Institute for International Economic Policy Working Papers Series
Elliott School of International Affairs
George Washington University

Alternative Measures of Homeownership Gaps Across Segregated Neighborhoods

IIEP-WP#7

Paul Carrillo
George Washington University

Anthony Yezer
George Washington University

February 2008

Institute for International Economic Policy
1957 E St. NW, Suite 501
Voice: (202) 994-5320
Fax: (202) 994-5477
Email: **iiep@gwu.edu**
Web: **www.gwu.edu/~iiep**

Alternative Measures of Homeownership Gaps Across Segregated Neighborhoods*

Paul Carrillo[†] Anthony Yezer[††]

Department of Economics

The George Washington University

February 2008

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 interest in the relation between credit conditions and homeownership gaps. Statistical analysis of these differences or "gaps" in homeownership between white and minority households follows what has become a highly stylized pattern. Essentially differences in homeownership at the mean or the conditional mean between groups are compared. This study implements a new decomposition technique that identifies the unexplained portion of the gap not only at the mean, but at every percentile of the distribution of the dependent variable. This method was first proposed by Machado and Mata (2005), extended by Albrecht et al. (2006), and has been used in several applications in labor economics. Similar to the labor market application, differences in homeownership gaps at the mean reflect a combination of non-significant differences at the upper end and much larger gaps at the lowest end of the distribution of homeowners.

Keywords: Homeownership; Race; Quantile regression decomposition

JEL Codes: R21; J15; C15

* An earlier version of this paper was presented at the ASSA meetings of AREUEA in New Orleans, Louisiana, January 4-7, 2008. We wish to thank Nathan Anderson, the discussant, and conference participants for helpful comments. Contact either author with correspondence regarding this paper.

[†] Address: Monroe Hall, 2115 G Street NW Suite 364, Washington DC, 20052; email: pcarrill@gwu.edu; phone: 202-994-7608.

[††] Address: Monroe Hall, 2115 G Street NW Suite 320, Washington DC, 20052; email: yezer@gwu.edu; phone: 202-994-6755.

Alternative Measures of Homeownership Gaps Across Segregated Neighborhoods

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 sharp decline in subprime lending will likely result in further changes in homeownership gaps.

Discussion of homeownership gaps has evolved over time and become highly stylized in the literature. First simple mean homeownership rates of different groups are compared. Second mean homeownership rates at various quintiles of the distribution of a single characteristic such as income or age are analyzed.[1] Third differences in the conditional mean homeownership rate adjusted for determinants of homeownership other than ethnicity are compared. In essence a tenure-choice equation is estimated with the size and significance of dummy variables for minority status providing the basis for

[1] 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 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.

measuring gaps.[2] Fourth, the Oaxaca-Blinder technique, sometimes modified because probit models are non-linear using Fairlie's (2005) approach[3], is 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 choice equation.[4] Fifth, Gabriel and Rosenthal (2005) have modified the Oaxaca-Blinder technique to decompose changes in homeownership gaps over time into one component 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.[5]

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

---

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

[3] 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.

[4] See, for example, Coulson (1999), Painter, Gabriel and Myers (2001), Wachter and Megbolugbe (1992), Myers and Chung (1996)

[5] See, for example, the application of this technique in Myers, and Lee (1998), Myers, Megbolugbe, and Lee (1998), and Myers, Painter, Yu, Ryu, and Wei, (2005).

technique. 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. Presumably, this displays 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.[6] 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 highest skilled women.[7]

The goal of this research is to apply the Machado-Mata technique to the measurement of homeownership gaps. The principle 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. The requirement that the dependent variable be continuous, forces some changes in the structure of the test conducted here. Specifically, the dependent variable is the fraction owner-occupants in highly segregated census block groups (CBG). The fraction owner-occupants is continuous and confining

[6] 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 2007) differentials.
[7] 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.

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. Changing from probit estimates of individual tenure-choice decisions to OLS estimates of the fraction owner-occupied from CBGs has two principle 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 choice equation rather than individual characteristics.

Given that the test proposed here for CBGs appears to differ substantially from tests using individual tenure-choice models, the first task will be to determine if the approach using segregated CBGs produces results comparable to the individual tenure choice models. This is done by testing to see if analysis of homeownership gaps using the GBG 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-choice 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. Accordingly, it appears 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. The last section concludes.

