ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 71-72 – 2003

Residential location theory and the measurement of segregation

Jeffrey S. ZAX *

ABSTRACT. – Conventional segregation indices indicate complete integration when the proportion of minorities in a metropolitan area is replicated in any subarea. However, this would not occur, even absent discrimination, because majorities and minorities have different distributions of other residence-determining characteristics. Segregation indices derived from predictions of commuting times for blacks, based on white commuting behavior, have a better theoretical foundation, are largely independent of conventional indices, and imply very different segregation rankings for metropolitan areas.

La théorie de la localisation géographique et la mesure de la ségrégation

RÉSUMÉ. – Les mesures traditionnelles de la ségrégation du logement définissent l'intégration comme se produisant lorsque la proportion de minorités ethniques dans une agglomération est répliquée dans n'importe quelle sous-zone. Ce papier propose des indices de ségregation nouveaux déduits des prédictions des temps de transport domicile-travail des personnes de couleur, sous l'hypothèse qu'ils devraient être identiques aux temps de transport domicile-travail des blancs qui ont les mêmes caractéristiques. Nous montrons que ces nouveaux indices entraînent un classement des agglomerations américaines en termes de ségregation très différents de celui effectué avec les indices standards.

* J. S. ZAX: University of Colorado at Boulder, Department of Economics, 256 UCB, Boulder, CO 80309-0256, USA.
Telephone: 303-492-8268, FAX: 303-492-8960 e-mail: zax@colorado.edu

A substantial industry is devoted to measuring and comparing segregation levels across American metropolitan areas. However, it relies upon measures of segregation that define "integration" as occurring when the proportions of minority and majority residents in any part of a metropolitan area are equal to their proportions in the metropolitan area as a whole.

This is an extreme assumption. The distributions of many socio-economic and demographic characteristics differ between majority and minority populations. Many of these characteristics may help determine the optimal choice of residential location. Therefore, the geographic distributions of these choices would typically differ between majority and minority populations, even if the latter were not subject to discrimination in housing markets.[1]

This paper proposes alternative segregation indices. These indices are based on predictions of the behaviors of individual minority workers, under the assumption that they are identical to those of otherwise similar majority workers. Differences between these predictions and the actual behavior of minority workers are, arguably, indications of the costs imposed by segregation.

In principle, this methodology is applicable to activity in any market where segregation might affect minority behavior. The housing market is a natural candidate, as the immediate locus of housing discrimination. Labor market behavior might also be affected, if restrictions on residential locations alter employment opportunities. Lastly, segregation is likely to affect commuting patterns, the connection between housing and labor markets.

As the initial implementation of this approach, this paper focuses on the comparison of white and black commuting patterns. This comparison is of interest, itself, and has rarely been studied. Moreover, it provides a convenient and tractable opportunity to experiment with the conceptual approach proposed here. Its results constitute a substantial challenge to conventional approaches.

Section 1 of this paper briefly describes conventional measures of segregation. Section 2 discusses the effects of segregation in the context of the theory of urban residential location choices. Section 3 introduces the data employed here. Section 4 summarizes regressions specific to each metropolitan area predicting the commuting times of white and black workers. Section 5 analyzes the implications of these regressions for differences between black and white commuting patterns. Section 6 derives metropolitan-level segregation indices from the differences in these patterns. Section 7 concludes.

1. BECKER ([1971], 78, footnote 4) makes this same point.

1 Conventional measures of segregation

The measurement of residential segregation has a long and multi-disciplinary history. TAEUBER and TAEUBER [1965] is an early contribution that measures segregation for many American cities in 1940, 1950 and 1960. Their work relies upon the index of dissimilarity:

$$D = \frac{1}{2} \sum_{i=1}^{n} \left| \frac{B_i}{B} - \frac{W_i}{W} \right|,$$

where i indexes different subareas within a metropolitan area, B represents the minority population in the metropolitan area, B_i the minority population in subarea i, W the majority population in the metropolitan area and W_i the majority population in subarea i.

The interpretation that TAEUBER and TAEUBER [1965, 29] offer for this index reveals a crucial theme that underlies almost all work in this area:

> "*Our segregation index is an index of dissimilarity, and its underlying rationale as a measure of residential segregation is simple: Suppose that whether a person was Negro or white made no difference in his choice of residence, and that his race was not related to any other factors affecting residential location (for instance, income level). Then no neighborhood would be all-Negro or all-white, but rather each race would be represented in each neighborhood in approximately the same proportion as in the city as a whole. … For this situation, the segregation index assumes a value of zero, indicating no residential segregation whatsoever.*"

The assumption that "race was not related to any other factors affecting residential location" is essential to the interpretation of this index. With it, the dissimilarity index has a natural and universal reference point at zero, indicating the complete absence of residential segregation.

Without this assumption, the index's value in the absence of segregation is unknown. Moreover, this value could vary across metropolitan areas, as the relationship between race and other determinants of residential location varies. Without a well-defined point of reference, it is impossible to determine whether observed values for the dissimilarity index indicate the presence of segregation, or whether a metropolitan area with a higher value for the index is actually more segregated, in any economic sense, than a metropolitan area with a lower value.

Nevertheless, the index of dissimilarity has served as the basis for many other studies, either solely (DENTON and MASSEY [1988], VALEY, ROOF and WILCOX [1977]), or jointly with some other measure (GOLDSTEIN and WHITE, [1985], MASSEY [1981], MASSEY and DENTON [1987], CUTLER, GLAESER and VIGDOR [1999]). At the same time, the literature has hosted a spirited discussion of the relative merits of this and many other indices (FARLEY [1984], JAKUBS [1977], SCHNARE [1980], TAEUBER and TAEUBER [1965], appendix A; WHITE [1986], ZELDER [1970]). Some contributions disregard the dissimilarity index entirely in favor of others (MASSEY and MULLAN [1984], MCKINNEY and SCHNARE [1989]).

MASSEY and DENTON [1988] attempt to consolidate this literature by providing a taxonomy for 20 different indices. They assert that these indices can be divided into five broad categories, “evenness”, “exposure”, “concentration”, “centralization” and “clustering”. However, despite the conceptual distinctions between these dimensions of segregation, all 20 measures share the same essential assumption; Integration occurs when the proportion of minorities in any sub-area is the same as their proportion in the metropolitan area as a whole.[2]

This reference point may perhaps be justified as an implication of a long-run, general equilibrium model in which discrimination and racial or ethnic distinctions of any sort disappear. If minorities enjoy the same educational and employment opportunities, and if social distinctions between groups vanish, then otherwise similar minority and majority workers would have similar jobs and incomes, similar consumption of housing services, and similar concerns regarding geographical proximity between their residence and non-work locations to which they require frequent access. Moreover, it would be reasonable to expect that the distributions of economic and social attributes within groups would be similar. In this case, there would be no reason to expect the geographic distributions of residence locations to vary across groups.

However, this justification suggests that segregation indices upon which it is based may be of limited practical value. This equilibrium does not appear to be on the immediate horizon. Moreover, available housing policy options are unlikely to effectively address any of the distortions in other markets that must be removed in order to achieve it. To the extent that segregation indices are valuable as a guide for housing policies, measures that indicate when improvements occur, absent any change in the extent of discrimination elsewhere in the economy, would be more useful.

Such measures have appeared in only a few instances. FARLEY [1975)] attempts to compare the actual distribution of black residences in the Detroit metropolitan area to that which would obtain if income were its sole determinant. More recently, DERANGO [2001, 15] defines segregation as “the difference in the average proportion of blacks in the neighborhoods of observationally equivalent black and white individuals”.

This paper makes several contributions. First, the economic theory of urban residential locational choice, rather than the assumption that integration implies equal population proportions throughout, provides the inspiration for its measures of segregation. Second, in comparison to FARLEY [1975], it imputes commuting times to black workers on the basis of many characteristics, not only income. Third, in comparison to DERANGO [2001], it presents a methodology that provides comparable segregation “scores” for all metropolitan areas.

2. This is not strictly accurate with respect to the exposure and isolation indices. However, both depend solely on the geographic distribution of minority and majority group members, and are independent of other characteristics of the members of either group. Moreover, neither embodies a natural definition of “integration”. For example, the interaction index achieves higher values when larger fractions of the minority population comprise smaller fractions of the population in the subareas in which they reside. As this is more likely when the minority population is, as a whole, less numerous, the interaction index has a tendency to assign higher values to metropolitan areas with smaller minority populations. However, metropolitan areas with this characteristic are not necessarily more “integrated”, according to any standard use of the term.

2 Segregation and the theory of residential location choice

According to the theory of urban residential location choice (MUTH [1969], for example), the choice of residential location balances two conflicting objectives. First, commutes are costly, especially in the consumption of time. All else equal, utility maximization requires that they be minimized. Second, utility increases with the consumption of housing services. Therefore, it also increases as the per-unit price of these services declines.

The conflict arises because accessibility is itself a good. Residential locations that have greater access will typically imply shorter commutes. Demand for land will be greater at these locations. Consequently, the equilibrium price of land will be greater as well. Any housing services produced at these locations must therefore be relatively expensive.

