Learning of business processes & application: An industry-ready approach

Yi Meng LAU
Singapore Management University, ymlau@smu.edu.sg

Yu Yee POON

Mike WEE

Regional Science and Urban Economics 42 (2012) 52–62

Contents lists available at ScienceDirect

Regional Science and Urban Economics

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

A household-level decomposition of the white–black homeownership gap

Eric Fesselmeyer [a,*], Kien T. Le [b], Kiat Ying Seah [c]

[a] Department of Economics, National University of Singapore, Block AS2, #05-02, 1 Arts Link, 117570 Singapore
[b] Social and Economic Survey Research Institute, Qatar University, P.O. Box 2713, Doha, Qatar
[c] Department of Real Estate and Institute of Real Estate Studies, National University of Singapore, 4 Architecture Drive, 117566 Singapore

ARTICLE INFO

Article history:
Received 18 November 2010
Received in revised form 24 May 2011
Accepted 26 May 2011
Available online 2 June 2011

JEL Classifications:
C14
R20
R21
R23

Keywords:
Homeownership
Race
Segregation

ABSTRACT

This paper uses a semiparametric homeownership model to estimate and to decompose the household-level white–black homeownership gap into an endowment component and a residual component across the distribution of homeownership rates. We find that the racial gap differs across homeownership rates and that studies that examine the gap only at the mean may be misleading. We also find that although household characteristics explain the homeownership gap for most households, there is a substantial portion of the gap that remains unexplained for households with a very low propensity to own homes. A comparison of the estimates from the semiparametric model and a probit model suggests that the semiparametric approach is able to capture the heterogeneity structure between the ethnic groups, particularly in the tails of the distribution. To illustrate the flexibility of our household-level approach, we decompose the homeownership gap in cities of varying levels of segregation.

1. Introduction

Against the backdrop of a considerable rise in the U.S. homeownership rate, from 64% in 1994 to 69% a decade later, a rather somber and poignant statistic persists: the white–black homeownership gap has remained large at around 25 percentage points.

There are several reasons why racial differences in homeownership rates are of interest. Homeownership has long been perceived as an integral component of the "American Dream" and since homeownership is related to the consumption of housing services, examining the racial differences in homeownership is critical to understand the economic well-being of the different racial groups. Aside from the fact that neighborhood quality – which is often associated with the proportion of homeowners – can have important ramifications on social and economic outcomes (education, crimes, and positive information spillovers), it has also been argued that homeownership may have farther-reaching benefits that are bestowed on children of homeowning parents (Green and White, 1997).

Many recent studies of the white–black homeownership gap, or *total gap*, decompose the gap at the conditional mean into an observable socio-economic factors component, typically referred to as the *characteristics gap*, and an unobservable residual component or the *residual gap*. Socio-economic factors include observed differences in income and wealth as well as differences in household demographics such as marital status, age, and educational attainment. The residual amount captures unmeasured factors such as differences in tastes for homeownership, discrimination, access to credit, and credit history.[1] These studies typically quantify how important each determinant is for the average household but leave the question of whether these determinants affect households with different homeownership rates differently unanswered.

In this paper we decompose the white–black homeownership gap at each percentile of the distribution of homeownership rates instead of only at the conditional mean. We find that analyzing the homeownership gap only at the conditional mean may be misleading as the homeownership gap and its causes differ across the distribution of homeownership rates. Our results suggest that racial differences in endowments (or differences in observable household characteristics) explain the racial gap well for most households except for households at the lower percentiles of the distribution. In other words, even if these black households were to have characteristics like their white counterparts, they would still be unlikely to own.

This paper builds on the literature that uncovers distributional information that would otherwise be masked at the means or

* Corresponding author. Tel.: +65 6516 4873; fax: +65 6775 2646.
E-mail address: ecsef@nus.edu.sg (E. Fesselmeyer).

[1] The residual gap could also potentially capture racial differences in measurement error if the differences were large.

0166-0462/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.regsciurbeco.2011.05.002

medians. Distributional analyses such as ours should prove useful to policy makers who are interested in understanding why certain segments of minority groups are persistently at the lower rung of the homeownership scale. In practice, identifying the determinants of the distribution of the homeownership gap serves as *prima facie* stylized evidence for more complex "inequality" studies on social and intergenerational mobility of the population.

The paper that is closest to ours in scope is Carrillo and Yezer (2009). Like us, Carrillo and Yezer's goal is to decompose the entire homeownership gap distribution. The method they use was introduced in Machado and Mata (2005) and was originally used to decompose changes in wages, a continuous variable.[2] Since homeownership is a binary response variable, a direct application of the Machado–Mata method is infeasible. Instead of decomposing the racial gap using household-level data, Carrillo and Yezer focus on average aggregate ownership rates at the census block group level. Doing so enables them to apply the Machado–Mata method, but at a cost. Carrillo and Yezer had to constrain their sample to include only highly segregated neighborhoods wherein the proportion of white households is either close to zero or close to 100%. It is therefore unclear whether their results apply to disaggregate data and to less segregated cities and neighborhoods. We propose a different method to decompose the homeownership gap. We use a binary response semiparametric method to estimate the probability function of the standard homeownership model by race with household-level data and then decompose the racial difference in the distributions of estimated homeownership probabilities using a Oaxaca–Blinder type approach. We do not view our approach as necessarily superior. In our opinion, our paper offers a complementary analysis in understanding the white–black gap puzzle.

There are two primary advantages of the semiparametric approach used in this study. First, our approach is capable of handling heterogeneity across different ethnic groups while averting the curse of dimensionality that plagues most nonparametric approaches. We show this by comparing our estimated homeownership densities with those obtained from a probit model and find that the normal distribution tends to place more mass at the center of the distribution and understates the effects of the tails in the presence of conditional heterogeneity.

Second, our approach is more flexible in that we are able to decompose the homeownership gap both at the national level and at the metropolitan-area level for cities with varying amounts of segregation. We see marked differences between the city-level analyses and the national data, thus highlighting a Tiebout-esque revelation that cities are innately heterogeneous and that effective socio-economic policies are to be mindful of this fact.

The rest of the paper is organized as follows. The next section discusses the relevant literature. Section 3 describes our methodology, and Section 4 the data. Section 5 contains our results, including a comparison of our results with a decomposition based on a probit homeownership model, and the last section concludes.

2. Related literature

2.1. Determinants of the homeownership gap

The size of the racial homeownership gap and its persistence over time are a source of concern among policy makers and researchers. This concern has led to a rich array of studies that attempt to understand the underlying causes of the racial gap. A natural conjecture is that discrimination exists in the process of homeownership, and minorities may be treated differently by realtors and mortgage lenders (for example, see Kain and Quigley (1972); Yinger (1995); Munnell et al. (1996).

Equally plausible is the explanation that the racial gap in homeownership is a result of racial differences in income and wealth. Wachter and Megbolugbe (1992) find that endowment differences explain a sizeable portion of the racial gap. Moreover, since buying a home requires a substantial commitment of one's wealth, downpayment constraints could effectively cripple a household's ability to own a home. Linneman and Wachter (1989) and Duca and Rosenthal (1994) find that downpayment and credit constraints are more important than income in influencing homeownership decisions. It is also widely believed that white households benefit from larger intergenerational transfers either through monetary gifts or bequeathed estate than do minority households (see for example, Gale and Scholz (1994)).

Differences in personal characteristics such as education level, race, age, gender, income levels, and marital status may also impact homeownership rates. Bostic and Surette (2001) conclude that changes in these characteristics explain more of the homeownership rates among the higher income families while contributing very little in explaining the changes in homeownership rates among lower income households. This suggests that household characteristics have differential impacts on homeownership decisions across income groups.

That tenure choice and location choice are joint decisions raises the possibility that persistent levels of segregation in American neighborhoods could potentially contribute to the racial gap in homeownership rates. In their study, Deng et al. (2003) documented empirical evidence that suggests that the opposite is true, that location decisions endogenous to tenure choice seem to mitigate the racial gap. The argument is that racial and income segregation in American cities could arise from self-sorting so that poorer minority households would choose to locate in affordable locales thus increasing their propensity to become homeowners.