2. Data and Measurement Issues

The first step in analyzing differences in tenure rates between segregated communities 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 [8] where the share of white population is above a

[8] 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.

threshold *R*, where $0<R<1$.[9] 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 Metropolitan Statistical Areas (MSA) that have at least 100 thousand residents was used to identify segregated areas.[10]

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 three different (high) choices of *R*. The first, second, and third pair of columns show characteristics of white and non-white neighborhoods when the threshold *R* equals 0.95, 0.97, and 0.97, respectively.[11] In all specifications, the (simple) average homeownership rate in white CBGs is about 82 percent while in non-white CBGs it is close to 48 percent. 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.

[Insert Table 1]

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 percent. This provides an incentive to pick a low threshold. It is possible, however, that low values of *R* may not accurately describe a segregated

[9] "whites" are defined as non-hispanic white individuals; based on the variable P7.3 from the US Census.
[10] There are 161,560 CBGs in these selected areas.
[11] Areas where homeownership rates were missing were dropped (in other words, areas with no population were excluded).

neighborhood. For this reason, the intermediate threshold of 0.97 is chosen.[12] Using this definition, there are 17,520 white and 12,017 minority CBGs, respectively.

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 choice equations have been estimated that incorporate several economic and household type controls, and it has been established that differences in endowments and household structure account for a large portion of the gap.[14]

The extent to which other characteristics of the neighborhoods explain the mean homeownership gap can be explored. The previous literature provides guidance for identifing the set of controls included in the specification. In particular, tenure choice equations in previous studies have included independent variables such as age, marital status, income and wealth, education, and immigration status.[15] Descriptive statistics for both white and non-white areas used in this study are found in Table 2.

---

[12] The main results of the paper, however, are similar regardless of this choice.

[13] For example, see Coulston (1999), Painter, Gabriel and Myers (2001), Deng, Ross, and Wachter (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, Hebert, and Rosenthal (2007) make 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, Gabriel and Myers (2001), Wachter and Megbolugbe (2002), and Gabriel and Rosenthal (2005).

White neighborhoods are less dense than non-white neighborhoods. For instance, the mean population density of a white CBG is about one ninth that of its counterpart in a non-white neighborhood. There are also important differences in income and employment indicators between white and non-white areas. For example, the mean of the median household incomes in white CBGs is more than twice as large as the mean of the median incomes in non-white CBGs. Furthermore, the mean unemployment rate is almost five times lower in white neighborhoods. Finally, it seems that white areas tend to have a higher proportion of people older than 65.

[Insert Table 2]

To explore how the characteristics of neighborhoods "affect" mean homeownership rates a linear model is used. 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.

[Insert Table 3]

The first three 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 characteristics. The second and third specifications include variables that explain economic and migration status, respectively. 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.

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. Indeed, once the full set of controls is included, the conditional mean gap virtually disappears. This result is consistent with other findings in the literature using individual tenure choice equations and suggests that the unexplained portion of the homeownership gap is very small.

The pooled regressions shown on Table 3 assume that the marginal effects of neighborhood characteristics on homeownership are the same in white and non-white neighborhoods. In the fourth and fifth columns this assumption is relaxed and separate regressions for each group are estimated. The Oaxaca-Blinder decomposition is then used to estimate the unexplained portion of the homeownership gap. That is, the estimated coefficients in column (4) 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 portion, 0.78 = (0.82 - 0.56)/0.34), 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, Hebert, and Rosenthal (2005). 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 choice 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 choice models.

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.

Unconditional differences [16]

Figure 1 displays the distribution of homeownership rates across CBGs for both white and non-white neighborhoods. Note that, in the bottom 10 percent of white neighborhoods homeownership rates are below 63 percent, while in the bottom 10 percent of non-white CBG homeownership rates are below 8 percent. Thus, the homeownership rate gap at the 10th percentile is 55 percentage points. Figure 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 36, 21 and 13 percentage points at the 50th, 75th, and 90th percentiles, respectively.

[16] McMillen and Singell (2007) use a similar approach to compare changes in the distribution of district-level real expenditures per student and class sizes over time.