At locations that are less accessible, demand for land and the per-unit price of housing services will be lower. All else equal, housing consumption will be greater. However, the typical commute to work will be longer.

From the perspective of consumption, then, each residential location presents a bundle of goods and bads, consisting of housing services and commuting time. In general, the goods and bads are positively correlated; residential locations providing more housing services also entail longer commutes.

The optimal choice of residential location balances the marginal utility gains from additional housing services against the marginal utility losses from longer commutes. It therefore determines simultaneously optimal housing consumption and optimal commuting time. If non-market forces impose constraints on this choice, either or both will be distorted.

Segregation is the obvious example of such non-market forces. Superficially, its consequence is given by its definition; It restricts the residences of those subject to it to particular geographic areas.

However, geographic area in itself is unlikely to be an important argument in the utility function.[3] The true cost of this restriction must arise from the constraints it places on the array of bundles of housing services and commutes from which its victims can choose. If those subject to it would choose the same bundle were it absent, its presence is inconsequential to them.

If, instead, segregation prevents the purchase of the bundle which maximizes utility subject to income, then its constraints are binding. They force those subject to segregation to choose a bundle of housing services and commuting time that is second-best. The difference between the utility associated with this choice and that which would be enjoyed with the first-best bundle is the penalty imposed by segregation. The most useful measure of this penalty would be its monetized value, the difference between actual income and the income that would be necessary to achieve the utility level associated with the first-best consumption bundle under the constraint of segregation.

3. Amenities and neighborhood attributes may well vary geographically. However, the discussion here subsumes them into “housing services”.

This paper represents the first attempt to consider the measurement of segregation from this perspective. As such, its aspirations are modest. The actual measurement of the monetary costs imposed by segregation requires a more rigorous theoretical treatment, at the very least, than the discussion presented here. It may also require richer data than those described below.

This paper pursues a more tractable alternative. The costs of segregation are attributable to distortions in the choices of commuting times, housing consumption and labor market activity. The following sections address the first of these distortions.[4] In the absence of segregation, otherwise similar white and black households would make similar choices of commuting times and bundles of housing services. Differences in these choices indicate that households of one type, presumably those of blacks, are operating under a constraint that does not affect households of the other type (ZAX [1991]).

Therefore, systematic differences in the commuting times of otherwise similar white and black workers should indicate the presence of segregation. The frequency and extent of these differences should be an indication of segregation's severity.

These assertions are based on the assumption that black and white utility functions are similar. Discussions of black self-segregation (CLARK [1986], [1988], [1989], [1991], GALSTER [1982], [1988], [1989], LIEBERSON and CARTER, [1982], ROSS [2002], SERMONS [2000]) address a contrary assumption: that blacks have stronger preferences for neighborhoods with higher concentrations of black residents than do whites. If this is true, then the measures of commuting disparity presented below may reflect the extent to which black workers are prepared to alter their commuting times in order to "consume" more black neighbors, as well as any consequences of discrimination in residential markets. Of course, if this is the case, interpretations of traditional measures of segregation must be similarly modified.

These assertions are also based on the assumption that the constraints of residential segregation largely bind black but not white households. If, in an unsegregated equilibrium, a white household would prefer the inner-city locations and restricted older housing stock typically available in black residential neighborhoods, then any barriers to exercising this preference would impose welfare losses on this household. The assumption here is that white households such as this are rare: The welfare losses that they suffer in any metropolitan area are small relative to those imposed on black households who, in an unsegregated equilibrium, would prefer access to the suburban locations and newer housing stock available in many white residential neighborhoods.[5]

The following sections estimate statistical models of white and black commuting patterns for each of 228 metropolitan areas. They then employ these patterns to compare the actual distributions of black commuting times

4. KAIN and QUIGLEY [1975] and COLLINS and MARGO [2001] are examples of analyses that investigate the consequences of segregation for black housing consumption. A large literature, including ZAX and KAIN [1996] and HOLZER [1991], investigate those for black labor market activity. However, none attempts to translate measured distortions into segregation indices.

5. It is theoretically possible that white households also suffer substantial welfare losses as a consequence of the constraints of segregation. However, this possibility has not provoked any known empirical investigation. This suggests an *a priori* consensus that its practical importance is negligible.

within each area to the distributions that would obtain if black commuting times were equal to those of otherwise similar white workers. The differences between the actual and predicted black patterns give rise to a number of alternative segregation measures for each metropolitan area.

The segregation measures below differ from those of the preceding literature in two important aspects. First, as described in the previous section, conventional segregation measures simply quantify differences in geographic distribution. The measures presented below examine behavioral differences between otherwise similar blacks and whites that should be at least partially reflective of differences in utility. Second, the comparisons here between "otherwise similar" workers contrast to the conventional comparisons between all blacks and all whites. Those here recognize that if black and white populations have different average characteristics, their average behaviors will differ even in the absence of segregation.

3 Data

The analyses in sections 4 and 5 are based on the A sample of the Public Use Microdata Samples (PUMS) from the 1990 Census of Population and Housing. Sample A of the PUMS consists of individual records for 5% of the U.S. population. These records contain all of the information available to the Census for each of these individuals, with the exception of those items that would compromise confidentiality.

In particular, the PUMS reports commuting time for all those at work in the week prior to the Census. As described in the previous section, the comparison between commuting times for white and black workers is the basis for this paper. This comparison begins with regression analyses, described in the next section, in which individual commuting time is the dependent variable.

Table 1 reports the additional individual-specific items, derived from the PUMS data, that serve as explanatory variables in the regressions described in section 4. These variables describe labor market activity, human capital, household structure and immigrant status. Those in the first two categories control for the extent of engagement in the labor market and the implicit value of time.[6] Those describing household structure control for determinants, apart from income, of the demand for housing, and the possibility that residential location determines commutes of other household members. Variables measuring immigrant status and mobility control for the possibility that resi-

6. While the specification of this equation is not the primary interest of this research, a brief discussion will illuminate the roles of some of the variables. "Prior military service" may affect returns to labor market activity, and therefore the implicit price of time. The variable "currently married" describes the marital status of the observed individual. The variable "Married-couple family household" describes the structure of the household in which that individual resides. They can differ for individuals who are not themselves household heads or spouses of household heads. In the long run, residential location, workplace location and labor supply may all be simultaneously determined. However, typical variations in commuting time imply very small variations in the return to labor. Therefore, it seems unlikely that the inclusion of "usual hours worked per week" among the explanatory variables engenders a substantive endogeneity bias.

TABLE 1
Explanatory variables for individual commuting times

Labor market characteristics	*Language status*
1989 Total individual earnings	Household language other than english
1989 Usual hours worked per week	Speaks english very well
1989 Received self-employment income	Speaks english well
Number of automobiles in household	Speaks english not well
Prior military service	Speaks english not at all
Educational characteristic	*Disability status*
Currently enrolled in school	Work limitation
Some high school, no degree	Mobility limitation
High school degree	Personal care limitation
Some college, no degree	
College degree	*Household structure*
Post-college degree	1989 Income of other family members
	Married-couple family household
Demographic characteristics	Other family household
Age	One other worker in household
Household head	At least two other workers in household
Spouse of household head	Presence of persons older than 64 in household
Child or grandchild of household head	Presence of subfamilies in household
Currently married	Number of related children in household
Black	
Female	
Migration and mobility	
Different house on 4/1/85	
Non-citizen	
Immigrant, entry 1987 to 1990	
Immigrant, entry 1985 to 1986	
Immigrant, entry 1980 to 1984	
Immigrant, entry 1970 to 1979	
Immigrant, entry before 1970	

dential location might depend upon familiarity with the area of residence and preferences for the characteristics of neighbors.

At the individual level these data are subject to one important limitation. The geographical information regarding residence and workplace locations is restricted to a single measure of the distance between them, commuting time. Therefore, the only comparison that is possible between black and white workers is in the duration of their commutes. Thus, the analysis below must construe black workers whose commuting time is identical to that of an otherwise similar white worker as unconstrained by segregation.

The possibility remains, however, that the black worker in this example might have a different residence location, workplace location, or both. If so, it is possible that these differences arise from restrictions imposed by segregation on the locational choices of black workers. In this case, the costs of segregation would take the form of suboptimal employment conditions or consumption of housing and neighborhood services.

These costs would not be recognized by the methodology employed below. A complete accounting would require a simultaneous comparison of commuting time, employment conditions, and housing consumption. While perhaps feasible, this accounting exceeds the ambitions of this paper. Therefore, the segregation indices presented in section 6 probably understate the true costs of segregation.

The PUMS data embody a second restriction, at the level of metropolitan areas, or MSAs.[7] The basic unit of geography in the PUMS is the Public Use Microdata Area, or PUMA. PUMAs must have no fewer than 100,000 residents (Census of Population and Housing [1992], page 1-1). The seventy-eight smallest metropolitan areas do not meet this threshold. Consequently, the PUMS places them entirely in PUMAs that include additional non-metropolitan populations.