2.2. Decomposing the homeownership gap

Cognizant of the fact that systematic differences in household characteristics across ethnic groups can have differential impacts on homeownership decisions, the latest studies have focused on *decomposing* the homeownership gap into an endowment component that is attributable to observable socioeconomic characteristics and a residual component, which soaks up any remaining differential impacts. The residual effect is widely thought to capture, though not exclusively, the effects of discrimination. The narrowing of this gap is generally attributed to changes in regulations governing the mortgage and housing markets whereby homeownership has been made more affordable and less discriminatory. Quantifying the relative magnitudes of either component is of interest to policy makers and researchers alike. Undoubtedly, identifying the size and the significance of the components that made up the gap allows for more meaningful policy analysis.

Silberman et al. (1982) is one of the earliest studies that analyzes the racial gap using the standard Oaxaca–Blinder decomposition evaluated at the sample mean of the attributes. They found that the residual portion of the gap was smaller than the endowment gap and that this residual portion has shrunk by over 30% between 1974 and 1978, suggesting a decline in discrimination. Long and Caudill (1992) undertake a comparable study based on a 1986 sample and finds that the residual portion only explains 34%, half of the endowment counterpart which explains 66% of the homeownership gap. Collins and Margo (2001) take a historical view of the homeownership gap using Census data for the years 1940 to 1990 to find that the widening of the gap between 1940 and 1960 was due to the migration of blacks

[2] The first use of the Machado–Mata method in the urban economics literature decomposed the changes in the distribution of house prices in Chicago (McMillen (2008)). For an alternative approach to the Machado–Mata method see [Cobb-clark and Sinning (2011)] who used the approach of DiNardo et al. (1996) to decompose changes in the distribution of home values in Australia.

to central city areas while the narrowing of the gap between 1960 and 1980 is explained by changes in the residual components. While informative, in all these studies, the Oaxaca–Blinder approach is limited to looking at the mean and thus could potentially mask distributional impacts. Bostic and Surette (2001) augmented the Oaxaca–Blinder approach by sorting their sample into income quintiles and racial groups before decomposing the racial gap within each group. Their results show that the relative importance of both the endowment effect and the residual effect changes with income level and ethnicity. Their results, while interesting *per se*, highlight the inadequacy of examining the homeownership trend only at the means, potentially ignoring the differential effects across the distribution.

3. Methodology

The goal of this paper is to decompose the difference in the distribution of household-level homeownership probabilities of the two races. We start with the standard approach that treats the unobserved utility of owning a home as a random variable and estimates the probability of homeownership conditional on an observed matrix of covariates X. Let the utility of homeownership be the latent variable y^*:

$$y^* \quad = \quad X\beta + \epsilon, \quad \epsilon \sim F_\epsilon \tag{1}$$

where β is a vector of parameters and X contains the standard observed household characteristics that existing studies have found to be important. These include variables such as household income, age, and the number of persons in the household.[3] The random term ε captures unobserved factors that affect ownership such as discrimination against blacks in the home mortgage market. A binary choice model can be written as

$$y = \begin{cases} 1 & \text{if } y^* \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then, for any given X, the conditional probability of homeownership is:

$$Pr(y = 1|X) = \int_{-\infty}^{\infty} \mathrm{I}(\epsilon \leq X\beta)\ dF_\epsilon = F_\epsilon(X\beta), \tag{2}$$

where I( ) is the indicator function.

Our decomposition is derived from the predicted homeownership probabilities of this conditional probability function. For each racial group, we compute the ξ^r that satisfies the sample moment

$$\frac{1}{N_r}\sum_i \mathrm{I}\left(\hat{F}^r_\epsilon(X^r_i\beta^r) \leq \xi^r(\alpha)\ \right) = \alpha,$$

where N_r is the number of observations of group r. The estimated *total* homeownership gap at the α-th percentile, Δ_α, is then $\Delta_\alpha = \xi^w(\alpha) - \xi^b(\alpha)$, where $\xi^r(\alpha)$ is the α-th percentile for the racial group $r \in \{b, w\}$, with b denoting black households and w white households. The decomposition of the homeownership gap at the α-th percentile is then

$$\Delta_\alpha = \left[\xi^b\left(\alpha; X^w\beta^b\right) - \xi^b(\alpha)\ \right] + \left[\xi^w(\alpha) - \xi^b\left(\alpha; X^w\beta^b\right)\right], \tag{3}$$

where $\xi^b(\alpha; X^w\beta^b)$ is the α-th percentile of the black counterfactual; it is the α-th percentile that would prevail if Eq. (2) of black households were evaluated using the attributes of white households.

The first term on the right-hand side of Eq. (3) represents the *characteristics gap* or the contribution of the racial difference in the covariates to the overall gap in homeownership. The covariates include housing services consumption and household characteristics. The second term is the *residual gap* which is the contribution of unobservable factors to the homeownership gap. Intuitively, we could think of the characteristics gap as capturing the difference in behavior between the two races as if their homeowning decisions are determined by observable characteristics alone, and the residual gap as capturing the racial difference in the manner by which these characteristics determine the respective propensity to own.

Note that we have explicitly allowed the CDF $F^r_\varepsilon(\cdot)$ to differ across the two racial groups to avoid possible misspecification that could cause our estimates to be inconsistent. Early racial gap studies tend to treat the CDF as known (either as a normal or a logistic distribution) and rely only on the inclusion of race dummies in their models to capture race-associated differences in homeownership probabilities. The problem with such an approach is the implicit assumption that both β and $F^r_\varepsilon(\cdot)$ are the same for the two racial groups. Such an assumption is usually not based on any *a priori* knowledge nor any economic theory.[4] In our study, we are able to estimate different CDFs for black and white households by using the Klein and Spady (1993) semiparametric single index binary choice model which avoids the 'curse of dimensionality' one would encounter in nonparametric regression but is capable of estimating conditional homeownership probability functions that are consistent with the data.[5]

3.1. Estimation

3.1.1. Home ownership model

We estimate the homeownership model using the semiparametric estimator of Klein and Spady (1993). In parametric binary models, estimation is typically carried out by optimizing the log-likelihood function with a pre-specified F_ε:

$$\max_\beta \sum_i y_i \log(F_\epsilon(X_i\beta)) + (1-y_i)\ \log(1-F_\epsilon(X_i\beta)) \tag{4}$$

Klein and Spady, (1993) propose an estimator of β that maximizes the log-likelihood function as in Eq. (4) but with a nonparametric estimate $\hat{F}_\epsilon(\cdot)$ in place of $F_\epsilon(\cdot)$.[6] Specifically, define $\hat{P}^r_{own} = N_r^{-1}\sum y^r_i$ as the sample proportion of homeowners in a particular racial group. Then for any real value ν,

$$\hat{F}^r_\epsilon(\nu) = \frac{\hat{P}^r_{own}\,\hat{q}^r(\nu|y=1)}{\hat{P}^r_{own}\,\hat{q}^r(\nu|y=1) + \left(1-\hat{P}^r_{own}\right)\hat{q}^r(\nu|y=0)} \tag{5}$$

where $\hat{q}(\cdot|y)$ is a kernel estimate of $q(\cdot|y)$, the conditional density of $X\beta$. The kernel estimates for either racial group are

$$\hat{q}^r(\nu|y=1) = \frac{\sum y^r_i K\left[(\nu - X^r_i\beta^r)/h_r\right]}{N_r\hat{P}^r_{own}h_r} \tag{6}$$

[3] Here we note that like much of the literature we do not formally model homeowner user costs. For a user cost type model it would be necessary to include locally measured variables such as quality adjusted price–rent ratios. The public version of the ACS, however, for confidentiality reasons, includes only very coarse uncensored household location information, such as regional dummies, which we include as a crude proxy of the regional differences in user costs. See Gabriel and Rosenthal (2005) for a more detailed discussion of this issue.

[4] Although subsequent decomposition studies specify different β across racial groups, these studies still assume the same underlying conditional distribution for black and white households. See Herbert et al. (2005) for an extensive survey of this literature.

