.
Wachter, S., Megbolugbe, I., 1992. Racial and ethnic disparities in homeownership. Housing Policy Debate 3 (2), 333–370.
Yun, M.S., 2007. An extension of the Oaxaca decomposition using generalized residuals. Journal of Economic and Social Measurement 15–22.