It is not possible to distinguish between workers in these PUMAs who are and are not residents of the included MSA. Accordingly, it is not possible to restrict analyses to the commuting times of only MSA residents. For this reason, section 4 omits these MSAs.[8]

The workers of interest here are those who worked elsewhere than in their home in the week prior to the administration of the Census questionnaire. The indices presented in section 6 below appear to be unreliable when based on small samples of black individuals of this type. Accordingly, section 4 also omits the remaining MSAs in which the PUMS contain fewer than 20 black workers with workplaces other than their residences, of which there are 25.[9]

Therefore, the analysis in section 6 compares the alternative segregation measures derived in section 5 to conventional measures reported by the U.S.

---

7. The metropolitan areas identified by the PUMS include both MSAs (metropolitan statistical areas) and PMSAs (primary metropolitan statistical areas). The latter are defined as the central metropolitan areas in CMSAs (consolidated metropolitan areas), which consist of at least one PMSA (Census of Population and Housing (1992), page A-9). Throughout this paper, "MSA" refers to both MSAs and PMSAs.

8. Albany GA, Athens GA, Bangor ME, Bismarck ND, Burlington VT, Casper WY, Charleston WV, Charlottesville VA, Cheyenne WY, Columbus GA-AL, Cumberland MD-WV, Dothan AL, Dubuque IA, Elmira NY, Enid OK, Evansville-Henderson IN-KY, Fargo-Moorhead ND-MN, Fitchburg-Leominster MA, Fort Smith AR-OK, Fort Walton Beach FL, Gadsden AL, Glens Falls NY, Grand Forks ND-MN, Great Falls MT, Huntsville AL, Iowa City IA, Jackson TN, Kankakee IL, Kokomo IN, La Crosse WI-MN, Lake Charles LA, Laredo TX, Lawrence KS, Lawton OK, Lewiston-Auburn ME, Lynchburg VA, Naples FL, Owensboro KY, Panama City FL, Parkersburg-Marietta WV-OH, Pine Bluff AR, Pittsfield MA, Portland ME, Portsmouth-Rochester NH-ME, Rapid City, SD, St. Joseph MO, San Angelo TX, Sherman-Denison TX, Sioux City IA-NE, Sioux Falls SD, Steubenville-Weirton OH-WV, Tallahassee FL, Texarkana TX, Texarkana AR, Topeka KS, Victoria TX and Wheeling WV-OH are identified in the PUMS but combined with non-metropolitan populations, and omitted from this analysis. Auburn-Opelika, AL, Barnstable-Yarmouth MA, Corvallis OR, Dover DE, Dutchess County NY, Flagstaff AZ-UT, Goldsboro NC, Grand Junction CO, Greenville NC, Hattiesburg MS, Jonesboro, AR, Missoula MT, Myrtle Beach SC, Newburgh NY-PA, Orange County,CA, Pocatello ID, Punta Gorda FL, Rocky Mount NC, San Luis Obispo-Atascadero-Paso Robles CA, Sumter SC, Ventura CA and Yolo CA are not explicitly identified in the PUMS and omitted.

9. These MSAs are Altoona, PA, Appleton-Oshkosh-Neenah WI, Bellingham WA, Billings MT, Boise City ID, Brownsville-Harlingen-San Benito TX, Duluth-Superior MN-WI, Eau Claire WI, Greeley CO, Green Bay WI, Las Cruces NM, Manchester NH, McAllen-Edinburg-Mission TX, Medford-Ashland OR, Provo-Orem UT, Redding CA, Richland-Kennewick-Pasco WA, Rochester MN, St. Cloud MN, Salem OR, Santa Fe NM, Sheboygan WI, Wausau WI, Williamsport PA and Yakima WA.

Census Bureau, (http://landview.census.gov/hhes/www/housing/resseg/ftp.htm) for 228 MSAs. In 1990, these MSAs contained 180,789,796 inhabitants. The sample examined in the next section is restricted to the 2,936,699 non-Hispanic white and black individuals who work at locations other than their residence within these MSAs and appear in the PUMS.[10]

4 MSA-specific commuting patterns

Table 2 summarizes average differences in black and white commuting patterns at the MSA level. In other words, it describes the distributions of MSA-level aggregates across the 228 MSAs included in this analysis. As an example, the "average white commuting time" of 20.5 minutes is the unweighted average of the 228 MSA-specific values for the average within-MSA white commuting time.[11]

This table demonstrates, first, that the average value for average within-MSA black commuting times in 1990 exceeded that for white commuting times. Average within-MSA black commuting times were also more variable across MSAs than were those for whites. The median value for average within-MSA black commuting times exceeded that for whites, as did the inter-quartile range for this statistic.

Table 2 gives the correlation between average black and white commuting times as .799. This might suggest that black and white commuting times did not differ greatly. However, this correlation is probably dominated by differences across MSAs in size and congestion, which would affect both white and black populations. Moreover, it does not address the essential question here, which is whether similar white and black residents in the same MSA have similar commutes.

MSA-specific regressions that estimate individual commuting times as functions of the explanatory variables in table 1 pursue this question. Table 3 summarizes the sample sizes and explanatory power of these regressions. The first panel of this table describes MSA-specific regressions which pool all

10. The U.S. Census Bureau, at the referenced website, also provides segregation indices for American Indians and Alaska Natives, Asians and Pacific Islanders, and Hispanics. This paper restricts itself to the analysis of black segregation in order to focus on its novel methodology. Harrison and Weinberg [1992], the precursor to the U.S. Census Bureau tabulations, exclude Hispanic whites from the white population in their calculations, but include Hispanic blacks with non-Hispanic blacks. They explain that "there is a trivial and unimportant difference in the index values between Blacks as a whole and non-Hispanic Blacks; here we concentrate on the figures for all Blacks" (page 6). Given the consistently lower levels of segregation indicated for Hispanics using conventional indices (Denton and Massey [1988], Harrison and Weinberg, [1992], Massey and Denton [1987], Massey and Mullan [1984]), it is at least arguable that, from the perspective of housing markets, the black racial identity of Hispanic blacks is their most salient minority characteristic. The analysis below adopts the Harrison and Weinberg [1992] classification in order to maintain comparability.

11. MSAs are the fundamental unit of analysis here, because segregation is a metropolitan-level phenomena. Therefore, tabulations of MSA-level statistics weight each MSA equally.

TABLE 2
Commuting times for white and black workers

	Average	Standard deviation	10th percentile	Median	90th percentile
Average white commuting time	20.5	2.95	17.2	20.0	24.8
Average black commuting time	20.8	4.31	16.0	20.3	26.5
Difference, average black commuting time minus average white commuting time	.294	2.64	-2.67	.295	2.96
Correlation across MSAs between average black and average white commuting times: .799					

Notes: Sample consists of 228 MSAs. Statistics describe the distribution of MSA-specific average commuting times.

TABLE 3
Explanatory power for MSA-specific commuting time regressions

	Average	Standard deviation	10th percentile	Median	90th percentile
Pooled samples of white and black workers:					
Sample sizes	12880.	17891.	2321.	6096.	32519.
R^2	.0427	.0215	.0218	.0374	.0699
Adjusted R^2	.0347	.0206	.0161	.0290	.0624
Stratified samples of white workers:					
Sample sizes	11331.	15032.	1941.	5747.	28062.
R^2	.0448	.0218	.0234	.0394	.0710
Adjusted R^2	.0362	.0211	.0163	.0309	.0621
Stratified samples of black workers:					
Sample sizes	1549.	3380.	49.	438.	3577.
R^2	.197	.225	.0256	.102	.549
Adjusted R2	.0556	.142	–.0148	.0225	.183

Note: Sample consists of 228 MSAs. Statistics describe the distribution of MSA-specific sample sizes and R^2s.

white and black workers, distinguishing the latter from the former only with a dummy variable. These 228 regressions have average sample sizes of 12,880, average R^2 values of .0427 and average adjusted R^2 values of .0347.

These R^2 values indicate that most of the variation in individual commuting times is independent of the explanatory variables employed here. This suggests that predicted commuting times derived from these regressions should be treated with some caution.[12] At the same time, these R^2 values are only slightly lower than those typically achieved in regressions that predict earnings in similar samples, from similar sets of explanatory variables.[13]

The second and third panels of table 3 present the same statistics for the 228 MSA-specific regressions that are restricted to the white sample members and the 220 MSA-specific regressions that are restricted to the black sample members in each MSA, respectively. As whites constitute most of the samples in most of the MSAs, the statistics in the second panel are similar to those of the first panel.

In contrast, those of the third panel indicate that explanatory power is much higher for regressions restricted to black workers. To some extent, this is an artifact of the small black sample sizes in some MSAs. However, relatively high average values for adjusted R^2 suggest that, in general, the explanatory variables may account for a larger proportion of the variance in black than in white commuting times.

Moreover, in general, relatively low explanatory power implies that idiosyncratic components of commuting time are large, rather than that commuting time does not depend on the explanatory variables. The R^2 values for the 228 MSA-specific regressions on pooled samples correspond to F-statistics that are significant at better than 1% for all but five MSAs, at better than 5% for three of those five and at better than 10% for another of the five. The F-statistic is insignificant only for the Ocala, FL MSA.

The R^2 values for the 228 white regressions yield a similar distribution of statistical significance levels. They are significant at better than 1% for 223 MSAs, at better than 5% for three additional MSAs, and at better than 10% for the remaining two MSAs. It is particularly important that all of these regressions are statistically significant, because they serve as the basis for the predictions of black commuting times in section 6.