[Insert Figure 1]

[Insert Figure 2]

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 tail of the distributions.[17]

Figure 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.

Quantile regressions

Quantile regressions can be used to assess what fraction of the homeownership gap observed in Figure 2 remains after adjusting for the effects of differences in endowments across these segregated areas. 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

$$Q_\theta[y|X] = X\delta_\theta, \qquad (1)$$

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.[18] Estimation of the quantile parameters, the δ_θ, is performed as the solution to

[17] One may argue that the estimated gap displayed in Figure 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, Figure 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 Figure 2. No significant differences were found. For details, please contact either author.

$$\arg\min_{\delta_\theta}\{\sum_{i:y_i>X_i\delta_\theta}\theta\,|\,y_i-X_i\delta_\theta\,|+\sum_{i:y_i\le X_i\delta_\theta}(1-\theta)\,|\,y_i-X_i\delta_\theta\,|\}. \qquad (2)$$

Quantile regression models were introduced by Koenker and Basset (1978). There have been many applications of quantile regression models in the literature including recent studies such as McMillen and Thorsnes (2007), Nguyen et al. (2007), Albrecht, Bjorklund and Vroman (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 Figure 1.

The form of the estimated quantile regressions is:

$$Q_\theta[\mathrm{y}|\mathrm{Z},\mathrm{W}] = \alpha_\theta + \beta_\theta \mathrm{W} + Z\gamma_\theta, \qquad (3)$$

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 $Q_\theta[\mathrm{y}|\mathrm{Z},\mathrm{W}]$ 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

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

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, and fifth rows incorporate education, income, and immigration variables, respectively.

[Insert Table 4]

The estimates of β_θ in the first row are, by construction, equivalent to those depicted in Figure 1. This coefficient decreases significantly for every quantile as explanatory variables are added. For example, it diminishes from 0.36 to 0.042 for the median case (θ=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 equation (2). 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.

[Insert Table 5]

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 test for heterogeneous effects, separate quantile regressions are estimated for each group and results are shown in Table 6. Clearly, there are important differences. 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. The share of population that does not speak English well has a positive (and large) association with homeownership rates in non-white neighborhoods.[19]

[Insert Table 6]

Because all the quantile coefficients differ between white and non-white areas, the estimates of β_θ in equation (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.

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) which identifies the sources of the differences between the means of two distributions. The advantage of the

[19] It may be that the foreign language effect distinguishes among groups within the non-white community, in particular, identifying immigrant communities.

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 2007) 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^W_θ be the coefficient of the θth conditional quantile regression of homeownership rates in white neighborhoods; that is, $Q_\theta[y^W|Z^W]= Z^W \gamma^W_\theta$. The counterfactual distribution is generated as follows:

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

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

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

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. (2006) have shown that the decomposition is consistent and asymptotically normal.

The Machado-Mata method is used to estimate $Q_\theta[y^*]$, 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 0.62 which is 6 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 Figure 2, quantile decompositions are more appropriate to estimate a decomposition of the homeownership gap.

The estimated $Q_\theta[y^*]$ 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 on homeownership rates in non-white areas. That is, a measure of $Q_\theta[y^*] - Q_\theta[y^{NW}]$ is estimated. This is the "unexplained" homeownership gap at each quantile of the distribution.

[Insert Figure 3]

The results are illustrated in Figure 3 and suggest that the unexplained portion of the gap is much larger at the left tail of the distribution. For instance, unexplained factors account for about 25 percentage points or about 48 percent 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 and represents about 40 and 25 percent of the total gap at the 50th and 75th percentile, respectively. Indeed, at the 75th percentile, the unexplained gap is no longer statistically different from zero.

4. 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 tenure choice equation literature. Furthermore, 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 non-significant while that at the lower end of the distribution is statistically significant and substantial.

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

References

Albrecht, J., A. Bjorklund and S. Vroman, (2003), "Is There a Glass Ceiling in Sweden? *Journal of Labor Economics*, 21(1), 145-177.

Albrecht, J., A. Vuuren and S. Vroman, (2006), "Counterfactual Distributions with Sample Selection Adjustments: Econometric Theory and an Application to the Netherlands", unpublished manuscript, Georgetown University.