[5] The Klein–Spady approach is 'semiparametric' because it achieves model identification by assuming that the CDF depends on the vector X through a single linear combination $X\beta$ but does not impose any other distributional assumptions.

[6] In single index binary choice models, the function F^r_ε includes location and level shift and so the vector X does not include a constant. In other words, the intercept component of β is subsumed in the function F^r_ε and is not identified; we need to impose the normalization that $\beta'\beta = 1$. Identification of β and F^r_ε also requires that X contains at least one continuously distributed variable which has a non-zero coefficient.

and

$$\hat{q}^r(\nu|y=0) = \frac{\sum(1-y_i^r)K[(\nu-X_i^r\beta^r)/h_r]}{N_r\left(1-\widehat{P}_{own}^r\right)h_r} \tag{7}$$

where K is the kernel function[7] and h_r the bandwidth, which varies with the racial group and satisfies the rate $N_r^{-1/6}<h_r<N_r^{-1/8}$.[8] Klein and Spady do not give any practical guidelines about how to select the bandwidth. In this paper, we used Silverman, (1986)'s rule of thumb. Our results change very little when we vary our bandwidth within a 10% band in either direction.

3.1.2. Counterfactual probabilities and decomposition

To compute the decomposition, we first estimate the counterfactual homeownership probabilities of blacks. Recall that the counterfactual of blacks is the probability of ownership if the black homeownership distribution was evaluated for white characteristics. These counterfactual probabilities are calculated from Eq. (5) using the black kernels, $\hat{q}^b(\nu|y=1)$ and $\hat{q}^b(\nu|y=0)$ of Eqs. (6) and (7), and evaluating ν at $\nu=X_i^w\beta^b$. The bandwidth for the counterfactual is h_b. The counterfactual percentiles are then computed, and the decomposition is carried out using Eq. (3).[9]

4. Data and results

We use data from the 2007 American Community Survey (ACS) which is a nationwide, annual survey that samples residents of over 3 million housing units. The ACS essentially collects the same information as the U.S. Census. It includes questions on demography (such as marital status, race, education, employment, and occupation) and also questions on housing (such as tenure choice, property value, housing type and cost of utilities).

We include in our sample only households that are headed by non-hispanic blacks or non-hispanic whites. We drop observations of households with more than one family, households whose head is in school, households whose head is a farmer, and households that live in a mobile home, trailer, boat, tent, or van. In the final sample, we are left with 695,038 white households and 85,454 black households.[10] We include household-level sample weights in our Klein–Spady estimation as well as in our decomposition. The main purpose of these weights is to allow the researcher to compute statistics that are representative of the full population since some household characteristics are over-weighted or under-weighted in the sample.[11]

Appendix A contains descriptive statistics of the variables we include in the model. The average white homeownership rate is 75% whereas it is 47% for blacks which translates into a homeownershap gap of 28% percentage points between the races. The averages of the explanatory variables are consistent with our beliefs about the economic differences of black and white households. Average household income is higher for white households. White heads are more likely to be college-educated, more likely to be male, and are more likely to be married. We also note that black households are more likely to reside in metropolitan areas than white ones, and a large proportion of black households live in the south.

Like Rosenthal, (1988) and Haurin et al., (1997), we have included duration of residence as a covariate. Actual duration of residence is a noisy measure of anticipated duration of residence, a critical determinant of a household's annualized user cost of homeownership, which in turn determines homeownership. At the same time, expected duration of residence is potentially endogenous if it is determined jointly with tenure choice (See for example, Boehm (1981)) and Henderson and Ionnides (1989)). The ideal approach in some cases would be to estimate a joint model using panel data to capture the dynamic nature of the decision. However, one of the goals of our paper is to relax the parametric assumptions of the homeownership model; the use of the Klein–Spady semiparametric method to do so precludes a binary dependent variable. Our analysis can therefore be interpreted as a reduced-form approach that is common in the literature in which the duration of residence covariate is interpreted as a control variable as well as a proxy for user cost.[12]

Estimates of β and their standard errors can be found in Appendix B. That all estimates have very small standard errors is not surprising given the size of our sample. Consistent with the previous studies, we find that having a higher income, being a larger household, being married with a spouse present, being older, earning more non-wage income, and having higher levels of education increase the likelihood of homeownership for both races. Also, households that live in metropolitan areas are less likely to own. There are some differences between the races to note. For example, having a college education has a bigger effect on blacks than on whites. Household size also has less of an impact on blacks than it does on whites. An unexpected result is that being an American citizen has a negative effect on the likelihood to own a home for blacks.

Fig. 1 provides snapshots of the stark contrast between the two races' propensities to own a home. Panel (a) contains kernel density estimates of the predicted homeownership probabilities of white and black households. A large amount of mass is concentrated in the upper range of the white density, reflecting the very high probability of homeownership for many of the white households. On the other hand, the black density is much more uniform. There is a nontrivial mass at the lower and middle range of the black density, indicating that a considerable number of black households have a low probability of homeownership. Panel (b) contains the CDFs of the predicted homeownership probabilities of white and black households. The horizontal difference between the black and white CDFs is the total homeownership gap. For example, at the 20th percentile of the black CDF, the homeownership probability is around 16% whereas for the white CDF, the homeownership probability is about 53%. The total homeownership gap at the 20th percentile is then the difference, about 37 percentage points.

A clearer picture of the total homeownership gap is in Fig. 2 where we graph the horizontal differences between the two CDFs by percentile. We see that the gap increases steadily in the lower range,

[7] In this study, we use the normal density function as the kernel function. The variation in density estimates is often negligible compared to the variation with bandwidth choice. See, for example, Cameron and Trivedi (2005), pp. 300 for comparisons.

[8] Klein and Spady (1993) showed that their estimator is consistent and achieves the asymptotic efficiency bound of Cosslett (1987).

[9] In principle, the decomposition can be done using the coefficients and distribution from either the white or black estimates. For example, Long and Caudill (1992) use the coefficients from the black regression whereas Collins and Margo (2001) use the coefficients from the white regression. We report results from both decompositions below.

[10] The advantage of using the ACS over other data sets such as the American Housing Survey is its large size. This is particularly important in our study because a semiparametric model such as the one we estimate is more data intensive than fully parameterized models.

[11] Details of these weights are discussed on the IPUMS website and in the Census Bureau's handbook, "A Compass for understanding and using the American Community Survey Data", February, 2009.

[12] The inclusion of a control variable that is potentially endogenous means that the variable will not have a causal interpretation; rather its inclusion is justified if not including it could cause omitted variable bias. See Angrist (2001) for a detailed discussion on this subject. As a check we estimated our model with and without the duration of residence variable. Since duration of residence is positively correlated with age and family size, omitting the duration variable should result in larger age and family size coefficients. This is indeed the case; we see a doubling to a 3.6 times increase in the white age coefficients and 1.3 times to 2.4 times increase in the black age coefficients as well as a doubling of family size coefficients for the white racial group and 1.3 times to 2.1 times increase in the black family size coefficients. We deem these large increases in estimates as evidence of omitted variable bias. Boehm (1981) and Herbert et al. (2005) make a similar point.