MSA-specific regressions for black worker are less likely to yield statistically significant explanations of their commuting times, in part because of the small sample sizes indicated in table 3. Eight MSAs contain too few observations to estimate these regressions. In 89 MSAs, the F-statistics corresponding to the R^2 values are insignificant. They are significant in the remaining 131 MSAs, at the 10% level in 18, at the 5% level in 27 and at the 1% level in 86. The relatively poorer performance of these regressions is not especially important, as their principal purpose is for comparison in this section.

To further summarize the results of these regressions, table 4 presents aggregated information about the coefficients in the 228 MSA-specific regressions on the pooled samples of white and black workers. A number of explanatory variables have largely consistent effects: significant coefficients of the same sign in most of these regressions.

Economic variables have significant effects on commuting times in many MSAs. Where significant, these effects are of largely consistent sign. For example, individual earnings are significantly associated with commuting

12. This issue reappears in section 7.

13. As discussed in section 6, the inspiration for the alternative segregation indices presented there derives in part from these earnings regressions.

TABLE 4
Summary of significant coefficients, pooled regressions

Explanatory variable	Number of significant coefficients	% of significant coefficients with postive values	Average value of significant coefficients
Intercept	228	100.%	17.1
Black	110	79.1%	1.63
1989 Total individual earnings, $1000s	127	89.8%	.0356
1989 Usual hours worked per week	159	100.%	.0547
1989 Received self-employment income	143	1.40%	– 1.99
Number of automobiles in household	139	86.3%	.445
Prior military service	111	98.2%	1.38
Currently enrolled in school	176	0.%	– 2.32
Some high school, no degree	111	13.5%	– 2.23
High school degree	107	29.9%	– 1.25
Some college, no degree	81	35.8%	– .673
College degree	76	10.5%	– 1.82
Post-college degree	176	0.%	– 4.32
Age	71	59.2%	.00514
Householder	67	86.6%	1.04
Spouse of household head	66	4.55%	– 1.87
Child or grandchild of household head	53	30.2%	– .788
Currently married	29	34.5%	– .0777
Female	180	0.%	– 2.00
Different house on 4/1/85	106	88.7%	1.23
Non-citizen	22	81.8%	6.19
Immigrant, entry 1987 to 1990	18	61.1%	4.02
Immigrant, entry 1985 to 1986	24	79.2%	14.4
Immigrant, entry 1980 to 1984	18	66.7%	3.73
Immigrant, entry 1970 to 1979	31	54.8%	1.13
Immigrant, entry before 1970	34	67.6%	.216
Household language other than english	57	15.8%	– 1.13
Speaks english very well	32	71.9%	1.20
Speaks english well	34	58.8%	1.45
Speaks english not well	33	60.6%	2.15
Speaks english not at all	19	63.2%	10.0
Work limitation	71	97.2%	1.93
Mobility limitation	72	94.4%	5.42
Personal care limitation	42	83.3%	2.23
1989 Income of other family members, $1000s	87	1.15%	– .0232
Married-couple family household	183	100.%	3.00
Other family household	163	100.%	2.58

TABLE 4 (continued)
Summary of significant coefficients, pooled regressions

Explanatory variable	Number of significant coefficients	% of significant coefficients with postive values	Average value of significant coefficients
One other worker in household	64	6.25%	– 1.21
At least two other workers in household	125	.800%	– 1.95
Presence of persons older than 64 years in household	52	28.8%	– .915
Presence of subfamilies in household	81	96.3%	2.29
Number of related children in household	77	57.1%	.0153

Notes: Significance is at the 10% level or better. The sample consists of 228 regression equations.

time in 127 out of 228 regressions. In 114 of these, its coefficient is positive. However, the average value of these 127 coefficients is quite small.

Holding constant hours worked, this indicates that commutes tend to increase slowly with wages. This result is consistent with the assumption that underlies most theoretical models of urban residential location choice; the elasticity of the demand for housing with respect to wages is greater than the elasticity of the shadow price of time with respect to wages. Here, the former apparently outweighs the latter, though slightly. In consequence, workers accept longer commutes in return for increased housing consumption as their wages increase.

Usual weekly hours worked are significantly associated with commuting time in 159 of the 228 regressions, and positively in all 159. Holding constant earnings, this appears to suggest, paradoxically, that commutes decline with wages. It seems more likely that this variable is capturing labor market attachment in most MSAs. Those who spend fewer hours at work probably have more non-work responsibilities. These responsibilities may require them to have more ready access to their residences.

Holding constant individual earnings, those who receive self-employment earnings tend to have commutes that are shorter by approximately two minutes. Those who are not self-employed must choose between the workplaces already established by potential employers. In contrast, the self-employed can choose the location at which to establish their workplace. The results here suggest that the latter take advantage of this additional flexibility to reduce their typical commuting times.

In nearly two out of five MSAs, individual commuting times decline significantly as other family members provide greater amounts of income. The most likely explanation seems to be that non-work responsibilities increase with those incomes, and therefore the labor force engagement of other family members. Consequently, the need for residential access increases as well.

Several demographic variables are also consistently influential. Female commuting times differ significantly from those of males in 180 MSAs. In all cases, they are shorter, typically by approximately two minutes. This again may reflect an underlying disparity in non-work responsibilities.

Workers currently enrolled in school have significantly different commute times from those who are not in 176 MSAs. Again, in all cases, their commutes are shorter, typically by a little more than two minutes. Presumably, this reduction improves access to the locations of educational institutions.

Perhaps surprisingly, the commutes of workers with post-college degrees regularly differ significantly from those whose education ended before high school, the omitted category. In the case of every significant difference, the former workers have shorter commutes, typically by more than four minutes. This may indicate differences in tastes, or perhaps a more restricted choice set of workplaces for those with the least education.

Lastly, household structure also significantly affects commuting time. The commuting times of individuals in married-couple households differ significantly from those in households without families, the omitted category, in 183 MSAs. Those of individuals in households containing other types of families differ significantly in 163 MSAs. In all cases of significant differences, those in households containing some sort of family have commutes that are longer, typically by a few minutes. The most likely explanation is that residence locations for these households compromise between the locational needs of multiple family members, leaving each with a longer commute than they would choose were they alone.

In sum, MSA-specific regressions for commuting times yield results that are largely plausible. At the same time, there is considerable heterogeneity across MSAs regarding individual results. This suggests that MSA-specific commuting patterns, as well as individual commuting times, embody substantial idiosyncratic components.

In other words, commuting patterns are partially determined by characteristics that are unique to each MSA, whether geographical, historical or administrative. These characteristics may include race and patterns of residential segregation. For example, automobile ownership significantly affects commute times in 139 MSAs, and positively in most. However, the regressions stratified by race demonstrate that the effects of this variable usually differ for black and white workers.

Automobile ownership significantly affects the commutes of white workers in 150 MSAs. It increases them in 138. This suggests that the flexibility of automobile commutes allows white workers to consume more distant and, presumably, less expensive housing.

Automobile ownership significantly affects the commutes of black workers in 85 MSAs, but increases them in only 14. In the 66 MSAs where automobile ownership significantly affects the commutes of both, it increases white commutes and decreases black commutes in 47. This suggests, perhaps, that restrictions on black residential choice prevent black workers from using commuting flexibility to improve their housing consumption. The next section explores the contrasts between commuting patterns of white and black workers more extensively.

5 Racial differences in commuting patterns

This section examines the implications of the regressions presented in section 4 regarding differences in the commuting patterns of black and white workers. Table 5 begins this examination by summarizing the characteristics for the coefficients on the black dummy variable in the 228 MSA-specific regressions pooled over white and black workers.

This table demonstrates that this dummy variable was statistically significant at 1% or better in 77 MSAs, at between 1% and 5% in 20 MSAs, and between 5% and 10% in 13 MSAs. Substantial numbers of both positive and negative coefficients appear in each significance category, although positive values become somewhat more prominent among those achieving higher levels of significance.[14] This heterogeneity suggests, in keeping with a theme recognized previously, that differences between black and white commuting times depend importantly on factors unique to each MSA.

In addition, table 5 indicates that the coefficient on the black dummy variable is not statistically significant in 118 of the 228 MSAs. However, this does not necessarily imply that the determinants of commuting times for black and white workers are similar in these MSAs. Rather, the restriction that these determinants differ only by a constant may be a misspecification.

TABLE 5
Characteristics of coefficients on the black dummy variable

	Number	Average	Standard deviation	10th percentile	Median	90th percentile
All coefficients	228	.746	2.39	– 1.93	.811	3.43
Coefficients not significant at 10%	118	– .0782	1.32	– 1.60	.109	1.26
Coefficients significant at 10% but not at 5%	13	– .453	2.50	– 4.03	.425	1.51
Coefficients significant at 5% but not at 1%	20	1.04	2.26	– 2.18	1.07	3.64
Coefficients significant at 1% or better	77	2.13	2.98	– 2.22	2.00	4.97

14. GABRIEL and ROSENTHAL [1996] estimate significantly longer commutes for blacks in a sample drawn from the American Housing Survey. ZAX [1990], using a sample of workers from a single establishment, estimates that blacks have shorter commutes than otherwise similar whites.