Arulampalam W., A. Booth, and M. Bryan, (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. and H. Chen, (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., S. Ross, and S. Wachter, (2003), "Racial differences in homeownership: the effect of residential location", *Regional Science and Urban Economics*, 33, 517-556.

Department of Housing and Urban Development, United States, (1991), "The President's National Urban Policy Report", Washington DC.

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., and G. Painter, (2003), "Paths to homeownership: An analysis of the residential location and homeownership choices of black households", *Journal of Real Estate Finance and Economics*, 27(1), 87-106.

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

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

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

Kain, J.F. and J.M. Quigley (1972). “Housing market discrimination, homeownership, and savings behavior”, *American EconomicReview*, 263-277.

Koenker R., (2005), “Quantile Regression”, Econometric Society Monograph, Cambridge University Press.

Koenker, R. and G. Bassett, (1978), “Regression quantiles”, *Econometrica*, 46, 33-50.

Koenker, R. and K. Hallock, (2001), “Quantile regression”, *Journal of EconomicPerspectives*, 15, 143-156.

Machado, J. and J. Mata, (2005), “Counterfactual decomposition of changes in wage distributions using quantile regression”, *Journal of Applied Econometrics*, 20, 445-465.

McMillen, D., (2007), “Changes in the distribution of house prices over time: structural characteristics, neighborhood, or coefficients?”, unpublished manuscript, University of Illinois at Chicago.

McMillen, D. and L. Singell, (2007), “A distributional difference-in-difference evaluation of school expenditures and class-sizes to reform and tax limits”, unpublished manuscript, University of Illinois at Chicago.

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

Myers, D. and S.W. Lee. (1998), “Immigrant trajectories into homeownership: a temporal analysis of residential assimilation”, *International Migration Review*, 593-624.

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

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

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

Nguyen, B., J. Albrecht, S. Vroman, and D. Westbrook, (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., S. Gabriel, and D. Myers, (2001), "Race, immigrant status, and housing tenure choice", *Journal of Urban Economics*, 49, 150-167.

Roistacher, E.A. and J.L. Goodman, (1976). "Race and home ownership: is discrimination disappearing?", *Economic Inquiry*, 59-70.

Wachter, S., and I. Megbolugbe, (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.

Table 1
Homeownership rate in segregated neighborhoods*

	(1)		(2)		(3)	
	White	**Non-white**	**White**	**Non-white**	**White**	**Non-white**
Mean rate (unweighted)[a]	0.823	0.476	0.827	0.483	0.828	0.485
Mean rate (population-weighted)[b]	0.832	0.469	0.836	0.481	0.838	0.493
Number CBGs	28,495	14,900	17,520	12,017	7,650	6,647
Total population (millions)	35.7	18.4	20.6	14.3	8.4	6.6

Note (*): A segregated "white neighborhood" is defined as a Census Block Group (CBG) where the share of white population exceeds 0.95, 0.97, and 0.99 in (1), (2), and (3), respectively. Accordingly, a "non-white" neighborhood is a CBG where the share of white population is below 0.05, 0.03, and 0.01 in (1), (2), and (3), respectively.

[a] Mean rates are average homeownership rates in each group of CBGs.

[b] Mean rates weighted by the total population of each CBG.

Table 2: Descriptive Statistics (mean and standard deviation)

Variables	All CBG	"White" CBG	"Non-White" CBG
HO rate **:** Number of owner-occupied housing units divided by the total occupied units in Census Block Group (CBG).	0.687 (0.268)	0.827 (0.147)	0.483 (0.274)
White : Equals one if CBG's share of white population is above 0.97.	0.593 (0.491)		
Density **:** Population density (total population per hectare).	33.062 (71.970)	8.387 (15.237)	69.068 (101.069)
Older 65 : Share of population older than 65.	0.145 (0.107)	0.167 (0.120)	0.113 (0.074)
Family **:** Share of family households.	0.733 (0.131)	0.732 (0.126)	0.736 (0.138)
Married : Proportion of population (15 and older) who is married.	0.543 (0.156)	0.632 (0.101)	0.413 (0.129)
HS dropout **:** Share of population (25 and older) that does not have a Highschool diploma.	0.253 (0.190)	0.141 (0.098)	0.418 (0.170)
Median income : Median household income in CBG ($ thousands) in 1999.	41.563 (23.849)	53.084 (22.757)	24.752 (12.912)
Unemployment rate **:** Share of population (16 and older) in the labor force that are unemployed.	0.090 (0.097)	0.038 (0.039)	0.165 (0.105)
Bad english : Share of population (between 18 and 64 years old) that does not speak English well.	0.072 (0.162)	0.004 (0.012)	0.171 (0.219)
Nobs **:** Number of observations (CBG) *	29,469	17,486	11,983