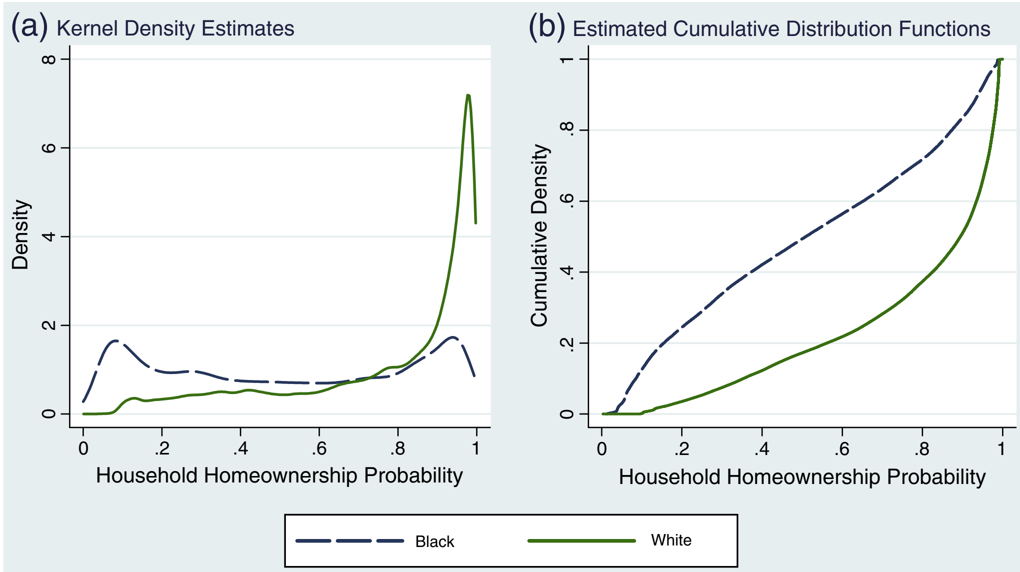

Fig. 1. Estimated homeownership probability functions.

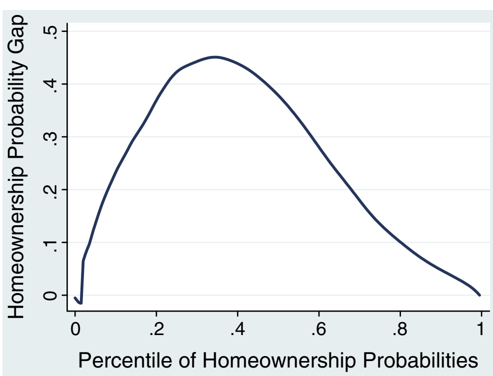

Fig. 2. Total white–black homeowership probability gap by percentile.

peaks at around the 35th percentile with a gap of approximately 47 percentage points, and decreases thereafter. The shape of the total gap is similar to the total gap depicted in the Carrillo–Yezer's census tract-level analysis but differs at the peaks. The total gap in the Carrillo–Yezer aggregate-level study peaks at about the 10th percentile with a value of around 56 percentage points. This suggests that at the household level, the difference in homeownership probability is less concentrated in the lower percentiles and more concentrated in the middle percentiles than in the census tract-level analysis.

5. Decomposition results

In Fig. 3 we have added the counterfactual kernel density estimate and the counterfactual CDF to the density functions and CDFs of Fig. 1. Recall that the counterfactual is the hypothetical outcome should black households have white characteristics. From panel (a) of Fig. 3

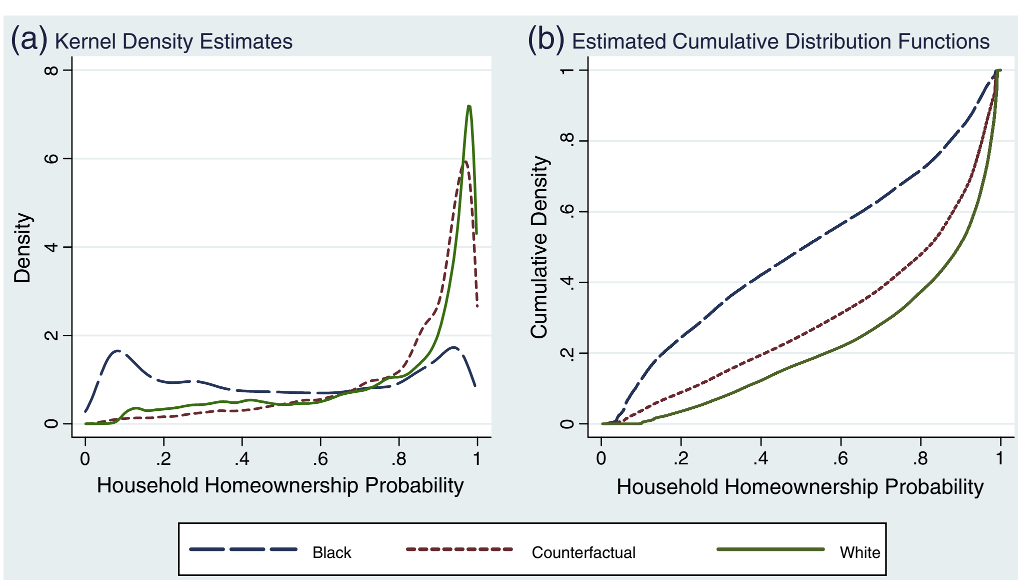

Fig. 3. Estimated homeownership probability functions (including counterfactual).

note that the black density is very different from the counterfactual density. In contrast, the shape and the level of the counterfactual density are similar to the white density. This suggests that differences in characteristics explain a substantial portion of the total homeownership gap, with the residual component explaining only a modest amount.

We saw in the previous section that the horizontal difference between the white and black CDFs is the total gap. Similarly, we can decompose the total gap into the residual gap and the characteristics gap by measuring the horizontal differences between the black and white CDFs and the counterfactual CDF. As the horizontal difference between the black CDF and the counterfactual CDF is due to differences in characteristics, it measures the characteristics gap. The horizontal difference between the counterfactual CDF and the white CDF is the residual gap as both CDFs are evaluated for the same characteristics. Again, to illustrate the calculation, we look at the 20th percentile. In the counterfactual world, the black household at the 20th percentile would have a homeownership rate of about 41 percentage points if they were endowed with white characteristics. This 25 percentage point increase compared to the 20th percentile of the black CDF can be attributed to differences in observable characteristics. In other words, the characteristics gap at the 20th percentile explains 25 percentage points of the total homeownership gap. The remaining portion of the total gap, about 12 percentage points, is the residual gap.

The decomposition of the entire total gap is depicted in Fig. 4 where we graph the horizontal differences between the CDFs by percentile. The shape of the characteristics gap across the percentiles is similar to that of the total gap in that it is small at the lower percentiles, peaks in the middle percentiles, and decreases at the upper percentiles. The residual gap is slightly larger than the characteristic gap at the lower percentiles (up to the 10th percentile), then becomes somewhat persistent up to the 35th percentile and decreases steadily thereafter. This tells us that factors other than household characteristics are important determinants of homeownership for blacks that are less likely to own a home. However, the greater the likelihood of homeownership, the more one can attribute the homeownership gap to the difference in household characteristics between the two ethnic groups. In general, compared to the characteristics gap, the residual gap is relatively small overall, meaning that the unobservable factors contribute little in explaining the racial gap especially at the higher percentiles. The finding is similar to that of Bostic and Surette, (2001) if one believes that with higher income, the probability of owning a home is higher. In that study, household characteristics explain more of homeownership rates among higher income families relative to lower income families.[13]

The individual covariates that explain much of the total gap are shown in Fig. 5. Income, education, marital status, and duration of residence mirror the shape of the characteristics gap with contributions that are relatively large. The contribution of income and of other income (interest, dividend, and rental income) is nearly 8 percentage points and 6 percentage points of the total gap at their peaks, reflecting the importance of the differences in income by race. Higher education levels among whites accounts for over 3 percentage points of the total gap at its peak. Similarly, the contribution of marital status is about 8 percentage points at its peak, reflecting the fact that white homeowners tend to be married. Duration of residence accounts for nearly 10 percentage points at its peak. This is consistent with the conventional wisdom that households that are more mobile are less likely to be homeowners and more likely to be renters.

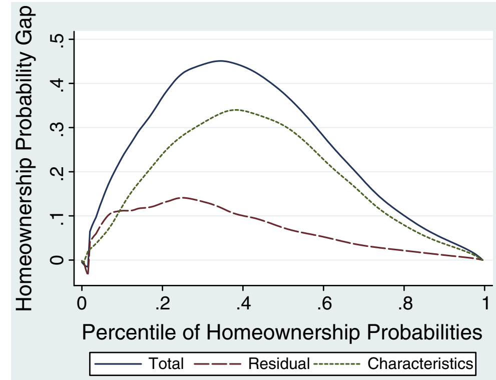

Fig. 4. Total gap and decomposition by percentile.