TABLE 6
Significance of black-white differences in commuting times

Significance of stratified model	Significance of coefficient for black dummy variable, pooled model: .1 < P-value	.1 ⩾ P-value > .05	.05 ⩾ P-value > .01	.01 ⩾ P-value	Total
No black regression, insufficient observations	7 (3.1%)	0	1 (.4%)	0	8 (3.51%)
.1 < P-value	74 (32.5%)	9 (4.0%)	7 (3.1%)	13 (5.7%)	103 (45.2%)
.1 ⩾ P-value > .05	6 (2.6%)	0	0	6 (2.6%)	12 (5.3%)
.05 ⩾ P-value > .01	13 (5.7%)	1 (.4%)	2 (.9%)	8 (3.5%)	24 (10.5%)
.01 ⩾ P-value	18 (7.9%)	3 (1.3%)	10 (4.4%)	50 (21.9%)	81 (35.5%)
Total	118 (51.8%)	13 (5.7%)	20 (8.8%)	77 (33.8%)	228

Note: Parentheses contain percentages of the entire sample.

Table 6 addresses this possibility. Stratified models, consisting of separate regressions for white and black workers in each MSA, allow the determinants of black and white commuting times to differ through different coefficients on each of the explanatory variables, rather than merely by a constant. The right-most column of this table presents the significance levels attained by these stratified models, compared to their associated pooled regressions. The bottom row of this table reproduces the count of coefficients on the black dummy variable by significance category from the first column of table 5. The body of this table interacts the significance levels from the two different estimations.

The right-most column reports that the stratified regressions are statistically superior to the pooled regressions in 81 MSAs with better than 1% significance, in 24 MSAs with between 5% and 1% significance, and in 12 MSAs with between 10% and 5% significance. These regressions are not statistically superior in 103 MSAs. They are not estimable, due to insufficient black observations, in 8 MSAs.

More importantly, the first column demonstrates that the stratified model is statistically superior to the pooled model in 37 MSAs in which the latter does not yield a significant coefficient on the black dummy variable. In other words, in these MSAs the pooled model suggests that there is no significant difference in the determinants of black and white commuting times, but the stratified model demonstrates that there is.

This table also demonstrates that, in one MSA where the stratified model is not estimable, the pooled model yields a statistically significant coefficient for the black dummy variable. This is also true for 29 MSAs in which the stratified model is not statistically superior to the pooled model. Lastly, in 80 MSAs both the pooled model and the comparison between it and the stratified model indicate that the determinants of black and white commuting times are statistically distinct.

This implies that the MSA-specific regressions for commuting time reveal no statistically significant differences between black and white workers in 81 MSAs. In all of these MSAs, the coefficient on the black dummy variable is not statistically significant. In seven, the stratified model is inestimable. In the remaining 74, this model is estimable but is not statistically distinguishable from the pooled model.

As a first approximation, from the perspective of this paper, these MSAs are "unsegregated". Differences in their black and white commuting patterns are not attributable to race. The remaining 154 are "segregated". Table 7 presents average characteristics of each subgroup.

Segregated MSAs tend to be large. The average segregated MSA had approximately 3.5 times as much population as did the average unsegregated MSA in 1990. Segregated MSAs are also more heavily black. Blacks constituted more than 13% of the population of the average segregated MSA, but less than eight percent of the population of the average unsegregated MSA.

TABLE 7
Average populations of segregated and unsegregated MSAs

	Unsegregated MSAs	Segregated MSAs
1990 population	301,346. (180,357.)	1,063,815. (1,382,506.)
1990 black population	22,712. (25,112.)	154,373. (280,950.)
1990 black % of population	7.63% (7.02%)	13.1% (9.75%)

Note: This table identifies an MSA as exhibiting significant segregation if either the coefficient for the dummy variable for black workers is statistically significant in the pooled regression, or if the regression stratified by race is statistically superior to the pooled regression.

Table 8 describes the geographical distribution of segregated and unsegregated MSAs, by this definition. Segregated MSAs appear to be most common in New England (Connecticut, Massachusetts, Maine, New Hampshire, Rhode Island and Vermont).[15] They are also abundant in the South Atlantic (Delaware, District of Columbia, Georgia, Florida, Maryland, North Carolina, South Carolina, Virginia and West Virginia) and Middle Atlantic (New York, New Jersey and Pennsylvania). In contrast, only one-third of the MSAs in the Mountain region (Arizona, Colorado, Idaho, Montana, New Mexico) are segregated.

15. These rankings should be treated with caution. In general, sample sizes are small. In particular, New England appears to have a large share of the MSAs that are not identified in the PUMS, and consequently are absent from the sample studied here.

TABLE 8
Distributions of segregation within Census regions

Region	No significant segregation	Significant segregation	Total number of MSAs	Percent of all MSAs
New England	2 (12.5%)	14 (87.5%)	16	7.0%
Middle Atlantic	8 (27.6%)	21 (72.4%)	29	12.7%
East North Central	17 (41.5%)	24 (58.5%)	41	19.0%
West North Central	6 (46.2%)	7 (53.9%)	13	5.7%
South Atlantic	11 (27.5%)	29 (72.5%)	40	17.5%
East South Central	8 (47.1%)	9 (52.9%)	17	7.5%
West South Central	10 (32.3%)	21 (67.7%)	31	13.6%
Mountain	8 (66.7%)	4 (33.3%)	12	5.3%
Pacific	11 (37.9%)	18 (62.1%)	29	12.7%
Total	81 (35.5%)	147 (64.5%)	228	100.%

Note: This table identifies an MSA as exhibiting significant segregation if either the coefficient for the dummy variable for black workers is statistically significant in the pooled regression, or if the regression stratified by race is statistically superior to the pooled regression.

6 Alternative measures of segregation

The dichotomization of MSAs into "unsegregated" and "segregated" is relatively coarse. This section uses the regressions of section 4 to develop much finer distinctions between MSAs. The methodology is analogous to the Oaxaca-Blinder analysis of sex and racial wage differentials (BLINDER [1973], OAXACA [1973], see CAIN [1986], for a consolidated exposition). The regression equation for the commuting times of white workers in an MSA predicts the commuting times for black workers, if they chose these times as did whites. The difference between these predicted times and actual black commuting times indicates the consequences for black workers of differences between their residential location options and those of otherwise similar white workers in the same MSA.[16]

16. BAYER, MCMILLAN and RUEBEN [2002] use a similar strategy of counterfactuals to explore the extent to which segregation in the San Francisco metropolitan area, conventionally measured, is attributable to characteristics other than race.

Table 9 provides summary statistics of these actual and predicted times. The first line of the first panel replicates the second line of table 2, and reiterates that the average of the MSA-specific average black commuting times across MSAs was 20.8 minutes. The first line of the second panel reports that, in comparison, the average predicted black commuting time according to the regression equation for white workers in the same MSA was 19.8 minutes. The average median actual and predicted black commuting times were 17.0 and 19.9 minutes, respectively.

The principal effect of these predictions is to reduce the skew in average black commuting times. The distributions of actual average and especially median black commuting times include a number of MSAs with very high values. Those of predicted average and median black commuting times are much more symmetrical.

In contrast, the effect of these predictions on typical black commuting times is ambiguous. The difference between the across-MSA average of the within-MSA median actual and predicted times is of opposite sign from the difference between the across-MSA average of the within-MSA average actual and predicted times.

According to the theory of urban residential location choice, deviations in either direction can be interpreted as distortions from the first-best optimum. If, as in the case of the within-MSA median measure, black workers have shorter commutes than otherwise similar white workers, they benefit from reduced commuting costs. However, this comes at the expense of reduced housing consumption. If the two worker types have similar utility functions, the revealed preferences of the white workers demonstrate that the utility gains from optimal consumption of housing services exceed the associated utility losses from longer commutes.

The case, as in the comparison of within-MSA averages, in which black workers have longer commutes than do otherwise similar white workers is analogous. Here, black workers suffer the additional costs of longer commutes, presumably compensated to some extent by the benefit of

TABLE 9

Characteristics of actual and predicted black commuting times

	Average	Standard deviation	10th percentile	Median	90th percentile
Average black commuting time within MSA	20.8	4.31	16.0	20.3	26.5
Median black commuting time within MSA	17.0	3.86	15.0	15.0	20.0
Average predicted black commuting time within MSA	19.9	2.85	16.6	19.7	23.3
Median predicted black commuting time within MSA	19.9	2.84	16.7	19.6	23.4

Notes: Sample consists of 228 MSAs. Statistics describe the distributions of MSA-specific average and median commuting times.

consuming housing services at a lower per-unit price.[17] Again, the revealed preferences of otherwise similar white workers demonstrate that this compensation is incomplete. The constraint of segregation again forces black workers to a suboptimal second-best.

The differences between the comparisons of within-MSA averages and medians suggest that outliers in the distributions of both actual and predicted times may be important. More broadly, they suggest that segregation, as measured within the dimension of commuting times, may have multiple dimensions. With this perspective, table 10 presents twelve MSA-level summary statistics derived from the underlying data: the joint distribution of actual and predicted black commuting times within each MSA.