Note: * We drop a few CBGs with missing values from our sample. Thus, the number of CBGs shown in this Table is slightly smaller than the ones on Table 1.

Table 3: Determinants of mean homeownership rates

Independent Variables	(1) Pooled	(2) Pooled	(3) Pooled	(4) White	(5) Non-white
Constant	-0.034 ***	-0.450 ***	-0.516 ***	-0.674 ***	-0.837 ***
	(0.010)	(0.039)	(0.040)	(0.052)	(0.055)
White	0.146 ***	0.020 ***	0.034 ***		
	(0.004)	(0.005)	(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	0.381 ***	0.400 ***	0.408 ***	0.255 ***	1.178 ***
	(0.014)	(0.014)	(0.014)	(0.014)	(0.042)
Family	0.458 ***	0.435 ***	0.436 ***	0.529 ***	0.518 ***
	(0.016)	(0.016)	(0.016)	(0.023)	(0.024)
Married	0.517 ***	0.449 ***	0.388 ***	0.203 ***	0.510 ***
	(0.017)	(0.017)	(0.019)	(0.028)	(0.028)
HS dropout		-0.304 ***	-0.328 ***	-0.034 **	-0.454 ***
		(0.011)	(0.012)	(0.017)	(0.015)
Log median household income		0.060 ***	0.068 ***	0.089 ***	0.083 ***
		(0.004)	(0.004)	(0.005)	(0.006)
Unemployment rate		-0.075 ***	-0.079 ***	0.028	0.020
		(0.022)	(0.022)	(0.048)	(0.025)
Bad english			0.095 ***	-0.454 ***	0.126 ***
			(0.012)	(0.097)	(0.016)
Observations	29,469	29,469	29,469	17,486	11,983
R-squared	0.635	0.683	0.684	0.487	0.544

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 parenthesis. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

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	0.561 *** (0.005)	0.509 *** (0.003)	0.362 *** (0.002)	0.211 *** (0.002)	0.133 *** (0.002)
Constant and Basic Demographics (Density, Older 65, Family, Married)	0.313 *** (0.006)	0.218 *** (0.004)	0.115 *** (0.003)	0.077 *** (0.003)	0.065 *** (0.003)
Constant, Basic Demographics, and Education (High School Dropouts)	0.171 *** (0.006)	0.093 *** (0.004)	0.031 *** (0.003)	0.005 (0.003)	0.005 (0.004)
Constant, Basic Demographics, Education, and Income (Median H. income, Unemployment)	0.163 *** (0.006)	0.079 *** (0.004)	0.016 *** (0.003)	-0.014 *** (0.003)	-0.016 *** (0.004)
Constant, Basic Demographics, Education, Income, and Immigration (Bad English)	0.148 *** (0.006)	0.075 *** (0.004)	0.042 *** (0.003)	0.025 *** (0.003)	0.011 *** (0.004)

Note: The Table displays estimates for the parameter "beta" in equation (3) for several specifications and quantiles. The dependent variable in each quantile regression is 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 29,469. Standard errors are in parenthesis. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