In summary, we find that the residual portion of the racial gap has less explanatory power than the characteristics gap overall but that for certain percentiles the size of the residual gap is large. Table 1 contains the contribution of the characteristics gap and residual gap to the total gap by percentage across the distribution as well as at the mean using the typical Oaxaca–Blinder decomposition. At the mean the residual gap explains only 29% of the total homeownership gap, less than half of the characteristics gap which explains about 71%. However, these average figures are not representative of the gaps across the distribution. Specifically, the residual portion is quite substantial at the 10th and 20th percentiles, at 48% and 35% of the total gap, respectively. In other words, a mean decomposition hides the fact that we are not fully accounting for many of the observable factors driving the white–black homeownership gap among those least likely to own.

5.1. Comparison with the probit model

In this section we compare the results of the Klein–Spady model with that of a probit model. One of the main advantages of a semiparametric approach is that we are able to relax parametric assumptions and yield consistent estimators. Monte Carlo studies have shown that probit estimators can be biased when the distribution F_ε is heteroskedastic or asymmetric (see for example, (Manski and Thompson, (1989) and Powell et al., (1989)). Notwithstanding the potential misspecification in using the probit model, researchers may prefer the probit model since nonparametric or semiparametric approaches such as the Klein–Spady are computationally more intensive.[14] Therefore, it seems instructive to examine the difference in the results of the two models.

First, we look at the estimated marginal effects of the two models in Appendix C.[15] Though the marginal effects tend to be larger for the

[13] Using the alternative counterfactual in the decomposition – that is, predicting homeownership probabilities for black households using the white distribution and coefficients – we find that the characteristics gap still explains most of the total gap, and there is no substantial change in the shape of the curves in Fig. 4. We do find, however, that the levels of the gaps are somewhat affected. In particular, the residual gap of the alternative decomposition is smaller than the residual gap of the reported decomposition for the percentiles below the 30th percentile, approximately equal at the 30th percentile, and larger for percentiles above the 30th percentile (and vice versa for the characteristics gap). For example, at the 10th percentile, the alternative residual gap is 11.2 percentage points compared to 7.3 percentage points for the reported decomposition. At the 30th percentile, the residual gap from both decompositions is approximately 13 percentage points. At the 90th percentile, the alternative residual gap is 2.6 percentage points compared to 11.1 percentage points for the reported decomposition.

[14] It took about 12 h to compute the Klein–Spady estimates for the white sample and 4 h for the black sample.

[15] The marginal effect for categorical variables is the change in homeownership probability when X_i changes from 0 to 1, holding all other variables constant at their observed values, averaged over the households. For continuous variables, the marginal effect is the change in predicted homeownership probability for a unit increase in the continuous variable, holding all other variables constant at their observed values, averaged over households.

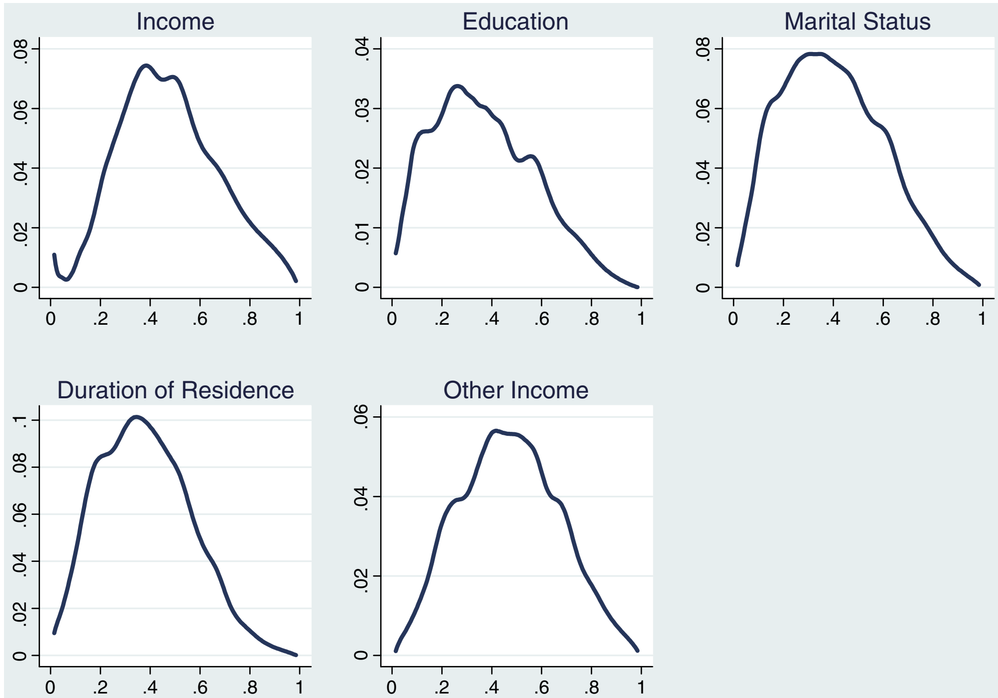

Fig. 5. Contribution of important variables to the total gap by percentile.

probit model, the signs of the marginal effects are the same across the models for all variables. If our comparison were to stop here, the probit model, due to its ease of computation, may be preferred to the Klein–Spady model.

Next, we examine the difference in predicted homeownership probabilities for the two models. Fig. 6 contains the kernel density estimates of the homeownership probabilities for the white and black households as well as the counterfactual homeownership probabilities. For the black density, the Klein–Spady model shifts the mass away from the center and places more weight at the lower and upper tails than the probit model does. For both the white and the counterfactual densities, the Klein–Spady model places more mass at the upper tails. This seems consistent with simulation evidence in Klein and Spady, (1993) that shows that a distribution with conditional heteroskedasticity has fatter tails than would a normal distribution — the main difference between a probit model and the Klein–Spady model would occur at the tails. Our figures show that the probit and Klein–Spady estimated homeownership probabilities are similar at the center of the distribution. In other words, the choice between the use of parametric model and a semi- or a nonparametric model may be of minor importance if we are interested in mean homeownership rates. But if we are interested in why certain minority groups are persistently at the lower rung of the homeownership scale, a semiparametric approach will be able to recover important distribution information from the sample data and a probit specification may not be the appropriate model.

We now compare the differences in the decomposition results. Table 2 contains the probit decomposition results. At most percentiles, the Klein–Spady model produces a larger total gap, particularly at the lower end of the distribution. At the 10th percentile, the Klein–Spady total gap is 23 percentage points and the probit total gap is only 15 percentage points. For the rest of the distribution, though, the difference in magnitudes between the two models is small. Overall, the contribution of the residual gap to the total gap is larger for the probit model than for the Klein–Spady except at the 80th and 90th percentiles where they are roughly similar. Consequently, the characteristics gap explains more of the total gap for the Klein–Spady model than for the probit model. Such a result is perhaps not all surprising since the probit model presupposes a parametric distribution which might not fit the data well. On the other hand, the Klein–Spady model allows more flexibility in estimation, hence increasing the fit of the conditional distribution and, consequently, the importance of the characteristics.

Table 1
Klein–Spady decomposition as a percentage of the total gap.

Percentile	Total gap	Characteristics gap	Residual gap	Characteristics gap (%)	Residual gap (%)
10th	0.23	0.12	0.11	52.0	48.0
20th	0.37	0.24	0.13	65.0	35.0
30th	0.44	0.31	0.13	70.1	29.9
40th	0.44	0.34	0.10	77.1	22.9
50th	0.38	0.31	0.07	80.7	19.3
60th	0.28	0.23	0.05	81.2	18.8
70th	0.18	0.15	0.03	81.7	18.3
80th	0.10	0.08	0.02	78.7	21.3
90th	0.05	0.04	0.01	76.6	23.4
Mean	0.27	0.19	0.08	70.9	29.1

6. City-level analysis

One of the benefits of modeling homeownership decisions at the household level is the flexibility to evaluate racial gaps at different levels of aggregation. Not only are we able to compare the racial gap at the national level, but also we are able to do so at the city level.[16] We illustrate this by looking at 9 primary metropolitan statistical areas

[16] Technically, we could only apply our methodology to relatively larger cities that contain a large number of observations.