TABLE 10
Characteristics of alternative MSA-specific segregation measures

	Average	Standard deviation	10th percentile	Median	90th percentile
% Black workers with predicted commuting times greater than actual commuting times	60.2	7.48	50.2	60.3	69.0
Average difference: predicted minus actual black commuting times	– .859	2.43	– 3.80	– .818	1.83
Average ratio: predicted/actual black commuting times	1.60	.241	1.36	1.55	1.89
Average annual value of difference between black predicted and actual commuting times	– 70.8	253.	– 295.	– 48.8	143.
Median difference: predicted minus actual black commuting times	2.53	2.09	.121	2.55	4.89
Median ratio: predicted/actual black commuting times	1.17	.149	1.01	1.16	1.35
Median annual value of difference between black predicted and actual commuting times	107.	101.	6.44	103.	225.
Correlation between actual and predicted black commuting times	.0691	.0740	– .000589	.0593	.154
Correlation between actual black commuting times and differences	– .833	.0583	– .891	– .851	– .748
Correlation between actual black commuting times and ratios	– .866	.0412	– .905	– .878	– .814
Correlation between actual black commuting times and annual value of differences	– .698	.0668	– .765	– .714	– .594
Commuting-time dissimilarity index	.799	.0390	.760	.796	.849

Note: Statistics describe the distribution of MSA-specific percentages, averages and medians.

17. GABRIEL and ROSENTHAL [1996] and ZAX [1991] estimate the extent of this compensation. ZAX [1991] examines the extent to which earnings also compensate for commuting costs.

The first statistic, the proportion of black workers with predicted commuting times that are greater than their actual commuting times, again dichotomizes the underlying information. It assigns equal weights to all those with predicted in excess of actual commutes, regardless of the magnitude of this difference. As given in table 10, approximately 60% of the black workers in a typical MSA would experience longer commutes were their commutes determined as are those of otherwise similar white workers. In almost all MSAs, a majority of black workers would experience longer commutes under these circumstances.

The second statistic, the within-MSA average difference, is analogous to the first but weights each black worker by the magnitude of the difference between their predicted and actual commutes. In the typical MSA, this statistic has values of approximately –.8. It is negative for the majority of MSAs, even thought the previous statistic demonstrates that the majority of commuters in most MSAs have longer predicted than actual commutes. Together, these observations imply that, among what is typically the minority of workers experiencing predicted reductions in commuting times, the magnitudes of those reductions would be much greater than the predicted increases for the majority of workers.

The third statistic, the average ratio, is based on the transformation of the joint distribution of predicted and actual black commute times into a single distribution describing the ratios of the two for commuters in an MSA. It consists of the average within-MSA value for this relative comparison. This statistic is greater than one for all MSAs. This last observation implies that, among what is typically the majority of workers experiencing predicted increases in commuting times, the relative magnitudes of those increases would be much greater than the relative reductions predicted for the minority of workers.

The implications of the second and third statistics are readily reconciled. Both appear to be artifacts of a pattern in which black workers with especially short actual commutes have predicted commutes that are somewhat longer in absolute terms but dramatically longer in relative terms. Similarly, black workers with especially long commutes have predicted commutes that are substantially shorter in absolute terms but less so in relative terms. Accordingly, the latter dominate the calculation of MSA-specific average differences between predicted and actual commutes. The former dominate the calculation of MSA-specific ratios of the two.

This pattern of differences between predicted and actual commuting distances is, in turn, consistent with received images of segregation in American MSAs. The residences of most of the workers in both groups of black commuters are, presumably, confined to central-city black neighborhoods. If segregation is a constraint, its release would allow many to move to more suburban residences.

Black commuters with central-city workplaces would thereby increase their previously short commutes. However, this increase would be limited by the need to remain accessible to the central city. Those suffering from the spatial mismatch of suburban workplaces would thereby reduce their previously long commutes. In this case, reductions would be encouraged by both increased savings in commuting times and increased access to suburban housing. The contrasts among the first three statistics of table 10 would be a natural consequence.

The fourth statistic represents an attempt to monetize the second statistic, the difference between predicted and actual black commuting times. Here, this difference is valued at the hourly wage rate implied by each black worker's reported earnings and hours of work in 1989. Twice this amount represents the total daily cost of the difference in actual and predicted commuting times. This daily estimate, multiplied by 250, an estimate of the number of working days in a year, estimates the annual monetary cost.

The average value of this annual cost across all black workers within an MSA constitutes the fourth statistic. The distribution of this statistic across MSAs has central values that are relatively small. However, MSAs at the extremes of the distribution have average distortions in black commuting times with annual values of several hundred dollars.

The fifth through seventh statistics replicate the second through fourth, substituting median values of the respective within-MSA distributions for the average values. As might be expected on the basis of table 9, statistics based on within-MSA median values indicate that predicted commutes are more likely to exceed actual commutes, but by relatively less, than are those based on within-MSA average values.

The next four statistics are within-MSA correlations between the actual commuting times for black workers and different functions of their predicted commuting times. The first relies on the untransformed predicted commuting times. If black commuting times were identical, or at least proportional, to those of otherwise similar white workers, this index would attain a maximum value of one. If they were unrelated, its value would be zero. If black workers had shorter commutes whenever otherwise similar white workers had longer commutes, its value would be negative.

In the event, this index demonstrates that actual and predicted black commuting times typically have a weak positive correlation. The positive sign indicates that, on average, black workers with longer actual commutes might expect to have longer predicted commutes, as well. The low values demonstrate that, in general, this expectation is quite weak. This suggests that, across all MSAs, actual and predicted black commuting patterns differ substantially.

The next three correlations in table 10 parallel the second, third and fourth, and again the fifth, sixth and seventh statistics. Each consists of the within-MSA correlation between actual black commuting times and the respective transformation of predicted black commuting times: the difference between predicted and actual times, the ratio of predicted to actual times, and the monetized value of the absolute difference.

Each of these three statistics is typically negative. These values indicate that, to varying extents in each MSA, black workers with long actual commutes are likely to have predicted commutes that are shorter. In contrast, those with relatively short actual commutes are likely to have predicted commutes that are longer. This confirms the presumption upon which the discussion of the first three statistics is based.

The final statistic of table 10 returns to the tradition of segregation indices established in previous literature. The commuting-time dissimilarity index is based on the conventional formula for the dissimilarity index given in section 1. However, instead of comparing the relative proportions of blacks and whites in each geographic area, it compares the relative proportions of blacks with predicted commutes of each duration to the relative proportions of blacks with actual commutes of the same duration.

In other words, *i* in the summation of the formula given in section 1 indexes commuting time, in minutes. Each term in that summation refers to commutes of the designated duration. The first ratio in each term is the fraction of black commuters with predicted commutes of this duration. The second ratio is the fraction of black commuters with actual commutes of this duration.

As with the conventional dissimilarity index, this statistic attains a minimum value of zero if the distributions of predicted and actual black commutes are identical. It attains a maximum value of one if these distributions do not overlap: if predicted and actual commutes have no durations in common.[18]

This commuting-time dissimilarity index indicates that these two distributions are largely disjoint for all MSAs. The average value of .799 is quite high. In comparison, the conventional, geographically-based dissimilarity index for these MSAs has an average value of only .582.

Moreover, the differences between these two distributions are surprisingly consistent across MSAs. The standard deviation for the commuting-time dissimilarity index is only .0390, while it is .132 for the conventional dissimilarity index.

The statistics of table 10 are all different transformations of the same joint distributions. Consequently, they are often correlated with each other. However, 54.5% of their inter-correlations are less than .5. Fewer than 20% are greater than .8. This suggests that each of these statistics may capture some independent aspect of residential segregation.[19]

Moreover, they capture aspects of residential segregation that are not addressed by conventional indices. Among the 228 correlations between the twelve statistics of table 10 and the nineteen conventional indices tabulated by the U.S. Census Bureau, only one exceeds .4. Of them, 140, or 61.4%, are less than .2. This suggests that the alternative and conventional indices are largely orthogonal.

Table 11 confirms this suggestion. It presents the R^2 values from twelve regressions calculated across the sample of 228 MSAs. In each, the dependent variable consists of MSA-specific values for one of the statistics of table 10. The explanatory variables consist of MSA-specific values for 17 of the 19 conventional indices provided by the U.S. Census.[20] The conventional indices together explain no more than 33% of the variation for any of the twelve statistics.

Table 11 suggests that conventional segregation indices and the statistics of table 10 quantify the differences between the within-MSA residential distributions of white and black workers in dramatically different ways. Tables 12 and 13 reiterate this point by comparing conventional and alternative rankings of MSAs by degree of segregation.

---

18. An anonymous referee provided the inspiration for this statistic.

19. For comparison, of the 171 correlations among the nineteen conventional segregation indices provided by the U.S. Census, 27.5% exceed .8 and 50.9% are less than .5 among the MSAs studied here.

20. The remaining two conventional indices are linearly dependent with those included in these regressions.

Any of the twelve statistics in table 10 may serve as alternative segregation indices. However, conventional measures generally address issues of relative black and white dispersion within MSAs. These same issues are addressed most directly by the last five statistics of that table.