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

Independent Variables	Quantile regression (percentile)				
	(10)	(25)	(50)	(75)	(90)
Constant	-0.800 ***	-0.519 ***	-0.350 ***	-0.330 ***	-0.204 ***
	(0.058)	(0.034)	(0.025)	(0.023)	(0.026)
White	0.148 ***	0.075 ***	0.042 ***	0.025 ***	0.011 ***
	(0.006)	(0.004)	(0.003)	(0.003)	(0.004)
Density	-0.001 ***	-0.001 ***	-0.001 ***	-0.001 ***	-0.001 ***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Older 65	0.288 ***	0.344 ***	0.374 ***	0.357 ***	0.286 ***
	(0.020)	(0.012)	(0.008)	(0.008)	(0.009)
Family	0.568 ***	0.535 ***	0.479 ***	0.374 ***	0.246 ***
	(0.021)	(0.011)	(0.009)	(0.008)	(0.010)
Married	0.489 ***	0.458 ***	0.343 ***	0.225 ***	0.136 ***
	(0.024)	(0.013)	(0.010)	(0.010)	(0.012)
HS dropout	-0.249 ***	-0.323 ***	-0.308 ***	-0.205 ***	-0.148 ***
	(0.015)	(0.009)	(0.007)	(0.007)	(0.008)
Log median income	0.059 ***	0.050 ***	0.053 ***	0.071 ***	0.078 ***
	(0.006)	(0.003)	(0.002)	(0.002)	(0.003)
Unemployment rate	-0.143 ***	-0.184 ***	-0.167 ***	-0.101 ***	-0.054 ***
	(0.025)	(0.015)	(0.012)	(0.012)	(0.013)
Bad english	-0.187 ***	-0.044 ***	0.170 ***	0.192 ***	0.169 ***
	(0.013)	(0.008)	(0.007)	(0.007)	(0.008)

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 29,469. Standard errors are in parenthesis. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

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	-0.578 ***	-1.087 ***	-0.451 ***	-0.958 ***	-0.244 ***	-0.753 ***
	(0.044)	(0.070)	(0.031)	(0.055)	(0.026)	(0.057)
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	0.240 ***	1.442 ***	0.257 ***	1.298 ***	0.232 ***	1.082 ***
	(0.010)	(0.045)	(0.007)	(0.031)	(0.007)	(0.030)
Family	0.705 ***	0.609 ***	0.537 ***	0.565 ***	0.354 ***	0.497 ***
	(0.015)	(0.025)	(0.010)	(0.019)	(0.010)	(0.021)
Married	0.254 ***	0.542 ***	0.188 ***	0.596 ***	0.118 ***	0.550 ***
	(0.017)	(0.029)	(0.012)	(0.023)	(0.011)	(0.025)
HS dropout	-0.080 ***	-0.489 ***	-0.062 ***	-0.484 ***	-0.050 ***	-0.395 ***
	(0.014)	(0.018)	(0.010)	(0.014)	(0.009)	(0.014)
Log median income	0.061 ***	0.087 ***	0.069 ***	0.089 ***	0.072 ***	0.085 ***
	(0.004)	(0.007)	(0.003)	(0.006)	(0.003)	(0.006)
Unemployment rate	0.026	0.012	0.050 **	0.007	0.009	0.052 **
	(0.027)	(0.029)	(0.020)	(0.023)	(0.019)	(0.023)
Bad english	-0.476 ***	0.085 ***	-0.297 ***	0.141 ***	-0.145 ***	0.155 ***
	(0.083)	(0.017)	(0.060)	(0.014)	(0.052)	(0.015)
Observations	17,486	11,983	17,486	11,983	17,486	11,983

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 parenthesis. *, **, and *** denote significance at the 10, 5, and 1 percent level, respectively.

Figure 1: Distribution of Homeownership Rates Across Racially Segregated Census Block Groups