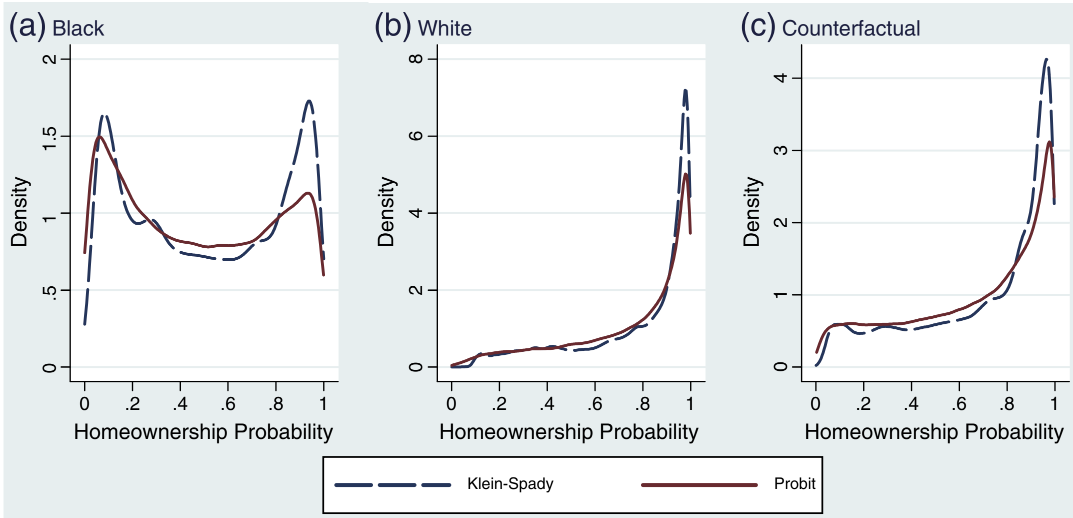

Fig. 6. Kernal density estimates of household homeownership probability.

(PMSAs). The population in these metropolitan areas range from 4.1 million to 9.5 million — the smallest is Atlanta, GA and the largest is Los Angeles-Long Beach, CA. The demographics of these 9 cities are shown in Table 3. We include in the summary statistics a Dissimilarity Index (DI), an index that is similar to a Gini coefficient and lies between zero and one with higher values corresponding to higher levels of segregation.[17] Our measure of DI is constructed from the year 2000 census-tract information.[18]

Fig. 7 summarizes the decomposition results. The homeownership gaps of Dallas and Houston are well-explained by the characteristics gap with the residual gap hovering at around 0 for both metropolitan areas. In contrast, characteristics gap explains less of the racial gaps in Chicago and New York. Both Chicago and New York have relatively large residual gaps that are fairly persistent across the percentiles, reaching a magnitude of around 15% at their peaks. The distribution of Chicago is heavier at the left-tail than New York's, suggesting that unobserved race-based factors are driving the low homeownership rates among Chicagoan blacks. When we look at the DI for both cities, we note that both Chicago and New York have larger values of DI than most cities. *Prima facie*, one would associate high levels of segregation with discrimination in the housing market. Whether or not this is true is an empirical question. Certainly in the case of Detroit, we see a high level of segregation but a somewhat lower level of residual gap across the distribution than the other two cities. This suggests that the DI offers only a partial look at discrimination, for example, it does not look at skill segregation (e.g., see Hellerstein and Neumark, (2008)) which directly affects the affordability of households to own homes and, in turn, impacts the size of the residual gap. Nonetheless, we speculate that the residual gap should increase with segregation but such a claim would require a deeper analysis that we defer to future research.[19] More importantly, graphs like those shown in Fig. 7 highlight the importance of analyzing social welfare at the local level, suggesting that policy analysis using national data might not fully reflect each individual city's racial structure.

7. Conclusion

In this paper, we use a semiparametric homeownership model to measure the contribution of characteristics to the household-level white–black homeownership gap distribution. We find that observable household characteristics are important determinants of homeownership. If blacks were to have similar characteristics as white households, their homeownership rate would improve by as much as 30 percentage points at the percentiles around the median. On the other hand, the residual gap is large at 35 to 48% of the homeownership gap at the lower percentiles. These black households are less likely to own than their white counterparts and this is not due to observable characteristics but attributable to purely race-based unobservable factors. An interesting implication arising from our approach is the flexibility of looking at disaggregated data at the household level without restraining the composition of the sample to segregated neighborhoods. We illustrate this by decomposing the racial gaps at the city-level. A useful application left for future research would be to examine the cross-section variation of residual gaps using proxies of racial discrimination, such as the Dissimilarity Index to determine important relationships between correlates of racial discrimination and the magnitude of residual gaps in homeownership.

Table 2
Probit decomposition as a percentage of the total gap.

Percentile	Total gap	Characteristics gap	Residual gap	Characteristics gap (%)	Residual gap (%)
10th	0.15	0.06	0.09	42.4	57.6
20th	0.30	0.16	0.14	53.6	46.4
30th	0.37	0.23	0.14	61.8	38.2
40th	0.39	0.26	0.13	67.1	32.9
50th	0.35	0.25	0.10	71.0	29.0
60th	0.29	0.21	0.08	73.8	26.2
70th	0.21	0.16	0.05	75.5	24.5
80th	0.13	0.10	0.03	78.2	21.8
90th	0.06	0.05	0.01	79.8	20.2
Mean	0.27	0.19	0.09	67.9	32.1

[17] See Cutler and Glaeser (1997) for further details.
[18] Available from the U.S. Census Bureau, Housing and Household Economic Statistics Division.
[19] It has been suggested to us that a possible explanation for the large residual gaps of New York and Chicago might be relatively lower rent–price ratios for blacks than for whites in these cities. A preliminary check of the data does not seem to support this. While we calculated the black–white ratio of the quality adjusted rent–price ratio for black and whites to be 51% in Chicago, the ratio was 136% in New York. This compares to the population-weighted average of 95% for the nine cities.

Table 3
City-level summary statistics.

Metro area (PMSA)	Total population	White population%	Black population%	White ownership rate	Black ownership rate	Mean ownership gap	Dissimilarity index
Atlanta, GA	4,112,198	0.61	0.31	0.81	0.56	0.25	0.65
Chicago, IL	8,272,768	0.68	0.17	0.80	0.45	0.35	0.80
Dallas-Fort Worth, TX	3,519,176	0.72	0.15	0.71	0.45	0.26	0.59
Detroit, MI	4,441,551	0.74	0.21	0.84	0.52	0.32	0.85
Houston-Brazoria, TX	4,177,646	0.65	0.18	0.72	0.49	0.23	0.67
Los Angeles-Long Beach, CA	9,519,338	0.59	0.08	0.61	0.40	0.20	0.67
New York-Northeastern NJ	9,314,235	0.63	0.17	0.64	0.33	0.31	0.82
Philadelphia, PA/NJ	5,100,931	0.74	0.19	0.79	0.53	0.26	0.72
Washington, DC/MD/VA	4,923,153	0.60	0.27	0.78	0.54	0.24	0.63