TABLE 11
Explanatory power of conventional segregation measures for alternative segregation measures

Alternative segregation measure	R^2	Adjusted R^2
% Black workers with predicted commuting times greater than actual commuting times	.289	.232
Average difference, predicted minus actual black commuting times	.208	.144
Average ratio, predicted/actual black commuting times	.270	.211
Average annual value of difference in black commuting times	.114	.0421
Median difference, predicted minus actual black commuting times	.251	.191
Median relative difference, predicted/ actual black commuting times	.277	.218
Median annual value of difference in black commuting times	.217	.154
Correlation between actual and predicted black commuting times	.202	.137
Correlation between actual black commuting times and differences	.329	.275
Correlation between actual black commuting times and ratios	.222	.159
Correlation between actual black commuting times and annual value of differences	.315	.259
Commuting-time dissimilarity index	.107	.0351

Note: Each row presents the R^2 values for a regression of the indicated alternative segregation measure on 17 conventional segregation measures from the U.S. Census. Each of the regressions omits the indices for isolation and distance-decay isolation because in this implementation they sum to one with, respectively, the indices for interaction and distance-decay interaction. The sample consists of 228 MSAs.

The correlation between actual and predicted black commuting times is the primary of the four correlations. The others simply reflect some transformation of the joint distribution of these two times. Therefore, this correlation is the basis for the ranking in table 12. The ranking of table 13 is based on the commuting-time dissimilarity index, because of its foundation in the original segregation index. The original dissimilarity index serves as the representative conventional measure in both tables.

The contrasts between MSA-level segregation rankings by conventional and alternative indices are startling. In table 12, agreement, to the extent that it can be said to be present, is restricted to the realm of the least segregated.

There, four of the ten least segregated MSAs by the correlation between actual and predicted commuting times are also ranked among the fifty least-segregated MSAs by the conventional dissimilarity index. All of the former ten are in the least-segregated half of the distribution by the conventional dissimilarity index. Similarly, four of the five least-segregated MSAs by the dissimilarity index are in the least-segregated half of the current sample, according to the actual-predicted correlation.

In contrast, of the ten most heavily segregated MSAs by the actual-predicted correlation, only two are among the hundred most segregated MSAs by the conventional dissimilarity index. None is ranked worse than 68. In contrast, the conventional dissimilarity index ranks four as among the 25 least-segregated MSAs.

TABLE 12

Example: MSAs ranked by the correlation between actual and predicted commuting times and by the conventional dissimilarity index

	Correlation between actual and predicted black commuting times		Dissimilarity index	
MSA name	Rank	Value	Rank	Value
Most segregated by correlation:				
Pueblo, CO	1	– .232	206	.394
New Bedford, MA	2	– .127	162	.508
Lincoln, NE	3	– .0957	195	.427
Spokane, WA	4	– .0928	204	.409
Yuba City, CA	5	– .0804	218	.352
Kenosha, WI	6	– .0795	105	.612
Santa Rosa, CA	7	– .0614	221	.341
Racine, WI	8	– .0607	86	.639
Sharon, PA	9	– .0521	68	.664
Salinas, CA	10	– .0491	122	.584
Least segregated by correlation:				
Middlesex-Somerset-Hunterdon, NJ	219	.223	143	.543
Worcester, MA-CT	220	.225	144	.540
Terre Haute, IN	221	.231	116	.597
Olympia, WA	222	.236	201	.413
Cedar Rapids, IA	223	.238	175	.478
Lafayette, IN	224	.280	208	.390
Chico-Paradise, CA	225	.282	186	.448
State College, PA	226	.299	169	.497
Bloomington-Normal, IL	227	.317	212	.368
Joplin, MO	228	.344	155	.524
Most segregated by dissimilarity index:				
Gary, IN	137	.0750	1	.899
Detroit, MI	105	.0546	2	.874
Chicago, IL	65	.0307	3	.838
Milwaukee-Waukesha, WI	12	– .0420	4	.826
Newark, NJ	119	.0622	5	.825
Least segregated by dissimilarity index:				
Fayetteville, NC	126	.0687	224	.313
Santa Cruz-Watsonville, CA	218	.222	225	.312
Nashua, NH	215	.197	226	.284
Boulder-Longmont, CO	174	.108	227	.272
Jacksonville, NC	76	.0395	228	.227

Of the five most heavily segregated MSAs by the dissimilarity index, only Milwaukee-Waukesha is ranked as highly segregated by the actual-predicted correlation. Three of the other four are not among the 100 most-segregated MSAs by this measure. Overall, the correlation between ranks according to the dissimilarity index and according to the actual-predicted correlation is only .211.

The contrast in table 13 is, if anything, starker. Again, four of the ten most-segregated MSAs by the commuting-time dissimilarity index are among the 25 least-segregated MSAs by the conventional dissimilarity index. However, in this case, three of the ten least-segregated MSAs by the commuting-time dissimilarity index are among the 100 most-segregated MSAs by the conventional dissimilarity index. Of the five most-segregated MSAs by the

TABLE 13

Example: MSAs ranked by commuting-time and conventional dissimilarity indices

	Commuting-time dissimilarity		Dissimilarity indexindex	
MSA name	Rank	Value	Rank	Value
Most segregated by commuting-time dissimilarity index:				
Santa Cruz-Watsonville, CA	1	.952	225	.312
Honolulu, HI	2	.908	196	.426
San Francisco, CA	3	.894	89	.638
Brazoria, TX	4	.879	180	.468
Pueblo, CO	5	.875	206	.394
Boulder-Longmont, CO	6	.875	227	.272
Dallas, TX	7	.874	98	.625
Chico-Paradise, CA	8	.870	186	.448
Spokane, WA	9	.869	204	.409
Miami, FL	10	.868	51	.690
Least segregated by commuting-time dissimilarity index:				
State College, PA	219	.731	169	.497
Merced, CA	220	.723	214	.364
Abilene, TX	221	.716	202	.412
Fort Collins-Loveland, CO	222	.714	217	.354
Visalia-Tulare-Porterville, CA	223	.712	168	.497
Waterbury, CT	224	.707	91	.635
Waterloo-Cedar Falls, IA	225	.692	50	.690
Fayetteville-Springdale-Rogers, AR	226	.680	72	.660
Eugene-Springfield, OR	227	.679	222	.323
Springfield, MO	228	.673	147	.537
Most segregated by dissimilarity index:				
Gary, IN	51	.824	1	.899
Detroit, MI	85	.808	2	.874
Chicago, IL	26	.841	3	.838
Milwaukee-Waukesha, WI	203	.762	4	.826
Newark, NJ	104	.802	5	.825
Least segregated by dissimilarity index:				
Fayetteville, NC	81	.810	224	.313
Santa Cruz-Watsonville, CA	1	.952	225	.312
Nashua, NH	11	.867	226	.284
Boulder-Longmont, CO	6	.875	227	.272
Jacksonville, NC	54	.823	228	.227

conventional dissimilarity index, none is ranked higher than twenty-sixth, and one is ranked among the 25 least-segregated MSAs, by the commuting-time dissimilarity index. All of the five least-segregated MSAs by the conventional dissimilarity index are among the 100 most-segregated MSAs by the commuting-time dissimilarity index. The correlation between ranks according to the two indices is only .0595.

7 Conclusion

The segregation indices suggested here are dramatically different from those employed in previous work. Their application to other minority groups in the United States is an obvious extension. Moreover, there are potential applications in other countries (Massey [1985]).

Apart from their novelty, the indices introduced here have three more important virtues. First, they have an explicit foundation in the theory of urban residential location choice. Second, they attempt to measure direct behavioral consequences of segregation. Third, the MSA-specific calculations upon which they are based are easier to implement than the block- and tract-level calculations required by conventional segregation measures.

At the same time, the regressions upon which the indices suggested here are based explain relatively little of the variation in actual commuting times. If some of the variation that remains is attributable to omitted explanatory variables, rather than to truly random components, the resulting predicted commuting times may be biased. The indices upon which they are based might consequently be misleading.

However, the conventional measures of segregation also embody implicit "predictions" of black residential locations, and therefore black commuting times, in the absence of segregation. Because these predictions are implicit, their statistical properties are unexamined. Moreover, they are based on the assumption that the distribution of black residences would replicate that of white residences in the absence of segregation, regardless of any other differences in black and white socio-economic and demographic characteristics. The extreme nature of this assumption almost surely guarantees that predictions derived from it "fit" the data less well, and are more seriously afflicted with omitted variable bias, than those based on the regressions in section 4.

Furthermore, the approach here can be refined. First, in principal it is possible to calculate confidence intervals around predicted values for black commute times. Those which contain actual black commute times could be construed as indicating differences between actual and predicted times that are statistically insignificant. Indices which disregard observations so construed would address the concern that predicted values are potentially subject to large variances.

Second, it may be possible to extend these indices to incorporate differences between characteristics of actual black housing and those of housing occupied by otherwise similar whites. It may also be possible to incorporate differences between black labor market outcomes and those of otherwise similar white workers. These elaborations would address the concern that the indices here summarize only one dimension of the possible costs of segregation.