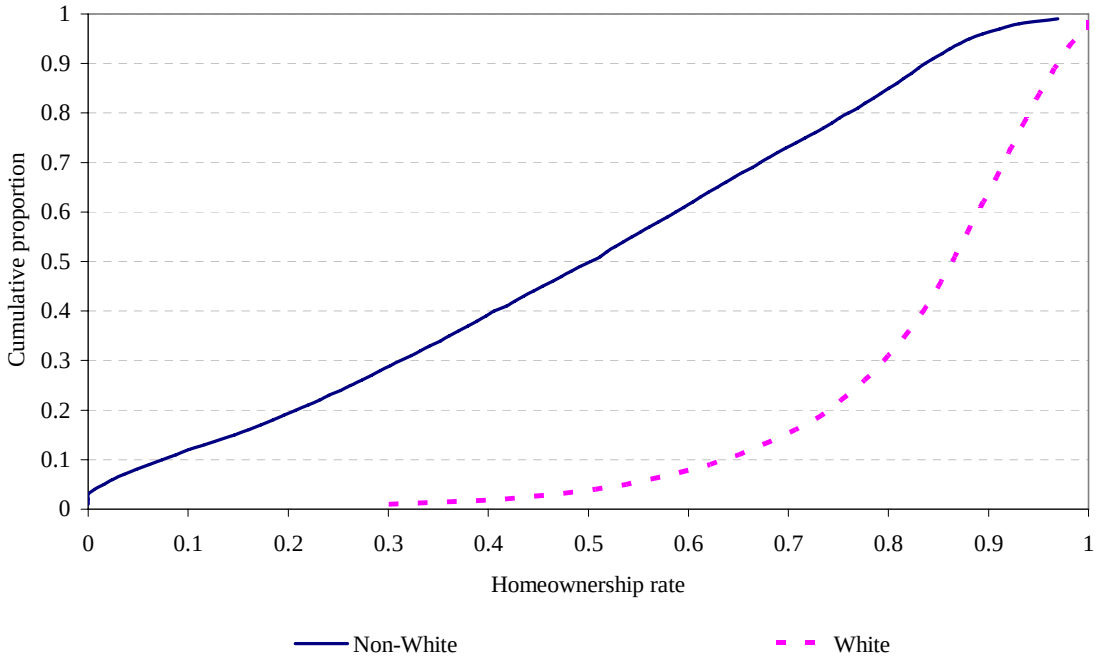

Figure 2: Difference in the Distribution of Homeownership Rates Between White and Nonwhite Neighborhoods

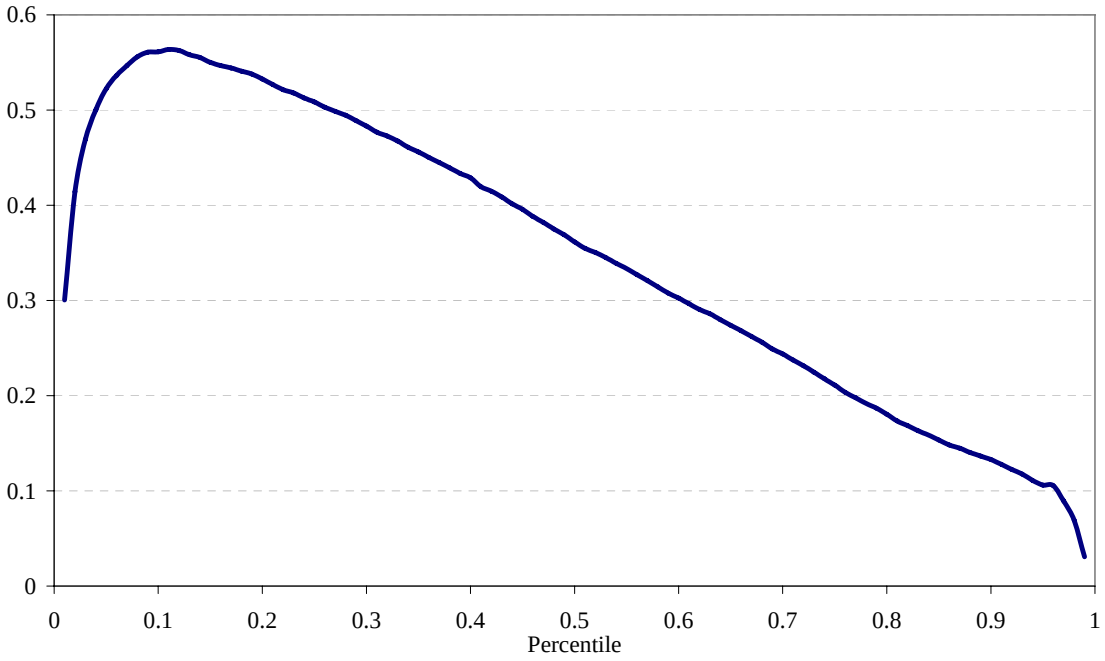

Figure 3: Difference between the counterfactual distribution of homeownership rates if white CBGs had non-white characteristics and the actual distribution of homeownership rates in non-white areas.