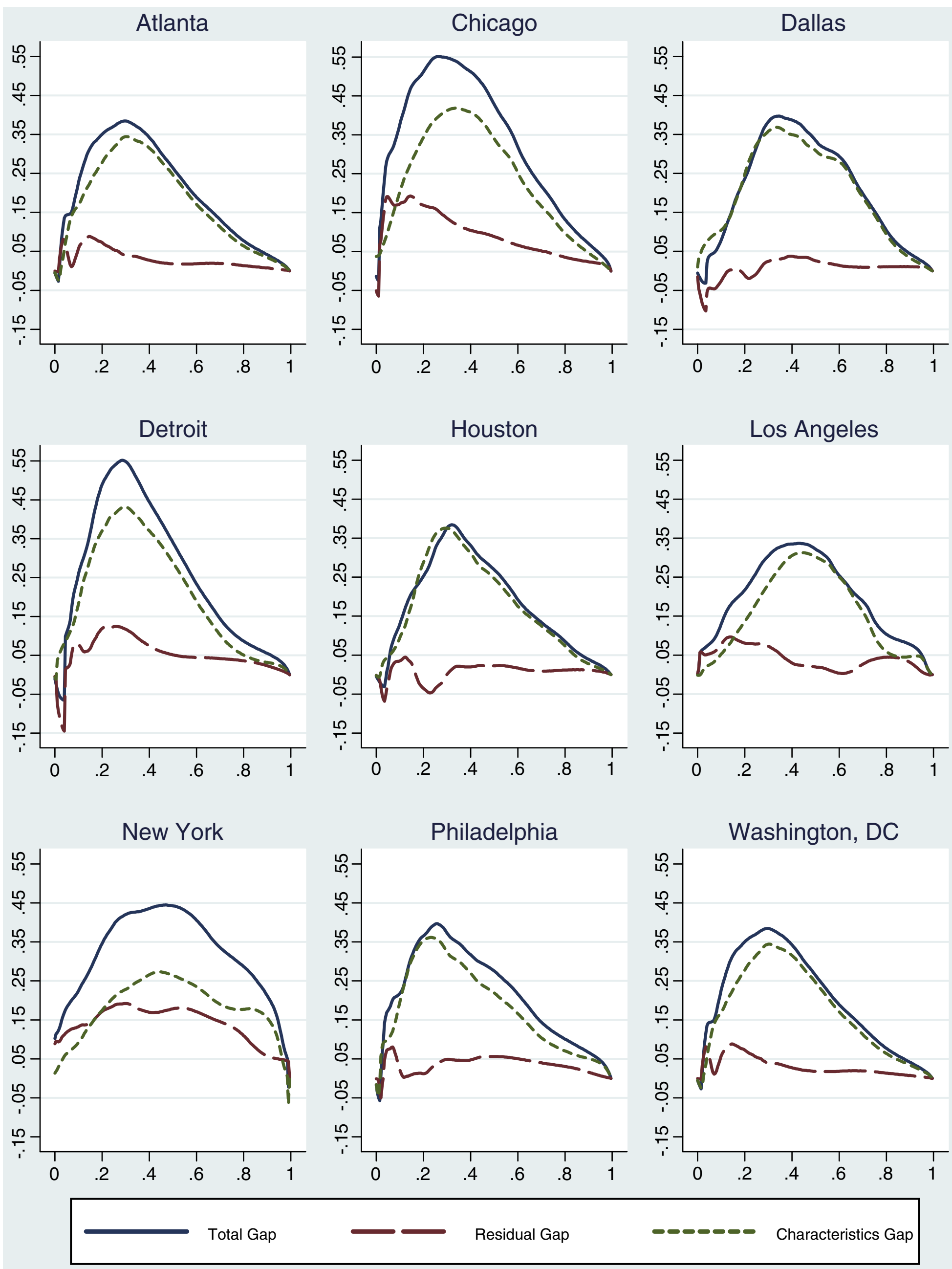

Fig. 7. City-level decompositions by percentile.

Acknowledgments

The authors thank Dan McMillen, Brian Bucks, two anonymous referees, and the participants of the 2010 AREUEA international conference, the 2010 ERES conference, and the 2011 AREUEA annual conference for helpful comments.

Appendix A. Summary statistics

	White		Black	
	Mean	Sd	Mean	Sd
Ownership rate	0.75	0.43	0.47	0.50
Log of household income	10.79	1.30	10.13	1.74
Age of head				
<30	0.09	0.28	0.13	0.34
30–40	0.17	0.37	0.20	0.40
40–50	0.22	0.41	0.24	0.43
50–60	0.21	0.41	0.20	0.40
60–70	0.14	0.35	0.12	0.32
70+	0.17	0.38	0.11	0.31
Educational attainment				
Less than grade 12	0.09	0.29	0.16	0.37
Grade 12	0.36	0.48	0.42	0.49
1 or 2 years of college	0.21	0.41	0.23	0.42
4 years of college	0.20	0.40	0.12	0.32
5+ years of college	0.13	0.34	0.07	0.25
Marital status of head				
Married — spouse present	0.58	0.49	0.31	0.46
Married — spouse absent	0.30	0.46	0.39	0.49
Single and never married	0.12	0.33	0.30	0.46
Head is American citizen	0.91	0.29	0.90	0.30
Head is male	0.57	0.50	0.41	0.49
Household size				
1	0.31	0.46	0.37	0.48
2	0.34	0.47	0.26	0.44
3	0.14	0.35	0.17	0.37
4	0.13	0.34	0.12	0.32
5+	0.08	0.27	0.09	0.29
Household earn self-employed income	0.12	0.33	0.06	0.24
Household has interest, dividend, or rental income	0.32	0.46	0.07	0.26
Number of earners in household				
1	0.27	0.44	0.29	0.46
2	0.39	0.49	0.46	0.50
3+	0.34	0.47	0.24	0.43
Head has difficulty working due to a disability	0.12	0.32	0.16	0.36
Head's occupation income score	23.45	16.17	20.23	13.89
Household moved into residence				
1 or 2 years ago	0.18	0.39	0.25	0.44
3 years ago	0.19	0.40	0.22	0.42
4 years ago	0.19	0.39	0.18	0.38
5 years ago	0.20	0.40	0.16	0.36
6 years ago	0.10	0.30	0.08	0.27
7+ years ago	0.13	0.33	0.11	0.31
Metropolitan area	0.82	0.38	0.90	0.29
Region				
New England	0.05	0.22	0.02	0.14
Middle Atlantic	0.15	0.36	0.16	0.36
East North Central	0.18	0.38	0.16	0.37
West North Central	0.08	0.27	0.03	0.18
Southern Atlantic	0.18	0.39	0.32	0.47
East South Central	0.05	0.23	0.10	0.29
West South Central	0.10	0.30	0.12	0.33
Mountain	0.07	0.26	0.02	0.13
Pacific	0.14	0.35	0.07	0.26

Appendix B. Klein–Spady estimates

	White		Black	
	Estimate	Std err	Estimate	Std err
Log of household income	0.12	0.001	0.14	0.002
Age of head (<30 excluded)				
30–40	0.06	0.002	0.06	0.006
40–50	0.09	0.002	0.08	0.006
50–60	0.11	0.002	0.09	0.006
60–70	0.14	0.003	0.13	0.007
70+	0.11	0.003	0.17	0.008

Appendix B (*continued*)

	White		Black	
	Estimate	Std err	Estimate	Std err
Educational attainment				
(Less than grade 12 excluded)				
Grade 12	0.05	0.002	0.04	0.004
1 or 2 years of college	0.07	0.002	0.07	0.005
4 years of college	0.09	0.002	0.12	0.006
5+ years of college	0.06	0.003	0.15	0.007
Marital status of head				
(Married — spouse present excluded)				
Married — spouse absent	−0.12	0.002	−0.11	0.004
Single and never married	−0.18	0.003	−0.18	0.005
Head is American citizen	0.07	0.002	−0.03	0.005
Head is male	−0.01	0.001	−0.01	0.003
Household size (1 excluded)				
2	0.03	0.002	0.02	0.004
3	0.05	0.003	0.01	0.005
4	0.07	0.003	0.02	0.006
5+	0.05	0.003	0.01	0.006
Household earn self-employed income	0.02	0.002	0.03	0.006
Household has interest, dividend, or rental income	0.10	0.002	0.13	0.006
Number of earners in household (1 excluded)				
2	−0.03	0.002	−0.04	0.004
3+	0.00	0.003	−0.04	0.006
Head has difficulty working due to a disability	−0.06	0.002	−0.02	0.004
Head's occupation income score	0.00	0.000	0.00	0.000
Household moved into residence				
(1 or 2 years ago excluded)				
3 years ago	0.15	0.002	0.12	0.004
4 years ago	0.26	0.002	0.23	0.004
5 years ago	0.35	0.002	0.33	0.005
6 years ago	0.45	0.002	0.42	0.006
7+ years ago	0.61	0.002	0.56	0.007
Metropolitan area	−0.04	0.002	−0.04	0.005
Region (New England excluded)				
Middle Atlantic	−0.01	0.003	−0.05	0.011
East North Central	0.10	0.003	0.12	0.010
West North Central	0.10	0.004	0.15	0.012
Southern Atlantic	0.09	0.003	0.15	0.009
East South Central	0.12	0.004	0.21	0.010
West South Central	0.07	0.003	0.15	0.010
Mountain	0.08	0.003	0.11	0.014
Pacific	−0.04	0.003	0.00	0.012