However, the final test for any segregation index should be its relevance to policy. Despite the energy devoted to calculating conventional indices, relatively little attention has been devoted to their application either to the question of whether housing policies affect segregation, or whether segregation affects other black outcomes. The first question has apparently been addressed only by SMITH [1989] and FURST [2001]. CUTLER and GLAESER [1997], COLLINS and MARGO [2000] and ZAX [1998] are the only known investigations of the second. This literature should be greatly expanded, and should address the additional question as to which segregation indices are most informative in these contexts. ▼

• References

BAYER P., MCMILLAN R., RUEBEN K. (2002). – « What drives racial segregation? Evidence from the San Francisco Bay Area using micro-census data », working paper, Yale University.

BECKER G. S. (1971). – *The Economics of Discrimination*, Second Edition, The University of Chicago Press, Chicago.

BLINDER A. S. (1973). – « Wage discrimination: Reduced form and structural variables », *Journal of Human Resources*, Vol. 8, p. 436-455.

CAIN G. G. (1986). – « The economic analysis of labor market discrimination: A survey », chapter 13 in Ashenfelter, Orley C. and Richard Layard, *Handbook of Labor Economics, Vol. 1*, Elsevier Science Publishers, New York, p. 693-785.

CENSUS OF POPULATION AND HOUSING (1990) (1992). – *Public Use Microdata Sample U.S. Technical Documentation*, Bureau of the Census, Department of Commerce, Washington, D.C.

CLARK W.A.V. (1986). – « Residential segregation in American cities: a review and interpretation », *Population Research and Policy Review*, Vol. 5, n°2, p. 95-127.

CLARK W.A.V. (1988). – « Understanding residential segregation in American cities: interpreting the evidence », *Population Research and Policy Review*, Vol. 7, n°2, p. 113-121.

CLARK W.A.V. (1989). – « Residential segregation in American cities: Common ground and differences in interpretation », *Population Research and Policy Review*, Vol. 8, n°2, May, p. 193-197.

CLARK W.A.V. (1991). – « Residential preferences and neighborhood racial segregation: A test of the Schelling segregation model », *Demography*, Vol. 28, n°1, February, p. 1-19.

COLLINS W. J., MARGO R. A. (2001). – « Race and home ownership: A century-long view », *Explorations in Economic History*, Vol. 38, p. 68-92.

COLLINS W. J., MARGO R. A. (2000). – « Residential segregation and socioeconomic outcomes: When did ghettos go bad? », *Economics Letters*, Vol. 69, p. 239-243.

CUTLER D. M., GLAESER E. L. (1997). – « Are ghettos good or bad? », *Quarterly Journal of Economics*, Vol. 112, 827-872.

CUTLER D. M., GLAESER E. L., VIGDOR J. L. (1999). – « The rise and decline of the American ghetto », *Journal of Political Economy*, Vol. 107, n°3, June, p. 455-506.

DENTON N. A., MASSEY D. S. (1988). – « Residential segregation of blacks, Hispanics, and Asians by socioeconomic status and generation », *Social Science Quarterly*, Vol. 69, n°4, December, p. 797-817.

DERANGO K. (2001). – *Black-white segregation, discrimination, and home ownership*, Staff Working Paper 01-71, W.E. Upjohn Institute for Employment Research, Kalamzoo.

FARLEY J. E. (1984). – « P* segregation indices: What can they tell us about housing segregation in 1980? », *Urban Studies*, Vol. 21, n°2, May, p. 331-336.

FARLEY R. (1975). – « Population trends and school segregation in the Detroit metropolitan area », *Wayne Law Review*, Vol. 21, March, p. 867-902.

FURST V. (2001). – *The Impact of Fair Housing Laws on Residential Segregation: An Econometric Study*, unpublished senior Honors Thesis, University of Colorado at Boulder.

GABRIEL S. A., ROSENTHAL S. S. (1996). – « Commutes, neighborhood effects, and earnings: An analysis of racial discrimination and compensating differentials », *Journal of Urban Economics*, Vol. 40, n°1, July, p. 61-83.

GALSTER G. (1982). – « Black and white preferences for neighborhood racial composition », *Journal of American Real Estate and Urban Economics Association*, Vol. 10, n°1, Spring, p. 39-66.

GALSTER G. (1988). – « Residential segregation in American cities: A contrary view », *Population Research and Policy Review*, Vol. 7, n°2, p. 93-112.

GALSTER G. (1989). – « Residential segregation in American cities: A further response to Clark », *Population Research and Policy Review*, Vol. 8, n°2, May, p. 181-192.

GOLDSTEIN I., WHITE C. (1985). – « Residential segregation and color stratification among Hispanics in Philadelphia: Comment on Massey and Mullan », *American Journal of Sociology*, Vol. 91, n°2, September, p. 391-396.

HARRISON R. J., WEINBERG D. M. (1992). – « Racial and ethnic residential segregation in 1990 », working paper, U.S. Bureau of the Census, Washington, D.C.

HOLZER H. (1991). – « The spatial mismatch hypothesis: What has the evidence shown? », *Urban Studies*, Vol. 28, n°1, February, p. 105-122.

JAKUBS J. F. (1977). – « Residential segregation: The Taeuber index reconsidered », *Journal of Regional Science*, Vol. 17, n°2, August, p. 281-303.

KAIN J. F., QUIGLEY J. M. (1975). – *Housing Markets and Racial Discrimination: A Microeconomic Analysis*, National Bureau of Economic Research and Columbia University Press, New York.

LIEBERSON S., CARTER D. K. (1982). – « A model for inferring the voluntary and involuntary causes of residential discrimination », *Demography*, Vol. 19, n°4, November, p. 511-526.

MASSEY D. S. (1981). – « Social class and ethnic segregation: A reconsideration of methods and conclusions », *American Sociological Review*, Vol. 46, n°5, October, p. 641-650.

MASSEY D. S. (1985). – « Ethnic residential segregation: A theoretical synthesis and empirical review », *Sociology and Social Research*, Vol. 69, n°3, April, p. 315-350.

MASSEY D. S., DENTON N. A. (1987). – « Trends in the residential segregation of blacks, Hispanics, and Asians: 1970-1980 », *American Sociological Review*, Vol. 52, n°6, December, p. 802-825.

MASSEY D. S., DENTON N. A., (1988). – « The dimensions of residential segregation », Social Force, Vol. 67, n°2, December, p. 281-315.

MASSEY D. S., MULLAN B. P., (1984). – « Processes of Hispanic and black spatial assimilation », *American Journal of Sociology*, Vol. 89, n°4, January, p. 836-873.

MCKINNEY S., SCHNARE A. B. (1989). – « Trends in residential segregation by race: 1960-1980 », *Journal of Urban Economics*, Vol. 26, n°2, September, p. 269-280.

MUTH R. (1969). – *Cities and Housing: The Spatial Pattern of Urban Residential Land Use*, University of Chicago Press, Chicago.

OAXACA R. L. (1973). – « Male-female wage differentials in urban labor markets », *International Economic Review*, Vol. 14, p. 693-709.

ROSS S. L. (2002). – « Segregation and racial preferences: An analysis of choice based on satisfaction and outcome measures », working paper, University of Connecticut.

SCHNARE A. B. (1980). – « Trends in residential segregation by race: 1960-1970 », *Journal of Urban Economics*, Vol. 7, n°3, May, p. 293-301.

SERMONS M. W. (2000). – « Influence of race on household residential utility », Geographical Analysis, Vol. 32, n°3, July, p. 225-246.

SMITH R. A. (1989). – « The effects of local fair housing ordinances on housing segregation: Their impact is small, but it's an important positive change toward integration », *American Journal of Economics and Sociology*, Vol. 48, n°2, p. 219-230.

TAEUBER K. E., TAEUBER A. F. (1965). – *Negroes in Cities*, Chicago, Aldine.

VAN VALEY T. L., ROOF W. C., WILCOX J. E. (1977). – « Trends in residential segregation: 1960-1970 », *American Journal of Sociology*, Vol. 82, n°4, January, p. 826-844.

WHITE M. J. (1986). – « Segregation and diversity measures in population distribution », *Population Index*, Vol. 52, n°2, Summer, p. 198-221.

ZAX J. S. (1990). – « Race and commutes », *Journal of Urban Economics*, Vol. 28, p. 336-348.

ZAX J. S. (1991). – « Compensation for commutes in labor and housing markets », *Journal of Urban Economics*, Vol. 30, n°2, September, p. 192-207.

ZAX J. S. (1998). – « Immigration, race and space », in Hamermesh, Daniel S. and Frank D. Bean, eds., *Help or Hindrance? The Economic Implications of Immigration for African Americans*, Russell Sage Foundation, New York, p. 222-252.

ZAX J. S., KAIN J. F. (1996). – « Moving to the suburbs: Do relocating companies leave their black workers behind? », *Journal of Labor Economics*, Vol. 14, n°3, July, p. 472-504.

ZELDER R. E. (1970). – « Racial segregation in urban housing markets », *Journal of Regional Science*, Vol. 10, n°1, April, p. 93-105.