Appendix C. Klein–Spady and probit marginal effects

	White		Black	
	Klein–Spady	Probit	Klein–Spady	Probit
Log of household income	0.07	0.03	0.12	0.03
Age of head (<30 excluded)				
30–40	0.04	0.05	0.05	0.09
40–50	0.05	0.08	0.07	0.11
50–60	0.06	0.09	0.08	0.13
60–70	0.08	0.11	0.11	0.17
70+	0.06	0.08	0.14	0.20
Educational attainment				
(Less than grade 12 excluded)				
Grade 12	0.03	0.05	0.03	0.06
1 or 2 years of college	0.04	0.08	0.06	0.11
4 years of college	0.05	0.11	0.10	0.18
5+ years of college	0.04	0.11	0.13	0.23
Marital status of head				
(Married — spouse present excluded)				
Married — spouse absent	−0.07	−0.11	−0.10	−0.13
Single and never married	−0.13	−0.16	−0.16	−0.19
Head is American citizen	0.05	0.07	−0.03	−0.01
Head is male	−0.01	0.00	−0.01	0.01

Appendix C (*continued*)

	White		Black	
	Klein–Spady	Probit	Klein–Spady	Probit
Household size (1 excluded)				
2	0.02	0.03	0.02	0.03
3	0.03	0.05	0.01	0.03
4	0.04	0.07	0.02	0.04
5+	0.03	0.06	0.01	0.03
Household earn self-employed income	0.01	0.01	0.03	0.02
Household has interest, dividend, or rental income	0.06	0.09	0.11	0.15
Number of earners in household (1 excluded)				
2	−0.02	−0.02	−0.03	0.00
3+	0.00	0.00	−0.03	0.05
Head has difficulty working due to a disability	−0.04	−0.07	−0.02	−0.03
Household moved into residence (1 or 2 years ago excluded)				
3 years ago	0.08	0.12	0.10	0.11
4 years ago	0.13	0.20	0.19	0.24
5 years ago	0.17	0.27	0.27	0.32
6 years ago	0.17	0.34	0.31	0.41
7+ years ago	0.19	0.44	0.40	0.53
Metropolitan area	−0.02	−0.02	−0.04	−0.03
Region (New England excluded)				
Middle Atlantic	−0.01	−0.01	−0.04	−0.04
East North Central	0.06	0.07	0.10	0.09
West North Central	0.05	0.06	0.13	0.10
Southern Atlantic	0.05	0.07	0.13	0.13
East South Central	0.06	0.08	0.17	0.16
West South Central	0.04	0.05	0.13	0.12
Mountain	0.05	0.06	0.09	0.09
Pacific	−0.03	−0.02	0.00	0.00

References

Angrist, J.D., 2001. Estimation of limited-dependent variable models with dummy endogenous regressors: simple strategies for empirical practice. Journal of Business & Economic Statistics 19 (1), 2–16.

Boehm, T.P., 1981. Tenure choice and expected mobility: a synthesis. Journal of Urban Economics 10 (3), 375–389.

Bostic, R.W., Surette, B.J., 2001. Have the doors opened wider? Trends in homeownership rates by race and income. Journal of Real Estate Finance and Economics 23 (3), 411–434.

Cameron, A.C., Trivedi, P.K., 2005. Microeconometrics: Methods and Applications. Cambridge University Press, New York.

Carrillo, P., Yezer, A., 2009. Alternative measures of homeownership gaps across segregated neighborhoods. Regional Science and Urban Economics 39 (5), 542–552.

Cobb-Clark, D.A., Sinning, M.G., 2011. Neighborhood diversity and the appreciation of native- and immigrant-owned homes. Regional Science and Urban Economics 24 (3), 214–226.

Collins, W.J., Margo, R.A., 2001. Race and home ownership: a century-long view. Explorations in Economic History 68–92.

Cosslett, S., 1987. Efficiency bounds for distribution-free estimators of the binary choice and the censored regression models. Econometrica 55, 559–585.

Cutler, D.M., Glaeser, E.L., 1997. Are ghettos good or bad? Quarterly Journal of Economics 112, 827–872.

Deng, Y., Ross, S.L., Wachter, S.M., 2003. Racial differences in homeownership: the effect of residential location. Regional Science and Urban Economics 33 (5), 517–556.

DiNardo, J., Fortin, N.M., Lemieux, T., 1996. Labor market institutions and the distribution of wages, 1973–1992: a semi-parametric approach. Econometrica 64, 1001–1044.

Duca, J.V., Rosenthal, S.S., 1994. Borrowing constraints and access to owner-occupied housing. Regional Science and Urban Economics 24 (3), 301–322.

Gabriel, S., Rosenthal, S., 2005. Homeownership in the 1980s and 1990s: aggregate trends and racial disparities. Journal of Urban Economics 57 (1), 101–127.

Gale, W.G., Scholz, J.K., 1994. Intergenerational transfers and the accumulation of wealth. Journal of Economic Perspectives 8 (4), 145–160.

Green, R.K., White, M.J., 1997. Measuring the benefits of homeowning: effects on children. Journal of Urban Economics 41 (3), 441–461.

Haurin, D.R., Henderschott, P.H., Wachter, S.M., 1997. Borrowing constraints and the tenure choice of American youth. Journal of Housing Research 8, 137–154.

Hellerstein, J., Neumark, D., 2008. Workplace segregation in the United States: race, ethnicity, and skill. The Review of Economics and Statistics 90 (3), 459–477.

Henderson, J.V., Ionnides, Y.M., 1989. Dynamic aspects of consumer decisions in housing markets. Journal of Urban Economics 26 (2), 212–230.

Herbert, C.E., Haurin, D.R., Rosenthal, S., Duda, M., 2005. Homeownership gaps among low-income and minority borrowers and neighborhoods. U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, DC.

Kain, J.F., Quigley, J.M., 1972. Housing market discrimination. American Economic Review 62 (3), 263–277.

Klein, R.W., Spady, R.H., 1993. An efficient semiparametric estimator for binary response models. Econometrica 61 (2), 387–421.

Linneman, P., Wachter, S., 1989. The impacts of borrowing constraints on homeownership. Real Estate Economics 17 (4), 389–402.

Long, J.E., Caudill, S.B., 1992. Racial differences in homeownership and housing wealth, 1970–1986. Economic Inquiry 30 (1), 83–100.

Machado, J.A.F., Mata, J., 2005. Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of Applied Econometrics 20 (4), 445–465.

Manski, C.F., Thompson, S.T., 1989. Estimation of best predictors of binary response. Journal of Econometrics 40 (1), 97–123.

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

Munnell, A.H., Tootell, G.M.B., Browne, L.E., McEneaney, J., 1996. Mortgage lending in Boston: interpreting HMDA Data. American Economic Review 86 (1), 25–53.

Powell, J.L., Stock, J., Stoker, T., 1989. Semiparametric estimation of selection models: some empirical results. Econometrica 57, 1403–1430.

Rosenthal, S., 1988. A residence time model of housing markets. Journal of Public Economics 36, 87–109.

Silberman, J., Yochum, G., Ihlanfeldt, K., 1982. Racial differentials in home purchase: the evidence from newly-formed households. Economic Inquiry 20 (3), 443–457.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman and Hall, London.

Wachter, S.M., Megbolugbe, I.F., 1992. Racial and ethnic disparities in homeownership. Housing Policy Debate 3 (2), 333–370.

Yinger, J., 1995. Closed doors, Opportunities Lost: the Continuing Costs of Housing Discrimination. Russell Sage Foundation Publications.